ECOLOGICAL AND EVOLUTIONARY GENETICS OF CHAGAS DISEASE VECTORS

by

TROY JASON KIERAN

(Under the Direction of Travis C. Glenn)

ABSTRACT

A variety of parasites and pathogens are responsible for human disease, many of which are vector-borne. Neglected Tropical Diseases (NTDs) burden the planet with more than one billion people and many are composed of complex multi-host systems. These complex biological and ecological systems present major challenges in understanding and controlling the transmission of disease. The field of molecular ecology offers a variety of genetic techniques to address issues related to ecological and evolutionary disease transmission, from species identification and diet analysis to landscape movements and speciation events. Next Generation Sequencing (NGS) technologies have advanced significantly over the past decade driving costs down, making them cost-effective for many researchers on the frontline of disease research and prevention. One particular vector-borne NTD, Chagas disease, is a parasite *Trypanosoma cruzi* transmitted by blood feeding vectors in the Triatominae subfamily between a wide range of potential mammalian hosts and humans. Despite widespread control programs, Chagas disease remains a major health threat to millions of Latin American residents. Knowledge of Triatominae microbiomes, population genetics (ecology), and phylogeny

(evolution) lags behind that of other insects and vector species. Studies involving these vectors' genetics can help us can gain further insights into the vector biology and ecology that may be applied to disease control and prevention efforts.

In this dissertation, I explore various aspects of triatomine ecology and evolution utilizing different NGS techniques. In the first study I develop a simple, cost-effective method for blood meal detection. In the second, I use Illumina 16S rRNA sequencing to examine the microbial composition of whole-bodies of *Rhodnius pallescens*, the major Chagas disease vector in Panama. The third project studies the population structure of *R. pallescens* among five populations in Panama. The last two projects utilize ultraconserved elements to test a bait set for hemipteran phylogenetics and then use them to examine the taxonomic relationships among Chagas disease vectors in the subfamily Triatominae.

INDEX WORDS: Triatominae, *Rhodnius pallescens*, *Trypanosoma cruzi*, Next-Generation Sequencing, molecular ecology, phylogenetics, microbiome, diet analysis, RADseq, ultraconserved elements

ECOLOGICAL AND EVOLUTIONARY GENETICS OF CHAGAS DISEASE VECTORS

by

TROY JASON KIERAN

BS, University of Maine, 2008

MS, Winthrop University, 2012

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2020

ECOLOGICAL AND EVOLUTIONARY GENETICS OF CHAGAS DISEASE VECTORS

by

TROY JASON KIERAN

Major Professor:	Travis C. Glenn
Committee:	Nicole L. Gottdenker
	Erin K. Lipp
	Vanessa Ezenwa

Electronic Version Approved:

Ron Walcott
Interim Dean of the Graduate School
The University of Georgia
May 2020

ACKNOWLEDGEMENTS

I would like to thank first and foremost Emily Bush for hanging in there with me for this journey. You made things a little easier and a little less stressful. Thank you, Natalia Bayona-Vasquez and Todd Pierson, for your many conversations be they scientific, contemporary, or personal. You have both been excellent friends and colleagues. Travis Glenn, thank you for putting up with me in the early years as your lab tech and giving me a chance. I have grown in experience and knowledge over the years under your advisor ship. I want to thank all my many collaborators for without whom this work would not be possible. Nicole Gottdenker, Azael Saldana, Jesse Thomas, Christina Varian, Kaylee Arnold, Jose Calzada, Christiane Weirauch, Eric Gordon, Alejandro Zaldivar-Riveron, Carlos Ibarra-Cerdena, Brant Faircloth. Your expertise, knowledge, and resources were irreplaceable.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

Introduction

Infectious diseases are estimated to cause the death of 15 million people per year with little reduction estimated in that number in the coming decades (Dye, 2014). A variety of parasites and pathogens are responsible for human disease, many of which are vector-borne, causing nearly 20% of deaths from infectious disease (European Center for Disease Prevention and Control 2014). A broad definition of vector is "any organism that functions as a carrier of an infectious agent between organisms of a different species" (Kuno & Chang, 2005) and encompasses all hematophagous (blood-feeding) arthropods, such as ticks and mosquitoes. Hematophagy has evolved independently many times over several million years in arthropods, with >14,000 species exhibiting this lifestyle (Adams, 1999; J. M. Ribeiro, 1995). Vector-borne diseases are a significant global health issue that affects the economy and human health, notably in tropical countries.

Neglected Tropical Diseases (NTDs) burden the planet with more than one billion people, greatly reducing their years of life (Hotez, Fenwick, Savioli, & Molyneux, 2009). Neglected zoonotic diseases (NZDs) are NTD's transmitted from animals to humans. Around 60-75% of emerging zoonotic diseases are composed of complex multi-host systems (Cleaveland, Laurenson, & Taylor, 2001). These complex biological and

ecological systems present major challenges in understanding and controlling the transmission of disease. As their name suggests, these diseases are often underfunded and under-researched compared to many others (Liese, Houghton, & Teplitskaya, 2014). The field of Molecular Ecology (Andrew et al., 2013) offers a variety of genetic techniques to address issues related to ecological and evolutionary disease transmission, from species identification and diet analysis to landscape movements and speciation events.

The amount of data that can be generated no longer constrains the scientific inquiry into fundamental biological questions (Jones & Good, 2016). Next Generation Sequencing (NGS) technologies have advanced significantly over the past decade, expanding their accessibility with numerous sequencing options (Ekblom & Galindo, 2011; Ellegren, 2014; Matz, 2018; McCormack, Hird, Zellmer, Carstens, & Brumfield, 2013). The continuing decrease in costs for sequencing and NGS technologies achieves further accessibility (Glenn, 2011; van Dijk, Auger, Jaszczyszyn, & Thermes, 2014), making them cost-effective for many researchers. Accessibility and affordability have expanded research into non-model organisms (Ekblom & Galindo, 2011; Ellegren, 2014; Matz, 2018; Wachi, Matsubayashi, & Maeto, 2018) that advance science as a whole, particularly in the fields of phylogenetics (McCormack et al., 2013) and molecular ecology (Andrew et al., 2013). These methodologies are promising to expand public health diagnostic research (Motro & Moran-Gilad, 2017; Radford et al., 2012) and infectious disease vector biology (Criscione, O'Brochta, & Reid, 2015; Rinker, Pitts, & Zwiebel, 2016). As these data are generated and made available to the public, additional and broader questions for system-wide application can be investigated.

Chagas disease, a zoonosis caused by the protozoan kinetoplastic parasite *Trypanosoma cruzi*, is transmitted by hematophagous bug vectors in the Triatominae subfamily between a wide range of potential mammalian hosts and humans. Despite widespread control programs, Chagas disease remains a major health threat to millions of Latin American residents, especially those living in poverty (Gurtler, Kitron, Cecere, Segura, & Cohen, 2007). Hundreds of thousands of people are migrating who suffer from Chagas or are infected with *T. cruzi* subclinically, creating a major challenge for U.S., European and other non-endemic public health systems (Bern, Kjos, Yabsley, & Montgomery, 2011; Bern & Montgomery, 2009; Gascon, Bern, & Pinazo, 2010; Gurtler et al., 2007). Chagas disease presents in two phases over the course of infection. The acute phase occurs about a week after infection and lasts for 4-8 weeks and often passes without any discernible clinical symptoms (Barrett M. P. et al., 2003). During this phase trypanosomes are easily detected in the blood, but as the disease reaches the chronic phase, parasitemia is significantly reduced. Many patients remain asymptomatic, however after about 10-25 years about 15-30% will suffer from clinical conditions related to organ damage, most often enlargement of the heart and gastrointestinal tract (Barrett M. P. et al., 2003). Currently only two drugs (nifurtimox and benzidazole) are effective against *T. cruzi* infection with earlier treatment being more successful, making early detection of infection important for treatment (Bern, 2015). Identifying and monitoring areas of high infection risk can make early detection more effective.

Knowledge of Triatominae population genetics (ecology) and phylogeny (evolution) lags behind that of other insects and vector species (Gourbiere, Dorn, Tripet, & Dumonteil, 2012). Studies involving these vectors' genetics can help with control

efforts, particularly when using methods enabled by single nucleotide polymorphisms (SNPs) and comprehensive genome comparisons (Gourbiere et al., 2012). We need to understand and manage NTD systems for the benefit of public health. By expanding our methodologies, we can gain further insights and begin to understand more basics of these systems with the hope of developing or improving rational vector / disease control strategies in the future.

Objective/Dissertation Overview

The overall objective of this dissertation is to examine ecological and evolutionary genetics of Chagas disease vectors. This is done by developing and validating NGS techniques as applied to the non-model Chagas disease vector *Rhodnius pallescens* and other Triatomines. The specific goals are:

I. Chapter 2 develops a reliable method for detecting bloodmeals that is accurate, inexpensive, easy to use, and reproducible.
II. Chapter 3 characterizes the microbiome of wild caught *Rhodnius pallescens* in Panama and observing patterns among composition and environmental variables.
III. Chapter 4 evaluates the population structure of *Rhodnius pallescens* in Panama.
IV. Chapter 5 validates a UCE bait set for Hemipteran UCEs that can be used to advance phylogenetic research in Triatominae, Chapter 6.

Below I describe the rationale and background for these Chapters.

Diet

Determining an organism's diet is a fundamental biological component for understanding interactions between species (Kartzinel et al., 2015), niche specializations (Kratina, LeCraw, Ingram, & Anholt, 2012), food webs (McCann, 2007), ecological dynamics (Poelen, Simons, & Mungall, 2014), and disease transmission (Kent, 2009) for vectors. In diet analysis, numerous methods of visual inspection, immunology, biomarkers, stable isotopes, and molecular have been employed (Nielsen et al., 2018). Many animals, however, have different diets from a range of temporal and spatial feeding requirements that make distinguishing food sources more difficult (McMeans, McCann, Humphries, Rooney, & Fisk, 2015). Consumed food, such as blood in hematophagous arthropods, breaks down and degrades due to digestion and time between feeding, impacting the rate of detection (Pinto et al., 2012; Puente, Ruiz, Soriguer, & Figuerola, 2013).

Effective diet analysis is based on high resolution (detection of more refined taxonomic levels i.e. species), the ability to sperate between different dietary items, and ideally the ability to quantify dietary components (Nielsen et al., 2018). However, in mixed diet samples, most diet analysis methods are unable to distinguish multiple sources. Unless coupled with additional cloning steps (Waleckx, Suarez, Richards, & Dorn, 2014), which are time- and cost-intensive, conventional PCR cannot identify multiple taxa. These methods depend on reference databases such as BOLD (Ratnasingham & Hebert, 2007) and Genbank (Benson et al., 2013), that must be accurate and well represented with reference taxa, which are key to obtaining data of high

resolution. These resources are currently incomplete but are expanding rapidly with molecular methods' wider use and cost reduction.

Amplicon primer choice is important for the design of the study and it is necessary to assess the benefits and limitations of each before use. There are many options, from single target species specific, to group specific, to the more universal. No primer set is truly universal, and they all have primer biases and amplification and sequencing artifacts (Clare, 2014; Deagle, Thomas, Shaffer, Trites, & Jarman, 2013). The selected primers and the potential biases will depend on the research study's objectives. For its accuracy and effectiveness, the quantification of diet mixtures is still under discussion. Many implications such as the order and the time frame items are consumed, digestive processes, amplification and sequencing bias, and taxonomically biased or deficient reference databases diminish the accuracy of quantifiable diet mixtures (Clare, 2014; Clare, Chain, Littlefair, & Cristescu, 2016; Deagle et al., 2013).

Most species of Triatominae feed on a wide range of vertebrate species and may be regarded as generalist (Galvao & Justi, 2015; Lent & Wygodzinsky, 1979). Some species, however, have a narrower diet such as preferring bird species in the genera *Psammolestes* (Salvatella, Basmadjian, Rosa, & Puime, 1992; Usinger, 1944) and preferring bats in the species *Cavernicola pilosa* (Oliveira, Ferreira, Carneiro, & Diotaiut, 2008; Usinger, 1944). Other reports have found evidence of Triatomines feeding off other arthropods (Garrouste, 2009; Kjos et al., 2013; C. M. Sandoval et al., 2010), including other kissing bugs (Claudia Magaly Sandoval et al., 2004), illustrating the diverse diet that this group has, making them an ideal study system for NGS diet analysis. Recently

NGS has been employed to distinguish between multiple blood meal sources in a mixed diet Triatomine sample (Kieran et al., 2017).

Understanding patterns of vector-host associations in Triatominae can determine host vertebrate species with significant roles in supporting populations of vectors and parasites. These vector-host associations are generally skewed towards a small subset of well-studied and documented primary vector species while others are overlooked (Carcavallo, Da Silva Rocha, & Galindez Giron, 1998). Improving methods for NGS diet analysis to make them affordable, efficient, and informative can help broaden the range of species and habitats surveyed to improve system-wide information on vectors for Chagas disease and the potential for disease transmission.

Microbiome

Diversification of insects has evolved over 400 myr within a wide range of ecological niches, resulting in various insect-microbe interactions that contribute to their varied success (Engel & Moran, 2013; Moran, McCutcheon, & Nakabachi, 2008). Interactions between insect and microbes can serve many functions including digestion, absorption of nutrients, development and reproduction, predator and pathogen defense (Oliver & Martinez, 2014). Many of these microbes are obligatory endosymbionts that provide the survival of insect hosts that evolved along with the host species with a vital function (Engel & Moran, 2013; Zindel, Gottlieb, & Aebi, 2011).

Microbial community structure and composition are important to understand the role and function they play in the ecological system. These communities were difficult to characterize in a thorough manner before the availability of culture-independent methods

(Nelson, Morrison, Benjamino, Grim, & Graf, 2014; Oliver & Martinez, 2014). With the advent of NGS, the breadth and importance of microbial community composition, or microbiomes, has greatly expanded our understanding of their role in many systems (Finney, Kamhawi, & Wasmuth, 2015; Nelson et al., 2014). However, relatively few invertebrate microbiomes have been characterized or understood while long been seen as a means of vector biocontrol, including in Triatomines, (Beard, 2002; Saldana, Hegde, & Hughes, 2017; Weiss & Aksoy, 2011) impacting research in vector-borne disease.

High-throughput and NGS technologies have expanded our microbiome knowledge of certain vector species, primarily mosquitoes, showing that the microbiome is often composed of relatively few taxa, high variability, and variation is influenced by factors such as the stage of host life, host sex, sampling methodology and the environment (Boissiere et al., 2012; Coon, Brown, & Strand, 2016; Coon, Vogel, Brown, & Strand, 2014; Gimonneau et al., 2014; Osei-Poku, Mbogo, Palmer, & Jiggins, 2012). This variability in a single group of vectors makes vector-microbe research difficult and specific to species or system, limiting the wide applicability of findings and requiring broader sampling of vector species and systems for further research.

16S rRNA amplicon sequencing is still the gold standard for bacteria-focused microbial ecology studies (Chakraborty, Doss, Patra, & Bandyopadhyay, 2014), providing a fast, relatively inexpensive means of microbial detection and compositional analysis. Other barcoding genes are often used for other microbial taxa, such as ITS and LSU for fungi, COI and 18S for protozoa, COI and rbcl for algae, and there is no current standard marker for viruses (Chakraborty et al., 2014). However, 16S PCR, has many known biases and limitations like any set of primers (Acinas, Sarma-Rupavtarm, Klepac-

Ceraj, & Polz, 2005; Berry, Ben Mahfoudh, Wagner, & Loy, 2011; Kennedy, Hall, Lynch, Moreno-Hagelsieb, & Neufeld, 2014; Suzuki & Giovannoni, 1996; Tremblay et al., 2015) and reference databases often have similar problems (Edgar 2018) (Edgar, 2018). Although amplicon sequencing studies of 16S rDNA may provide a broad picture of microbial communities, at lower taxonomic levels they may present limited resolution and sensitivity (Poretsky, Rodriguez, Luo, Tsementzi, & Konstantinidis, 2014).

Metagenomic sequencing methods, where total genomic DNA is extracted from an environmental sample, may include whole microbial communities that gets away from the limitations of 16S amplicons, are not without their own limitations. Genomic DNA extracts may contain non-target DNA from sources other than microbiota of interest, which may affect downstream analyzes and lead to misassembly of sequence contigs, spurious reads, and therefore skewed conclusions (Koutsovoulos et al., 2016; Schmieder & Edwards, 2011)and filtering of metagenomic contaminant DNA may be particularly problematic. Compared to 16S rRNA amplicon sequencing, metagenomics is also much more expensive for library preparation and sequencing, and more labor intensive in the laboratory and bioinformatically. Amplicon gene studies based on Illumina have become increasingly common due to significant cost advantages, higher performance, scalability, and fewer sequencing errors (Glenn, 2011; Logares et al., 2014). In addition to user-friendly advances in bioinformatics tools such as Mothur (Schloss et al., 2009) and Qiime (Bolyen et al., 2018; Caporaso et al., 2010), the number of researchers involved in 16S microbiome analysis has increased.

Few studies have examined the microbiomes of Triatominae using culture independent methods, and only 13 of the more than 150 species have been examined,

including *Dipetalogaster maximus*, *Panstrongylus megitus*, *Rhodnius neglectus*, *R. pallescens*, *R. prolixus*, *Triatoma brasiliensis*, *T. dimidiata*, *T. infestans*, *T. Juazeirensis*, *T. sherlocki*, *T. maculata*, *T. pseudomaculata*, and *T. vitticeps* (da Mota et al., 2012; Diaz, Villavicencio, Correia, Costa, & Haag, 2016; Dumonteil et al., 2018; Gumiel et al., 2015; Montoya-Porras, Omar, Alzate, Moreno-Herrera, & Cadavid-Restrepo, 2018; Vieira et al., 2015). In addition, these studies relied heavily on insects reared in colonies and used a small number of individuals per species. So far, though, a few trends have emerged. Bacillales, Actinomycetales, Enterobacteriales and Burkholderiales (da Mota et al., 2012; Diaz et al., 2016; Dumonteil et al., 2018; Gumiel et al., 2015; Montoya-Porras et al., 2018; Vieira et al., 2015) are commonly observed bacterial orders in Triatominae. Triatominae carry many secondary symbiotic bacterial, but the two primary symbionts are *Wolbachia sp*. and *Arsenophonus sp*. (Jimenez-Cortes et al., 2018).

Understanding microbiome patterns across different variables (i.e. location, habitat, infection status, etc) is important for acquiring biological knowledge that can be used as a means of vector disease control. Chapter 3 expands knowledge in this area by utilizing a larger sample size of three populations of wild *R. pallescens* and compares the composition of microbiomes across various biological and environmental variables. As research in this area expands with more systems, species, and circumstances being examined, it is possible to further explore effective means of biocontrol options.

Population Genetics

Dispersal is an important characteristic of species evolution, diversification and proliferation, but remains one of the most misunderstood concepts in ecology and

evolutionary biology (Edelaar & Bolnick, 2012; Kokko & Lopez-Sepulcre, 2006). Dispersal is crucial for pathogens because it determines the rate of disease spread (Brown & Hovmoller, 2002; Viboud et al., 2006) and the potential for evolution (Gandon, Capowiez, Dubois, Michalakis, & Olivieri, 1996; Morgan, Gandon, & Buckling, 2005), with significant implications for public health, agriculture and biodiversity (Daszak, Cunningham, & Hyatt, 2000; Fisher et al., 2012; Pennisi, 2010). Anthropogenic changes in landscape are recognized as drivers of infectious disease prevalence and a threat to public health (Biek & Real, 2010; Meentemeyer, Haas, & Vaclavik, 2012). However, we lack fundamental knowledge on how the spread of disease can be affected by genotype, space, and environment.

Environmental changes lead to vector-borne disease research to investigate the spatial and temporal transmission dynamics of host, vector and pathogen populations (Reisen, 2010). Landscape changes can affect the transmission of disease by changing patterns of contact between hosts and vectors (Patz et al., 2004), increasing the selection pressure on a particular pathogen or vector to a new modified landscape (Reisen, 2010). While an important issue, quantitative methods and data collection are still lacking spatially and temporarily to understand key components in pathogen-vector responses to landscape changes (Brearley et al., 2013), particularly in tropical regions where changes are accelerating (Barretto, Berndes, Sparovek, & Wirsenius, 2013). Quantifying the dispersal of host and vector populations has proved useful in developing effective pathogen control strategies (Cecere, Vasquez-Prokopec, Gurtler, & Kitron, 2006; Tabachnick & Black, 1995).

Gene flow in parasites in believed to be dependent on the host with the most mobility (Blouin, Yowell, Courtney, & Dame, 1995). However, the host with the highest dispersal rate is thought to drive parasite gene flow in complex, multi-host species systems (Prugnolle et al., 2005). Several studies examined the population structures of the parasite and one of the hosts, and the relationship between the two structures was inconsistent (Brouat et al., 2011; Bruyndonckx, Biollaz, Dubey, Goudet, & Christe, 2010; Dharmarajan et al., 2016; Dybdahl & Lively, 1996; Geist & Kuehn, 2008; McCoy, Boulinier, & Tirard, 2005; Pennings, Achenbach, & Foitzik, 2011; Prugnolle et al., 2008; van Schaik, Kerth, Bruyndonckx, & Christe, 2014). However, given the implications of multi-host systems, the issue has been addressed directly by few studies (Witsenburg et al., 2015). A meta-analysis has shown that parasite dispersal rates actually depend on their hosts' dispersal. However, interpretations of host-parasite genetic co-structure and therefore relative dispersal rates are not so straightforward with other potential evolutionary and ecological forces also involved as drivers of the spatial distribution of genetic diversity in hosts and parasites (Maze-Guilmo, Blanchet, McCoy, & Loot, 2016). In a system with hundreds of potential reservoir species, such as Chagas, this is likely to be more apparent. However, we have a reasonable starting point for understanding Chagas parasite dispersal and gene flow, initial research with a commonly abundant species that is frequently fed on and infected with *T. cruzi*.

Genetic studies on the Triatominae (Hemiptera: Reduviidae) are limited in number and sporadic on the species covered. Random Amplification of Polymorphic DNA (RAPD) PCR (Garcia et al., 1998) and SSCP (Stothard, Frame, & Miles, 1999) were the first molecular markers used. Some researchers have carried out studies using

microsatellite markers to examine the phylogeography at larger spatial scales of different triatomines (Breniere et al., 2012; Perez de Rosas, Segura, Fichera, & Garcia, 2008; Perez de Rosas, Segura, & Garcia, 2011). However, this provides only large-scale identification of meta-populations without any local knowledge. Microsatellites also have disadvantages, including a limited number of genetic markers and limited spatial resolution over varying sample area degrees (Zink, 2010). While others relied on mitochondrial markers of DNA, which often have limited population-level variation. To date, only two studies have used RADseq approaches in Triatominae vectors (Hernandez-Castro et al., 2017; Orantes et al., 2018), leaving this particular system and methodologies open to research.

Previous studies of triatomine population genetics found a variety of population structures including limited dispersal potential (Ramirez et al., 2005), range expansion (Piccinali et al., 2009), suggestive of dispersal blockers in a highly local, urban environment (Khatchikian et al., 2015), local structure with recolonization events (Stevens et al., 2015), and no structure over larger area (Hernandez-Castro et al., 2017). Research in this area has employed different markers and methods, different geographic scales, different species, and habitat diversity. All suggest that the structure that we may find in each individual system is unique. Research on vector population genetics in important for identifying cryptic species, diversity and local adaptations, gene flow and colonization potential, and population structure and transmission (McCoy, 2008). Therefore, in triatomine population genetics, there is a need for more systemic research and standardized methods. Only one such study has been conducted in Columbia and Panama on the population structure of *Rhodnius pallescens* (Gomez-Palacio et al., 2012).

For non-model vector species and systems, the era of NGS has provided an expanded and rapid accumulation of genomic data (Rinker et al., 2016). Combined with advances in computer disease modeling, NGS can result in more accurate risk of disease transmission (Chiyaka et al., 2013). One NGS method, Restriction site Associated DNA Sequencing (RADseq) is one of the most important scientific breakthroughs in the last 10-15 years (Andrews, Good, Miller, Luikart, & Hohenlohe, 2016). RADseq is a reduced-representation approach that uses enzymes to digest genomic DNA and bind to the cut-site sequence-based adapters. This enables researchers to find hundreds to thousands of polymorphic loci across an easy, cost-effective, and reproducible genome. By reducing the sequenced sub-set of the genome, we can achieve greater genotype sequencing depth and confidence at each locus, while reducing costs per sample, allowing more sequencing of samples. RADseq approaches have the ability to provide efficient, versatile, low-cost research into ecological and evolutionary questions. Chapter 4 of this dissertation uses a RADseq approach to expand research on the population structure of *Rhodnius pallescens* in Panama.

Phylogentics

Comprehension of classification (taxonomy) and evolutionary relationships (phylogeny) between organisms is a key component of understanding biology. Clear differences between species, morphological characteristics, distributions, and ecology all play a role in determining insect vector capacity. Well-defined knowledge of these relationships will help inform strategies for vector control by allowing for clear taxonomic distinctions, aid in insecticide resistance and re-infestation potential, and

understanding infection cycles (Gourbiere et al., 2012). Using morphological features alone can lead to invalid taxon, as these features can arise independently due to similar niches. Triatominae are a group plagued by complex taxonomic classifications that are often contradictory. Various levels of taxonomic groupings such as sub-family, tribe, species complex, and genus, contribute to a messy understanding of evolutionary relationships in this group of disease vectors.

Evolutionary phylogenetic research in Triatominae, similar to population genetics, has lagged behind those of other insects and vector species (Gourbiere et al., 2012) and could use an update. Several studies on Triatominae phylogeny have been conducted recently (Acosta et al., 2013; Hwang & Weirauch, 2012; Ibarra-Cerdena, Zaldivar-Riveron, Peterson, Sanchez-Cordero, & Ramsey, 2014; Justi, Galvao, & Schrago, 2016; Justi, Russo, Mallet, Obara, & Galvao, 2014; Weirauch & Munro, 2009; Zhang et al., 2016), but no consensus has been reached as to whether all Triatominae classes are mono-, poly-, or paraphyletic (Gourbiere et al., 2012; Otálora-Luna, Pérez-Sánchez, Sandoval, & Aldana, 2015). Even the most recent phylogenetic studies in this group show Triatominae as monophyletic (Justi et al., 2016) and paraphyletic (Hwang & Weirauch, 2012; Zhang et al., 2016).

Triatominae, order Hemiptera, is a subfamily of the tropical predatory insect family Reduviidae, commonly referred to as assassin bugs. There are approximately 6,800 species in 25 Reduviidae subfamilies (Hwang & Weirauch, 2012). Triatominae consists of 149 species described in 15 genera and 5 tribes (Justi et al., 2016). They are further divided into 11 or 8 complexes of species (Lent & Wygodzinsky, 1979; Schofield & Galvão, 2009) and 8 sub-complexes (Schofield & Galvão, 2009). Especially with

respect to complex species, there is a lot of confusion. A' species' as traditionally defined by the Biological Species Concept is a group of actual or potential individuals interbreeding with gene flow barriers (Mayr, 1963). In Triatominae systematics, species complexes are common and are generally used to refer to groups with morphological similarities (Lent & Wygodzinsky, 1979; Schofield & Galvão, 2009). These groupings, however, often do not collaborate with molecular evidence. Due to the reliance on morphology (Lent & Wygodzinsky, 1979), the focus on "epidemiological types" (Gourbiere et al., 2012), and the use of limited molecular markers to organize species in this group, these phylogenetic issues are not resolved.

Hematophagy is a major defining characteristic of triatomes. Many Hemiptera species are plant pests with piercing-sucking parts of the mouth for plant extraction of sap and phloem. The Reduviidae subfamily evolved from plants to feed on insect hemolymph, a feature also found in Triatominae (Lent & Wygodzinsky, 1979; C. M. Sandoval et al., 2010). Triatominae evolved further to feed on vertebrate blood, with changes in the structure of the mouth (Cobben, 1978; Weirauch, 2008), salivary glands (J. M. C. Ribeiro, Assumpção, & Francischetti, 2012) and genetic traits (Mesquita et al., 2015). Recent estimates put Triatominae's appearance at 20-40 MYA (Hwang & Weirauch, 2012; Ibarra-Cerdena et al., 2014; Justi et al., 2016), coinciding with South America's habitat diversification and mammal and bird radiation (Justi & Galvao, 2017). There is some discussion, however, as to whether this group once or twice evolved the trait over evolutionary history (Otálora-Luna et al., 2015). Taxonomic clarification is an important key step for surveillance and vector control (Gourbiere et al., 2012) due to the differences in vector vs. non-vector species within this group.

Previous studies of Triatominae phylogenetics primarily used a handful of markers for the analysis of mitochondrial (i.e. COI, CytB, 16S, 12S) and/or nuclear genes (i.e. ITS, 18S, 28S) (Justi & Galvao, 2017)). Although the standard may still be these techniques, they often lack resolution. Previously, the most likely limitation in this area of research was limited genetic resources. Limited genetic resources will become less common with decreasing costs and expanded access, and improved phylogenetic studies should follow. Anchored hybrid enrichment (Lemmon, Emme, & Lemmon, 2012) and ultraconserved elements (UCEs) (Faircloth et al., 2012) are capable of resolving phylogenies at varying taxonomic levels with a greater number of orthologous genetic markers (increased resolution) and relatively low costs.

Ultraconserved Elements are highly conserved DNA regions across divergent taxa (Bejerano et al., 2004). This makes them ideal for target enrichment and proved useful for arthropod phylogenetic studies (Baca, Alexander, Gustafson, & Short, 2017; Faircloth, Branstetter, White, & Brady, 2015; Starrett et al., 2017). A recent UCE bait set has been designed for Hemiptera phylogenetics (Faircloth, 2017). We test and validate this test set on various Hemiptera taxa, including Triatominae, in Chapter 5 (Kieran et al., 2019). These UCE baits are being used by ongoing research (Chapter 6) to elicit relationships within Triatominae.

References

Acinas, S. G., Sarma-Rupavtarm, R., Klepac-Ceraj, V., & Polz, M. F. (2005). PCR-induced sequence artifacts and bias: insights from comparison of two 16S rRNA clone libraries constructed from the same sample. *Appl Environ Microbiol, 71*(12), 8966-8969. doi:10.1128/AEM.71.12.8966-8969.2005

Acosta, I. D. L., da Costa, A. P., Nunes, P. H., Gondim, M. F. N., Gatti, A., Rossi , J. L., . . . Marcili, A. (2013). Morphological and molecular characterization and phylogenetic relationships of a new species of trypanosome in *Tapirus terrestris* (lowland tapir), *Trypanosoma terrestris* sp nov., from Atlantic Rainforest of southeastern Brazil. *Parasit Vectors, 6*. doi:10.1186/1756-3305-6-349

Adams, T. S. (1999). Hematophagy and hormone release. *Annals of the Entomological Society of America, 92*(1), 1-13.

Andrew, R. L., Bernatchez, L., Bonin, A., Buerkle, C. A., Carstens, B. C., Emerson, B. C., . . . Rieseberg, L. H. (2013). A road map for molecular ecology. *Mol Ecol, 22*(10), 2605-2626. doi:10.1111/mec.12319

Andrews, K. R., Good, J. M., Miller, M. R., Luikart, G., & Hohenlohe, P. A. (2016). Harnessing the power of RADseq for ecological and evolutionary genomics. *Nat Rev Genet, 17*(2), 81-92. doi:10.1038/nrg.2015.28

Baca, S. M., Alexander, A., Gustafson, G. T., & Short, A. E. Z. (2017). Ultraconserved elements show utility in phylogenetic inference of Adephaga (Coleoptera) and

suggest paraphyly of ‘Hydradephaga’. *SYSTEMATIC ENTOMOLOGY, 42*(4), 786-795. doi:10.1111/syen.12244

Barrett M. P., Bruchmore R. J. S., Stich A., Lazzari J. O., Frash A. C., Cazzulo J. J., & S., K. (2003). The trypanosomiasis. *Lancet, 362*, 1469–1480.

Barretto, A. G. O. P., Berndes, G., Sparovek, G., & Wirsenius, S. (2013). Agricultural intensification in Brazil and its effects on land use patterns: An analysis of the 1975-2006 period. *Glob Change Biol, 19*, 1804-1815.

Beard, C. B. (2002). Bacterial symbionts of the Triatominae and their potential use in control of Chags disease transmission. *Annu Rev Entomol, 47*.

Bejerano, G., Pheasant, M., Makunin, I., Stephen, S., Kent, W. j., Mattick, J. S., & Haussler, D. (2004). Ultraconserved Elements in the Human Genome. *SCIENCE, 304*, 1321-1325.

Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., & Sayers, E. W. (2013). GenBank. *Nucleic Acids Research, 41*(D1), D36-D42. doi:10.1093/nar/gks1195

Bern, C. (2015). Chagas' Disease. *N Engl J Med, 373*(5), 456-466. doi:10.1056/NEJMra1410150

Bern, C., Kjos, S., Yabsley, M. J., & Montgomery, S. P. (2011). Trypanosoma cruzi and Chagas' Disease in the United States. *Clin Microbiol Rev, 24*(4), 655-681. doi:10.1128/CMR.00005-11

Bern, C., & Montgomery, S. P. (2009). An estimate of the burden of Chagas disease in the United States. *Clin Infect Dis, 49*(5), e52-54. doi:10.1086/605091

Berry, D., Ben Mahfoudh, K., Wagner, M., & Loy, A. (2011). Barcoded primers used in multiplex amplicon pyrosequencing bias amplification. *Appl Environ Microbiol, 77*(21), 7846-7849. doi:10.1128/AEM.05220-11

Biek, R., & Real, L. A. (2010). The landscape genetics of infectious disease emergence and spread. *Molecular Ecology, 19*, 3515-3531.

Blouin, M. S., Yowell, C. A., Courtney, C. H., & Dame, J. B. (1995). Host movement and the genetic structure of populations of parasitic nematodes. *Genetics, 141*(3), 1007-1014.

Boissiere, A., Tchioffo, M. T., Bachar, D., Abate, L., Marie, A., Nsango, S. E., . . . Morlais, I. (2012). Midgut microbiota of the malaria mosquito vector *Anopheles gambiae* and interactions with *Plasmodium falciparum* infection. *PLoS Pathog, 8*(5), e1002742. doi:10.1371/journal.ppat.1002742

Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C., Al-Ghalith, G. A., . . . Caporaso, J. G. (2018). Qiime2: Reproducible, interactive, scalable, and extensible microbiome data science. *PeerJ Preprints*. doi:10.7287/peerj.preprints.27295v2

Brearley, G., Rhodes, J., Bradley, A., Baxter, G., Seabrook, L., Lunney, D., . . . McAlpine, C. (2013). Wildlife disease prevalence in human-modified landscapes. *Biol Rev Camb Philos Soc, 88*(2), 427-442. doi:10.1111/brv.12009

Breniere, S. F., Waleckx, E., Magallon-Gastelum, E., Bosseno, M. F., Hardy, X., Ndo, C., . . . Kengne, P. (2012). Population genetic structure of *Meccus longipennis* (Hemiptera, Reduviidae, Triatominae), vector of Chagas disease in West Mexico. *Infect Genet Evol, 12*(2), 254-262. doi:10.1016/j.meegid.2011.11.003

Brouat, C., Tatard, C., Machin, A., Kane, M., Diouf, M., Ba, K., & Duplantier, J. M. (2011). Comparative population genetics of a parasitic nematode and its host community: the trichostrongylid Neoheligmonella granjoni and Mastomys rodents in southeastern Senegal. *Int J Parasitol, 41*(12), 1301-1309. doi:10.1016/j.ijpara.2011.07.014

Brown, J. K., & Hovmoller, M. S. (2002). Aerial dispersal of pathogens on the global and continental scales and its impact on plant disease. *SCIENCE, 297*(5581), 537-541. doi:10.1126/science.1072678

Bruyndonckx, N., Biollaz, F., Dubey, S., Goudet, J., & Christe, P. (2010). Mites as biological tags of their hosts. *Mol Ecol, 19*, 2270-2778.

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., . . . Knight, R. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat Methods, 7*(5), 335-336. doi:10.1038/nmeth.f.303

Carcavallo, R. U., Da Silva Rocha, D., & Galindez Giron, I. (1998). Feeding sources and patterns. In R. U. Carcavallo, I. Galindez Giron, & J. Jurberg (Eds.), *Atlas of Chagas' disease vectors in the Americas.* (Vol. 2, pp. 537-560). Rio de Janeiro: Editora Fiocruz.

Cecere, M. C., Vasquez-Prokopec, G. M., Gurtler, R. E., & Kitron, U. (2006). Reinfestation sources for Chagas disease vector, *Triatoma infestans*, Argentina. *Emerg Infect Dis, 12*(7), 1096-1102. doi:10.3201/eid1207.051445

Chakraborty, C., Doss, C. G., Patra, B. C., & Bandyopadhyay, S. (2014). DNA barcoding to map the microbial communities: current advances and future directions. *Appl Microbiol Biotechnol, 98*(8), 3425-3436. doi:10.1007/s00253-014-5550-9

Chiyaka, C., Tatem, A. J., Cohen, J. M., Gething, P. W., Johnston, G., Gosling, R., . . . Smith, D. L. (2013). Infectious disease. The stability of malaria elimination. *SCIENCE, 339*(6122), 909-910. doi:10.1126/science.1229509

Clare, E. L. (2014). Molecular detection of trophic interactions: Emerging trends, distinct advantages, significant considerations and conservation applications. *Ecological Applications, 7*, 1144-1157.

Clare, E. L., Chain, F. J., Littlefair, J. E., & Cristescu, M. E. (2016). The effects of parameter choice on defining molecular operational taxonomic units and resulting ecological analyses of metabarcoding data. *Genome, 59*(11), 981-990. doi:10.1139/gen-2015-0184

Cleaveland, S., Laurenson, M. K., & Taylor, L. H. (2001). Diseases of humans and their domestic mammals: pathogen characteristics, host range and the risk of emergence. *PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY OF LONDON SERIES B-BIOLOGICAL SCIENCES, 356*(1411), 991-999.

Cobben, R. H. (1978). Evolutionary trends in Heteroptera, Part II, Mouth-part structures and feeding stratergies. *Meded. Landbouwhogeschool Wageningen, 78*, 1-407.

Coon, K. L., Brown, M. R., & Strand, M. R. (2016). Mosquitoes host communities of bacteria that are essential for development but vary greatly between local habitats. *Molecular Ecology, 25*(22), 5806-5826. doi:10.1111/mec.13877

Coon, K. L., Vogel, K. J., Brown, M. R., & Strand, M. R. (2014). Mosquitoes rely on their gut microbiota for development. *Mol Ecol, 23*(11), 2727-2739. doi:10.1111/mec.12771

Criscione, F., O'Brochta, D. A., & Reid, W. (2015). Genetic technologies for disease vectors. *Current Opinion in Insect Science, 10*, 90-97. doi:10.1016/j.cois.2015.04.012

da Mota, F. F., Marinho, L. P., Moreira, C. J., Lima, M. M., Mello, C. B., Garcia, E. S., . . . Azambuja, P. (2012). Cultivation-independent methods reveal differences among bacterial gut microbiota in triatomine vectors of Chagas disease. *PLoS Negl Trop Dis, 6*(5), e1631. doi:10.1371/journal.pntd.0001631

Daszak, P., Cunningham, A. A., & Hyatt, A. D. (2000). Emerging infectious diseases of wildlife--threats to biodiversity and human health. *SCIENCE, 287*(5452), 443-449.

Deagle, B. E., Thomas, A. C., Shaffer, A. K., Trites, A. W., & Jarman, S. N. (2013). Quantifying sequence proportions in a DNA-based diet study using Ion Torrent amplicon sequencing: which counts count? *Molecular Ecology Resources, 13*(4), 620-633. doi:10.1111/1755-0998.12103

Dharmarajan, G., Beasley, J. C., Beatty, W. S., Olson, Z. H., Fike, J. A., Rhodes, O. E., & Park, A. (2016). Genetic co-structuring in host-parasite systems: Empirical data from raccoons and raccoon ticks. *Ecosphere, 7*(3). doi:10.1002/ecs2.1269

Diaz, S., Villavicencio, B., Correia, N., Costa, J., & Haag, K. L. (2016). Triatomine bugs, their microbiota and *Trypanosoma cruzi*: asymmetric responses of bacteria to an infected blood meal. *Parasit Vectors, 9*(1), 636. doi:10.1186/s13071-016-1926-2

Dumonteil, E., Ramirez-Sierra, M. J., Perez-Carrillo, S., Teh-Poot, C., Herrera, C., Gourbiere, S., & Waleckx, E. (2018). Detailed ecological associations of triatomines revealed by metabarcoding and next-generation sequencing:

implications for triatomine behavior and *Trypanosoma cruzi* transmission cycles. *Sci Rep, 8*(1), 4140. doi:10.1038/s41598-018-22455-x

Dybdahl, M. F., & Lively, C. M. (1996). The Geography of Coevolution: Comparative Population Structures for a Snail and Its Trematode Parasite. *EVOLUTION, 50*(6), 2264-2275. doi:10.1111/j.1558-5646.1996.tb03615.x

Dye, C. (2014). After 2015: infectious diseases in a new era of health and development. *Philos Trans R Soc Lond B Biol Sci, 369*(1645), 20130426. doi:10.1098/rstb.2013.0426

Edelaar, P., & Bolnick, D. I. (2012). Non-random gene flow: an underappreciated force in evolution and ecology. *Trends Ecol Evol, 27*(12), 659-665. doi:10.1016/j.tree.2012.07.009

Edgar, R. (2018). Taxonomy annotation and guide tree errors in 16S rRNA databases. *PeerJ, 6*, e5030. doi:10.7717/peerj.5030

Ekblom, R., & Galindo, J. (2011). Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity (Edinb), 107*(1), 1-15. doi:10.1038/hdy.2010.152

Ellegren, H. (2014). Genome sequencing and population genomics in non-model organisms. *Trends Ecol Evol, 29*, 51-63.

Engel, P., & Moran, N. A. (2013). The gut microbiota of insects - diversity in structure and function. *FEMS Microbiol Rev, 37*(5), 699-735. doi:10.1111/1574-6976.12025

Faircloth, B. C. (2017). Identifying conserved genomic elements and designing universal bait sets to enrich them. *Methods in Ecology and Evolution, 8*(9), 1103-1112. doi:10.1111/2041-210x.12754

Faircloth, B. C., Branstetter, M. G., White, N. D., & Brady, S. G. (2015). Target enrichment of ultraconserved elements from arthropods provides a genomic perspective on relationships among Hymenoptera. *Mol Ecol Resour, 15*(3), 489-501. doi:10.1111/1755-0998.12328

Faircloth, B. C., McCormack, J. E., Crawford, N. G., Harvey, M. G., Brumfield, R. T., & Glenn, T. C. (2012). Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. *Syst Biol, 61*(5), 717-726. doi:10.1093/sysbio/sys004

Finney, C. A., Kamhawi, S., & Wasmuth, J. D. (2015). Does the arthropod microbiota impact the establishment of vector-borne diseases in mammalian hosts? *PLoS Pathog, 11*(4), e1004646. doi:10.1371/journal.ppat.1004646

Fisher, M. C., Henk, D. A., Briggs, C. J., Brownstein, J. S., Madoff, L. C., McCraw, S. L., & Gurr, S. J. (2012). Emerging fungal threats to animal, plant and ecosystem health. *Nature, 484*(7393), 186-194. doi:10.1038/nature10947

Galvao, C., & Justi, S. A. (2015). An overview on the ecology of Triatominae (Hemiptera:Reduviidae). *Acta Trop, 151*, 116-125. doi:10.1016/j.actatropica.2015.06.006

Gandon, S., Capowiez, Y., Dubois, Y., Michalakis, Y., & Olivieri, I. (1996). Local adaptation and gene-for-gene coevolution in a metapopulation model.

PROCEEDINGS OF THE ROYAL SOCIETY B-BIOLOGICAL SCIENCES, 263, 1003-1009.

Garcia, A. L., Carrasco, H. J., Schofield, C. J., Stothard, J. R., Frame, I. A., Valente, S. A. S., & Miles, M. A. (1998). Random amplification of polymorphic DNA as a tool for taxonomic studies of triatomine bugs (Hemiptera : Reduviidae). *J Med Entomol, 35*(1), 38-45. doi:DOI 10.1093/jmedent/35.1.38

Garrouste, R. (2009). La première observation in natura de l'entomophagie de *Panstrongylus geniculatus* (Latreille 1811) hématophage vecteur de la maladie de Chagas (Hemiptera: Reduviidae). *Annales De La Societe Entomologique De France, 45*, 302-304.

Gascon, J., Bern, C., & Pinazo, M. J. (2010). Chagas disease in Spain, the United States and other non-endemic countries. *Acta Trop, 115*(1-2), 22-27. doi:10.1016/j.actatropica.2009.07.019

Geist, J., & Kuehn, R. (2008). Host-parasite interactions in oligotrophic stream ecosystems: the roles of life-history strategy and ecological niche. *Mol Ecol, 17*(4), 997-1008. doi:10.1111/j.1365-294X.2007.03636.x

Gimonneau, G., Tchioffo, M. T., Abate, L., Boissiere, A., Awono-Ambene, P. H., Nsango, S. E., . . . Morlais, I. (2014). Composition of *Anopheles coluzzii* and *Anopheles gambiae* microbiota from larval to adult stages. *Infect Genet Evol, 28*, 715-724. doi:10.1016/j.meegid.2014.09.029

Glenn, T. C. (2011). Field guide to next-generation DNA sequencers. *Mol Ecol Resour, 11*(5), 759-769. doi:10.1111/j.1755-0998.2011.03024.x

Gomez-Palacio, A., Jaramillo, O. N., Caro-Riano, H., Diaz, S., Monteiro, F. A., Perez, R., . . . Triana, O. (2012). Morphometric and molecular evidence of intraspecific biogeographical differentiation of *Rhodnius pallescens* (HEMIPTERA: REDUVIIDAE: RHODNIINI) from Colombia and Panama. *Infect Genet Evol, 12*(8), 1975-1983. doi:10.1016/j.meegid.2012.04.003

Gourbiere, S., Dorn, P., Tripet, F., & Dumonteil, E. (2012). Genetics and evolution of triatomines: from phylogeny to vector control. *Heredity (Edinb), 108*(3), 190-202. doi:10.1038/hdy.2011.71

Gumiel, M., da Mota, F. F., Rizzo Vde, S., Sarquis, O., de Castro, D. P., Lima, M. M., . . . Azambuja, P. (2015). Characterization of the microbiota in the guts of *Triatoma brasiliensis* and *Triatoma pseudomaculata* infected by *Trypanosoma cruzi* in natural conditions using culture independent methods. *Parasit Vectors, 8*, 245. doi:10.1186/s13071-015-0836-z

Gurtler, R. E., Kitron, U., Cecere, M. C., Segura, E. L., & Cohen, J. E. (2007). Sustainable vector control and management of Chagas disease in the Gran Chaco, Argentina. *Proc Natl Acad Sci U S A, 104*(41), 16194-16199.

Hernandez-Castro, L. E., Paterno, M., Villacis, A. G., Andersson, B., Costales, J. A., De Noia, M., . . . Llewellyn, M. S. (2017). 2b-RAD genotyping for population genomic studies of Chagas disease vectors: *Rhodnius ecuadoriensis* in Ecuador. *PLoS Negl Trop Dis, 11*(7), e0005710. doi:10.1371/journal.pntd.0005710

Hotez, P. J., Fenwick, A., Savioli, L., & Molyneux, D. H. (2009). Rescuing the bottom billion through control of neglected tropical diseases. *Lancet, 373*(9674), 1570-1575. doi:Doi 10.1016/S0140-6736(09)60233-6

Hwang, W. S., & Weirauch, C. (2012). Evolutionary history of assassin bugs (insecta: hemiptera: Reduviidae): insights from divergence dating and ancestral state reconstruction. *PLoS One, 7*(9), e45523. doi:10.1371/journal.pone.0045523

Ibarra-Cerdena, C. N., Zaldivar-Riveron, A., Peterson, A. T., Sanchez-Cordero, V., & Ramsey, J. M. (2014). Phylogeny and niche conservatism in North and Central American triatomine bugs (Hemiptera: Reduviidae: Triatominae), vectors of Chagas' disease. *PLoS Negl Trop Dis, 8*(10), e3266. doi:10.1371/journal.pntd.0003266

Jimenez-Cortes, J. G., Garcia-Contreras, R., Bucio-Torres, M. I., Cabrera-Bravo, M., Cordoba-Aguilar, A., Benelli, G., & Salazar-Schettino, P. M. (2018). Bacterial symbionts in human blood-feeding arthropods: Patterns, general mechanisms and effects of global ecological changes. *Acta Trop, 186*, 69-101. doi:10.1016/j.actatropica.2018.07.005

Jones, M. R., & Good, J. M. (2016). Targeted capture in evolutionary and ecological genomics. *Mol Ecol, 25*(1), 185-202. doi:10.1111/mec.13304

Justi, S. A., & Galvao, C. (2017). The Evolutionary Origin of Diversity in Chagas Disease Vectors. *Trends Parasitol, 33*(1), 42-52. doi:10.1016/j.pt.2016.11.002

Justi, S. A., Galvao, C., & Schrago, C. G. (2016). Geological Changes of the Americas and their Influence on the Diversification of the Neotropical Kissing Bugs (Hemiptera: Reduviidae: Triatominae). *PLoS Negl Trop Dis, 10*(4), e0004527. doi:10.1371/journal.pntd.0004527

Justi, S. A., Russo, C. A., Mallet, J. R., Obara, M. T., & Galvao, C. (2014). Molecular phylogeny of Triatomini (Hemiptera: Reduviidae: Triatominae). *Parasit Vectors, 7*, 149. doi:10.1186/1756-3305-7-149

Kartzinel, T. R., Chen, P. A., Coverdale, T. C., Erickson, D. L., Kress, W. J., Kuzmina, M. L., . . . Pringle, R. M. (2015). DNA metabarcoding illuminates dietary niche partitioning by African large herbivores. *Proceedings of the National Academy of Sciences of the United States of America, 112*, 8019-8024.

Kennedy, K., Hall, M. W., Lynch, M. D., Moreno-Hagelsieb, G., & Neufeld, J. D. (2014). Evaluating bias of illumina-based bacterial 16S rRNA gene profiles. *Appl Environ Microbiol, 80*(18), 5717-5722. doi:10.1128/AEM.01451-14

Kent, R. J. (2009). Molecular methods for arthropod bloodmeal identification and applications to ecological and vector-borne disease studies. *Mol Ecol Resour, 9*(1), 4-18. doi:10.1111/j.1755-0998.2008.02469.x

Khatchikian, C. E., Foley, E. A., Barbu, C. M., Hwang, J., Ancca-Juarez, J., Borrini-Mayori, K., . . . Chagas Disease Working Group in, A. (2015). Population structure of the Chagas disease vector *Triatoma infestans* in an urban environment. *PLoS Negl Trop Dis, 9*(2), e0003425. doi:10.1371/journal.pntd.0003425

Kieran, T. J., Gordon, E. R. L., Forthman, M., Hoey-Chamberlain, R., Kimball, R. T., Faircloth, B. C., . . . Glenn, T. C. (2019). Insight from an ultraconserved element bait set designed for hemipteran phylogenetics integrated with genomic resources. *Mol Phylogenet Evol, 130*, 297-303. doi:10.1016/j.ympev.2018.10.026

Kieran, T. J., Gottdenker, N. L., Varian, C. P., Saldana, A., Means, N., Owens, D., . . . Glenn, T. C. (2017). Blood Meal Source Characterization Using Illumina Sequencing in the Chagas Disease Vector *Rhodnius pallescens* (Hemiptera: Reduviidae) in Panama. *J Med Entomol, 54*(6), 1786-1789. doi:10.1093/jme/tjx170

Kjos, S. A., Marcet, P. L., Yabsley, M. J., Kitron, U., Snowden, K. F., Logan, K. S., . . . Dotson, E. M. (2013). Identification of bloodmeal sources and *Trypanosoma cruzi* infection in triatomine bugs (Hemiptera: Reduviidae) from residential settings in Texas, the United States. *J Med Entomol, 50*(5), 1126-1139. doi:10.1603/me12242

Kokko, H., & Lopez-Sepulcre, A. (2006). From individual dispersal to species ranges: perspectives for a changing world. *SCIENCE, 313*(5788), 789-791. doi:10.1126/science.1128566

Koutsovoulos, G., Kumar, S., Laetsch, D. R., Stevens, L., Daub, J., Conlon, C., . . . Blaxter, M. (2016). No evidence for extensive horizontal gene transfer in the genome of the tardigrade Hypsibius dujardini. *Proceedings of the National Academy of Sciences of the United States of America, 113*(18), 5053-5058. doi:10.1073/pnas.1600338113

Kratina, P., LeCraw, R. M., Ingram, T., & Anholt, B. R. (2012). Stability and persistence of food webs with omnivory: Is there a general pattern? *Ecosphere, 3*, 50.

Kuno, G., & Chang, G. J. (2005). Biological transmission of arboviruses: reexamination of and new insights into components, mechanisms, and unique traits as well as

their evolutionary trends. *Clin Microbiol Rev, 18*(4), 608-637. doi:10.1128/CMR.18.4.608-637.2005

Lemmon, A. R., Emme, S. A., & Lemmon, E. M. (2012). Anchored hybrid enrichment for massively high-throughput phylogenomics. *Syst Biol, 61*(5), 727-744. doi:10.1093/sysbio/sys049

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*, 123-520.

Liese, B. H., Houghton, N., & Teplitskaya, L. (2014). Development assistance for neglected tropical diseases: progress since 2009. *International Health, 6*(3), 162-171. doi:10.1093/inthealth/ihu052

Logares, R., Sunagawa, S., Salazar, G., Cornejo-Castillo, F. M., Ferrera, I., Sarmento, H., . . . Acinas, S. G. (2014). Metagenomic 16S rDNA Illumina tags are a powerful alternative to amplicon sequencing to explore diversity and structure of microbial communities. *Environ Microbiol, 16*(9), 2659-2671. doi:10.1111/1462-2920.12250

Matz, M. V. (2018). Fantastic Beasts and How To Sequence Them: Ecological Genomics for Obscure Model Organisms. *Trends Genet, 34*(2), 121-132. doi:10.1016/j.tig.2017.11.002

Mayr, E. (1963). *Animal Species and Evolution*. Cambridge, MA: Harvard University Press.

Maze-Guilmo, E., Blanchet, S., McCoy, K. D., & Loot, G. (2016). Host dispersal as the driver of parasite genetic structure: a paradigm lost? *Ecol Lett, 19*(3), 336-347. doi:10.1111/ele.12564

McCann, K. (2007). Protecting biostructure. *Nature, 446*(7131), 29. doi:10.1038/446029a

McCormack, J. E., Hird, S. M., Zellmer, A. J., Carstens, B. C., & Brumfield, R. T. (2013). Applications of next-generation sequencing to phylogeography and phylogenetics. *Mol Phylogenet Evol, 66*(2), 526-538. doi:10.1016/j.ympev.2011.12.007

McCoy, K. D. (2008). The population genetic structure of vectors and our understanding of disease epidemiology. *Parasite, 15*(3), 444-448. doi:10.1051/parasite/2008153444

McCoy, K. D., Boulinier, T., & Tirard, C. (2005). Comparative host-parasite population structures: disentangling prospecting and dispersal in the black-legged kittiwake *Rissa tridactyla*. *Mol Ecol, 14*(9), 2825-2838. doi:10.1111/j.1365-294X.2005.02631.x

McMeans, B. C., McCann, K. S., Humphries, M., Rooney, N., & Fisk, A. T. (2015). Food Web Structure in Temporally-Forced Ecosystems. *Trends Ecol Evol, 30*(11), 662-672. doi:10.1016/j.tree.2015.09.001

Meentemeyer, R. K., Haas, S. E., & Vaclavik, T. (2012). Landscape epidemiology of emerging infectious diseases in natural and human-altered ecosystems. *Annu Rev Phytopathol, 50*, 379-402. doi:10.1146/annurev-phyto-081211-172938

Mesquita, R. D., Vionette-Amaral, R. J., Lowenberger, C., Rivera-Pomar, R., Monteiro, F. A., Minx, P., . . . Oliveira, P. L. (2015). Genome of *Rhodnius prolixus*, an

insect vector of Chagas disease, reveals unique adaptations to hematophagy and parasite infection. *Proc Natl Acad Sci U S A, 112*(48), 14936-14941. doi:10.1073/pnas.1506226112

Montoya-Porras, L. M., Omar, T. C., Alzate, J. F., Moreno-Herrera, C. X., & Cadavid-Restrepo, G. E. (2018). 16S rRNA gene amplicon sequencing reveals dominance of Actinobacteria in *Rhodnius pallescens* compared to *Triatoma maculata* midgut microbiota in natural populations of vector insects from Colombia. *Acta Trop, 178*, 327-332. doi:10.1016/j.actatropica.2017.11.004

Moran, N. A., McCutcheon, J. P., & Nakabachi, A. (2008). Genomics and evolution of heritable bacterial symbionts. *Annu Rev Genet, 42*, 165-190. doi:10.1146/annurev.genet.41.110306.130119

Morgan, A. D., Gandon, S., & Buckling, A. (2005). The effect of migration on local adaptation in a coevolving host-parasite system. *Nature, 437*(7056), 253-256. doi:10.1038/nature03913

Motro, Y., & Moran-Gilad, J. (2017). Next-generation sequencing applications in clinical bacteriology. *Biomol Detect Quantif, 14*, 1-6. doi:10.1016/j.bdq.2017.10.002

Nelson, M. C., Morrison, H. G., Benjamino, J., Grim, S. L., & Graf, J. (2014). Analysis, optimization and verification of Illumina-generated 16S rRNA gene amplicon surveys. *PLoS One, 9*(4), e94249. doi:10.1371/journal.pone.0094249

Nielsen, J. M., Clare, E. L., Hayden, B., Brett, M. T., Kratina, P., & Gilbert, M. T. P. (2018). Diet tracing in ecology: Method comparison and selection. *Methods in Ecology and Evolution, 9*(2), 278-291. doi:10.1111/2041-210x.12869

Oliveira, M. A., Ferreira, R. L., Carneiro, M. A., & Diotaiut, L. (2008). Ecology of *Cavernicola pilosa* Barber, 1937 (Hemiptera: Reduviidae: Triatominae) in the Boa Esperanc ̧a Cave, Tocantins, Brazil. *Ecotropica, 14*, 63-68.

Oliver, K. M., & Martinez, A. J. (2014). How resident microbes modulate ecologically-important traits of insects. *Curr Opin Insect Sci, 4*, 1-7. doi:10.1016/j.cois.2014.08.001

Orantes, L. C., Monroy, C., Dorn, P. L., Stevens, L., Rizzo, D. M., Morrissey, L., . . . Helms Cahan, S. (2018). Uncovering vector, parasite, blood meal and microbiome patterns from mixed-DNA specimens of the Chagas disease vector *Triatoma dimidiata*. *PLoS Negl Trop Dis, 12*(10), e0006730. doi:10.1371/journal.pntd.0006730

Osei-Poku, J., Mbogo, C. M., Palmer, W. J., & Jiggins, F. M. (2012). Deep sequencing reveals extensive variation in the gut microbiota of wild mosquitoes from Kenya. *Mol Ecol, 21*(20), 5138-5150. doi:10.1111/j.1365-294X.2012.05759.x

Otálora-Luna, F., Pérez-Sánchez, A. J., Sandoval, C., & Aldana, E. (2015). Evolution of hematophagous habit in Triatominae (Heteroptera: Reduviidae). *Revista Chilena De Historia Natural, 88*(1). doi:10.1186/s40693-014-0032-0

Patz, J. A., Daszak, P., Tabor, G. M., Aguirre, A. A., Pearl, M., Epstein, J., . . . Bradley, D. J. (2004). Unhealthy landscapes: Policy recommendations on land use change and infectious disease emergence. *Environ Health Perspect, 112*(10), 1092-1098.

Pennings, P. S., Achenbach, A., & Foitzik, S. (2011). Similar evolutionary potentials in an obligate ant parasite and its two host species. *J Evol Biol, 24*(4), 871-886. doi:10.1111/j.1420-9101.2010.02223.x

Pennisi, E. (2010). Armed and dangerous. *SCIENCE, 327*(5967), 804-805. doi:10.1126/science.327.5967.804

Perez de Rosas, A. R., Segura, E. L., Fichera, L., & Garcia, B. A. (2008). Macrogeographic and microgeographic genetic structure of the Chagas' disease vector *Triatoma infestans* (Hemiptera: Reduviidae) from Catamarca, Argentina. *Genetica, 133*(3), 247-260. doi:10.1007/s10709-007-9208-8

Perez de Rosas, A. R., Segura, E. L., & Garcia, B. A. (2011). Molecular phylogeography of the Chagas' disease vector *Triatoma infestans* in Argentina. *Heredity (Edinb), 107*(1), 71-79. doi:10.1038/hdy.2010.159

Piccinali, R. V., Marcet, P. L., Noireau, F., Kitron, U., Gurtler, R. E., & Dotson, E. M. (2009). Molecular population genetics and phylogeography of the Chagas disease vector *Triatoma infestans* in South America. *J Med Entomol, 46*(4), 796-809.

Pinto, J., Roellig, D. M., Gilman, R. H., Calderón, M., Bartra, C., Salazar, R., . . . Cama, V. (2012). Temporal differences in blood meal detection from the midguts of *Triatoma infestans*. *Rev Inst Med Trop Sao Paulo, 54*(2), 83-88. doi:10.1590/s0036-46652012000200005

Poelen, J. H., Simons, J. D., & Mungall, C. J. (2014). Global biotic interactions: An open infrastructure to share and analyze species-interaction datasets. *Ecological Informatics, 24*, 148-159. doi:10.1016/j.ecoinf.2014.08.005

Poretsky, R., Rodriguez, R. L., Luo, C., Tsementzi, D., & Konstantinidis, K. T. (2014). Strengths and limitations of 16S rRNA gene amplicon sequencing in revealing temporal microbial community dynamics. *PLoS One, 9*(4), e93827. doi:10.1371/journal.pone.0093827

Prugnolle, F., Durand, P., Jacob, K., Razakandrainibe, F., Arnathau, C., Villarreal, D., . . . Renaud, F. (2008). A comparison of *Anopheles gambiae* and *Plasmodium falciparum* genetic structure over space and time. *Microbes Infect, 10*(3), 269-275. doi:10.1016/j.micinf.2007.12.021

Prugnolle, F., Theron, A., Pointier, J. P., Jabbour-Zahab, R., Jarne, P., Durand, P., & de Meeus, T. (2005). Dispersal in a parasitic worm and its two hosts: consequence for local adaptation. *EVOLUTION, 59*(2), 296-303.

Puente, J. M., Ruiz, S., Soriguer, R., & Figuerola, J. (2013). Effect of blood meal digestion and DNA extraction protocol on the success of blood meal source determination in the malaria vector *Anopheles atroparvus*. *Malaria Journal, 12*.

Radford, A. D., Chapman, D., Dixon, L., Chantrey, J., Darby, A. C., & Hall, N. (2012). Application of next-generation sequencing technologies in virology. *J Gen Virol, 93*(Pt 9), 1853-1868. doi:10.1099/vir.0.043182-0

Ramirez, C. J., Jaramillo, C. A., del Pilar Delgado, M., Pinto, N. A., Aguilera, G., & Guhl, F. (2005). Genetic structure of sylvatic, peridomestic and domestic populations of *Triatoma dimidiata* (Hemiptera: Reduviidae) from an endemic zone of Boyaca, Colombia. *Acta Trop, 93*(1), 23-29. doi:10.1016/j.actatropica.2004.09.001

Ratnasingham, S., & Hebert, P. D. (2007). bold: The Barcode of Life Data System (http://www.barcodinglife.org). *Mol Ecol Notes, 7*(3), 355-364. doi:10.1111/j.1471-8286.2007.01678.x

Reisen, W. K. (2010). Landscape epidemiology of vector-borne diseases. *Annu Rev Entomol, 55*, 461-483. doi:10.1146/annurev-ento-112408-085419

Ribeiro, J. M. (1995). Blood-feeding arthropods: live syringes or invertebrate pharmacologists? *Infect Agents Dis, 4*(3), 143-152.

Ribeiro, J. M. C., Assumpção, T. C., & Francischetti, I. M. B. (2012). An Insight into the Sialomes of Bloodsucking Heteroptera. *Psyche: A Journal of Entomology, 2012*, 1-16. doi:10.1155/2012/470436

Rinker, D. C., Pitts, R. J., & Zwiebel, L. J. (2016). Disease vectors in the era of next generation sequencing. *Genome Biol, 17*(1), 95. doi:10.1186/s13059-016-0966-4

Saldana, M. A., Hegde, S., & Hughes, G. L. (2017). Microbial control of arthropod-borne disease. *Mem Inst Oswaldo Cruz, 112*(2), 81-93. doi:10.1590/0074-02760160373

Salvatella, A. R., Basmadjian, Y., Rosa, R., & Puime, A. (1992). Triatoma delpontei Romana & Abalos, 1947 (Hemiptera, Tratominae) in the Brazilian State of Rio Grande do Sul. *Rev Inst Med Trop Sao Paulo, 35*, 73-76.

Sandoval, C. M., Duarte, R., Gutierrez, R., Rocha, D. d. S., Angulo, V. M., Esteban, L., . . . Galvao, C. (2004). Feeding sources and natural infection of Belminus herreri (Hemiptera, Reduviidae, Triatominae) from dwellings in Cesar, Colombia. *Mem Inst Oswaldo Cruz, 99*(2), 137-140.

Sandoval, C. M., Ortiz, N., Jaimes, D., Lorosa, E., Galvao, C., Rodriguez, O., . . . Gutierrez, R. (2010). Feeding behaviour of *Belminus ferroae* (Hemiptera: Reduviidae), a predaceous Triatominae colonizing rural houses in Norte de Santander, Colombia. *Med Vet Entomol, 24*(2), 124-131. doi:10.1111/j.1365-2915.2010.00868.x

Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., . . . Weber, C. F. (2009). Introducing mothur: open-source, platform-independent,

community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol, 75*(23), 7537-7541. doi:10.1128/AEM.01541-09

Schmieder, R., & Edwards, R. (2011). Quality control and preprocessing of metagenomic datasets. *Bioinformatics, 27*(6), 863-864. doi:10.1093/bioinformatics/btr026

Schofield, C. J., & Galvão, C. (2009). Classification, evolution, and species groups within the Triatominae. *Acta Trop, 110*(2-3), 88-100. doi:10.1016/j.actatropica.2009.01.010

Starrett, J., Derkarabetian, S., Hedin, M., Bryson, R. W., Jr., McCormack, J. E., & Faircloth, B. C. (2017). High phylogenetic utility of an ultraconserved element probe set designed for Arachnida. *Mol Ecol Resour, 17*(4), 812-823. doi:10.1111/1755-0998.12621

Stevens, L., Monroy, M. C., Rodas, A. G., Hicks, R. M., Lucero, D. E., Lyons, L. A., & Dorn, P. L. (2015). Migration and Gene Flow Among Domestic Populations of the Chagas Insect Vector *Triatoma dimidiata* (Hemiptera: Reduviidae) Detected by Microsatellite Loci. *J Med Entomol, 52*(3), 419-428. doi:10.1093/jme/tjv002

Stothard, J., Frame, I., & Miles, M. (1999). Genetic diversity and genetic exchange in *Trypanosoma cruzi*: dual drug-resistant "progeny" from episomal transformants. *Mem Inst Oswaldo Cruz, 94 Suppl 1*, 189-193.

Suzuki, M. T., & Giovannoni, S. J. (1996). Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR. *Appl Environ Microbiol, 62*(2), 625-630.

Tabachnick, W. J., & Black, W. C. (1995). Making a Case for Molecular Population Genetic-Studies of Arthropod Vectors. *PARASITOLOGY TODAY, 11*(1), 27-30. doi:Doi 10.1016/0169-4758(95)80105-7

Tremblay, J., Singh, K., Fern, A., Kirton, E. S., He, S., Woyke, T., . . . Tringe, S. G. (2015). Primer and platform effects on 16S rRNA tag sequencing. *Front Microbiol, 6*, 771. doi:10.3389/fmicb.2015.00771

Usinger, R. L. (1944). *The Triatominae of North and Central America and the West Indies and their public health significance.* Washington: US Public Health Service.

van Dijk, E. L., Auger, H., Jaszczyszyn, Y., & Thermes, C. (2014). Ten years of next-generation sequencing technology. *Trends Genet, 30*(9), 418-426. doi:10.1016/j.tig.2014.07.001

van Schaik, J., Kerth, G., Bruyndonckx, N., & Christe, P. (2014). The effect of host social system on parasite population genetic structure: comparative population genetics of two ectoparasitic mites and their bat hosts. *BMC Evol Biol, 14*, 18. doi:10.1186/1471-2148-14-18

Viboud, C., Bjornstad, O. N., Smith, D. L., Simonsen, L., Miller, M. A., & Grenfell, B. T. (2006). Synchrony, waves, and spatial hierarchies in the spread of influenza. *SCIENCE, 312*(5772), 447-451. doi:10.1126/science.1125237

Vieira, C. S., Mattos, D. P., Waniek, P. J., Santangelo, J. M., Figueiredo, M. B., Gumiel, M., . . . Azambuja, P. (2015). *Rhodnius prolixus* interaction with *Trypanosoma rangeli*: modulation of the immune system and microbiota population. *Parasit Vectors, 8*, 135. doi:10.1186/s13071-015-0736-2

Wachi, N., Matsubayashi, K. W., & Maeto, K. (2018). Application of next-generation sequencing to the study of non-model insects. *Entomological Science, 21*(1), 3-11. doi:10.1111/ens.12281

Waleckx, E., Suarez, J., Richards, B., & Dorn, P. L. (2014). *Triatoma sanguisuga* blood meals and potential for Chagas disease, Louisiana, USA. *Emerg Infect Dis, 20*(12), 2141-2143. doi:10.3201/eid2012.131576

Weirauch, C. (2008). Cladistic analysis of Reduviidae (Heteroptera : Cimicomorpha) based on morphological characters. *SYSTEMATIC ENTOMOLOGY, 33*(2), 229-274. doi:10.1111/j.1365-3113.2007.00417.x

Weirauch, C., & Munro, J. B. (2009). Molecular phylogeny of the assassin bugs (Hemiptera: Reduviidae), based on mitochondrial and nuclear ribosomal genes. *Mol Phylogenet Evol, 53*(1), 287-299. doi:10.1016/j.ympev.2009.05.039

Weiss, B., & Aksoy, S. (2011). Microbiome influences on insect host vector competence. *Trends Parasitol, 27*(11), 514-522. doi:10.1016/j.pt.2011.05.001

Witsenburg, F., Clement, L., Lopez-Baucells, A., Palmeirim, J., Pavlinic, I., Scaravelli, D., . . . Christe, P. (2015). How a haemosporidian parasite of bats gets around: the genetic structure of a parasite, vector and host compared. *Mol Ecol, 24*(4), 926-940. doi:10.1111/mec.13071

Zhang, J., Gordon, E. R., Forthman, M., Hwang, W. S., Walden, K., Swanson, D. R., . . . Weirauch, C. (2016). Evolution of the assassin's arms: insights from a phylogeny of combined transcriptomic and ribosomal DNA data (Heteroptera: Reduvioidea). *Sci Rep, 6*, 22177. doi:10.1038/srep22177

Zindel, R., Gottlieb, Y., & Aebi, A. (2011). Arthropod symbioses: a neglected parameter in pest- and disease-control programmes. *JOURNAL OF APPLIED ECOLOGY, 48*(4), 864-872. doi:10.1111/j.1365-2664.2011.01984.x

Zink, R. M. (2010). Drawbacks with the use of microsatellites in phylogeography: the song sparrow *Melospiza melodiaas* a case study. *Journal of Avian Biology, 41*(1), 1-7. doi:10.1111/j.1600-048X.2009.04903.x

CHAPTER 2

BLOODMEAL SOURCE CHARACTERIZATION USING ILLUMINA SEQUENCING IN THE CHAGAS DISEASE VECTOR *RHODNIUS PALLESCENS* (HEMIPTERA: REVUVIIDAE) IN PANAMA[*]

[*] Kieran TJ, Gottdenker NL, Varian CP, Saldana A, Means N, Owens D, Calzada JE, Glenn TC. 2017. *Journal of Medical Entomology* 54(6): 1786-1789. Reprinted here with permission of the publisher.

Abstract

Accurate blood meal identification is critical to understanding hematophagous vector-host relationships. This study describes a customizable Next-Generation Sequencing (NGS) approach to identify blood meals from *Rhodnius pallescens* triatomines using multiple barcoded primers and existing software to pick operational taxonomic units and match sequences for blood meal identification. We precisely identified all positive control samples using this method and further examined 74 wild-caught *R. pallescens* samples. With this novel blood meal identification method, we detected 13 vertebrate species in the blood meals, as well as single and multiple blood meals in individual bugs. Our results demonstrate the reliability and descriptive uses of our method.

Introduction

Identifying blood meals of hematophagous zoonotic disease vectors is important to understand vector-host preferences and ecological relationships, host availability, and contact rates. This information is critical for prediction and prevention of vector-borne zoonotic diseases. A variety of molecular (Kent, 2009) and immunological methods (Christensen & Vasquez, 1981; Dias, Bezerra, Machado, Casanova, & Diotaiuti, 2008; Gomes et al., 2001; Pineda et al., 2008) that vary in sensitivity and specificity have been developed to identify blood meals from zoonotic disease vectors. Unfortunately, many of these methods cannot easily identify multiple blood meals from different species fed upon by an individual vector. Identification of multiple blood meals within a single vector by conventional PCR and sequencing often requires additional cloning steps

(Waleckx, Suarez, Richards, & Dorn, 2014). Next-generation sequencing (NGS) facilitates identification within mixed samples because each read derives from a single source DNA-molecule (Metzker, 2010). In this study, we identify blood meal sources of a triatomine Chagas disease vector using an NGS technique.

Triatomine vectors of *Trypanosoma cruzi,* etiologic agent of Chagas disease, are hematophagous members of the Reduviidae, subfamily Triatominae. In many areas where Chagas disease is endemic, sylvatic *T. cruzi* transmission between triatomines and wild mammalian host species predominates. In some areas of the Neotropics, *T. cruzi* transmission is associated with palm-inhabiting *Rhodnius* triatomines and a wide range of mammal reservoir hosts; humans can become infected by contact with bugs that fly from palm trees into domiciles, habitation by some *Rhodnius* species living in palm-thatched roofs, and accidental ingestion of bug-contaminated food (Abad-Franch et al., 2015). In this study, we describe a novel NGS method to identify blood meals in wild-caught *Rhodnius pallescens*, the principal vector of *T. cruzi* in Panama (Christensen & Vasquez, 1981; Whitlaw & Chanoitis, 1978).

Methods

R. pallescens vectors (N=74 total) were caught using Noireau traps (Noireau et al., 2002) from *Attalea butyracea* palm trees in two locations in Panamá: Trinidad de las Minas, Capira (N 8.46713°, W 79.59451°) (N=40 bugs from 2 palms,) and Las Pavas, La Chorrera (N 9.05448°, W 79.53352°) (N=34 bugs from 2 palms), and placed in 95% molecular grade ethanol. DNA from 9 mammals (Table 2.1) collected from capture of wild mammals (IACUC approval number 2013 05-001-43-Y3-A0) was used as positive

controls. We used a purified DNA-free water as a negative control for every PCR reaction and subsequent sequencing.

The whole insect was macerated and digested overnight in 400uL digest buffer (100 mM Tris, 50 mM EDTA, 150 mM NaCl) with 5µL 10mg/µL Proteinase K. DNA was extracted with Phenol-Chloroform-Isoamyl alcohol. Completed extractions were reconstituted in TLE (10 mM Tris 0.2 mM EDTA) and impurities were cleaned from the extracted samples with SPRI-beads (Thermo-Scientific, Waltham, MA, USA) using a 1:1 ratio. Cleaned extractions were reconstituted in 30µL TLE. DNA was amplified using two sets of primers: the first amplified the target loci and the second converted the amplicon into a library ready for Illumina sequencing. The first primer pair targeted a 145bp region of the 12S rRNA gene commonly used for barcoding and detecting vertebrates (F-5'-CAAACTGGGATTAGATACC-3', R-5'-AGAACAGGCTCCTCTAG-3' (Humair et al., 2007). The vertebrate 12S rRNA gene was chosen because this region is relatively short (145bp), allowing for identification of partially digested blood meals within vectors, and providing high sensitivity for species identification (Gottdenker et al. 2012). On the 5' end, we added Illumina TruSeq Read 1 to the forward and Illumina TruSeq Read 2 to the reverse primer. We synthesized 8 forward and 12 reverse fusion primers, each with a unique variable length (5-8bp) index sequence between the 12S and TruSeq sequences (Table 2.2).

For the first-round PCR, we used 12.5 µL reactions of KAPA HiFi HotStart Kits (Kapa Biosystems, Wilmington, Massachusetts, USA). We created a master mix for all samples using 2.5µL of 5x Buffer, 0.375µL of 10mM dNTPs, 0.25µL hot start Taq, and 5.4µL molecular grade water. We aliquoted out 8.5µL of master mix to each well of a

PCR plate and individually added 1µL of 5µM forward primer, 1uL of 5µM reverse primer, and 2µL of DNA. The ranges of DNA sample concentrations were from 10 ng/µL to 60 ng/µl. Each DNA sample was paired with a unique primer-index combination with the following thermocycler conditions: 98°C for 3 min, followed by 30 cycles at 95°C for 30s, 63°C for 1 min, 72°C for 1 min and a final extension at 72°C for 5 min. Amplification success was verified on a 1.5% agarose gel. Successful amplicons were pooled in equal concentrations and cleaned using a 1:1 ratio of SPRI-beads and reconstituted in 25µL TLE.

The second-round PCR primers consisted of Illumina TruSeqHT compatible 8 nt indexed primers (Glenn et al., 2016). We used 25 µL reaction of KAPA HiFi HotStart Kits using 5µL of 5x Buffer, 0.75µL of 10mM dNTPs, 0.5µL HotStart, 3.75µL molecular grade water, 2.5µL of 5µM forward primer, 2.5µL 5µM reverse primer, and 10µL of 12S amplicon pool. We performed two replicate PCRs with the following thermocycler conditions: 98°C for 2 min, followed by 10 cycles at 98°C for 30s, 60°C for 30s, 72°C for 30s and a final extension at 72°C for 5 min. Library product was cleaned and primers were removed with SPRI-beads (1:1 ratio) and pooled with other uniquely indexed samples prior to sequencing. All libraries were sent to the Georgia Genomics Facility (http://dna.uga.edu) for sequencing on an Illumina MiSeq using a PE300 kit (Illumina, San Diego, CA).

Sequencing data were demultiplexed according to outer indexes using bcl2fastq (Illumina, v1.8.4). The 12S amplicon pool was demultiplexed by internal barcodes and primers removed using Mr. Demuxy v1.2.0 (https://pypi.python.org/pypi/Mr_Demuxy/1.2.0). Paired-end sequencing reads were

imported into Geneious v8.1 (Biomatters Limited, NJ), set as paired-reads with an expected insert size of 145 bp, and trimmed to remove low quality bases using default settings and a quality score of 0.001. Paired-end sequencing reads were then merged using the FLASH v1.2.9 plugin (Magoc & Salzberg, 2011) and read lengths below 80bp were excluded. Data was exported from Geneious as FASTA files and imported into the software package QIIME v1.9.1 (Caporaso et al., 2010) for Operational Taxonomic Unit (OTU) designation and identification. We downloaded 12S DNA sequences of Panamanian vertebrate species from Genbank to compile a custom 12S reference database and taxonomy file. QIIME facilitates sequence identification procedures as the NGS process results in thousands of reads for many samples simultaneously, but it requires a reference sequence database and taxonomy file. For missing reference sequences, we downloaded closely related species. Reference sequences were trimmed to the corresponding 145bp region of 12S in Geneious so no non-target sequence may potentially match. We used QIIME's default UCLUST (Edgar, 2010) OTU picking strategy and taxonomic identification was defined using QIIME's BLAST with a ≥98% similarity to reference sequences. Resulting OTU identifications for each sample were output showing absolute abundance (total number of reads). For three samples that had largely unidentified reads, we clustered sequences at 99% similarity and BLASTed the most abundant sequences on Genbank. For a conservative estimate of blood meal source, we eliminated species hits receiving ≤10% of total read hits for the sample.

Results

All positive controls were correctly identified using our method (Table 2.1). From *R. pallescens* samples, we identified a total of 13 unique vertebrate blood meals (11 mammals, 1 bird, 1 reptile) across all samples with two species comprising 75.5% of total reads, *Didelphis marsupialis* (Common opossum) (34.1%) and *Coendou* sp. (Prehensile-tailed porcupine) (41.4%) (Table 2.3). The Trinidad de las Minas site had more diverse blood meal sources (N=12 species) compared to the Las Pavas site (N=5 species). Human DNA was present in 6.9% of reads (N=28). We also found *R. pallescens* individuals with multiple blood meals including dual (N= 27), triple (N= 13), and quadruple (N=3) species detections, whereas 30 had a single blood meal (Figure 2.1). All blood meals were identified animals ranging in local environments.

Discussion

We precisely identified species that are found at the collection sites, including two common arboreal species (the common opossum, *Didelphis marsupialis* and the prehensile-tailed porcupine, *Coendou sp.*), which comprise most blood meal sources in the collected samples, with the remaining source species at lower frequencies. Identifying these source species is an important step to better understanding *T. cruzi* transmission ecology and improved evaluation of human disease risk. Currently, this method is limited to the availability of reference sequences for the target loci of interest. For instance, in our study, *Coendou sp.* matches to the only available sequence from GenBank (*Coendou bicolor*), yet a different species is in Panama and is the most likely source (*Coendou rothschildi*). Our study improves blood meal detection and reliability from previous

studies. We detected blood meals in 100% of samples tested and identified 98% of blood meal sources. Similar studies had detection rates of 40.3% using 12S rRNA gene conventional PCR followed by conventional sequencing (Gottdenker, Fernando Chaves, Calzada, Saldana, & Carroll, 2012) and 27.3% using an IgG dot-blot protein assay (Pineda et al., 2008). The most significant improvement of our method over previous studies is the finer taxonomic resolution and ability to detect multiple blood meal sources in a single vector. Gottdenker and colleagues (Gottdenker et al., 2012) previously identified 26 out of 42 taxa to the species level, while Pineda and colleagues (Pineda et al., 2008) blood meal identifications corresponded to approximately the family level.

One sample had a majority reads that did not correspond to any of the species in our reference database or reliable sources manually using the NCBI website BLAST, demonstrating the gaps in reference material. Based on the BLAST results, the sample is likely an unidentified reptile. Over time, as more wild host species are genetically characterized in the areas of study, these gaps in reference material will likely diminish, making this method more accurate. Furthermore, multiple blood meals observed in younger nymphal stages (N1, N2) in our samples are unexpected, and may be due to hematoklepty, defined as a bug feeding from the blood meal ingested by another bug, or unexpected contamination, although we followed all standard procedures (negative control samples) to avoid contamination. Furthermore, no gel bands were detected on negative control PCRs and negative control samples sent to sequencing indicated no contamination during sample processing or library preparation. Studies on younger instar triatomine blood meal preferences in the wild are rare, and further research on the

ecology of this vector is needed to investigate these observations and identify additional potential key reservoir species.

An understanding of triatomine feeding preferences is important for the development and implementation of Chagas disease control and surveillance, particularly for sylvatic and peridomestic Chagas transmission cycles. This study describes an accurate NGS method for characterizing blood meal composition from multiple sources of mixed DNA samples in individual triatomine vectors. Fusion primers are easily constructed for other target loci, and may be constructed with or without internal indexes, allowing for a versatile and customizable system for blood meal identification in additional hematophagous vector species. While this method is efficient and precise at blood meal identification, extreme care must be exercised to reduce contamination at all project stages, and effort will be necessary to ensure a set of proper reference sequences for all possible vertebrate species in a study area are available.

Acknowledgements

We thank Sr. Jose Montenegro with help in field capture of triatomines and Sr. Roberto Rojas with identification of triatomine species and stage. Anamaría Santamaría and Vanessa Pineda of the ICGES helped with DNA extraction from vectors. The authors also thank the Ministerio de Ambiente in Panama (MiAMBIENTE), the Smithsonian Tropical Research Institute (STRI). This study was funded by the UGA Office of the Vice President for Research, Secretaria Nacional de Ciencia Tecnología e Inovación (SENACYT), grant COL11-043 awarded to ICGES.

References

Abad-Franch, F., Lima, M. M., Sarquis, O., Gurgel-Goncalves, R., Sanchez-Martin, M., Calzada, J., . . . Gottdenker, N. L. (2015). On palms, bugs, and Chagas disease in the Americas. *Acta Trop, 151*, 126-141. doi:10.1016/j.actatropica.2015.07.005

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., . . . Knight, R. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat Methods, 7*(5), 335-336. doi:10.1038/nmeth.f.303

Christensen, H. A., & Vasquez, A. M. (1981). Host Feeding Profiles of Rhodnius pallescens (Hemiptera: Reduviidae) in Rural Villages of Central Panama. *Am J Trop Med Hyg, 30*(1), 278-283.

Dias, F. B., Bezerra, C. M., Machado, E. M., Casanova, C., & Diotaiuti, L. (2008). Ecological aspects of Rhodnius nasutus Stal, 1859 (Hemiptera: Reduviidae: Triatominae) in palms of the Cahpada do Araripe in Ceara, Brazil. *Mem Inst Oswaldo Cruz, 103*(8), 824-830.

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics, 26*(19), 2460-2461. doi:10.1093/bioinformatics/btq461

Glenn, T. C., Nilsen, R., Kieran, T. J., Finger, J. W., Pierson, T. W., Bentley, K. E., . . . Faircloth, B. C. (2016). Adapterama I: Universal stubs and primers for thousands of dual-indexed Illumina libraries (iTru & iNext)

. *BioRxiv[http://biorxiv.org/content/early/2016/06/15/049114]*. doi:https://doi.org/10.1101/049114

Gomes, L. A., Duarte, R., Lima, D. C., Diniz, B. S., Serrao, M. L., & Labarthe, N. (2001). Comparison between precipitin and ELISA tests in the bloodmeal detection of Aedes aegypti (Linnaeus) and Aedes fluviatilis (Lutz) mosquitoes experimentally fed on feline, canine and human hosts. *Mem Inst Oswaldo Cruz, 96*(5), 693-695.

Gottdenker, N. L., Fernando Chaves, L., Calzada, J. E., Saldana, A., & Carroll, C. R. (2012). Host Life History Strategy, Species Diversity, and Habitat Influence Trypanosoma cruzi Vector Infection in Changing Landscapes. *Plos Neglected Tropical Diseases, 6*(11), e1884. doi:10.1371/journal.pntd.0001884

Humair, P.-F., Douet, V., Cadenas, F. M., Schouls, L. M., Van de Pol, I., & Gern, L. (2007). Molecular Identification of Bloodmeal Source in Ixodes ricinus Ticks Using 12S rDNA As a Genetic Marker. *J. Med. Entomol., 44*(5), 869-880.

Kent, R. J. (2009). Molecular methods for arthropod bloodmeal identification and applications to ecological and vector-borne disease studies. *Mol Ecol Resour, 9*(1), 4-18. doi:10.1111/j.1755-0998.2008.02469.x

Magoc, T., & Salzberg, S. L. (2011). FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics, 27*(21), 2957-2963. doi:10.1093/bioinformatics/btr507

Metzker, M. L. (2010). Sequencing technologies - the next generation. *Nat Rev Genet, 11*(1), 31-46. doi:10.1038/nrg2626

Noireau, F., Abad-Franch, F., Valente, S. A., Dias-Lima, A., Lopes, C., Cunha, V., . . . Jurberg, J. (2002). Trapping Triatominae in Silvatic Habitats. *Mem Inst Oswaldo Cruz, 97*(1), 61-63.

Pineda, V., Montalvo, E., Alvarez, D., Sanatamaria, A. M., Calzada, J. E., & Saldaña, A. (2008). Feeding sources and trypanosome infection index of Rhodnius pallescens in a Chagas disease endemic area of Amador County, Panama. *Rev Inst Med Trop Sao Paulo, 50*(2), 113-116.

Waleckx, E., Suarez, J., Richards, B., & Dorn, P. L. (2014). Triatoma sanguisuga blood meals and potential for Chagas disease, Louisiana, USA. *Emerging Infectious Disease, 20*(12), 2141-2143.

Whitlaw, J. T., & Chanoitis, B. N. (1978). Palm Trees and Chagas Disease Panama. *Am J Trop Med Hyg, 27*(5), 873-881.

Tables

Table 2.1. Total reads after 10% cutoff of positive control samples (columns) with species match (rows). Unidentified reads numbers were retained for reference. Positive control sample ID key: 1 = monkey; 2 = calf; 3, 6, 7 = sloth; 4, 5 = dog; 8 = armadillo; 9, 10 = human; 11, 12 = rat; 13 = *Marmosa* (mouse opossum); 14, 15 = opossum.

#OTU ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Owl monkey (*Aotus sp.*)	6403														
Domestic cow (*Bos taurus*)		11110													
Three toed sloth *(Bradypus variegatus)*			140												
Domestic dog (*Canis lupus familiaris)*				3321	862										
Two toed sloth *(Choloepus hoffmanni)*						10950	7704								
Six banded armadillo *(Euphract*								6466							

us sexcinctus)														
Human *(Homo sapiens)*		80[a]						6223[b]	4739[b]					
Armored rat (*Hoplomys gymnurus*)										8958	4097			
Mouse opossum (*Marmosa sp.*)												3637		
Brown four-eyed opossum (*Metachirus nudicaudatus*)													2381	
Gray four-eyed opossum (*Philander opossum*)														5360
Spiny Rat (*Proechimys semispinosus*)										1196[b]	444[b]			
Tamandua		43												

(*Tamandua mexicana*)			[a]													
Unidentified	2827	1155	2	133	38	2170	629	431	7237	5074	674	165	323	74	2826	
Total Reads	9230	11225	263	3454	901	1320	833	697	1360	9813	10826	4706	3960	2455	8186	

a. Pairwise similarities between *Bradypus*, *Homo* and *Tamandua* range from 75-80% meaning non-*Bradypus* hits are likely due to contamination.

b. Pairwise similarities between *Hoplomys* and *Proechimys* are 91-93% meaning *Proechimys* hits are likely due to close similarity clustering both species during UCLUST.

Table 2.2. Fusion-indexed 12S primers used in this study showing the variable length barcodes in bold type.

Primer Name	Sequence (fusion, barcode, primer) 5'-3'
iTru_A_12S_6F	ACACTCTTTCCCTACACGACGCTCTTCCGATCT **GGTAC** CAAACTGGGATTAGATACC
iTru_B_12S_6F	ACACTCTTTCCCTACACGACGCTCTTCCGATCT **CAACAC** CAAACTGGGATTAGATACC
iTru_C_12S_6F	ACACTCTTTCCCTACACGACGCTCTTCCGATCT **ATCGGTT** CAAACTGGGATTAGATACC
iTru_D_12S_6F	ACACTCTTTCCCTACACGACGCTCTTCCGATCT **TCGGTCAA** CAAACTGGGATTAGATACC
iTru_E_12S_6F	ACACTCTTTCCCTACACGACGCTCTTCCGATCT **AAGCG** CAAACTGGGATTAGATACC
iTru_F_12S_6F	ACACTCTTTCCCTACACGACGCTCTTCCGATCT **GCCACA** CAAACTGGGATTAGATACC
iTru_G_12S_6F	ACACTCTTTCCCTACACGACGCTCTTCCGATCT **CTGGATG** CAAACTGGGATTAGATACC
iTru_H_12S_6F	ACACTCTTTCCCTACACGACGCTCTTCCGATCT **TGATTGAC** CAAACTGGGATTAGATACC
iTru_1_12S_9R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT **AGGAA** AGAACAGGCTCCTCTAG
iTru_2_12S_9R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT **GAGTGG** AGAACAGGCTCCTCTAG
iTru_3_12S_9R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT **CCACGTC** AGAACAGGCTCCTCTAG
iTru_4_12S_9R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT **TTCTCAGC** AGAACAGGCTCCTCTAG
iTru_5_12S_9R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT **CTAGG** AGAACAGGCTCCTCTAG
iTru_6_12S_9R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT **TGCTTA** AGAACAGGCTCCTCTAG
iTru_7_12S_9R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT **GCGAAGT** AGAACAGGCTCCTCTAG
iTru_8_12S_9R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT **AATCCTAT** AGAACAGGCTCCTCTAG
iTru_9_12S_9R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT **ATCTG** AGAACAGGCTCCTCTAG
iTru_10_12S_9R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT **GAGACT** AGAACAGGCTCCTCTAG
iTru_11_12S_9R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT **CGATTCC** AGAACAGGCTCCTCTAG
iTru_12_12S_9R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT **TCTCAATC** AGAACAGGCTCCTCTAG

Table 2.3. Percentage (%) of total sequence reads and total number of insect samples (No.) matching each species at each site. “a” = High proportion of unidentified of Las Pavas Palm 1 likely the result of lower sequence reads (N= 609) for these samples.

	Trinidad de las Minas (Total N = 86)			Las Pavas (Total N = 61)		
	Site Totals	Palm 1	Palm 2	Site Totals	Palm 1	Palm 2

	% (No.)	% (No.)	% (No.)	% (No.)	% (No.)	% (No.)
Fruit Bat (*Artibeus sp.*)	10.5 (2)	10.5 (2)				
Domestic cow (*Bos taurus*)	0.1 (1)	0.1 (1)				
Domestic dog (*Canis lupus familiaris*)	0.3 (1)	0.3 (1)				
Prehensile-tailed porcupine (*Coendou bicolor*)	21.0 (7)	22.6 (7)		53.8 (19)		53.8 (19)
Common opossum (*Didelphis marsupialis*)	24.5 (18)	26.0 (16)	5.0 (2)	39.9 (26)	53.0 (6)	39.9 (20)
Human (*Homo sapiens*)	15.5 (22)	11.8 (16)	63.9 (6)	1.7 (6)	11.0 (2)	1.7 (4)
Armored rat (*Hoplomys gymnurus)*	0.3 (1)	0.3 (1)				
Long-tailed skink (*Mabuya sp.*)	0.4 (1)		6.0 (1)			
Brown four-eyed opossum (*Metachirus nudicaudatus*)	12.0 (5)	12.8 (4)	2.5 (1)			
House mouse (*Mus musculus)*	0.2 (2)	0.2 (2)				
Black myotis bat (*Myotis*	9.3 (16)	9.1 (12)	11.0 (4)	3.4 (7)	1.5 (1)	3.4 (6)

nigricans)						
Summer tanager-bird *Piranga rubra*	0.8 (2)		10.8 (2)			
Tamandua (*Tamandua mexicana*)				1.2 (1)		1.2 (1)
Unidentified	4.2 (8)	5.5 (7)	0.9 (1)	0.0 (2)	34.5[a] (2)	

Figures

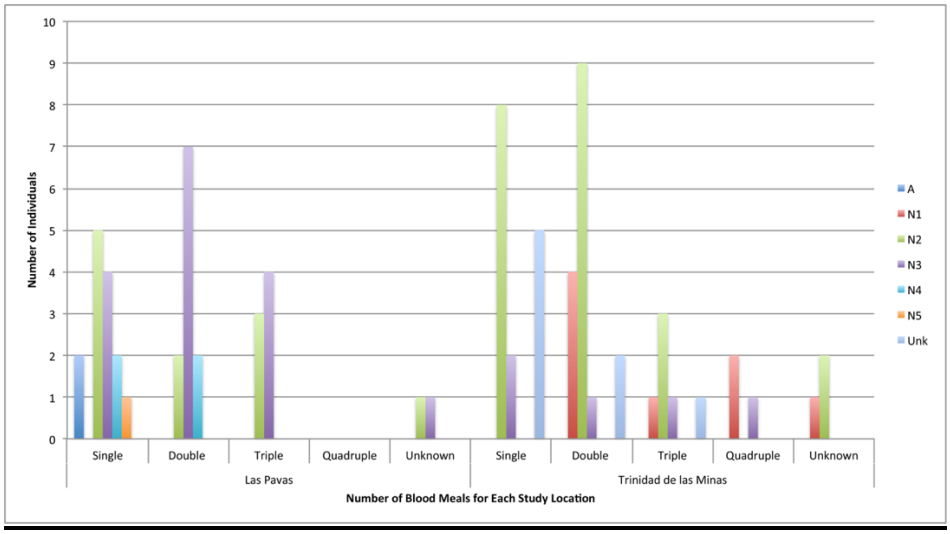


Figure 2.1. Bar graph showing the number of individual samples containing one (Single) to four (Quadruple) unique blood meals for each location grouped by *R. pallescens* age class. Age class ID key: A = adult; N1–N5 = nymphal stages 1–5; Unk = unknown.

CHAPTER 3

REGIONAL BIOGEOGRAPHY OF MICROBIOTA COMPOSITION IN THE CHAGAS DISEASE VECTOR *RHODNIUS PALLESCENS*[*]

[*] Kieran TJ, Arnold KMH, Thomas IV JC, Varian CP, Saldana A, Calzada JE, Glenn TC, Gottdenker NL. Submitted to: *Parasites & Vectors*.

Abstract

Triatomine bugs are vectors of the protozoan parasite *Trypanosoma cruzi*, which causes Chagas disease. *Rhodnius pallescens* is a major vector of Chagas disease in Panama. Understanding the microbial ecology of disease vectors is important in the development of vector management strategies that target vector survival and fitness. In this study we examine the whole-body microbial composition of *R. pallescens* from three locations in Panama. We collected 89 *R. pallescens* specimens using Noireau traps in *Attalea buytracea* palms. We then extracted total DNA from whole-bodies of specimens and amplified bacterial microbiota using 16S rRNA metabarcoding PCR. The 16S libraries were sequenced on an Illumina MiSeq and analyzed using QIIME2 software. We found Proteobacteria, Actinobacteria, Bacteroidetes, and Firmicutes to be the most abundant bacterial phyla across all samples. Geographic location showed the largest difference in microbial composition with northern Veraguas Province having the most diversity and Panama Oeste Province localities being most similar to each other. *Wolbachia* was detected in high abundance (48-72%) at Panama Oeste area localities with a complete absence of detection in Veraguas Province. No significant differences in microbial composition were detected between triatomine age class, primary blood meal source, or *T. cruzi* infection status. We found biogeographic regions differ in microbial composition among *R. pallescens* populations in Panama. While overall the microbiota has bacterial taxa consistent with previous studies in triatomine microbial ecology, locality differences are an important observation for future studies. Geographic heterogeneity in microbiomes of vectors is an important consideration for future developments that leverage microbiomes for disease control.

Introduction

Insect microbiota are composed of a wide variety of microbial species (Engel & Moran, 2013; Moran, McCutcheon, & Nakabachi, 2008), that serve as commensals, pathogens, or have mutualistic benefits that impact the reproduction, nutrition, and immune systems of the insect host (Engel & Moran, 2013; Moran et al., 2008; Oliver & Martinez, 2014; Weiss & Aksoy, 2011). The symbiotic relationship between an insect disease vector and its microbiota can have an important influence on the competence and transmission potential of human diseases (Cirimotich, Ramirez, & Dimopoulos, 2011; Weiss & Aksoy, 2011), including in blood feeding species (Minard, Mavingui, & Moro, 2013; Oliver & Martinez, 2014). The composition of microbial species associated with arthropod vectors of infectious diseases may also impact vector competence, increasing or decreasing pathogen transmission from vector to host (Finney, Kamhawi, & Wasmuth, 2015). Insect microbiota research can lead to improved methods of vector control (Crotti et al., 2012; M. A. Saldana, Hegde, & Hughes, 2017), but limited research, often with conflicting results, leaves many questions unanswered (Finney et al., 2015). Vector life stage, distribution, species, methods/sampling strategies, and environment (e.g., habitat type or geographic region) may influence vector microbiota (Finney et al., 2015).

In this study, we evaluate patterns of whole-body microbiota of a Chagas disease vector. Chagas disease, caused by the kinetoplastid protozoan parasite *Trypanosoma cruzi*, is transmitted between a wide range of potential mammalian hosts and humans by hematophagous (blood feeding) triatomine insect vectors. Despite widespread control programs, Chagas disease remains a significant health threat to millions of inhabitants in

Latin America, particularly those that live in poverty (Gurtler, Kitron, Cecere, Segura, & Cohen, 2007). The idea of using bacterial symbionts of triatomine bugs to control Chagas disease has long been proposed (Beard, 2002). Recent studies describe microbial community composition within triatomines (Castro et al., 2012; da Mota et al., 2012; Diaz, Villavicencio, Correia, Costa, & Haag, 2016; Dumonteil et al., 2018; Orantes et al., 2018; Rodriguez-Ruano et al., 2018; Waltmann et al., 2019), including *R. pallescens* from Colombia (Montoya-Porras, Omar, Alzate, Moreno-Herrera, & Cadavid-Restrepo, 2018) and Panama (Espino et al., 2009), and a sister species, *R. prolixus* (Vieira et al., 2015). Studies thus far describe triatomine microbiota as having low complexity in terms of diversity and species-specific patterns (da Mota et al., 2012; Diaz et al., 2016; Rodriguez-Ruano et al., 2018), yet the microbiota for many taxa remain to be studied. Infection of triatomines with trypanosomes has been associated with reduction in gut microbial diversity (Castro et al., 2012; Vieira et al., 2015), and blood meal identity may influence composition of the predominant bacterial taxa (Diaz et al., 2016; Dumonteil et al., 2018). However, other important comparisons among triatomines are lacking, such as differences between location and habitat type. Microbial composition variation between different geographic locations has been observed in ticks (Carpi et al., 2011; Fryxell & DeBruyn, 2016; Gurfield, Grewal, Cua, Torres, & Kelley, 2017; Van Treuren et al., 2015; Williams-Newkirk, Rowe, Mixson-Hayden, & Dasch, 2014) and with mixed observations in mosquitoes (Coon, Brown, & Strand, 2016; Novakova et al., 2017). Habitat has also been shown to be a main driver of microbial species composition in mosquitos (Dada et al., 2014; Dickson et al., 2017).

There more than 150 species of triatomines, with different distributions, habitat requirements, life histories, and vectoral capacities that can impact microbiota. This complexity requires extensive research into different triatomine microbiomes. Currently, we still lack basic microbial community composition descriptions for many triatomine species and these large gaps in our knowledge make informed research for vector biocontrol difficult. Therefore, advancing research in triatomine microbiota is crucial for gaining a better understanding of *T. cruzi* infection, triatomine vector capacity, and developmental biology.

Here, we describe the whole-body bacterial microbiota of wild caught *R. pallescens* from three separate geographic locations in Panama. We used the entire triatomine body to encompass all potential microbial taxa relevant to *R. pallescens* that could affect their fitness and survival as a benchmark for future studies of localized anatomy microbiota. We hypothesize that both habitat type and geographic location will be associated with differences in whole-body microbiota composition. We further hypothesize that complex environments, such as forest patches, will be associated with a more diverse microbiota composition than more homogeneous environments, such as cattle pastures. We use Illumina 16S rRNA amplicon sequencing to characterize and evaluate the bacterial microbiota of *R. pallescens* between different locations and habitats, comparing infection status, age class, and primary blood meal source to evaluate a range of variables that may be associated with whole-body bacterial community composition.

Methods

Sample Collection & DNA Extraction

All *R. pallescens* evaluated specimens (N=89) were collected in Panama, Central America using Noireau traps (Noireau et al., 2002) in *Attalea butryacea* palms (the main habitat of this species) and placed directly in 95% molecular grade ethanol before use. We sampled from a total of 8 palms, in three habitats (pasture, peridomestic, peridomestic-forest), from three geographic locations in lowland moist tropical forest (Las Pavas, Trinidad de las Minas) and moist tropical forest (Santa Fe, Veraguas) (Figure 3.1). We consider peridomestic to be home yards or areas within 100 meters of a dwelling and peridomestic-forest to be patches of regenerated forest within a peridomestic landscape matrix. Samples from Las Pavas, La Chorrera District (N 9.104167°, W 79.885833°) (N=27 from two habitats) and Trinidad de las Minas, Capira District (N 8.775556°, W 79.995833°) (N=32, from one habitat) were from a previous study examining bloodmeals (Kieran et al., 2017). We further collected 30 samples from four sites comprising three habitats located in Santa Fe District, Veraguas (N 8.509232°, W 81.077800°) from 8-11 July 2017. The Santa Fe region has recently been described as a new endemic focus for Chagas disease in Panama, where a dark morph of *R. pallescens* predominates (A. Saldana et al., 2012; A. Saldana et al., 2018). All specimens were nymphs, primarily N3 and below (92%, 95% CI, 84.4-96.39%), with the exception of one male from Trinidad de las Minas (Appendix A). DNA was extracted from whole specimens following Kieran and colleagues (Kieran et al., 2017). Briefly, samples were macerated and digested overnight in digest buffer with Protenaise K and extracted with Phenol-Chloroform-Isoamyl alcohol. Extractions were reconstituted in TLE buffer (10

mM Tris pH 8, 0.1 mM EDTA), and impurities were removed with Sera-Mag SpeedBeads™ (Fisher Scientific, Waltham, MA; (Faircloth & Glenn, 2012)) with a final reconstitution in 30 µL TLE.

DNA Amplification & Sequencing

We amplified bacterial 16S rRNA DNA using the S-D-Bact-0341-b-S-17 (5'-CCTACGGGNGGCWGCAG-3') forward and S-D-Bact-0785-a-A-21 (5'-GACTACHVGGGTATCTAATCC-3') reverse primer pair (Klindworth et al., 2013) to which we added a modifications following previous studies (Glenn et al., 2019. https://www.biorxiv.org/content/10.1101/619544v1; Kieran et al., 2017; Wang, Tang, Glenn, & Wang, 2016). We added Illumina TruSeq sequences to the 5' end of the forward (Read 1) and reverse (Read 2) primer creating fusion primers. We synthesized 8 forward and 12 reverse fusion primers, each with a unique variable length (5-8bp) index sequence between the 16S and TruSeq sequences. We then performed two rounds of PCR. For the first-round we performed replicate PCRs using 12.5 µl reactions of KAPA HiFi HotStart Kits (Kapa Biosystems, Wilmington, MA) consisting of 2.5 µl of 5× Buffer, 0.375 µl of 10 mM dNTPs, 0.25 µl hot start Taq, 5.4 µl molecular grade water, 1 µl of 5µM forward primer, 1 ul of 5 µM reverse primer, and 2 µl of DNA. The ranges of DNA sample concentrations were from 10 ng/µl to 60 ng/µl. Each DNA sample had a unique primer-index combination with the following thermocycler conditions: 98°C for 3 min, followed by 30 cycles at 95°C for 30s, 63°C for 1 min, 72°C for 1 min and a final extension at 72°C for 5 min. Amplification success was verified on a 1.5% agarose gel.

Amplicons were pooled in equal concentrations and cleaned using a 1:1 ratio of SPRI-beads and reconstituted in 25 µl TLE. The second-round PCR primers consisted of Illumina TruSeqHT compatible 8 nt indexed primers (Glenn et al., 2019). We used 25 µl reaction of KAPA HiFi HotStart Kits using 5 µl of 5× Buffer, 0.75 µl of 10 mM dNTPs, 0.5 µl HotStart, 3.75 µl molecular grade water, 2.5 µl of 5 µM forward primer, 2.5 µl 5 µM reverse primer, and 10 µl of 16S amplicon pool. We performed two replicate PCRs with the following thermocycler conditions: 98°C for 2 min, followed by 10 cycles at 98°C for 30 s, 60°C for 30 s, 72°C for 30 s and a final extension at 72°C for 5 min. Library product was cleaned with Sera-Mag SpeedBeads™ (1:1 ratio) and pooled with other uniquely indexed samples prior to sequencing.

For blood meal source data, we used previous data from Kieran and colleagues (Kieran et al., 2017) and for newly collected samples, we amplified 12S rRNA DNA following Kieran and colleagues (Kieran et al., 2017). All libraries were sent to the Georgia Genomics and Bioinformatics Core (http://dna.uga.edu) for sequencing on an Illumina MiSeq using a v3 PE300 kit (Illumina, San Diego, CA). We also screened for the presence of *T. cruzi* and *T. rangeli* amplifying telomeric kinetoplastid DNA with Tc189 and Tr primers (Chiurillo et al., 2003). Samples were also verified for *T. cruzi* using 121/122 primers targeting the kinetoplastid minicircle (Wincker et al., 1994). Amplification success was verified on a 1.5% agarose gel.

Data Processing & Analysis

We demultiplexed the amplicon indices using Mr. Demuxy 1.2.0 (https://pypi.org/project/Mr_Demuxy/) and resulting fastq files were imported into

Geneious 10.0.1 ((Kearse et al., 2012); https://www.geneious.com) where we trimmed primers, paired and merged the reads using FLASH (Magoc & Salzberg, 2011). Subsequent data were exported as fastq files for importation into Qiime2 (Bolyen et al., 2018). The quality of the sequences was checked and filtered using QIIME2 v. 2018.8 plugin DADA2 (Callahan et al., 2016), and chimeric sequences were removed. The remaining forward sequences were truncated to a final length of 292 and the reverse sequences were truncated to a final length of 240. Amplicon sequence variants (ASV) were analyzed using the q2-diversity Qiime2 plugin to calculate multiple alpha diversity metrics, including Shannon's index H', Simpson's index D_s, Chao1, faith's phylogenetic diversity, and observed-ASV's.

The Qiime2 plugin q2-phylogeny was used to complete a multiple sequence alignment and create rooted and unrooted phylogenetic trees from the filtered alignment. Alpha and beta diversity metrics were assigned using the q2-diversity Qiime2 plugin. An alpha rarefaction was used to evaluate sampling depth, and the data was rarefied at 1000 sequences per sample, removing two samples and retaining 87 samples for final analyses. The Qiime2 plugin q2-feature-classifier was used to align the sequences against the Greengenes 13.8 database (DeSantis et al., 2006). OTUs were identified from phyla down to genera level, we removed archaea, chloroplasts, mitochondria, not available (NAs), and uncharacterized taxa at the kingdom level.

Alpha diversity (species diversity) was calculated using Shannon (species richness), Simpson's (evenness or relative abundance), and Chao1 (estimate of diversity from abundance) diversity metrics. To compare alpha diversities from individuals across location, habitat type, and infection status, a one-way analysis of variance (ANOVA) and

post-hoc Tukey's honest significant difference tests for multiple comparisons were performed to evaluate differences in taxonomic abundance and alpha diversities. A *p* value less than 0.05 was considered statistically significant.

Beta diversity (compositional variation) was calculated for the whole-body microbiota comparison between triatomines across location, habitat type, and infection status using Bray-Curtis dissimilarity. Bray-Curtis is based on shared OTU counts between individuals. Finally, we used nonmetric multidimensional scaling (nMDS) (Kruskal, 1964), to visualize differences between the microbial communities, and a permutational MANOVA for hypothesis testing (Anderson, 2001). All diversity analyses and visualizations were conducted using qiime2 artifact outputs in R (v. 3.5.1) and with the packages *phyloseq* (McMurdie & Holmes, 2013)*, vegan* (Dixon, 2003)*, dplyr* (Wickham, François, Henry, & Müller, 2018)*, ggplot2* (Wickham, Chang, et al., 2018), *metacoder* (Foster, Sharpton, & Grunwald, 2017).

Results

16S rRNA sequences and classification of entire microbiota community

We obtained a total of 4,995,733 16S rRNA V3-V4 region sequences from 101 samples, including the negative controls. After quality filtering, the number of sequences obtained per sample ranged from 1000 to 34,792 reads, with a mean frequency of 9,811.63. The total number of OTUs within the 89 final samples was 4,033, with the top 4 phyla consisting of Proteobacteria (60.67%), Actinobacteria (16.93%), Bacteroidetes (9.55%), Firmicutes (4.11%) out of the total phyla present in the dataset. *R. pallescens*, in Las Pavas and Trinidad de las Minas respectively, is primarily composed of

Proteobacteria (71.63%, 76.11%), Actinobacteria (6.13%, 15.88%), and Bacteroidetes (13.98%, 2.66%) (Figure 3.2, Appendix B). This contrasts with specimens from northern Veraguas with the most abundant phylum shifting from Proteobacteria (26.43%) to Actinobacteria (27.56%) and introducing more Firmicutes (11.48%). At the family-level (Figure 3.3), the top 3 taxa overall are Anaplasmataceae (45.76%), Pseudonocardiaceae (6.04%), Moraxellaceae (2.77%), although these relative proportions differ by location (Appendix C). Most notably, as seen in Figure 3.3, Anaplasmataceae is the dominant family throughout samples from Las Pavas (48.30%) and Trinidad de las Minas (72.51%) (Figure 3.4, Appendix D) but is not present within samples collected in northern Veraguas. The high abundance of Anaplasmataceae is due to a single genus, *Wolbachia* spp., comprising greater than 70% and 42% of the composition in more than half the specimens from Trinidad de las Minas and Las Pavas, respectively (Table 3.1). No differences in microbial composition were detected between triatomine age class (Shannon, $F = 1.07$, $p > 0.09$) or primary blood meal source (Shannon, $F = 1.07$, $p > 0.38$).

Infection Rates

Trypanosma cruzi and *T. rangeli* were detected in sampled vectors at all locations (Appendix A). Rates of positive *T. cruzi* infection were 7.41% (2/27, 95% CI, 0.96-24.47%) at Las Pavas (N palms = 2), 75% (24/32, 95% CI, 57.67-86.97%) at Trinidad de las Minas (N palms = 2), and 46.67% (14/30, 95% CI, 30.23-63.46%) at Sante Fe District, Veraguas (N palms = 4). For *T. rangeli*, positive detection rates were 66.67% (18/27, 95% CI, 47.71-81.47%) in Las Pavas, 78.13% (25/32, 95% CI, 60.96-89.27%) in

Trinidad de las Minas, and 30% (9/30, 95% CI, 16.52-48.02) in Veraguas. There was an overall rate of coinfections of both trypanosomes of 31.46% (28/89, 95% CI, 22.72-41.73%). Coinfections were less abundant at the lower infection sites of Las Pavas (3.7%, 95% CI, < 0.01-19.8%) and Veraguas (13.33%, 95% CI, 4.7-30.3%) compared to Trinidad de las Minas which had a higher coinfection rate (71.88%, 95% CI, 54.46-84.6%).

Location, Habitat Type, and Infection Status on Microbial Composition

Alpha diversity

Alpha richness between locations was significantly different using three different diversity metrices (Observed ANOVA, $F = 12.16$, $p < 0.001$; Shannon's index H' ANOVA, $F = 11.7$, $p < 0.001$; Chao1 ANOVA, $F = 12.04$, $p < 0.001$). Individuals from Veraguas had significantly greater alpha richness when compared to individuals from Trinidad de las Minas (Observed, Chao1, Shannon TukeyHSD, $p < 0.0001$; Simpson Tukey HSD $p = 0.008$) and Las Pavas (Observed TukeyHSD, $p < 0.0004$, Chao1 TukeyHSD, $p = 0.0004$; Shannon TukeyHSD, $p = 0.0096$). Alpha richness across habitat type showed significance for one metric (Shannon TukeyHSD, $F = 4.72$, $p = 0.011$) between peridomestic and peridomestic-forest types ($p = 0.009$). *Trypanosoma cruzi* infection status, however, was not significantly different (Simpson, $F = 3.54$, $p > 0.063$).

Beta diversity

The community differences (beta diversity) of triatomine microbiota showed significant differences at pasture sites across Trinidad de las Minas and Veraguas (PERMANOVA: $p = 0.001$ using Bray-Curtis dissimilarity indices) and at peridomestic

sites across all three locations (PERMANOVA: $p = 0.001$ using Bray-Curtis dissimilarity indices). Community differences across the three habitat types also showed some significant differences within Veraguas (PERMANOVA: $p = 0.001$ using Bray-Curtis dissimilarity indices) and Las Pavas (PERMANOVA: $p = 0.001$ using Bray-Curtis dissimilarity indices). There was no observed difference between pasture habitat composition between sites (PERMANOVA p = 0.818). When examining *T. cruzi* infection status, the only significant differences in microbial composition between *T. cruzi* positive and negative samples were observed at Las Pavas (PERMANOVA: $p = 0.029$ using Bray-Curtis dissimilarity indices) and between peridomestic habitats among all locations (PERMANOVA: $p = 0.01$ using Bray-Curtis dissimilarity indices). Sites at Trinidad de las Minas (PERMANOVA: $p = 0.65$ using Bray-Curtis dissimilarity indices) and Veraguas (PERMANOVA: $p = 0.76$ using Bray-Curtis dissimilarity indices) did not show significant compositional difference between infected and non-infected samples.

Discussion

Here, we characterized the bacterial microbiota of 87 wild individuals of the Chagas disease vector *Rhodnius pallescens* from three populations in Panama. We explored comparisons in composition between location, microhabitat, nymphal stage, *T. cruzi* infection, and blood meal status. Overall, the microbiota of *R. pallescens* exhibits relatively low complexity in its bacterial composition which is consistent with other triatomine studies (da Mota et al., 2012; Diaz et al., 2016; Rodriguez-Ruano et al., 2018). Proteobacteria has also been found to be the most abundant phylum in other vector species (Azambuja, Garcia, & Ratcliffe, 2005; Boissiere et al., 2012; Engel & Moran,

2013; Gumiel et al., 2015; Vivero, Jaramillo, Cadavid-Restrepo, Soto, & Herrera, 2016; Yun et al., 2014) including the triatomines *R. neglectus, R. prolixus, Triatoma vitticeps, T. infestans, T. brasiliensis, T. pseudomaculata, Dipetalogaster maximus*, and *Panstrongylus megistus* (da Mota et al., 2012; Diaz et al., 2016; Gumiel et al., 2015), while a predominance of Actinobacteria has been found previously in *R. pallescens* (Montoya-Porras et al., 2018), both consistent with our findings.

Common bacterial genera found in other triatomines include *Burkholderia*, *Dietzia*, *Gordonia*, *Williamsia* (Diaz et al., 2016; Gumiel et al., 2015; Montoya-Porras et al., 2018), *Actinomycetospora*, *Arsenophonus*, *Corynebacterium*, *Rhodococcus*, *Staphylococcus* (Diaz et al., 2016; Dumonteil et al., 2018; Rodriguez-Ruano et al., 2018), and *Enterococcus*, *Enterobacteriaceae*, *Bacillus* (Rodriguez-Ruano et al., 2018; Waltmann et al., 2019). Of these only *Actinomycetospora* was one of the top 20 genera found across all studied sites (Appendix D). *Enterobacteriaceae* and *Bacillus* were found at all sites, but at much lower abundance (0.34-6.03% and 0.21-0.76% respectively) than found by Waltmann and colleagues (Waltmann et al., 2019). *Dietzia* and *Gordonia* were each in the top 20 taxa for Las Pavas and Trinidad de las Minas (Appendix D). *Arsenophonus* was detected a very low abundance in only a single specimen from northern Veraguas province. All other taxa were found across all sites, but at lower abundance than other studies.

High levels of Proteobacteria observed in specimens from Panama Oeste province localities are due to the very high levels of *Wolbachia* sp. Specimens from northern Veraguas province, interestingly, do not have any *Wolbachia* present. It has been estimated that *Wolbachia* infects 52% of all insect species (Sazama, Bosch, Shouldis,

Ouellette, & Wesner, 2017) and can infect a high proportion of the number of individuals in a species (Jiggins, Hurst, Schulenburg, & Majerus, 2001). However, while many arthropod species may be infected with *Wolbachia*, a majority of the individuals within a species may not be. In a comparative study, Sazama and colleagues (Sazama et al., 2017) found that less than half the individuals were infected in most (69%) *Wolbachia* infected species. *Wolbachia* has been found previously in *Rhodnius* sp. (da Mota et al., 2012; Espino et al., 2009), and is common in hematophagous insects (Jimenez-Cortes et al., 2018), but has not been found in other triatomines (da Mota et al., 2012; Dumonteil et al., 2018; Gumiel et al., 2015). In triatomines (da Mota et al., 2012) and sandflies (Fraihi et al., 2017; Monteiro et al., 2016; Vivero et al., 2016), the role of *Wolbachia* remains unknown. In mosquitoes, *Wolbachia* can affect reproduction and insecticide resistance among others (Minard et al., 2013) creating opportunities for vector biocontrol. However, without further understanding of the role of *Wolbachia* in triatomines, and further research on characterizing microbiomes under various environmental conditions paired with functional analysis of microbial taxa, is much needed for further identification of microbes that may serve as effective control agents for triatomines.

In our study, geographic location was associated with differences between microbial communities of *R. pallescens*. This observation is most evident between the two most disparate geographical locations (northern Veraguas vs Panama Oeste localities). This observation may be the result of quite different environments between these locations. Veraguas province is located in the highlands of the western isthmus of Panama where the climate is cooler (mean 21°C) and more humid with mountainous topography. In the two Panama Oeste area locations (Las Pavas, Trinidad de las Minas)

the topography is flatter with warmer temperatures (mean 27°C). Of particular interest is that the evaluated specimens from Santa Fe District in northern Veraguas correspond with a darker chromatic variation of *R. pallescens* infected by specific genetic groups of *Trypanosoma rangeli* and *T. cruzi* (A. Saldana et al., 2018). Although the genetic characteristics of this population have not been studied, the reported phenotypic differences and the differences found in their microbial composition could be explained by the presence of this dark chromatic variant in this geographical region. However, it is not known if this dark variant represents a separate geographical population, a new subspecies, or a new separate species of *R. pallescens*.

Geographic differences in microbiota have been observed in ticks (Carpi et al., 2011; Fryxell & DeBruyn, 2016; Gurfield et al., 2017; Van Treuren et al., 2015; Williams-Newkirk et al., 2014), but not in mosquitoes where species-specific microbiota is thought to be stable (Novakova et al., 2017). One study in triatomines did not observe any difference in the microbiota between three distant locations in the southern US for *Triatoma protracta* (Rodriguez-Ruano et al., 2018), which is a trend that has been confirmed (Sudakaran, Salem, Kost, & Kaltenpoth, 2012) and opposed (Welch, Macias, & Bextine, 2015) in other hemipterans. As in other insect taxa, clade-wide stability of microbiota with regards to one variable or another does not appear to be consistent. Generalizations about geographic variation in insect microbiomes will have to remain at the species level for now. However, this is still very much an open question in triatomine microbiota research.

Contrary to our expectations, there was no significant difference in bacterial community composition between *T. cruzi* infected and uninfected individuals. This

contrasts with previous studies that have found significant differences between *T. cruzi* positive and negative microbiomes (Diaz et al., 2016; Rodriguez-Ruano et al., 2018). However, our small sample size, limited number of palms sampled (N = 8) with skewed infection ratios, and a skewed abundance of younger stage nymphs (N1-N3), may confound true observable differences. Furthermore, detectable levels of differences may be localized to a portion of the triatomine gut where *T. cruzi* develops and deserves further study and experimental controls. A similar situation may occur during infection with *T. rangeli*, which can colonize not only the insect's intestine but also the hemocoel and salivary glands (Azambuja & Garcia, 2005). We also did not find any significant difference in the microbiota as a result of the dominant blood meal source found, as previously observed (Dumonteil et al., 2018). This could be due to the complexity of variables that influence the microbiota. While blood meal source potentially has an effect on the bacterial composition in the gut, these samples often have mixed blood meal sources with differing amounts of abundance (Appendix A; (Kieran et al., 2017)), making discrete differences difficult to observe. Habitat was found to be different between peridomestic and peridomestic-forest, however this observation represents a few peridomestic-forest samples from northern Veraguas only and likely is an artifact of location difference. Similarly, age class showed no differences, but since this dataset is highly skewed toward a couple of nymphal classes, distinctions are impossible to detect. This study examined the whole-body microbiota, which may obscure more anatomically localized differences observed in other studies and, on the whole, result in a more holistic microbial composition where local environment has a bigger impact.

Conclusion

In conclusion, we examined the whole-body microbiota of *Rhodnius pallescens*, which can serve as a benchmark for future comparative studies examining the microbiota of specific organs or anatomical regions. Interestingly, the largest difference in *R. pallescens* microbial community composition was between geographic location. While we did not find any definitive differences between other variables (e. g. habitat type, age class, blood meal, infection status) these remain important aspects of vector biology that require further study. The effects of geographic environmental diversity can be minimized through the use of more comparative studies using laboratory reared insects and controlled studies to tease apart more complex variables such as blood meals and infection status.

Acknowledgements

We thank Franklyn Samudio and Jose Montenegro for assistance in the field. The authors also thank the Ministerio de Ambiente in Panama (MiAMBIENTE), the Smithsonian Tropical Research Institute (STRI), and the University of Georgia Graduate School.

References

Anderson, M. J. (2001). A new method for non-parametric multivariate analysis of variance. *Austral Ecology, 26*(1), 32-46. doi:DOI 10.1111/j.1442-9993.2001.01070.pp.x

Azambuja, P., & Garcia, E. S. (2005). *Trypanosoma rangeli* interactions within the vector *Rhodnius prolixus* - A mini review. *Mem Inst Oswaldo Cruz, 100*(5), 567-571. doi:Doi 10.1590/S0074-02762005000500019

Azambuja, P., Garcia, E. S., & Ratcliffe, N. A. (2005). Gut microbiota and parasite transmission by insect vectors. *Trends Parasitol, 21*(12), 568-572. doi:10.1016/j.pt.2005.09.011

Beard, C. B. (2002). Bacterial symbionts of the Triatominae and their potential use in control of Chags disease transmission. *Annu Rev Entomol, 47*.

Boissiere, A., Tchioffo, M. T., Bachar, D., Abate, L., Marie, A., Nsango, S. E., . . . Morlais, I. (2012). Midgut microbiota of the malaria mosquito vector *Anopheles gambiae* and interactions with *Plasmodium falciparum* infection. *PLoS Pathog, 8*(5), e1002742. doi:10.1371/journal.ppat.1002742

Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C., Al-Ghalith, G. A., . . . Caporaso, J. G. (2018). Qiime2: Reproducible, interactive, scalable, and extensible microbiome data science. *PeerJ Preprints*. doi:10.7287/peerj.preprints.27295v2

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods, 13*(7), 581-583. doi:10.1038/nmeth.3869

Carpi, G., Cagnacci, F., Wittekindt, N. E., Zhao, F., Qi, J., Tomsho, L. P., . . . Schuster, S. C. (2011). Metagenomic profile of the bacterial communities associated with *Ixodes ricinus* ticks. *PLoS One, 6*(10), e25604. doi:10.1371/journal.pone.0025604

Castro, D. P., Moraes, C. S., Gonzalez, M. S., Ratcliffe, N. A., Azambuja, P., & Garcia, E. S. (2012). *Trypanosoma cruzi* immune response modulation decreases microbiota in *Rhodnius prolixus* gut and is crucial for parasite survival and development. *PLoS One, 7*(5), e36591. doi:10.1371/journal.pone.0036591

Chiurillo, M. A., Crisante, G., Rojas, A., Peralta, A., Dias, M., Guevara, P., . . . Ramirez, J. L. (2003). Detection of *Trypanosoma cruzi* and *Trypanosoma rangeli* infection by duplex PCR assay based on telomeric sequences. *Clinical and Diagnostic Lanoratory Immunology, 10*(5), 775-779.

Cirimotich, C. M., Ramirez, J. L., & Dimopoulos, G. (2011). Native microbiota shape insect vector competence for human pathogens. *Cell Host Microbe, 10*(4), 307-310. doi:10.1016/j.chom.2011.09.006

Coon, K. L., Brown, M. R., & Strand, M. R. (2016). Mosquitoes host communities of bacteria that are essential for development but vary greatly between local habitats. *Molecular Ecology, 25*(22), 5806-5826. doi:10.1111/mec.13877

Crotti, E., Balloi, A., Hamdi, C., Sansonno, L., Marzorati, M., Gonella, E., . . . Daffonchio, D. (2012). Microbial symbionts: a resource for the management of

insect-related problems. *Microb Biotechnol, 5*(3), 307-317. doi:10.1111/j.1751-7915.2011.00312.x

da Mota, F. F., Marinho, L. P., Moreira, C. J., Lima, M. M., Mello, C. B., Garcia, E. S., . . . Azambuja, P. (2012). Cultivation-independent methods reveal differences among bacterial gut microbiota in triatomine vectors of Chagas disease. *PLoS Negl Trop Dis, 6*(5), e1631. doi:10.1371/journal.pntd.0001631

Dada, N., Jumas-Bilak, E., Manguin, S., Seidu, R., Stenstrom, T. A., & Overgaard, H. J. (2014). Comparative assessment of the bacterial communities associated with *Aedes aegypti* larvae and water from domestic water storage containers. *Parasit Vectors, 7*, 391. doi:10.1186/1756-3305-7-391

DeSantis, T. Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E. L., Keller, K., . . . Andersen, G. L. (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol, 72*(7), 5069-5072. doi:10.1128/AEM.03006-05

Diaz, S., Villavicencio, B., Correia, N., Costa, J., & Haag, K. L. (2016). Triatomine bugs, their microbiota and *Trypanosoma cruzi*: asymmetric responses of bacteria to an infected blood meal. *Parasit Vectors, 9*(1), 636. doi:10.1186/s13071-016-1926-2

Dickson, L. B., Jiolle, D., Minard, G., Moltini-Conclois, I., Volant, S., Ghozlane, A., . . . Lambrechts, L. (2017). Carryover effects of larval exposure to different environmental bacteria drive adult trait variation in a mosquito vector. *Sci Adv, 3*(8), e1700585. doi:10.1126/sciadv.1700585

Dixon, P. (2003). VEGAN, a package of R functions for community ecology. *Journal of Vegetation Science, 14*(6), 927-930. doi:DOI 10.1111/j.1654-1103.2003.tb02228.x

Dumonteil, E., Ramirez-Sierra, M. J., Perez-Carrillo, S., Teh-Poot, C., Herrera, C., Gourbiere, S., & Waleckx, E. (2018). Detailed ecological associations of triatomines revealed by metabarcoding and next-generation sequencing: implications for triatomine behavior and *Trypanosoma cruzi* transmission cycles. *Sci Rep, 8*(1), 4140. doi:10.1038/s41598-018-22455-x

Engel, P., & Moran, N. A. (2013). The gut microbiota of insects - diversity in structure and function. *FEMS Microbiol Rev, 37*(5), 699-735. doi:10.1111/1574-6976.12025

Espino, C. I., Gomez, T., Gonzalez, G., do Santos, M. F., Solano, J., Sousa, O., . . . Osuna, A. (2009). Detection of Wolbachia bacteria in multiple organs and feces of the triatomine insect *Rhodnius pallescens* (Hemiptera, Reduviidae). *Appl Environ Microbiol, 75*(2), 547-550. doi:10.1128/AEM.01665-08

Faircloth, B. C., & Glenn, T. C. (2012). Not all sequence tags are created equal: designing and validating sequence identification tags robust to indels. *PLoS One, 7*(8), e42543. doi:10.1371/journal.pone.0042543

Finney, C. A., Kamhawi, S., & Wasmuth, J. D. (2015). Does the arthropod microbiota impact the establishment of vector-borne diseases in mammalian hosts? *PLoS Pathog, 11*(4), e1004646. doi:10.1371/journal.ppat.1004646

Foster, Z. S., Sharpton, T. J., & Grunwald, N. J. (2017). Metacoder: An R package for visualization and manipulation of community taxonomic diversity data. *PLoS Comput Biol, 13*(2), e1005404. doi:10.1371/journal.pcbi.1005404

Fraihi, W., Fares, W., Perrin, P., Dorkeld, F., Sereno, D., Barhoumi, W., . . . Zhioua, E. (2017). An intergrated overview of the midgut bacterial flora composition of *Phlebotomus pernicoisus*, a vector of zoonotic visceral leishmaniasis in the Western Mediterranean Basin. *PLoS Negl Trop Dis, 11*.

Fryxell, R. T. T., & DeBruyn, J. M. (2016). The Microbiome of Ehrlichia-Infected and Uninfected Lone Star Ticks (*Amblyomma americanum*). *PLoS One, 11*(1), e0146651. doi:10.1371/journal.pone.0146651

Glenn, T. C., Nilsen, R. A., Kieran, T. J., Sanders, J. G., Bayona-Vásquez, N. J., Finger, J. W., . . . Faircloth, B. C. (2019). *Adapterama I: Universal stubs and primers for 384 unique dual-indexed or 147,456 combinatorially-indexed Illumina libraries (iTru & iNext)*. bioRxiv. Retrieved from https://www.biorxiv.org/content/10.1101/049114v2

Glenn, T. C., Pierson, T. W., Bayona-Vásquez, N. J., Kieran, T. J., Hoffberg, S. L., Thomas, J. C., . . . Faircloth, B. C. (2019. https://www.biorxiv.org/content/10.1101/619544v1). *Adapterama II: Universal amplicon sequencing on Illumina platforms (TaggiMatrix)*. bioRxiv. Retrieved from https://www.biorxiv.org/content/10.1101/619544v1

Gumiel, M., da Mota, F. F., Rizzo Vde, S., Sarquis, O., de Castro, D. P., Lima, M. M., . . . Azambuja, P. (2015). Characterization of the microbiota in the guts of *Triatoma brasiliensis* and *Triatoma pseudomaculata* infected by *Trypanosoma cruzi* in

natural conditions using culture independent methods. *Parasit Vectors, 8*, 245. doi:10.1186/s13071-015-0836-z

Gurfield, N., Grewal, S., Cua, L. S., Torres, P. J., & Kelley, S. T. (2017). Endosymbiont interference and microbial diversity of the Pacific coast tick, *Dermacentor occidentalis*, in San Diego County, California. *PeerJ, 5*, e3202. doi:10.7717/peerj.3202

Gurtler, R. E., Kitron, U., Cecere, M. C., Segura, E. L., & Cohen, J. E. (2007). Sustainable vector control and management of Chagas disease in the Gran Chaco, Argentina. *Proc Natl Acad Sci U S A, 104*(41), 16194-16199.

Jiggins, F. M., Hurst, G. D. D., Schulenburg, J. H. G. V. D., & Majerus, M. E. N. (2001). Two male-killing Wolbachia strains coexist within a population of the butterfly *Acraea encedon*. *Heredity, 86*, 161-166. doi:DOI 10.1046/j.1365-2540.2001.00804.x

Jimenez-Cortes, J. G., Garcia-Contreras, R., Bucio-Torres, M. I., Cabrera-Bravo, M., Cordoba-Aguilar, A., Benelli, G., & Salazar-Schettino, P. M. (2018). Bacterial symbionts in human blood-feeding arthropods: Patterns, general mechanisms and effects of global ecological changes. *Acta Trop, 186*, 69-101. doi:10.1016/j.actatropica.2018.07.005

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., . . . Drummond, A. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics, 28*(12), 1647-1649. doi:10.1093/bioinformatics/bts199

Kieran, T. J., Gottdenker, N. L., Varian, C. P., Saldana, A., Means, N., Owens, D., . . . Glenn, T. C. (2017). Blood Meal Source Characterization Using Illumina Sequencing in the Chagas Disease Vector *Rhodnius pallescens* (Hemiptera: Reduviidae) in Panama. *J Med Entomol, 54*(6), 1786-1789. doi:10.1093/jme/tjx170

Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M., & Glockner, F. O. (2013). Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res, 41*(1), e1. doi:10.1093/nar/gks808

Kruskal, J. B. (1964). Multidimensional Scaling by Optimizing Goodness-of-Fit to a Nonmetric Hypothesis. *Psychometrika, 29*, 1-28.

Magoc, T., & Salzberg, S. L. (2011). FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics, 27*(21), 2957-2963. doi:10.1093/bioinformatics/btr507

McMurdie, P. J., & Holmes, S. (2013). phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One, 8*(4), e61217. doi:10.1371/journal.pone.0061217

Minard, G., Mavingui, P., & Moro, C. V. (2013). Diversity and function of bacterial microbiota in the mosquito holobiont. *Parasit Vectors, 6*, 146. doi:10.1186/1756-3305-6-146

Monteiro, C. C., Villegas, L. E. M., Campolina, T. B., Pires, A. C. M. A., Miranda, J. C., Pimenta, P. F. P., & Secundino, N. F. C. (2016). Bacterial diversity of the

American sand fly *Lutzomyia intermedia* using high-throughput metagenomic sequencing. *Parasit Vectors, 9*. doi:ARTN 480 10.1186/s13071-016-1767-z

Montoya-Porras, L. M., Omar, T. C., Alzate, J. F., Moreno-Herrera, C. X., & Cadavid-Restrepo, G. E. (2018). 16S rRNA gene amplicon sequencing reveals dominance of Actinobacteria in *Rhodnius pallescens* compared to *Triatoma maculata* midgut microbiota in natural populations of vector insects from Colombia. *Acta Trop, 178*, 327-332. doi:10.1016/j.actatropica.2017.11.004

Moran, N. A., McCutcheon, J. P., & Nakabachi, A. (2008). Genomics and evolution of heritable bacterial symbionts. *Annu Rev Genet, 42*, 165-190. doi:10.1146/annurev.genet.41.110306.130119

Noireau, F., Abad-Franch, F., Valente, S. A., Dias-Lima, A., Lopes, C. M., Cunha, V., . . . Jurberg, J. (2002). Trapping Triatominae in silvatic habitats. *Mem Inst Oswaldo Cruz, 97*(1), 61-63.

Novakova, E., Woodhams, D. C., Rodriguez-Ruano, S. M., Brucker, R. M., Leff, J. W., Maharaj, A., . . . Scott, J. (2017). Mosquito Microbiome Dynamics, a Background for Prevalence and Seasonality of West Nile Virus. *Front Microbiol, 8*, 526. doi:10.3389/fmicb.2017.00526

Oliver, K. M., & Martinez, A. J. (2014). How resident microbes modulate ecologically-important traits of insects. *Curr Opin Insect Sci, 4*, 1-7. doi:10.1016/j.cois.2014.08.001

Orantes, L. C., Monroy, C., Dorn, P. L., Stevens, L., Rizzo, D. M., Morrissey, L., . . . Helms Cahan, S. (2018). Uncovering vector, parasite, blood meal and microbiome

patterns from mixed-DNA specimens of the Chagas disease vector *Triatoma dimidiata*. *PLoS Negl Trop Dis, 12*(10), e0006730. doi:10.1371/journal.pntd.0006730

Rodriguez-Ruano, S. M., Skochova, V., Rego, R. O. M., Schmidt, J. O., Roachell, W., Hypsa, V., & Novakova, E. (2018). Microbiomes of North American Triatominae: The Grounds for Chagas Disease Epidemiology. *Front Microbiol, 9*, 1167. doi:10.3389/fmicb.2018.01167

Saldana, A., Pineda, V., Martinez, I., Santamaria, G., Santamaria, A. M., Miranda, A., & Calzada, J. E. (2012). A new endemic focus of Chagas disease in the northern region of Veraguas Province, Western Half Panama, Central America. *PLoS One, 7*(4), e34657. doi:10.1371/journal.pone.0034657

Saldana, A., Santamaria, A. M., Pineda, V., Vasquez, V., Gottdenker, N. L., & Calzada, J. E. (2018). A darker chromatic variation of *Rhodnius pallescens* infected by specific genetic groups of *Trypanosoma rangeli* and *Trypanosoma cruzi* from Panama. *Parasit Vectors, 11*(1), 423. doi:10.1186/s13071-018-3004-4

Saldana, M. A., Hegde, S., & Hughes, G. L. (2017). Microbial control of arthropod-borne disease. *Mem Inst Oswaldo Cruz, 112*(2), 81-93. doi:10.1590/0074-02760160373

Sazama, E. J., Bosch, M. J., Shouldis, C. S., Ouellette, S. P., & Wesner, J. S. (2017). Incidence of Wolbachia in aquatic insects. *Ecol Evol, 7*(4), 1165-1169. doi:10.1002/ece3.2742

Sudakaran, S., Salem, H., Kost, C., & Kaltenpoth, M. (2012). Geographical and ecological stability of the symbiotic mid-gut microbiota in European firebugs,

Pyrrhocoris apterus (Hemiptera, Pyrrhocoridae). *Mol Ecol, 21*(24), 6134-6151. doi:10.1111/mec.12027

Van Treuren, W., Ponnusamy, L., Brinkerhoff, R. J., Gonzalez, A., Parobek, C. M., Juliano, J. J., . . . Meshnick, S. R. (2015). Variation in the Microbiota of Ixodes Ticks with Regard to Geography, Species, and Sex. *Appl Environ Microbiol, 81*(18), 6200-6209. doi:10.1128/AEM.01562-15

Vieira, C. S., Mattos, D. P., Waniek, P. J., Santangelo, J. M., Figueiredo, M. B., Gumiel, M., . . . Azambuja, P. (2015). *Rhodnius prolixus* interaction with *Trypanosoma rangeli*: modulation of the immune system and microbiota population. *Parasit Vectors, 8*, 135. doi:10.1186/s13071-015-0736-2

Vivero, R. J., Jaramillo, N. G., Cadavid-Restrepo, G., Soto, S. I., & Herrera, C. X. (2016). Structural differences in gut bacteria communities in developmental stages of natural populations of *Lutzomyia evansi* from Colombia's Caribbean coast. *Parasit Vectors, 9*, 496. doi:10.1186/s13071-016-1766-0

Waltmann, A., Willcox, A. C., Balasubramanian, S., Borrini Mayori, K., Mendoza Guerrero, S., Salazar Sanchez, R. S., . . . Bowman, N. M. (2019). Hindgut microbiota in laboratory-reared and wild *Triatoma infestans*. *PLoS Negl Trop Dis, 13*(5), e0007383. doi:10.1371/journal.pntd.0007383

Wang, J., Tang, L., Glenn, T. C., & Wang, J. S. (2016). Aflatoxin B1 Induced Compositional Changes in Gut Microbial Communities of Male F344 Rats. *Toxicol Sci, 150*(1), 54-63. doi:10.1093/toxsci/kfv259

Weiss, B., & Aksoy, S. (2011). Microbiome influences on insect host vector competence. *Trends Parasitol, 27*(11), 514-522. doi:10.1016/j.pt.2011.05.001

Welch, E. W., Macias, J., & Bextine, B. (2015). Geographic patterns in the bacterial microbiome of the glassy-winged sharpshooter, *Homalodisca vitripennis* (Hemiptera: Cicadellidae). *Symbiosis, 66*(1), 1-12. doi:10.1007/s13199-015-0332-4

Wickham, H., Chang, W., Henry, L., Pedersen, T. L., Takahashi, K., Wilke, C., & Woo, K. (2018). ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics. R package. Retrieved from https://CRAN.R-project.org/package=ggplot2

Wickham, H., François, R., Henry, L., & Müller, K. (2018). dplyr: A Grammar of Data Manipulation. R package. Retrieved from https://CRAN.R-project.org/package=dplyr

Williams-Newkirk, A. J., Rowe, L. A., Mixson-Hayden, T. R., & Dasch, G. A. (2014). Characterization of the bacterial communities of life stages of free living lone star ticks (*Amblyomma americanum*). *PLoS One, 9*(7), e102130. doi:10.1371/journal.pone.0102130

Wincker, P., Britto, C., Pereria, J. B., Cardoso, M. A., Oelemann, W., & Morel, C. M. (1994). Use of a simplified polymerase chain reaction procedure to detect *Trypanosoma cruzi* in blood samples from chronic chagasic patients in a rural endemic area. *Am J Trop Med Hyg, 51*(6), 771-777.

Yun, J. H., Roh, S. W., Whon, T. W., Jung, M. J., Kim, M. S., Park, D. S., . . . Bae, J. W. (2014). Insect gut bacterial diversity determined by environmental habitat, diet, developmental stage, and phylogeny of host. *Appl Environ Microbiol, 80*(17), 5254-5264. doi:10.1128/AEM.01226-14

Tables

Table 3.1. Total number reads and the number and proportion of Wolbachia reads for each sample across all locations.

SampleID	Location	Wolbachia	Total Reads	% Wolbachia
066p	pavas	10,795	12,933	83.47%
065p	pavas	6,831	8,795	77.67%
062p	pavas	4,458	5,916	75.35%
050p	pavas	10,173	13,627	74.65%
061p	pavas	4,631	6,295	73.57%
067p	pavas	5,070	7,035	72.07%
063p	pavas	5,174	7,541	68.61%
011p	pavas	5,049	7,721	65.39%
008p	pavas	4,739	7,914	59.88%
060p	pavas	1,932	3,429	56.34%
064p	pavas	1,864	3,331	55.96%
049p	pavas	4,598	10,264	44.80%
010p	pavas	2,702	6,071	44.51%
007p	pavas	3,414	7,974	42.81%
037p	pavas	3,681	8,683	42.39%
036p	pavas	3,073	8,725	35.22%
051p	pavas	2,733	8,032	34.03%
040p	pavas	2,986	8,918	33.48%
046p	pavas	3,299	9,971	33.09%
038p	pavas	2,632	8,228	31.99%
009p	pavas	2,755	8,763	31.44%
013p	pavas	2,302	7,634	30.15%
039p	pavas	2,283	8,636	26.44%
048p	pavas	1,720	7,168	24.00%
014p	pavas	1,127	5,083	22.17%
045p	pavas	1,571	7,119	22.07%
012p	pavas	1,493	7,618	19.60%
063TM	minas	14,633	17,274	84.71%

064TM	minas	17,233	20,729	83.13%
045TM	minas	16,109	19,730	81.65%
068TM	minas	10,150	12,437	81.61%
048TM	minas	12,607	15,578	80.93%
050TM	minas	12,686	15,744	80.58%
031TM	minas	12,523	15,652	80.01%
030TM	minas	11,151	13,960	79.88%
044TM	minas	13,118	16,488	79.56%
073TM	minas	10,815	13,757	78.61%
065TM	minas	10,663	13,603	78.39%
067TM	minas	12,441	15,967	77.92%
039TM	minas	14,365	18,545	77.46%
035TM	minas	15,411	19,914	77.39%
061TM	minas	9,439	12,357	76.39%
032TM	minas	10,181	13,650	74.59%
036TM	minas	10,761	14,469	74.37%
071TM	minas	9,088	12,263	74.11%
069TM	minas	11,386	15,861	71.79%
040TM	minas	10,531	14,771	71.30%
034TM	minas	12,715	18,134	70.12%
041TM	minas	11,329	16,504	68.64%
037TM	minas	4,893	8,365	58.49%
062TM	minas	3,207	5,831	55.00%

033TM	minas	2,980	5,572	53.48%
029TM	minas	3,474	6,604	52.60%
074TM	minas	4,024	7,661	52.53%
042TM	minas	3,763	7,251	51.90%
047TM	minas	3,390	6,565	51.64%
066TM	minas	3,660	7,244	50.52%
046TM	minas	4,272	11,642	36.69%
070TM	minas	2,572	7,268	35.39%
203	veraguas	0	1,023	0.00%
204	veraguas	0	11,759	0.00%
205	veraguas	0	6,370	0.00%
206	veraguas	0	7,821	0.00%
207	veraguas	0	8,557	0.00%
208	veraguas	0	8,242	0.00%
209	veraguas	0	13,200	0.00%
349	veraguas	0	11,269	0.00%
349.1	veraguas	0	5,935	0.00%
350	veraguas	0	1,079	0.00%
351	veraguas	0	1,084	0.00%
352	veraguas	0	30,654	0.00%
354	veraguas	0	9,178	0.00%
355	veraguas	0	5,028	0.00%
357	veraguas	0	3,674	0.00%
358	veraguas	0	8,169	0.00%
363	veraguas	0	10,931	0.00%

364	veragu as	0	988	0.00%
365	veragu as	0	1,149	0.00%
366	veragu as	0	1,757	0.00%
369	veragu as	0	1,887	0.00%
374	veragu as	0	1,029	0.00%
378	veragu as	0	613	0.00%
386	veragu as	0	16,597	0.00%
387	veragu as	0	6,150	0.00%
388	veragu as	0	34,792	0.00%
390	veragu as	0	12,993	0.00%
392	veragu as	0	17,923	0.00%
394	veragu as	0	10,700	0.00%
404	veragu as	0	7,629	0.00%

Figures

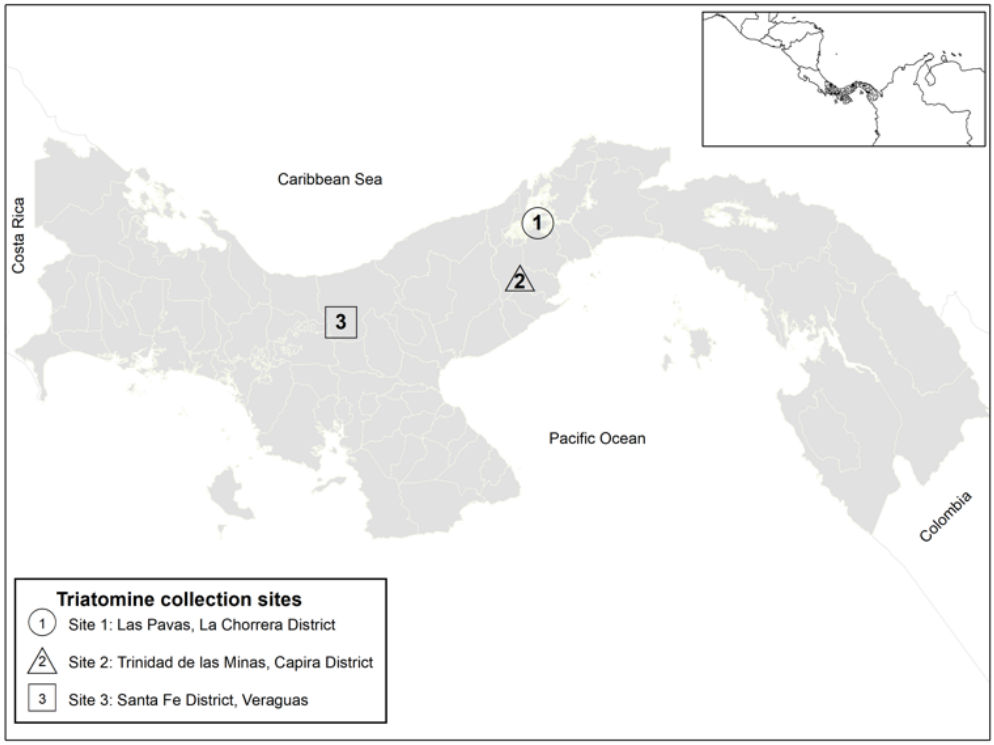


Figure 3.1. Map of Panama showing the locations of the three collection sites.

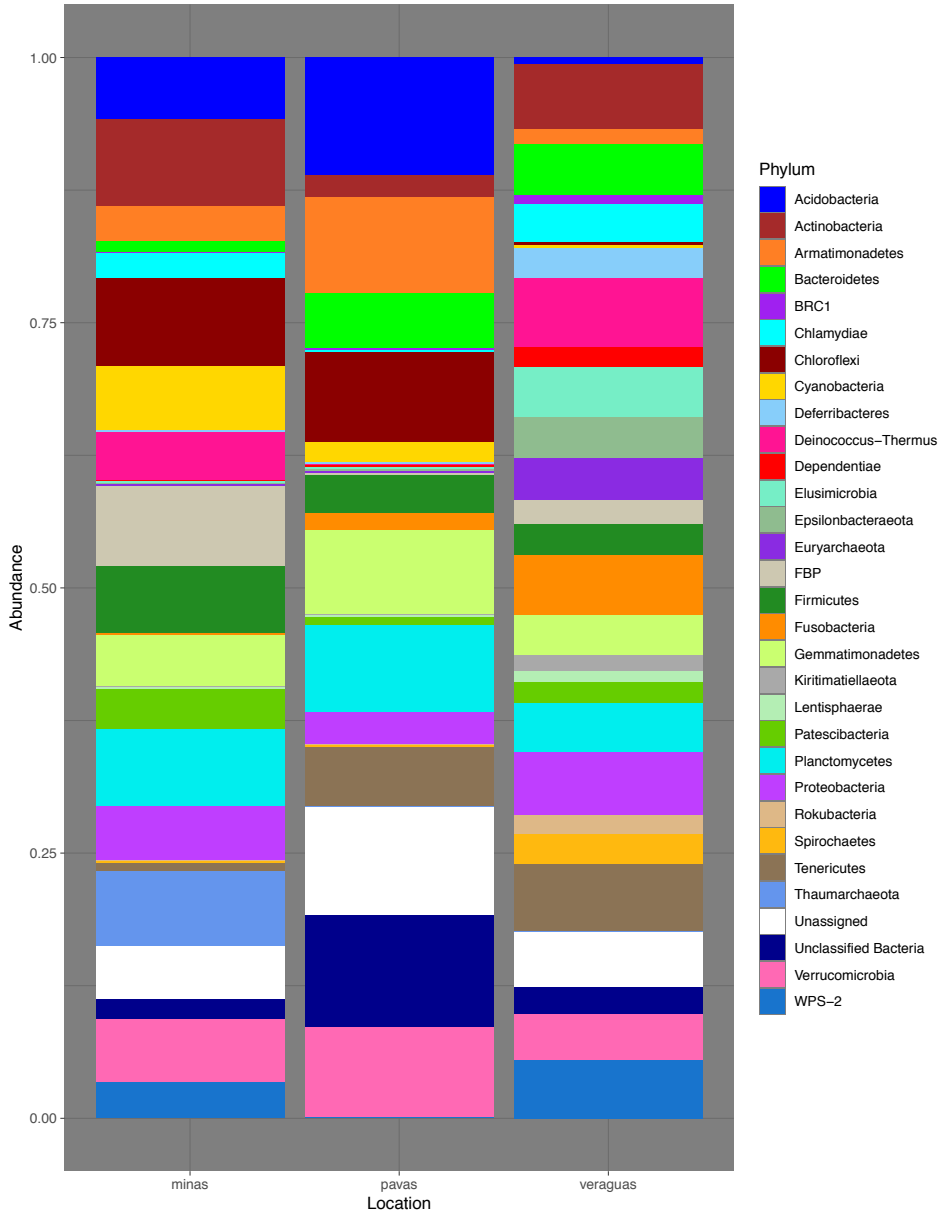


Figure 3.2. Taxonomic composition at the phylum level by location.

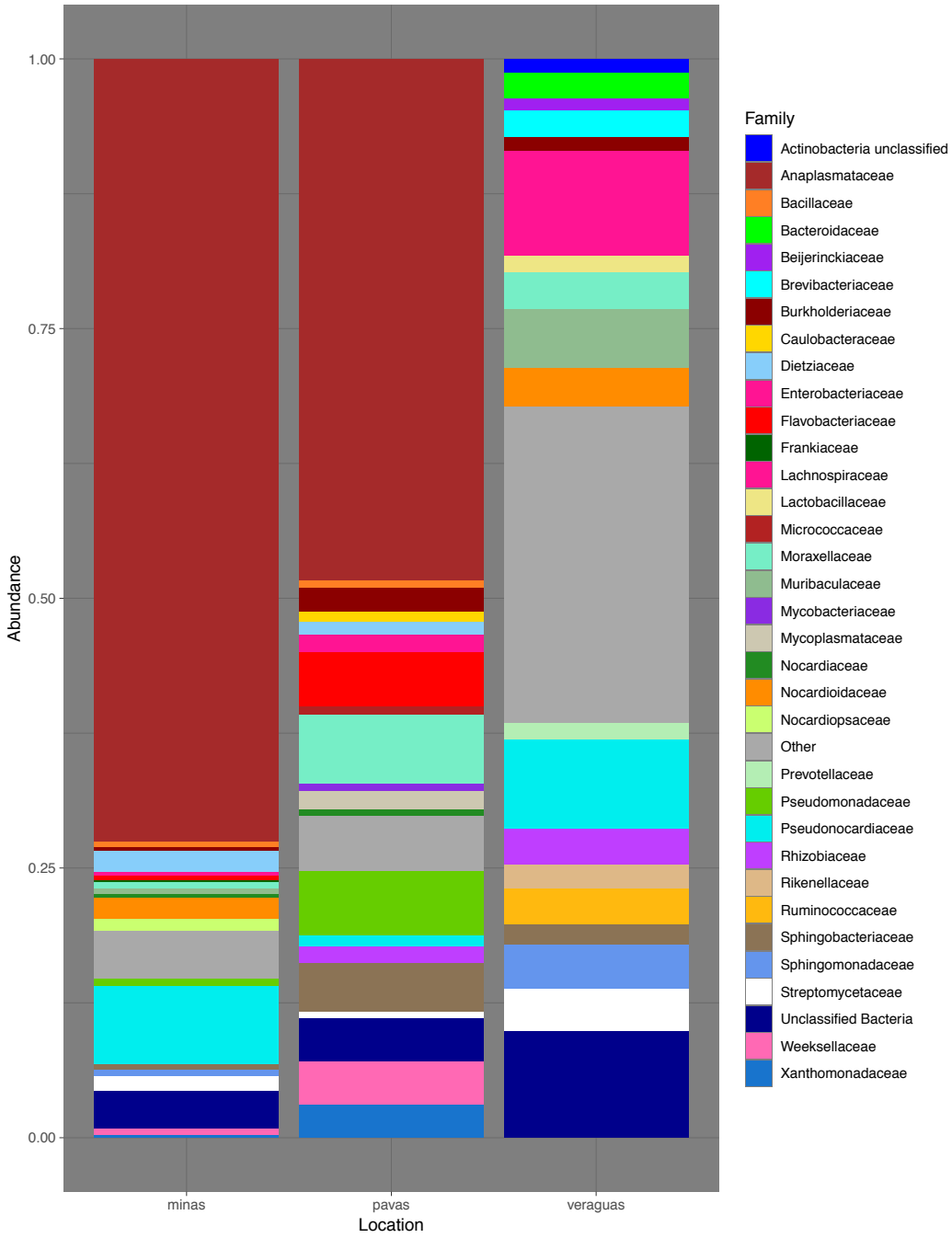


Figure 3.3. Top 20 taxonomic composition per location at the family level.

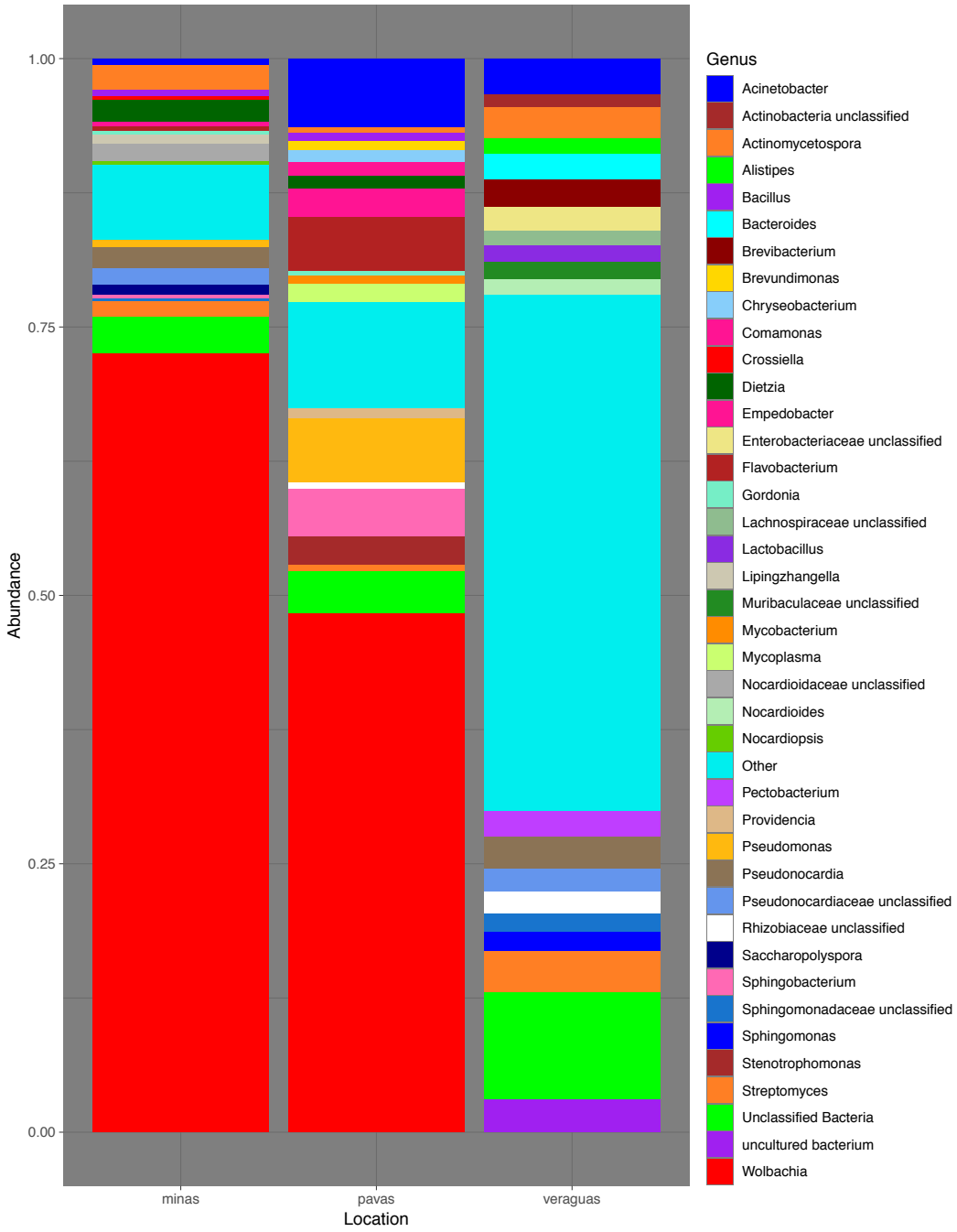


Figure 3.4. Top 20 taxonomic composition per location at genus-level.

CHAPTER 4

POPULATION GENETICS OF TWO CHROMATIC MORPHS OF THE CHAGAS DISEASE VECTOR *RHODNIUS PALLESCENS* IN PANAMA[1]

[1] Kieran TJ, Bayona-Vásquez NJ, Varian CP, Saldaña A, Calzada JE, Gottdenker NL, Glenn TC. Submitted to: *Infection, Genetics, and Evolution*.

Abstract

Rhodnius pallescens in the principal vector of Chagas disease in Panama. Recently a second darker chromatic morph has been discovered in the highlands of Veraguas Province. Limited genetic studies have been conducted with regards to the population structure and dispersal potential of Triatominae vectors, particularly in *R. pallescens*. Next Generation Sequencing technologies allow for more widespread adoption of these analyses in a time and cost-effective manner. RADseq is one such technique with great potential for examining vector biology across space and time. Here we utilize a RADseq method (3RAD), along with complete mitochondrial genomes, to examine the population structure of the two chromatic morpho types of *R. pallescens* in Panama. We generated a 2,216 SNP dataset and 6 complete mtDNA genomes. Both data sets (RADseq and mitochondrial) showed highly differentiated clades with essentially no gene flow between the dark and light chromatic morphs from Veraguas and central Panama respectively. We comment on the growing evidence showing clear distinctions between these two morpho types with the possibility that these are separate species. An area of research the requires further investigation. We also discuss the cost-effectiveness of 3RAD compared to other RADseq methods used recently in Chagas disease vector research.

Introduction

Understanding the population biology and transmission ecology of multi-host parasites is a top priority (Woolhouse & Dye, 2001) and often neglected area of disease research (Webster, Gower, Knowles, Molyneux, & Fenton, 2016). As part of these goals,

the identification of genetic diversity and structure of arthropod vectors of human diseases is essential to understand the transmission of disease, with the aim of vector-borne disease elimination. Vectors are the most prevalent form of disease transmission so ecological studies of their biology are of great importance. Research on vector population genetics is important for identifying cryptic species, genetic diversity, local adaptations, gene flow and colonization potential, population structure, dispersal, and transmission (McCoy, 2008). The decreasing costs of Next-Generation Sequencing (NGS) and the vast amount of genetic data generated by NGS have opened the door to vector ecology research that was previously unfeasible for many researchers.

The Reduviidae subfamily, Triatominae, are important vectors of Chagas disease, a parasitic zoonosis caused by the kinetoplastid protozoan *Trypanosoma cruzi*, transmitted by hematophagous insects. In many parts of the Americas, the disease is endemic and is estimated to affect 6-7 million people (WHO, 2019). Research into the genetic variation and population structure of triatomines is key to understanding factors important for vector control, including vectoral capacity, anthropogenic adaptation, and dispersal potential. There are more than 150 recognized Triatominae species which are widespread across the Americas (Justi, Galvao, & Schrago, 2016) with *Rhodnius prolixus*, *Triatoma dimidiata*, and *T. infestans* recognized as the major global vectors of Chagas disease. Other species, however, are important regional and local vectors. In Panamá, the principal vector of Chagas disease is *R. pallescens* (Calzada et al., 2010; Calzada et al., 2006; Rodriguez & Loaiza, 2017) and new endemic regions are being described (Calzada et al., 2010; Saldana et al., 2012; Saldana et al., 2018). A darker chromatic variation of *R. pallescens* infected by specific genetic groups of *Trypanosoma rangeli* and *T. cruzi* has

been found recently in Santa Fe District, Veraguas Province, Panama (Saldana et al., 2018) but has not yet been genetically characterized, opening further avenues of research in this area.

Studies examining triatomine genetic structure at larger spatial scales have consistently found significant levels of genetic differentiation suggesting minimal gene flow between geographically distant populations (Bargues et al., 2008; A. Gomez-Palacio et al., 2012; A. Gomez-Palacio, Triana, Jaramillo, Dotson, & Marcet, 2013; Monsalve, Panzera, Herrera, Triana-Chavez, & Gomez-Palacio, 2016; Peretolchina et al., 2018; Perez de Rosas, Segura, Fichera, & Garcia, 2008; Perez de Rosas, Segura, & Garcia, 2011; Pfeiler, Bitler, Ramsey, Palacios-Cardiel, & Markow, 2006; Roden, Champagne, & Forschler, 2011; Villacis et al., 2017; Waleckx et al., 2011). The amount of genetic structure tends to decrease with diminishing spatial scale, yet the actual amount of genetic variability can vary by specific locality (A. Gomez-Palacio et al., 2012; Orantes et al., 2018; Perez de Rosas et al., 2008; R. V. Piccinali et al., 2009). Gene flow has also been detected between sylvatic and domiciliary populations suggestive of good dispersal capability and reinfestation potential (Almeida, Faucher, Lavina, Costa, & Harry, 2016; Breniere et al., 2013; Breniere et al., 2012; Ceballos et al., 2011; Fitzpatrick, Dora Feliciangeli, Sanchez-Martin, Monteiro, & Miles, 2008; Stevens et al., 2015).

Triatomines often exhibit phenotypic differences within a species across their geographical distribution, in aspects such as size, anatomical morphology, and coloration (Abrahan, Hernandez, Gorla, & Catala, 2008; Catala et al., 2005; Dujardin, Costa, Bustamante, Jaramillo, & Catala, 2009; Hernandez, Abrahan, Moreno, Gorla, & Catala, 2008; Saldana et al., 2012; Saldana et al., 2018; Schofield & Galvão, 2009; Villacís,

Grijalva, & Catalá, 2010). This variability can be introduced by phenotypic plasticity, environmental differences (Dujardin et al., 2009; Dujardin, Panzera, & Schofield, 1999), or genetics (Panzera et al., 2004; Pires, Abrao, Machado, Schofield, & Diotaiuti, 2002). These processes contribute to the diversity of triatomine species complexes complicating taxonomy and identification in this group (Schofield & Galvão, 2009) as current species classification is largely based on morphological and chromatic features (Lent & Wygodzinsky, 1979). In part, the observed variability is likely due to the limited dispersal capability of triatomines which may lead to a strong biogeographic genetic structure. Processes of dispersal and barriers to gene flow are still not well understood to explain these patterns and how triatomines contribute to disease spread over different landscape compositions and scales.

To contain disease transmission, understanding triatomine population structure and gene flow is important and needs further study (Abad-Franch & Monteiro, 2005). Different methods have been used on triatomines, including allozymes, microsatellites, nuclear and mitochondrial DNA (Abad-Franch & Monteiro, 2005; Monteiro, Marcet, & Dorn, 2010). In the last 10 - 15 years, genetic studies in triatomines have been dominated by microsatellites and mitochondrial markers (primarily Cyt-b) (e. g. (Almeida et al., 2016; Breniere et al., 2012; A. Gomez-Palacio et al., 2012; Perez de Rosas et al., 2008; Perez de Rosas et al., 2011)). Such markers, however, may not be suitable for fine-scale population genetic analyses. Mitochondrial markers are a single locus with limited power from single gene surveys (Ballard & Rand, 2005; Rubinoff & Holland, 2005). Whole mitochondrial genomes have increased power, but are still limited by their mode of inheritance and lack of recombination (Kivisild, 2015). Microsatellites can be obtained

from multiple loci and have high power per locus which make them excellent markers as very fine-scales, but their high mutation rates can limit resolution among divergent taxa (Zink, 2010).

More recent work has shown that restriction site associated DNA sequencing (RADseq) markers to be very useful for ecological and population genetic studies (Andrews, Good, Miller, Luikart, & Hohenlohe, 2016) including disease vectors (Brown et al., 2014; Gulia-Nuss et al., 2016; Hernandez-Castro et al., 2017; Kotsakiozi et al., 2017; Orantes et al., 2018). RADseq enables hundreds to thousands of polymorphic loci to be discovered in non-model organisms using simple and cost-effective methods (Bayona-Vasquez et al., 2019; Davey et al., 2011). RADseq is also able to examine divergence of populations or species with varying degrees of signal strength (Andrews et al., 2016). However, in triatomines, only a few studies have used RADseq to date (Hernandez-Castro et al., 2017; Orantes et al., 2018).

Here we expand RADseq and genetic variation research in triatomines by examining the population genetic structure of *Rhodnius pallescens* from five localities in Panamá. Our aim was to determine the feasibility of a RADseq method (3RAD) in this vector, detect polymorphic loci, and report resources for further and future studies in the species, to determine whether there is population structure at regional and fine scale (i.e., provinces vs palms within a location). As an early indication of dispersal capacity, we also estimated migration rates among populations. Similarly, and for comparative analyses with previous studies, we evaluated the genetic variation of complete mitochondrial genomes, here sequenced, assembled and annotated, from a small number of individuals

from each locality. Finally, we evaluated the potential of sequence data generated from 3RAD of infected hosts for parasite population genetic analyses.

Methods

Sample Collection & DNA Extraction

Rhodnius pallescens specimens were collected from *Attalea butryacea* palms in Panamá using Noireau traps (Noireau et al., 2002). We sampled a total 105 bugs from 12 palms in five geographic locations (Figure 4.1). Sampling included; 2 palms in Las Pavas, La Chorrera District (N 9.104167°, W 79.885833°) (n=27); 3 palms in Santa Rita, La Chorrera District (N 8.905248°, W 79.882956°) (n=9); 3 palms in Trinidad de las Minas, Capira District (N 8.775556°, W 79.995833°) (n=32), all in Panama Oeste Province; 2 palms in Chilibre, Panama District, Panama Province east of the canal (N 9.098966°, W 79.569524°) (n=7); 4 palms in Santa Fe District, northern Veraguas Province (N 8.509232°, W 81.077800°) (n=30). Specimens from Las Pavas, Trinidad de las Minas, and Santa Fe were previously used in other studies examining bloodmeals (Kieran et al., 2017) and microbiome composition (Kieran, Arnold, et al., 2019). DNA was extracted from whole specimens following Kieran et al. (2017). Briefly, macerated samples were placed in digest buffer with Proteinase K and digested overnight before extraction with Phenol-Chloroform-Isoamyl alcohol. Extracted DNA were reconstituted in 30 µL TLE buffer (10 mM Tris pH 8, 0.1 mM EDTA) and impurities were removed with Sera-Mag SpeedBeads™ (Fisher Scientific, Waltham, MA; (Faircloth & Glenn, 2012)).

Library Preparation & Sequencing

We constructed 3RAD (RADseq) libraries following established protocols (Bayona-Vasquez et al., 2019; Graham et al., 2015). DNA was digested using three restriction enzymes (EcoRI-HF, NheI-HF, XbaI; New England Biolabs® Inc.) followed immediately by ligation of adaptors with internal variable length barcodes (Bayona-Vasquez et al., 2019) and amplified with iTru-indexed primers (Glenn et al., 2019). Samples were then pooled and sent for paired-end 75 and 150 bp sequencing on either a NextSeq 500 (Georgia Genomics and Bioinformatics Core), HiSeq 3000 (Oklahoma Medical Research Foundation) or a NovaSeq (University of Kansas Medical Center). In addition, we prepared genomic libraries for six individuals (two from Las Pavas, one from Trinidad de las Minas, and three from Santa Fe) using KAPA Hyper Plus Kit (KAPA Biosystems) with Y-yoke adaptors and indexed primers (Glenn et al., 2019) as previously described (Kieran, Gordon, et al., 2019) for assembly of mitochondrial genomes.

Generating a catalog of probes for polymorphic loci in *R. pallescens*

With a subset of samples from our RADseq study, we generated a catalog of polymorphic loci and designed baits that can be used for future studies in the species by using sequence capture (i.e. RADcap, (Hoffberg et al., 2016)). To generate the catalog of polymorphic loci, we analyzed 90 samples in STACKS v.2.0beta7, we processed raw fastq files with *process_radtags*, and then used a *de novo*-based assembly with parameters M = 4 and n = 4. Then, using the pipeline *populations*, we output a fasta file with each allele sequence per individual for only biological plausible loci (--fasta_strict, max two alleles

for diploids, -p and -r = 80%). Using EMBOSS 6.5.7, we created two new fasta files, one containing the consensus sequence for each locus, where the different alleles (SNPs) for that locus are encoded as ambiguities (consambig), and another file with the same consensus sequence per locus, but instead of ambiguities at polymorphic sites, the base with the highest score (most frequent) is used (cons). Then, using the R package BIOSTRINGS (Pagès, Aboyoun, Gentleman, & DebRoy, 2019), we counted the number of ambiguities in our consambig file, to determine the number of SNPs per locus, we filtered out loci with more than five SNPs and length < 230 bp to have enough space to potentially synthesize two 90 nucleotide baits per locus.

We repeat-masked against bilateral lineage repeats, then 90 nt baits, four per locus (2 for Read1 and two for Read2) were designed with ~20 nt flexible spacing. Then each bait candidate was blasted against *R. prolixus* genome (GCA_000181055.3) and a hybridization melting temperature T_m was estimated for each hit assuming standard myBaits® buffers and conditions. For each bait, one BLAST hit with the highest T_m was first discarded, and then only the tip 500 hits were considered. Based on the distribution of remaining calculated Tm's we then filtered out non-specific baits using the following criteria: A) stringent (only specific baits pass) if they satisfy one of the following conditions, i) no hits with T_m above 60 °C, ii) at most 2 hits 62.5 – 65 °C, iii) at most 10 hits 62.5 – 65 °C and at least 1 failing flanking bait, iv) at most 10 hits 62.5 – 65 °C, 2 hits 65 – 67.5 °C, and fewer than 2 passing flanking baits, v) at most 2 hits 62.5 – 65 °C, 1 hit 65 – 67 °C, 1 hit 70 °C or above, and < 2 passing flanking baits; B) moderate (some non-specific baits pass) if they have at most 10 hits 62.5 – 65 °C and 2 hits above 65 °C, and fewer than 2 passing baits on each flank; and C) relaxed (more non-specific baits

pass) if they have at most 10 hits 62.5 – 65 °C and 4 hits above 65 °C, and fewer than 2 passing baits on each flank.

From this list, we removed baits with repeats, we selected baits that passed our BLAST with relaxed conditions (C from paragraph above), and selected baits with GC content > 30% and < 60%. We kept one bait in the Read1 and one bait in the Read2 per locus, and only considered those where both, Read1 and Read2, presented baits that passed the filters.

RADseq Data Processing, Bioinformatics & Data Filtering

We used STACKS v. 2.3e (Rochette, Rivera-Colon, & Catchen, 2019) to process raw fastq reads, assemble loci, map catalog loci to a reference genome and create output files. In brief we, demultiplexed, quality-filtered, checked for enzyme cut-sites and trimmed all reads to 64 bases using *process_radtags*. Then, we ran a subset of samples using the *denovo_map* program to optimize the number of mismatches allowed between stacks within and between individuals (-M and -n, respectively) according to Paris et al. (Paris, Stevens, Catchen, & Johnston, 2017). After optimization, we ran the complete dataset with the optimized parameters and used the *stacks-integrate-alignments* pipeline to map the *de novo* assembled loci in our catalog to the genome of a closely related species *R. prolixus* (GCA_000181055.3). We ran the *populations* pipeline setting the parameters to output only those loci present in 80% of the individuals within one location and at least in four out of the five locations sampled. The genepop file generated was imported in RSTUDIO v. 1.1463 (RStudio Team, 2015) and transformed to *genind* object using ADEGENET v. 2.1.1 (Jombart & Ahmed, 2011). We filtered out loci with > 25%

missing data, and then we filtered out samples with > 25% missing data using the function *missingno* from the R package POPPR v. 2.8.2 (Kamvar, Tabima, & Grunwald, 2014).

Population Genetics Analyses in *R. pallescens*

Global and per sampling site observed and expected heterozygosity were estimated using the *summary* function from ADEGENET v. 2.1.1 (Jombart & Ahmed, 2011). Similarly, we estimated identity-based gene diversity indices, F_{IS}, and Hardy-Weinberg equilibrium (HWE) using the probability test with 1000 dememorization, 100 batches and 1000 iterations per batch in the web version of GENEPOP 4.2 (Raymond & Rousset, 1995).

Genetic differentiation was assessed through the estimation of pairwise F_{ST} (Weir & Cockerham, 1984) between sampling sites and between palms using ARLEQUIN v. 3.5.2.2 (Excoffier & Lischer, 2010), calculating the statistical significance using 1000 permutations and allowing 5% of missing data. Similarly, an AMOVA hierarchical analysis was computed in the same software by grouping all palms within sites.

To examine genetic clusters in our data, we investigated using the model-based clustering method implemented in STRUCTURE v.2.3.4 (Pritchard, Stephens, & Donnelly, 2000), randomly selecting one SNP per locus of our filtered dataset. We ran 100,000 permutations for burnin and 1,000,000 MCMC chains. We permuted K= 1 to 6 and for each value of K we ran three replicates under an admixture ancestry model and correlated allele frequencies. Then, we implemented the Evanno method from STRUCTURE HARVESTER v0.6.94 (Earl & vonHoldt, 2011) to select the number of genetic clusters in our data. Similarly, we used the *find.clusters* function from ADEGENET v. 2.1.1, and

assessed the optimal number of cluster (*K*) based on the Bayesian Information Criterion (BIC) value.

We tested isolation by distance (IBD) using a Mantel test between a matrix of Edward's genetic distances and Euclidean geographic distances between palms across all sites. Edward's genetic distances between palms where estimated using the function *dist.genepop* from the ADEGENET v. 2.1.1 package, and the test was performed using the *mantel.randtest* function from the R package ADE4 v. 1.7-13 (Dray & Dufour, 2007), using 999 replicates to estimate significance. To validate the IBD pattern we performed the same test removing palms from SF site.

To further investigate the direction and magnitude of migration between sampling sites, we generated a migration network using the *divMigrate* function from the R package *diveRsity* (Keenan, 2014) to estimate the effective number of migrants per generation (*Nm*) (Alcala, Goudet, & Vuilleumier, 2014). We then performed the same migration analyses limited to just the individual palms within the Santa Fe site. We also estimated Nei's genetic distances between samples and calculated a UPGMA tree estimating the bootstrap support (1,000 samplings) on the nodes by using the function *aboot* with from POPPR v. 2.8.2.

Mitochondrial Genome Assembly

Genomic reads were demultiplexed by iTru indexes using bcl2fastq and adaptors were trimmed in GENEIOUS v 10.0.1 (Kearse et al., 2012). Trimmed reads were assembled using NOVOPLASTY v2.7.2 (Dierckxsens, Mardulyn, & Smits, 2017) using a *Rhodnius prolixus* COI sequence from GenBank (AF449138.1) as a seed input for all six

samples. Contigs from NOVOPLASTY were imported into GENEIOUS for verification by using *Map to Reference* with the quality trimmed reads (length 80-151bp) against their respective mitochondrial genome and checked for circularization. Confirmed complete mitogenomes were then annotated using MITOS2 WEB SERVER (Bernt et al., 2013; Donath et al., 2019) with RefSeq 63 Metazoa reference and Invertebrate genetic code settings. Complete annotated genomes were aligned with MAFFT v 1.3.6 (Katoh & Standley, 2013) as implemented in GENEIOUS. Sequences for Cyt-b and 16S rRNA from assembled mitogenomes were each extracted and aligned with *Rhodnius sp*. sequences from GenBank using MAFFT and manually trimming alignments to consistent lengths. A maximum-likelihood tree with a GTRGAMMA model and 100 bootstraps was constructed using RAxML v 4.0 (Stamatakis, 2014) as implemented in GENEIOUS.

Detecting 3RAD loci from *Trypanosoma cruzi and T. rangeli* in the samples

Similar to the analyses in *R. pallescens*, we used *stacks-integrate-alignments* pipeline to map the loci in our *de novo* catalog independently to the genomes of the protozoan parasites *Trypanosoma cruzi* (ASM20906v1) and *T. rangeli* (ASM371947v1). In the *populations* pipeline we used parameters to output loci present in at least 10% of the total samples. We then compared both maps to identify loci exclusive to each of the *Trypanosoma* species and only considered those in subsequent analyses. Then, we filtered out loci with > 50% of missing data and then samples with > 50% of missing data. We compared the samples that are genotyped for these loci with those reported in the study from Kieran et al. (2017), who used PCR methods to detect, in some of the same samples used here, the presence of *T. cruzi* and *T. rangeli*.

We estimated the number of samples with *Trypanosoma* loci detected, number of alleles, observed and expected heterozygosity. We estimated the alpha-score and built a DAPC analysis to plot posterior probability values using the function *compoplot* from ADEGENET v. 2.1.1 for each *Trypanosoma* species. We also estimated a Nei's genetic distance tree and calculated an UPGMA tree estimating the bootstrap support (1,000 samplings) on the nodes by using the function *aboot* from POPPR v. 2.8.2.

Results

3RAD summary

We sequenced 105 *R. pallescens* samples from five localities in Panama, resulting in a total of 135,285,123 paired reads with an average of 1,288,430 paired-reads per sample. After filtering out low-quality reads, specifically reads without cut sites or ambiguous adapters, we recovered a total of 131,329,744 paired reads (97.07%), with an average of 1,250,759 paired-reads per sample (Appendix E). During the *de novo* assembly, we recovered 298,894 loci, 11,608 polymorphic loci with 27,438 SNPs.

Generating a catalog of polymorphic loci and baits for *R. pallescens* and its holobiome

From 90 samples (excluding samples from Santa Fé, not yet obtained) we obtained 52,985 loci shared in 80% of samples from our *de novo* assembly, of which 21,195 had 1-5 SNPs. Of these, 15,714 had a length > 230 bp. At the designing baits stage, 36,460 baits were designed for 15,713 loci, but after filtering out baits that didn't pass our thresholds (in repeats, GC content, BLAST criteria and one bait per read) we kept 13,460 baits for 6,730 polymorphic loci (Appendix F).

R. pallescens population genetics

For the 105 samples, in total, 75,153 loci of those assembled *de novo* (25.1%) mapped to the reference genome of *R. prolixus*. Of these 1,468 were polymorphic and shared in 80% of samples and sites and having a total of 3,933 SNPs. We filtered out 1,717 SNPs that contained > 25% missing data, and then we filtered out 15 samples (AB104_1, 012p, 040p, 045p, 049p, 036TM, 039TM, 040TM, 041TM, 046TM, 068TM, 205, 349, 358, and 363) that contained > 25% missing data, resulting in a final dataset containing 90 individuals and 2,216 SNPs.

Across sampling sites, number of samples ranged from 7 (CH) to 26 (TM and SF); number of loci or hereafter SNPs ranged from 2,097 (SR) to 2,216 (LP, TM, and SF); total number of alleles ranged from 2,643 (CH) to 3,279 (TM); mean allelic richness, estimated as the average number of alleles in a rarefied sample per SNP ranged from 1.23 (SR) to 1.31 (SF); mean values of observed heterozygosity ranged from 0.084 (CH) to 0.095 (LP); mean values of expected heterozygosity ranged from 0.083 (LP) to 0.120 (TM). These estimates can be visualized per sampling site in Figure 4.1 and Table 4.1. Also, LP and SF show the highest values of gene diversity (within and among individuals) and observed heterozygosity, and SR shows the lowest values. However, expected heterozygosity (H_E) are higher for CH, SR and SF. The inbreeding coefficient F_{IS} was higher for SF, and negative for CH, LP and TM. Deviations between observed and expected heterozygosity were significant for LP, TM, and SF sites and also when grouping all sites, according to the HWE probability test.

In total, 1,794 SNPs passed the 5% missing data threshold. With these, pairwise F_{ST} estimates across sampling sites were higher in the comparisons between the SF site

and any other collection site (0.79–0.80) and all of these were statistically significant after Bonferroni correction ($p = 0.005$). SR showed the lowest values of genetic differentiation when compared to CH and TM (Table 4.2). When comparing among palms from the different sampling sites (Bonferroni $p = 0.0007$), those palms from SF present higher and significant F_{ST} values when compared to palms at other sites. Similarly, palm 4 from SF presented high F_{ST} values compared to the other three palms at the same site. Also, individuals from the two palms in the LP site were significantly different (Table 4.2). Our AMOVA analysis shows the variance is maximized when comparing sites, but also shows a significant component of the variance within sites between palms (Table 4.3).

Genetic clustering analysis performed in STRUCTURE v.2.3.4 reveal that the number of clusters with the highest posterior probability is $K = 2$ (Mean LnP = -24710.53), and this result was supported by the Evanno method with the highest value of Delta K (Appendix G; Figure 4.3A). The membership probabilities reveal almost a 100% assignment of individuals to each of the two clusters, and partitions genetic structure in one cluster that groups CH, LP, SR and TM sites, and a second cluster with individuals sampled from SF. The alternative analysis using the BIC value suggests $K = 3$ (Appendix H). The assignment of each individual to these three cluster, reveal that samples from CH, LP, SR, and TM belong to one cluster, and those samples from SF are split in two different clusters. When looking closer to samples from SF site, those samples from palms 1, 2 and 3 belong to one cluster, and samples from palm 4 belongs to the third genetic cluster (Figure 4.3B). When testing for IBD across all palms, we found a significant pattern of isolation by distance (R = 0.95, p-value = 0.001); however, when

this test is performed without samples from SF this pattern is lost (R = -0.13, *p*-value = 0.664).

In the estimation of the *Nm* between sampling sites, we found high levels of migration between SR–TM–LP (*Nm* ranged from 0.53 to 1), then lower rates of migration between these locations and CH (*Nm* ranged from 0.53 to 0.83) and very low migration rates between SF and all of the other localities (*Nm* = 0.01-0.02, Figure 4.4).

The UPGMA tree based on Nei's genetic distance reveals two main clades with high support that are separated by a Nei's distance of 0.27, the first clade contains exclusively samples from SF and the second clade contains exclusively samples from CH, LP, SR and TM (Figure 4.44A). Within the first clade, a ~0.03 Nei's genetic distance separates the samples from Palm 4 from the other palms within SF. In the second clade, smaller genetic distances separate sites and palms, and only CH form a monophyletic group, the other sites are mixed among clades.

Mitochondrial Genomes

Mitochondrial genomes varied in length by locality from 15,888 bp in SF and 16,391 bp in LP and TM. This is due to a 502 bp insert from 9,850-10,351 in Panama Oeste samples. All genomes consist of 13 PCGs, 2 rRNAs, and 22 tRNAs (Appendix I). Similarity of individual genes between northern Veraguas and Panama Oeste province localities ranged from 96.29% (NAD3) - 98.89% (16S) (Appendix J). Maximum likelihood phylogenetic trees are presented in Appendix K for Cyt-b (A) and 16S (B).

Detection of *T. cruzi* and *T. rangeli* DNA

In total, 1,829 loci of those assembled *de novo* (0.6%) mapped to the reference genome of *T. cruzi*. And 1,367 loci (0.46%) mapped to the reference genome of *T. rangeli*. When comparing mapping sequences, we found 901 loci shared by both species (i.e., reads mapped to both species due to low divergence).

For *T. cruzi*, 806 loci were polymorphic and occurred in at least 10% of samples (103 samples), mapped to 230 scaffolds in the reference, composed of 107,112 sites having a total of 1,613 SNPs, all biallelic (2,126 alleles). Of these, 223 loci with 641 SNPs were exclusive for *T. cruzi*. After filtering for loci and samples with >50% of missing data, we recovered 17 SNPs from five loci shared in 68 samples. The locality with the smallest number of samples was CH (2) and the one with the most was TM (27). The locality with the highest number of SNPs genotyped were LP and TM (17) and the one with fewest was CH (13), SR and SF were genotyped for 15. The DAPC shows a pattern of differentiation in *T. cruzi* in samples from SF (Appendix L). A similar pattern is observed in the UPGMA tree using Nei's distance, where most samples from SF belong to a highly distant clade from the rest of the samples (Figure 4.5B).

If assuming that PCR positive samples can be recorded as *positive* for the detection of *T. cruzi*, we compared our results with those from previously published studies (Kieran et al. 2017, Kieran et al. 2019a) (Appendix M). We found that 31 out of the 62 (50%) samples were consistently positive in at least one of the two PCR tests for the same pathogen, but also other 31 (50%) samples were negative compared to PCR tests, and for PCR negative samples we found 11 of 27 (40.7%) samples 3RAD data was

not consistent with PCR tests, and 16 (59.3%) samples were consistently negative across PCR and 3RAD data. Indicating a ratio of true vs false of 1.2:1.

For T. rangeli, 569 loci were polymorphic and shared in at least 10% of samples (out of 105 samples), mapped to 167 scaffolds in the reference, composed of 75,994 sites having a total of 505 SNPs, all biallelic (1,010 alleles). Of these, 48 loci with 81 SNPs were exclusive for *T. rangeli*. After filtering these for > 50% missing data, only five SNPs from one locus (304370) were shared in 54 samples. The locality with the smallest number of samples was CH (4), followed by SF (5) and SR (6), those with the most samples were LP (14) and TM (27). All localities were recorded for that one locus and five SNPs. The DAPC shows a pattern of differentiation in *T. rangeli* in samples from SF (Appendix I). A similar pattern is observed in the UPGMA tree using Nei's distance, where most samples from SF belong to a highly distant clade from the rest of the sample (Figure 4.5C).

Again, assuming PCR positive samples can be recorded as *positive* for the detection of *T. rangeli*, we compared our results from previously published studies (Kieran, Arnold, et al., 2019; Kieran et al., 2017) (Appendix J). We found that 25 of the 42 (59.5%) samples were consistently positive in at least one of the two PCR tests for the same pathogen, but another 17 samples (40.5%) were negative compared to PCR tests, and for PCR negative samples we found 27 of 47 (57.4%) 3RAD data were not consistent with PCR tests, and 20 (42.6%) samples were consistently negative across PCR and 3RAD data. Indicating a ratio of true vs false of 1.02:1.

Discussion

This study examined the genetic differentiation of *Rhodnius pallescens* in Panama at fine and regional spatial scales utilizing a RADseq method (3RAD, (Bayona-Vasquez et al., 2019)) and complete mitochondrial genomes. In our study, we found little genetic differentiation among sites in Panama (CH) and Panama Oeste (LP, TM, SR) provinces with high rates of migration among these sites. Chilibre (CH), east of the canal, exhibited high rates of migration to Panama Oeste. While the rates are lower than among Panama Oeste localities, the still high rates indicate that the canal may have minimal impact on the movement of *R. pallescens* in central Panama. This is in contrast to samples in Santa Fe (SF), Veraguas Province, which exhibited very low rates of migration and very high rates of genetic differentiation with all other localities in central Panama. This could be due to the greater distance between SF and the other locations (~ 150 km) which was significant for IBD, as well as some topographical/elevational effect. This dynamic corresponds to the clustering support for two populations (K = 2), which is consistent with the results found in the sister species *R. ecuadoriensis* in Ecuador with similar distances between collected samples surveyed with RADseq (Hernandez-Castro et al., 2017) and microsatellite (Villacis et al., 2017) data.

The results from our RADseq data are consistent with the mitochondrial genome data, which showed variation between Veraguas (SF) and Panama Oeste (LP, TM) localities. Most notable is the 502 bp insert in Panama Oeste samples that is missing from Veraguas specimens. The genetic differentiation between Veraguas and central Panama localities is consistent with previously observed levels of genetic differentiation for other triatomine species at larger spatial scales (Bargues et al., 2008; A. Gomez-Palacio et al.,

2012; Monsalve et al., 2016; Peretolchina et al., 2018; Perez de Rosas et al., 2008; Perez de Rosas et al., 2011; Pfeiler et al., 2006; R. V. Piccinali et al., 2009; Roden et al., 2011; Villacis et al., 2017; Waleckx et al., 2011). This suggest the possibility that these two color-morphs may be different species of *Rhodnius pallescens*, which we discuss further below.

Previous genetic studies of triatomine populations sampled at smaller spatial scales characterized a variety of factors that contribute to population structure, including: limited dispersal potential (Ramirez et al., 2005), dispersal blockers in a highly local, urban environment (Khatchikian et al., 2015), and local structure (i.e. (Breniere et al., 2013; Romina Valeria Piccinali & Gürtler, 2015; Stevens et al., 2015). While most genetic variance is within populations (palms in our study), we observed some variance between palms within localities, with LP and SF palms exhibiting significant differentiation. Between LP palms IBD was not observed, and differences in genetic variance may possibly be due to ecological factors related to the environment that affect mating and dispersal. In this case one palm was in pasture habitat and the other in peridomestic. However, this potential microhabitat pattern was not found among SF palms.

In SF we observe high rates of migration between three of the four palms, with low to moderate rates between these and the fourth palm (SF-4). Physical distance does not seem to be the primary driver of these lower rates as IBD between these palms was not detected; however, the Santa Maria River is separating this fourth palm from the others. This potential riverine barrier to dispersal is in contrast to the apparent lack of a barrier between Chilibre and Panama Oeste samples due to the Panama Canal. These

findings leave open the possibility that rivers and similar topographical features can act as barriers to triatomine dispersal, which have been found to limit the range expansion of triatomine species (Costa, Dornak, Almeida, & Peterson, 2014). The Chagres River priory to the construction of the Panama Canal (> 100 years ago) could have acted as a dispersal barrier. In which case, the geneflow between opposite sides of the canal could be the result of recent anthropogenic factors (Dujardin, Schofield, & Tibayrenc, 1998; Forattini, Rocha e Silva, Ferreira, Rabello, & Pattoli, 1971; Lent & Wygodzinsky, 1979; Schofield, 1994) and not merely the presence of a water barrier. However, whether this geneflow is a result of ongoing dispersal or an artifact of previous dispersal across a non-barrier Chagres River before the construction of the Panama Canal is an open question.

Our study highlights the finding that *R. pallescens* specimens from Santa Fe, Veraguas are genetically differentiated from other specimens in Panama as evidenced by SNPs and complete mitochondrial genome differences. Furthermore, phylogenetic trees for Cyt-b and 16S genes show a separate clade for *R. pallescens* specimens from Veraguas (Appendix M). This finding is consistent with results from Cyt-b analyses that showed Santa Fe specimens forming a separate clade from other specimens in Panama (including La Chorrera) and western Colombia, which together form a clade sister to the rest of Colombia (A. Gomez-Palacio et al., 2012).

Rhodnius pallescens specimens from Veraguas are morphologically very similar to central Panama samples except for a slightly larger size and darker chromatic variation (Saldana et al., 2012; Saldana et al., 2018). Morphological characteristics are the primary means of Triatominae species identification (Lent & Wygodzinsky, 1979), however there is a lot of phenotypic variability seen in many populations (Abrahan et al., 2008; Catala

et al., 2005; Dujardin et al., 2009; Hernandez et al., 2008; Saldana et al., 2012; Saldana et al., 2018; Schofield & Galvão, 2009; Villacís et al., 2010). Chromatic or color variation is found in triatomines (Almeida, Pacheco, Noireau, & Costa, 2002; Costa, Correia, Neiva, Goncalves, & Felix, 2013; Noireau, Flores, Gutierrez, & Dujardin, 1997) and in *Rhodnius sp.* in particular (Abad-Franch et al., 2001; F. B. Dias, Jaramillo, & Diotaiuti, 2014; F. B. S. Dias, Bezerra, Machado, Casanova, & Diotaiuti, 2008; Gaunt & Miles, 2000; Saldana et al., 2012; Saldana et al., 2018).

This variability can arise through ecological and geographical variation/isolation in Triatominae populations (Dujardin et al., 2009; Dujardin et al., 1999; Schofield & Galvão, 2009). The occurrence of a darker chromatic variation of *R. pallescens* in the cooler, mountainous region of Veraguas, Panama could be an example of this (Saldana et al., 2018). This is consistent with the variation found in the microbiome (Kieran, Arnold, et al., 2019), trypanosomes present (Saldana et al., 2018), and the present genetic diversity observed. Our results are also consistent with divergences observed in cytotypes (Andres Gomez-Palacio et al., 2008), previous mitochondrial marker analysis, and morphometrics (A. Gomez-Palacio et al., 2012) of individuals from Santa Fe, Veraguas compared to other Panamanian samples.

While chromatic variation has been used as an important trait for species descriptions in *Rhodnius* (da Rosa et al., 2017; Souza et al., 2016), the observed genetic divergence between *R. pallescens* in Veraguas vs central Panama could be interpreted as separate species, sub-species, or representative of a strong biogeographic signal. This is most evident by the mitochondrial marker phylogenetic trees which show a consistent *R. pallescens* clade, bifurcating between Veraguas and central Panama specimens. However,

however caution should be taken when taxonomically splitting Triatominae, particularly *Rhodnius* (Nascimento et al., 2019). Therefore, further research on genetics, morphology, and crossbreeding trials between these localities that take a more holistic approach are needed to resolve this divergence question. Research on vector population genetics is important for identifying cryptic species, diversity and local adaptations, gene flow, colonization potential, and population structure and transmission (McCoy, 2008). This is particularly important for a group like Triatominae which exhibits a lot of phenotypic variability and is composed of many species' complexes (Lent & Wygodzinsky, 1979; Schofield & Galvão, 2009).

This study adds to the growing list of RADseq methodologies being employed in triatomine vectors for population and genetic diversity analyses (Bayona-Vasquez et al., 2019; Hernandez-Castro et al., 2017; Orantes et al., 2018). We encourage the scientific community to use the same set of enzymes and RADseq methods to produce comparable datasets across Triatominae species and populations. The 3RAD method proved effective and affordable with < $1.50 per library in preparation costs (Bayona-Vasquez et al., 2019) and approximately $1.50 (NovaSeq) to $3.50 (HiSeq) in sequencing cost per million reads. In terms of actual money spent, our 135.3 million reads generated for 105 samples cost a total of $6.01 per sample with HiSeq pricing. The total cost is substantially lower than similar RADseq methods that reported $18 (Hernandez-Castro et al., 2017) and $30 (Toonen et al., 2013) per sample in library and sequencing costs. 3RAD greatly reduces costs by extensive multiplexing and pooling of samples, that allows for proportional sequencing for the amount of data desired (Bayona-Vasquez et al., 2019; Glenn et al., 2019). With RADcap baits designed herein (Appendix F), cost

could be reduced further as sample size increases using target enrichment (Hoffberg et al., 2016). As costs continue to fall (Davey et al., 2011; Glenn, 2011), these genetic methodologies can be more widely applied to disease vectors for numerous lines on inquiry for effective disease control and public health improvement (Rinker, Pitts, & Zwiebel, 2016).

Acknowledgements

We thank Franklyn Samudio and Jose Montenegro for assistance in the field. The authors also thank the Ministerio de Ambiente in Panama (MiAMBIENTE), the Smithsonian Tropical Research Institute (STRI), and the University of Georgia Graduate School. University of Kansas Medical Center Genomics Core for core support services using supporting grants: Smith Intellectual and Developmental Disabilities Research Center (NIH U54 HD 090216), the Molecular Regulation of Cell Development and Differentiation – COBRE (5P20GM104936-10) and the S10 High-End Instrumentation Grant (NIH S10OD021743).

References

Abad-Franch, F., & Monteiro, F. A. (2005). Molecular research and the control of Chagas disease vectors. *Anais da Academia Brasileira de Ciências, 77*(3), 437-454.

Abad-Franch, F., Paucar, A., Carpio, C., Cuba, C. A., Aguilar, H. M., & Miles, M. A. (2001). Biogeography of Triatominae (Hemiptera: Reduviidae) in Ecuador: implications for the design of control strategies. *Mem Inst Oswaldo Cruz, 96*(5), 611-620.

Abrahan, L., Hernandez, L., Gorla, D., & Catala, S. (2008). Phenotypic diversity of *Triatoma infestans* at the microgeographic level in the gran chaco of Argentina and the Andean valleys of Bolivia. *J Med Entomol, 45*(4), 660-666. doi:Doi 10.1603/0022-2585(2008)45[660:Pdotia]2.0.Co;2

Alcala, N., Goudet, J., & Vuilleumier, S. (2014). On the transition of genetic differentiation from isolation to panmixia: What we can learn from Gst and D. *THEORETICAL POPULATION BIOLOGY, 93*, 75-84.

Almeida, C. E., Faucher, L., Lavina, M., Costa, J., & Harry, M. (2016). Molecular Individual-Based Approach on Triatoma brasiliensis: Inferences on Triatomine Foci, Trypanosoma cruzi Natural Infection Prevalence, Parasite Diversity and Feeding Sources. *PLoS Negl Trop Dis, 10*(2), e0004447. doi:10.1371/journal.pntd.0004447

Almeida, C. E., Pacheco, R. S., Noireau, F., & Costa, J. (2002). *Triatoma rubrovaria* (Blanchard, 1843) (Hemiptera : Reduviidae) I: Isoenzymatic and chromatic

patterns of five populations from the state of Rio Grande do Sul, Brazil. *Mem Inst Oswaldo Cruz, 97*(6), 829-834. doi:Doi 10.1590/S0074-02762002000600013

Andrews, K. R., Good, J. M., Miller, M. R., Luikart, G., & Hohenlohe, P. A. (2016). Harnessing the power of RADseq for ecological and evolutionary genomics. *Nat Rev Genet, 17*(2), 81-92. doi:10.1038/nrg.2015.28

Ballard, J. W. O., & Rand, D. M. (2005). The population biology of mitochondrial DNA and its phylogenetic implications. *Annual Review of Ecology Evolution and Systematics, 36*, 621-642. doi:10.1146/annurev.ecolsys.36.091704.175513

Bargues, M. D., Klisiowicz, D. R., Gonzalez-Candelas, F., Ramsey, J. M., Monroy, C., Ponce, C., . . . Mas-Coma, S. (2008). Phylogeography and genetic variation of Triatoma dimidiata, the main Chagas disease vector in Central America, and its position within the genus Triatoma. *PLoS Negl Trop Dis, 2*(5), e233. doi:10.1371/journal.pntd.0000233

Bayona-Vasquez, N. J., Glenn, T. C., Kieran, T. J., Pierson, T. W., Hoffberg, S. L., Scott, P. A., . . . Faircloth, B. C. (2019). Adapterama III: Quadruple-indexed, double/triple-enzyme RADseq libraries (2RAD/3RAD). *PeerJ, 7*, e7724. doi:10.7717/peerj.7724

Bernt, M., Donath, A., Juhling, F., Externbrink, F., Florentz, C., Fritzsch, G., . . . Stadler, P. F. (2013). MITOS: improved de novo metazoan mitochondrial genome annotation. *Mol Phylogenet Evol, 69*(2), 313-319. doi:10.1016/j.ympev.2012.08.023

Breniere, S. F., Salas, R., Buitrago, R., Bremond, P., Sosa, V., Bosseno, M. F., . . . Barnabe, C. (2013). Wild populations of *Triatoma infestans* are highly connected

to intra-peridomestic conspecific populations in the Bolivian Andes. *PLoS One, 8*(11), e80786. doi:10.1371/journal.pone.0080786

Breniere, S. F., Waleckx, E., Magallon-Gastelum, E., Bosseno, M. F., Hardy, X., Ndo, C., . . . Kengne, P. (2012). Population genetic structure of *Meccus longipennis* (Hemiptera, Reduviidae, Triatominae), vector of Chagas disease in West Mexico. *Infect Genet Evol, 12*(2), 254-262. doi:10.1016/j.meegid.2011.11.003

Brown, J. E., Evans, B. R., Zheng, W., Obas, V., Barrera-Martinez, L., Egizi, A., . . . Powell, J. R. (2014). Human impacts have shaped historical and recent evolution in Aedes aegypti, the dengue and yellow fever mosquito. *Evolution, 68*(2), 514-525. doi:10.1111/evo.12281

Calzada, J. E., Pineda, V., Garisto, J. D., Samudio, F., Santamaria, A. M., & Saldana, A. (2010). Human trypanosomiasis in the eastern region of the Panama Province: new endemic areas for Chagas disease. *Am J Trop Med Hyg, 82*(4), 580-582. doi:10.4269/ajtmh.2010.09-0397

Calzada, J. E., Pineda, V., Montalvo, E., Alvarez, D., Santamaria, A. M., Samudio, F., . . . Saldana, A. (2006). Human trypanosome infection and the presence of intradomicile Rhodnius pallescens in the Western Border of the Panama Canal, Panama. *American Journal of Tropical Medicine and Hygiene, 74*(5), 762-765.

Catala, S., Sachetto, C., Moreno, M., Rosales, R., Salazar-Schetrino, P. M., & Gorla, D. (2005). Antennal phenotype of Triatoma dimidiata populations and its relationship with species of phyllosoma and protracta complexes. *J Med Entomol, 42*(5), 719-725. doi:10.1093/jmedent/42.5.719

Ceballos, L. A., Piccinali, R. V., Marcet, P. L., Vazquez-Prokopec, G. M., Cardinal, M. V., Schachter-Broide, J., . . . Gurtler, R. E. (2011). Hidden sylvatic foci of the main vector of Chagas disease Triatoma infestans: threats to the vector elimination campaign? *PLoS Negl Trop Dis, 5*(10), e1365. doi:10.1371/journal.pntd.0001365

Costa, J., Correia, N. C., Neiva, V. L., Goncalves, T. C., & Felix, M. (2013). Revalidation and redescription of *Triatoma brasiliensis macromelasoma* Galvao, 1956 and an identification key for the *Triatoma brasiliensis* complex (Hemiptera: Reduviidae: Triatominae). *Mem Inst Oswaldo Cruz, 108*(6), 785-789. doi:10.1590/0074-0276108062013016

Costa, J., Dornak, L. L., Almeida, C. E., & Peterson, A. T. (2014). Distributional potential of the *Triatoma brasiliensis* species complex at present and under scenarios of future climate conditions. *Parasit Vectors, 7*(238). doi:Artn 238 10.1186/1756-3305-7-238

da Rosa, J. A., Justino, H. H. G., Nascimento, J. D., Mendonca, V. J., Rocha, C. S., de Carvalho, D. B., . . . de Oliveira, J. (2017). A new species of Rhodnius from Brazil (Hemiptera, Reduviidae, Triatominae). *Zookeys, 675*(675), 1-25. doi:10.3897/zookeys.675.12024

Davey, J. W., Hohenlohe, P. A., Etter, P. D., Boone, J. Q., Catchen, J. M., & Blaxter, M. L. (2011). Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nat Rev Genet, 12*(7), 499-510. doi:10.1038/nrg3012

Dias, F. B., Jaramillo, O. N., & Diotaiuti, L. (2014). Description and characterization of the melanic morphotype of *Rhodnius nasutus* Stal, 1859 (Hemiptera: Reduviidae:

Triatominae). *Rev Soc Bras Med Trop, 47*(5), 637-641. doi:10.1590/0037-8682-0007-2014

Dias, F. B. S., Bezerra, C. M., Machado, E. M. d. M., Casanova, C., & Diotaiuti, L. (2008). Ecological aspects of Rhodnius nasutus Stal, 1859 (Hemiptera: Reduviidae: Triatominae) in palms of the Chapada do Araripe in Ceara, Brazil. *Mem Inst Oswaldo Cruz, 103*(8), 824-830.

Dierckxsens, N., Mardulyn, P., & Smits, G. (2017). NOVOPlasty: de novo assembly of organelle genomes from whole genome data. *Nucleic Acids Res, 45*(4), e18. doi:10.1093/nar/gkw955

Donath, A., Juhling, F., Al-Arab, M., Bernhart, S. H., Reinhardt, F., Stadler, P. F., . . . Bernt, M. (2019). Improved annotation of protein-coding genes boundaries in metazoan mitochondrial genomes. *Nucleic Acids Res*. doi:10.1093/nar/gkz833

Dray, S., & Dufour, A. B. (2007). The ade4 package: Implementing the duality diagram for ecologists. *Journal of Statistical Software, 22*(4), 1-20. doi:10.18637/jss.v022.i04

Dujardin, J. P., Costa, J., Bustamante, D., Jaramillo, N., & Catala, S. (2009). Deciphering morphology in Triatominae: the evolutionary signals. *Acta Trop, 110*(2-3), 101-111. doi:10.1016/j.actatropica.2008.09.026

Dujardin, J. P., Panzera, P., & Schofield, C. J. (1999). Triatominae as a model of morphological plasticity under ecological pressure. *Mem Inst Oswaldo Cruz, 94*, 223-228. doi:Doi 10.1590/S0074-02761999000700036

Dujardin, J. P., Schofield, C. J., & Tibayrenc, M. (1998). Population structure of Andean Triatoma infestans: allozyme frequencies and their epidemiological relevance. *Med Vet Entomol, 12*(1), 20-29. doi:10.1046/j.1365-2915.1998.00076.x

Earl, D. A., & vonHoldt, B. M. (2011). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources, 4*(2), 359-361. doi:10.1007/s12686-011-9548-7

Excoffier, L., & Lischer, H. E. (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour, 10*(3), 564-567. doi:10.1111/j.1755-0998.2010.02847.x

Faircloth, B. C., & Glenn, T. C. (2012). Not all sequence tags are created equal: designing and validating sequence identification tags robust to indels. *PLoS One, 7*(8), e42543. doi:10.1371/journal.pone.0042543

Fitzpatrick, S., Dora Feliciangeli, M., Sanchez-Martin, M. J., Monteiro, F. A., & Miles, M. A. (2008). Molecular Genetics Reveal That Silvatic Rhodnius prolixus Do Colonise Rural Houses. *PLoS Negl Trop Dis, 2*(4).

Forattini, O. P., Rocha e Silva, E. O. d., Ferreira, O. A., Rabello, E. X., & Pattoli, D. G. B. (1971). Aspectos ecológicos da tripanossomose americana: III - dispersão local de triatomíneos, com especial referência ao Triatoma sordida. *Rev Saude Publica, 5*(2), 193-205. doi:10.1590/s0034-89101971000200002

Gaunt, M., & Miles, M. (2000). The ecotopes and evolution of triatomine bugs (triatominae) and their associated trypanosomes. *Mem Inst Oswaldo Cruz, 95*(4), 557-565.

Glenn, T. C. (2011). Field guide to next-generation DNA sequencers. *Mol Ecol Resour, 11*(5), 759-769. doi:10.1111/j.1755-0998.2011.03024.x

Glenn, T. C., Nilsen, R. A., Kieran, T. J., Sanders, J. G., Bayona-Vasquez, N. J., Finger, J. W., . . . Faircloth, B. C. (2019). Adapterama I: universal stubs and primers for 384 unique dual-indexed or 147,456 combinatorially-indexed Illumina libraries (iTru & iNext). *PeerJ, 7*, e7755. doi:10.7717/peerj.7755

Gomez-Palacio, A., Jaramillo, O. N., Caro-Riano, H., Diaz, S., Monteiro, F. A., Perez, R., . . . Triana, O. (2012). Morphometric and molecular evidence of intraspecific biogeographical differentiation of *Rhodnius pallescens* (HEMIPTERA: REDUVIIDAE: RHODNIINI) from Colombia and Panama. *Infect Genet Evol, 12*(8), 1975-1983. doi:10.1016/j.meegid.2012.04.003

Gomez-Palacio, A., Jaramillo-Ocampo, N., Triana-Chavez, O., Saldana, A., Calzada, J., Perez, R., & Panzera, F. (2008). Chromosome variability in the chagas disease vector *Rhodnius pallescens* (Hemiptera, Reduviidae, Rhodniini). *Mem Inst Oswaldo Cruz, 103*(2), 160-164.

Gomez-Palacio, A., Triana, O., Jaramillo, O. N., Dotson, E. M., & Marcet, P. L. (2013). Eco-geographical differentiation among Colombian populations of the Chagas disease vector *Triatoma dimidiata* (Hemiptera: Reduviidae). *Infect Genet Evol, 20*, 352-361. doi:10.1016/j.meegid.2013.09.003

Graham, C. F., Glenn, T. C., McArthur, A. G., Boreham, D. R., Kieran, T., Lance, S., . . . Somers, C. M. (2015). Impacts of degraded DNA on restriction enzyme associated DNA sequencing (RADSeq). *Mol Ecol Resour*. doi:10.1111/1755-0998.12404

Gulia-Nuss, M., Nuss, A. B., Meyer, J. M., Sonenshine, D. E., Roe, R. M., Waterhouse, R. M., . . . Hill, C. A. (2016). Genomic insights into the Ixodes scapularis tick vector of Lyme disease. *Nat Commun, 7*, 10507. doi:10.1038/ncomms10507

Hernandez, L., Abrahan, L., Moreno, M., Gorla, D., & Catala, S. (2008). Phenotypic variability associated to genomic changes in the main vector of Chagas disease in the southern cone of South America. *Acta Trop, 106*(1), 60-67. doi:10.1016/j.actatropica.2008.01.006

Hernandez-Castro, L. E., Paterno, M., Villacis, A. G., Andersson, B., Costales, J. A., De Noia, M., . . . Llewellyn, M. S. (2017). 2b-RAD genotyping for population genomic studies of Chagas disease vectors: *Rhodnius ecuadoriensis* in Ecuador. *PLoS Negl Trop Dis, 11*(7), e0005710. doi:10.1371/journal.pntd.0005710

Hoffberg, S. L., Kieran, T. J., Catchen, J. M., Devault, A., Faircloth, B. C., Mauricio, R., & Glenn, T. C. (2016). RADcap: sequence capture of dual-digest RADseq libraries with identifiable duplicates and reduced missing data. *Mol Ecol Resour, 16*(5), 1264-1278. doi:10.1111/1755-0998.12566

Jombart, T., & Ahmed, I. (2011). adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics, 27*(21), 3070-3071. doi:10.1093/bioinformatics/btr521

Justi, S. A., Galvao, C., & Schrago, C. G. (2016). Geological Changes of the Americas and their Influence on the Diversification of the Neotropical Kissing Bugs (Hemiptera: Reduviidae: Triatominae). *PLoS Negl Trop Dis, 10*(4), e0004527. doi:10.1371/journal.pntd.0004527

Kamvar, Z. N., Tabima, J. F., & Grunwald, N. J. (2014). Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ, 2*, e281. doi:10.7717/peerj.281

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol, 30*(4), 772-780. doi:10.1093/molbev/mst010

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., . . . Drummond, A. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics, 28*(12), 1647-1649. doi:10.1093/bioinformatics/bts199

Keenan, K. (2014). **Package 'diveRsity'**. *https://cran.r-project.org/web/packages/diveRsity/diveRsity.pdf*. Retrieved from https://cran.r-project.org/web/packages/diveRsity/diveRsity.pdf

Khatchikian, C. E., Foley, E. A., Barbu, C. M., Hwang, J., Ancca-Juarez, J., Borrini-Mayori, K., . . . Chagas Disease Working Group in, A. (2015). Population structure of the Chagas disease vector *Triatoma infestans* in an urban environment. *PLoS Negl Trop Dis, 9*(2), e0003425. doi:10.1371/journal.pntd.0003425

Kieran, T. J., Arnold, K. M. H., Thomas, J. C. t., Varian, C. P., Saldana, A., Calzada, J. E., . . . Gottdenker, N. L. (2019). Regional biogeography of microbiota composition in the Chagas disease vector Rhodnius pallescens. *Parasit Vectors, 12*(1), 504. doi:10.1186/s13071-019-3761-8

Kieran, T. J., Gordon, E. R. L., Forthman, M., Hoey-Chamberlain, R., Kimball, R. T., Faircloth, B. C., . . . Glenn, T. C. (2019). Insight from an ultraconserved element bait set designed for hemipteran phylogenetics integrated with genomic resources. *Mol Phylogenet Evol, 130*, 297-303. doi:10.1016/j.ympev.2018.10.026

Kieran, T. J., Gottdenker, N. L., Varian, C. P., Saldana, A., Means, N., Owens, D., . . . Glenn, T. C. (2017). Blood Meal Source Characterization Using Illumina Sequencing in the Chagas Disease Vector *Rhodnius pallescens* (Hemiptera: Reduviidae) in Panama. *J Med Entomol, 54*(6), 1786-1789. doi:10.1093/jme/tjx170

Kivisild, T. (2015). Maternal ancestry and population history from whole mitochondrial genomes. *Investig Genet, 6*(3), 3. doi:10.1186/s13323-015-0022-2

Kotsakiozi, P., Richardson, J. B., Pichler, V., Favia, G., Martins, A. J., Urbanelli, S., . . . Caccone, A. (2017). Population genomics of the Asian tiger mosquito, Aedes albopictus: insights into the recent worldwide invasion. *Ecol Evol, 7*(23), 10143-10157. doi:10.1002/ece3.3514

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*, 123-520.

McCoy, K. D. (2008). The population genetic structure of vectors and our understanding of disease epidemiology. *Parasite, 15*(3), 444-448. doi:10.1051/parasite/2008153444

Monsalve, Y., Panzera, F., Herrera, L., Triana-Chavez, O., & Gomez-Palacio, A. (2016). Population differentiation of the Chagas disease vector Triatoma maculata

(Erichson, 1848) from Colombia and Venezuela. *J Vector Ecol, 41*(1), 72-79. doi:10.1111/jvec.12196

Monteiro, F., Marcet, P., & Dorn, P. (2010). Population Genetics of Triatomines. In J. Telleria & M. Tibayrenc (Eds.), *American Trypanosomiasis* (1st ed., pp. 169-208): Elsevier Inc.

Nascimento, J. D., da Rosa, J. A., Salgado-Roa, F. C., Hernandez, C., Pardo-Diaz, C., Alevi, K. C. C., . . . Ramirez, J. D. (2019). Taxonomical over splitting in the Rhodnius prolixus (Insecta: Hemiptera: Reduviidae) clade: Are R. taquarussuensis (da Rosa et al., 2017) and R. neglectus (Lent, 1954) the same species? *PLoS One, 14*(2), e0211285. doi:10.1371/journal.pone.0211285

Noireau, F., Abad-Franch, F., Valente, S. A., Dias-Lima, A., Lopes, C. M., Cunha, V., . . . Jurberg, J. (2002). Trapping Triatominae in silvatic habitats. *Mem Inst Oswaldo Cruz, 97*(1), 61-63. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/11992149

Noireau, F., Flores, R., Gutierrez, T., & Dujardin, J. P. (1997). Detection of sylvatic dark morphs of *Triatoma infestans* in the Bolivian Chaco. *Mem Inst Oswaldo Cruz, 92*(5), 583-584. doi:10.1590/s0074-02761997000500003

Orantes, L. C., Monroy, C., Dorn, P. L., Stevens, L., Rizzo, D. M., Morrissey, L., . . . Helms Cahan, S. (2018). Uncovering vector, parasite, blood meal and microbiome patterns from mixed-DNA specimens of the Chagas disease vector *Triatoma dimidiata*. *PLoS Negl Trop Dis, 12*(10), e0006730. doi:10.1371/journal.pntd.0006730

Pagès, H., Aboyoun, P., Gentleman, R., & DebRoy, S. (2019). Biostrings: Efficient manipulation of biological strings. R package. (Version 2.52.0.).

Panzera, F., Dujardin, J. P., Nicolini, P., Caraccio, M. N., Rose, V., Tellez, T., . . . Perez, R. (2004). Genomic changes of Chagas disease vector, South America. *Emerg Infect Dis, 10*(3), 438-446. doi:10.3201/eid1003.020812

Paris, J. R., Stevens, J. R., Catchen, J. M., & Johnston, S. (2017). Lost in parameter space: a road map for stacks. *Methods in Ecology and Evolution, 8*(10), 1360-1373. doi:10.1111/2041-210x.12775

Peretolchina, T., Pavan, M. G., Correa-Antonio, J., Gurgel-Goncalves, R., Lima, M. M., & Monteiro, F. A. (2018). Phylogeography and demographic history of the Chagas disease vector Rhodnius nasutus (Hemiptera: Reduviidae) in the Brazilian Caatinga biome. *PLoS Negl Trop Dis, 12*(9), e0006731. doi:10.1371/journal.pntd.0006731

Perez de Rosas, A. R., Segura, E. L., Fichera, L., & Garcia, B. A. (2008). Macrogeographic and microgeographic genetic structure of the Chagas' disease vector *Triatoma infestans* (Hemiptera: Reduviidae) from Catamarca, Argentina. *Genetica, 133*(3), 247-260. doi:10.1007/s10709-007-9208-8

Perez de Rosas, A. R., Segura, E. L., & Garcia, B. A. (2011). Molecular phylogeography of the Chagas' disease vector *Triatoma infestans* in Argentina. *Heredity (Edinb), 107*(1), 71-79. doi:10.1038/hdy.2010.159

Pfeiler, E., Bitler, B. G., Ramsey, J. M., Palacios-Cardiel, C., & Markow, T. A. (2006). Genetic variation, population structure, and phylogenetic relationships of Triatoma rubida and T. recurva (Hemiptera: Reduviidae: Triatominae) from the

Sonoran Desert, insect vectors of the Chagas' disease parasite Trypanosoma cruzi. *Mol Phylogenet Evol, 41*(1), 209-221. doi:10.1016/j.ympev.2006.07.001

Piccinali, R. V., & Gürtler, R. E. (2015). Fine-scale genetic structure of *Triatoma infestans* in the Argentine Chaco. *Infection, Genetics and Evolution, 34*, 143-152. doi:10.1016/j.meegid.2015.05.030

Piccinali, R. V., Marcet, P. L., Noireau, F., Kitron, U., Gurtler, R. E., & Dotson, E. M. (2009). Molecular population genetics and phylogeography of the Chagas disease vector *Triatoma infestans* in South America. *J Med Entomol, 46*(4), 796-809. doi:10.1603/033.046.0410

Pires, H. H. R., Abrao, D. O., Machado, E. M. D., Schofield, C. J., & Diotaiuti, L. (2002). Eye colour as a genetic marker for fertility and fecundity of Triatoma infestans (Klug, 1834) Hemiptera, Reduviidae, Triatominae. *Mem Inst Oswaldo Cruz, 97*(5), 675-678. doi:Doi 10.1590/S0074-02762002000500016

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of Population Structure Using Multilocus Genotype Data. *Genetics, 155*(2), 945-959.

Ramirez, C. J., Jaramillo, C. A., del Pilar Delgado, M., Pinto, N. A., Aguilera, G., & Guhl, F. (2005). Genetic structure of sylvatic, peridomestic and domestic populations of *Triatoma dimidiata* (Hemiptera: Reduviidae) from an endemic zone of Boyaca, Colombia. *Acta Trop, 93*(1), 23-29. doi:10.1016/j.actatropica.2004.09.001

Raymond, M., & Rousset, F. (1995). Genepop (Version-1.2) - Population-Genetics Software for Exact Tests and Ecumenicism. *Journal of Heredity, 86*(3), 248-249. doi:DOI 10.1093/oxfordjournals.jhered.a111573

Rinker, D. C., Pitts, R. J., & Zwiebel, L. J. (2016). Disease vectors in the era of next generation sequencing. *Genome Biol, 17*(1), 95. doi:10.1186/s13059-016-0966-4

Rochette, N. C., Rivera-Colon, A. G., & Catchen, J. M. (2019). *Stacks 2: Analytical Methods for Paired-end Sequencing Improve RADseq-based Population Genomics*. bioRxiv. Retrieved from https://www.biorxiv.org/content/10.1101/615385v1

Roden, A. E., Champagne, D. E., & Forschler, B. T. (2011). Biogeography of Triatoma sanguisuga (Hemiptera: Reduviidae) on two barrier islands off the coast of Georgia, United States. *J Med Entomol, 48*(4), 806-812. doi:10.1603/me11049

Rodriguez, I. G., & Loaiza, J. R. (2017). American trypanosomiasis, or Chagas disease, in Panama: a chronological synopsis of ecological and epidemiological research. *Parasit Vectors, 10*(1), 459. doi:10.1186/s13071-017-2380-5

RStudio Team. (2015). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA URL http://www.rstudio.com/.

Rubinoff, D., & Holland, B. S. (2005). Between two extremes: mitochondrial DNA is neither the panacea nor the nemesis of phylogenetic and taxonomic inference. *Syst Biol, 54*(6), 952-961. doi:10.1080/10635150500234674

Saldana, A., Pineda, V., Martinez, I., Santamaria, G., Santamaria, A. M., Miranda, A., & Calzada, J. E. (2012). A new endemic focus of Chagas disease in the northern region of Veraguas Province, Western Half Panama, Central America. *PLoS One, 7*(4), e34657. doi:10.1371/journal.pone.0034657

Saldana, A., Santamaria, A. M., Pineda, V., Vasquez, V., Gottdenker, N. L., & Calzada, J. E. (2018). A darker chromatic variation of *Rhodnius pallescens* infected by

specific genetic groups of *Trypanosoma rangeli* and *Trypanosoma cruzi* from Panama. *Parasit Vectors, 11*(1), 423. doi:10.1186/s13071-018-3004-4

Schofield, C. J. (1994). Triatominae: Biología y control. In (pp. 80). London, UK: Eurocommunica Publications.

Schofield, C. J., & Galvão, C. (2009). Classification, evolution, and species groups within the Triatominae. *Acta Trop, 110*(2-3), 88-100. doi:10.1016/j.actatropica.2009.01.010

Souza, E. D., Von Atzingen, N. C., Furtado, M. B., de Oliveira, J., Nascimento, J. D., Vendrami, D. P., . . . da Rosa, J. A. (2016). Description of *Rhodnius marabaensis sp*. n. (Hemiptera, Reduviidae, Triatominae) from Para State, Brazil. *Zookeys, 621*(621), 45-62. doi:10.3897/zookeys.621.9662

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics, 30*(9), 1312-1313. doi:10.1093/bioinformatics/btu033

Stevens, L., Monroy, M. C., Rodas, A. G., Hicks, R. M., Lucero, D. E., Lyons, L. A., & Dorn, P. L. (2015). Migration and Gene Flow Among Domestic Populations of the Chagas Insect Vector *Triatoma dimidiata* (Hemiptera: Reduviidae) Detected by Microsatellite Loci. *J Med Entomol, 52*(3), 419-428. doi:10.1093/jme/tjv002

Toonen, R. J., Puritz, J. B., Forsman, Z. H., Whitney, J. L., Fernandez-Silva, I., Andrews, K. R., & Bird, C. E. (2013). ezRAD: a simplified method for genomic genotyping in non-model organisms. *PeerJ, 1*, e203. doi:10.7717/peerj.203

Villacís, A. G., Grijalva, M. J., & Catalá, S. S. (2010). Phenotypic Variability of Rhodnius ecuadoriensis Populations at the Ecuadorian Central and Southern Andean Region. *J Med Entomol, 47*(6), 1034-1043. doi:10.1603/me10053

Villacis, A. G., Marcet, P. L., Yumiseva, C. A., Dotson, E. M., Tibayrenc, M., Breniere, S. F., & Grijalva, M. J. (2017). Pioneer study of population genetics of Rhodnius ecuadoriensis (Hemiptera: Reduviidae) from the central coastand southern Andean regions of Ecuador. *Infect Genet Evol, 53*, 116-127. doi:10.1016/j.meegid.2017.05.019

Waleckx, E., Salas, R., Huaman, N., Buitrago, R., Bosseno, M. F., Aliaga, C., . . . Breniere, S. F. (2011). New insights on the Chagas disease main vector *Triatoma infestans* (Reduviidae, Triatominae) brought by the genetic analysis of Bolivian sylvatic populations. *Infect Genet Evol, 11*(5), 1045-1057. doi:10.1016/j.meegid.2011.03.020

Webster, J. P., Gower, C. M., Knowles, S. C., Molyneux, D. H., & Fenton, A. (2016). One health - an ecological and evolutionary framework for tackling Neglected Zoonotic Diseases. *Evol Appl, 9*(2), 313-333. doi:10.1111/eva.12341

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-Statistics for the Analysis of Population Structure. *Evolution, 38*(6), 1358-1370. doi:10.1111/j.1558-5646.1984.tb05657.x

WHO. (2019). Chagas Disease (American trypanosomiasis) fact sheet. Retrieved from http://www.who.int/mediacentre/factsheets/fs340/en/

Woolhouse, M. E. J., & Dye, C. (2001). Population biology of emerging and re-emerging pathogens - Preface. *PHILOSOPHICAL TRANSACTIONS OF THE ROYAL*

SOCIETY OF LONDON SERIES B-BIOLOGICAL SCIENCES, 356(1411), 981-982. doi:DOI 10.1098/rstb.2001.0899

Zink, R. M. (2010). Drawbacks with the use of microsatellites in phylogeography: the song sparrow *Melospiza melodiaas* a case study. *Journal of Avian Biology, 41*(1), 1-7. doi:10.1111/j.1600-048X.2009.04903.x

Tables

Table 4.1. Gene and genetic diversity indices per sampling site over all SNPs for *R. pallescens*. Asterisks represent highly significant values ($p < 0.05$).

	Gene Diversity within Individuals	Gene Diversity among Individuals	F_{IS}	H_O (SD)	H_E (SD)	HWE
CH	0.084	0.080	-0.056	0.084 (0.17)	0.120 (0.24)	1.0
LP	0.094	0.088	-0.069	0.095 (0.17)	0.087 (0.14)	*
SR	0.073	0.076	0.034	0.073 (0.16)	0.121 (0.25)	1.0
TM	0.084	0.084	-0.003	0.085 (0.15)	0.083 (0.13)	*
SF	0.092	0.118	0.222	0.092 (0.15)	0.116 (0.18)	*
GLOBAL				0.090	0.090	*

Table 4.2. *R. pallescens* genetic differentiation estimates between palms and between sampling sites on 1,794 SNPs. Below diagonals, pairwise F_{ST} values. Above diagonals, *p*-value after 1000 permutations (significant values in bold).

		Palms												Sites				
F_{ST}		**CH-1**	**CH-2**	**LP-1**	**LP-2**	**SR-1**	**SR-2**	**TM-1**	**TM-2**	**SF-1**	**SF-2**	**SF-3**	**SF-4**	**CH**	**LP**	**SR**	**TM**	**SF**
	CH-1		0.059	0.037	0.024	0.013	0.024	0.009	0.005	0.011	0.012	0.006	0.009					
	CH-2	0.018		0.003	**<0.001**	0.002	0.002	0.018	0.004	0.004	0.008	0.002	0.001					
	LP-1	0.010	**0.001**		**0.001**	0.005	0.002	**0.001**	**0.001**	0.001	0.002	0.001	**0.001**					
	LP-2	0.024	**<0.001**	**0.001**		0.000	0.000	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**					
	SR-1	0.013	0.002	0.005	0.000		0.029	0.009	0.007	0.007	0.004	0.002	0.004					
	SR-2	0.024	0.002	0.002	0.000	0.029		0.015	0.036	0.002	0.003	0.004	0.003					
	TM-1	0.009	0.018	**0.001**	**<0.001**	0.009	0.015		0.000	**<0.001**	**<0.001**	**<0.001**	**<0.001**					
	TM-2	0.005	0.004	**0.001**	**<0.001**	0.007	0.036	0.000		**<0.001**	**<0.001**	**<0.001**	**<0.001**					
	SF-1	0.011	0.004	0.001	**<0.001**	0.007	0.002	**<0.001**	**<0.001**		0.000	0.000	0.000					
	SF-2	0.012	0.008	0.002	**<0.001**	0.004	0.003	**<0.001**	**<0.001**	0.000		0.000	0.000					
	SF-3	0.006	0.002	0.001	**<0.001**	0.002	0.004	**<0.001**	**<0.001**	0.000	0.000		0.000					
	SF-4	0.009	0.001	**0.001**	**<0.001**	0.004	0.003	**<0.001**	**<0.001**	0.000	0.000	0.000						
Palms	**CH-1**		**0.059**	**0.037**	**0.024**	**0.013**	**0.024**	**0.009**	**0.005**	**0.011**	**0.012**	**0.006**	**0.009**					
	CH-2	0.018		**0.003**	**<0.001**	**0.002**	**0.002**	**0.018**	**0.004**	**0.004**	**0.008**	**0.002**	**0.001**					
	LP-1	0.010	0.001		**0.001**	**0.005**	**0.002**	**0.001**	**0.001**	**0.001**	**0.002**	**0.001**	**0.001**					
	LP-2	0.024	<0.001	0.001		**0.000**	**0.000**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**					
	SR-1	0.013	0.002	0.005	0.000		**0.029**	**0.009**	**0.007**	**0.007**	**0.004**	**0.002**	**0.004**					
	SR-2	0.024	0.002	0.002	0.000	0.029		**0.015**	**0.036**	**0.002**	**0.003**	**0.004**	**0.003**					
	TM-1	0.009	0.018	0.001	<0.001	0.009	0.015		**0.000**	**<0.001**	**<0.001**	**<0.001**	**<0.001**					
	TM-2	0.005	0.004	0.001	<0.001	0.007	0.036	0.000		**<0.001**	**<0.001**	**<0.001**	**<0.001**					
	SF-1	0.011	0.004	0.001	<0.001	0.007	0.002	<0.001	<0.001		**0.000**	**0.000**	**0.000**					
	SF-2	0.012	0.008	0.002	<0.001	0.004	0.003	<0.001	<0.001	0.000		**0.000**	**0.000**					
	SF-3	0.006	0.002	0.001	<0.001	0.002	0.004	<0.001	<0.001	0.000	0.000		**0.000**					
	SF-4	0.009	0.001	0.001	<0.001	0.004	0.003	<0.001	<0.001	0.000	0.000	0.000						

	M	36	04	05	45	03	00	08		**00**	01	**00**	**00**					
	-2		2	3		0	6			**1**		**1**	**1**					
			0.	0.		0.	0.						**<0.**					
	SF	0.8	82	82	0.8	82	82	0.8	0.8		0.0	0.0	**00**					
	-1	19	1	0	35	5	7	35	22		11	01	**1**					
			0.	0.		0.	0.						**<0.**					
	SF	0.8	81	81	0.8	81	81	0.8	0.8	0.0		0.0	**00**					
	-2	09	2	2	30	6	8	30	15	42		05	**1**					
			0.	0.		0.	0.						**<0.**					
	SF	0.8	80	81	0.8	81	81	0.8	0.8	0.0	0.0		**00**					
	-3	07	9	0	28	4	5	28	12	58	47		**1**					
			0.	0.		0.	0.											
	SF	0.8	85	84	0.8	85	85	0.8	0.8	0.3	0.3	0.3						
	-4	51	1	4	51	4	5	49	46	43	35	51						
															<0.		**<0.**	**<0.**
	C														**00**	0.0	**00**	**00**
	H														**1**	06	**1**	**1**
																<0.	**<0.**	**<0.**
														0.0		**00**	**00**	**00**
	LP													28		**1**	**1**	**1**
Sit																		**<0.**
es														0.0	0.0		0.0	**00**
	SR													18	31		01	**1**
																		<0.
	T													0.0	0.0	0.0		**00**
	M													50	43	19		**1**
														0.7	0.7	0.7	0.8	
	SF													87	99	93	03	

Table 4.3. Hierarchical AMOVA analyses for *R. pallescens*.

	df	Variance (%)	*F-statistic*	*p*-value
Among groups (CH-SR-LP-TM-SF)	4	69.46	$F_{CT} = 0.695$	0.004
Among locations within groups (Palms)	7	4.20	$F_{SC} = 0.137$	< 0.001
Within locations (Within Palms)	78	-1.14	$F_{IS} = -0.043$	0.969

Figures

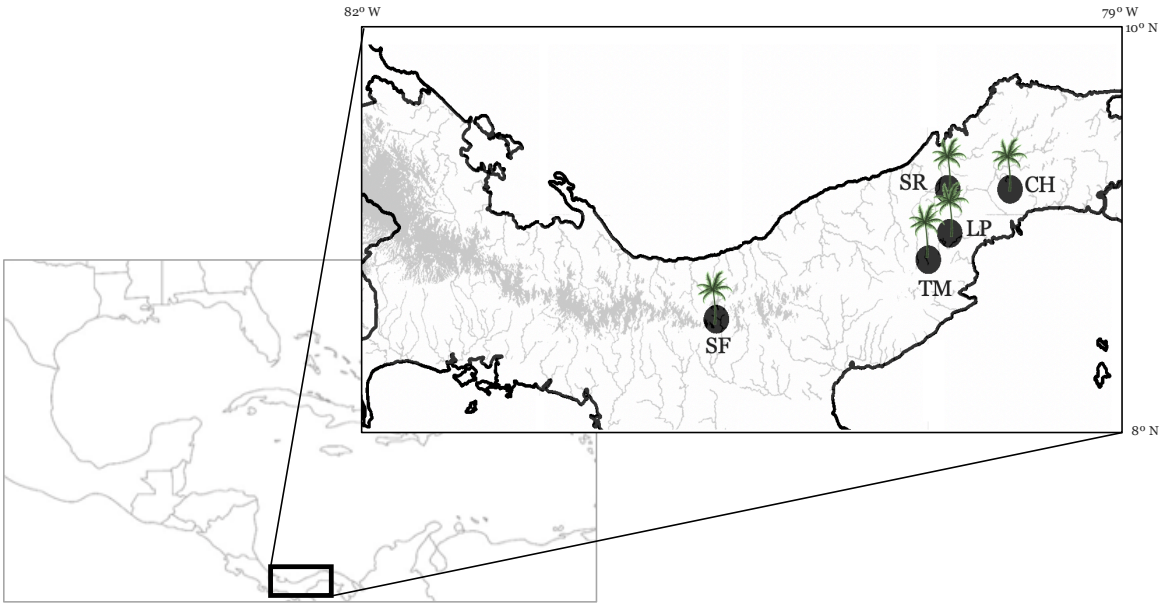


Figure 4.1. Sampling locations of *Rhodnius pallecens* in Panama. SF: Santa Fe (Veraguas), TM: Trinidad de las Minas, LP: Las Pavas, SR: Santa Rita, CH: Chilibre.

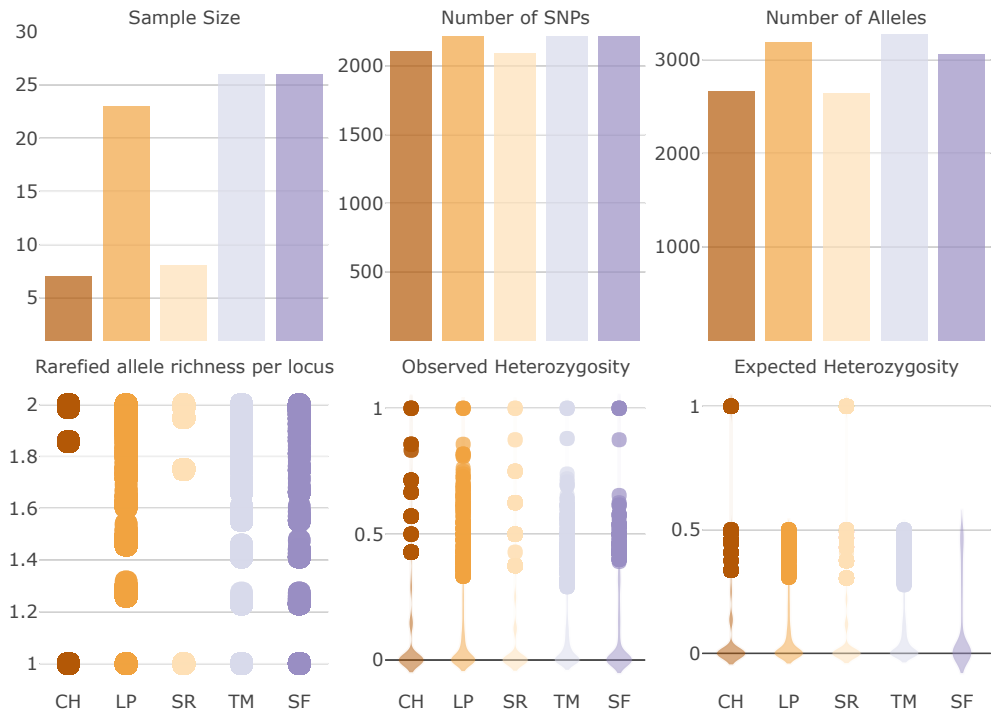


Fig. 4.2. Number of samples, number of SNPs, number of alleles, allele richness, and observed and expected heterozygosity per sampling site in *R. pallescens*.

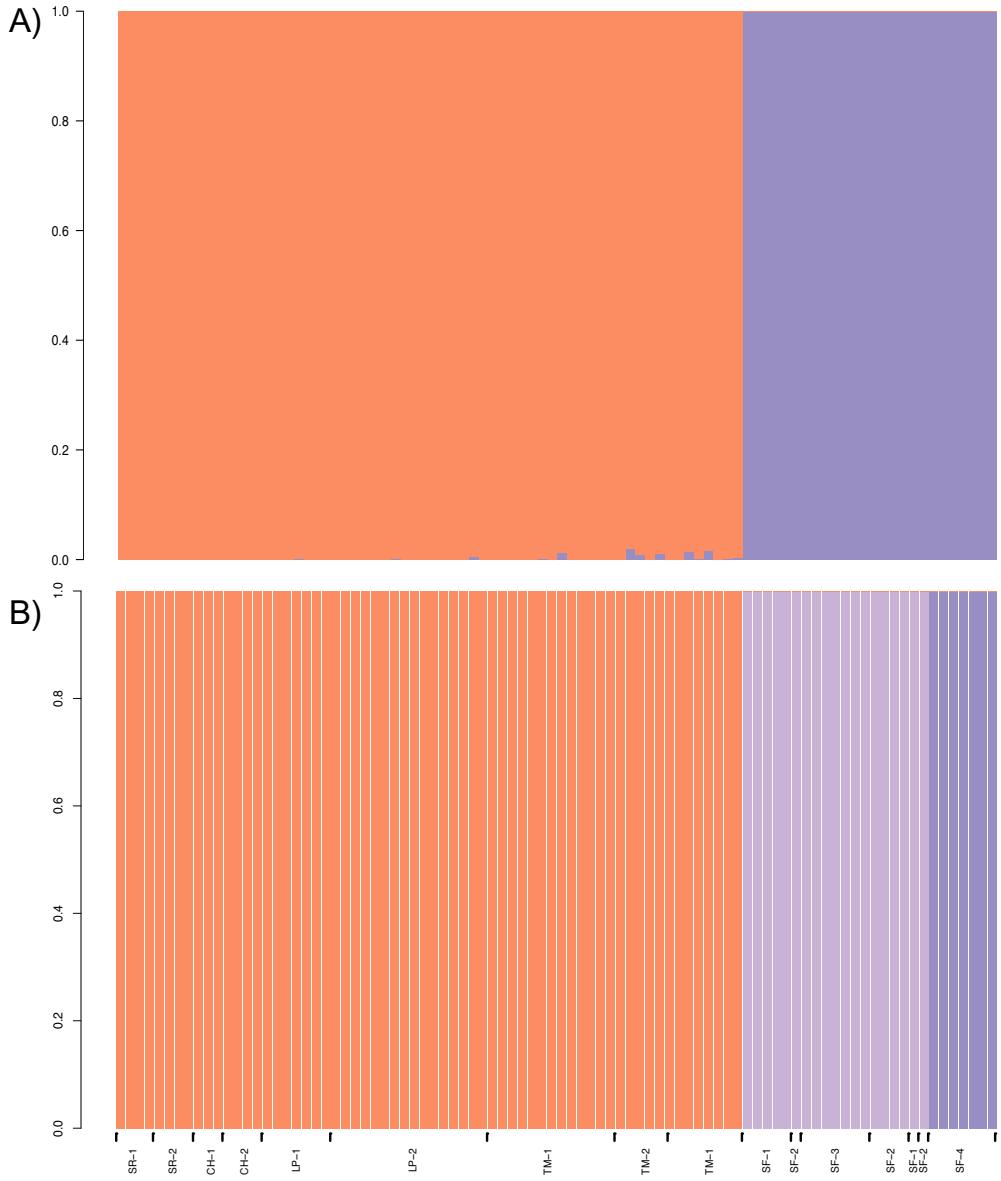


Fig. 4.3. *Rhodnius pallescens* individual assignment probabilities to *K* clusters inferred by **A)** STRUCTURE analysis where the *K* with highest posterior probability and Delta K is *K* = 2 using 867 SNPs (one randomly selected SNP per locus); and **B)** using the BIC score calculated with the *find.cluster* function in ADEGENET v. 2.1.1 (Jombart et al. 2011) using 2,216 SNPs (see main text for selection criteria). Tick labels correspond to palms within sampling sites

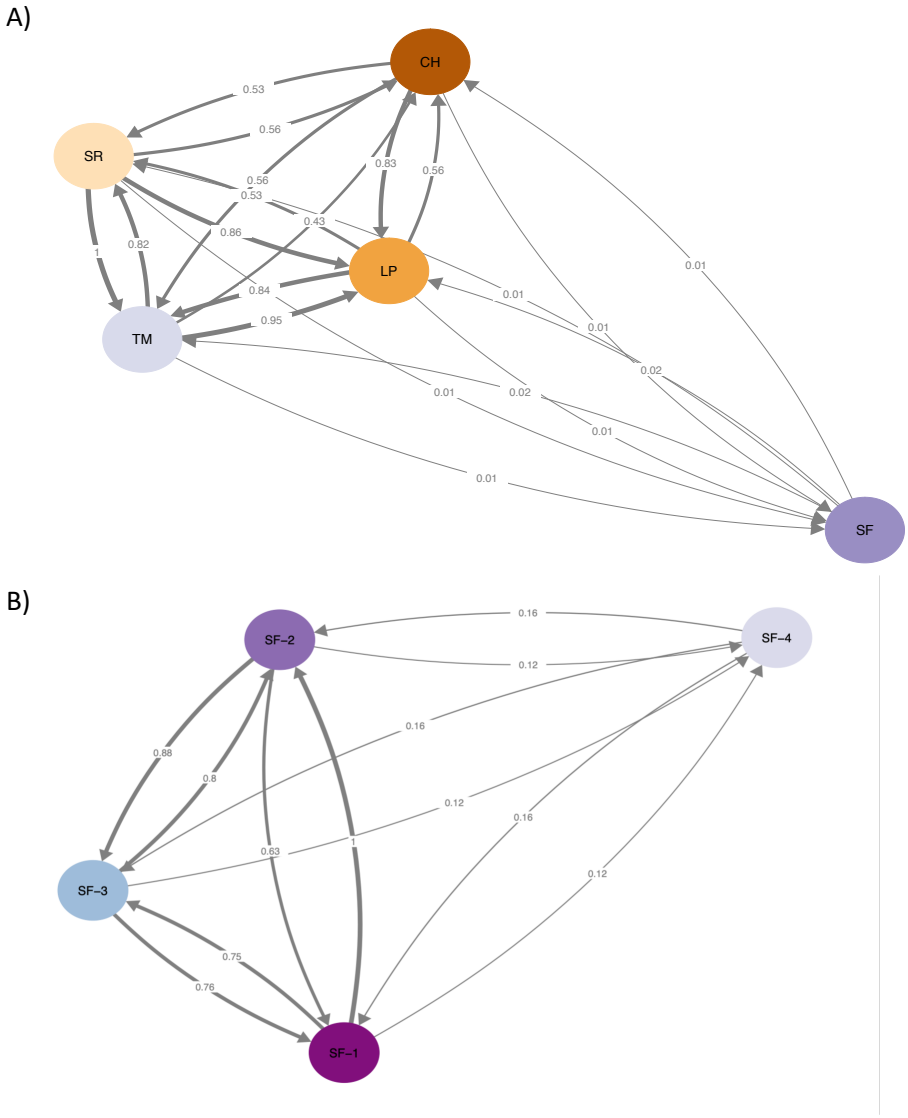


Fig. 4.4. Relative migration (*Nm*) network between A) sampling sites and B) palms within Santa Fe, for the kissing bug *Rhodnius pallescens* using 2,216 SNPs. Thicker arrows indicate stronger migration relationships compared to lighter arrows.

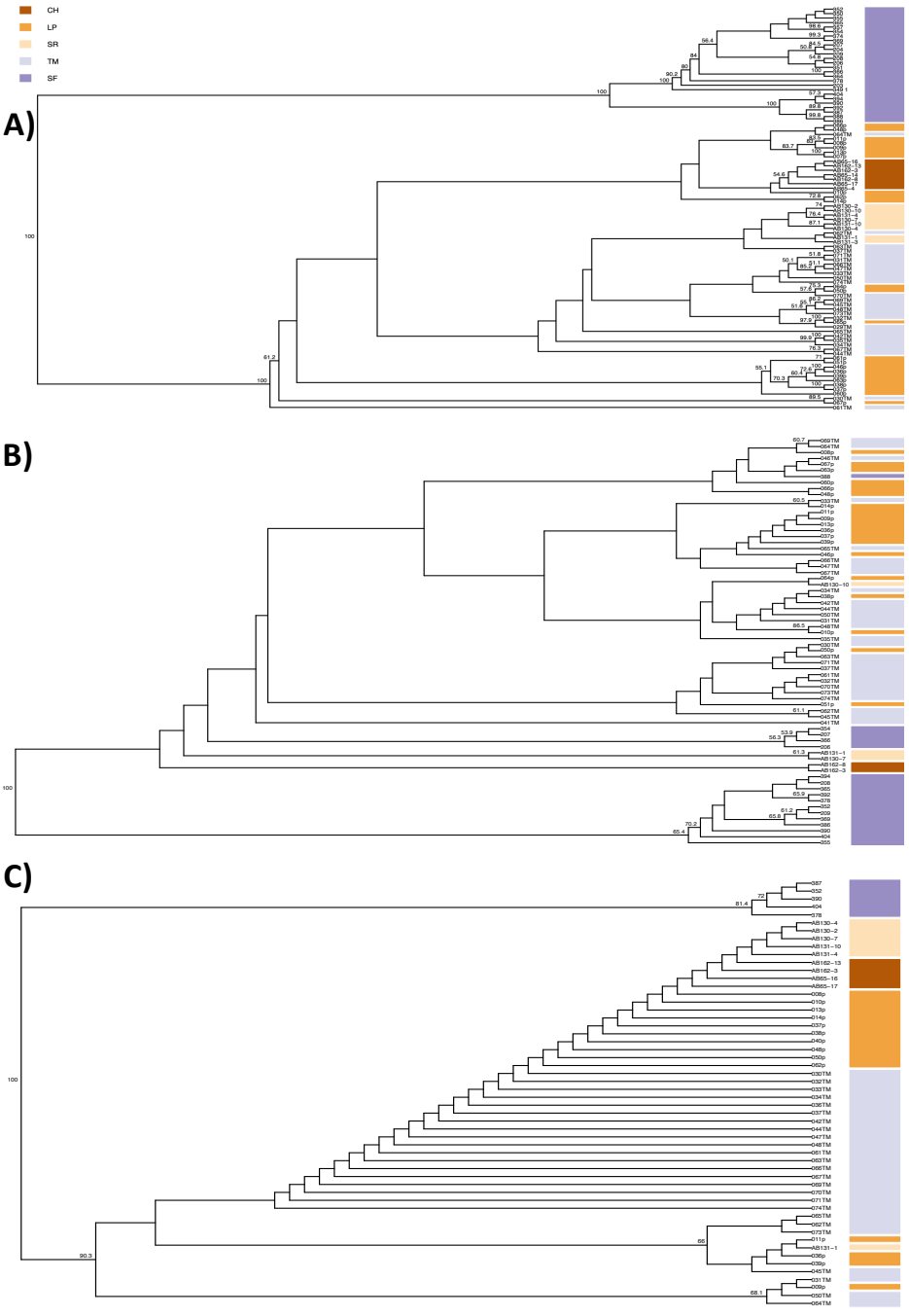


Fig. 4.5. UPGMA trees based on Nei's genetic distance with bootstrap support values for **A)** *Rhodnius pallescens* (2,216 SNPs); **B)** *Trypanosoma cruzi* (17 SNPs); and **C)** *Trypanosoma rangeli* (5 SNPs)

CHAPTER 5

INSIGHT FROM AN ULTRACONSERVED ELEMENT BAIT SET DESIGNED FOR HEMIPTERAN PHYLOGENETICS INTERGRATED WITH GENOMIC RESOURCES*

* Kieran TJ, Gordon ERL, Forthman M, Hoey-Chamberlain R, Kimball RT, Faircloth BC, Weirauch C, Glenn TC. 2019. *Molecular Phylogenetics and Evolution* 130: 297-303. Reprinted here with permission of the publisher.

Abstract

Target enrichment of conserved genomic regions facilitates collecting sequences of many orthologous loci from non-model organisms to address phylogenetic, phylogeographic, population genetic, and molecular evolution questions. Bait sets for sequence capture can simultaneously target thousands of loci, which opens new avenues of research on speciose groups. Current phylogenetic hypotheses on the >103,000 species of Hemiptera have failed to unambiguously resolve major nodes, suggesting that alternative datasets and more thorough taxon sampling may be required to resolve relationships. We here use a recently designed ultraconserved element (UCE) bait set for Hemiptera, with a focus on the suborder Heteroptera, or the true bugs, to test previously proposed relationships. We present newly generated UCE data for 36 samples representing three suborders, all seven heteropteran infraorders, 23 families, and 34 genera of Hemiptera and one thysanopteran outgroup. To improve taxon sampling, we also mined additional UCE loci in silico from published hemipteran genomic and transcriptomic data. We obtained 2,271 UCE loci for newly sequenced hemipteran taxa, ranging from 265 to 1,696 (average 904) per sample. These were similar in number to the data mined from transcriptomes and genomes, but with fewer loci overall. The amount of missing data correlates with greater phylogenetic divergence from taxa used to design the baits. This bait set hybridizes to a wide range of hemipteran taxa and specimens of varying quality, including dried specimens as old as 1973. Our estimated phylogeny yielded topologies consistent with other studies for most nodes and was strongly-supported. We also demonstrate that UCE loci are almost exclusively from the transcribed portion of the genome, thus data can be successfully integrated with existing

genomic and transcriptomic resources for more comprehensive phylogenetic sampling, an important feature in the era of phylogenomics. UCE approaches can be used by other researchers for additional studies on hemipteran evolution and other research that requires well resolved phylogenies.

Introduction

Hemiptera is the largest order of non-holometabolous insects with >103,000 described species. They are characterized by distinctive piercing-sucking mouthparts, which allow exploitation of plant vascular tissue in many species, including some economically important agricultural pests. While some hemipteran species can be beneficial predators, one group includes vectors of human diseases, and others are nuisance pests. Hemiptera are thought to have diverged from their sister taxon Thysanoptera, the thrips, more than 300 mya (Misof et al., 2014). Although ancestral herbivorous feeding habits have been retained in several hemipteran lineages including aphids, whiteflies and relatives (Sternorrhyncha), cicadas and relatives (Auchenorrhyncha), and moss bugs (Coleorrhyncha), life history strategies diversified in a fourth lineage, the suborder Heteroptera, to include predacious, hematophagous, mycetophagous and mixed-feeding habits (Weirauch, Schuh, Cassis, & Wheeler, 2018). Despite the diversity and economic importance of Hemiptera, phylogenetic relationships among and within major lineages, i.e., the suborders Sternorrhyncha, Auchenorrhyncha, Coleorrhyncha, and Heteroptera, have remained contentious (Cryan & Urban, 2012; H. Li et al., 2015; Song et al., 2016).

Based on increasingly extensive datasets with respect to taxon sampling and/or characters (few loci to complete mitochondrial genomes and transcriptomes), these analyses have converged on congruent topologies in certain parts of the tree (i.e., establishment of Auchenorrhyncha as more closely related to Heteroptera than to Sternorrhyncha). However, other major questions have remained unsolved, such as relationships among the early diverging lineages within Heteroptera (M. Li, Tian, Zhao, & Bu, 2012; Wang et al., 2016; Wang et al., 2017; Weirauch et al., 2018; Wheeler, Schuh, & Bang, 1993). Generating and analyzing comprehensive datasets with respect to both taxonomic and character sampling by using a large number of universal markers from throughout the genome has the potential to greatly advance our understanding of phylogenetic relationships across Hemiptera. Approaches using anchored hybrid enrichment (Lemmon, Emme, & Lemmon, 2012) or ultraconserved elements (UCEs) (B. C. Faircloth et al., 2012) generate such data for relatively low costs, making them feasible for taxon-rich phylogenetic analyses.

UCEs are highly conserved across divergent taxa (Bejerano et al., 2004) which make them useful as anchors for target enrichment. They have been shown to be useful markers for comparison across diverse taxa in vertebrates (Alexander et al., 2017; Crawford et al., 2012; Crawford et al., 2015; B. C. Faircloth et al., 2012; Gilbert et al., 2015; McCormack et al., 2013; Moyle et al., 2016; Smith, Harvey, Faircloth, Glenn, & Brumfield, 2014) and more recently in arthropods (Baca, Alexander, Gustafson, & Short, 2017; B. C. Faircloth, Branstetter, White, & Brady, 2015; Starrett et al., 2017; Van Dam et al., 2017). Sequence variability increases with distance from the conserved UCE (B. C. Faircloth et al., 2012), which allows for analyses at different phylogenetic scales, from

deep divergence (B. C. Faircloth, Sorenson, Santini, & Alfaro, 2013) to population level (Harvey, Smith, Glenn, Faircloth, & Brumfield, 2016; Manthey, Campillo, Burns, & Moyle, 2016). While most UCE studies in arthropods have focused on the Hymenoptera (Blaimer et al., 2015; Blaimer, Lloyd, Guillory, & Brady, 2016; S. Bossert, Murray, Blaimer, & Danforth, 2017; M. G. Branstetter et al., 2017; Branstetter, Longino, Ward, Faircloth, & Price, 2017; B. C. Faircloth et al., 2015), bait sets designed for other groups have recently been used in empirical studies (Baca et al., 2017; Starrett et al., 2017; Van Dam et al., 2017), which show the promising utility and effectiveness of UCEs in non-vertebrates. Having a universal set of genetic markers for a diverse group like Hemiptera can help standardize the phylogenetic data available and make comparative studies across multiple projects, questions, and scales easier for researchers. Faircloth (2017) recently designed and *in silico* tested UCE bait sets for several arthropod orders including Hemiptera. Only two of the designed bait sets have been used to generate UCE loci from samples and evaluated [Arachnida; (Starrett et al., 2017) and Coleoptera; (Baca et al., 2017)]. Here, we expand such testing to the Hemiptera UCE bait set.

Methods

Bait and Taxon Sampling

Hemiptera UCE capture baits (Brant C. Faircloth & Gilbert, 2017) consisting of 40,207 baits for 2,731 loci were tested on 36 hemipteran samples representing three suborders and the seven heteropteran infraorders (Table 5.1, Appendix N). Taxa were chosen to include a mix of species closely related to the taxa used for bait design and more distantly related taxa to assess UCE efficacy across Hemiptera. We also included

multiple individuals within two families, Coreidae (n=9) and Reduviidae: Triatominae (n=6), to assess the utility of the bait set for recovering shallower phylogenetic relationships.

DNA extraction and library preparation

For most specimens, genomic DNA was extracted from ethanol preserved and recently pinned specimens using a Qiagen DNeasy kit, and older pinned specimens using a Qiagen QIAquick PCR Clean Up kit. Specimens of Coreidae were extracted using a Puregene Solid Tissue kit (Appendix N). DNA concentration and quality were assessed on a Qubit, fragment analyzer, and a 1.5% agarose gel. Samples with higher molecular weight were fragmented on a Bioruptor UCD-300 sonication device (Diagenode) based on quality for 2–9 cycles of 30 s on/30 s off. Resulting fragments were in the range of 200–1000 bp.

Libraries were prepared with a KAPA Hyper Prep Kit (Kapa Biosystems) following manufacturer's protocol with a few modifications. Half volume reactions were performed on all samples. Universal TruSeq compatible adaptor stubs were ligated onto A-tailed DNA fragments. Adapter-ligated product was amplified using Illumina TruSeq compatible dual-indexed primers with modified 8 bp indexes (Glenn et al., 2016). PCR reactions were 25 µL consisting of 10 µL of adapter-ligated DNA, 12.5 µL 2X KAPA HiFi HotStart ReadyMix, and 2.5 µL each of the 5 µM dual-indexed primers. Thermocycler conditions were 98°C for 45 s, followed by 14 cycles of 98°C for 30 s, 60°C for 30 s, and 72°C for 30 s, and then a final extension of 72°C for 1 min. All clean-up steps used Sera-Mag magnetic beads (Thermo-Scientific, Waltham, MA, USA). Post-

PCR cleaned product was quantified on Qubit and equimolar amounts of 9–12 samples were combined into 500 ng pools.

UCE enrichment and sequencing

Enrichments of library pools were performed using the MYbaits kit (MYcroarray, now Arbor Biosciences) following the manufacturer's protocol v3.01. Hybridizations were performed at 65°C for 24 hours. After hybridization, library pools were bound to Dynabeads M-280 Streptavidin magnetic beads (Life Technologies) for enrichment. Post-hybridization enrichments were amplified in a 25 µL volume reaction consisting of 10 µL enriched DNA, 12.5 µL 2X KAPA HiFi HotStart ReadyMix, and 2.5 µL each of 5 µM of Illumina P5/P7 primers. Amplification conditions were 98°C for 45 s, followed by 16 cycles of 98°C for 20 s, 60°C for 30 s, and 72°C for 60 s, and then a final extension of 72°C for five minutes. Enriched and amplified library pools were quantified on Qubit and pooled in equimolar ratios. Libraries were sequenced using paired-end 150 bp reads on an Illumina HiSeq 3000 (Oklahoma Medical Research Foundation).

Data processing and analysis

For clarification, we make use of the following terms to distinguish between the different data sets analyzed as part of this study: 1) empirical – newly generated UCE data for this study, 2) *in silico* – UCE data retrieved from Faircloth (2017), 3) transcriptome – UCE data retrieved from publicly available transcriptomes and protein-encoding portions of genomes.

Raw sequencing data were processed using PHYLUCE v1.5.0 (B. C. Faircloth, 2016) with associated software as incorporated in the pipeline. We used default values unless otherwise noted. Adaptors and low-quality bases were removed using Illumiprocessor (https://github.com/faircloth-lab/illumiprocessor). Reads were assembled using Trinity v2.0.6 (Grabherr et al., 2011). We aligned UCE loci with MAFFT (Katoh & Standley, 2013), changing the max divergence from 20% (for empirical + *in silico* dataset) to 40% (empirical dataset), and trimmed with GBLOCKS (Castresana, 2000; Talavera & Castresana, 2007). Data matrices that were 50% and 60% complete (i.e., single locus alignments contain at least this percentage of total taxa), were used for further maximum likelihood (ML) phylogenetic analysis using RAxML v8.1.20 (Stamatakis, 2014). To complement our taxon sampling and assess the capacity to integrate our data with existing genomic data, we reassembled transcriptomic data from nine paraneopteran taxa from Misof *et al*. (2014) (https://doi.org/10.5061/dryad.3c0f1) using Trinity v2.0.6 (Grabherr et al., 2011). We also downloaded the protein-coding sequences of the genome of *Bemisia tabaci* from GenBank (Xie et al., 2017). Next, we used a custom pipeline (https://github.com/AlexKnyshov/main_repo) that uses tblastx to search for homologous loci in transcriptomes. We extracted the best matching amino-acid coding portion of sequences from transcriptomes that matched UCE loci with an e-value of 1e-10 or less, which allowed for inclusion of even short matching sequences in amino acid space. After excluding *Xenophysella greensladeae*, the taxon with the fewest UCE loci recovered and the worst assembly, we realigned the data using the MAFFT E-INS-i algorithm and trimmed the alignments from the end to include at least 80% taxon representation. We used the best of 20 ML trees, followed by 100 bootstrap replicates, using the

GTRGAMMA model with genes partitioned by locus. Furthermore, for the four genomes with annotated coding sequences on GenBank that were used in bait design, we used blastn to assess whether the UCE loci corresponded with coding regions. We used an e-value cutoff of 1e-30 (equivalent to an exact match of a string of ~75 base pairs). We also conducted a cross-species check for the pair of the most closely related genome and transcriptome included in our analysis, using tblastx and an e-value cutoff of 1e-10 (as used in our analysis to find corresponding loci in transcriptomes). Analyses were conducted on the University of Georgia and the University of Connecticut high-performance computer clusters.

Results

UCE recovery

Summary results for newly generated UCE data are presented in Table 5.1. We produced 2,114,434 raw paired-end reads per sample on average, with an average of 1,955,078 (91.49%) passing filter. Assemblies resulted in an average of 3,641 contigs per sample. We recovered a total of 2,721 UCE loci across all taxa. Loci per sample ranged between 265 and 1,696 (average=904) for hemipteran taxa, with 117 loci from the thysanopteran outgroup for which new data was gathered. From the 50% and 60% complete data matrices, we recovered 532 and 220 UCE loci from the empirical data set, respectively, while the inclusion of *in silico* data increased recovery to 744 and 325 UCE loci, respectively. We found an average of 34.44% (33.13% for Hemiptera) missing data between UCE alignments, with a range of 10.03% to 62.46% within Hemiptera and 81.45% for the outgroup. We recovered more UCE loci than average for two dried

specimens used in this study (Table 5.1, Appendix N). Summary results of empirically generated UCE data processed with *in silico* data are presented in Appendix O. Summary numbers for parsimony and invariant sites for each data matrix are reported in Appendix P.

Within Heteroptera, UCE loci numbers varied between the infraorders, e.g., from a low of 268 loci (Dipsocoromorpha) to a high of 1,042 loci (Cimicomorpha), corresponding to the amount of missing data within each group (Appendix Q). Overall, the number of loci recovered was less than expected when compared to the Faircloth (2017) *in silico* study. We found the average number of loci, compared to the *in silico* study, to be 73.33% within the same genus, 68.02% within the same family, 52.74% within the same suborder, 51.28% within the order Hemiptera, and 13.54% within the Thysanoptera (Appendix R).

UCEs from transcriptome data

The results of extracting UCE loci from 10 transcriptomes are shown in Appendix S. We recovered the most UCE loci from the coding sequences of the *Bemisia tabaci* genome, most likely due to its completeness compared to transcriptomes based only on cDNA sequencing (88%, or 287 of the 325 loci, used in recovering the tree). We excluded *Xenophysella greensladeae* due to the low N50 of the assembly and the few UCE loci we were able to recover. On average, we recovered 71.5% of the 325 loci used in reconstructing the phylogeny across the remaining eight transcriptomes. For the four annotated genomes, we found that an average of 96.5% UCE loci of the ~1,500-2,300 per taxon contained a match to an annotated protein-coding sequence (Table 5.2). With a

cross-species check of a closely related genome and transcriptome pair, we found that 70.5% of the 2,266 UCE loci designed for *Gerris buenoi* could be found in the transcriptome of *Velia caprai* (Table 5.2).

Phylogenetic trees and taxa relationships

The phylogenetic tree for the full set of empirical, *in silico*, and transcriptome data is shown in Figure 5.1. Bootstrap values were 100% for all but two (*Glycaspis* + *Pachypsylla* and *Brochymena* + *Halyomorpha*) of the shallow evolutionary relationships (Figure 5.1) with a trend of decreasing support values with increased evolutionary depth. Trees for each data set showed mostly consistent topologies with similar support, trending toward more support at deeper phylogenetic nodes when additional taxa are added (i.e., addition of *in silico* and transcriptome data). The inclusion of this additional data did help to recover the well-supported and uncontentious relationship of Sternorrhyncha as sister to Auchenorrhyncha + Heteroptera which was not recovered otherwise in analyzing the newly acquired data in combination with *in silico* data or by itself (Figures 5.2 and 5.3).

Our analysis recovered a monophyletic Hemiptera, with Sternorrhyncha highly supported as the sister group to Auchenorrhyncha + Heteroptera. Support for Auchenorrhyncha + Heteroptera was weak (68%). Inter-infraorder support within Heteroptera ranged from 77–100%. We recovered with high support a monophyletic Geoheteroptera (Leptopodomorpha + (Cimicomorpha + Pentatomomorpha)), the land bugs, which are sister to a clade comprising the four remaining heteropteran infraorders. A strongly supported clade comprising Enicocephalomorpha, Dipsocoromorpha, and

Gerromorpha) (GED clade; 100%) was recovered and moderately supported (77%) as the sister group of Nepomorpha. No topological differences were observed between phylogenetic trees produced using 50% versus 60% data matrices.

The results of intra-familiar level sampling of Reduviidae that focused on the subfamily Triatominae strongly supported (both 100%) a *Rhodnius* + *Psammolestes* clade as sister to *Panstrongylus* + (*Dipetalogaster* + *Triatoma*). For Coreidae, our analysis, which includes two subfamilies and six tribes, recovered all relationships with 100% support. Both species of *Acanthocephala* (Acanthocephalini) were recovered as sister to one another. The genera *Mygdonia* and *Anoplocnemis* were also recovered as sister taxa, supporting a monophyletic Mictini, which is sister to *Anisoscelis* + (*Stenoeurilla* + *Acanthocephala*). The only sampled representative of the subfamily Meropachyinae, *Lycambes sargi*, was nested within the coreine tribe Nematopodini, which together formed a clade sister to all other sampled coreids.

Discussion

Study Rationale

Ultraconserved elements have been widely used for phylogenetic research among vertebrate groups during the past several years, with arthropod UCEs being developed comparatively more recently. Research in the area of arthropod UCEs is still largely open and untested for the vast majority of taxonomic groups. While several UCE bait sets have been empirically evaluated (Baca et al., 2017; B. C. Faircloth et al., 2015; Starrett et al., 2017; Van Dam et al., 2017), the designs for Diptera, Hemiptera, and Lepidoptera (Brant C. Faircloth & Gilbert, 2017) have yet to be similarly evaluated. We demonstrate the

utility of the Hemiptera UCE bait set (Brant C. Faircloth & Gilbert, 2017) across divergent hemipteran taxa.

Recovery of UCE loci

We obtained UCE loci from all taxa sequenced and the number of loci recovered for each sample was correlated with sequencing read depth, consistent with previous UCE studies. We were also able to recover a large number of loci from 25+ year old museum specimens, which will facilitate studies of heteropteran taxa in the future. The taxa for which we recovered the most UCE loci were frequently those with the closest relationship to the species used for bait design, e.g., *Oncopeltus* sp. and *Rhodnius robustus* are congeneric with *Oncopeltus fasciatus* and *Rhodnius prolixus*, respectively, and were the two taxa with the most UCE loci recovered. The amount of missing UCE data for taxa sampled is correlated (Pearson $r = 0.477$, $p = 0.003$) with greater phylogenetic divergence from taxa used to design baits (Appendix 1), which along with some nodes at deeper evolutionary depths having lower support, highlights the importance of including phylogenetically diverse taxa when developing baits for lineages as old as Hemiptera (300 mya; (Misof et al., 2014). Despite the missing data, however, most nodes had 100% support, and all but four exceeded 70% (Figure 5.1).

On average, we obtained about two-thirds the number of loci we expected when compared to the *in silico* study (Brant C. Faircloth & Gilbert, 2017) and 7.4 times less sequencing data than the generated *in silico* data. The amount of UCE loci recovered positively trended (Pearson $r = 0.328$, $p = 0.054$) with the amount of coverage. With increased sequencing depth for certain samples, it is likely we would obtain more unique

UCE loci, and the gap between *in silico* expected and empirically obtained would diminish, though even at this level of coverage the data can resolve most relationships. Changes to the assembly methods and enrichment stringency may also increase the number of loci collected across Hemiptera.

We recovered matching sequences for about 90% of the 325 UCE loci from the empirical + *in silico* dataset for which we conducted a search on amino acid coding sequences of the *Bemisia tabaci* genome. We also found an average of 96.5% UCE loci matching annotated protein-coding sequences with the corresponding percentages in two of the four examined taxa as high as 99.6%. We further investigated the only nine loci with no matches in the annotations of the aphid genome, which was the genome with the fewest number of loci without matches to clarify the nature of these UCE loci. We found that seven of these nine loci matched proteins annotated in other aphid species and two loci corresponded with spliceosomal RNAs (U11 and U12). Thus for the pea aphid at least 2,057 of 2,059 UCE loci (99.9%) contain a protein-encoding core. We suspect that lower percentages found in some transcriptome and genome assemblies can be attributed to incomplete annotations, assemblies, or limited sequencing efforts. This reflects a fundamental difference of UCE loci in vertebrates where UCEs are primarily noncoding yet conserved elements versus invertebrates, where they are primarily protein-coding, as is being increasingly recognized (Silas Bossert & Danforth, 2018).

Systematics of Hemiptera

Recent published studies have proposed several alternative hypotheses for relationships within Hemiptera, and particularly within the Heteroptera (Wang et al.,

2017; Weirauch et al., 2018). Using a large molecular dataset of loci not previously employed that samples broadly throughout the genome, our analyses test these relationships and corroborate some. For example, our phylogenetic hypothesis is congruent with many recently published topologies (Misof et al., 2014; Wang et al., 2017; Weirauch et al., 2018) in supporting a monophyletic Auchenorrhyncha as sister to the Heteroptera (Coleorrhyncha was not included in this analysis, so their position was not evaluated). Consistent with most analyses in recent decades (reviewed in Weirauch *et al.* 2018), relationships within the more densely sampled Heteroptera strongly support the monophyly of Geoheteroptera, the land bugs, which include the great majority of the extant species diversity in this suborder. Also congruent with some recently published phylogenetic hypotheses is the well-supported clade formed by the Gerromorpha, Enicocephalomorpha, and Dipsocoromorpha [GED clade; (Wang et al., 2017; Weirauch et al., 2018)], although relationships within this clade differ between analyses with either Dipsocoromorpha (this study; Wang *et al.* 2017) or Gerromorpha (Weirauch *et al.* 2018) being recovered as sister group to the two remaining infraorders. However, one of the most controversial issues that has significant impact on our understanding of character evolution within the Heteroptera remains unresolved: in contrast to recent phylogenies that either supported the GED clade (Wang et al., 2017) or the aquatic Nepomorpha (Weirauch et al., 2018) as the sister group to all remaining Heteroptera, the current analysis modestly (77%) supported Nepomorpha as sister to the GED clade, putting forward a third alternative hypothesis. Taxon sampling within the Geoheteroptera in our analysis is limited, but relationships conform with currently accepted hypotheses (e.g., Aradidae as sister to Trichophora, Pentatomoidea sister to all other Trichophora,

Miroidea and Cimicoidea+Naboidea clade are sister taxa within Cimicomorpha [Weirauch *et al.* 2018]).

Intra-familial relationships within the reduviid subfamily Triatominae are consistent with previous phylogenetic analyses of the group based on fewer loci (Georgieva, Gordon, & Weirauch, 2017; Justi, Russo, Mallet, Obara, & Galvao, 2014). For the Coreidae, relationships among and within the four subfamilies and 37 tribes have remained unresolved across morphological and single-gene phylogenetic studies (X. Li, 1997; Pan, Guan, & Su, 2007; Souza, Marchesin, & Itoyama, 2016). We expected and recovered a sister group relationship between the two sampled species of *Acanthocephala* (Acanthocephalini). Phylogenetic analyses during the past couple decades have also supported the monophyly of the Micitini, albeit with different taxon sampling compared to our study that included *Anoplocnemis* and *Mygdonia* (Li, 1997; Pan et al., 2007). Our analysis recovered a paraphyletic Coreinae with respect to *Lycambes* (Meropachyinae), a result that has been supported in some previous analyses (X. Li, 1996, 1997). However, in these previous studies, Meropachyinae was supported as the sister group to Chariesterini (not included in our analysis), whereas our study finds the subfamily to be nested within the Nematopodini (*Thasus* and *Mozena*).

Furthermore, our results show congruent topologies across data sets, indicating the usefulness of UCEs even with a relatively small number of samples. However, increasing the taxonomic representation with the addition of *in silico* and transcriptome data improved support for many deep phylogenetic nodes, which may improve with further additions. More importantly, the ability to sample UCE loci from other genomic

resources expands possibilities of taxonomic representation and improves the utility of UCEs for phylogenomic studies.

Conclusion

Our study adds to the accumulating evidence that custom UCE bait sets can resolve most phylogenetic nodes with high bootstrap support, including baits designed for invertebrate groups. We also have shown the capability of integrating our invertebrate UCE loci with protein-coding data from transcriptomes. As phylogenomic datasets become more common and varied in structure, the capability of combining large genetic datasets with others from different sources will become more important to generate a complete tree of life, which represents another strength of this approach. Because the number of loci recovered empirically is significantly lower than expected, we recommend researchers explore various options to improve loci recovery as needed based on study objectives. For example, comparing different assembly methods, less stringent enrichment conditions, and incorporating additional taxa to improve phylogenetic relationships. Certain taxonomic clades within Hemiptera may also benefit from a designed subset of UCE baits as more genomic resources become available. However, we have shown that established relationships within Hemiptera can be recovered with relatively few loci, which provide broad application of the current bait set to researchers.

Acknowledgements

Work was supported by the National Science Foundation [DEB-1136626 to B.C.F. and T.C.G. and IOS-1553100 awarded to Christine W. Miller]. We thank Arbor

Biosciences for providing complimentary MYbaits kit for evaluations. We thank Alexander Knyshov for use of his scripts. Authors declare no conflicts of interest.

Data Accessibility

Raw sequencing reads are deposited in Sequence Read Archive (https://www.ncbi.nlm.nih.gov/sra/SRP161492). All configuration files and scripts, milestone data used for analyses, and a workflow are available on Dryad Digital Repository (https://doi.org/10.5061/dryad.425vg4m). Scripts used for transcriptome analyses are on the github repository of Alexander Knyshov (https://github.com/AlexKnyshov/main_repo).

References

Alexander, A. M., Su, Y. C., Oliveros, C. H., Olson, K. V., Travers, S. L., & Brown, R. M. (2017). Genomic data reveals potential for hybridization, introgression, and incomplete lineage sorting to confound phylogenetic relationships in an adaptive radiation of narrow-mouth frogs. *Evolution, 71*(2), 475-488. doi:10.1111/evo.13133

Baca, S. M., Alexander, A., Gustafson, G. T., & Short, A. E. Z. (2017). Ultraconserved elements show utility in phylogenetic inference of Adephaga (Coleoptera) and suggest paraphyly of 'Hydradephaga'. *SYSTEMATIC ENTOMOLOGY, 42*(4), 786-795. doi:10.1111/syen.12244

Bejerano, G., Pheasant, M., Makunin, I., Stephen, S., Kent, W. j., Mattick, J. S., & Haussler, D. (2004). Ultraconserved Elements in the Human Genome. *SCIENCE, 304*, 1321-1325.

Blaimer, B. B., Brady, S. G., Schultz, T. R., Lloyd, M. W., Fisher, B. L., & Ward, P. S. (2015). Phylogenomic methods outperform traditional multi-locus approaches in resolving deep evolutionary history: a case study of formicine ants. *BMC Evol Biol, 15*, 271. doi:10.1186/s12862-015-0552-5

Blaimer, B. B., Lloyd, M. W., Guillory, W. X., & Brady, S. G. (2016). Sequence Capture and Phylogenetic Utility of Genomic Ultraconserved Elements Obtained from Pinned Insect Specimens. *PLoS One, 11*(8), e0161531. doi:10.1371/journal.pone.0161531

Bossert, S., & Danforth, B. N. (2018). On the universality of target-enrichment baits for phylogenomic research. *Methods in Ecology and Evolution*. doi:10.1111/2041-210x.12988

Bossert, S., Murray, E. A., Blaimer, B. B., & Danforth, B. N. (2017). The impact of GC bias on phylogenetic accuracy using targeted enrichment phylogenomic data. *Mol Phylogenet Evol, 111*, 149-157. doi:10.1016/j.ympev.2017.03.022

Branstetter, M. G., Jesovnik, A., Sosa-Calvo, J., Lloyd, M. W., Faircloth, B. C., Brady, S. G., & Schultz, T. R. (2017). Dry habitats were crucibles of domestication in the evolution of agriculture in ants. *Proc Biol Sci, 284*(1852). doi:10.1098/rspb.2017.0095

Branstetter, M. G., Longino, J. T., Ward, P. S., Faircloth, B. C., & Price, S. (2017). Enriching the ant tree of life: enhanced UCE bait set for genome-scale phylogenetics of ants and other Hymenoptera. *Methods in Ecology and Evolution, 8*(6), 768-776. doi:10.1111/2041-210x.12742

Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol, 17*(4), 540-552.

Crawford, N. G., Faircloth, B. C., McCormack, J. E., Brumfield, R. T., Winker, K., & Glenn, T. C. (2012). More than 1000 ultraconserved elements provide evidence that turtles are the sister group of archosaurs. *Biol Lett, 8*(5), 783-786. doi:10.1098/rsbl.2012.0331

Crawford, N. G., Parham, J. F., Sellas, A. B., Faircloth, B. C., Glenn, T. C., Papenfuss, T. J., . . . Simison, W. B. (2015). A phylogenomic analysis of turtles. *Mol Phylogenet Evol, 83*, 250-257. doi:10.1016/j.ympev.2014.10.021

Cryan, J. R., & Urban, J. M. (2012). Higher-level phylogeny of the insect order Hemiptera: is Auchenorrhyncha really paraphyletic? *SYSTEMATIC ENTOMOLOGY, 37*(1), 7-21. doi:10.1111/j.1365-3113.2011.00611.x

Faircloth, B. C. (2016). PHYLUCE is a software package for the analysis of conserved genomic loci. *Bioinformatics, 32*(5), 786-788. doi:10.1093/bioinformatics/btv646

Faircloth, B. C., Branstetter, M. G., White, N. D., & Brady, S. G. (2015). Target enrichment of ultraconserved elements from arthropods provides a genomic perspective on relationships among Hymenoptera. *Mol Ecol Resour, 15*(3), 489-501. doi:10.1111/1755-0998.12328

Faircloth, B. C., & Gilbert, M. (2017). Identifying conserved genomic elements and designing universal bait sets to enrich them. *Methods in Ecology and Evolution, 8*(9), 1103-1112. doi:10.1111/2041-210x.12754

Faircloth, B. C., McCormack, J. E., Crawford, N. G., Harvey, M. G., Brumfield, R. T., & Glenn, T. C. (2012). Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. *Syst Biol, 61*(5), 717-726. doi:10.1093/sysbio/sys004

Faircloth, B. C., Sorenson, L., Santini, F., & Alfaro, M. E. (2013). A Phylogenomic Perspective on the Radiation of Ray-Finned Fishes Based upon Targeted Sequencing of Ultraconserved Elements (UCEs). *PLoS One, 8*(6), e65923. doi:10.1371/journal.pone.0065923

Georgieva, A. Y., Gordon, E. R. L., & Weirauch, C. (2017). Sylvatic host associations of Triatominae and implications for Chagas disease reservoirs: a review and new

host records based on archival specimens. *PeerJ, 5*, e3826. doi:10.7717/peerj.3826

Gilbert, P. S., Chang, J., Pan, C., Sobel, E. M., Sinsheimer, J. S., Faircloth, B. C., & Alfaro, M. E. (2015). Genome-wide ultraconserved elements exhibit higher phylogenetic informativeness than traditional gene markers in percomorph fishes. *Mol Phylogenet Evol, 92*, 140-146. doi:10.1016/j.ympev.2015.05.027

Glenn, T. C., Nilsen, R., Kieran, T. J., Finger, J. W., Pierson, T. W., Bentley, K. E., . . . Faircloth, B. C. (2016). *Adapterama I: Universal stubs and primers for thousands of dual-indexed Illumina libraries (iTru & iNext)*. BioRxiv.

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., . . . Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol, 29*(7), 644-652. doi:10.1038/nbt.1883

Harvey, M. G., Smith, B. T., Glenn, T. C., Faircloth, B. C., & Brumfield, R. T. (2016). Sequence Capture versus Restriction Site Associated DNA Sequencing for Shallow Systematics. *Syst Biol, 65*(5), 910-924. doi:10.1093/sysbio/syw036

Justi, S. A., Russo, C. A., Mallet, J. R., Obara, M. T., & Galvao, C. (2014). Molecular phylogeny of Triatomini (Hemiptera: Reduviidae: Triatominae). *Parasit Vectors, 7*, 149. doi:10.1186/1756-3305-7-149

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol, 30*(4), 772-780. doi:10.1093/molbev/mst010

Lemmon, A. R., Emme, S. A., & Lemmon, E. M. (2012). Anchored hybrid enrichment for massively high-throughput phylogenomics. *Syst Biol, 61*(5), 727-744. doi:10.1093/sysbio/sys049

Li, H., Shao, R., Song, N., Song, F., Jiang, P., Li, Z., & Cai, W. (2015). Higher-level phylogeny of paraneopteran insects inferred from mitochondrial genome sequences. *Sci Rep, 5*, 8527. doi:10.1038/srep08527

Li, M., Tian, Y., Zhao, Y., & Bu, W. (2012). Higher level phylogeny and the first divergence time estimation of Heteroptera (Insecta: Hemiptera) based on multiple genes. *PLoS One, 7*(2), e32152. doi:10.1371/journal.pone.0032152

Li, X. (1996). Cladistic analysis and higher classification of Coreoidea (Heteroptera). *Entomologia Sinica, 3*, 283-292.

Li, X. (1997). Cladistic analysis of the phylogenetic relationships among the tribal rank taxa of Coreidae (Hemiptera-Heteroptera: Coreoidea). *Acta Zootaxonomica Sinica, 22*, 60-68.

Manthey, J. D., Campillo, L. C., Burns, K. J., & Moyle, R. G. (2016). Comparison of Target-Capture and Restriction-Site Associated DNA Sequencing for Phylogenomics: A Test in Cardinalid Tanagers (Aves, Genus: Piranga). *Syst Biol, 65*(4), 640-650. doi:10.1093/sysbio/syw005

McCormack, J. E., Harvey, M. G., Faircloth, B. C., Crawford, N. G., Glenn, T. C., & Brumfield, R. T. (2013). A phylogeny of birds based on over 1,500 loci collected by target enrichment and high-throughput sequencing. *PLoS One, 8*(1), e54848. doi:10.1371/journal.pone.0054848

Misof, B., Liu, S., Meusemann, K., Peters, R. S., Donath, A., Mayer, C., . . . Zhou, X. (2014). Phylogenomics resolves the timing and pattern of insect evolution. *Science, 346*(6210), 763-767. doi:10.1126/science.1257570

Moyle, R. G., Oliveros, C. H., Andersen, M. J., Hosner, P. A., Benz, B. W., Manthey, J. D., . . . Faircloth, B. C. (2016). Tectonic collision and uplift of Wallacea triggered the global songbird radiation. *Nat Commun, 7*, 12709. doi:10.1038/ncomms12709

Pan, X. L., Guan, J., & Su, F. K. (2007). Discussion on the phylogeny of partial species of Coreinae and Mictinae based on sequences of cytochrome b gene (Hemiptera: Coreidae). *Sichuan Journal of Zoology, 26*, 516-519.

Smith, B. T., Harvey, M. G., Faircloth, B. C., Glenn, T. C., & Brumfield, R. T. (2014). Target capture and massively parallel sequencing of ultraconserved elements for comparative studies at shallow evolutionary time scales. *Syst Biol, 63*(1), 83-95. doi:10.1093/sysbio/syt061

Song, N., Li, H., Cai, W., Yan, F., Wang, J., & Song, F. (2016). Phylogenetic relationships of Hemiptera inferred from mitochondrial and nuclear genes. *Mitochondrial DNA A DNA Mapp Seq Anal, 27*(6), 4380-4389. doi:10.3109/19401736.2015.1089538

Souza, H. V., Marchesin, S. R., & Itoyama, M. M. (2016). Analysis of the mitochondrial COI gene and its informative potential for evolutionary inferences in the families Coreidae and Pentatomidae (Heteroptera). *Genet Mol Res, 15*(1). doi:10.4238/gmr.15017428

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics, 30*(9), 1312-1313. doi:10.1093/bioinformatics/btu033

Starrett, J., Derkarabetian, S., Hedin, M., Bryson, R. W., Jr., McCormack, J. E., & Faircloth, B. C. (2017). High phylogenetic utility of an ultraconserved element probe set designed for Arachnida. *Mol Ecol Resour, 17*(4), 812-823. doi:10.1111/1755-0998.12621

Talavera, G., & Castresana, J. (2007). Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst Biol, 56*(4), 564-577. doi:10.1080/10635150701472164

Van Dam, M. H., Lam, A. W., Sagata, K., Gewa, B., Laufa, R., Balke, M., . . . Riedel, A. (2017). Ultraconserved elements (UCEs) resolve the phylogeny of Australasian smurf-weevils. *PLoS One, 12*(11), e0188044. doi:10.1371/journal.pone.0188044

Wang, Y.-h., Cui, Y., Rédei, D., Baňař, P., Xie, Q., Štys, P., . . . Bu, W.-j. (2016). Phylogenetic divergences of the true bugs (Insecta: Hemiptera: Heteroptera), with emphasis on the aquatic lineages: the last piece of the aquatic insect jigsaw originated in the Late Permian/Early Triassic. *Cladistics, 32*(4), 390-405. doi:10.1111/cla.12137

Wang, Y.-H., Wu, H.-Y., Rédei, D., Xie, Q., Chen, Y., Chen, P.-P., . . . Bu, W.-J. (2017). When did the ancestor of true bugs become stinky? Disentangling the phylogenomics of Hemiptera-Heteroptera. *Cladistics*, 1-25. doi:10.1111/cla.12232

Weirauch, C., Schuh, R. T., Cassis, G., & Wheeler, W. C. (2018). Revisiting habitat and lifestyle transitions in Heteroptera (Insecta: Hemiptera): insights from a combined morphological and molecular phylogeny. *Cladistics*. doi:10.1111/cla.12233

Wheeler, W. C., Schuh, R. T., & Bang, R. (1993). Cladistic relationships among higher groups of Heteroptera congruence between morphological and molecular data sets. *Entomologica scandinavica, 24*(2), 121-137.

Xie, W., Chen, C., Yang, Z., Guo, L., Yang, X., Wang, D., . . . Zhang, Y. (2017). Genome sequencing of the sweetpotato whitefly Bemisia tabaci MED/Q. *Gigascience, 6*(5), 1-7. doi:10.1093/gigascience/gix018

Tables

Table 5.1. Summary results of each sample in the empirical data set. Suborder names are abbreviated to the first three letters. Infraorder names are abbreviated by removing -morpha.

Suborder	Infraorder	Family	Genus	Species	Reads Passed	QC	Contigs	UCE Loci	On-Target	Missing Data
Auc.	Cicado.	Cicadellidae	*Stephanolla*	*rufoapicata*	930,754	91.23%	4,209	1,059	25.16%	41.22%
Het.	Cimico.	Anthocoridae	*Xylastocoris*	*sp.*	800,230	90.14%	2,594	697	26.87%	30.46%
Het.	Cimico.	Cimicidae	*Cimex*	*adjunctus*	6,127,852	96.72%	5,587	1,216	21.76%	31.42%
Het.	Cimico.	Miridae	*nr. Sophianus*	*sp.*	3,164,378	92.44%	2,643	305	11.54%	61.73%
Het.	Cimico.	Nabidae	*Alloeorhynchus*	*sp.*	2,590,442	89.28%	2,278	570	25.02%	34.36%
Het.	Cimico.	Reduviid	*Dipetalog*	*maximu*	2,04	96.	4,3	913	21.1	47.59

		ae	*aster*	*s*	0,34	88	17		5%	%
					4	%				
					1,17	89.				
		Reduviid	*Panstron*	*genicula*	7,56	61	3,0	1,17	38.9	
Het.	Cimico.	ae	*gylus*	*tus*	6	%	26	7	0%	24.6%
					8,48	97.				
		Reduviid	*Psammol*		4,29	83	4,1	1,58	38.1	
Het.	Cimico.	ae	*estes*	*arthuri*	6	%	64	8	4%	21.5%
					3,19	89.				
		Reduviid			5,09	93	3,7	1,50	39.7	13.69
Het.	Cimico.	ae	*Rhodnius*	*robustus*	2	%	93	8	6%	%
					2,97	92.				
		Reduviid		*dimidiat*	1,87	37	4,7	1,40	29.3	22.83
Het.	Cimico.	ae	*Triatoma*	*a*	4	%	68	1	8%	%
						91.				
	Dipsocor	Ceratoco	*Trichoton*		920,	64	1,3		19.0	58.74
Het.	o.	mbidae	*annus*	*sp.*	424	%	89	265	8%	%
						90.				
	Dipsocor	Schizopt	*Hoplonan*		657,	31	1,0		24.9	62.46
Het.	o.	eridae	*nus*	*sp.*	912	%	85	271	8%	%
					1,09	90.				
	Enicocep	Enicocep	*Oncyloco*		9,20	89	1,2		28.1	50.32
Het.	halo.	halidae	*tis*	*sp.*	2	%	35	347	0%	%
Het.	Gerro.	Gerridae	*Gerris*	*sp.*	1,72	91.	4,0	1,29	31.6	34.35

					1,476	89 %	80	0	2%	%
					2,034,830	87.07 %	2,4 53	481	19.61%	45.38 %
	Leptopodo.	Leptopodidae	*Valleriola*	*sp.*	1,956,488	90.98 %	2,3 26	601	25.84%	40.85 %
	Nepo.	Belostomatidae	*Abedus*	*indentatus*	1,742,898	88.58 %	1,9 08	577	30.24%	38.01 %
	Nepo.	Corixidae	*Micronectus*	*sp.*	2,470,014	91.99 %	2,2 70	502	22.11%	45.21 %
	Pentatomo.	Aradidae	*Mezira*	*sp.*	1,199,180	93.37 %	1,3 34	381	28.56%	57.49 %
	Pentatomo.	Coreidae	*Anisoscelis*	*flavolineatus*	1,344,146	89.15 %	5,3 81	698	12.97%	34.55 %
	Pentatomo.	Coreidae	*Anoplocnemis*	*sp.*	1,433,772	92.32 %	6,7 01	1,035	15.45%	15.16 %
	Pentato	Coreidae	*Mozena*	*nr.*	1,526,488	88.	6,0	967	16.1	22.92

	mo.			*lineolata*	8,344	13 %	01		1%	%
					2,764	92.				
	Pentato		*Acanthoc*		4,968	07	8,9	1,21	13.6	16.93
Het.	mo.	Coreidae	*ephala*	*thomasi*		%	11	5	3%	%
						91.				
	Pentato		*Acanthoc*	*femorat*	839,	11	4,0		20.1	25.95
Het.	mo.	Coreidae	*ephala*	*a*	558	%	37	814	6%	%
					1,08	91.				
	Pentato		*Lycambe*		1,110	00	4,8		18.1	19.28
Het.	mo.	Coreidae	*s*	*sargi*		%	90	887	4%	%
					1,30	92.				
	Pentato		*Mygdoni*	*tubercul*	1,796	40	5,8	1,04	17.8	10.03
Het.	mo.	Coreidae	*a*	*osa*		%	55	6	7%	%
						91.				
	Pentato		*Stenoeuri*	*nr.*	722,	30	3,1		29.9	
Het.	mo.	Coreidae	*lla*	*prolixa*	064	%	52	944	5%	18.1%
					1,77	90.				
	Pentato			*neocalif*	9,776	48	5,8	1,16	19.8	14.71
Het.	mo.	Coreidae	*Thasus*	*ornicus*		%	62	3	4%	%
					1,76	90.				
	Pentato				9,802	76	2,1		44.3	26.49
Het.	mo.	Cydnidae				%	35	946	1%	%
Het.	Pentato	Lygaeida	*Oncopelt*	*sp.*	1,26	91.	4,3	1,69	39.4	21.45

	mo.	e	*us*		7,356	94 %	01	6	3%	%
					1,768	92.				
	Pentato	Pentato	*Brochym*		4,158	78 %	4,7	1,46	30.5	26.68
Het.	mo.	midae	*ena*	*sp.*			84	0	2%	%
					1,518	94.				
	Pentato	Pentato	*Euschistu*	*latimarg*	1,668	13 %	4,6	1,43	30.8	24.88
Het.	mo.	midae	*s*	*inatus*			65	7	0%	%
					2,149	91.				
	Pentato	Pachygro	*Oedancal*		9,304	70 %	2,2		27.2	33.88
Het.	mo.	nthidae	*a*	*sp.*			23	605	2%	%
	[S.F.]					89.				
	Psylloide	Aphalari		*brimblec*	989,0	77 %	3,4		22.5	35.76
Ste.	a	dae	*Glycaspis*	*ombei*	64		36	776	8%	%
	[S.F.]				3,858	92.				
	Aphidoi	Aphidida			8,732	57 %	3,0	1,24	40.6	27.98
Ste.	dea	e	*Aphis*	*fabae*			53	0	3%	%
	[S.F.]					91.				
	Psylloide		*Heterops*		775,3	82 %	2,9		16.1	55.69
Ste.	a	Psyllidae	*ylla*	*texana*	36		67	479	4%	%
						88.				
		Phlaeoth	*Klamboth*		171,6	50 %			13.1	81.45
Thy.		ripidae	*rips*	*myopori*	72		887	117	9%	%
				Average	**1,95**	**91.**	**3,6**	**883**	**25.3**	**34.44**

		s	**5,07 8**	**49 %**	**41**	**2%**	**%**

Table 5.2. Summary results of UCE loci found in annotated protein-coding sequences of genomes (top), and UCE loci designed for *Gerris buenoi* that matched to the *Velia caprai* transcriptome (bottom).

UCE loci vs. CDS of self				
Species	*# of target UCE loci*	*# of transcripts*	*# of blastn hits with 1e-30 cutoff*	*% match*
Acyrthosiphon pisum	2,059	30,790	2,050	99.56
Cimex lectularius	2,283	26,626	2,273	99.56
Diaphorina citri	1,545	21,652	1,364	88.28
Halyomorpha halys	2,257	27,675	2,233	98.94
UCE loci vs. Velia transcriptome				
Species	*# of target UCE loci*	*# of transcripts*	*# of tblastx hits with 1e-10 cutoff*	*% match*
Gerris buenoi	2,266	46,481	1,599	70.56

Figures

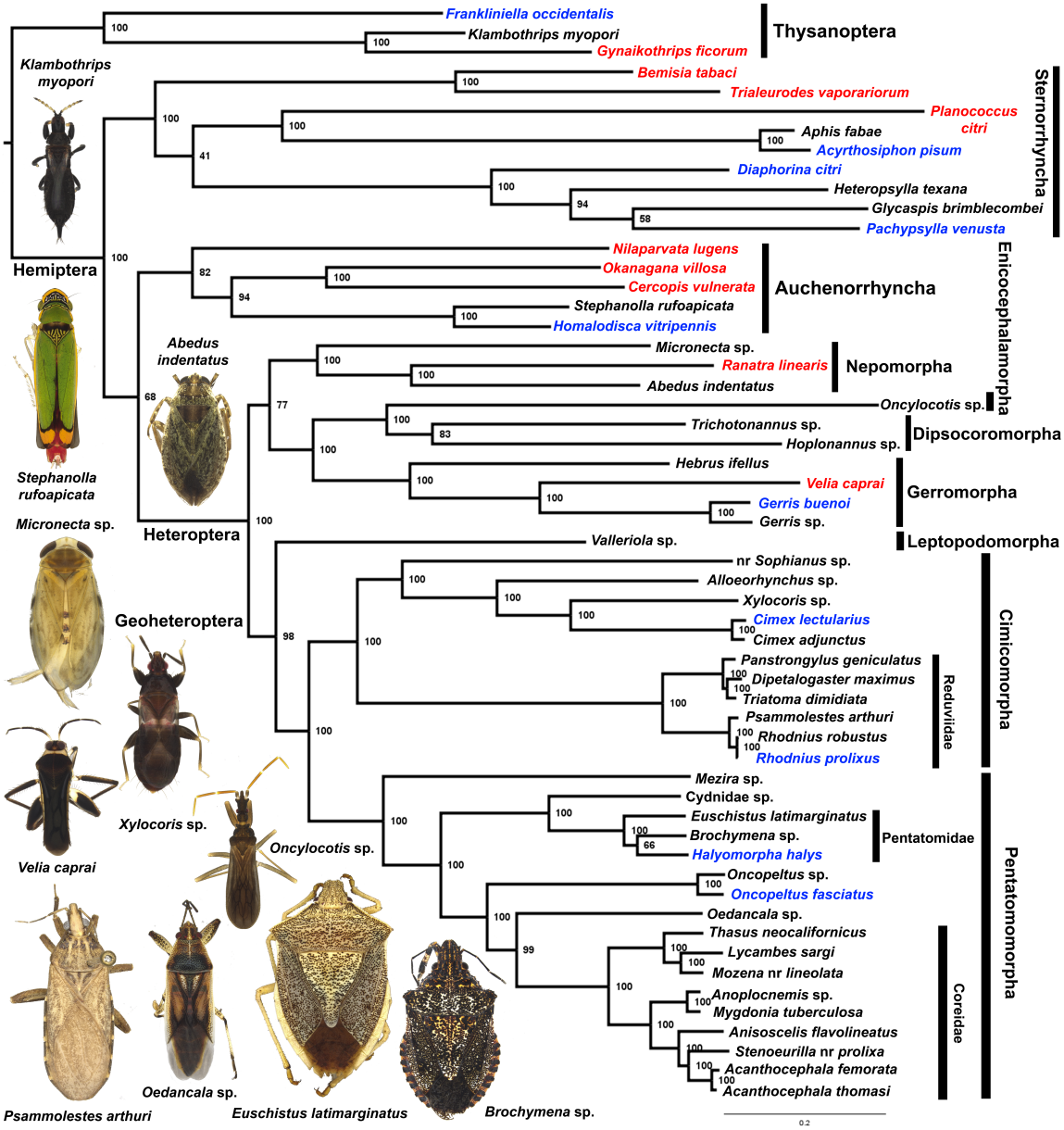


Figure 5.1. Best Maximum Likelihood tree from a search of 20 trees with 100 bootstraps of the 80% data matrix of samples using empirical, *in silico* (blue), and transcriptome UCE data (red).

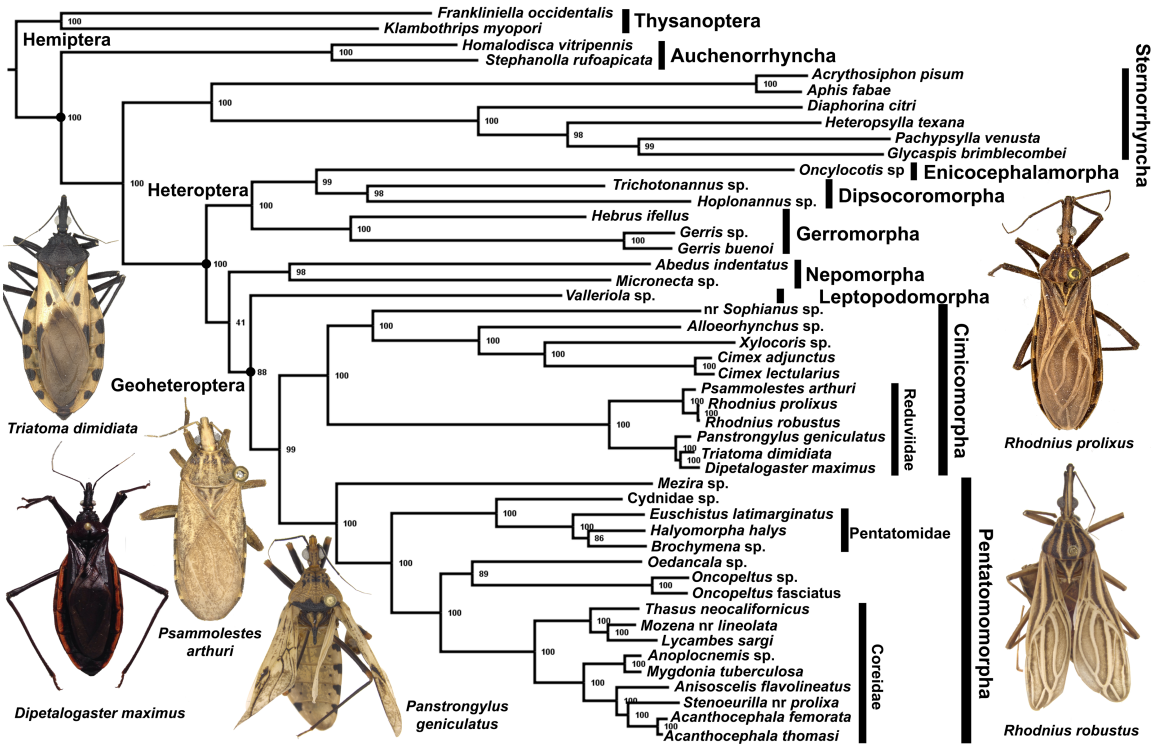


Figure 5.2. Phylogenetic tree of samples with in silico and newly acquired UCE data based on the 60% data matrix.

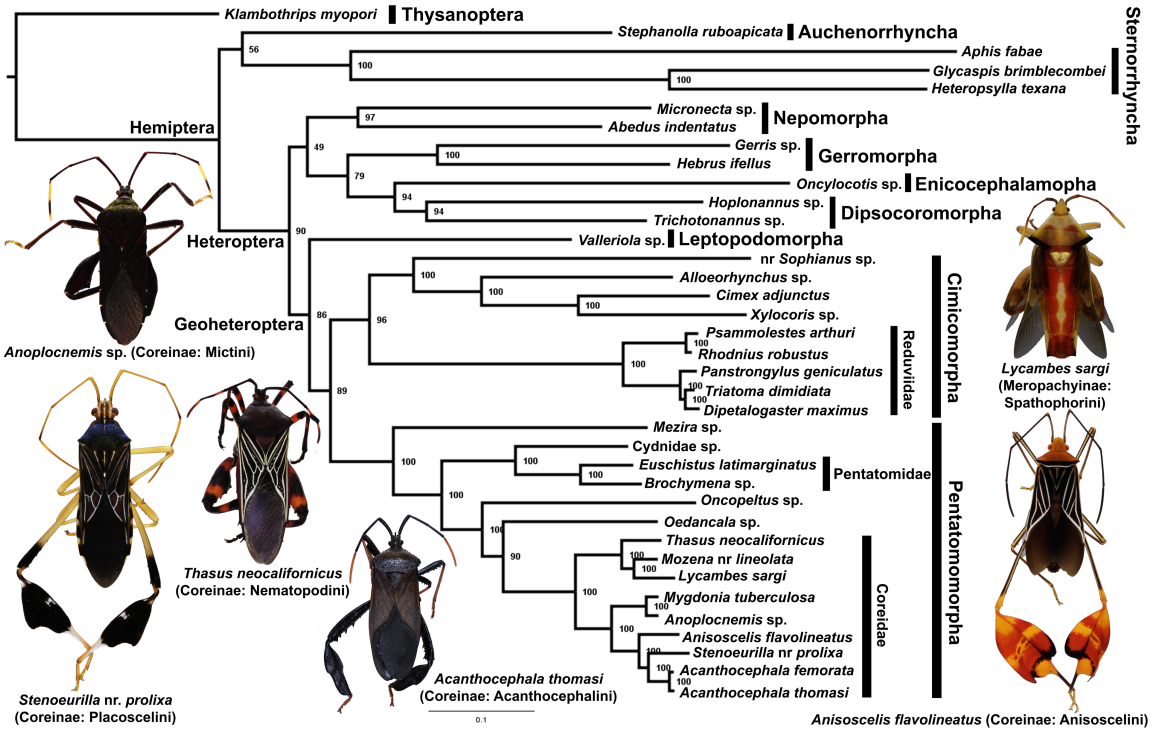


Figure 5.3. Phylogenetic tree of samples with newly acquired UCE data based on the 60% data matrix.

CHAPTER 6

PHYLOGENETICS OF THE SUBFAMILY TRIATOMINAE USING ULTRACONSERVED ELEMENTS[1]

[1] * Kieran TJ. To be Submitted to: *Systematic Entomology*.

Abstract

Triatominae, the kissing bugs, are the biggest radiation of hematophagous species in the Hemiptera with approximately 150 species. Kissing bugs are the sole vectors of the causative agent of Chagas disease (*Trypanosoma cruzi*), a neglected tropical disease that affects millions, primarily in Central and South America. Surprisingly, considering the medical significance of this group, Triatominae's evolutionary origin from predatory assassin bug ancestors is still under discussion and phylogenetic relationships are poorly understood among and within the five tribes of Triatominae. We use ultraconserved elements (UCE) generated from ethanol-preserved and pinned museum specimens to produce the first data-rich and taxonomically densely sampled, well-supported phylogenetic hypothesis for this group of important human disease vectors. This study is the first to include multiple species and/or genera of four of the five currently recognized tribes and is significant in being the first phylogeny to include substantial diversity of Old-World Triatominae. We examine tribal and generic concepts as well as species groups, subgroups, complexes commonly referred to in the epidemiological literature on kissing bugs concluding that based on this dataset: 1) Triatominae are monophyletic and *Opisthacidius* Berg is their predatory sister taxon; 2) (Cavernicolini + *Microtriatoma* [Bolboderini] + Rhodniini) is the sister lineage of (*Belminus* [Bolboderini] + Triatomini); 3) the three large genera (*Rhodnius* Stål, *Triatoma* Laporte, and *Panstrongylus* Berg) are paraphyletic, as previously suggested; 4) Triatomini fall into nine well-supported clades, only two of which are identical in composition to previously recognized groups; 5) the Old World clade is nested within a clade also comprising the *rubida* and *protracta* clades, and is sister to the *protracta* clade. These results highlight the importance of continued

research in this area for further clade wide study of the epidemiolocal importance of this group of vectors.

Introduction

The family Reduviidae (Hemiptera: Heteroptera) is the largest and one of the most diverse clades of predatory insects with ~6,800 described species (Froeschner & Kormilev, 1989; Maldonado, 1990; Weirauch, 2008). The sub-family Triatominae is composed of 152 extant species described in 15 genera and five tribes (Dorn et al., 2018; Justi & Galvao, 2017; Justi, Galvao, & Schrago, 2016; Lima-Cordon et al., 2019; Oliveira, Ayala, Justi, da Rosa, & Galvao, 2018). Triatominae (kissing bugs) is an especially notable group that feed on vertebrate blood and are known vectors of Chagas disease. Chagas is one of the major neglected tropical disease affecting Latin America caused by the protozoan parasite *Trypanosoma cruzi* (Hotez, Bottazzi, Franco-Paredes, Ault, & Periago, 2008). All species of Triatominae are considered potential vectors of *T. cruzi* (Galvao, Carcavallo, Rocha, & Jurberg, 2003). Well characterized knowledge of evolutionary relationships can help inform vector control strategies.

Despite the attention Triatominae have received for their epidemiological and public health importance, phylogentic relationships within the group remain problematic. The issue of whether this group in mono-, poly-, or para-phyletic has not been firmly established. Recent studies have found evidence of both monophyly (Justi et al., 2016; Weirauch & Munro, 2009) and paraphyly (Hwang & Weirauch, 2012). Cryptic species complexes are also quite prevalent in Triatominae adding to the phylogenetic confusion.

The two largest genera, *Rhodnius* and *Triatoma*, have 3 and 8 complexes respectively with *Triatoma* having a further 8 subcomplexes (Schofield & Galvão, 2009).

Evolutionary phylogenetic research in Triatominae has lagged compared to other insect and vector species (Gourbiere, Dorn, Tripet, & Dumonteil, 2012; Justi & Galvao, 2017). Part of the problem is that many evolutionary studies have focused on important "epidemiological types" or a small subset of Triatominae rather that the entire clade (Gourbiere et al., 2012). Previous studies have also been limited by genetic resources and primarily use a small number of mitochondrial (i.e. COI, CytB, 16S, 12S) and/or nuclear (i.e. ITS, 18S, 28S) gene markers for analyses (Justi & Galvao, 2017). Ultraconserved elements (UCEs) can help eliminate both of these restrictions by allowing for the capture of thousands of orthologous loci at and relatively reduced cost for hundreds of taxa. Steps employed in designing UCEs bait sets are to help ensure homology while reducing paralogs (Faircloth, 2017), important considerations for phylogenetic research. A UCE bait set for Hemipteran phylogenetics has recently been designed (Faircloth, 2017) and tested (Kieran et al., 2019, Chapter 5) with promising results. In this study we examine the evolutionary relationships of Triatominae using UCEs in combination with additional genetic markers from Genbank to produce the most comprehensive phylogeny to date.

Methods

Taxonomic sampling

Two datasets were generated, one consisting of the 74 terminals (69 ingroup and 5 outgroups) and for which UCE data were obtained (dataset 1), the other combining UCE data with existing ribosomal data (dataset 2) for a total of 194 terminals (169 ingroup and

25 outgroup taxa). The two datasets include 51 and 105 putative species of Triatominae, respectively. Voucher information for the two datasets is presented as Appendix T. Specimens were identified using a combination of taxonomic keys and authoritatively identified specimens deposited in major natural history museums.

Laboratory methods

We extracted DNA from 70 specimens representing 12 Genera and 53 species that were recently collected or from museum collections (Appendix T). We used either a Qiagen DNeasy kit, Qiagen QIAquick PCR Clean Up kit, or a Phenol Chloroform Isoamyl extraction methods (Appendix T) for all samples. The concentration and quality of DNA was assessed on a Qubit 2.0, fragment analyzer, and a 1.5% agarose gel. Samples with intact, high molecular weight were fragmented on a Bioruptor UCD-300 sonication device (Diagenode). We varied the number of cycles of 30 s on/30 s off from 2-9 based on the DNA quality. Fragmented DNA was run on a 1.5% agarose gel showing a size range from 200-1000bp.

Libraries were prepped following Kieran et al (2019, Chapter 5). Briefly, we used a KAPA Hyper Prep Kit (Kapa Biosystems) with Universal TruSeq compatible adaptor stubs and Illumina TruSeq compatible dual-indexed primers with modified 8 bp indexes (Glenn et al., 2019) to construct the libraries. Post-PCR product was cleaned using Sera-Mag magnetic beads (Thermo-Scientific, Waltham, MA, USA) and quantified with Qubit 2.0. Samples were combined in equimolar amounts based on library size, quality, and taxonomic relatedness, for 11-17 samples per 500ng pool. Pools of libraries were then enriched using previously designed (Faircloth, 2017) and tested (Kieran et al., 2019,

Chapter 5) Hemiptera v1 UCE baits. We used a myBaits kit (Arbor Biosciences) following the manufacturers protocol. Enriched libraries were sequenced on an Illumina HiSeq 3000 using paired-end 150 bp reads (Oklahoma Medical Research Foundation).

Bioinformatic analysis

Sequenced data were processed using PHYLUCE v1.6.1 (Faircloth, 2016) incorporated software. Adaptors and low-quality bases were removed using Illumiprocessor (https://github.com/faircloth-lab/illumiprocessor). Reads were assembled using Trinity v1 r20140717 (Grabherr et al., 2011). We aligned UCE loci using MAFFT (Katoh & Standley, 2013), changing the max divergence to 40%, and trimmed with GBLOCKS (Castresana, 2000; Talavera & Castresana, 2007). We created 60% and 85% data matrices for downstream analyses. As previously performed (Kieran et al., 2019, Chapter 5) we incorporated UCE loci from transcriptome data for non-triatominae reduviids (Appendix T).

We performed a maximum likelihood (ML) phylogenetic analysis using RAxML v8.1.20 (Stamatakis, 2014) with the 60% and 85% UCE data matrices. We used the best of 20 ML trees, followed by 100 bootstrap replicates, using the GTRGAMMA model on concatenated loci (60%, 85% matrix) and partitioned by locus (85% matrix). Partitioned we determined using PartitionFinder2 (Lanfear, Frandsen, Wright, Senfeld, & Calcott, 2017). The best likelihood tree of the partitioned analysis of the 85% matrix is presented as Figure 6.1 (see Appendices L and M for topologies derived from the unpartitioned 60% and 85%). Using all recovered UCE loci we performed a genetree analysis using ASTRAL-III v5.6.1 (Zhang, Rabiee, Sayyari, & Mirarab, 2018) (Appendix W). To

extend taxonomic coverage for a densely-sampled phylogenetic hypothesis, we added 16S rDNA, 18S rDNA and 28S rDNA from GenBank (Table 6.1) and extracted from our UCE enrichment data to the 85% UCE matrix and performed RAxML partitioned (Figure 6.2) and unpartitioned (Appendix X) analyses following the procedures outlined above. We also performed a Bayesian analysis using MrBayes v 3.2.6 (Ronquist et al., 2012) with 100 bootstraps, included as Appendix Y (UCE-only) and Appendix Q (UCE+ribosomal). The ML tree derived from the partitioned analysis that includes 107 species of Triatominae is at the core of results and discussion below (Figure 6.2).

Results

UCE recovery

We generated an average of 4,242,725 raw paired-end reads per sample with 89.23% passing filter (Appendix T). We recovered a total of 2,544 UCE loci with a range of 273 – 1943 per sample (mean = 1470, median = 1613.5) with an average of 44.65% on-target. We obtained a total of 1539 and 341 UCE loci per 60% and 85% data matrix respectively (Table 6.2). Average amount of missing data was for the 85% matrix was half (11.74%) that of the 60% matrix (22.12%).

Phylogenetic trees and taxa relationships

The UCE-only and UCE+ribosomal 60% and 85% matrices as well as partitioned and unpartitioned analyses using RAxML of the concatenated matrix, ASTRAL gene tree approaches, and MrBayes analyses produced overall consistent topologies and support values (Table 6.3; Figures 1, 2, Appendices L-Q). While the majority of nodes are

supported in all or most analyses, mostly with absolute or near-absolute support values, several conflicting or less-well supported nodes suggest that about a handful of deeper-level relationships within Triatominae will benefit from further testing using expanded datasets.

All but one of the analyses recovered Triatominae as monophyletic with high or absolute branch support values (Table 6.3). *Opisthacidius* was recovered as sister taxon to the monophyletic Triatominae in all analyzes except one with similarly high support. In the single diverging topology (ASTRAL; Appendix W), *Opisthacidius* was inferred to be the sister taxon to the Cavernicolini + *Microtriatoma* + Rhodniini clade, rendering Triatominae paraphyletic.

In all eight analyses, Triatominae are deeply split into two clades that both were recovered with full support, (Cavernicolini (*Microtriatoma* ["Bolboderini"] + Rhodniini)) and (*Belminus* ["Bolboderini"] + Triatomini). Bolboderini that in our analyses were represented by three species in two genera are polyphyletic, while Cavernicolini (100%), Rhodniini (100%), and Triatomini (five analyses with 100%) are monophyletic. Within the (Cavernicolini (*Microtriatoma* ["Bolboderini"] + Rhodniini)) clade, *Psammolestes* renders *Rhodnius* paraphyletic in all analyses and is recovered as sister taxon to the *prolixus* group of *Rhodnius*. The three recognized species groups of Rhodniini were recovered in all analyses, receiving full support in the UCE-only and UCE+ribosomal MrBayes analyses, with support values in the high 90s in the remaining analyses. The *pallescens* and *prolixus* groups were strongly supported as sister taxa in all analyses (five with 100%, remaining 97-99%), with the *pictipes* group as sister lineage to that clade.

Within Triatomini, we consistently recovered nine fully or highly supported clades that at least in part correspond to groups of species also recovered in previously published analyses. Among these are what we here refer to as the *dispar* clade (all 100%) that corresponds to the *dispar* group of Schofield & Galvão (2009). The *infestans* clade in our analyses only partially aligns with the *infestans* group sensu Schofield & Galvão (2009). The majority of *infestans* group species form a moderately well supported clade in two of the UCE+ribosomal analyses (80 and 82%; only two taxa included in the UCE-only analyses, all with 100%), but the *spinolai* complex of the *infestans* group always forms a separate lineage from our *infestans* clade and *Triatoma tibiamaculata* (*brasiliensis* subcomplex of *infestans* group and complex) is recovered as part of what we refer to as *flavida* clade (that also comprises most *Panstrongylus* spp.). Only the *infestans* complex is monophyletic, while all remaining subcomplexes (*brasiliensis*, *rubrovaria*, *sordida*, *matogrossensis*, and *maculata*) are paraphyletic or polyphyletic in our analyses.

Species of the *spinolai* complex, sometimes also treated as a separate genus, *Mepraia*, form a monophyletic group in the UCE+ribosomal analyses (88%-100%; only one species included in the UCE-only analysis). The two species of *Eratyrus* included in all analyses always are fully supported sister taxa. All species of *Panstrongylus* except *Panstrongylus rufotuberculatus* are recovered in a clade with species of the *flavida* group (sometimes also treated as a separate genus, *Nesotriatoma*) in both UCE-only and UCE+ribosomal analyses; in the UCE+ribosomal analyses, this clade in addition includes *T. tibiamaculata* (*infestans* group). We refer to this monophyletic group as the *flavida* clade; it was recovered with full support in all UCE-only analyses with values in the UCE+ribosomal analyses ranging between 76 and 100%. *Panstrongylus rufotuberculatus*

is excluded from this clade in all analyses and always forms the sister lineage to the *phyllosoma*, *protracta*, *rubrofasciata*, and *rubida* clades (100% in UCE-only, 93-100% in UCE+ribosomal analyses).

The remaining four clades are fully supported in almost all analyses (Table 6.3) and correspond to the *rubrofasciata* group sensu Schofield & Galvão (2009). Our *phyllosoma* clade (fully supported in all analyses) combines most species of the *lecticularia* and *phyllosoma* complexes, neither of which is monophyletic in our analyses, but excludes *Triatoma bolivari*, *Triatoma rubida*, and *T. ryckmani*. We refer to these three species together as *rubida* clade (fully supported in seven analyses, one with 99%). The *rubrofasciata* complex sensu Schofield & Galvão (2009) that we here refer to as *rubrofasciata* clade is fully supported in all but two analyses which received values of 97 and 98%; this clade comprises the two (UCE+ribosomal) or one (UCE-only) included species of *Linshcosteus*, the cosmopolitan *Triatoma rubrofasciata*, and the three (UCE+ribosomal) or one (UCE-only) Old World endemic species of *Triatoma* included in our analyses. Support values for the protracta clade are similar (full for six analyses, 98% and 99% for the remaining two); in addition to species of the protracta complex sensu Schofield & Galvão (2009), it also includes *Dipetalogaster maximus*, *Paratriatoma hirsuta*, and two of the species classified in the *lecticularia* complex, *Triatoma indictiva* and *Triatoma lecticularia*.

While these clades are highly supported, relationships between some of them will require additional scrutiny. While the *protracta* + *rubrofasciata* clade received full support in two of the analyses (Table 6.3), others did not recover this clade, or only with low support. The *rubida* + *protracta* + *rubrofasciata* clade is overall much better

supported (between 80% and 100%) but was not recovered in the ASTRAL analysis. Similarly, while the clade consisting of *Eratyrus*, the *flavida* clade, *P. rufotuberculatus* and the *rubrofasciata* group was highly or fully supported in most analyses, it was not inferred in the ASTRAL analysis. Even less well supported is the monophyletic group that includes the above-mentioned clade plus the *spinolai* clade that was found in only five of the eight analyses, all with low to moderate support.

Discussion

Ultraconserved Element methods performed exceedingly well in these specimens as was shown previously with Triatominae and other hemipterans (Kieran et al., 2019, Chapter 5). Consistent topologies across all analyses supports the high quality and accuracy of the data presented here. These results contribute new information to the phylogenetic clades of Triatominae vector species, while confirming other previously observed relationships. One major take away is the finding that Triatominae are monophyletic, with *Opisthacidius spp*. as a well-supported sister taxon consistent with Justi et al. (2016). We observed a close relationship between Rhodniini and Cavernicolini previously proposed (e.g., Hwang & Weirauch, 2012, Justi et al., 2016) and a new finding of a paraphyletic Bolboderini.

All large genera groups are paraphyletic (*Panstrongylus*, *Rhodnius*, and *Triatoma*). *Panstrongylus* is interspersed with several *Triatoma* species (*T. bruneri, T. flavida, T. obscura, T. tibiamaculata*) while *P. rufrotuberculatus* is in a sister clade from all other *Panstrongylus*. Justi et al. 2016 partially observed this with *T. bruneri* and *T. tibiamaculata* but lack the additional taxonomic sampling presented here. *Psammolestes*

spp. render the genera *Rhodnius* paraphyletic, as previously reported (Justi et al., 2016; Monteiro, Wesson, Dotson, Schofield, & Beard, 2000). In the case of *Triatom*a, the genera are made paraphyletic by several taxa (*Dipetalogaster, Eratyrus, Linshcosteus, Mepraia, and Paratriatoma*). This was partially shown by Justi et al. 2016, but that study lacked the diversity of taxa and included only one species of *Linshcosteus* and *P. hirsuta*. With additional taxonomic sampling (*i.e*. *Linshcosteus, Dipetalogaster, and Eratyrus*) this result is made clear and is a significant divergence from all previous studies.

The situation for *Rhodnius* species groups is more mixed. Both the UCE only and the UCE+ribosomal analyses found *pictipes* group + (*pallescens* group + *prolixus* group). This differs from Justi et al. 2016 who found *pallescens* group + (*pictipes* group + *prolixus* group) and Abad-Franch & Monteiro (Abad-Franch & Monteiro, 2007) with *pallescens* and *pictipes* groups as sister taxa, which questions the previous notion of a monophyletic *prolixus* clade. The Rhodniini were hypothesized to have evolved in the Amazon region and radiated eastward (Paula, Diotaiuti, & Galvao, 2007; Schofield & Galvão, 2009). However, in our analysis the Rhodniini are sister clade to the widespread *Microtriatoma* clade, which would suggest that the ancestral range of their common ancestor may have been fairly large.

Within the Triatomini, only two clades were supported by other studies, the pairing of the *protracta* and the *rubrofasciata* subgroups, and the monophyletic *dispar* clade as sister to all other Triatomini (Justi et al., 2016). Our results found the *infestans* group paraphyletic, with *spinolai* subgroup more closely related to remaining Triatomini (very different topology from Just et al. 2016: their *spinolai* clade includes *breyeri* and

eratyrusiformis, but also *protracta* and *barberi*, which in our analyses are where they should be, i.e. part of the *protracta* subgroup).

Justi et al. 2016 found the Old-World clade to be sister to all Central and North American taxa. In contrast, our analyses, shows the Old-World clade nested within a clade also comprising the *rubida* and *protracta* clades, and is sister to the *protracta* clade. This finding opens new avenues questions of how this group dispersed to the Old-World and further biogeographic research in this area should be investigated.

This data set provides a new valuable resource for current and future Triatominae research. Ultraconserved elements proved to be a valuable addition to the phylogenetic toolkit of Triatominae. With a goal towards additional taxonomic sampling in the future, these data can contribute to future studies that will ultimately result in a truly complete phylogeny of the subfamily Triatominae.

Acknowledgments

Many thanks are offered to Christiane Weirauch, Alejandro Zaldívar-Riverón, Carlos Ibarra-Cerdeña, Azael Saldaña, and Nicole Gottdenker for providing the taxonomic specimens. Eric R. L. Gordon and Rochelle Hoey-Chamberlain for laboratory assistance. Arbor Biosciences for providing the UCE baits used in this study.

References

Abad-Franch, F., & Monteiro, F. A. (2007). Biogeography and evolution of Amazonian triatomines (Heteroptera : Reduviidae): implications for Chagas disease surveillance in humid forest ecoregions. *Mem Inst Oswaldo Cruz, 102*, 57-69. doi:Doi 10.1590/S0074-02762007005000108

Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution, 17*(4), 540-552. doi:DOI 10.1093/oxfordjournals.molbev.a026334

Dorn, P. L., Justi, S. A., Dale, C., Stevens, L., Galvao, C., Lima-Cordon, R., & Monroy, C. (2018). Description of Triatoma mopan sp. n. from a cave in Belize (Hemiptera, Reduviidae, Triatominae). *Zookeys*(775), 69-95. doi:10.3897/zookeys.775.22553

Faircloth, B. C. (2016). PHYLUCE is a software package for the analysis of conserved genomic loci. *Bioinformatics, 32*(5), 786-788. doi:10.1093/bioinformatics/btv646

Faircloth, B. C. (2017). Identifying conserved genomic elements and designing universal bait sets to enrich them. *Methods in Ecology and Evolution, 8*(9), 1103-1112. doi:10.1111/2041-210x.12754

Froeschner, R. C., & Kormilev, N. A. (1989). Phymatidae or ambush bugs of the world: a synonymic list with keys to species, except Lophoscutus and Phymata (Hemiptera). *Entomography, 6*, 1-76.

Galvao, C., Carcavallo, R. U., Rocha, D. S., & Jurberg, J. (2003). A Checklist of the current valid species of the subfamily Triatominae Jeannel, 1919 (Hemiptera: Reduviidae) and their geographical distribution, with nomenclatural and taxonomic notes. *Zootaxa, 202*, 1-36.

Glenn, T. C., Nilsen, R. A., Kieran, T. J., Sanders, J. G., Bayona-Vásquez, N. J., Finger, J. W., . . . Faircloth, B. C. (2019). *Adapterama I: Universal stubs and primers for 384 unique dual-indexed or 147,456 combinatorially-indexed Illumina libraries (iTru & iNext)*. bioRxiv. Retrieved from https://www.biorxiv.org/content/10.1101/049114v2

Gourbiere, S., Dorn, P., Tripet, F., & Dumonteil, E. (2012). Genetics and evolution of triatomines: from phylogeny to vector control. *Heredity (Edinb), 108*(3), 190-202. doi:10.1038/hdy.2011.71

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., . . . Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol, 29*(7), 644-652. doi:10.1038/nbt.1883

Hotez, P. J., Bottazzi, M. E., Franco-Paredes, C., Ault, S. K., & Periago, M. R. (2008). The neglected tropical diseases of Latin America and the Caribbean: a review of disease burden and distribution and a roadmap for control and elimination. *PLoS Negl Trop Dis, 2*(9), e300. doi:10.1371/journal.pntd.0000300

Hwang, W. S., & Weirauch, C. (2012). Evolutionary history of assassin bugs (insecta: hemiptera: Reduviidae): insights from divergence dating and ancestral state reconstruction. *PLoS One, 7*(9), e45523. doi:10.1371/journal.pone.0045523

Justi, S. A., & Galvao, C. (2017). The Evolutionary Origin of Diversity in Chagas Disease Vectors. *Trends Parasitol, 33*(1), 42-52. doi:10.1016/j.pt.2016.11.002

Justi, S. A., Galvao, C., & Schrago, C. G. (2016). Geological Changes of the Americas and their Influence on the Diversification of the Neotropical Kissing Bugs (Hemiptera: Reduviidae: Triatominae). *PLoS Negl Trop Dis, 10*(4), e0004527. doi:10.1371/journal.pntd.0004527

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol, 30*(4), 772-780. doi:10.1093/molbev/mst010

Kieran, T. J., Gordon, E. R. L., Forthman, M., Hoey-Chamberlain, R., Kimball, R. T., Faircloth, B. C., . . . Glenn, T. C. (2019). Insight from an ultraconserved element bait set designed for hemipteran phylogenetics integrated with genomic resources. *Mol Phylogenet Evol, 130*, 297-303. doi:10.1016/j.ympev.2018.10.026

Lanfear, R., Frandsen, P. B., Wright, A. M., Senfeld, T., & Calcott, B. (2017). PartitionFinder 2: New Methods for Selecting Partitioned Models of Evolution for Molecular and Morphological Phylogenetic Analyses. *Mol Biol Evol, 34*(3), 772-773. doi:10.1093/molbev/msw260

Lima-Cordon, R. A., Monroy, M. C., Stevens, L., Rodas, A., Rodas, G. A., Dorn, P. L., & Justi, S. A. (2019). Description of Triatomahuehuetenanguensis sp. n., a potential Chagas disease vector (Hemiptera, Reduviidae, Triatominae). *Zookeys*(820), 51-70. doi:10.3897/zookeys.820.27258

Maldonado, C. J. (1990). Systematic Catalogue of the Reduviidae of the World (Insecta: Heteroptera). *Caribbean Journal of Science*, 694.

Monteiro, F. A., Wesson, D. M., Dotson, E. M., Schofield, C. J., & Beard, C. B. (2000). Phylogeny and molecular taxonomy of the Rhodniini derived from mitochondrial and nuclear DNA sequences. *Am J Trop Med Hyg, 62*(4), 460-450.

Oliveira, J., Ayala, J. M., Justi, S. A., da Rosa, J. A., & Galvao, C. (2018). Description of a new species of Nesotriatoma Usinger, 1944 from Cuba and revalidation of synonymy between Nesotriatoma bruneri (Usinger, 1944) and N. flavida (Neiva, 1911) (Hemiptera, Reduviidae, Triatominae). *Journal of Vector Ecology, 43*(1), 148-157.

Paula, A. S., Diotaiuti, L., & Galvao, C. (2007). Systematics and biogeography of Rhodniini (Heteroptera: Reduviidae: Triatominae) based on 16S mitochondrial rDNA sequences. *Journal of Biogeography, 34*(4), 699-712.

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Hohna, S., . . . Huelsenbeck, J. P. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol, 61*(3), 539-542. doi:10.1093/sysbio/sys029

Schofield, C. J., & Galvão, C. (2009). Classification, evolution, and species groups within the Triatominae. *Acta Trop, 110*(2-3), 88-100. doi:10.1016/j.actatropica.2009.01.010

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics, 30*(9), 1312-1313. doi:10.1093/bioinformatics/btu033

Talavera, G., & Castresana, J. (2007). Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst Biol, 56*(4), 564-577. doi:10.1080/10635150701472164

Weirauch, C. (2008). Cladistic analysis of Reduviidae (Heteroptera : Cimicomorpha) based on morphological characters. *SYSTEMATIC ENTOMOLOGY, 33*(2), 229-274. doi:10.1111/j.1365-3113.2007.00417.x

Weirauch, C., & Munro, J. B. (2009). Molecular phylogeny of the assassin bugs (Hemiptera: Reduviidae), based on mitochondrial and nuclear ribosomal genes. *Mol Phylogenet Evol, 53*(1), 287-299. doi:10.1016/j.ympev.2009.05.039

Zhang, C., Rabiee, M., Sayyari, E., & Mirarab, S. (2018). ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics, 19*(Suppl 6), 153. doi:10.1186/s12859-018-2129-y

Tables

Table 6.1. Taxa and Genbank accession numbers for all ribosomal samples used in this study.

Species	16S	18S	28S
Cavernicola_pilosa	JQ897785.1	JQ897550.1	JQ897627.1
Dipetalogaster_maximus	KC248968.1		KC249134.1
Eratyrus_mucronatus	JQ897794.1	AJ421953.1	JQ897635.1
Linshcosteus_sp.	AF394595.1	AJ421954.1	
Mepraia_spinolai	AF324518.1	AJ421961.1	
Panstrongylus_geniculatus	JQ897822.1	JQ897583.1	JQ897655.1
Panstrongylus_herreri	AY185833.1		
Panstrongylus_lignarius	JQ897823.1	JQ897584.1	KX109906.1
Panstrongylus_lutzi	KC248969.1		KC249135.1
Panstrongylus_megistus	KC248972.1	AJ243336.1	KC249136.1 KC249137.1 KC249138.1 KC249139.1 KC249140.1 KC249141.1
Panstrongylus_rufotuberculatus	KY748239.1	AJ421955.1	
Panstrongylus_tupynambai	KC248977.1		KC249142.1
Psammolestes_tertius	AY035439.1	Y18751.1	
Rhodnius_brethesi	KC248980.1		
Rhodnius_colombiensis	AY035438.1		KC543516.1
Rhodnius_domesticus	AY035440.1		

Rhodnius_ecuadoriensis	AF028746.1		KC543517.1
Rhodnius_nasutus	AF028749.1		AF435856.1
Rhodnius_neglectus	JQ897839.1	JQ897601.1	JQ897670.1
Rhodnius_neivai	AY035441.1		
Rhodnius_pictipes	KC248982.1	KC249094.1	JQ897756.1
Rhodnius_prolixus	AF324519.1	AJ421962.1	AF435862.1
Rhodnius_robustus	MF966358.1		AF435858.1
Rhodnius_stali	KC248984.1	AJ243335.1	KY111675.1
Triatoma_arthurneivai	AY035460.1		
Triatoma_baratai	KC571991.1		KC249143.1
Triatoma_barberi	JX872241.1	AJ421958.1	
Triatoma_brasiliensis	KC248985.1	AJ421957.1	GQ853395.1
Triatoma_breyeri	KC248988.1		
Triatoma_bruneri	KC248989.1		KC249146.1
Triatoma_carcavalloi	KC248990.1	KC249097.1	
Triatoma_circummaculata	KC248992.1	KC249098.1 KC249099.1	KC249147.1 KC249148.1
Triatoma_costalimai	KC248997.1	KC249101.1	KC249149.1
Triatoma_delpontei	KC248999.1		KC249151.1
Triatoma_dimidiata	KC249003.1	AJ243328.1	KC249152.1
Triatoma_flavida	AY035451.1	AJ421959.1	
Triatoma_garciabesi	KC249006.1	KC249102.1	KC249158.1
Triatoma_guasayana	KC249009.1	KC249103.1	KC249162.1 KC249163.1
Triatoma_guazu	KC249013.1	KC249105.1	KC249164.1

			KC249166.1
Triatoma_infestans	KC249023.1	Y18750.1	KC249168.1 KC249169.1
			KC249172.1
Triatoma_jatai	KT601154.1		
Triatoma_juazeirensis	KF769453.1		KC249173.1
Triatoma_jurbergi	KC249027.1	KC249110.1	KC249174.1
Triatoma_klugi	KC249028.1		
Triatoma_lecticularia	AY185837.1	KC249111.1	KC249175.1
Triatoma_lenti	KY576788.1		
Triatoma_longipennis	KC249031.1	AJ243331.1	KC249177.1
Triatoma_maculata	KC249034.1		KX109904.1
			GQ853398.1
Triatoma_matogrossensis	KC249036.1	KC249114.1	KC249180.1 KC249181.1
			KC249182.1
Triatoma_mazzottii	AY035446.1	AJ243333.1	AY860392.1
Triatoma_melanica	KC249041.1		KC249183.1
Triatoma_melanocephala	KF769451.1		
Triatoma_melanosoma	KC249042.1		
Triatoma_mexicana	JX872251.1		
Triatoma_nitida	JX872239.1		
Triatoma_pallidipennis	AY167618.1	AJ243330.1	KC249184.1
Triatoma_patagonica	AF324528.1		
Triatoma_petrochii	KY654073.1		

Triatoma_picturata	AY185840.1	AJ243332.1	AY860404.1
Triatoma_pintodiasi	MG264738.1		
Triatoma_platensis	KC249047.1		GQ853400.1 KC249186.1
Triatoma_protracta	KT231827.1	FJ230520.1	KC249187.1 KC249189.1
Triatoma_pseudomaculata	KC249051.1		KC249190.1 KC249192.1
Triatoma_recurva	FJ230417.1	FJ230496.1	FJ230577.1
Triatoma_rubida	AY185842.1		AY860389.1
Triatoma_rubrofasciata	KY420176.1	AJ421960.1	KR632546.1
Triatoma_rubrovaria	KC249065.1	KC249116.1	KC249197.1 KC249204.1
Triatoma_ryckmani	JX872249.1		
Triatoma_sanguisuga	HQ141281.1		GQ853392.1
Triatoma_sherlocki	KC249068.1		KC249205.1 KC249207.1
Triatoma_sordida	KC249071.1	AJ421956.1	KC249209.1 KC249210.1
Triatoma_tibiamaculata	KC249081.1	KC249127.1	KC249215.1
Triatoma_vandae	KC249083.1	KC249129.1	KC249216.1 KC249218.1
Triatoma_venosa	JQ897850.1	JQ897611.1	JQ897681.1
Triatoma_vitticeps	KC249085.1	KC249130.1	KC249220.1 KC249221.1

Triatoma_williami	KC249089.1		
Triatoma_wygodzinskyi	KC249090.1	KC249133.1	KC249222.1

Table 6.2. Summary statistics of data matrices used for phylogenetic analysis, with number (#) and percentages (%). Ribosomal DNA (rl).

	# taxa	# loci	# sites	# (%) informative	# (%) uninformative	# (%) invariant
UCE 60%	74	1539	337,745	105,407 (31.2)	39,811 (11.8)	192,527 (57.0)
UCE 85%	74	341	81,892	25,406 (31.0)	9,851 (12.0)	46,635 (56.9)
UCE 85% + rl	176	344	85,293	25,986 (30.5)	10,049 (11.8)	49,258 (57.8)

Table 6.3. Compilation of support values for each major node across all trees/analyses. * = North America/Central America/ Old-World, cl. = clade, cn = conflicting node (highlighted), na = not applicable.

	RAxML 85 P	RAxML 85 P	RAxML 60 UP	RAxML 85 UP	ASTRAL	RAxML 85 UP	MrBayes 85	MrBayes 85
	Figure 1	Figure 2	App. L	App. M	App. N	App. O	App. P	App. Q
Triatominae + *Opisthacidius*	100	96	100	100	cn	100	100	100
Triatominae	95	98	100	97	cn	100	100	93
Cavernicolini + Microtriatoma + Rhodniini	100	100	100	100	100	100	100	100
Microtriatoma + Rhodniini	100	98	100	100	100	97	100	100
Rhodniini	100	100	100	100	100	100	100	100
(pallescens gr. + prolixus gr.)	100	99	100	98	100	97	100	100
Belminus + Triatomini	100	100	100	100	100	100	100	100
Triatomini	100	96	100	100	100	94	100	50
dispar clade	100	100	100	100	100	100	100	100
Triatomini minus dispar clade	100	96	100	100	100	94	100	50
infestans clade	100	82	100	100	100	80	100	cn
spinolai cl. + Eratyrus + flavida + P. rufotuberculatus *	83	62	65	85	cn	64	cn	cn
spinolai clade	na	91	na	na	na	88	na	100
Eratyrus + flavida + P. rufotuberculatus *	100	95	100	100	cn	92	100	100
Eratyrus	100	100	100	100	100	100	100	100
flavida + P. rufotuberculatus *	97	91	99	96	cn	84	100	100
flavida clade	100	76	100	100	100	76	100	100
(P. rufotuberculatus + *	100	94	100	95	100	93	100	100
(phyllosoma+(rubida+(protacta+rubrofasciata)))	100	98	100	100	91	97	100	100
(rubida cl. + (protacta cl. + rubrofasciata cl.))	97	94	97	80	na	90	100	100
rubida clade	100	99	100	100	100	100	100	100
(protacta cl. + rubrofasciata cl.)	66	66	100	100	40	72	cn	cn
rubrofasciata clade	100	98	100	100	100	97	100	100
protracta clade	100	98	100	100	100	99	100	100
phyllosoma clade	100	100	100	100	100	100	100	100

Figures

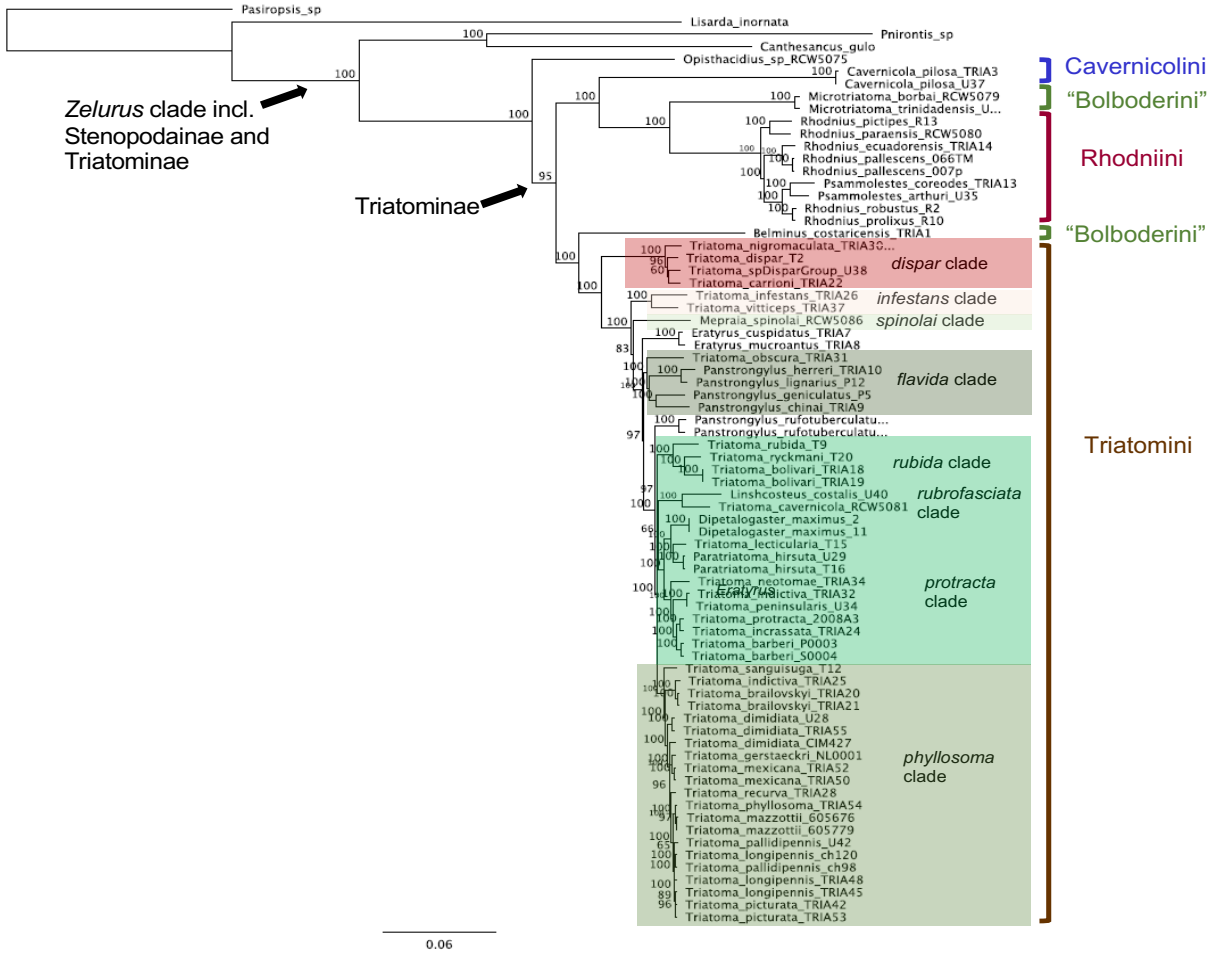


Figure 6.1. Maximum Likelihood tree from the partitioned 85% UCE dataset.

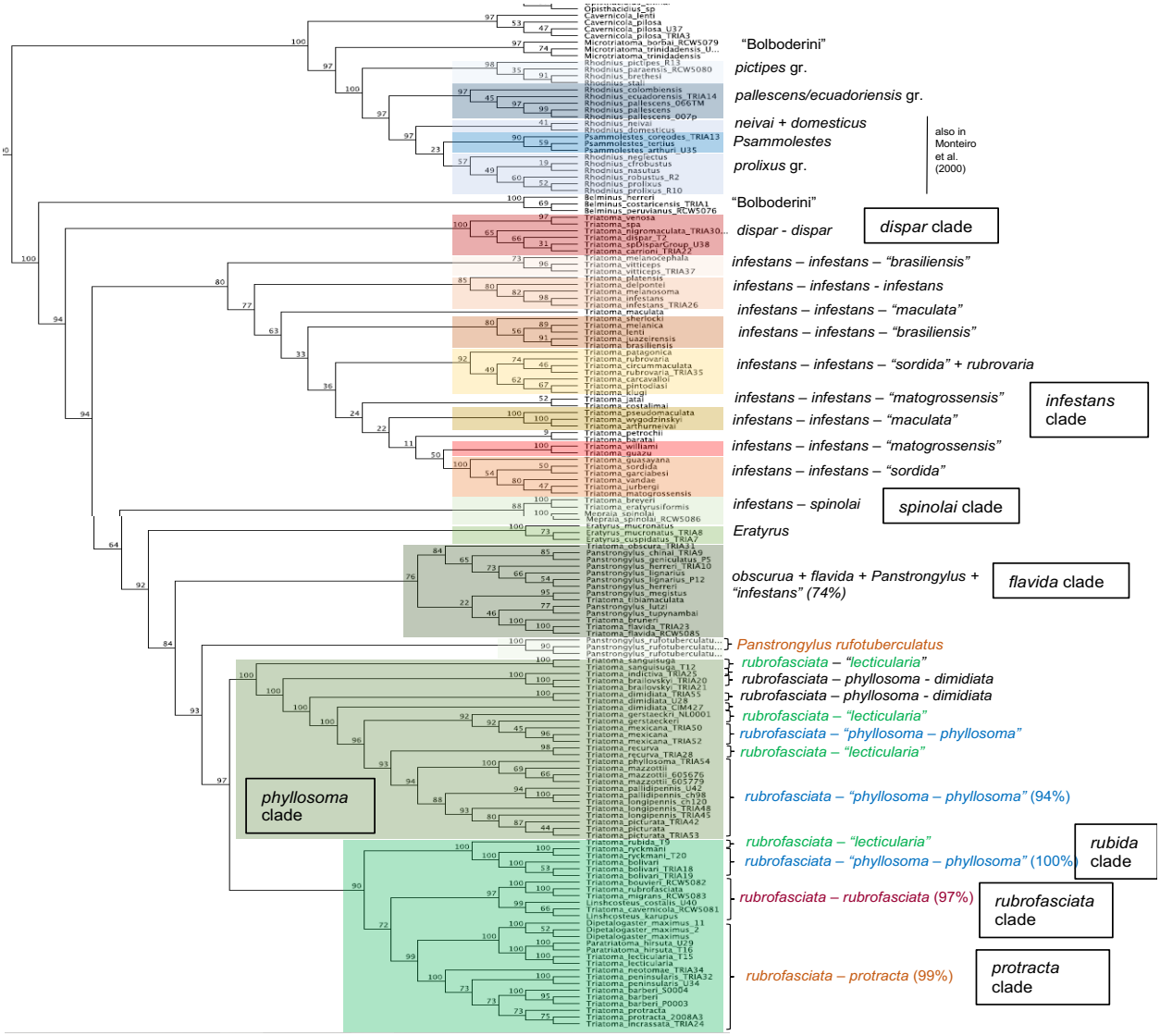


Figure 6.2. Maximum Likelihood tree from the partitioned 85% UCE+ribosomal dataset.

CHAPTER 7

CONCLUSIONS

Vector-borne disease pose a significant to public health today through a combination of new and resurgent infectious diseases. Many of these diseases are multi-host zoonoses, making attempts at control or elimination difficult, and new spillover events hard to predict. These threats are connected to the increase in global human activity which has led to increased environmental change. Following behind these rapidly increase threats are the expanded research into the ecology and evolution of these diseases, processes inherently connect to the natural environment. But even with this increased research, there are still many gaps in systems (i.e. Neglected Tropical Diseases) and questions (i.e. genetic diversity and spread of disease). In the last decade or two, rapid advances in genomic technologies, particularly Next-Generation Sequencing (NGS), have led to an increase in reproducible output with decreasing costs, opening the door to further research.

Historically, vector control strategies have focused on biological information about vector species to limit population growth or human contact. This basic strategy is only enhanced with more fine scale resolution in the big data, genomics era. While numerous genetic technologies exist, there widespread use and application to real-world biocontrol efforts is not as widespread. By expanding cost-effective, easy to use, and reproducible research in Next-Generation Sequencing of disease vectors, we can begin to bridge this gap. Chagas disease, reviewed in this dissertation, presents a good test of

application for this idea in a disease system that is often low on funds and genetic resources for these kinds of research questions. In this dissertation I have validated the use of these affordable and ease of use methods toward questions in the ecological and evolutionary genetics of Chagas disease vectors.

In Chapter 2, I develop an NGS method to examine bloodmeals of a Chagas disease vector (*Rhodnius pallescens*) that fits the criteria of an ideal NGS method. The method is easy to use with a simple two-step PCR system, reproducible, with interchangeable components for a customizable method that can help reduce initial primer costs in addition to sequencing costs. The bioinformatics uses existing publicly available data (i.e. GenBank) and supported software (i.e. QIIME) that allows for straight forward analyses even for beginner users. Chapter 2 allows for widespread use of bloodmeal/diet analyses in complex mixed samples that can open the does to use in many different taxa and/or systems resulting a extensive datasets that identify reservoir host species. Further research such as network analyses can then be applied to the data for a more complete understanding of disease reservoirs.

In Chapter 3, I investigate the whole-body microbiome of *Rhodnius pallescens*. Microbiome research in insects, particularly disease vectors, is still in the early stages. Microbiomes have been seen as one potential route to vector biocontrol but first we must understand what taxa are present and in what context before we can determine functions and how those functions may be utilized for disease control. In Chapter 3, I used a standard 16S rRNA amplicon to describe the microbiome of *R. pallescens* in two geographically distinct areas in Panama. The microbiota composition was distinct between these localities, particularly for *Wolbachia sp*. *Wolbachia* has been proposed as a

taxon of interest for disease vector control. *Wolbachia* in this study (Chapter 3) was found in very high abundance in the Panama Oeste samples and not at all in the Veraguas specimens. This study helps provide a baseline for future research examining the microbiome of *R. pallescens* in various environmental/ecological contexts.

In Chapter 4, I investigate the population genetics of *R. pallescens* in Panama using the same samples from Chapter 3. This RADseq (3RAD) method illustrates the low-cost, ease of use method for low resourced labs. In this study there was a very distinct separation of Panama Oeste vs Veraguas specimens with essentially zero gene flow. This compounding information on distinctiveness of the Veraguas populations has spark ideas of possible speciation, which require further investigation employing many of the methods used within this dissertation. This method also shows the wide range of questions and vectors this simple method could be applied to, as further research in population genetics of disease vectors is needed to combat there spread.

In Chapter 5, I validate a set of previously designed target enrichment baits. This set of Ultraconserved Elements (UCE) loci for hemipteran taxa opens the door to many research applications in the ecology and evolution of this insect group. In particular, I use this same bait set in Chapter 6. Here, I examine the phylogenetics of the Hemiptera subfamily Triatominae, vectors of Chagas disease. This UCE phylogeny, paired with ribosomal 16S, 18S, and 28S sequences from GenBank represents the most comprehensive phylogeny of Triatominae to date. This data can be reused in the future with new monetary resources being devoted to the collection of UCE data from new Triatominae and related taxa for a truly complete phylogeny of this important group of

disease vectors. Understanding the evolutionary history of vectors and non-vectors is equally important for control efforts as the ecological knowledge.

In summary, this dissertation provides an overview of several different NGS methods that can be applied to disease vector systems, encompassing the range of ecological (diet, microbiome, population genetic, Chapters 2-4) and evolutionary (UCE phylogeny, Chapters 5-6) types of research questions that can be examined. In addition, many are simple to use, and cost-effective for potential wide-spread adoption in low resource labs. By harnessing the combined use of these methods and others, a deeper understanding of disease vectors systems can be obtained. For example, combining species microbiome or diet analysis with phylogenetic to examine the evolutionary history of this ecological attributes.

APPENDIX A

Metadata for all samples used in Chapter 3

Metadata for all samples used in this study including NCBI accession numbers.

#BarcodeID	SampleID	InputFileName	Location	MicroHab	Palm	AgeClass	NumBM	TypeBM	Human	1stBM	2ndBM	3rdBM	4thBM	121	122	Duplex_Trangeli	Duplex_Tcruzi	Coinfection	Tcruzi_SingleInf	Trangeli_SingleInf	Accession	NCBI link
A1	007p	Merged A1_R.fastq	pavas	Pasture	Palma1	N3	one	Mammal	No	Didelphis				negative	negative	negative	negative	no	negative	negative	SAMN11673035	https://www.ncbi.nlm.nih.gov/biosample/11673035
B1	008p	Merged B1_R.fastq	pavas	Pasture	Palma1	N3	one	Mammal	No	Didelphis				negative	negative	negative	negative	no	negative	negative	SAMN11673036	https://www.ncbi.nlm.nih.gov/biosample/11673036
C1	009p	Merged C1_R.fastq	pavas	Pasture	Palma1	N2	one	Mammal	No	Didelphis				negative	negative	negative	negative	no	negative	negative	SAMN11673037	https://www.ncbi.nlm.nih.gov/biosample/11673037
D1	010p	Merged D1_R.fastq	pavas	Pasture	Palma1	N2	one	Mammal	No	Didelphis				negative	negative	negative	negative	no	negative	negative	SAMN11673038	https://www.ncbi.nlm.nih.gov/biosample/11673038
E1	011p	Merged E1_R.fastq	pavas	Pasture	Palma1	N2	one	Mammal	No	Didelphis				negative	negative	negative	negative	no	negative	negative	SAMN11673039	https://www.ncbi.nlm.nih.gov/biosample/11673039
F1	012p	Merged F1_R.fastq	pavas	Pasture	Palma1	N2	unk	unk	No	Other				negative	negative	negative	negative	no	negative	negative	SAMN11673040	https://www.ncbi.nlm.nih.gov/biosample/11673040
G1	013p	Merged G1_R.fastq	pavas	Pasture	Palma1	N3	unk	unk	No	Other				negative	negative	positive	negative	no	negative	negative	SAMN11673041	https://www.ncbi.nlm.nih.gov/biosample/11673041
H1	014p	Merged H1_R.fastq	pavas	Pasture	Palma1	N2	two	Mammal	Yes	Human	Didelphis			positive	negative	negative	negative	no	single	negative	SAMN11673042	https://www.ncbi.nlm.nih.gov/biosample/11673042
A2	036p	Merged A2_R.fastq	pavas	Peridomestic	Palma2	N3	three	Mammal	No	Tamandua	Didelphis	Coendu		negative	positive	negative	negative	no	negative	single	SAMN11673043	https://www.ncbi.nlm.nih.gov/biosample/11673043
B2	037p	Merged B2_R.fastq	pavas	Peridomestic	Palma2	N3	two	Mammal	No	Didelphis	Coendu			negative	positive	negative	negative	no	negative	single	SAMN11673044	https://www.ncbi.nlm.nih.gov/biosample/11673044
C2	038p	Merged C2_R.fastq	pavas	Peridomestic	Palma2	N3	two	Mammal	No	Coendu	Didelphis			negative	positive	negative	negative	no	negative	single	SAMN11673045	https://www.ncbi.nlm.nih.gov/biosample/11673045
D2	039p	Merged D2_R.fastq	pavas	Peridomestic	Palma2	N3	two	Mammal	No	Didelphis	Coendu			negative	positive	negative	negative	no	negative	single	SAMN11673046	https://www.ncbi.nlm.nih.gov/biosample/11673046
E2	040p	Merged E2_R.fastq	pavas	Peridomestic	Palma2	N3	two	Mammal	No	Didelphis	Coendu			negative	positive	negative	negative	no	negative	single	SAMN11673047	https://www.ncbi.nlm.nih.gov/biosample/11673047
F2	045p	Merged F2_R.fastq	pavas	Peridomestic	Palma2	N3	one	Mammal	No	Coendu				negative	positive	negative	negative	no	negative	single	SAMN11673048	https://www.ncbi.nlm.nih.gov/biosample/11673048
G2	046p	Merged G2_R.fastq	pavas	Peridomestic	Palma2	N3	two	Mammal	No	Didelphis	Coendu			negative	positive	negative	negative	no	negative	single	SAMN11673049	https://www.ncbi.nlm.nih.gov/biosample/11673049
H2	048p	Merged H2_R.fastq	pavas	Peridomestic	Palma2	N2	two	Mammal	No	Coendu	Didelphis			negative	positive	negative	negative	no	negative	single	SAMN11673050	https://www.ncbi.nlm.nih.gov/biosample/11673050
A3	049p	Merged A3_R.fastq	pavas	Peridomestic	Palma2	N2	one	Mammal	No	Didelphis				negative	positive	negative	negative	no	negative	single	SAMN11673051	https://www.ncbi.nlm.nih.gov/biosample/11673051
B3	050p	Merged B3_R.fastq	pavas	Peridomestic	Palma2	N2	three	Mammal	Yes	Myotis	Human	Coendu		positive	positive	positive	yes	coinfection	coinfection	SAMN11673052	https://www.ncbi.nlm.nih.gov/biosample/11673052
C3	051p	Merged C3_R.fastq	pavas	Peridomestic	Palma2	N2	three	Mammal	No	Myotis	Coendu	Didelphis		negative	positive	negative	negative	no	negative	single	SAMN11673053	https://www.ncbi.nlm.nih.gov/biosample/11673053
D3	060p	Merged D3_R.fastq	pavas	Peridomestic	Palma2	N3	three	Mammal	No	Coendu	Didelphis	Myotis		negative	positive	negative	negative	no	negative	single	SAMN11673054	https://www.ncbi.nlm.nih.gov/biosample/11673054
E3	061p	Merged E3_R.fastq	pavas	Peridomestic	Palma2	N3	two	Mammal	No	Didelphis	Coendu			negative	positive	negative	negative	no	negative	single	SAMN11673055	https://www.ncbi.nlm.nih.gov/biosample/11673055
F3	062p	Merged F3_R.fastq	pavas	Peridomestic	Palma2	N3	three	Mammal	Yes	Myotis	Human	Coendu		negative	positive	negative	negative	no	negative	single	SAMN11673056	https://www.ncbi.nlm.nih.gov/biosample/11673056
G3	063p	Merged G3_R.fastq	pavas	Peridomestic	Palma2	N3	one	Mammal	No	Didelphis				negative	positive	negative	negative	no	negative	single	SAMN11673057	https://www.ncbi.nlm.nih.gov/biosample/11673057
H3	064p	Merged H3_R.fastq	pavas	Peridomestic	Palma2	N3	two	Mammal	No	Didelphis	Coendu			negative	positive	negative	negative	no	negative	single	SAMN11673058	https://www.ncbi.nlm.nih.gov/biosample/11673058
A4	065p	Merged A4_R.fastq	pavas	Peridomestic	Palma2	N3	three	Mammal	Yes	Human	Myotis	Didelphis		negative	positive	negative	negative	no	negative	single	SAMN11673059	https://www.ncbi.nlm.nih.gov/biosample/11673059
B4	066p	Merged B4_R.fastq	pavas	Peridomestic	Palma2	N2	three	Mammal	Yes	Myotis	Human	Didelphis		negative	positive	negative	negative	no	negative	single	SAMN11673060	https://www.ncbi.nlm.nih.gov/biosample/11673060
C4	067p	Merged C4_R.fastq	pavas	Peridomestic	Palma2	N2	one	Mammal	No	Didelphis				negative	negative	negative	negative	no	negative	negative	SAMN11673061	https://www.ncbi.nlm.nih.gov/biosample/11673061
D4	029TM	Merged D4_R.fastq	minas	Peridomestic	Palma3	N3	one	Mammal	No	Didelphis				positive	positive	positive	yes	coinfection	coinfection	SAMN11673062	https://www.ncbi.nlm.nih.gov/biosample/11673062
E4	030TM	Merged E4_R.fastq	minas	Peridomestic	Palma3	N3	one	Mammal	No	Didelphis				positive	positive	positive	yes	coinfection	coinfection	SAMN11673063	https://www.ncbi.nlm.nih.gov/biosample/11673063
F4	031TM	Merged F4_R.fastq	minas	Peridomestic	Palma3	N3	two	Mammal	Yes	Myotis	Human			positive	positive	negative	yes	coinfection	coinfection	SAMN11673064	https://www.ncbi.nlm.nih.gov/biosample/11673064
G4	032TM	Merged G4_R.fastq	minas	Peridomestic	Palma3	N2	three	Mammal	Yes	Human	Myotis	Didelphis		positive	positive	positive	yes	coinfection	coinfection	SAMN11673065	https://www.ncbi.nlm.nih.gov/biosample/11673065
H4	033TM	Merged H4_R.fastq	minas	Peridomestic	Palma3	N2	three	Mammal	Yes	Didelphis	Human	Myotis		positive	positive	positive	yes	coinfection	coinfection	SAMN11673066	https://www.ncbi.nlm.nih.gov/biosample/11673066
A5	034TM	Merged A5_R.fastq	minas	Peridomestic	Palma3	N2	three	Mammal	Yes	Myotis	Didelphis	Human		positive	positive	positive	yes	coinfection	coinfection	SAMN11673067	https://www.ncbi.nlm.nih.gov/biosample/11673067
B5	035TM	Merged B5_R.fastq	minas	Peridomestic	Palma3	N2	one	Mammal	No	Didelphis				positive	positive	positive	yes	coinfection	coinfection	SAMN11673068	https://www.ncbi.nlm.nih.gov/biosample/11673068
C5	036TM	Merged C5_R.fastq	minas	Peridomestic	Palma3	N2	two	Mammal	No	Didelphis	Myotis			positive	positive	positive	yes	coinfection	coinfection	SAMN11673069	https://www.ncbi.nlm.nih.gov/biosample/11673069
D5	037TM	Merged D5_R.fastq	minas	Peridomestic	Palma3	N2	two	Mammal	No	Didelphis	Myotis			positive	positive	positive	yes	coinfection	coinfection	SAMN11673070	https://www.ncbi.nlm.nih.gov/biosample/11673070
E5	039TM	Merged E5_R.fastq	minas	Peridomestic	Palma3	N2	one	Mammal	No	Didelphis				positive	positive	positive	yes	coinfection	coinfection	SAMN11673071	https://www.ncbi.nlm.nih.gov/biosample/11673071
F5	040TM	Merged F5_R.fastq	minas	Peridomestic	Palma3	N2	unk	Mammal	No	Other				positive	positive	positive	yes	coinfection	coinfection	SAMN11673072	https://www.ncbi.nlm.nih.gov/biosample/11673072
G5	041TM	Merged G5_R.fastq	minas	Peridomestic	Palma3	N2	unk	Mammal	No	Other				positive	positive	positive	yes	coinfection	coinfection	SAMN11673073	https://www.ncbi.nlm.nih.gov/biosample/11673073
H5	042TM	Merged H5_R.fastq	minas	Peridomestic	Palma3	N2	two	Mammal	Yes	Human	Didelphis			positive	positive	positive	yes	coinfection	coinfection	SAMN11673074	https://www.ncbi.nlm.nih.gov/biosample/11673074
A6	044TM	Merged A6_R.fastq	minas	Peridomestic	Palma3	N2	one	Mammal	No	Metachirus				positive	positive	positive	yes	coinfection	coinfection	SAMN11673075	https://www.ncbi.nlm.nih.gov/biosample/11673075
B6	045TM	Merged B6_R.fastq	minas	Peridomestic	Palma3	N2	two	Mammal	Yes	Human	Myotis			positive	positive	positive	yes	coinfection	coinfection	SAMN11673076	https://www.ncbi.nlm.nih.gov/biosample/11673076
C6	046TM	Merged C6_R.fastq	minas	Peridomestic	Palma3	N2	two	Mammal	Yes	Human	Didelphis			negative	positive	negative	no	negative	single	SAMN11673077	https://www.ncbi.nlm.nih.gov/biosample/11673077
D6	047TM	Merged D6_R.fastq	minas	Peridomestic	Palma3	N2	two	Mammal	No	Didelphis	Metachirus			positive	negative	positive	no	single	negative	SAMN11673078	https://www.ncbi.nlm.nih.gov/biosample/11673078
E6	048TM	Merged E6_R.fastq	minas	Peridomestic	Palma3	N2	one	Mammal	No	Didelphis				positive	positive	positive	yes	coinfection	coinfection	SAMN11673079	https://www.ncbi.nlm.nih.gov/biosample/11673079
F6	050TM	Merged F6_R.fastq	minas	Peridomestic	Palma4	M	one	Mammal	No	Metachirus				positive	positive	negative	yes	coinfection	coinfection	SAMN11673080	https://www.ncbi.nlm.nih.gov/biosample/11673080
G6	061TM	Merged G6_R.fastq	minas	Peridomestic	Palma4	N3	four	Bird_Mammal	Yes	Human	Piranga	Myotis	Didelphis	negative	negative	negative	no	negative	negative	SAMN11673081	https://www.ncbi.nlm.nih.gov/biosample/11673081
H6	062TM	Merged H6_R.fastq	minas	Peridomestic	Palma4	N3	three	Mammal	Yes	Human	Metachirus	Myotis		negative	negative	negative	no	negative	negative	SAMN11673082	https://www.ncbi.nlm.nih.gov/biosample/11673082
A7	063TM	Merged A7_R.fastq	minas	Peridomestic	Palma4	N2	two	Bird_Mammal	Yes	Piranga	Human			negative	negative	negative	no	negative	negative	SAMN11673083	https://www.ncbi.nlm.nih.gov/biosample/11673083
B7	064TM	Merged B7_R.fastq	minas	Peridomestic	Palma4	N2	two	Reptile_Mammal	Yes	Human	Mabuya			negative	negative	negative	no	negative	negative	SAMN11673084	https://www.ncbi.nlm.nih.gov/biosample/11673084
C7	065TM	Merged C7_R.fastq	minas	Peridomestic	Palma4	N2	two	Mammal	Yes	Myotis	Human			negative	negative	negative	no	negative	negative	SAMN11673085	https://www.ncbi.nlm.nih.gov/biosample/11673085
D7	066TM	Merged D7_R.fastq	minas	Peridomestic	Palma4	N1	three	Mammal	Yes	Human	Didelphis	Myotis		positive	positive	negative	yes	coinfection	coinfection	SAMN11673086	https://www.ncbi.nlm.nih.gov/biosample/11673086
E7	067TM	Merged E7_R.fastq	minas	Peridomestic	Palma3	N1	three	Mammal	Yes	Metachirus	Human	Other		positive	positive	positive	yes	coinfection	coinfection	SAMN11673087	https://www.ncbi.nlm.nih.gov/biosample/11673087
F7	068TM	Merged F7_R.fastq	minas	Peridomestic	Palma3	N1	two	Mammal	Yes	Myotis	Human			positive	positive	positive	yes	coinfection	coinfection	SAMN11673088	https://www.ncbi.nlm.nih.gov/biosample/11673088
G7	069TM	Merged G7_R.fastq	minas	Peridomestic	Palma3	N1	unk	unk	No	Other				positive	positive	positive	yes	coinfection	coinfection	SAMN11673089	https://www.ncbi.nlm.nih.gov/biosample/11673089
H7	070TM	Merged H7_R.fastq	minas	Peridomestic	Palma3	N1	two	Mammal	Yes	Human	Myotis			negative	positive	negative	no	negative	single	SAMN11673090	https://www.ncbi.nlm.nih.gov/biosample/11673090
A8	071TM	Merged A8_R.fastq	minas	Peridomestic	Palma3	N1	two	Mammal	Yes	Human	Myotis			positive	positive	positive	yes	coinfection	coinfection	SAMN11673091	https://www.ncbi.nlm.nih.gov/biosample/11673091
B8	073TM	Merged B8_R.fastq	minas	Peridomestic	Palma3	N1	four	Mammal	Yes	Human	Myotis	Didelphis	Coendu	positive	positive	positive	yes	coinfection	coinfection	SAMN11673092	https://www.ncbi.nlm.nih.gov/biosample/11673092
C8	074TM	Merged C8_R.fastq	minas	Peridomestic	Palma3	N1	four	Mammal	Yes	Didelphis	Coendu	Human	Myotis	negative	negative	negative	no	negative	negative	SAMN11673093	https://www.ncbi.nlm.nih.gov/biosample/11673093
B7	203	Merged B7_R.fastq	veraguas	Peridomestic_Forest	SF1	N5	zero	Mammal	Unk	NA				negative	positive	negative	no	negative	single	SAMN11673094	https://www.ncbi.nlm.nih.gov/biosample/11673094
C7	204	Merged C7_R.fastq	veraguas	Peridomestic_Forest	SF1	N5	two	Mammal	Yes	Didelphis	Human			negative	positive	negative	no	negative	single	SAMN11673095	https://www.ncbi.nlm.nih.gov/biosample/11673095
D7	205	Merged D7_R.fastq	veraguas	Peridomestic_Forest	SF1	N5	four	Mammal	Yes	Didelphis	Oecomys	Human	Canis	negative	positive	negative	no	negative	single	SAMN11673096	https://www.ncbi.nlm.nih.gov/biosample/11673096
E7	206	Merged E7_R.fastq	veraguas	Peridomestic_Forest	SF1	N3	two	Mammal	Yes	Human	Didelphis			positive	positive	positive	yes	coinfection	coinfection	SAMN11673097	https://www.ncbi.nlm.nih.gov/biosample/11673097
F7	207	Merged F7_R.fastq	veraguas	Peridomestic_Forest	SF1	N3	one	Mammal	No	Didelphis				positive	negative	negative	no	single	negative	SAMN11673098	https://www.ncbi.nlm.nih.gov/biosample/11673098
G7	208	Merged G7_R.fastq	veraguas	Peridomestic_Forest	SF1	N2	one	Mammal	Yes	Human				negative	positive	positive	no	negative	single	SAMN11673099	https://www.ncbi.nlm.nih.gov/biosample/11673099
H7	209	Merged H7_R.fastq	veraguas	Peridomestic_Forest	SF1	N2	two	Mammal	Yes	Human	Myotis			positive	positive	positive	yes	coinfection	coinfection	SAMN11673100	https://www.ncbi.nlm.nih.gov/biosample/11673100
A5	349	Merged A5_R.fastq	veraguas	Pasture	SF7	N5	four	Mammal	Yes	Oecomys	Human	Other	Mus	positive	negative	positive	no	single	negative	SAMN11673101	https://www.ncbi.nlm.nih.gov/biosample/11673101
B5	349.1	Merged B5_R.fastq	veraguas	Pasture	SF7	N1	zero	Mammal	Unk	NA				negative	negative	negative	no	negative	negative	SAMN11673102	https://www.ncbi.nlm.nih.gov/biosample/11673102
C5	350	Merged C5_R.fastq	veraguas	Pasture	SF7	N4	three	Mammal	Yes	Mus	Human	Oecomys		positive	negative	positive	no	single	negative	SAMN11673103	https://www.ncbi.nlm.nih.gov/biosample/11673103
D5	351	Merged D5_R.fastq	veraguas	Pasture	SF7	N2	two	Mammal	Yes	Mus	Human			positive	negative	positive	no	single	negative	SAMN11673104	https://www.ncbi.nlm.nih.gov/biosample/11673104
E5	352	Merged E5_R.fastq	veraguas	Pasture	SF7	N2	two	Mammal	Yes	Mus	Human			positive	negative	positive	no	single	negative	SAMN11673105	https://www.ncbi.nlm.nih.gov/biosample/11673105
F5	354	Merged F5_R.fastq	veraguas	Pasture	SF7	N1	two	Mammal	Yes	Mus	Human			negative	negative	negative	no	negative	negative	SAMN11673106	https://www.ncbi.nlm.nih.gov/biosample/11673106
G5	355	Merged G5_R.fastq	veraguas	Pasture	SF7	N1	three	Reptile_Mammal	Yes	Lepidodactylus	Human	Myotis		negative	negative	negative	no	negative	negative	SAMN11673107	https://www.ncbi.nlm.nih.gov/biosample/11673107
H5	357	Merged H5_R.fastq	veraguas	Pasture	SF7	N1	one	Mammal	Yes	Human				negative	negative	negative	no	negative	negative	SAMN11673108	https://www.ncbi.nlm.nih.gov/biosample/11673108
A6	358	Merged A6_R.fastq	veraguas	Pasture	SF7	N2	two	Mammal	Yes	Human	Myotis			negative	positive	negative	no	negative	single	SAMN11673109	https://www.ncbi.nlm.nih.gov/biosample/11673109
B6	363	Merged B6_R.fastq	veraguas	Pasture	SF8	N5	zero	Mammal	Unk	NA				positive	negative	positive	no	single	negative	SAMN11673110	https://www.ncbi.nlm.nih.gov/biosample/11673110
C6	364	Merged C6_R.fastq	veraguas	Pasture	SF8	N3	three	Mammal	Yes	Human	Didelphis	Mus		positive	negative	positive	no	single	negative	SAMN11673111	https://www.ncbi.nlm.nih.gov/biosample/11673111
D6	365	Merged D6_R.fastq	veraguas	Pasture	SF8	N2	two	Mammal	No	Didelphis	Mus			positive	negative	positive	no	single	negative	SAMN11673112	https://www.ncbi.nlm.nih.gov/biosample/11673112
E6	366	Merged E6_R.fastq	veraguas	Pasture	SF8	N2	one	Mammal	No	Mus				positive	negative	positive	no	single	negative	SAMN11673113	https://www.ncbi.nlm.nih.gov/biosample/11673113
F6	369	Merged F6_R.fastq	veraguas	Pasture	SF8	N1	one	Mammal	No	Mus				negative	negative	negative	no	negative	negative	SAMN11673114	https://www.ncbi.nlm.nih.gov/biosample/11673114
G6	374	Merged G6_R.fastq	veraguas	Pasture	SF8	N1	one	Mammal	No	Mus				negative	negative	negative	no	negative	negative	SAMN11673115	https://www.ncbi.nlm.nih.gov/biosample/11673115
A7	378	Merged A7_R.fastq	veraguas	Pasture	SF8	N1	two	Mammal	Yes	Didelphis	Human			negative	negative	negative	no	negative	negative	SAMN11673116	https://www.ncbi.nlm.nih.gov/biosample/11673116
A8	386	Merged A8_R.fastq	veraguas	Peridomestic	SF11	N3	two	Mammal	No	Didelphis	Philander			negative	negative	negative	no	negative	negative	SAMN11673117	https://www.ncbi.nlm.nih.gov/biosample/11673117
B8	387	Merged B8_R.fastq	veraguas	Peridomestic	SF11	N2	one	Mammal	No	Didelphis				positive	positive	positive	yes	coinfection	coinfection	SAMN11673118	https://www.ncbi.nlm.nih.gov/biosample/11673118
C8	388	Merged C8_R.fastq	veraguas	Peridomestic	SF11	N1	one	Mammal	No	Didelphis				positive	positive	positive	yes	coinfection	coinfection	SAMN11673119	https://www.ncbi.nlm.nih.gov/biosample/11673119
D8	390	Merged D8_R.fastq	veraguas	Peridomestic	SF11	N2	one	Mammal	No	Didelphis				negative	positive	negative	no	negative	negative	SAMN11673120	https://www.ncbi.nlm.nih.gov/biosample/11673120
E8	392	Merged E8_R.fastq	veraguas	Peridomestic	SF11	N1	one	Mammal	No	Didelphis				negative	negative	negative	no	negative	negative	SAMN11673121	https://www.ncbi.nlm.nih.gov/biosample/11673121
F8	394	Merged F8_R.fastq	veraguas	Peridomestic	SF11	N1	two	Mammal	No	Didelphis	Mus			negative	negative	negative	no	negative	negative	SAMN11673122	https://www.ncbi.nlm.nih.gov/biosample/11673122
G8	404	Merged G8_R.fastq	veraguas	Peridomestic	SF11	N1	three	Mammal	Yes	Didelphis	Mus	Human		positive	negative	positive	no	single	negative	SAMN11673123	https://www.ncbi.nlm.nih.gov/biosample/11673123

APPENDIX B

Phylum Level Read Counts For Each Location in Chapter 3

Phylum level read counts and proportions for each of the three locations.

	Las Pavas		Trinidad de las Minas		Veraguas	
Phylum	**# Reads**	**% Reads**	**# Reads**	**% Reads**	**# Reads**	**% Reads**
Acidobacteria	96	0.04%	364	0.09%	1,420	0.55%
Actinobacteria	13,091	6.13%	66,898	15.88%	71,158	27.56%
Armatimonadetes	6	0.00%	21	0.00%	144	0.06%
Bacteroidetes	29,844	13.98%	11,190	2.66%	44,251	17.14%
BRC1	0	0.00%	0	0.00%	114	0.04%
Chlamydiae	0	0.00%	15	0.00%	322	0.12%
Chloroflexi	5	0.00%	67	0.02%	1,057	0.41%
Cyanobacteria	129	0.06%	389	0.09%	1,160	0.45%
Deferribacteres	0	0.00%	0	0.00%	289	0.11%
Deinococcus-Thermus	0	0.00%	3	0.00%	83	0.03%
Dependentiae	0	0.00%	0	0.00%	177	0.07%
Elusimicrobia	0	0.00%	0	0.00%	40	0.02%
Epsilonbacteraeota	0	0.00%	0	0.00%	357	0.14%
Euryarchaeota	0	0.00%	0	0.00%	36	0.01%
FBP	0	0.00%	57	0.01%	207	0.08%
Firmicutes	2,602	1.22%	4,433	1.05%	29,635	11.48%
Fusobacteria	12	0.01%	0	0.00%	66	0.03%
Gemmatimonadetes	40	0.02%	30	0.01%	324	0.13%
Kiritimatiellaeota	0	0.00%	0	0.00%	149	0.06%
Lentisphaerae	0	0.00%	0	0.00%	120	0.05%
Patescibacteria	1,232	0.58%	282	0.07%	2,405	0.93%
Planctomycetes	452	0.21%	544	0.13%	4,737	1.83%
Proteobacteria	152,869	71.63%	320,707	76.11%	68,241	26.43%
Rokubacteria	0	0.00%	0	0.00%	17	0.01%
Spirochaetes	0	0.00%	0	0.00%	286	0.11%
Tenericutes	3,810	1.79%	1,265	0.30%	815	0.32%
Thaumarchaeota	0	0.00%	5	0.00%	0	0.00%
Unassigned;__	700	0.33%	32	0.01%	569	0.22%
Verrucomicrobia	5	0.00%	368	0.09%	4,337	1.68%
WPS-2	0	0.00%	23	0.01%	65	0.03%
Unclassified Bacteria	8,531	4.00%	14,697	3.49%	25,599	9.92%
	213,424	100.00%	421,390	100.00%	258,180	100.00%

APPENDIX C

Top 20 Bacterial Families for Each Location in Chapter 3

Top 20 Family level read counts and proportions for each of the three locations.

	Las Pavas			Trinidad de las Minas			Veraguas	
Family	**# Reads**	**% Reads**	**Family**	**# Reads**	**% Reads**	**Family**	**# Reads**	**% Reads**
Anaplasmataceae	103,085	48.30%	Anaplasmataceae	305,570	72.51%	Unclassified Bacteria	25,599	9.92%
Moraxellaceae	13,487	6.32%	Pseudonocardiaceae	30,279	7.19%	Pseudonocardiaceae	21,512	8.33%
Pseudomonadaceae	12,785	5.99%	Unclassified Bacteria	14,697	3.49%	Enterobacteriaceae	15,577	6.03%
Flavobacteriaceae	10,752	5.04%	Dietziaceae	8,444	2.00%	Muribaculaceae	14,070	5.45%
Sphingobacteriaceae	9,694	4.54%	Nocardioidaceae	8,101	1.92%	Sphingomonadaceae	10,518	4.07%
Unclassified Bacteria	8,531	4.00%	Streptomycetaceae	5,815	1.38%	Streptomycetaceae	10,101	3.91%
Weeksellaceae	8,431	3.95%	Nocardiopsaceae	4,831	1.15%	Lachnospiraceae	9,579	3.71%
Xanthomonadaceae	6,672	3.13%	Pseudomonadaceae	3,078	0.73%	Nocardioidaceae	9,082	3.52%
Burkholderiaceae	4,679	2.19%	Sphingobacteriaceae	2,511	0.60%	Ruminococcaceae	8,783	3.40%
Mycoplasmataceae	3,704	1.74%	Moraxellaceae	2,490	0.59%	Moraxellaceae	8,732	3.38%
Enterobacteriaceae	3,395	1.59%	Weeksellaceae	2,490	0.59%	Rhizobiaceae	8,560	3.32%
Rhizobiaceae	3,276	1.53%	Bacillaceae	2,308	0.55%	Brevibacteriaceae	6,526	2.53%
Dietziaceae	2,609	1.22%	Sphingomonadaceae	2,185	0.52%	Bacteroidaceae	6,348	2.46%
Pseudonocardiaceae	2,115	0.99%	Muribaculaceae	1,923	0.46%	Rikenellaceae	5,536	2.14%
Caulobacteraceae	1,880	0.88%	Nocardiaceae	1,703	0.40%	Sphingobacteriaceae	4,792	1.86%
Micrococcaceae	1,711	0.80%	Flavobacteriaceae	1,539	0.37%	Lactobacillaceae	4,173	1.62%
Bacillaceae	1,614	0.76%	Enterobacteriaceae	1,443	0.34%	Prevotellaceae	3,826	1.48%
Mycobacteriaceae	1,501	0.70%	Xanthomonadaceae	1,246	0.30%	Actinobacteria unclassified	3,169	1.23%
Streptomycetaceae	1,354	0.63%	Burkholderiaceae	1,206	0.29%	Burkholderiaceae	3,120	1.21%
Nocardiaceae	1,295	0.61%	Frankiaceae	1,073	0.25%	Beijerinckiaceae	2,648	1.03%
Other	10,854	5.09%	Other	18,458	4.38%	Other	75,929	29.41%
Totals	213,424	100.00%		421,390	100.00%		258,180	100.00%

APPENDIX D

Top 20 Bacterial Genera for Each Location in Chapter 3

Top 20 Genus level read counts and proportions for each of the three locations.

	Las Pavas			Trinidad de las Minas			Veraguas	
Genus	**# Reads**	**% Reads**	**Genus**	**# Reads**	**% Reads**	**Genus**	**# Reads**	**% Reads**
Wolbachia	103,085	48.30%	Wolbachia	305,570	72.51%	Unclassified Bacteria	25,599	9.92%
Acinetobacter	13,487	6.32%	Unclassified Bacteria	14,697	3.49%	Streptomyces	10,018	3.88%
Pseudomonas	12,785	5.99%	Actinomycetospora	9,706	2.30%	Acinetobacter	8,408	3.26%
Flavobacterium	10,746	5.04%	Dietzia	8,444	2.00%	uncultured bacterium	8,018	3.11%
Sphingobacterium	9,420	4.41%	Pseudonocardia	7,964	1.89%	Pseudonocardia	7,715	2.99%
Unclassified Bacteria	8,531	4.00%	Nocardioidaceae unclassified	6,996	1.66%	Actinomycetospora	7,573	2.93%
Stenotrophomonas	5,656	2.65%	Pseudonocardiaceae unclassified	6,727	1.60%	Brevibacterium	6,526	2.53%
Empedobacter	5,611	2.63%	Streptomyces	5,815	1.38%	Bacteroides	6,348	2.46%
Mycoplasma	3,704	1.74%	Saccharopolyspora	4,032	0.96%	Pectobacterium	6,279	2.43%
Dietzia	2,609	1.22%	Lipingzhangella	3,691	0.88%	Enterobacteriaceae unclassified	5,750	2.23%
Comamonas	2,583	1.21%	Pseudomonas	3,078	0.73%	Rhizobiaceae unclassified	5,467	2.12%
Chryseobacterium	2,362	1.11%	Acinetobacter	2,490	0.59%	Pseudonocardiaceae unclassified	5,350	2.07%
Providencia	2,052	0.96%	Bacillus	2,308	0.55%	Sphingomonas	4,727	1.83%
Brevundimonas	1,880	0.88%	Empedobacter	1,852	0.44%	Sphingomonadaceae unclassified	4,209	1.63%
Bacillus	1,614	0.76%	Crossiella	1,788	0.42%	Lactobacillus	4,173	1.62%
Mycobacterium	1,501	0.70%	Flavobacterium	1,521	0.36%	Muribaculaceae unclassified	4,026	1.56%
Streptomyces	1,284	0.60%	Gordonia	1,436	0.34%	Nocardioides	3,784	1.47%
Actinomycetospora	1,239	0.58%	Sphingomonadaceae unclassified	1,377	0.33%	Alistipes	3,567	1.38%
Rhizobiaceae unclassified	1,210	0.57%	Sphingobacterium	1,222	0.29%	Lachnospiraceae unclassified	3,388	1.31%
Gordonia	1,023	0.48%	Nocardiopsis	1,140	0.27%	Actinobacteria unclassified	3,169	1.23%
Other	21,042	9.86%	Other	29,536	7.01%	Other	124,086	48.06%
Totals	213,424	100.00%		421,390	100.00%		258,180	100.00%

APPENDIX E

Chapter 4 Sample Metadata

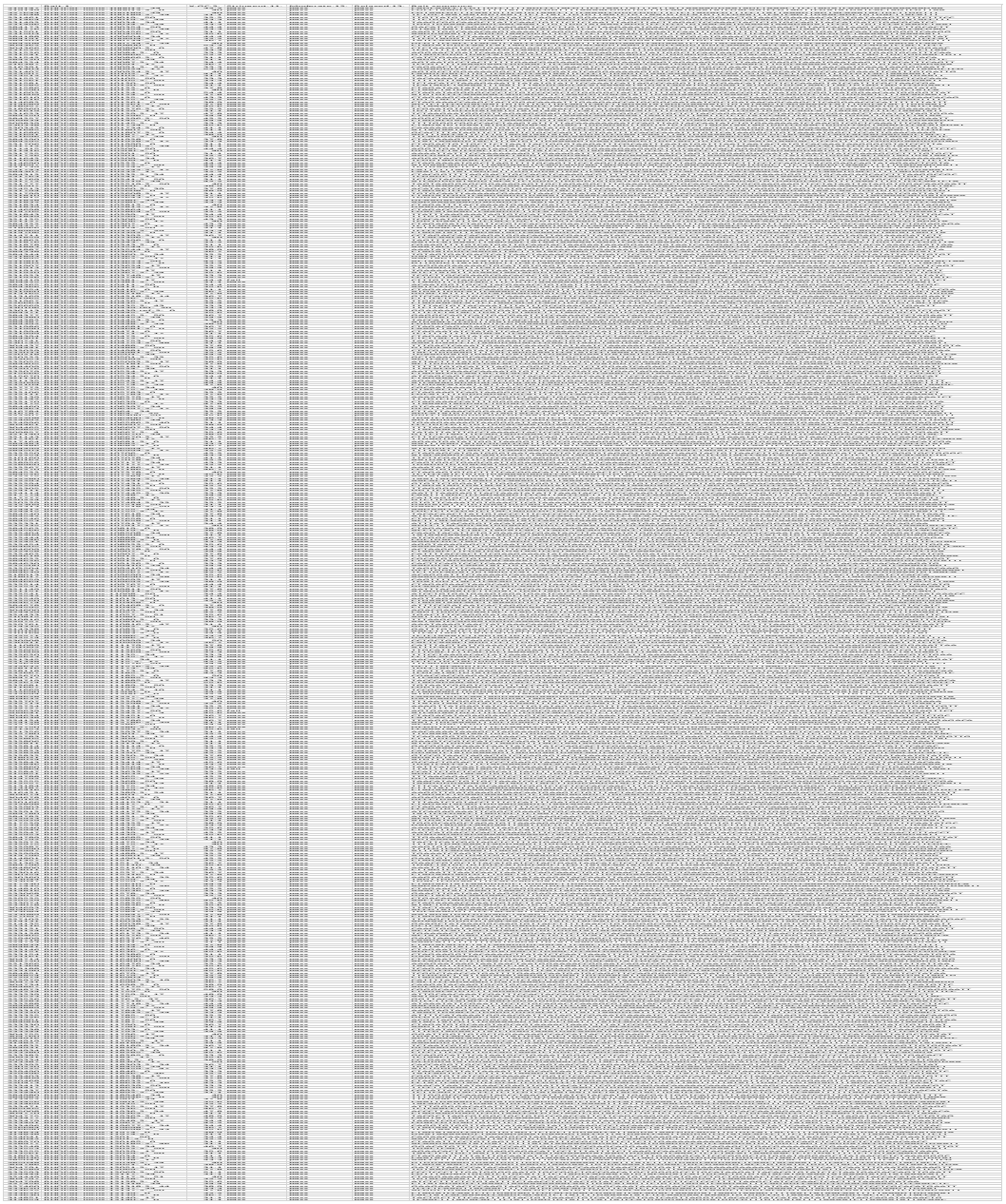
Table of sample metadata including sampling locations, sequencing reads numbers, coverage, and accession numbers for data deposited in the Sequence Read Archive (SRA).

Sample	Locality	Province	Canal	Palm Code	Palm Number	Total Reads Raw	Filtered Reac	Covarage (x)	BioProject Accession	BioSample Accession	SRA Accession
AB130-2	SR	Panama_Oes	West	AB130	SR-1	1262894	1212624	20.48	PRJNA595751	SAMN13567134	SRR10694799
AB130-4	SR	Panama_Oes	West	AB130	SR-1	1805108	1673573	33.55	PRJNA595751	SAMN13567135	SRR10694798
AB130-7	SR	Panama_Oes	West	AB130	SR-1	1183370	1094162	18.84	PRJNA595751	SAMN13567136	SRR10694782
AB131-10	SR	Panama_Oes	West	AB131	SR-2	1651000	1453122	25.04	PRJNA595751	SAMN13567137	SRR10694771
AB131-1	SR	Panama_Oes	West	AB131	SR-2	1265832	1216568	20.61	PRJNA595751	SAMN13567138	SRR10694760
AB131-3	SR	Panama_Oes	West	AB131	SR-2	1589572	1420708	24.47	PRJNA595751	SAMN13567139	SRR10694749
AB131-4	SR	Panama_Oes	West	AB131	SR-2	1739134	1609953	26.48	PRJNA595751	SAMN13567140	SRR10694738
AB162-13	Ch	Panama	East	AB162	CH-1	1347828	1252229	21.1	PRJNA595751	SAMN13567141	SRR10694727
AB162-3	Ch	Panama	East	AB162	CH-1	1312996	1183360	15.72	PRJNA595751	SAMN13567142	SRR10694716
AB162-8	Ch	Panama	East	AB162	CH-1	993138	875150	13.07	PRJNA595751	SAMN13567143	SRR10694705
AB65-14	Ch	Panama	East	AB65	CH-2	1910960	1809111	34.75	PRJNA595751	SAMN13567144	SRR10694797
AB65-16	Ch	Panama	East	AB65	CH-2	1835232	1719662	27.85	PRJNA595751	SAMN13567145	SRR10694791
AB65-17	Ch	Panama	East	AB65	CH-2	2102812	1905950	38.58	PRJNA595751	SAMN13567146	SRR10694790
AB65-4	Ch	Panama	East	AB65	CH-2	1169900	1086486	11.81	PRJNA595751	SAMN13567147	SRR10694789
AB130-10	SR	Panama_Oes	West	AB130	SR-1	1270496	1132656	23.19	PRJNA595751	SAMN13567148	SRR10694788
007p	LP	Panama_Oes	West	Palma1	LP-1	1371782	1257907	11	PRJNA595751	SAMN11673035	SRR10694787
008p	LP	Panama_Oes	West	Palma1	LP-1	1561152	1436364	11.09	PRJNA595751	SAMN11673036	SRR10694786
009p	LP	Panama_Oes	West	Palma1	LP-1	3573132	3515382	36.58	PRJNA595751	SAMN11673037	SRR10694785
010p	LP	Panama_Oes	West	Palma1	LP-1	3451828	3379211	31.06	PRJNA595751	SAMN11673038	SRR10694784
011p	LP	Panama_Oes	West	Palma1	LP-1	3636078	3547676	27.06	PRJNA595751	SAMN11673039	SRR10694783
012p	LP	Panama_Oes	West	Palma1	LP-1	758906	734738	37.82	PRJNA595751	SAMN11673040	SRR10694781
013p	LP	Panama_Oes	West	Palma1	LP-1	2588852	2521404	23.59	PRJNA595751	SAMN11673041	SRR10694780
014p	LP	Panama_Oes	West	Palma1	LP-1	922596	891009	16.5	PRJNA595751	SAMN11673042	SRR10694779
036p	LP	Panama_Oes	West	Palma2	LP-2	5234172	5155625	40.12	PRJNA595751	SAMN11673043	SRR10694778
037p	LP	Panama_Oes	West	Palma2	LP-2	3961906	3853109	27.26	PRJNA595751	SAMN11673044	SRR10694777
038p	LP	Panama_Oes	West	Palma2	LP-2	5343010	5252288	38.12	PRJNA595751	SAMN11673045	SRR10694776
039p	LP	Panama_Oes	West	Palma2	LP-2	4681934	4606122	40.64	PRJNA595751	SAMN11673046	SRR10694775
040p	LP	Panama_Oes	West	Palma2	LP-2	1128696	1008454	8.84	PRJNA595751	SAMN11673047	SRR10694774
045p	LP	Panama_Oes	West	Palma2	LP-2	1690692	1465382	8.27	PRJNA595751	SAMN11673048	SRR10694773
046p	LP	Panama_Oes	West	Palma2	LP-2	1267366	1155493	10	PRJNA595751	SAMN11673049	SRR10694772
048p	LP	Panama_Oes	West	Palma2	LP-2	3614156	3540662	26.99	PRJNA595751	SAMN11673050	SRR10694770
049p	LP	Panama_Oes	West	Palma2	LP-2	1179066	1089174	11.01	PRJNA595751	SAMN11673051	SRR10694769
050p	LP	Panama_Oes	West	Palma2	LP-2	2690118	2646998	25.62	PRJNA595751	SAMN11673052	SRR10694768
051p	LP	Panama_Oes	West	Palma2	LP-2	177334	142046	7.82	PRJNA595751	SAMN11673053	SRR10694767
060p	LP	Panama_Oes	West	Palma2	LP-2	184322	176189	10.97	PRJNA595751	SAMN11673054	SRR10694766
061p	LP	Panama_Oes	West	Palma2	LP-2	83854	72716	6.32	PRJNA595751	SAMN11673055	SRR10694765
062p	LP	Panama_Oes	West	Palma2	LP-2	228290	202763	8.11	PRJNA595751	SAMN11673056	SRR10694764
063p	LP	Panama_Oes	West	Palma2	LP-2	203152	190858	9.97	PRJNA595751	SAMN11673057	SRR10694763
064p	LP	Panama_Oes	West	Palma2	LP-2	440162	416914	14.13	PRJNA595751	SAMN11673058	SRR10694762
065p	LP	Panama_Oes	West	Palma2	LP-2	404932	394661	7.27	PRJNA595751	SAMN11673059	SRR10694761
066p	LP	Panama_Oes	West	Palma2	LP-2	311364	296937	12.36	PRJNA595751	SAMN11673060	SRR10694759
067p	LP	Panama_Oes	West	Palma2	LP-2	170440	139124	7.18	PRJNA595751	SAMN11673061	SRR10694758
029TM	TM	Panama_Oes	West	Palma3	TM-1	954156	919170	9.53	PRJNA595751	SAMN11673062	SRR10694757
030TM	TM	Panama_Oes	West	Palma3	TM-1	3955686	3925818	27.91	PRJNA595751	SAMN11673063	SRR10694756
031TM	TM	Panama_Oes	West	Palma3	TM-1	4918856	4884350	32.37	PRJNA595751	SAMN11673064	SRR10694755
032TM	TM	Panama_Oes	West	Palma3	TM-1	4294570	4269660	34.13	PRJNA595751	SAMN11673065	SRR10694754
033TM	TM	Panama_Oes	West	Palma3	TM-1	3301954	3265417	25.54	PRJNA595751	SAMN11673066	SRR10694753
034TM	TM	Panama_Oes	West	Palma3	TM-1	2068304	1995596	16.96	PRJNA595751	SAMN11673067	SRR10694752
035TM	TM	Panama_Oes	West	Palma3	TM-1	1796830	1651584	12.06	PRJNA595751	SAMN11673068	SRR10694751
036TM	TM	Panama_Oes	West	Palma3	TM-1	1327112	1215187	9.31	PRJNA595751	SAMN11673069	SRR10694750
037TM	TM	Panama_Oes	West	Palma3	TM-1	1983734	1897552	16.58	PRJNA595751	SAMN11673070	SRR10694748
039TM	TM	Panama_Oes	West	Palma3	TM-1	1258532	1148138	10.19	PRJNA595751	SAMN11673071	SRR10694747
040TM	TM	Panama_Oes	West	Palma3	TM-1	1722038	1560457	11.55	PRJNA595751	SAMN11673072	SRR10694746
041TM	TM	Panama_Oes	West	Palma3	TM-1	1089590	993720	8.4	PRJNA595751	SAMN11673073	SRR10694745
042TM	TM	Panama_Oes	West	Palma3	TM-1	3613820	3555139	29.03	PRJNA595751	SAMN11673074	SRR10694744
044TM	TM	Panama_Oes	West	Palma3	TM-1	5408306	5308396	42.61	PRJNA595751	SAMN11673075	SRR10694743
045TM	TM	Panama_Oes	West	Palma3	TM-1	4579764	4476903	33.67	PRJNA595751	SAMN11673076	SRR10694742
046TM	TM	Panama_Oes	West	Palma3	TM-1	4389878	4327763	32.73	PRJNA595751	SAMN11673077	SRR10694741
047TM	TM	Panama_Oes	West	Palma3	TM-1	2948796	2910900	37.76	PRJNA595751	SAMN11673078	SRR10694740
048TM	TM	Panama_Oes	West	Palma3	TM-1	4816130	4728236	37.78	PRJNA595751	SAMN11673079	SRR10694739
050TM	TM	Panama_Oes	West	Palma4	TM-2	3496932	3355917	23.54	PRJNA595751	SAMN11673080	SRR10694737
061TM	TM	Panama_Oes	West	Palma4	TM-2	5692648	5671852	70	PRJNA595751	SAMN11673081	SRR10694736
062TM	TM	Panama_Oes	West	Palma4	TM-2	6660320	6630866	67.08	PRJNA595751	SAMN11673082	SRR10694735
063TM	TM	Panama_Oes	West	Palma4	TM-2	4747764	4722513	43.43	PRJNA595751	SAMN11673083	SRR10694734
064TM	TM	Panama_Oes	West	Palma4	TM-2	3734570	3708875	34.67	PRJNA595751	SAMN11673084	SRR10694733
065TM	TM	Panama_Oes	West	Palma4	TM-2	4354312	4323875	39	PRJNA595751	SAMN11673085	SRR10694732
066TM	TM	Panama_Oes	West	Palma4	TM-2	2710062	2692032	32.75	PRJNA595751	SAMN11673086	SRR10694731
067TM	TM	Panama_Oes	West	Palma3	TM-1	2780798	2755985	29.27	PRJNA595751	SAMN11673087	SRR10694730
068TM	TM	Panama_Oes	West	Palma3	TM-1	127914	119564	5.6	PRJNA595751	SAMN11673088	SRR10694729
069TM	TM	Panama_Oes	West	Palma3	TM-1	4305222	4262649	35.4	PRJNA595751	SAMN11673089	SRR10694728
070TM	TM	Panama_Oes	West	Palma3	TM-1	3160918	3145073	41.84	PRJNA595751	SAMN11673090	SRR10694726
071TM	TM	Panama_Oes	West	Palma3	TM-1	1372014	1361848	15.91	PRJNA595751	SAMN11673091	SRR10694725
073TM	TM	Panama_Oes	West	Palma3	TM-1	1117016	1106261	16.09	PRJNA595751	SAMN11673092	SRR10694724
074TM	TM	Panama_Oes	West	Palma3	TM-1	658256	642541	15.44	PRJNA595751	SAMN11673093	SRR10694723
AB104_1	SR	Panama_Oes	West	AB104	SR-3	1624460	1501647	11.63	PRJNA595751	SAMN13567149	SRR10694722
349	SF	Veraguas	West	SF7	SF-3	275856	271343	20.12	PRJNA595751	SAMN11673101	SRR10694721
349_1	SF	Veraguas	West	SF7	SF-3	458698	451019	26.73	PRJNA595751	SAMN11673102	SRR10694720
350	SF	Veraguas	West	SF7	SF-3	2858574	2797444	103.94	PRJNA595751	SAMN11673103	SRR10694719
351	SF	Veraguas	West	SF7	SF-3	3005690	2949890	116.67	PRJNA595751	SAMN11673104	SRR10694718
352	SF	Veraguas	West	SF7	SF-3	6050938	5972317	180.41	PRJNA595751	SAMN11673105	SRR10694717
354	SF	Veraguas	West	SF7	SF-3	2739958	2692245	115.05	PRJNA595751	SAMN11673106	SRR10694715
355	SF	Veraguas	West	SF7	SF-3	4876704	4797808	150.89	PRJNA595751	SAMN11673107	SRR10694714
357	SF	Veraguas	West	SF7	SF-3	1810450	1739393	91.17	PRJNA595751	SAMN11673108	SRR10694713
358	SF	Veraguas	West	SF7	SF-3	2602510	2570253	76.12	PRJNA595751	SAMN11673109	SRR10694712
363	SF	Veraguas	West	SF8	SF-2	153534	149855	13.68	PRJNA595751	SAMN11673110	SRR10694711
364	SF	Veraguas	West	SF8	SF-2	1651382	1617207	67.4	PRJNA595751	SAMN11673111	SRR10694710
365	SF	Veraguas	West	SF8	SF-2	4770070	4682338	162.17	PRJNA595751	SAMN11673112	SRR10694709
366	SF	Veraguas	West	SF8	SF-2	2947170	2871829	37.26	PRJNA595751	SAMN11673113	SRR10694708
369	SF	Veraguas	West	SF8	SF-2	5721204	5578612	189.28	PRJNA595751	SAMN11673114	SRR10694707
374	SF	Veraguas	West	SF1	SF-1	2537326	2479259	104.46	PRJNA595751	SAMN11673115	SRR10694706
203	SF	Veraguas	West	SF1	SF-1	755632	739034	35.91	PRJNA595751	SAMN11673094	SRR10694704
204	SF	Veraguas	West	SF1	SF-1	1510434	1468611	58.99	PRJNA595751	SAMN11673095	SRR10694703
205	SF	Veraguas	West	SF1	SF-1	227742	781	8.12	PRJNA595751	SAMN11673096	SRR10694702
206	SF	Veraguas	West	SF1	SF-1	3024022	2961338	114.9	PRJNA595751	SAMN11673097	SRR10694701
207	SF	Veraguas	West	SF1	SF-1	4697006	4576308	131.16	PRJNA595751	SAMN11673098	SRR10694700
208	SF	Veraguas	West	SF1	SF-1	4074694	3982646	124.58	PRJNA595751	SAMN11673099	SRR10694699
209	SF	Veraguas	West	SF8	SF-2	4046744	3850586	128.82	PRJNA595751	SAMN11673100	SRR10694698
378	SF	Veraguas	West	SF11	SF-2	3781006	3715809	143.08	PRJNA595751	SAMN11673116	SRR10694697
386	SF	Veraguas	West	SF11	SF-4	3314650	3233046	69.99	PRJNA595751	SAMN11673117	SRR10694696
387	SF	Veraguas	West	SF11	SF-4	3944088	3812616	27.05	PRJNA595751	SAMN11673118	SRR10694695
388	SF	Veraguas	West	SF11	SF-4	2788080	2712479	110.66	PRJNA595751	SAMN11673119	SRR10694796
390	SF	Veraguas	West	SF11	SF-4	7260826	7088943	227.1	PRJNA595751	SAMN11673120	SRR10694795
392	SF	Veraguas	West	SF11	SF-4	5074412	4984096	184.2	PRJNA595751	SAMN11673121	SRR10694794
394	SF	Veraguas	West	SF11	SF-4	4065488	3979611	151.54	PRJNA595751	SAMN11673122	SRR10694793
404	SF	Veraguas	West	SF11	SF-4	5274532	5162768	169.39	PRJNA595751	SAMN11673123	SRR10694792
			Average (paired reads)			1288429.743	1250759.47	46.8746667			
			Minimum (paired reads)			41927	390.5	5.6			
			Maximum (paired reads)			3630413	3544471.5	227.1			
			Total (paired reads)			135285123	13132744				
			Std. Dev. (paired reads)			870081.7604	866688.808	49.5485255			

APPENDIX F

Chapter 4 RADcap Bait Design

Loci used for RADcap bait design, with GC content values, pass/fail call at a given filtration level (see main text), and bait sequence.



APPENDIX G

Chapter 4 STRUCTURE HARVESTER Output

Output graphs and tables from STRUCTURE HARVESTER for *Rhodnius pallescens*.

Structure Harvester

Job shy-wood-08d3

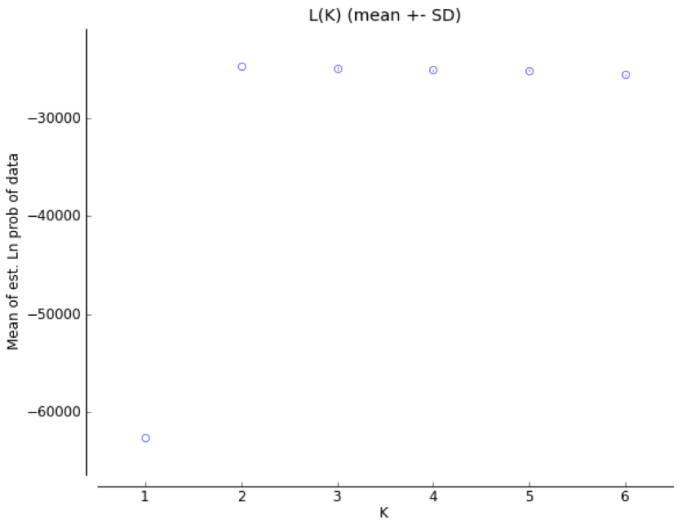
This output file was generated at:

2019-Jul-30 14:25:06 PDT

This document is not permanent. It will automatically be removed from the server in seven (7) days. Please save or print it for your records. If images are missing, try reloading; this sometimes happens under heavy server load.

Single file archive including this page, all images, all clumpp files: download. [.tar.gz]

L(K)



L(K): pdf eps

Clumpp files

K = 1 Clumpp indfile K = 1 Clumpp popfile

K = 2 Clumpp indfile K = 2 Clumpp popfile

K = 3 Clumpp indfile K = 3 Clumpp popfile

K = 4 Clumpp indfile K = 4 Clumpp popfile

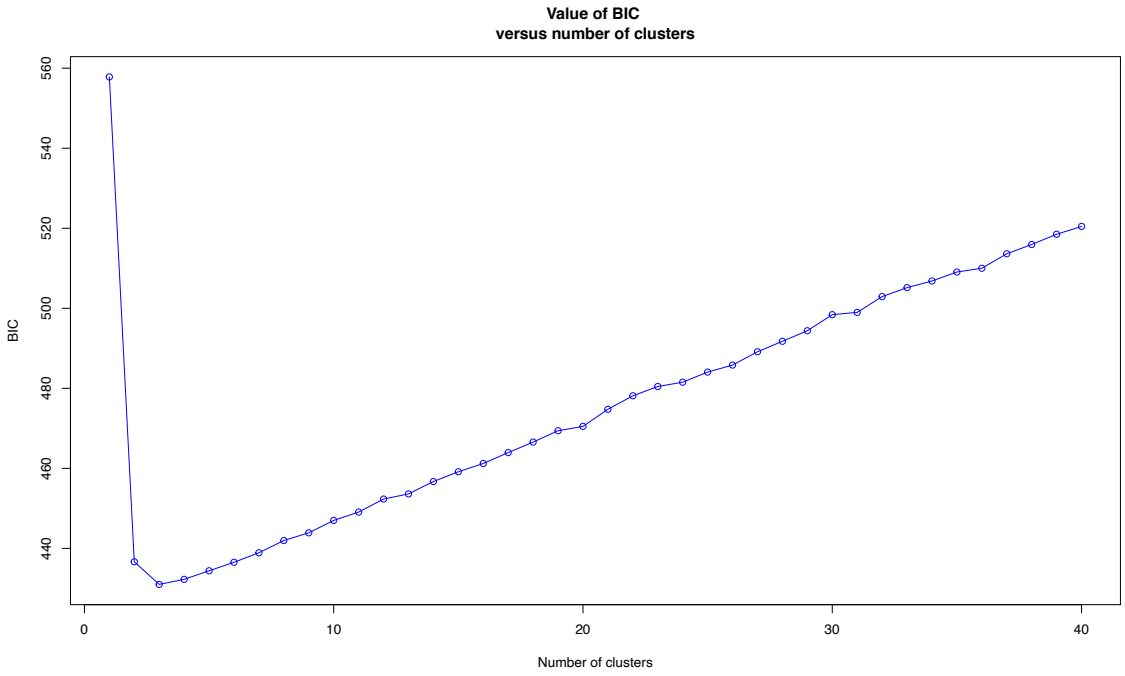
K = 5 Clumpp indfile K = 5 Clumpp popfile

K = 6 Clumpp indfile K = 6 Clumpp popfile

APPENDIX H

Chapter 4 Bayesian Information Criterion

Graph of Bayesian Information Criterion (BIC) values for each number of clusters from the *find.clusters* function from ADEGENET v. 2.1.1, where three clusters have the highest posterior probability.



APPENDIX I

Chapter 4 Mitogenome Statistics

Table of mitochondrial genome statistics, including number of sequencing reads obtained, length, number of genes, number of tRNAs, number of rRNAs, nucleotide content, and GenBank Accession numbers.

Sample ID	Location	Age Class	Raw Paired Reads	Paired Reads 80-151bp	Genome Length	# PCGs	# rRNAs	# tRNAs		A %	T %	C %	G %	Accesion #
007p	Las Pavas	N3	6,045,350	4,172,252	16,391	13	2	22	23?	40.9	31.3	17.0	10.7	MN824454
045p	Las Pavas	N3	4,728,748	3,826,616	16,391	13	2	22	23?	40.9	31.3	17.0	10.7	MN824455
066TM	Trinidad de las Minas	N1	4,552,990	3,919,260	16,391	13	2	22	23?	40.9	31.3	17.0	10.7	MN824456
203	Veraguas	N5	8,292,758	4,996,190	15,888	13	2	22		40.8	31.4	17.0	10.8	MN824451
349	Veraguas	N5	9,761,254	6,401,880	15,888	13	2	22		40.8	31.4	17.0	10.8	MN824452
363	Veraguas	N5	8,412,318	5,186,316	15,888	13	2	22		40.8	31.4	17.0	10.8	MN824453

APPENDIX J

Chapter 4 Mitochondrial Gene Comparisons

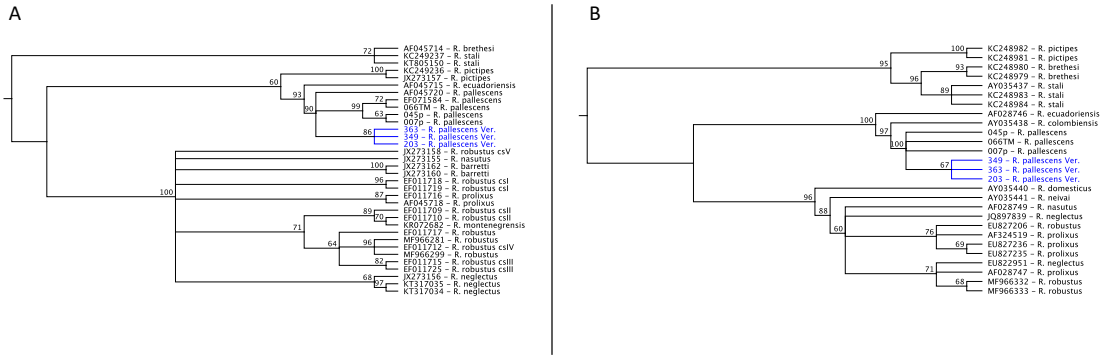
Table of mitochondrial genes showing length and percent similarity between Panama Oeste and Veraguas province localities.

Gene	bp Panama Oeste	bp Veraguas	bp difference	Panama Oeste vs Veraguas
12S	771	772	1	98.057%
16S	1,258	1,258	0	98.887%
ATP6	684	684	0	96.491-637%
ATP8	150	150	0	96.667%
CYTB	1,134	1,134	0	97.443-531%
COI	1,545	1,545	0	97.346%
COII	678	678	0	97.493-640%
COIII	813	834	21	96-679-925%
NAD1	930	930	0	96.559-667%
NAD2	996	996	0	98.092%
NAD3	351	351	0	96.296%
NAD4	1,332	1,332	0	97.222-297%
NAD4L	261	261	0	96.552-935%
NAD5	1,701	1,701	0	97.061%
NAD6	495	495	0	96.768%

APPENDIX K

Chapter 4 Phylogenetic Trees for Cytb and 16S

Phylogenetic gene trees from Cytb (A) and 16S (B) analyses. *Rhodnius pallescens* samples from Veraguas (SF) are highlighted in blue.



APPENDIX L

Chapter 4 Trypanosome Summary Statistics

Figures of summary genetic diversity 3RAD statistics (I & III) and ADEGENET v. 2.1.1 plots of individual assignments to each population (*compoplot*, II & IV) for *Trypanosoma cruzi* and *T. rangeli*, respectively.

Trypanosoma cruzi fine-scale

population genetics

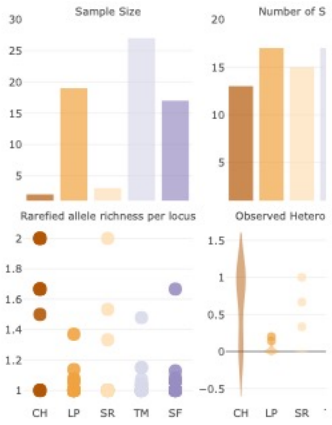


Figure I. Number of samples, number of SNPs, number of alleles, allelic richness, observed and expected heterozygosity for 17 SNPs from five exclusive loci for *Trypanosoma cruzi*.

APPENDIX M

Chapter 4 Infection Comparison Data

Tables of PCR tests and 3RAD infection data for *Trypanosoma cruzi* and *T. rangeli* in *R. pallescens* samples.

Locality	SampleID	Tcruzi_121_122	Duplex_Tcruzi	Coincidence 121-122&Duplex_Tcruzi	Tcruzi_PCR	Tcruzi_3RAD	Duplex_Trangeli	Trangeli_3RAD
SF	203	negative	negative	YES	negative		positive	*
SF	204	negative	negative	YES	negative		positive	*
SF	205	negative	negative	YES	negative		positive	*
SF	206	positive	positive	YES	positive	positive	positive	*
SF	207	positive	negative	NO	positive	positive	negative	
SF	208	negative	positive	NO	positive	positive	positive	*
SF	209	positive	positive	YES	positive	positive	positive	*
SF	349	positive	positive	YES	positive	*	negative	
SF	349.1	negative	negative	YES	negative		negative	
SF	350	positive	positive	YES	positive	*	negative	
SF	351	positive	positive	YES	positive	*	negative	
SF	352	positive	positive	YES	positive	positive	negative	positive
SF	354	negative	negative	YES	negative	positive	negative	
SF	355	negative	negative	YES	negative	positive	negative	
SF	357	negative	negative	YES	negative		negative	
SF	358	negative	negative	YES	negative		positive	*
SF	363	positive	positive	YES	positive	*	negative	
SF	364	positive	positive	YES	positive	*	negative	

SF	365	positive	positive	YES	positive	positive	negative	
SF	366	positive	positive	YES	positive	positive	negative	
SF	369	negative	negative	YES	negative	positive	negative	
SF	374	negative	negative	YES	negative		negative	
SF	378	negative	negative	YES	negative	positive	negative	positive
SF	386	negative	negative	YES	negative	positive	negative	
SF	387	positive	positive	YES	positive	*	positive	positive
SF	388	positive	positive	YES	positive	positive	positive	*
SF	390	negative	negative	YES	negative	positive	negative	positive
SF	392	negative	negative	YES	negative	positive	negative	
SF	394	negative	negative	YES	negative	positive	negative	
SF	404	positive	positive	YES	positive	positive	negative	positive
LP	007p	negative	negative	YES	negative		negative	
LP	008p	negative	negative	YES	negative	positive	negative	positive
LP	009p	negative	negative	YES	negative	positive	negative	positive
LP	010p	negative	negative	YES	negative	positive	negative	positive
LP	011p	negative	negative	YES	negative		negative	positive
LP	012p	negative	negative	YES	negative		negative	
LP	013p	negative	positive	YES	positive	positive	negative	positive
LP	014p	positive	negative	NO	positive	positive	negative	positive
LP	036p	negative	negative	YES	negative	positive	positive	positive
LP	037p	negative	negative	YES	negative	positive	positive	positive
LP	038p	negative	negative	YES	negative	positive	positive	positive
LP	039p	negative	negative	YES	negative	positive	positive	positive
LP	040p	negative	negative	YES	negative		positive	positive
LP	045p	negative	negative	YES	negative		positive	*
LP	046p	negative	negative	YES	negative	positive	positive	*
LP	048p	negative	negative	YES	negative	positive	positive	positive

LP	049p	negative	negative	YES	negative		positive	*
LP	050p	positive	positive	YES	positive	positive	positive	positive
LP	051p	negative	negative	YES	negative	positive	positive	*
LP	060p	negative	negative	YES	negative	positive	positive	*
LP	061p	negative	negative	YES	negative		positive	*
LP	062p	negative	negative	YES	negative		positive	positive
LP	063p	negative	negative	YES	negative	positive	positive	*
LP	064p	negative	negative	YES	negative	positive	positive	*
LP	065p	negative	negative	YES	negative		positive	*
LP	066p	negative	negative	YES	negative	positive	positive	*
LP	067p	negative	negative	YES	negative	positive	negative	
TM	029TM	positive	positive	YES	positive	*	positive	*
TM	030TM	positive	positive	YES	positive	positive	positive	positive
TM	031TM	positive	negative	NO	positive	positive	positive	positive
TM	032TM	positive	positive	YES	positive	positive	positive	positive
TM	033TM	positive	positive	YES	positive	positive	positive	positive
TM	034TM	positive	positive	YES	positive	positive	positive	positive
TM	035TM	positive	positive	YES	positive	positive	positive	*
TM	036TM	positive	positive	YES	positive	*	positive	positive
TM	037TM	positive	positive	YES	positive	positive	positive	positive
TM	039TM	positive	positive	YES	positive	*	positive	*
TM	040TM	positive	positive	YES	positive	*	positive	*
TM	041TM	positive	positive	YES	positive	positive	positive	*
TM	042TM	positive	positive	YES	positive	positive	positive	positive
TM	044TM	positive	positive	YES	positive	positive	positive	*
TM	045TM	positive	positive	YES	positive	positive	positive	positive
TM	046TM	negative	negative	YES	negative	positive	positive	*
TM	047TM	positive	positive	YES	positive	positive	negative	positive

TM	048 TM	positive	positive	YES	positive	positive	positive	positive
TM	050 TM	positive	negative	NO	positive	positive	positive	positive
TM	061 TM	negative	negative	YES	negative	positive	negative	positive
TM	062 TM	negative	negative	YES	negative	positive	negative	positive
TM	063 TM	negative	negative	YES	negative	positive	negative	positive
TM	064 TM	negative	negative	YES	negative	positive	negative	positive
TM	065 TM	negative	negative	YES	negative	positive	negative	positive
TM	066 TM	positive	negative	NO	positive	positive	positive	positive
TM	067 TM	positive	positive	YES	positive	positive	positive	positive
TM	068 TM	positive	positive	YES	positive	*	positive	*
TM	069 TM	positive	positive	YES	positive	positive	positive	positive
TM	070 TM	negative	negative	YES	negative	positive	positive	*
TM	071 TM	positive	positive	YES	positive	positive	positive	positive
TM	073 TM	positive	positive	YES	positive	positive	positive	positive
TM	074 TM	negative	negative	YES	negative	positive	negative	positive

121-122
just T cruzi

duplex check for both, so t cruzi douplex ideally correspond to 121-122

coinfection: from duplex

TRUE POSITIVE	31	TRUE POSITIVE	25
FALSE POSITIVE	31	FALSE POSITIVE	17
FALSE NEGATIVE	11	FALSE NEGATIVE	27
TRUE NEGATIVE	16	TRUE NEGATIVE	20

APPENDIX N

Chapter 5 Sample Metadata

Table with empirical sample collection, storage, extraction, and DNA quality information.

Sample ID	Infraorder	Family	Subfamily	Genus	Species	Collection Locality	Collection Date	Storage Method	DNA Extraction Method	DNA [ng/ul]	DNA (ng)	Average Size	Bioruptor Cycles
CMF_0018	Pentatomomorpha	Coreidae	Coreinae	Anisoscelis	flavolineatus	Panama	2008	100% EtOH	Puregene Solid Tissue Protocol	>24000	>1440		
CMF_0020	Pentatomomorpha	Coreidae	Coreinae	*Anoplocnemis*	*sp.*	Cameroon	2013	100% EtOH	Puregene Solid Tissue Protocol	>24000	>1440		
CMF_0026	Pentatomomorpha	Coreidae	Coreinae	*Mozena*	*nr. lineolata*	Costa Rica	2010	100% EtOH	Puregene Solid Tissue Protocol	>24000	>1440		
CMF_0028	Pentatomomorpha	Coreidae	Coreinae	*Acanthocephala*	*thomasi*	USA:Arizona	2014	100% EtOH	Puregene Solid Tissue Protocol	>24000	>1200		
CMF_0032	Pentatomomorpha	Coreidae	Coreinae	*Acanthocephala*	*femorata*	USA:Texas	2015	100% EtOH	Puregene Solid Tissue Protocol	>24000	>1440		
CMF_0049	Pentatomomorpha	Coreidae	Meropachyinae	*Lycambes*	*sargi*	Costa Rica	2010	100% EtOH	Puregene Solid Tissue Protocol	>24000	>1440		
CMF_0053	Pentatomomorpha	Coreidae	Coreinae	*Mygdonia*	*tuberculosa*	Nigeria	2010	100% EtOH	Puregene Solid Tissue Protocol	>24000	>1440		
CMF_0083	Pentatomomorpha	Coreidae	Coreinae	*Stenoeurilla*	*nr. prolixa*	Costa Rica	2008	100% EtOH	Puregene Solid Tissue Protocol	>24000	>1440		
CMF_0190	Pentatomomorpha	Coreidae	Coreinae	*Thasus*	*neocalifornicus*	USA:Arizona	2016	100% EtOH	Puregene Solid Tissue Protocol	>24000	>1440		
R2	Cimicomorpha	Reduviidae	Triatominae	*Rhodnius*	*robustus*	Peru	2007	100% EtOH	Qiagen DNeasy	9.02	360.8	~2344	3
U1	Pentatomorpha	Pentatomidae	Pentatominae	*Brochymena*	*sp.*	USA:Texas	2015	100% EtOH	Qiagen DNeasy	32.6	3260	2628	3
U2	Pentatomorpha	Pentatomidae	Pentatominae	*Euschistus*	*latimarginatus*	USA:Texas	2015	100% EtOH	Qiagen DNeasy	43.2	4320	3939	3
U3	Gerromorpha	Gerridae	Gerrinae	*Gerris*	*sp.*	Mexico	2009	100% EtOH	Qiagen DNeasy	29.8	2980	3950	3
U4	Superfamily Psylloidea	Psyllidae	Ciriacreminae	*Heteropsylla*	*texana*	USA: California	2016	100% EtOH	Qiagen DNeasy	53.2	5320	8781	6
U5	Superfamily Psylloidea	Aphalaridae	Spondyliaspidinae	*Glycaspis*	*brimblecombei*	USA: California	2016	100% EtOH	Qiagen DNeasy	29.8	2980	11997	9
U6	Superfamily Aphidoidea	Aphididae	Aphidinae	*Aphis*	*fabae*	USA: California	2016	100% EtOH	Qiagen DNeasy	19.7	1970	7201	6
L38	Pentatomomorpha	Lygaeidae	Lygaeinae	*Oncopeltus*	*sp.*	Peru	2010	100% EtOH	Qiagen DNeasy	17.8	890	7518	6
U8	Cicadomorpha	Cicadellidae	Cicadellinae	*Stephanolla*	*rufoapicata*	Costa Rica	2014	100% EtOH	Qiagen DNeasy	13.6	1360	8522	6
U9	Cimicomorpha	Anthocoridae	Tribe Xylocorini	*Xylastocoris*	*sp.*	Brunei	2010	100% EtOH	Qiagen DNeasy	4.12	412	~6352	6
U10		Phlaeothripidae	Phlaeothripinae	*Klambothrips*	*myopori*	USA: California	2016	100% EtOH	Qiagen DNeasy	5.5	550	~9527	9
U11	Cimicomorpha	Cimicidae	Cimicinae	*Cimex*	*adjunctus*	USA: North Carolina	2009	100% EtOH	Qiagen DNeasy	4.6	460	~1780	0
MM68	Cimicomorpha	Miridae	Isometopinae	*nr. Sophianus*	*sp.*	Thailand	2007	100% EtOH	Qiagen DNeasy	2.04	91.8	~3000	0
U15	Nepomorpha	Belostomatidae	Belostomatinae	*Abedus*	*indentatus*	USA: California	2014	100% EtOH	Qiagen DNeasy	10.6	1060	9303	9
ED7323	Dispsocoromorpha	Ceratocombidae	Trichotonannninae	*Trichotonannus*	*sp.*	Thailand	2008	100% EtOH	Qiagen DNeasy	27.8	2224	5872	6
U16	Enicocephalomorpha	Enicocephalidae	Enicocephalinae	*Oncylocotis*	*sp.*	Cameroon	2013	100% EtOH	Qiagen DNeasy	14	1120	5879	6
U18	Leptopodomorpha	Leptopodidae	Leptopodinae	*Valleriola*	*sp.*	Thailand	2006	100% EtOH	Qiagen DNeasy	11.7	1170	4585	6
ED7324	Dipsocoromorpha	Schizopteridae	Schizopterinae	*Haplonannus*	*sp.*	Trinidad	2013	100% EtOH	Qiagen DNeasy	2.22	177.6	~7200	6
U19	Pentatomomorpha	Cydnidae	Cydinae			Cameroon	2013	100% EtOH	Qiagen DNeasy	26.6	2660	5841	6
U20	Gerromorpha	Hebridae	Hebrinae	*Hebrus*	*ilellus*	Cameroon	2013	100% EtOH	Qiagen DNeasy	4.1	410	13781	9
U21	Pentatomomorpha	Aradidae	Mezirinae	*Mezira*	*sp.*	Brunei	2010	100% EtOH	Qiagen DNeasy	5.02	502	4586	6
U22	Cimicomorpha	Nabidae	Prostemmatinae	*Alloeorhynchus*	*sp.*	Democratic Republic of the Congo	2010	100% EtOH	Qiagen DNeasy	7.9	790	6207	6
U23	Pentatomomorpha	Pachygronthidae	Pachygronthinae	*Oedancala*	*sp.*	Colombia	2010	100% EtOH	Qiagen DNeasy	64.4	6440	12943	9
U24	Nepomorpha	Corixidae	Micronectinae	*Micronecta*	*sp.*	Cameroon	2013	100% EtOH	Qiagen DNeasy	4.04	404	9290	9
P5	Cimicomorpha	Reduviidae	Triatominae	*Panstrongylus*	*geniculatus*	French Guyana	2010	100% EtOH	Qiagen DNeasy	39.4	985	5794	6
(U28; UCR_ ENT 00003026)	Cimicomorpha	Reduviidae	Triatominae	*Triatoma*	*dimidiata*	Belize	2007	pinned	Qiagen QIAQUICK PCR Clean up kit	4.46	446	3125	6
(U33; UCR_ENT 00003019)	Cimicomorpha	Reduviidae	Triatominae	*Dipetalogaster*	*maximus*	Mexico	1989	pinned	Qiagen QIAQUICK PCR Clean up kit	50.8	5080	~100	0
U35	Cimicomorpha	Reduviidae	Triatominae	*Psammolestes*	*arthuri*	Venezuala	1973	pinned	Qiagen QIAQUICK PCR Clean up kit	<2		~200	0

APPENDIX O

Chapter 5 Summary Results

Table showing summary results for the empirical and Faircloth (2017) *in silico* data set.

Suborder	Infraorder	Family	Subfamily	Genus	Species	Raw Paired Reads	Reads Passed QC		Contigs	UCE Loci	On-Target	Missing Data	Reference
Auchenorrhyncha	Cicadomorpha	Cicadellidae	Cicadellinae	*Homalodisca*	*vitripennis*	NA	23,852,717	NA	2,100	1,914	91.14%	15.62%	Faircloth 2017
Auchenorrhyncha	Cicadomorpha	Cicadellidae	Cicadellinae	*Stephanolla*	*rufoapicata*	1,020,182	930,754	91.23%	4,209	1,059	25.16%	42.68%	
Heteroptera	Cimicomorpha	Anthocoridae	Tribe Xylocorini	*Xylastocoris*	*sp.*	887,724	800,230	90.14%	2,594	697	26.87%	34.7%	
Heteroptera	Cimicomorpha	Cimicidae	Cimicinae	*Cimex*	*adjunctus*	6,335,458	6,127,852	96.72%	5,587	1,216	21.76%	34.39%	
Heteroptera	Cimicomorpha	Cimicidae	Cimicinae	*Cimex*	*lectularius*	NA	10,051,932	NA	2,283	1,648	72.19%	11.35%	Faircloth 2017
Heteroptera	Cimicomorpha	Miridae	Isometopinae	*nr. Sophianus*	*sp.*	3,423,332	3,164,378	92.44%	2,643	305	11.54%	69.06%	
Heteroptera	Cimicomorpha	Nabidae	Prostemmatinae	*Alloeorhynchus*	*sp.*	2,901,484	2,590,442	89.28%	2,278	570	25.02%	42.12%	
Heteroptera	Cimicomorpha	Reduviidae	Triatominae	*Dipetalogaster*	*maximus*	2,105,974	2,040,344	96.88%	4,317	913	21.15%	52.22%	
Heteroptera	Cimicomorpha	Reduviidae	Triatominae	*Panstrongylus*	*geniculatus*	1,314,062	1,177,566	89.61%	3,026	1,177	38.90%	28.83%	
Heteroptera	Cimicomorpha	Reduviidae	Triatominae	*Psammolestes*	*arthuri*	8,672,768	8,484,296	97.83%	4,164	1,588	38.14%	22.95%	
Heteroptera	Cimicomorpha	Reduviidae	Triatominae	*Rhodnius*	*prolixus*	NA	11,054,664	NA	2,329	1,919	82.40%	9.58%	Faircloth 2017
Heteroptera	Cimicomorpha	Reduviidae	Triatominae	*Rhodnius*	*robustus*	3,552,754	3,195,092	89.93%	3,793	1,508	39.76%	16.03%	
Heteroptera	Cimicomorpha	Reduviidae	Triatominae	*Triatoma*	*dimidiata*	3,217,196	2,971,874	92.37%	4,768	1,401	29.38%	26.43%	
Heteroptera	Dipsocoromorpha	Ceratocombidae	Trichotonanninae	*Trichotonannus*	*sp.*	1,004,360	920,424	91.64%	1,389	265	19.08%	65.81%	
Heteroptera	Dipsocoromorpha	Schizopteridae	Schizopterinae	*Haplonannus*	*sp.*	728,500	657,912	90.31%	1,085	271	24.98%	69.98%	
Heteroptera	Enicocephalomorpha	Enicocephalidae	Enicocephalinae	*Oncylocotis*	*sp.*	1,209,390	1,099,202	90.89%	1,235	347	28.10%	61.57%	
Heteroptera	Gerromorpha	Gerridae	Gerrinae	*Gerris*	*buenoi*	NA	11,588,088	NA	2,266	1,908	84.20%	12%	Faircloth 2017
Heteroptera	Gerromorpha	Gerridae	Gerrinae	*Gerris*	*sp.*	1,873,498	1,721,476	91.89%	4,080	1,290	31.62%	38.2%	
Heteroptera	Gerromorpha	Hebridae	Hebrinae	*Hebrus*	*ifellus*	2,337,088	2,034,830	87.07%	2,453	481	19.61%	55.64%	
Heteroptera	Leptopodomorpha	Leptopodidae	Leptopodinae	*Valleriola*	*sp.*	2,150,432	1,956,488	90.98%	2,326	601	25.84%	47.79%	
Heteroptera	Nepomorpha	Belostomatidae	Belostomatinae	*Abedus*	*indentatus*	1,967,514	1,742,898	88.58%	1,908	577	30.24%	44.66%	
Heteroptera	Nepomorpha	Corixidae	Micronectinae	*Micronecta*	*sp.*	2,685,132	2,470,014	91.99%	2,270	502	22.11%	49.68%	
Heteroptera	Pentatomomorpha	Aradidae	Mezirinae	*Mezira*	*sp.*	1,284,302	1,199,180	93.37%	1,334	381	28.56%	62.68%	
Heteroptera	Pentatomomorpha	Coreidae	Coreinae	*Anisoscelis*	*flavolineatus*	1,507,764	1,344,146	89.15%	5,381	698	12.97%	41.53%	
Heteroptera	Pentatomomorpha	Coreidae	Coreinae	*Anoplocnemis*	*sp.*	1,553,072	1,433,772	92.32%	6,701	1,035	15.45%	21.34%	
Heteroptera	Pentatomomorpha	Coreidae	Coreinae	*Mozena*	*nr. lineolata*	1,734,122	1,528,344	88.13%	6,001	967	16.11%	24.82%	
Heteroptera	Pentatomomorpha	Coreidae	Coreinae	*Acanthocephala*	*thomasi*	3,002,966	2,764,968	92.07%	8,911	1,215	13.63%	17.3%	
Heteroptera	Pentatomomorpha	Coreidae	Coreinae	*Acanthocephala*	*femorata*	921,522	839,558	91.11%	4,037	814	20.16%	28.12%	
Heteroptera	Pentatomomorpha	Coreidae	Meropachyinae	*Lycambes*	*sargi*	1,188,020	1,081,110	91.00%	4,890	887	18.14%	25.29%	
Heteroptera	Pentatomomorpha	Coreidae	Coreinae	*Mygdonia*	*tuberculosa*	1,408,812	1,301,796	92.40%	5,855	1,046	17.87%	13.79%	
Heteroptera	Pentatomomorpha	Coreidae	Coreinae	*Stenoeurilla*	*nr. prolixa*	790,910	722,064	91.30%	3,152	944	29.95%	22.52%	
Heteroptera	Pentatomomorpha	Coreidae	Coreinae	*Thasus*	*neocalifornicus*	1,967,124	1,779,776	90.48%	5,862	1,163	19.84%	13.02%	
Heteroptera	Pentatomomorpha	Cydnidae	Cydninae			1,950,008	1,769,802	90.76%	2,135	946	44.31%	32.1%	
Heteroptera	Pentatomomorpha	Lygaeidae	Lygaeinae	*Oncopeltus*	*sp.*	1,378,500	1,267,356	91.94%	4,301	1,696	39.43%	19.82%	
Heteroptera	Pentatomomorpha	Lygaeidae	Lygaeinae	*Oncopeltus*	*fasciatus*	NA	NA	NA	2,201	1,990	90.41%	7.72%	Faircloth 2017
Heteroptera	Pentatomomorpha	Pentatomidae	Pentatominae	*Brochymena*	*sp.*	1,901,530	1,764,158	92.78%	4,784	1,460	30.52%	24.23%	
Heteroptera	Pentatomomorpha	Pentatomidae	Pentatominae	*Euschistus*	*latimarginatus*	1,605,918	1,511,668	94.13%	4,665	1,437	30.80%	26.07%	
Heteroptera	Pentatomomorpha	Pentatomidae	Pentatomnae	*Halyomorpha*	*halys*	NA	19,393,098	NA	2,257	1,926	85.33%	7.74%	Faircloth 2017
Heteroptera	Pentatomomorpha	Pachygronthidae	Pachygronthinae	*Oedancala*	*sp.*	2,343,848	2,149,304	91.70%	2,223	605	27.22%	39.89%	
Sternorrhyncha		Aphalaridae	Pachypsyllinae	*Pachypsylla*	*venusta*	NA	NA	NA	2,034	1,786	87.81%	18.71%	Faircloth 2017
Sternorrhyncha		Aphalaridae	Spondyliaspidinae	*Glycaspis*	*brimblecombei*	1,101,814	989,064	89.77%	3,436	776	22.58%	39.78%	
Sternorrhyncha		Aphididae	Aphidinae	*Acyrthosiphon*	*pisum*	NA	9,788,132	NA	2,059	1,414	68.67%	22.16%	Faircloth 2017
Sternorrhyncha		Aphididae	Aphidinae	*Aphis*	*fabae*	4,168,544	3,858,732	92.57%	3,053	1,240	40.63%	28.17%	
Sternorrhyncha		Liviidae	Euphyllurinae	*Diaphorina*	*citri*	NA	NA	NA	1,545	1,369	88.61%	31.63%	Faircloth 2017
Sternorrhyncha		Psyllidae	Ciriacreminae	*Heteropsylla*	*texana*	844,452	775,336	91.82%	2,967	479	16.14%	59.7%	
Order Thysanoptera		Phlaeothripidae	Phlaeothripinae	*Klambothrips*	*myopori*	193,984	171,672	88.50%	887	117	13.19%	84.65%	
Order Thysanoptera		Thripidae	Thripinae	*Frankliniella*	*occidentalis*	NA	NA	NA	2,197	864	39.33%	35.39%	Faircloth 2017
					Averages	**2,114,434**	**3,195,566**	**91.49%**	**3,371**	**1,036**	**35.48%**	**25.32%**	**34.44%**

APPENDIX P

Chapter 5 Data Matrices Summary

Summary values for each data matrix showing the number and percent of, parsimony- informative and uninformative, and invariant sites.

	# taxa	**# loci**	**# sites**	**# (%) informative**	**# (%) uninformative**	**# (%) invariant**
50% empirical	37	532	99,747	41,876 (42.0)	10,418 (10.4)	47,4538 (47.6)
50% empirical + in silico	47	744	136,130	65,790 (48.3)	12,393 (9.1)	57,947 (42.6)
60% empirical	37	220	40,794	16,912 (41.5)	4,274 (10.5)	19,608 (48.0)
60% empirical + in silico	47	325	59,154	27,858 (47.1)	5,273 (8.9)	26,023 (44.0)
80% empirical + in silico + transcriptome	56	232	53,927	24,933 (46.3)	4,022 (7.5)	24,912 (46.2)

APPENDIX Q

Chapter 5 UCE Loci Comparisons

Table comparing the number of loci and the amount of missing data within five of the heteropteran infraorders.

Infraorder	Family	Subfamily	Genus	Species	UCE Loci	Avg. Loci	% Missing Data	Avg. Missing Data
Cimicomorpha	Anthocoridae	Tribe Xylocorini	*Xylastocoris*	*sp.*	697		30.46	
Cimicomorpha	Cimicidae	Cimicinae	*Cimex*	*adjunctus*	1,216		31.42	
Cimicomorpha	Miridae	Isometopinae	*nr. Sophianus*	*sp.*	305		61.73	
Cimicomorpha	Nabidae	Prostemmatinae	*Alloeorhynchus*	*sp.*	570		34.36	
Cimicomorpha	Reduviidae	Triatominae	*Dipetalogaster*	*maximus*	913	1,042	47.59	32.02
Cimicomorpha	Reduviidae	Triatominae	*Panstrongylus*	*geniculatus*	1,177		24.60	
Cimicomorpha	Reduviidae	Triatominae	*Psammolestes*	*arthuri*	1,588		21.50	
Cimicomorpha	Reduviidae	Triatominae	*Rhodnius*	*robustus*	1,508		13.69	
Cimicomorpha	Reduviidae	Triatominae	*Triatoma*	*dimidiata*	1,401		22.83	
Dipsocoromorpha	Ceratocombidae	Trichotonanninae	*Trichotonannus*	*sp.*	265	268	58.74	60.60
Dipsocoromorpha	Schizopteridae	Schizopterinae	*Hoplonannus*	*sp.*	271		62.46	
Gerromorpha	Gerridae	Gerrinae	*Gerris*	*sp.*	1,290	886	34.35	39.87
Gerromorpha	Hebridae	Hebrinae	*Hebrus*	*ifellus*	481		45.38	
Nepomorpha	Belostomatidae	Belostomatinae	*Abedus*	*indentatus*	577	540	38.01	41.61
Nepomorpha	Corixidae	Micronectinae	*Micronecta*	*sp.*	502		45.21	
Pentatomomorpha	Aradidae	Mezirinae	*Mezira*	*sp.*	381		57.49	
Pentatomomorpha	Coreidae	Coreinae	*Anisoscelis*	*flavolineatus*	698		34.55	
Pentatomomorpha	Coreidae	Coreinae	*Anoplocnemis*	*sp.*	1,035		15.16	
Pentatomomorpha	Coreidae	Coreinae	*Mozena*	*nr. lineolata*	967		22.92	
Pentatomomorpha	Coreidae	Coreinae	*Acanthocephala*	*thomasi*	1,215		16.93	
Pentatomomorpha	Coreidae	Coreinae	*Acanthocephala*	*femorata*	814		25.95	
Pentatomomorpha	Coreidae	Meropachyinae	*Lycambes*	*sargi*	887		19.28	
Pentatomomorpha	Coreidae	Coreinae	*Mygdonia*	*tuberculosa*	1,046	1,020	10.03	24.57
Pentatomomorpha	Coreidae	Coreinae	*Stenoeurilla*	*nr. prolixa*	944		18.10	
Pentatomomorpha	Coreidae	Coreinae	*Thasus*	*neocalifornicus*	1,163		14.71	
Pentatomomorpha	Cydnidae	Cydninae			946		26.49	
Pentatomomorpha	Lygaeidae	Lygaeinae	*Oncopeltus*	*sp.*	1,696		21.45	
Pentatomomorpha	Pentatomidae	Pentatominae	*Brochymena*	*sp.*	1,460		26.68	
Pentatomomorpha	Pentatomidae	Pentatominae	*Euschistus*	*latimarginatus*	1,437		24.88	
Pentatomomorpha	Pachygronthidae	Pachygronthinae	*Oedancala*	*sp.*	605		33.88	
				Averages	**935**		**31.36**	

APPENDIX R

Chapter 5 UCE and In Silico Loci Comparisons

Table comparing the number of loci between the empirical data and the Faircloth (2017) *in silico* data for various taxonomic levels.

ORDER	Genus	Species	Loci	Reference	% Loci	Avg % Loci
Hemiptera			904	This Study	51.28%	
Hemiptera			1,763	Faircloth 2017		32.41%
Thysanoptera	*Klambothrips*	*myopori*	117	This Study	13.54%	
Thysanoptera	*Frankliniella*	*occidentalis*	864	Faircloth 2017		
SUBORDER			Loci	Reference	% Loci	Avg % Loci
Auchenorrhyncha			1,059	This Study	55.33%	
Auchenorrhyncha			1,914	Faircloth 2017		
Heteroptera			883	This Study	46.87%	52.74%
Heteroptera			1,884	Faircloth 2017		
Sternorrhyncha			832	This Study	56.03%	
Sternorrhyncha			1,485	Faircloth 2017		
FAMILY	Genus	Species	Loci	Reference	% Loci	Avg % Loci
Cicadellidae	*Stephanolla*	*rufoapicata*	1,059	This Study	55.33%	
Cicadellidae	*Homalodisca*	*vitripennis*	1,914	Faircloth 2017		
Reduviidae	*Dipetalogaster*	*maximus*	913	This Study	47.58%	
Reduviidae	*Rhodnius*	*prolixus*	1,919	Faircloth 2017		
Reduviidae	*Panstrongylus*	*geniculatus*	1,177	This Study	61.33%	
Reduviidae	*Rhodnius*	*prolixus*	1,919	Faircloth 2017		
Reduviidae	*Psammolestes*	*arthuri*	1,588	This Study	82.75%	
Reduviidae	*Rhodnius*	*prolixus*	1,919	Faircloth 2017		
Reduviidae	*Triatoma*	*dimidiata*	1,401	This Study	73.01%	68.02%
Reduviidae	*Rhodnius*	*prolixus*	1,919	Faircloth 2017		
Lygaeidae	*Oncopeltus*	*sp.*	1,696	This Study	85.23%	
Lygaeidae	*Oncopeltus*	*fasciatus*	1,990	Faircloth 2017		
Pentatomidae	*Brochymena*	*sp.*	1,460	This Study	75.80%	
Pentatomidae	*Halyomorpha*	*halys*	1,926	Faircloth 2017		
Aphididae	*Aphis*	*fabae*	1,240	This Study	87.69%	
Aphididae	*Acyrthosiphon*	*pisum*	1,414	Faircloth 2017		
Aphalaridae	*Glycaspis*	*brimblecombei*	776	This Study	43.45%	
Aphalaridae	*Pachypsylla*	*venusta*	1,786	Faircloth 2017		
Family	**GENUS**	Species	Loci	Reference	% Loci	Avg % Loci
Cimicidae	*Cimex*	*adjunctus*	1,216	This Study	73.79%	
Cimicidae	*Cimex*	*lectularius*	1,648	Faircloth 2017		
Reduviidae	*Rhodnius*	*robustus*	1,508	This Study	78.58%	73.33%
Reduviidae	*Rhodnius*	*prolixus*	1,919	Faircloth 2017		
Gerridae	*Gerris*	*sp.*	1,290	This Study	67.61%	
Gerridae	*Gerris*	*buenoi*	1,908	Faircloth 2017		

APPENDIX S

Chapter 5 Transcriptome Data Summary

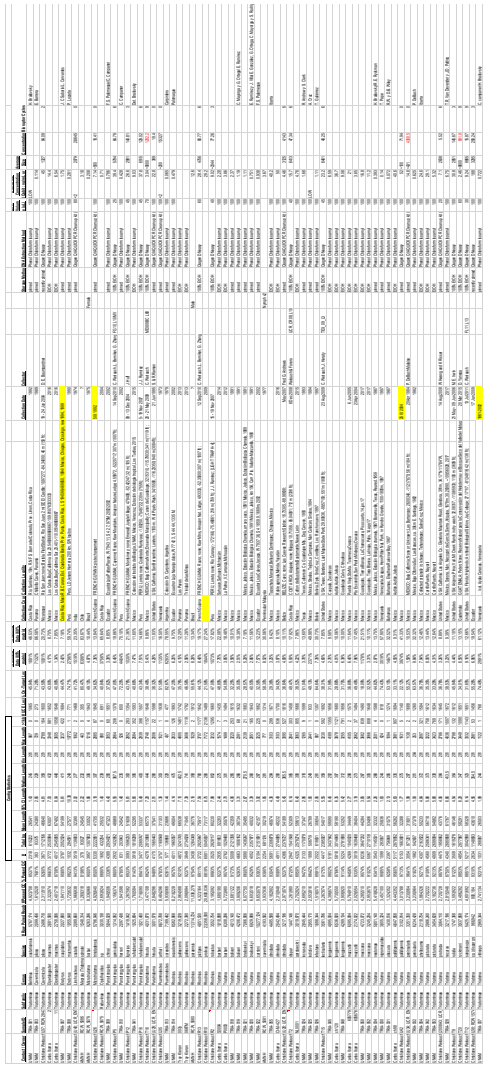
Summary results of transcriptome assembly data used for analyses for each BLAST e-value cutoff value. Asterisks indicate assembly was removed from analysis.

										Contig Length				Loci Recovered BLAST cut off		
Order	Suborder	Infraorder	Superfamily	Family	Species	Accession #	Isoforms	Genes	% GC	Median	Avg.	Assembled Bases	N50	1-e10**	1-e15	1-e20
Hemiptera	Auchenorrhyncha	Cicadomorpha	Cercopoidea	Cercopidae	*Cercopis vulnerata*	SRR921578	77,139	54,778	36.66	437	694.91	38,066,042	955	242	211	182
Hemiptera	Auchenorrhyncha	Cicadomorpha	Cicadoidea	Cicadidae	*Okanagana villosa*	SRR921625	147,189	94,409	36.94	395	622.5	58,770,023	806	254	225	186
Hemiptera	Auchenorrhyncha	Fulgoromorpha	Fulgoroidea	Delphacidae	*Nilaparvata lugens*	SRR921622	104,882	70,383	40.6	435	691.32	48,656,976	940	262	237	203
Hemiptera*	Coleorrhyncha		Peloridioidea	Peloridiidae	*Xenophysella greensladeae*	SRR921658	227,646	208,683	40.56	293	434.17	90,604,745	458		171	
Hemiptera	Sternorrhyncha		Aleyrodoidea	Aleyrodidae	*Bemisia tabaci*	GCF_001854935.1_ASM185493v1		24,428						287	263	216
Hemiptera	Sternorrhyncha		Aleyrodoidea	Aleyrodidae	*Trialeurodes vaporariorum*	SRR921651	125,563	96,957	38.92	372	597.56	57,937,670	763	248	215	169
Hemiptera	Sternorrhyncha		Coccoidea	Pseudococcidae	*Planococcus citri*	SRR921633	82,155	59,661	37.27	410	690.81	41,214,240	1048	225	191	155
Hemiptera	Heteroptera	Gerromorpha	Gerroidea	Veliidae	*Velia caprai*	SRR921656	63,492	46,481	38.91	406	609.76	28,342,354	792	201	176	151
Hemiptera	Heteroptera	Nepomorpha	Nepoidea	Nepidae	*Ranatra linearis*	SRR921639	74,737	57,141	39.69	395	628.65	35,921,536	857	206	183	150
Thysanoptera	Tubulifera		Phlaeothripoidea	Phlaeothripidae	*Gyanaikothrips ficorum*	SRR921603	219,185	174,210	41.47	325	488.9	85,171,532	577	222	184	144

APPENDIX T

Chapter 6 UCE Summary Results

Table of summary results for data collected and analyzed. Includes information on the samples used (i.e. ascension numbers, locality), summary statistics on the sequence reads processed, and the UCE data.



APPENDIX U

Chapter 6 RAxML Concatenated 60% UCE Tree

APPENDIX V

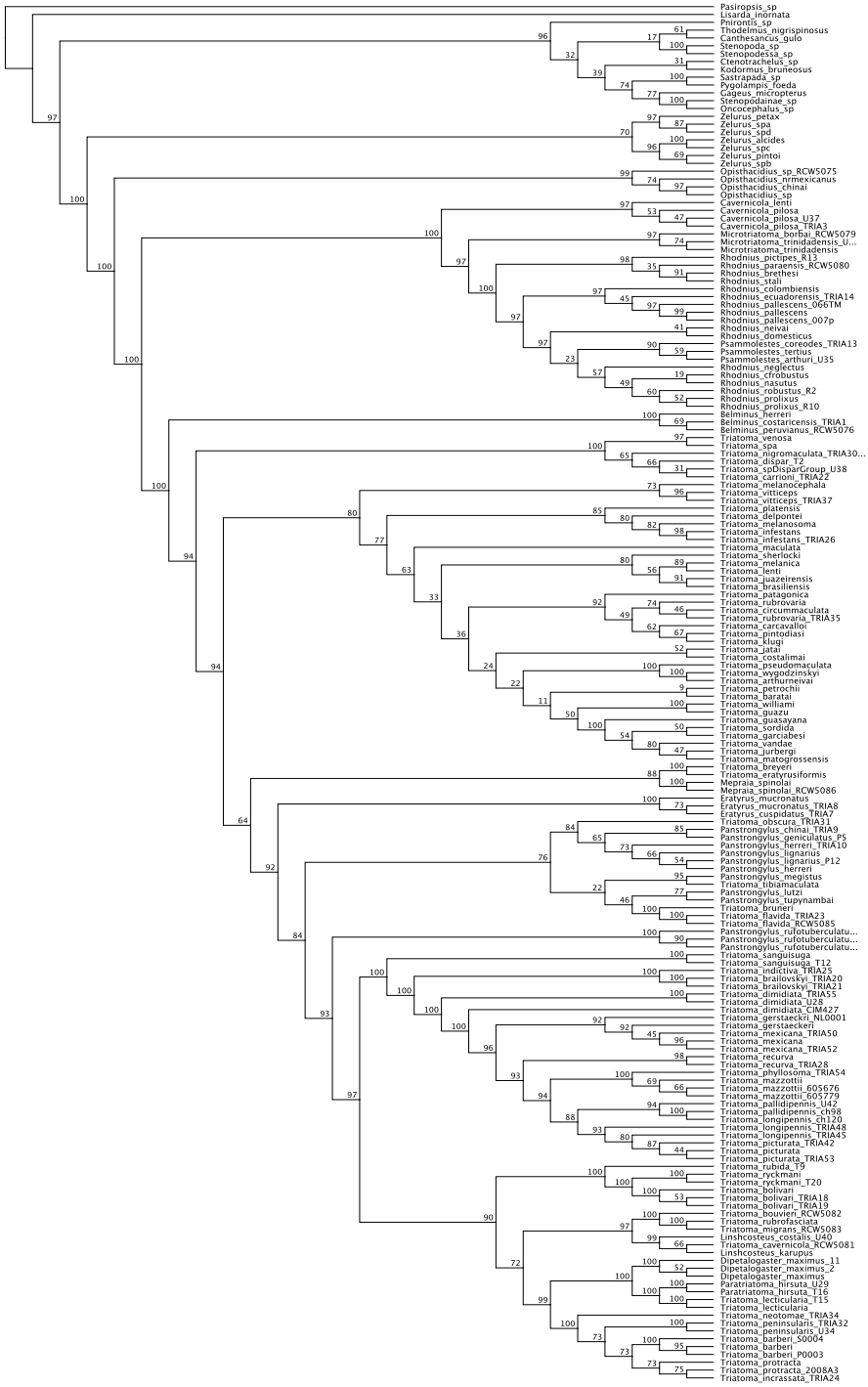
Chapter 6 RAxML Concatenated 85% UCE Tree

APPENDIX W

Chapter 6 ASTRAL UCE Gene Tree

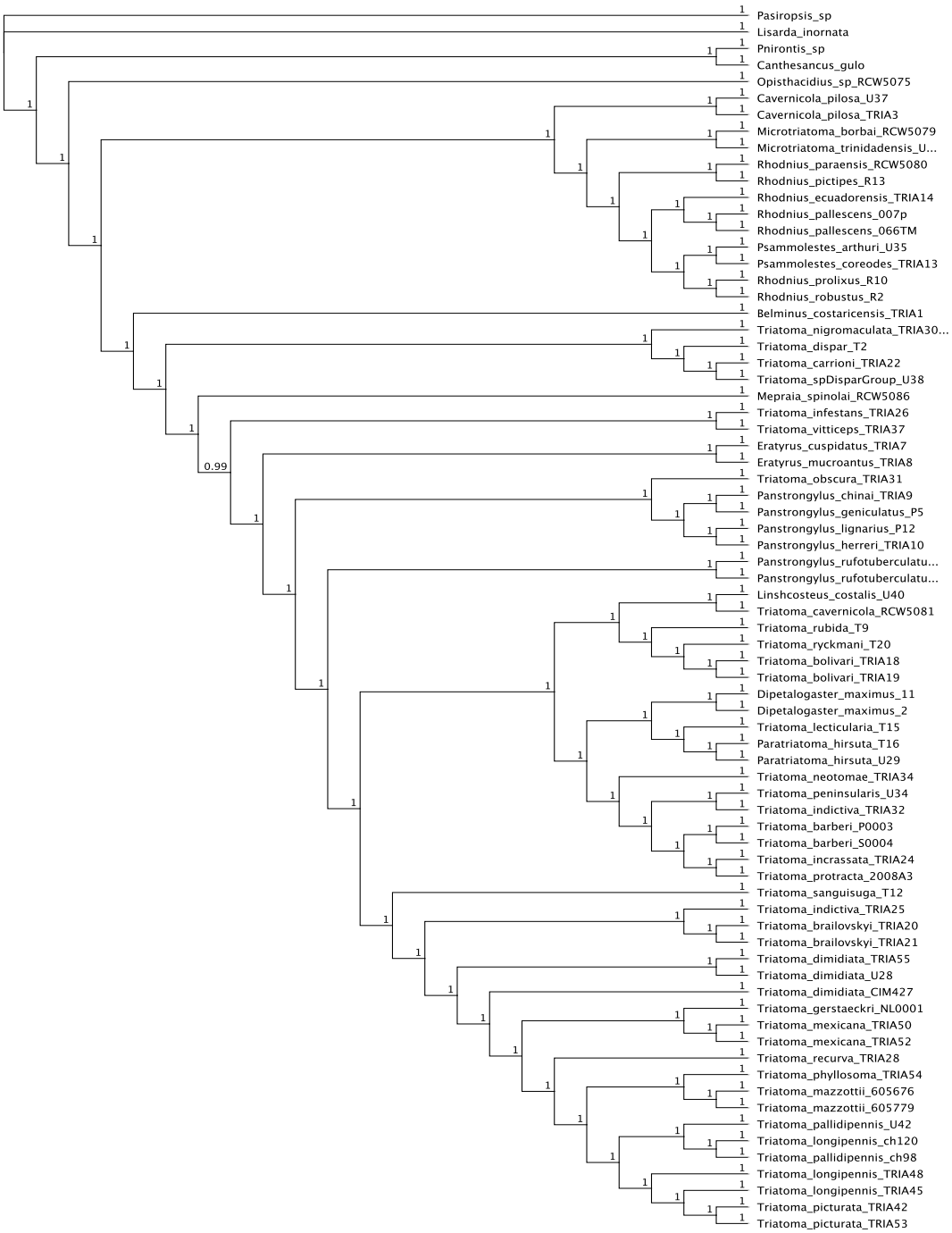
APPENDIX X

Chapter 6 RAxML 85% UCE + Ribosomal Tree



APPENDIX Y

Chapter 6 MrBayes 85% UCE Tree



APPENDIX Z

Chapter 6 MrBayes 85% UCE + Ribosomal Tree