Distribution Agreement

In presenting this dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this dissertation. I retain all ownership rights to the copyright of the dissertation. I also retain the right to use in future works (such as articles or books) all or part of this dissertation.

Signature:

_____ _____

Amanda Jo Williams-Newkirk Date

Bacterial Community Ecology of the Lone Star Tick (*Amblyomma americanum*)

By

Amanda Jo Williams-Newkirk
Doctor of Philosophy

Graduate Division of Biological and Biomedical Science
Population Biology, Ecology and Evolution

Gregory A. Dasch, Ph.D. Advisor	Thomas R. Gillespie, Ph.D. Advisor
Nicole M. Gerardo, Ph.D. Committee Member	Uriel D. Kitron, Ph.D., M.P.H. Committee Member
Timothy D. Read, Ph.D. Committee Member	James Taylor, Ph.D. Committee Member

Accepted:

Lisa A. Tedesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

Date

Bacterial Community Ecology of the Lone Star Tick (*Amblyomma americanum*)

By

Amanda Jo Williams-Newkirk
A.B., Bryn Mawr College, 2004
M.S., The University of Southern Mississippi, 2008

Advisor: Gregory A. Dasch, Ph.D.
Advisor: Thomas R. Gillespie, Ph.D.

An abstract of
A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in the Graduate Division of Biological and Biomedical Science
Population Biology, Ecology and Evolution
2015

ABSTRACT

Bacterial Community Ecology of the Lone Star Tick (*Amblyomma americanum*)
By Amanda Jo Williams-Newkirk

Vector-borne diseases dominate the list of emerging infections from the last century. Among arthropod vectors, ticks (Acari, Ixodida) transmit a greater diversity of disease causing organisms than any other known order. In the southeastern United States, the lone star tick (*Amblyomma americanum*) is the most abundant and aggressive human-biting tick and a vector of multiple known or suspected pathogens, including *Rickettsia, Ehrlichia, Borrelia,* and Heartland virus. The long life cycle of *A. americanum* and its broad host range complicate studies aimed at understanding the ecology of these diseases. Furthermore, abundant symbiotic bacteria, many of which are closely related to known pathogens, are a part of the tick's microbial community. Studies in this and other systems have robustly demonstrated that the non-pathogenic component of a vector's microbiome is capable of affecting the ability of the arthropod to acquire, maintain, and transmit pathogens to vertebrates. As a result, it is now recognized that pathogen ecology must be described in the context of the vector and the vector's microbiome. The aim of my dissertation was to provide the necessary context for further studies of *A. americanum* disease ecology. In my survey of *A. americanum* bacterial communities, I characterized the bacteria from 131 ticks collected from five sites in three states distributed throughout the species' eastern range. I then compared *A. americanum*'s bacterial community to that of sympatric tick species of varying relatedness to determine if species with the opportunity to share hosts also shared similar microbiota and to look for evidence of the evolutionary history of the ticks in their bacterial communities. During these studies I detected multiple genotypes of a novel symbiont, *Candidatus* 'Midichloria mitochondrii', in two *Amblyomma* species. In addition, I provided the first reference mitochondrial genome sequence for *A. americanum,* a tool that will be very important to future population genetics studies of this tick. My work provides a foundation for future studies in *A. americanum* by quantifying both intra and interspecific differences in bacterial communities, detecting new members of the bacterial community, and describing interactions between members which may now be targeted for characterization, metabolic evaluation, and functional studies.

Bacterial Community Ecology of the Lone Star Tick (*Amblyomma americanum*)

By

Amanda Jo Williams-Newkirk
A.B., Bryn Mawr College, 2004
M.S., The University of Southern Mississippi, 2008

Advisor: Gregory A. Dasch, Ph.D.
Advisor: Thomas R. Gillespie, Ph.D.

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in the Graduate Division of Biological and Biomedical Science
Population Biology, Ecology and Evolution
2015

ACKNOWLEDGEMENTS

I was extremely fortunate to have had the support of two advisors while at Emory University. Dr. Gregory Dasch of the Rickettsial Zoonoses Branch (RZB) at the Centers for Disease Control and Prevention (CDC) provided me with an indescribable amount of professional and personal support. I will always be grateful to him for providing me with an amazing environment in which to develop as a scientist. Dr. Thomas Gillespie gave me a space at Emory to call home, and I thank him for the many opportunities to work with the incredible undergraduate students in his laboratory.

I am also very grateful to my dissertation committee members: Drs. Nicole Gerardo, Uriel Kitron, Timothy Read, and James Taylor. They were all exceedingly generous with their time and knowledge, and it would have been extremely difficult to design and execute such an interdisciplinary dissertation without their support.

Funding for this work was provided by many generous organizations. The RZB at CDC supported me from 2010 – 2014 through an Oak Ridge Institute for Science and Education Fellowship. In 2013 – 2014, I was supported by a National Institutes of Health Training Grant awarded to the Emory Graduate Program in Population Biology, Ecology and Evolution. I was supported in 2014 – 2015 by the Association of Public Health Laboratories Bioinformatics in Public Health Practical Fellowship. Additionally, in 2013 I was awarded a National Science Foundation Doctoral Dissertation Improvement Grant to support a portion of my tick ecology research. I was also fortunate in 2009 to receive funding from the Emory Global Health Institute to support tick collections in Uganda. I thank all of these organizations for their support of my research and my education.

I was assisted during my dissertation work by many undergraduate students to whom I am grateful for their hard work and endless enthusiasm for all things tick-related. I thank Ellen Dugan, Deema Elchoufi, Ian Fried, Dana Greenlee, Jasmine Hensley, Ho Joon Hwang, Benjamin Kramer-Roach, Gwen Parker, Karen Wu, Audrey Zeis, and Gene Zhou.

I made many friends at both Emory and CDC that brought much joy and many laughs to my studies, in addition to contributing their substantial intellectual might to my work. Many thanks to Johanna Salzer, Maria Zambrano, Ryan Lash, Shari Lydy, Sandor Karpathy, Marina Eremeeva, Lori Rowe, and Seth Sims. I also thank all of the members of the RZB for their constant support and encouragement throughout this process.

It would have been impossible for me to reach this goal without the support, guidance, and example of my parents, Kim and Jimmie Williams. I will always be thankful for their love and belief in me.

Finally, I thank my husband, Rob Newkirk, for always saying “I want you to be whatever you want to be” and going above and beyond every day to make that promise a reality.

TABLE of CONTENTS

TABLE of FIGURES

TABLE of TABLES

CHAPTER 1

The evolution of tick microbiome research

Ticks (Acari, Ixodida) are an order of hematophagous arthropods closely related to mites. As a group, ticks transmit a greater diversity of pathogenic microorganisms to humans and domestic animals than any other group of arthropod vectors (Jongejan & Uilenberg 2004). The financial burden of these diseases is significant; in the United States alone, treatment of Lyme disease is estimated to cost $712 million - $1.3 billion annually (Adrion *et al.* 2015), while vaccination campaigns for tick-borne encephalitis in parts of Europe have saved tens of millions of euros (Šmit & Postma 2014).

Ixodid (hard-bodied) ticks frequently have long life cycles and feed from multiple and diverse hosts, creating complex transmission cycles which maintain zoonotic disease agents in the environment. In a typical three host tick life cycle, each of the three post-egg life stages takes a single blood meal from one host before leaving the host, digesting the blood meal, and molting into the next life stage (Figure 1.1). Adults mate prior to or during the female's blood meal; depending on the species, adult males may never feed or take multiple small blood meals. Ticks may acquire TBD agents either vertically from their mother or horizontally from their host during any of the three feedings (Figure 1.1). While feeding, ticks may transmit acquired pathogens to their vertebrate host (Figure 1.1) or to other ticks simultaneously feeding in close physical proximity on the same host (*i.e.* co-feeding transmission) (Randolph *et al.* 1996). While much effort has gone into tracing the transmission of TBD agents between tick and host(s) and to understanding the

specific environmental survival requirements of different tick species, it is still difficult to accurately predict the distribution and prevalence of TBD agents in tick populations and where human tick-borne infections will occur.

Two neglected components of TBD ecology are the interactions between ticks and the members of their greater microbial community (*i.e.* microbiome) and the interactions between tick-borne pathogens and the commensal microbiome. This gap is a historical artifact of technological limitations in the cultivation and detection of tick-borne bacteria and the slow scientific recognition that vectors are much more than mechanical vessels for transmitting pathogens between two hosts. The biology of higher organisms results from both eukaryotic and prokaryotic capabilities and their intimate interactions in a given species (Moran *et al.* 2008; Douglas 2009).

1.1 Known effects of the tick microbiome on disease ecology

Ticks may act as both vectors of pathogens, transmitting the agent between vertebrates with little amplification occurring within the tick, and as reservoirs of pathogens, serving as the primary source and an amplification host for the agent. The efficiency with which a tick performs these tasks for a given pathogen is referred to as its vector and reservoir capacity, respectively. Members of the tick microbiome may affect both of these abilities, either positively or negatively. Few examples of these effects exist in the literature due to the tendency of most studies to focus on one or two pathogens at a time; these agents may be rare in the tick population (~1-5% infection prevalence) and vary widely in their abundance within ticks when present (*e.g.* Mixson *et al.* 2006; Loftis *et al.* 2008; Fritzen

et al. 2011). However, when taken in aggregate these studies provide a compelling argument for the importance of the tick microbiome in tick-borne disease ecology throughout the infection-transmission cycle.

The first opportunity for microbiome effects arise when a pathogen attempts to establish an infection in the tick. De la Fuente *et al.* observed first in tick cell lines (2002) and later in *Dermacentor variabilis* fed on splenectomized, infected calves (2003) that a pre-existing infection with one *Anaplasma marginale* genotype prevented the superinfection of the tick with a second genotype. Whether this is due to changes in the tick cells which prevented the entry of a second *Anaplasma* genotype as suggested by the authors or some form of resource based competitive exclusion remains to be determined. Horizontal transmission via the salivary glands is the second potential point of intervention by the microbiome. Pathogens must colonize the tick salivary glands before they can be secreted into vertebrate host tissue and achieve transmission. The vertebrate pathogen *Coxiella burnetii* can be transmitted by some tick species; Šuťáková and Řeháček (1990) inoculated *Dermacentor reticulatus* ticks with *C. burnetii* (Nine Mile phase I) and *Rickettsiella phytoseiuli*. *Rickettsiella* increased the ability of *C. burnetii* to invade tick tissues, including the salivary glands, in all mixed infection trials. Thirdly, the microbiome may also interfere with the vertical transmission of pathogens within tick lineages, which reduces the tick's reservoir capacity for a given pathogen. Using laboratory colonies of ticks that had been infected with either *Rickettsia rhipicephali* or *R. montanensis* (*montana*) through capillary feeding, Macaluso *et al.* (2002) performed a reciprocal challenge on adult females from each lineage by feeding them on blood containing the paired *Rickettsia* and testing their egg clutches for both rickettsiae. They

found no evidence that females with existing *Rickettsia* infection were able to transovarially transmit a second *Rickettsia* to their offspring, suggesting that the established infection prevented colonization of the oocytes by novel *Rickettsia*. A similar scenario has been described in the Bitterroot Valley, Montana for the tick *D. andersoni*, the pathogen *R. rickettsii*, and the non-pathogenic *R. peacockii* (the "east side agent"). Burgdorfer and colleagues (1981) classically demonstrated *R. peacockii*'s ability to block vertical transmission of *R. rickettsii* in co-infected ticks, a mechanism which was used to explain the apparent lack of Rocky Mountain spotted fever cases (caused by *R. rickettsii*) among residents of the eastern half of the valley where *R. peacockii* was prevalent in ticks. However, these conclusions have since been called into question largely because of limitations in the power of the sample sizes (Telford 2009).

Tick-borne pathogens have also been shown to directly enhance the survival of their vectors and therefore increase both reservoir and vector capacity. *Ixodes ricinus* ticks naturally infected with *B. burgdorferi* sensu lato genotypes survived longer in the laboratory under suboptimal temperature and humidity conditions than did uninfected ticks (Herrmann & Gern 2010). Nymphs of this species were also found to have higher fat reserves when infected with *B. burgdorferi* (Herrmann *et al.* 2013), a trait that is likely to significantly prolong the amount of time ticks can quest for a new host before dying. Natural *Rickettsia* infections have been shown to increase larval motility in the laboratory, another trait which is likely to increase host finding rates (Kagemann & Clay 2013). *Haemaphysalis leporispalustris* ticks that obtained *Rickettsia rickettsii* from infected laboratory rabbits fed longer and produced more progeny than uninfected ticks (Freitas *et al.* 2008). This was an unusual finding; in general, virulent strains of *R.*

rickettsii are costly to their tick hosts and appear to be maintained by horizontal rather than vertical transmission (*e.g.* Niebylski *et al.* 1999).

Non-pathogenic bacteria have also been found to enhance tick survival. Antibiotic treatment of the tick *Amblyomma americanum* to reduce the load of its likely obligate symbiont *Coxiella* prolonged oviposition time for fed females and reduced the viability of their eggs (Zhong *et al.* 2007). *Candidatus* 'Midichloria mitochondrii' (hereafter *M. mitochondrii*) is found at 100% prevalence in free-living *I. ricinus* immatures and females, but drops to less than 35% within several generations of laboratory maintenance (Lo *et al.* 2006). This pattern strongly suggests that *M. mitochondrii* is a beneficial facultative symbiont under natural conditions but is too costly to maintain without selective pressure (Pistone *et al.* 2012).

Tick host microbiomes may also enhance pathogen transmission. Recently, Dunn *et al.* (2014) demonstrated in the laboratory that mice (*Peromyscus leucopus*) co-infected with *Borrelia burgdorferi* and *Babesia microti* transmitted *B. microti* to more feeding *Ixodes scapularis* larvae than did mice infected with *B. microti* alone. It is unclear from the experiment whether co-infection caused changes to the host or tick physiology are responsible for the higher transmission rate, although higher parasitemias in the co-infected mice suggest that the primary effect was on the mouse host. The authors then used field data to parameterize a model showing how *B. burgdorferi*, which typically occurs in a region newly colonized by *I. scapularis* before *B. microti*, may be necessary to raise *B. microti*'s basic reproductive number (R_0) above one and allow *B. microti* to establish and persist in the host and tick populations.

1.2 Next generation sequencing as a tool for descriptive and experimental microbiome studies

Studies of whole tick microbiomes require a methodology which is suitable for the simultaneous detection and characterization of all members of the microbiota or at least a particular component of it, *e.g.* bacteria, protozoa, or viruses. This is a fundamentally different approach than those used in the majority of tick studies prior to 2005, which used targeted PCR assays for specific known tick-borne agents. The few studies that were performed typically Sanger sequenced a few hundred cloned bacterial 16S rRNA gene (hereafter 16S) fragments (*e.g.* Benson *et al.* 2004; Clay *et al.* 2008) or used gradient gel electrophoresis to "finger print" bacterial communities for comparison (*e.g.* Schabereiter-Gurtner *et al.* 2003; Moreno *et al.* 2006; van Overbeek *et al.* 2008). These methods were insufficient to fully characterize the bacterial component of the tick microbiome.

After 2005, new types of technologies became widely available which allowed the simultaneous sequencing of heterogeneous pools of amplicons with greatly reduced time and effort. These Next Generation sequencing (NGS) platforms made it possible to sequence individual samples to a greater depth than had been previously achieved across entire sample sets. The first application of NGS to ticks was the deep sequencing of 16S amplicons from a few members of two tick species, *Rhipicephalus microplus* (Andreotti *et al.* 2011) and *I. ricinus* (Carpi *et al.* 2011). These studies provided important initial glimpses into the minority members of the tick bacterial community, but were too limited in their sample size to evaluate the range of intraspecific variation. Within a few years additional papers were published which significantly increased sample sizes when surveying microbiome composition (Lalzar *et al.* 2012; Nakao *et al.* 2013; Ponnusamy *et*

al. 2014; Budachetri *et al.* 2014) and began to test hypotheses relating to the ecology of both the tick and TBDs (Hawlena *et al.* 2013; Menchaca *et al.* 2013; Zhang *et al.* 2014; Narasimhan *et al.* 2014).

My dissertation was performed in this same time period and encapsulates this critical transition point in the evolution of tick and TBD ecology research. The series of papers presented here focus on the lone star tick, *A. americanum*, an extremely abundant and aggressive human-biting tick in the southeastern United States (Estrada-Peña & Jongejan 1999). *Amblyomma americanum* has a three host life cycle with few host restrictions at any stage (Guglielmone *et al.* 2014) and is known to transmit multiple zoonotic agents causing ehrlichiosis, rickettsiosis, and tularemia (Childs & Paddock 2003; Loftis *et al.* 2006) in humans and domestic animals. This tick has also been associated with Heartland viral infection (Savage *et al.* 2013), Southern Tick-Associated Rash Illness (STARI) (Masters *et al.* 2008), and an unusual allergic response to red meat ingestion in some people (Commins *et al.* 2011). The range of *A. americanum* has expanded into the Northeastern and Midwestern states in the last century (Ginsberg *et al.* 1991; Means & White 1997; Keirans & Lacombe 1998; Childs & Paddock 2003; Cortinas & Spomer 2013), greatly increasing the number of people potentially exposed to the pathogens it transmits.

I first used NGS of 16S amplicons to survey the bacterial communities of individual *A. americanum* from seven sites in five states along the eastern extent of the species' range (Chapters 2 and 3). In Chapter 2, I focused on a detailed examination of a potential symbiont first discovered in *A. americanum* during my work, *Candidatus* 'Midichloria mitochondrii' (hereafter *M. mitochondrii*). I detected three novel genotypes of *M.*

mitochondrii in this data set, one of which is geographically discontinuous with the other two strains. A phylogenetic analysis was used to place the novel genotypes into context with related agents from other tick species.

In Chapter 3, I selected a subset of the samples of nymphs and adult males and females from five sites in three states for whole bacterial community analysis. This chapter achieved several aims: (1) to survey the intraspecific variation in bacterial community composition and structure contained within the nymphal and adult life stages, (2) to test the hypothesis that bacterial communities differed between life stages and sexes, and (3) to test for significance in the co-occurrence or lack thereof between the more common members of the bacterial community.

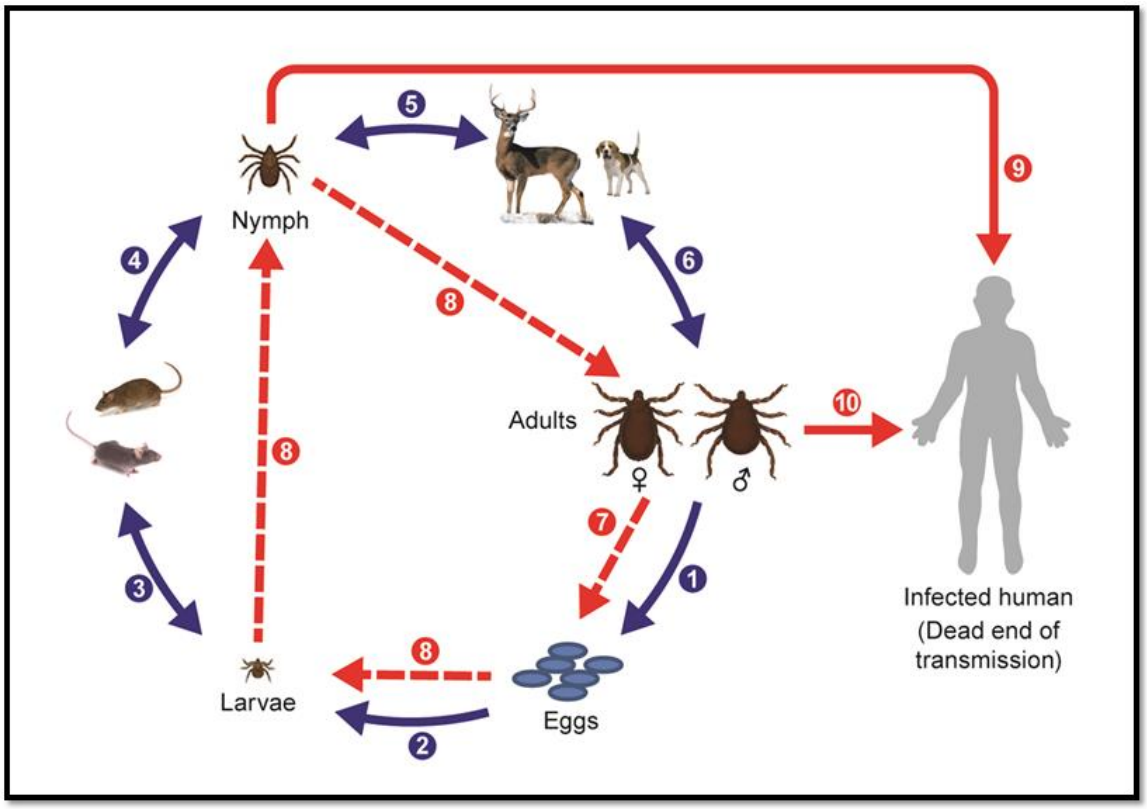
My in-depth analysis of the *A. americanum* bacterial community revealed several life stage and sex-specific differences that had not been described in previous NGS studies of this tick. To determine if these differences were common across the hard ticks and to contextualize the structure and composition of the *A. americanum* bacterial community relative to other tick species, I used NGS to characterize the communities of three species (*Amblyomma maculatum*, *D. variabilis*, and *I. scapularis*) sympatric with *A. americanum* and representative of the two major branches of the Ixodidae, Prostriata (*Ixodes*) and Metastriata (*Amblyomma, Dermacentor*). In Chapter 4, I showed that despite shared ranges and the opportunity for co-feeding on the same hosts there is little overlap in the specific bacterial 16S genotypes harbored by these species. I also tested for sex-specific bacterial communities and examined the relationships among the bacterial communities of each species for evidence that the phylogenetic history of the tick species has affected their composition or structure.

In Chapter 5, I described the first reference sequence for the complete mitochondrial genome of *A. americanum*. These data will enable the development of population genetic markers at multiple loci that are maternally inherited. In addition to examining population structure and changes related to the *A. americanum* population expansion, future studies using these markers will also permit analysis of the relationship between tick lineages and any associated variation in the microbiome, including vector capacity.

The series of papers in this dissertation represent the successful application of NGS technology to a tick of significant ecological and public health importance. From an initial survey of the bacterial community I detected and characterized several important novel symbionts, evaluated the significance of some community interactions, and identified some variables related to community differences within and between species. These findings furthered the field of TBD ecology by broadening our understanding of these complex systems and by providing targets and tools for future experimental studies.

1.3 Figures

Figure 1.1 Generalized ixodid tick three host life cycle and natural transmission of pathogens. Blue arrows indicate main steps of tick natural cycle: (1) oviposition by engorged female; (2) eggs hatched into larvae; (3) larvae feed on small animals; (4) engorged larvae hatch into nymphs; (5) nymphs feed on large or small animals; and (6) nymphs molt into adult ticks that feed on large animals or bite humans. Broken red arrows indicate transovarial (7) and transstadial transmission (8) of pathogens, and solid red arrows indicate transmission of a pathogen to humans through a bite of a nymph (9) or an adult tick (10). Figure and text reproduced from Eremeeva and Dasch (2015).



1.4 References

Adrion ER, Aucott J, Lemke KW, Weiner JP (2015) Health care costs, utilization and patterns of care following Lyme Disease. *PLoS ONE*, **10**, e0116767.

Andreotti R, León AAP de, Dowd SE, Guerrero FD, Bendele KG, Scoles GA (2011) Assessment of bacterial diversity in the cattle tick *Rhipicephalus* (*Boophilus*) *microplus* through tag-encoded pyrosequencing. *BMC Microbiology*, **11**, 6.

Benson MJ, Gawronski JD, Eveleigh DE, Benson DR (2004) Intracellular symbionts and other bacteria associated with deer ticks (*Ixodes scapularis*) from Nantucket and Wellfleet, Cape Cod, Massachusetts. *Applied Environmental Microbiology*, **70**, 616–620.

Budachetri K, Browning RE, Adamson SW, Dowd SE, Chao C-C, Ching W-M, Karim S (2014) An insight into the microbiome of the *Amblyomma maculatum* (Acari: Ixodidae). *Journal of Medical Entomology*, **51**, 119–129.

Burgdorfer W, Hayes SF, Mavros AJ (1981) Nonpathogenic rickettsiae in *Dermacentor andersoni*: a limiting factor for the distribution of *Rickettsia rickettsii*. In: *Rickettsiae and Rickettsial Diseases* (eds Burgdorfer W, Anacker RL), pp. 585–594. Academic Press, Inc., New York.

Carpi G, Cagnacci F, Wittekindt NE, Zhao F, Qi J, Tomsho LP, Drautz DI, Rizzoli A, Schuster SC (2011) Metagenomic profile of the bacterial communities associated with *Ixodes ricinus* ticks. *PLoS ONE*, **6**, e25604.

Childs JE, Paddock CD (2003) The ascendancy of *Amblyomma americanum* as a vector of pathogens affecting humans in the United States. *Annual Review of Entomology*, **48**, 307–337.

Clay K, Klyachko O, Grindle N, Civitello D, Oleske D, Fuqua C (2008) Microbial communities and interactions in the lone star tick, *Amblyomma americanum*. *Molecular Ecology*, **17**, 4371–4381.

Commins SP, James HR, Kelly LA, Pochan SL, Workman LJ, Perzanowski MS, Kocan KM, Fahy JV, Nganga LW, Ronmark E, Cooper PJ, Platts-Mills TAE (2011) The relevance of tick bites to the production of IgE antibodies to the mammalian oligosaccharide galactose-α-1,3-galactose. *Journal of Allergy and Clinical Immunology*, **127**, 1286–1293.e6.

Cortinas R, Spomer S (2013) Lone star tick (Acari: Ixodidae) occurrence in Nebraska: historical and current perspectives. *Journal of Medical Entomology*, **50**, 244–251.

Douglas AE (2009) The microbial dimension in insect nutritional ecology. *Functional Ecology*, **23**, 38–47.

Dunn JM, Krause PJ, Davis S, Vannier EG, Fitzpatrick MC, Rollend L, Belperron AA, States SL, Stacey A, Bockenstedt LK, Fish D, Diuk-Wasser MA (2014) *Borrelia burgdorferi* promotes the establishment of *Babesia microti* in the Northeastern United States. *PLoS ONE*, **9**, e115494.

Eremeeva ME, Dasch GA (2015) Challenges posed by tick-borne rickettsiae: eco-epidemiology and public health implications. *Frontiers in Public Health*, **3**.

Estrada-Peña A, Jongejan F (1999) Ticks feeding on humans: a review of records on human-biting Ixodoidea with special reference to pathogen transmission. *Experimental & Applied Acarology*, **23**, 685–715.

Freitas LHT, Faccini JLH, Labruna MB (2008) Experimental infection of the rabbit tick, *Haemaphysalis leporispalustris*, with the bacterium *Rickettsia rickettsii*, and

comparative biology of infected and uninfected tick lineages. *Experimental and Applied Acarology*, **47**, 321–345.

Fritzen CM, Huang J, Westby K, Freye JD, Dunlap B, Yabsley MJ, Schardein M, Dunn JR, Jones TF, Moncayo AC (2011) Infection prevalences of common tick-borne pathogens in adult lone star ticks (*Amblyomma americanum*) and American dog ticks (*Dermacentor variabilis*) in Kentucky. *The American Journal of Tropical Medicine and Hygiene*, **85**, 718–723.

De la Fuente J, Blouin EF, Kocan KM (2003) Infection exclusion of the rickettsial pathogen *Anaplasma marginale* in the tick vector *Dermacentor variabilis*. *Clinical and Diagnostic Laboratory Immunology*, **10**, 182–184.

De la Fuente J, Garcia-Garcia JC, Blouin EF, Saliki JT, Kocan KM (2002) Infection of tick cells and bovine erythrocytes with one genotype of the intracellular ehrlichia *Anaplasma marginale* excludes infection with other genotypes. *Clinical and Diagnostic Laboratory Immunology*, **9**, 658–668.

Ginsberg HS, Ewing CP, O'Connell AF, Bosler EM, Daley JG, Sayre MW (1991) Increased population densities of *Amblyomma americanum* (Acari: Ixodidae) on Long Island, New York. *The Journal of Parasitology*, **77**, 493–495.

Guglielmone AA, Robbins RG, Apanaskevich DA, Petney TN, Estrada-Pena A, Horak IG (2014) *The Hard Ticks of the World (Acari: Ixodida: Ixodidae)*. Springer, London.

Hawlena H, Rynkiewicz E, Toh E, Alfred A, Durden LA, Hastriter MW, Nelson DE, Rong R, Munro D, Dong Q, Fuqua C, Clay K (2013) The arthropod, but not the

vertebrate host or its environment, dictates bacterial community composition of fleas and ticks. *The ISME Journal*, **7**, 221–223.

Herrmann C, Gern L (2010) Survival of *Ixodes ricinus* (Acari: Ixodidae) under challenging conditions of temperature and humidity is influenced by *Borrelia burgdorferi* sensu lato infection. *Journal of Medical Entomology*, **47**, 1196–1204.

Herrmann C, Voordouw MJ, Gern L (2013) *Ixodes ricinus* ticks infected with the causative agent of Lyme disease, *Borrelia burgdorferi* sensu lato, have higher energy reserves. *International Journal for Parasitology*, **43**, 477–483.

Jongejan F, Uilenberg G (2004) The global importance of ticks. *Parasitology*, **129**, S3–S14.

Kagemann J, Clay K (2013) Effects of infection by *Arsenophonus* and *Rickettsia* bacteria on the locomotive ability of the ticks *Amblyomma americanum*, *Dermacentor variabilis*, and *Ixodes scapularis*. *Journal of Medical Entomology*, **50**, 155–162.

Keirans JE, Lacombe EH (1998) First records of *Amblyomma americanum*, *Ixodes* (*Ixodes*) *dentatus*, and *Ixodes* (*Ceratixodes*) *uriae* (Acari: Ixodidae) from Maine. *The Journal of Parasitology*, **84**, 629–631.

Lalzar I, Harrus S, Mumcuoglu KY, Gottlieb Y (2012) Composition and seasonal variation of *Rhipicephalus turanicus* and *Rhipicephalus sanguineus* bacterial communities. *Applied and Environmental Microbiology*, **78**, 4110–6.

Lo N, Beninati T, Sassera D, Bouman EAP, Santagati S, Gern L, Sambri V, Masuzawa T, Gray JS, Jaenson TGT, Bouattour A, Kenny MJ, Guner ES, Kharitonenkov IG, Bitam I, Bandi C (2006) Widespread distribution and high prevalence of an alpha-

proteobacterial symbiont in the tick *Ixodes ricinus*. *Environmental Microbiology*, **8**, 1280–1287.

Loftis A, Mixson T, Stromdahl E, Yabsley M, Garrison L, Williamson P, Fitak R, Fuerst P, Kelly D, Blount K (2008) Geographic distribution and genetic diversity of the *Ehrlichia* sp. from Panola Mountain in *Amblyomma americanum*. *BMC Infectious Diseases*, **8**, 54.

Loftis AD, Reeves WK, Spurlock JP, Mahan SM, Troughton DR, Dasch GA, Levin ML (2006) Infection of a goat with a tick-transmitted *Ehrlichia* from Georgia, U.S.A., that is closely related to *Ehrlichia ruminantium*. *Journal of Vector Ecology*, **31**, 213–223.

Macaluso KR, Sonenshine DE, Ceraul SM, Azad AF (2002) Rickettsial infection in *Dermacentor variabilis* (Acari: Ixodidae) inhibits transovarial transmission of a second *Rickettsia*. *Journal of Medical Entomology*, **39**, 809–813.

Masters EJ, Grigery CN, Masters RW (2008) STARI, or Masters disease: lone star tick–vectored Lyme-like illness. *Infectious Disease Clinics of North America*, **22**, 361–376.

Means RG, White DJ (1997) New distribution records of *Amblyomma americanum* (L.) (Acari : Ixodidae) in New York State. *Journal of Vector Ecology*, **22**, 133–145.

Menchaca AC, Visi DK, Strey OF, Teel PD, Kalinowski K, Allen MS, Williamson PC (2013) Preliminary assessment of microbiome changes following blood-feeding and survivorship in the *Amblyomma americanum* nymph-to-adult transition using semiconductor sequencing. *PLoS ONE*, **8**, e67129.

Mixson TR, Campbell SR, Gill JS, Ginsberg HS, Reichard MV, Schulze TL, Dasch GA (2006) Prevalence of *Ehrlichia*, *Borrelia*, and rickettsial agents in *Amblyomma americanum* (Acari: Ixodidae) collected from nine states. *Journal of Medical Entomology*, **43**, 1261–1268.

Moran NA, McCutcheon JP, Nakabachi A (2008) Genomics and evolution of heritable bacterial symbionts. *Annual Review of Genetics*, **42**, 165–190.

Moreno CX, Moy F, Daniels TJ, Godfrey HP, Cabello FC (2006) Molecular analysis of microbial communities identified in different developmental stages of *Ixodes scapularis* ticks from Westchester and Dutchess Counties, New York. *Environmental Microbiology*, **8**, 761–772.

Nakao R, Abe T, Nijhof AM, Yamamoto S, Jongejan F, Ikemura T, Sugimoto C (2013) A novel approach, based on BLSOMs (Batch Learning Self-Organizing Maps), to the microbiome analysis of ticks. *The ISME Journal*.

Narasimhan S, Rajeevan N, Liu L, Zhao YO, Heisig J, Pan J, Eppler-Epstein R, DePonte K, Fish D, Fikrig E (2014) Gut microbiota of the tick vector *Ixodes scapularis* modulate colonization of the Lyme disease spirochete. *Cell Host & Microbe*, **15**, 58–71.

Niebylski ML, Peacock MG, Schwan TG (1999) Lethal effect of *Rickettsia rickettsii* on its tick vector (*Dermacentor andersoni*). *Applied and Environmental Microbiology*, **65**, 773–778.

Van Overbeek L, Gassner F, Plas CL van der, Kastelein P, Rocha UN, Takken W (2008) Diversity of *Ixodes ricinus* tick-associated bacterial communities from different forests. *FEMS Microbiology Ecology*, **66**, 72–84.

Pistone D, Sacchi L, Lo N, Epis S, Pajoro M, Favia G, Mandrioli M, Bandi C, Sassera D (2012) *Candidatus* Midichloria mitochondrii: symbiont or parasite of tick mitochondrii? In: *Manipulative Tenants: Bacteria Associated With Arthropods* Frontiers in Microbiology. (eds Zchori-Fein E, Bourtzis K), pp. 113–126. CRC Press, Boca Raton.

Ponnusamy L, Gonzalez A, Treuren WV, Weiss S, Parobek CM, Juliano JJ, Knight R, Roe RM, Apperson CS, Meshnick SR (2014) Diversity of Rickettsiales in the microbiome of the lone star tick, *Amblyomma americanum*. *Applied and Environmental Microbiology*, **80**, 354–359.

Randolph SE, Gern L, Nuttall PA (1996) Co-feeding ticks: epidemiological significance for tick-borne pathogen transmission. *Parasitology Today*, **12**, 472–479.

Savage HM, Godsey MS, Lambert A, Panella NA, Burkhalter KL, Harmon JR, Lash RR, Ashley DC, Nicholson WL (2013) First detection of Heartland virus (Bunyaviridae: Phlebovirus) from field collected arthropods. *The American Journal of Tropical Medicine and Hygiene*, **89**, 445–452.

Schabereiter-Gurtner C, Lubitz W, Rölleke S (2003) Application of broad-range 16S rRNA PCR amplification and DGGE fingerprinting for detection of tick-infecting bacteria. *Journal of Microbiological Methods*, **52**, 251–260.

Šmit R, Postma MJ (2014) Review of tick-borne encephalitis and vaccines: clinical and economical aspects. *Expert Review of Vaccines*, 1–11.

Sutáková G, Rehácek J (1990) Mixed infection of Rickettsiella phytoseiuli and Coxiella burnetii in Dermacentor reticulatus female ticks: electron microscope study. *Journal of Invertebrate Pathology*, **55**, 407–416.

Telford SR (2009) Status of the "East side hypothesis" (transovarial interference) twenty five years later. *Annals of the New York Academy of Sciences*, **1166**, 144–150.

Zhang X-C, Yang Z-N, Lu B, Ma X-F, Zhang C-X, Xu H-J (2014) The composition and transmission of microbiome in hard tick, *Ixodes persulcatus*, during blood meal. *Ticks and Tick-borne Diseases*, **5**, 864–870.

Zhong JM, Jasinskas A, Barbour AG (2007) Antibiotic treatment of the tick vector *Amblyomma americanum* reduced reproductive fitness. *PLoS One*, **2**, e405.

CHAPTER 2

Presence, genetic variability, and potential significance of "*Candidatus* Midichloria mitochondrii" in the lone star tick *Amblyomma americanum*

This chapter was originally published in Experimental and Applied Acarology and is in the public domain. See Williams-Newkirk AJ, Rowe LA, Mixson-Hayden TR, Dasch GA (2012) Presence, genetic variability, and potential significance of "*Candidatus* Midichloria mitochondrii" in the lone star tick *Amblyomma americanum*. Experimental and Applied Acarology, 58, 291–300.

2.1 Summary

I used Next Generation sequencing to detect the bacterium "*Candidatus* Midichloria mitochondrii" for the first time in lone star ticks (*Amblyomma americanum*) from the eastern United States. 177 individuals and 11 tick pools from seven sites in four states were tested by pyrosequencing with barcoded 16S rRNA gene eubacterial primers targeting variable regions 5-3. Average infection prevalence was 0.15 across all surveyed populations (range 0-0.29) and only the site with the smallest sample size (n=5) was negative. Three genotypes differing by 2.58-4.06% in a 271 bp region of 16S rRNA gene were identified. Two variants co-occurred in sites in North Carolina and New York, but were not observed in the same tick at those sites. The third genotype was found only in Georgia. Phylogenetic analysis of this fragment indicated that the three variants are more

closely related to "*Candidatus* Midichloria mitochondrii" genotypes from other tick species than to each other. This variation suggests that multiple independent introductions occurred in *A. americanum* which may provide insight into bacterial spread within its ecosystem and parasitism on this tick. Whether the presence of this bacterium affects acquisition or maintenance of other pathogens and symbionts in *A. americanum* or the survival, biology and evolution of the tick itself is unknown.

2.2 Introduction

"*Candidatus* Midichloria mitochondrii" (hereafter *M. mitochondrii*) is a newly described Rickettsiales from ixodid ticks that has an unusual tropism for mitochondria in some hosts (Beninati *et al.* 2004; Epis *et al.* 2008). In ticks where mitochondria infection has been observed, the bacterium replicates within the organelles and may cause extensive damage to them without observable impact on host cell function (Sacchi *et al.* 2004). Efficient vertical transmission of and high infection rates with *M. mitochondrii* (females, nymphs, and larvae 100%; males 44%) have been demonstrated in *Ixodes ricinus* (Lo *et al.* 2006; Sassera *et al.* 2008). The infection prevalence of *M. mitochondrii* remains imprecisely known in most other surveyed tick species (7-100%) due to small sample sizes (Epis *et al.* 2008; Venzal *et al.* 2008).

Neither the significance of *M. mitochondrii* as a tick symbiont nor its potential for vertebrate pathogenicity has been established, although evidence supporting its importance in both roles exists. Declining prevalence in laboratory-reared *I. ricinus* strongly suggests the bacterium is a facultative symbiont providing benefit to ticks under

some environmental conditions (Lo *et al.* 2006), but is too energetically costly to maintain once environmental selection pressure is removed (Pistone *et al.* 2012). Understanding its benefit to tick hosts may provide a novel method of control for ticks that vector disease.

Many Rickettsiales, including the *Ixodes persulcatus* tick-associated relative of *M. mitochondrii* known as the 'Montezuma agent' (Figure 2.1), are pathogenic for man (Mediannikov *et al.* 2004). *Midichloria mitochondrii* DNA has been detected in roe deer blood (Skarphédinsson *et al.* 2005) and tick salivary glands (Pistone *et al.* 2012), indicating vertebrates are, at minimum, exposed to the organism. Consequently, the potential of *M. mitochondrii* as a vertebrate pathogen cannot be ignored, although its active infection of a vertebrate has not yet been documented. The discordant phylogenies of tick and *M. mitochondrii* strains are most parsimoniously explained by horizontal transmission through a shared vertebrate host (Epis *et al.* 2008), thus also suggesting some vertebrates may be susceptible to infection. Considering that numerous tick species host this bacterium and its high abundance in the average infected individual tick (Sassera *et al.* 2008), it is important that the host range, role, and potential for pathogenicity of *M. mitochondrii* be defined.

Epis and colleagues recently used PCR to survey 21 tick species for the presence of *M. mitochondrii* (Epis *et al.* 2008). The lone star tick (*Amblyomma americanum*), the most abundant and aggressive ixodid tick in the southeastern United States and a vector of several known and potential human pathogens (Childs & Paddock 2003; Goddard & Varela-Stokes 2009), was reported as negative for *M. mitochondrii* based on a screen of 10 individual adults and 5 pools from Georgia. However, while conducting a

metagenomic survey of the bacterial community of *A. americanum* from four states in the United States, I detected *M. mitochondrii* DNA in multiple tick samples. Here I describe the distribution and prevalence of the bacterium in *A. americanum*, as well as its unusually high level of genetic variation compared to strains previously characterized from other ticks. I further consider its potential significance for the biology and control of this tick.

2.3 Materials and methods

Tick collection and DNA extraction

Host-seeking *A. americanum* were collected by dragging a 1 m^2 white cotton cloth through ground-level vegetation (Ginsberg & Ewing 1989). Collection sites, tick preservation, and DNA extraction are described in Mixson et al. (2006) for all samples collected in New York (1999, 2003), New Jersey (2003), and North Carolina (2002); DNAs were stored at 4°C until tested in this study. Ticks were collected in 2010 from Panola Mountain and Sweetwater Creek State Parks near Atlanta, Georgia. Most ticks collected in 2010 were preserved in 70% ethanol immediately after collection and stored at 4°C. A subset of adults were placed live into individual, sterile vials and washed with lysis buffer from the QIAamp DNA Mini Kit (Qiagen, Valencia, California) for 20 minutes to collect surface bacteria for metagenomic analysis. DNA was extracted from the lysis buffer rinse following the manufacturer's instructions. Following the surface wash, ticks were preserved in 70% ethanol. DNA was extracted from all ethanol

preserved nymphs and adults individually and pools of 50 larvae from single egg masses as described in Bermúdez et al. (2009).

16S rRNA gene pyrosequencing

PCR primer and barcode designs were from the Human Microbiome Project's Provisional 16S 454 Protocol (Broad primer sequences for variable regions 5->3) (HMP Website) and modified to be compatible with the Roche (Indianapolis, Indiana) 454 sequencer's Titanium chemistry. Modified primer sequences were F: 5'-CTA TGC GCC TTG CCA GCC CGC TCA GCC TAC GGG AGG CAG CAG-3' and R: 5'-CGT ATC GCC TCC CTC GCG CCA TCA G[barcode] CCC GTC AAT TCM TTT RAG T-3'. For PCR, each 20 μL reaction mix contained 1x AccuPrime PCR Buffer II (Invitrogen, Carlsbad, California, USA), 0.5 U AccuPrime Taq High Fidelity, 13.5 μL water, 0.3 μM each forward and reverse primer, and 2 μL DNA. All reactions were performed in an Eppendorf Master Gradient thermocycler (Brinkmann Instruments, Westbury, New York, USA) under the following program: one cycle of 94°C (2 min), 35 cycles of 94°C (30 s), 50°C (30 s), and 68°C (1 min), and one cycle of 68°C (5 min). Reaction products were analyzed by electrophoresis on 1% agarose gels stained with ethidium bromide to ensure reaction success, then quantitated using the Quant-iT PicoGreen dsDNA kit (Invitrogen) modified from the manufacturer's protocol to 50% of the recommended final assay volume. Amplicons were pooled in equal concentrations, purified using the Wizard SV Gel and PCR Clean-Up System (Promega, Madison, Wisconsin), then sequenced at the Centers for Disease Control and Prevention (CDC) Biotechnology Core Facility. Amplicons from 50 samples collected prior to 2010 were sequenced on one quarter of a 454 Titanium plate. Amplicons from 132 ticks collected in 2010 were sequenced on a

whole plate; 49 samples were amplified and sequenced twice. Amplicons from the external rinses of 6 adults were also sequenced on this plate.

Pyrosequence analysis

Raw sequences were processed to remove low quality reads using the software mothur v.1.22.0 (Schloss *et al.* 2009) following the trim.seqs alternative approach described in mothur's online documentation (Schloss 2011). The remaining trimmed, high quality reads were assigned to operational taxonomic units (OTUs) by clustering sequences at ≥97% identity, and OTUs were identified by comparison to reference databases. Because this report is primarily concerned with the presence or absence of *M. mitochondrii* within ticks, the number of sequences analyzed across samples was not standardized.

Phylogenetic trees

I retrieved all GenBank sequences that overlapped 100% and shared ≥94% BLAST identity with my 16S rDNA fragment variants in addition to representative sequences from other clades within the Rickettsiales. All sequences were aligned and trimmed to the same length in the software package Geneious (Drummond *et al.* 2010). Using the Tamura-Nei genetic distance model, a consensus neighbor-joining tree was constructed based on 1,000 bootstrap iterations with a 50% support threshold.

2.4 Results

M. mitochondrii *abundance and prevalence*

Not all 188 tick DNA samples generated a sufficient number of sequence reads to reliably detect low numbers of *M. mitochondrii*. Of the 50 DNA samples from New York, New Jersey, and North Carolina and the 138 samples from Georgia, only 35 and 96 DNA samples, respectively, produced ≥1000 high quality reads each. Across all samples surveyed, 27 individuals, four pools of 5 nymphs, and the exterior wash of one tick contained at least one *M. mitochondrii* sequence (mean=118 sequences/sample, range=1-555). The minimum infection rate was calculated for pooled nymphs, but each pool of larvae was considered a single sample because each pool originated from a single female. Total infection prevalence was 0.15 and within site prevalence ranged from 0-0.29 (Table 1). The only site with no evidence of infection was also the site with the smallest sample size (Fire Island, NY, n=5 nymphs, probability of detection 0.11).

M. mitochondrii *genetic variation*

Three distinct sequence variants of the 16S rRNA gene fragment were identified from the pyrosequencing data (Figure 2.2); these differed in pairwise comparisons by 2.58% (variant A-B), 3.69% (A-C), and 4.06% (B-C). Genotypes differed in their distribution between sites and samples. Variant A (GenBank accession number JQ678693) was found exclusively in ticks from Georgia and was the sole genotype found there. Variants B (JQ678691) and C (JQ678692) were identified individually in North Carolina and New Jersey sites, respectively, and also together in North Carolina and New York sites (Table 2). The only DNA that contained multiple variants was extracted from a pool of 5

nymphs from Shelter Island, New York, where both genotypes B and C were also frequently observed in individual ticks (variant B: 1 male and 3 nymph pools; variant C: 2 females, 3 males, and 2 nymph pools).

Phylogenetic analysis

The phylogenetic tree based on the 271 bp 16S rRNA gene sequence fragments (Figure 2.1) was highly similar in its organization within the *M. mitochondrii* clade to published trees based on 1000 bp to near full length 16S rRNA gene fragments (Epis *et al.* 2008; Vannini *et al.* 2010). However, unlike other *M. mitochondrii* sequences obtained from the same tick species, the *A. americanum*-derived variants were more closely related to those from other ticks than to each other. The position of several other *M. mitochondrii*-like organisms and some related clades did not fully resolve with the amount of sequence data available.

2.5 Discussion

Midichloria mitochondrii has been previously reported in 11 ixodid tick species from Europe, Asia, North America, South America, and Iceland (Parola *et al.* 2003; Epis *et al.* 2008; Venzal *et al.* 2008; Beninati *et al.* 2009; Dergousoff & Chilton 2011; Pistone *et al.* 2012) and from horse flies (*Tabanus bovinus* and *Tabanus tergestinus*) and bed bugs (*Cimex lectularius*) collected in Europe (Hornok *et al.* 2008; Richard *et al.* 2009). The wide distribution of this bacterium in haematophagous arthropods and the lack of concordance between the phylogenies of *M. mitochondrii* and its tick hosts strongly suggest that the bacterium is transmitted horizontally between tick species or into tick

species from some other host(s) (Epis *et al.* 2008). The data reported here provide further support for this hypothesis by identifying three distinct genotypes of *M. mitochondrii* in *A. americanum*. While the full 16S rRNA gene sequence is not yet available for these new variants, previous work on *M. mitochondrii* within a single tick species has shown variation in this gene to be limited or absent in all ticks (Pistone *et al.* 2012) except *I. holocyclus* where two strains are known that differ by 2.5% (Lo *et al.* 2006; Beninati *et al.* 2009).

The structure of the *M. mitochondrii* clade in the phylogenetic tree shown in Figure 2.1 is very similar to the trees based on the longer 16S rRNA gene sequences published previously (Epis *et al.* 2008; Vannini *et al.* 2010). The most notable characteristic of the tree was the distribution of *A. americanum*-derived genotypes throughout the *M. mitochondrii* clade, rather than clustering together as is seen in other tick species. This suggests that *M. mitochondrii* may have been introduced into *A. americanum* multiple times from different vertebrate or invertebrate hosts. However, of the three tick species hosting closely related strains, only the geographic range of *Amblyomma tuberculatum* overlaps with that of *A. americanum*, indicating that additional undiscovered hosts of *M. mitochondrii* likely exist in North America.

As yet, there is no evidence that individual ticks were co-infected by multiple variants despite indications that infection prevalence may be relatively high in populations hosting multiple strains. Several mechanisms may give rise to this pattern. One possibility is that, like in *I. ricinus* (Sassera *et al.* 2008), the dominant mode of transmission for *M. mitochondrii* in *A. americanum* may be vertically from female to eggs, creating very little opportunity for the acquisition of multiple infections by superinfection. Additional

mechanisms to prevent or reduce the incidence of co-infections in ticks may be similar to those seen in other arthropod-bacterial symbiont systems. Examples include, but are not limited to, a bacterial population bottleneck during the colonization of host oocytes (Mira & Moran 2002), cytoplasmic incompatibility between hosts infected with different bacterial strains (Stouthamer *et al.* 1999), and competitive exclusion of additional strains by previous host infection (Macaluso *et al.* 2002). Additional work is needed to confirm the pattern of single genotype infections within individual ticks and to determine its significance either at the population level of *A. americanum* or by its mode of acquisition and maintenance.

The detection of *M. mitochondrii* DNA in the external rinse of a single male was interesting, but also puzzling because the male's corpuscular DNA sample did not test positive. Nearly three times the number of sequence reads were obtained from the male's rinse compared to its residual DNA sample. This suggests that the buffers used to rinse the tick may have extracted large amounts of DNA from the body cavity of the tick. Further work is needed to determine if *M. mitochondrii* is a component of the external bacterial community of some *A. americanum*.

The relationship dynamics between *M. mitochondrii* and its tick hosts are still under investigation. Other bacteria have been shown to provide a competitive advantage to ticks in specific environmental conditions. For example, *Borrelia burgdorferi* increases the tolerance of *I. ricinus* for hot, dry habitats (Herrmann & Gern 2010). The fluctuating infection prevalence of *M. mitochondrii* in *I. ricinus* (Lo *et al.* 2006; Sassera *et al.* 2008) and its frequent discovery in other tick species may indicate a similarly beneficial association, possibly involving a role in energy metabolism given the preference of *M.*

mitochondrii for tick cell mitochondria in some hosts (Pistone *et al.* 2012). This is especially interesting considering the ongoing expansion of *A. americanum* out of its historical range in the southeastern United States and into the northeastern and midwestern states (Ginsberg *et al.* 1991; Means & White 1997; Keirans & Lacombe 1998; Merten & Durden 2000). *Amblyomma americanum* is the most frequently reported tick attached to humans in the southeast and Atlantic states (Merten & Durden 2000), and although once considered merely a nuisance species, the importance of *A. americanum* as a disease vector in the United States is now well recognized (Childs & Paddock 2003). It is important that future research addresses the question of *M. mitochondrii*'s potential contribution to the range expansion of *A. americanum* so that manipulation of the relationship may be explored as a method of control for this important vector. Genetic modification of this agent to suppress acquisition of pathogenic agents vectored by *A. americanum* may be possible if one or more *M. mitochondrii* variants are stably inherited both transovarially and transstadially, similar to the introduction of *Wolbachia* into *Aedes aegypti* to suppress arbovirus and protozoa transmission (Moreira *et al.* 2009).

This is the first report of *M. mitochondrii* in *A. americanum*. Estimates of prevalence were highly variable between sites, an unsurprising finding given the small within-site sample sizes. With my limited sample sizes and number of sites tested, the true prevalence, genetic diversity and distribution of genotypes of *M. mitochondrii* in *A. americanum* is likely only roughly approximated. However, my results have demonstrated that *M. mitochondrii* is widely distributed, genetically diverse, and common in *A. americanum*.

2.6 Figures

Figure 2.1 Phylogenetic tree of *Midichloria mitochondrii* 16S rRNA gene fragments obtained from *Amblyomma americanum* and related sequences from GenBank. Bold type is used to highlight the position of the *A. americanum* variants. The tree was constructed from a 271 bp fragment using neighbor-joining with *Rhizobium leguminosarum* as the out group. Unless noted, all terminal node labels indicate the tick host from which the *M. mitochondrii* sequence was obtained. GenBank accession numbers are provided in parentheses. Branch support is given for bootstraps greater than 75%. Scale bar indicates the number of substitution per site.

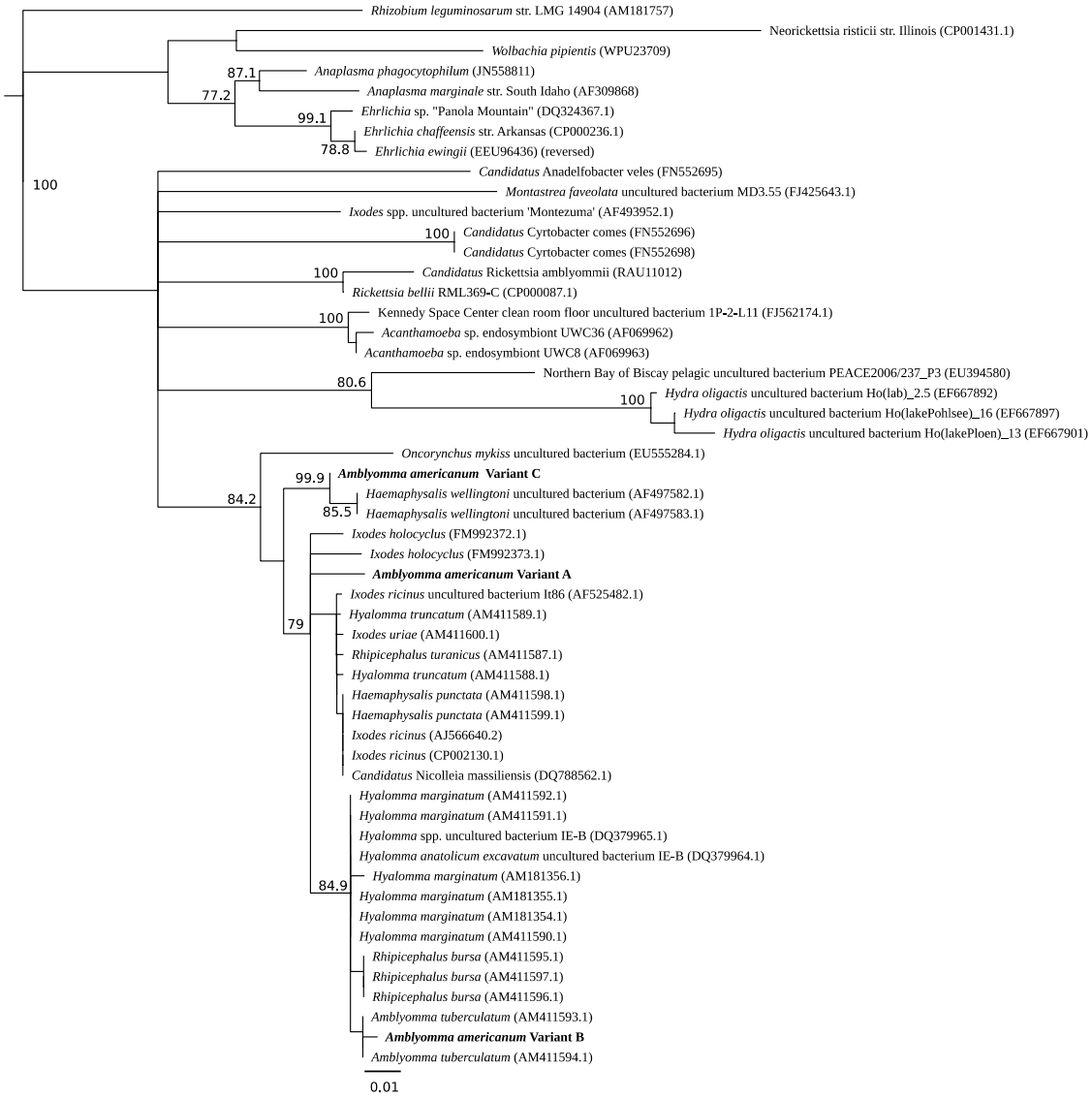
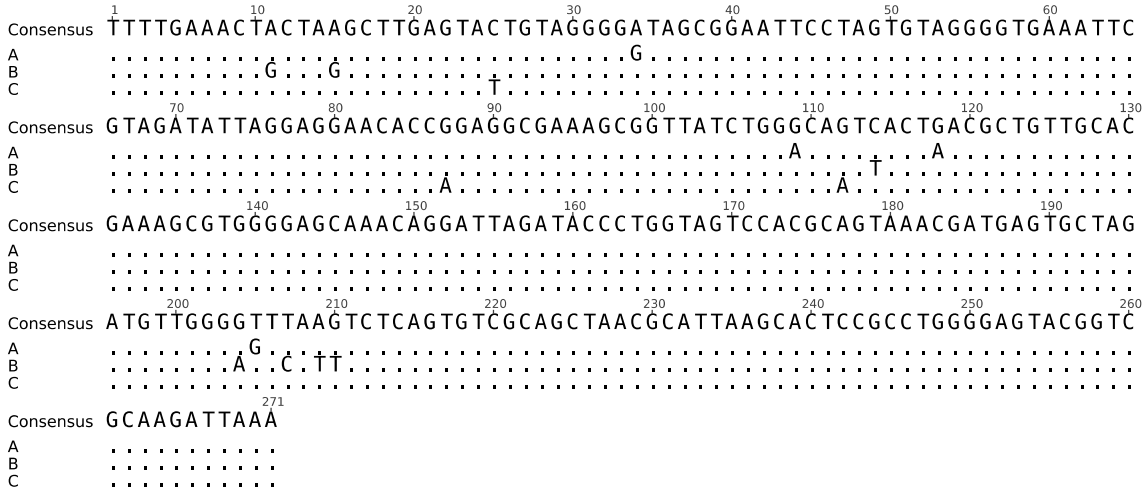

Rhizobium leguminosarum str. LMG 14904 (AM181757)
Neorickettsia risticii str. Illinois (CP001431.1)
Wolbachia pipientis (WPU23709)
87.1
Anaplasma phagocytophilum (JN558811)
77.2
Anaplasma marginale str. South Idaho (AF309868)
99.1
Ehrlichia sp. "Panola Mountain" (DQ324367.1)
Ehrlichia chaffeensis str. Arkansas (CP000236.1)
78.8
Ehrlichia ewingii (EEU96436) (reversed)
100
Candidatus Anadelfobacter veles (FN552695)
Montastrea faveolata uncultured bacterium MD3.55 (FJ425643.1)
Ixodes spp. uncultured bacterium 'Montezuma' (AF493952.1)
100
Candidatus Cyrtobacter comes (FN552696)
Candidatus Cyrtobacter comes (FN552698)
100
Candidatus Rickettsia amblyommii (RAU11012)
Rickettsia bellii RML369-C (CP000087.1)
100
Kennedy Space Center clean room floor uncultured bacterium 1P-2-L11 (FJ562174.1)
Acanthamoeba sp. endosymbiont UWC36 (AF069962)
Acanthamoeba sp. endosymbiont UWC8 (AF069963)
Northern Bay of Biscay pelagic uncultured bacterium PEACE2006/237_P3 (EU394580)
80.6
100
Hydra oligactis uncultured bacterium Ho(lab)_2.5 (EF667892)
Hydra oligactis uncultured bacterium Ho(lakePohlsee)_16 (EF667897)
Hydra oligactis uncultured bacterium Ho(lakePloen)_13 (EF667901)
Oncorynchus mykiss uncultured bacterium (EU555284.1)
99.9
***Amblyomma americanum* Variant C**
84.2
Haemaphysalis wellingtoni uncultured bacterium (AF497582.1)
85.5
Haemaphysalis wellingtoni uncultured bacterium (AF497583.1)
Ixodes holocyclus (FM992372.1)
Ixodes holocyclus (FM992373.1)
***Amblyomma americanum* Variant A**
Ixodes ricinus uncultured bacterium It86 (AF525482.1)
Hyalomma truncatum (AM411589.1)
79
Ixodes uriae (AM411600.1)
Rhipicephalus turanicus (AM411587.1)
Hyalomma truncatum (AM411588.1)
Haemaphysalis punctata (AM411598.1)
Haemaphysalis punctata (AM411599.1)
Ixodes ricinus (AJ566640.2)
Ixodes ricinus (CP002130.1)
Candidatus Nicolleia massiliensis (DQ788562.1)
Hyalomma marginatum (AM411592.1)
Hyalomma marginatum (AM411591.1)
Hyalomma spp. uncultured bacterium IE-B (DQ379965.1)
Hyalomma anatolicum excavatum uncultured bacterium IE-B (DQ379964.1)
Hyalomma marginatum (AM181356.1)
84.9
Hyalomma marginatum (AM181355.1)
Hyalomma marginatum (AM181354.1)
Hyalomma marginatum (AM411590.1)
Rhipicephalus bursa (AM411595.1)
Rhipicephalus bursa (AM411597.1)
Rhipicephalus bursa (AM411596.1)
Amblyomma tuberculatum (AM411593.1)
***Amblyomma americanum* Variant B**
Amblyomma tuberculatum (AM411594.1)
0.01

Figure 2.2 Three variants of the *Midichloria mitochondrii* 16S rRNA gene fragment identified in *Amblyomma americanum*. The first row of each section represents the consensus sequence, and the numbers above the consensus indicate the nucleotide position. Variants are labeled A, B, and C with their differences from the consensus indicated. Total fragment length is 271 bp and spans a portion of variable region 4 and all of variable region 5.



2.7 Tables

Table 2.1 Prevalence of *Midichloria mitochondrii* infection in *Amblyomma americanum* collected from the eastern United States.

Site	Pooled Larvae	Nymph	Pooled Nymphs	Male	Female	Adult Rinse	Site Prevalence
Bodie Island, NC	NT	NT	NT	1/5	1/5	NT	0.20 (n=10)
Buxton Woods, NC	NT	NT	NT	0/5	2/5	NT	0.20 (n=10)
Earle, NJ	NT	NT	NT	2/5	0/5	NT	0.20 (n=10)
Fire Island, NY	NT	0/5	NT	NT	NT	NT	0 (n=5)
Shelter Island, NY	NT	NT	4/5	4/5	2/5	NT	0.29 (n=35)
Panola Mountain, GA	0/3	6/60	NT	1/3	0/3	1/6	0.11 (n=75)
Sweetwater Creek, GA	0/3	8/60	NT	NT	NT	NT	0.13 (n=63)
Life Stage Prevalence	0 (n=6)	0.11 (n=125)	0.16 (n=25)	0.35 (n=23)	0.22 (n=23)	0.17 (n=6)	

Minimum infection rate is given for pooled nymphs. Fraction indicates the number positive / number sampled. Numbers in parentheses are the sample size for a given site or life stage. NT = not tested.

Table 2.2 Distribution of *Midichloria mitochondrii* 16S rRNA gene variants at *Amblyomma americanum* sample sites.

	16S rDNA Variant			
Site	A	B	C	Total No. Samples
Bodie Island, NC	0	2	0	2
Buxton Woods, NC	0	1	1	2
Earle, NJ	0	0	2	2
Fire Island, NY	0	0	0	0
Shelter Island, NY	0	4	7	10
Panola MTN, GA	8	0	0	8
Sweetwater Creek, GA	8	0	0	8

Values indicate the number of samples where a variant was detected. Note that one pool of 5 nymphs from Shelter Island, NY contained large numbers of both B and C sequences, so the pool was counted in both columns.

2.8 References

Beninati T, Lo N, Sacchi L, Genchi C, Noda H, Bandi C (2004) A novel alpha-proteobacterium resides in the mitochondria of ovarian cells of the tick *Ixodes ricinus*. *Applied and Environmental Microbiology*, **70**, 2596–2602.

Beninati T, Riegler M, Vilcins I-ME, Sacchi L, McFadyen R, Krockenberger M, Bandi C, O'Neill SL, Lo N (2009) Absence of the symbiont *Candidatus* Midichloria mitochondrii in the mitochondria of the tick *Ixodes holocyclus*. *FEMS Microbiology Letters*, **299**, 241–247.

Bermúdez SE, Eremeeva ME, Karpathy SE, Samudio F, Zambrano ML, Zaldivar Y, Motta JA, Dasch GA (2009) Detection and identification of rickettsial agents in ticks from domestic mammals in eastern Panama. *Journal of Medical Entomology*, **46**, 856–861.

Childs JE, Paddock CD (2003) The ascendancy of *Amblyomma americanum* as a vector of pathogens affecting humans in the United States. *Annual Review of Entomology*, **48**, 307–337.

Dergousoff SJ, Chilton NB (2011) Novel genotypes of *Anaplasma bovis*, "*Candidatus* Midichloria" sp. and *Ignatzschineria* sp. in the Rocky Mountain wood tick, *Dermacentor andersoni*. *Veterinary Microbiology*, **150**, 100–106.

Drummond AJ, Ashton B, Buxton S, Cheung M, Cooper A, Duran C, Field M, Heled J, Kearse M, Markowitz S, Moir R, Stones-Havas S, Sturrock S, Thierer T, Wilson A (2010) *Geneious*.

Epis S, Sassera D, Beninati T, Lo N, Beati L, Piesman J, Rinaldi L, McCoy KD, Torina A, Sacchi L, Clementi E, Genchi M, Magnino S, Bandi C (2008) *Midichloria*

mitochondrii is widespread in hard ticks (Ixodidae) and resides in the mitochondria of phylogenetically diverse species. *Parasitology*, **135**, 485–494.

Ginsberg HS, Ewing CP (1989) Comparison of flagging, walking, trapping, and collecting from hosts as sampling methods for northern deer ticks, *Ixodes dammini*, and lone-star ticks, *Amblyomma americanum* (Acari: Ixodidae). *Experimental and Applied Acarology*, **7**, 313–322.

Ginsberg HS, Ewing CP, O'Connell AF, Bosler EM, Daley JG, Sayre MW (1991) Increased population densities of *Amblyomma americanum* (Acari: Ixodidae) on Long Island, New York. *The Journal of Parasitology*, **77**, 493–495.

Goddard J, Varela-Stokes AS (2009) Role of the lone star tick, *Amblyomma americanum* (L.), in human and animal diseases. *Veterinary Parasitology*, **160**, 1–12.

Herrmann C, Gern L (2010) Survival of *Ixodes ricinus* (Acari: Ixodidae) under challenging conditions of temperature and humidity is influenced by *Borrelia burgdorferi* sensu lato infection. *Journal of Medical Entomology*, **47**, 1196–1204.

[HMP website] The Human Microbiome Project Generalized Draft form of HMP Data Generation Working Group 16S 454 Default Protocol - Version 4.2—pilot study P.1 (15). Available at: www.hmpdacc.org/doc/HMP_MDG_454_ 16S_Protocol_V4_2_102109.pdf. Accessed Sept 2010.

Hornok S, Földvári G, Elek V, Naranjo V, Farkas R, de la Fuente J (2008) Molecular identification of *Anaplasma marginale* and rickettsial endosymbionts in blood-sucking flies (Diptera: Tabanidae, Muscidae) and hard ticks (Acari: Ixodidae). *Veterinary Parasitology*, **154**, 354–359.

Keirans JE, Lacombe EH (1998) First records of *Amblyomma americanum*, *Ixodes* (*Ixodes*) *dentatus*, and *Ixodes* (*Ceratixodes*) *uriae* (Acari: Ixodidae) from Maine. *The Journal of Parasitology*, **84**, 629–631.

Lo N, Beninati T, Sacchi L, Bandi C (2006) An alpha-proteobacterium invades the mitochondria of the tick *Ixodes ricinus*. In: *Insect Symbiosis, Volume 2* Contemporary Topics in Entomology. (eds Bourtzis K, Miller TA), pp. 25–37. CRC Press.

Macaluso KR, Sonenshine DE, Ceraul SM, Azad AF (2002) Rickettsial infection in *Dermacentor variabilis* (Acari: Ixodidae) inhibits transovarial transmission of a second *Rickettsia*. *Journal of Medical Entomology*, **39**, 809–813.

Means RG, White DJ (1997) New distribution records of *Amblyomma americanum* (L.) (Acari : Ixodidae) in New York State. *Journal of Vector Ecology*, **22**, 133–145.

Mediannikov OI, Ivanov LI, Nishikawa M, Saito R, Sidel'nikov IN, Zdanovskaia NI, Mokretsova EV, Tarasevich IV, Suzuki H (2004) Microorganism "Montezuma" of the order Rickettsiales: the potential causative agent of tick-borne disease in the Far East of Russia. *Zhurnal Mikrobiologii, Epidemiologii, I Immunobiologii*, **1**, 7–13.

Merten HA, Durden LA (2000) A state-by-state survey of ticks recorded from humans in the United States. *Journal of Vector Ecology*, **25**, 102–113.

Mira A, Moran NA (2002) Estimating population size and transmission bottlenecks in maternally transmitted endosymbiotic bacteria. *Microbial Ecology*, **44**, 137–143.

Mixson TR, Campbell SR, Gill JS, Ginsberg HS, Reichard MV, Schulze TL, Dasch GA (2006) Prevalence of *Ehrlichia*, *Borrelia*, and rickettsial agents in *Amblyomma*

americanum (Acari: Ixodidae) collected from nine states. *Journal of Medical Entomology*, **43**, 1261–1268.

Moreira LA, Iturbe-Ormaetxe I, Jeffery JA, Lu G, Pyke AT, Hedges LM, Rocha BC, Hall-Mendelin S, Day A, Riegler M, Hugo LE, Johnson KN, Kay BH, McGraw EA, van den Hurk AF, Ryan PA, O'Neill SL (2009) A *Wolbachia* symbiont in *Aedes aegypti* limits infection with Dengue, Chikungunya, and Plasmodium. *Cell*, **139**, 1268–1278.

Parola P, Cornet J-P, Sanogo YO, Miller RS, Thien HV, Gonzalez J-P, Raoult D, Telford III SR, Wongsrichanalai C (2003) Detection of *Ehrlichia* spp., *Anaplasma* spp., *Rickettsia* spp., and other eubacteria in ticks from the Thai-Myanmar border and Vietnam. *Journal of Clinical Microbiology*, **41**, 1600–1608.

Pistone D, Sacchi L, Lo N, Epis S, Pajoro M, Favia G, Mandrioli M, Bandi C, Sassera D (2012) *Candidatus* Midichloria mitochondrii: symbiont or parasite of tick mitochondrii? In: *Manipulative Tenants: Bacteria Associated With Arthropods* Frontiers in Microbiology. (eds Zchori-Fein E, Bourtzis K), pp. 113–126. CRC Press, Boca Raton.

Richard S, Seng P, Parola P, Raoult D, Davoust B, Brouqui P (2009) Detection of a new bacterium related to "*Candidatus* Midichloria mitochondrii" in bed bugs. *Clinical Microbiology and Infection*, **15**, 84–85.

Sacchi L, Bigliardi E, Corona S, Beninati T, Lo N, Franceschi A (2004) A symbiont of the tick *Ixodes ricinus* invades and consumes mitochondria in a mode similar to that of the parasitic bacterium *Bdellovibrio bacteriovorus*. *Tissue and Cell*, **36**, 43–53.

Sassera D, Lo N, Bouman EAP, Epis S, Mortarino M, Bandi C (2008) "*Candidatus* Midichloria" endosymbionts bloom after the blood meal of the host, the hard tick *Ixodes ricinus*. *Applied and Environmental Microbiology*, **74**, 6138 –6140.

Schloss PD (2011) Schloss SOP. *Mothur Wiki*. Available at: www.mothur.org/wiki/Schloss_SOP. Accessed Nov 2011.

Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ, Weber CF (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology*, **75**, 7537–7541.

Skarphédinsson S, Jensen PM, Kristiansen K (2005) Survey of tickborne infections in Denmark. *Emerging Infectious Diseases*, **11**, 1055–1061.

Stouthamer R, Breeuwer JAJ, Hurst GDD (1999) *Wolbachia pipientis*: microbial manipulator of arthropod reproduction. *Annual Review of Microbiology*, **53**, 71–102.

Vannini C, Ferrantini F, Schleifer K-H, Ludwig W, Verni F, Petroni G (2010) "*Candidatus* Anadelfobacter veles" and "*Candidatus* Cyrtobacter comes", two new Rickettsiales hosted by the protist ciliate *Euplotes harpa* (Ciliophora, Spirotrichea). *Applied and Environmental Microbiology*, **76**, 4047–4054.

Venzal J, Estrada-Peña A, Portillo A, Mangold A, Castro O, de Souza C, Félix M, Pérez-Martínez L, Santibánez S, Oteo J (2008) Detection of alpha and gamma-proteobacteria in *Amblyomma triste* (Acari: Ixodidae) from Uruguay. *Experimental and Applied Acarology*, **44**, 49–56.

CHAPTER 3

Characterization of the bacterial communities of life stages of free living lone star ticks (*Amblyomma americanum*)

This chapter was originally published in PLoS ONE and is in the public domain. See Williams-Newkirk AJ, Rowe LA, Mixson-Hayden TR, Dasch GA (2014) Characterization of the bacterial communities of life stages of free living lone star ticks (*Amblyomma americanum*). PLoS ONE, 9, e102130.

3.1 Summary

The lone star tick (*Amblyomma americanum*) is an abundant and aggressive biter of humans, domestic animals, and wildlife in the southeastern-central USA and an important vector of several known and suspected zoonotic bacterial pathogens. However, the biological drivers of bacterial community variation in this tick are still poorly defined. Knowing the community context in which tick-borne bacterial pathogens exist and evolve is required to fully understand the ecology and immunobiology of the ticks and to design effective public health and veterinary interventions. I performed a metagenomic survey of the bacterial communities of questing *A. americanum* and tested 131 individuals (66 nymphs, 24 males, and 41 females) from five sites in three states. Pyrosequencing was performed with barcoded eubacterial primers targeting variable 16S rRNA gene regions 5-3. The bacterial communities were dominated by *Rickettsia* (likely

R. amblyommii) and an obligate *Coxiella* symbiont, together accounting for 6.7-100% of sequences per tick. DNAs from *Midichloria*, *Borrelia*, *Wolbachia*, *Ehrlichia*, *Pseudomonas*, or unidentified Bacillales, Enterobacteriaceae, or Rhizobiales groups were also detected frequently. *Wolbachia* and *Midichloria* significantly co-occurred in Georgia ($p<0.00001$), but not in other states. The significance of the *Midichloria*-*Wolbachia* co-occurrence is unknown. Among ticks collected in Georgia, nymphs differed from adults in both the composition ($p=0.002$) and structure ($p=0.002$) of their bacterial communities. Adults differed only in their community structure ($p=0.002$) with males containing more *Rickettsia* and females containing more *Coxiella*. Comparisons among adult ticks collected in New York and North Carolina supported the findings from the Georgia collection despite differences in geography, collection date, and sample handling, implying that the differences detected are consistent attributes. The data also suggest that some members of the bacterial community change during the tick life cycle and that some sex-specific attributes may be detectable in nymphs.

3.2 Introduction

Ticks transmit a greater diversity of pathogens to humans and domestic animals than any other vector group (Jongejan & Uilenberg 2004), and the lone star tick (*Amblyomma americanum*) is the most common human-biting tick in the southeastern United States (Merten & Durden 2000). During the last century *A. americanum* has expanded its range into the northeastern and mid-western states (Ginsberg *et al.* 1991; Means & White 1997;

Keirans & Lacombe 1998), further increasing the number of people and domestic animals exposed to the tick and the pathogens it transmits.

Amblyomma americanum is a three-host tick that is nonspecific in its host use in the immature (larval and nymphal) life stages and has a preference for white-tailed deer (*Odocoileus virginianus*) as adults (reviewed in Childs and Paddock (2003)). As many pathogens are acquired by ticks from their vertebrate hosts, the lack of host specificity throughout much of its life perfectly positions this tick as a vector of multiple zoonotic diseases, including ehrlichioses (Buller *et al.* 1999), rickettsioses, tularemia, and perhaps even Southern Tick-Associated Rash Illness (reviewed in Childs and Paddock (2003)). Other bacteria of undefined pathogenicity (*Candidatus* "Midichloria mitochondrii" (Williams-Newkirk *et al.* 2012; Bazzocchi *et al.* 2013), *Borrelia lonestari* (James *et al.* 2001)) and several viruses (Lone Star virus (Kokernot *et al.* 1969), Heartland virus (Savage *et al.* 2013)) are also associated with this species. Larvae, nymphs, and adult females feed once from a single host before becoming quiescent in the leaf litter and either molting into the next life stage (immatures) or ovipositing (adult females). Adult males take multiple smaller blood meals while seeking feeding females (Gladney & Drummond 1970).

Tick-borne disease ecology motivates many tick microbiological studies, which frequently use specific assays to survey for the presence of known vertebrate pathogens. Yet many vertebrate pathogens are neither prevalent in the vector population (*e.g.* Mixson *et al.* 2006) nor abundant within pathogen-infected ticks (*e.g.* Corrigan 2012) and may represent only a minority population within the tick microbiome. This suggests that the interactions between vertebrate pathogens and other more common and abundant

bacteria found in ticks may be more important to the abundance and distribution of pathogen-infected ticks in the environment. Several previous surveys of the bacterial communities of *A. americanum* have provided useful inventories of community members. Sanger sequencing by Clay *et al.* (2008) and Heise *et al.* (2010) and pyrosequencing by Ponnusamy *et al.* (2014) and Yuan (2010) of bacterial 16S rRNA gene fragments from *A. americanum* confirmed the common presence of a *Coxiella* that is a likely obligate symbiont (Jasinskas *et al.* 2007; Zhong *et al.* 2007). They also detected the low pathogenicity spotted fever group *Rickettsia Candidatus* "Rickettsia amblyommii" (hereafter *R. amblyommii*) as an abundant and common bacterial community member. Additional *Rickettsia* species (Heise *et al.* 2010) and a novel *Arsenophonus* were also identified (Clay *et al.* 2008; Yuan 2010), as well as a number of diverse taxa frequently found both environmentally and as animal associates, most commonly *Pseudomonas* (Clay *et al.* 2008; Heise *et al.* 2010; Yuan 2010), Enterobacteracea (Clay *et al.* 2008; Heise *et al.* 2010), and Bacillaceae (Heise *et al.* 2010; Yuan 2010).

While the *A. americanum* bacterial community has been well documented, the nature and basis for intraspecific differences are less well explored. Menchaca *et al.* (2013) used semi-conductor sequencing to describe changes in the bacterial communities of laboratory colony derived nymphs and adults to test the effects of engorgement, molting, age, and environmental conditions on bacterial communities. They found an interesting trend towards reduction in community diversity with tick age and environmental stressors under extremely controlled conditions. When pyrosequencing bacterial 16S amplicons from a small sample of 12 *A. americanum*, Ponnusamy *et al.*(2014) found more unique bacterial taxa in males and nymphs than in females. In other ixodid tick species, tick life

stage and sex have been used to explain variation in whole bacterial communities (Moreno *et al.* 2006; Hawlena *et al.* 2013) and focal bacterial taxa (Lo *et al.* 2006; Halos *et al.* 2010; Lalzar *et al.* 2012), although the mechanisms behind these differences remain unclear.

Here I characterize the life stage and sex specific bacterial communities of wild caught *A. americanum* collected at five sites in three states using pyrosequencing of variable regions 5-3 of the bacterial *rrs* (16S rRNA) gene. Samples of nymphs and adults from two sites in Georgia were compared to detect differences in community composition and structure between life stages and sexes from a single geographic area. Archived DNAs from adult *A. americanum* were also compared to see if sex-specific bacterial communities were a general phenomenon in this species or a region-specific occurrence. The geographical distribution of my sample sites in combination with my larger sample sizes enabled the identification of significant life stage and sex specific patterns of bacterial community composition and structure that have not been reported from previous studies.

3.3 Materials and methods

Ethics statement

Tick collection in Georgia was performed in state parks with permission from the Georgia Department of Natural Resources (permits #29-WBH-10-135 and #29-WBH-11-49) and park management. Ticks from other states were collected as described in Mixson *et al.* (2006).

Tick collection and DNA extraction

All *A. americanum* ticks were collected by running a 1 m^2 flannel cloth over vegetation. The primary tick DNA sample set (hereafter referred to as the Georgia collection) was collected in 2010-2011 from two ecologically similar forested state parks very near to Atlanta, Georgia (Table 3.1) and stored in 70% ethanol at 4°C. DNA extraction was performed as described in Bermúdez *et al.* (2009) with the Promega (Madison, WI, USA) Wizard SV 96 Genomic DNA Purification System. Ticks were treated with sequential pre-extraction surface washes of 10% bleach, 70% ethanol, and distilled water to reduce surface contamination. DNAs were stored at 4°C until used. Tick DNAs from my archives (hereafter referred to as the archival collection) were obtained from ticks from three sites on barrier islands in New York (1) and North Carolina (2) in 2002-2003 (Table 3.1), stored whole at -20°C, and then individually processed for DNA extraction using the QIAamp Mini Kit (Qiagen, Hilden, Germany) as described in Mixson *et al.* (2006). These tick samples were not pre-treated prior to DNA extraction to reduce surface contaminants as they were collected and processed under older protocols for a different purpose than this experiment. DNA extraction was performed during 2002-2004, and the DNAs were stored at 4°C until used in this study. DNAs extracted following this manufacturer's protocols are stable at least 16 years at 4°C (Hartmann *et al.* 2012).

454 library preparation and sequencing

PCR primer and barcode designs were obtained from the Human Microbiome Project's 16S 454 Sequencing Protocol. The primer sequences included eubacterial *rrs* primers 357F and 926R, which target variable regions 5-3 (HMP 2010). Five prime modifications

of primers were made for compatibility with the Roche (Indianapolis, Indiana, USA) 454 sequencer's Titanium chemistry. Modified primer sequences were F: 5'-CTA TGC GCC TTG CCA GCC CGC TCA GCC TAC GGG AGG CAG CAG-3' and R: 5'-CGT ATC GCC TCC CTC GCG CCA TCA G (barcode) CCC GTC AAT TCM TTT RAG T-3'. Barcode sequences are provided in the supporting information (Table 3.S1). The expected amplicon size was approximately 643 bp.

For PCR, each 20 μL reaction mix contained either 1x AccuPrime PCR Buffer II (Invitrogen, Carlsbad, California, USA), 0.5 U AccuPrime Taq High Fidelity, 13.5 μL water, 0.3 μM each forward and reverse primer, and 2 μL DNA or 15 μL Platinum PCR SuperMix High Fidelity (Invitrogen), and 1 μL water. I have found no significant differences in bacterial community composition when using these two enzymes for primary PCR (data comparing tick DNAs amplified with both polymerases are described further below and in the results). All reactions were performed in an Eppendorf Master Gradient thermocycler (Brinkmann Instruments, Westbury, New York, USA) with the following program: one step of 94 °C (2 min), 35 cycles of 94 °C (30 s), 50 °C (30 s), and 68 °C (1 min), and one step of 68 °C (5 min), followed by holding at 4 °C. Amplicons were analyzed by electrophoresis on 1 % agarose gels stained with ethidium bromide to ensure reaction success and then quantitated using the Quant-iT PicoGreen dsDNA kit (Invitrogen) modified from the manufacturer's protocol to 50 % of the recommended final assay volume. Equal quantities of each PCR product were pooled, and the pool was purified using the Wizard SV Gel and PCR Clean-Up System (Promega). The pool was sequenced by the Centers for Disease Control and Prevention (CDC) Biotechnology Core Facility Branch on a Roche 454 GS-FLX sequencer.

To determine if the two polymerases that I used differentially biased the bacterial communities identified from these tick DNAs, I amplified 50 tick DNAs belonging to the archival collection once with each of the polymerases used in this experiment as described above. Tick DNA and barcode pairings were held constant between the libraries. A single library was produced from the AccuPrime amplicons, and two libraries were produced from the Platinum amplicons to control for variability introduced by random sampling, pipetting error, and bead loading on the 454 sequencer. Each library was sequenced on one quarter of a plate (Table 3.S2). Twenty-six of 50 DNAs were sequenced to a depth of at least 1000 high quality reads across all three treatments and were used in the analysis of polymerase effects.

Life stage effects were analyzed using amplicons from six quarters from three sequencing plates (Table 3.S2). Some DNAs were sequenced more than once, but no duplicates were used in the analyses. Samples not analyzed here were replicates of reported samples, belonged to variable groups that were not sufficiently sampled, or returned insufficient numbers of high quality sequences for further analyses.

Bioinformatic analysis

The bioinformatic pipeline was executed in the software package mothur (version 1.31.2) (Schloss *et al.* 2009). The quality control pipeline was modeled on the Schloss SOP pipeline available on the mothur website (Schloss 2013). Briefly, flow data were denoised and the resulting sequences were discarded if the reverse primer sequence or barcode could not be identified, or if barcode or primer sequences had more than 1 or 2 errors, respectively. Sequences were also discarded that contained homopolymer runs >8

bp or ambiguous nucleotide calls. Sequence trimming was performed based on a sliding window of 50 bp with a threshold quality score of 35. The remaining sequences were aligned to the SILVA SEED bacterial 16S database (Schloss 2009) using 8mer searching to choose a template sequence and the Needleman-Wunsch algorithm to perform the alignment. The SILVA SEED database was modified to include several sequences of interest, eg. *Arsenophonus* and *Candidatus* 'Midichloria mitochondrii'. Aligned sequences were then screened by starting positions to eliminate sequences that mapped outside the target region of the *rrs*. Finally, chimeric sequences were identified using the chimera.uchime function in mothur and then removed from further analyses.

All sequences passing the quality control pipeline were assigned to operational taxonomic units (OTUs) as described in the Schloss SOP. To summarize, all high quality sequence reads were ranked by abundance and rare reads differing from abundant reads by a single base were assumed to be sequencing errors and clustered with their similar sequence. An uncorrected pairwise distance matrix was calculated between the aligned sequence reads; both internal and terminal gaps were penalized once. Reads were then clustered into OTUs using the average neighbor method at a 3% distance. Taxonomic assignment of individual reads was determined by using mothur's Bayesian classifier and requiring a bootstrap confidence of 80% on 100 iterations (kmer size = 8). A consensus taxonomy was determined for each OTU based on the individual sequence taxonomies within each. Genus level identification is provided whenever possible; otherwise the lowest taxonomic rank applying to all members of the OTU is given. Singleton OTUs were removed prior to statistical analyses (Zhou *et al.* 2011). Representative highest abundance sequences from OTUs are deposited in GenBank under accession numbers

KJ130495-KJ130517; additional information can be found in Table 3.S3. The full data set is available in the NCBI SRA under study accession number SRP042723.

Bacterial diversity analyses

Good's community coverage estimate (Esty 1986), rarefaction curves, and alpha diversity were calculated in the software mothur using a subsample size of 1000 sequences and 1000 iterations. The inverse Simpson diversity index (1/D) was used for alpha diversity because it has a clear biological interpretation and is less affected by sample size (Hill *et al.* 2003) and the presence of spurious OTUs (Pinto & Raskin 2012) than similar measures. Analysis of variance (ANOVA) and Tukey's honestly significant difference (HSD) post hoc tests were performed on the R statistical platform (version 3.0.1) (R Core Team 2013).

Bacterial community analyses

I used community ecology methods to analyze both the OTU composition (presence/absence) and structure (abundance) of bacterial communities observed within each tick. As no one metric adequately compares all aspects of community assemblages, I measured the distance between bacterial communities using two metrics that each emphasize different community characteristics. The Jaccard dissimilarity index performs pairwise comparisons of communities based on the presence or absence of OTUs (Jaccard 1908), while the Bray-Curtis dissimilarity index weights the abundance of OTUs in its calculations (Bray & Curtis 1957). Each tick bacterial community was rarefied to 1000 sequences, and Jaccard and Bray-Curtis calculations were performed in R with the package "vegan" (version 2.0-7) (Oksanen *et al.* 2013) using the vegdist command.

To partition variation within each distance matrix, I used a non-parametric permutational multivariate analysis of variance (PERMANOVA) as implemented in the vegan function adonis (permutations=1000) (Oksanen *et al.* 2013). As PERMANOVA can sometimes be affected by differences in within group variation, each matrix was also tested for homogeneity of group dispersions via the vegan function betadisper (Oksanen *et al.* 2013). Significance values were obtained by permutation (n=1000). Corrections for multiple comparisons in the PERMANOVA and dispersion tests were performed using Holm's method (Holm 1979).

To visualize data sets and to corroborate the results of the PERMANOVAs, a non-metric multi-dimensional scaling (NMDS) was performed on each distance matrix. This ordination technique represents highly dimensional data by maximizing the correlation in rank-order between the original data set and a two dimensional representation (Faith *et al.* 1987; Minchin 1987). The relative location of each tick's bacterial community within the ordination space can be interpreted with the addition of species (treating OTUs as species for this purpose) scores to the plot. Bacterial communities closer to a given species score in ordination space have greater values for that species than those communities located farther away (Kindt & Coe 2005). The ordinations were produced using the function metaMDS in vegan (maximum permutations=1000) with square root and Wisconsin double standardization transformations used according to default settings (Oksanen *et al.* 2013). The goodness of fit for each variable's group centroids were evaluated using the command envfit (Oksanen *et al.* 2013), and the species scores were calculated as weighted averages of the bacterial community scores (Oksanen *et al.* 2013). Groups are

considered significantly different for $\alpha = 0.05$ if the 95% confidence intervals did not overlap.

To determine which OTUs differed in abundance between groups I used the Metastats method of White *et al.* (2009) as implemented in mothur. In an effort to control type II error rates, I only interpreted those *p*-values that had an associated false discovery rate (i.e., the proportion of false positives expected in a set of reported significant results) of $q \leq 0.05$ (Pawitan *et al.* 2005; White *et al.* 2009).

Polymerase effects analysis

For the archival tick DNAs successfully sequenced across the AccuPrime, Platinum 1, and Platinum 2 libraries, I processed the data as described above in bioinformatic analysis, rarefied each community to 1000 sequences, and compared the bacterial communities from each library originating from the same tick DNA. Jaccard and Bray-Curtis dissimilarity values were calculated pairwise for each tick's three bacterial communities. For each dissimilarity metric an ANOVA was used to compare the means of each of the three comparison groups (Platinum 1 and Platinum 2, Platinum 1 and Accuprime, Platinum 2 and Accuprime) across ticks. The means would differ only if a polymerase produced a differentially biased bacterial community composition or structure as measured by the Jaccard or Bray-Curtis metrics, respectively. Bartlett's test (Bartlett 1937) was used to ensure homogeneity of group variances. All statistics were performed on the R platform.

Co-infection analysis

The co-occurrence of the most abundant OTUs (≥100 sequences in the rarefied data) was evaluated using a probabilistic model to detect pairs that occurred together significantly more or less often than if OTU assemblages were random (Veech 2013). The analysis software was kindly provided by the test's author (v1.0). The Pearson correlation coefficient (r) was calculated using the cor.test function in the R statistical platform (R Core Team 2013) with a two-sided alternative hypothesis.

3.4 Results

Effect of Polymerase on Bacterial Community Diversity

The Accuprime polymerase library produced 114,250 raw sequences compared to the 82,843 and 93,834 sequences produced by the Platinum polymerase libraries. The distribution of raw sequence lengths (Figure 3.S1), the proportion of sequences removed as sequencing errors, alignment errors, chimeras, and contaminants (Figure 3.S2), and the proportion of sequences retained as high quality sequences (Figure 3.S2) were similar between the polymerases. Bartlett's test for homogeneity of variances was non-significant for both the Jaccard and Bray-Curtis metrics. ANOVA detected no difference in the mean level of dissimilarity between bacterial communities originating from the same tick DNA but amplified with different polymerases (Platinum 1 and Accuprime, Platinum 2 and Accuprime) and bacterial communities that originated from the same amplification of the same tick DNA (Platinum 1 and Platinum 2) (Jaccard: $F(2,75)=2.74$, $p>0.07$; Bray-Curtis: $F(2,75)=1.40$, $p>0.25$). Because the AccuPrime and Platinum SuperMix polymerases did not differentially bias bacterial community detection, bacterial

communities amplified using these polymerases were pooled in all life stage analyses below.

Life stage library characteristics

I obtained 350,501 high quality 16S rRNA gene sequence reads from 131 *A. americanum* tick samples from five sites (Tables 3.1 and 3.S2). These sequences belonged to two tick DNA sample sets, referred to as the archival and Georgia tick DNA sets. Additional tick DNA samples sequenced on the same plates were not included in the analyses because they were inferior replicates of retained tick samples, produced low sequence yields (<1000), or belonged to variable groups that were insufficiently sampled for analysis. The data were dominated by sequences from high-abundance OTUs belonging to *Rickettsia* and *Coxiella*, which comprised 35.0% and 59.4% of sequences, respectively, and were found in 88.5% and 99.2% of the ticks, respectively. The sequence reads averaged 255.6 bp in length after trimming and comprised 408 operational taxonomic units (OTUs) when clustered at 97% identity. One hundred seventy OTUs were comprised of a single sequence (singletons) and were removed (Zhou *et al.* 2011). To facilitate comparison of bacterial communities between samples, each bacterial community was then rarefied to 1000 sequences. This reduced the total number of OTUs to 212, of which 42 contained only a single sequence. The 170 OTUs containing >1 sequence represented a minimum of 99 bacterial families and 99 genera.

Bacterial diversity

The Georgia DNA set's rarefaction curve approached an asymptote, indicating that the available bacterial diversity was well sampled (Figure 3.S3). This conclusion was

supported by the Good's coverage estimate for each tick sample, which had a mean of 0.998 (range 0.994-1.00). In contrast, the bacterial diversity of the archival DNAs, from ticks which were not surface decontaminated prior to DNA extraction, was less well sampled and the rarefaction curve had a steeper slope. However, while adding additional tick DNAs samples to this set would have detected additional OTUs, the communities of each tick in the data set were individually well sampled (Good's coverage estimate, mean=0.993, range=0.976-1.00).

Alpha diversity was estimated using the inverse Simpson index and compared between DNA sample sets and across life stages and collection sites (Figure 3.1). Significant differences were found for all variables considered (ANOVA with Holm's correction for multiple comparisons, DNA collections: $F(1,129)=15.9$, $p=0.00023$; life stages: $F(4,126)=5.21$, $p=0.00064$; sites: $F(4,126)=6.83$, $p=0.00016$). Among life stages from the different DNA collection sets, adult females from Georgia had lower alpha diversity than adult males (Tukey's HSD, $p=0.013$) and females ($p=0.0017$) from the archival DNA collection. Comparisons across collection sites found a higher alpha diversity among ticks from Shelter Island compared to those from Bodie Island (Tukey's HSD, $p=0.022$), Sweetwater Creek ($p=0.00059$) and Panola Mountain ($p=6.9\text{x}10^{-5}$).

Bacterial community composition and structure

Of the 212 OTUs analyzed across both DNA collections, only *Rickettsia*, *Coxiella*, and a Bacillales group were observed in more than one third of all standardized samples (Figure 3.2). An additional 17 OTUs composed ≥1% of the community when detected, but were detected in one third or fewer of the tick DNA samples. No individual OTU was

present in all rarefied samples, including *Coxiella*, a proposed obligate symbiont of *A. americanum* (Zhong *et al.* 2007). Of the two negative samples, the male from Panola Mountain was *Coxiella* positive prior to rarefaction, while the nymph from Sweetwater Creek was positive in a replicate sequencing run (data not shown).

DNAs from ticks collected in Georgia differed from those in the archival DNA collection in the age of the samples, the tick life stages represented, and the DNA extraction protocols used. Bacterial community structure was quite different between these two groups of ticks (PERMANOVA on Bray-Curtis distance matrix, F=5.87, R^2=0.044, p=0.012) (Figure 3.S4), but they did not differ in their dispersion (F=1.09, p=0.28). The bacterial communities of the Georgia tick DNAs contained 12 OTUs not found in the archival tick DNA bacterial communities, while the archival tick DNAs contained 61 OTUS not found among the Georgia tick bacterial communities. An additional 7 OTUs were present in both DNA collections but differentially abundant (Table 3.S4). Greater than 93% (75/80) of the OTUs that were differentially abundant between the Georgia and archival tick DNAs have no confirmed association with the internal microbiota of ticks and therefore may have been derived from the external environment.

Twenty-two positive associations were detected among the 12 most abundant (>100 sequences) OTUs from the Georgia DNA collection (Figure 3.3). The Bacillales 1, *Rhodobaca*, and Rhizobiales 1 OTUs had the largest number of associations. The *Wolbachia* and *Midichloria* 1 OTUs were strongly positively associated. Two positive associations were found among the most abundant OTUs from the archival DNA collection (Rhizobiales and Acidobacteria Gp1, Burkholderiaceae and *Methylobacterium*, p<0.04). In general no significant associations were detected between the *Rickettsia* and

Coxiella OTUs and the other OTUs due to the near ubiquity of *Rickettsia* and *Coxiella* in these samples. However, there was a significant inverse relationship between the abundance of *Rickettsia* and *Coxiella* across both sample sets (Georgia: n=104, Pearson's r= -0.976, p=2.2x10^{-16}; archival: n=27, Pearson's r= -0.689, p=6.97x10^{-6}). The life stages differed in the dominant bacteria (ANOVA on percentage *Rickettsia* out of total *Rickettsia* and *Coxiella* sequences, Georgia: F(2,101)=26.8, p=4.68x10^{-10}; archival: F(1,25)=14.0, p=0.00097) (Figure 3.4). Within the Georgia collection, females differed from males and nymphs (Tukey's HSD, males: p=1.0x10^{-7}, nymphs: p<1.0x10^{-7}) (Figure 3.4).

Life stage and sex specific bacterial communities

Ticks from the Georgia collection were analyzed to detect differences between female, male, and nymphal bacterial communities (n=104). Samples from Sweetwater Creek and Panola Mountain, Georgia were pooled because no differences were detected between the sites (data not shown). PERMANOVA found significant differences between life stages and sexes using both distance metrics (Table 3.2). Differences in within group variation (i.e. dispersion) were also detected with both distance matrices (Table 3.2) (Anderson 2001). Results from the PERMANOVA were corroborated using ordination by NMDS. When only the presence or absence of OTUs was considered (Figure 3.S5), nymphal communities were different from adults, but male and female communities did not differ. However, differences were detected between males and females when the ordination was weighted for OTU abundance (Figure 3.5). Comparison of the mean relative abundance of OTUs detected 8 of 13 which were differentially abundant between the tick life stages and sexes (Figure 3.S6).

To determine if an effect of sex on tick bacterial communities could be detected in other parts of the *A. americanum* range, I applied PERMANOVA to Jaccard and Bray-Curtis distance matrices with data from the archival DNA samples from adult ticks collected outside of Georgia (n=27) (Table 3.1). Ticks from all three sites were pooled for analysis because no site effect was detected (data not shown). PERMANOVA found significant differences for both distance matrices (Table 3.2), but NMDS found there was a difference between males and females only in the relative abundance of bacterial community members (Bray-Curtis metric, Figure 3.6) and not in community membership (Jaccard metric, Figure 3.S7). Analysis of the OTUs containing ≥100 sequences each found 2 of 18 were differentially abundant between groups, with females possessing significantly higher *Coxiella* abundance and having higher *Francisella* abundance ($p<0.001$).

3.5 Discussion

454 pyrosequencing of barcoded *rrs* variable region 5-3 amplicons was used here to characterize the bacterial communities of 131 individual *A. americanum* ticks, including nymphs and adult males and females. The ticks were all questing, field-collected individuals obtained from five sites in three states; as such they are probably at least representative of the diversity of bacterial communities that occur in *A. americanum* found in the eastern states comprising its range.

The exact relationship of 454 sequence read abundance to the biological abundance of a given bacteria in the community is difficult to prove as amplification efficiency varies

both between bacterial targets (*e.g.* template GC content (Pinto & Raskin 2012), primer-template matching (Hong *et al.* 2009), 16S copy number variation per chromosome and per cell) and laboratory protocols (*e.g.* primer bar-code bias (Berry *et al.* 2011), PCR conditions (Acinas *et al.* 2005)). In this system, while previous work has shown a good correlation between 454 read abundance and targeted qPCR quantitation (Clay & Fuqua 2010), it is not currently possible to confirm such a relationship for all potential members of the *A. americanum* bacterial community. In general, while comparing communities across protocols is problematic due to the many potential sources of bias, 454 metagenome results are reproducible for a given protocol (Kauserud *et al.* 2012; Pilloni *et al.* 2012; Pinto & Raskin 2012); however, some authors have found that this is not always true (Zhou *et al.* 2011).

I document here significant differences in the bacterial communities of the life stages and sexes of questing *A. americanum*. Males and females in both the Georia and archival DNA collections differed in their bacterial community structure (Figures 3.5 and 3.6) but not composition (Figures 3.S5 and 3.S7). However, nymphs differed from adults in both bacterial community composition and structure (Figures 3.S5 and 3.5, respectively), with many of the more abundant OTUs appearing primarily in nymphs (Figure 3.S7). Many, but not all, of these OTUs belonged to taxa with both environmental and arthropod-associated members. While the nymphs from Georgia were surface decontaminated prior to DNA extraction, my method was unlikely to remove all exoskeleton-associated bacteria. The higher surface to volume ratio of nymphs compared to adults also provides a larger relative area for contamination. However, Menchaca *et al.* (2013) found high

levels of Bradyrhizobiaceae in both nymphs and adults derived from laboratory colonies and held indoors, suggesting a more persistent association for at least one of these taxa.

As I expected from previous investigations of the bacterial community of *A. americanum* (Mixson *et al.* 2006; Clay *et al.* 2008; Yuan 2010; Moncayo *et al.* 2010; Fritzen *et al.* 2011; Menchaca *et al.* 2013) (Table 3.S3), *Rickettsia* and *Coxiella* were by far the dominant OTUs identified in my samples across all life stages, sexes, and DNA collections. The *A. americanum*-associated *Coxiella* is regarded as an obligate symbiont of the lone star tick due to its ubiquitous presence in previous surveys, limited evidence for a reduced genome size relative to the free-living relative (*Coxiella burnetii*), transovarial and transstadial maintenance, and the reduced viability and fecundity of antibiotic treated ticks (Jasinskas *et al.* 2007; Zhong *et al.* 2007; Klyachko *et al.* 2007; Clay *et al.* 2008). While one tick had no detectable *Coxiella* sequences in the data presented here, I later detected low levels of *Coxiella* in a replicate data set. This suggests that the abundance of *Coxiella* was at the limit of detection for 454 in this sample. My data provide further support for the hypothesis that *Coxiella* is an obligate symbiont of *A. americanum*, although its population size in some individuals may be very low relative to other members of the bacterial community.

Fragments of the *rrs* are insensitive for the identification of different species of *Rickettsia*. However, abundant previous work has shown that the most common *Rickettsia* in *A. americanum* is *R. amblyommii* (*e.g.* Mixson *et al.* 2006; Clay *et al.* 2008; Jiang *et al.* 2010; Zhang *et al.* 2012), a member of the spotted fever group with poorly defined human pathogenicity (Dasch *et al.* 1993; Billeter *et al.* 2007; Apperson *et al.* 2008). Despite efficient vertical (Stromdahl *et al.* 2008) and probable horizontal

(reviewed in Smith *et al.* 2010) transmission in *A. americanum*, infection rates in populations across the tick's range vary widely (0-84%) (Mixson *et al.* 2006; Clay *et al.* 2008; Jiang *et al.* 2010; Zhang *et al.* 2012). My archival tick DNAs (Table 3.1) had been previously tested for *R. amblyommii* using direct PCR of the gene encoding the rOmpA protein followed by restriction fragment length polymorphism analysis; techniques and data were described in Mixson *et al.* (2006). Of these 27 adult ticks, two were negative by direct PCR but contained at least a few *Rickettsia* sequences when analyzed here (n=1-160 sequences per tick, original unrarefied data). While *R. amblyommii* generally occurs as a high density, disseminated infection in *A. americanum* (Zanetti *et al.* 2008), taken together these data indicate that some *R. amblyommii* infections are below the detection limit of direct PCR and may not always be reliably detected by that insensitive technique (this has been affirmed by qPCR comparison of varying and often low *R. amblyommii* levels in other tick samples, Dasch unpublished results). *Rickettsia amblyommii* infection rates are therefore likely to be somewhat higher than has been previously reported, although whether or not the bacterium is actually ubiquitous or transcriptionally active in all samples remains to be determined. If naturally uninfected ticks exist, these low level infections may represent individuals that had acquired the bacterium during the preceding blood meal.

There was a strong inverse relationship between the abundance of *Rickettsia* and *Coxiella* in the bacterial communities of the adult *A. americanum* ticks from both the Georgia and archival DNA collections. Interestingly, the dominant bacterium varied with the sex of adults such that females produced more *Coxiella* sequences and males more *Rickettsia* sequences (Figure 3.4). The relatively higher *Coxiella* to *Rickettsia* ratio observed in

females is supported by the qPCR data of Jasinskas *et al.* (2007), but differs from the report of Ponnusamy *et al.* (2014), who detected an increase in *Rickettsia* in females and low levels of *Coxiella* across all *A. americanum* samples using pyrosequenced V1-3 *rrs* fragments. Given the extensive evidence for *Coxiella*'s abundance in most *A. americanum* samples (Clay *et al.* 2008; Heise *et al.* 2010; Yuan 2010; Menchaca *et al.* 2013) and the small sample analyzed by Ponnusamy *et al.*, it is likely that either some bias against *Coxiella* existed in the Ponnusamy *et al.* protocol or that a non-representative sample was drawn from their population. If *Coxiella* is obligate for the survival of *A. americanum* (Zhong *et al.* 2007), the number of bacteria may increase in females to ensure 100% transmission to eggs. An alternative, but not mutually exclusive, explanation is that males may contain higher numbers of *R. amblyommii* to facilitate increased locomotion (Kagemann & Clay 2013), which would increase reproductive fitness if it enabled more mating opportunities for males. The distribution of nymphal bacterial communities across the gradient from *Rickettsia* dominated to *Coxiella* dominated suggests that sex-biased bacterial communities may arise in unfed questing nymphs (Figure 3.4). It will be difficult to prove this hypothesis until molecular sexing tools are available for *A. americanum*.

The only previously documented difference in *A. americanum* bacterial communities was decreased species richness in females relative to males and nymphs (Ponnusamy *et al.* 2014), but similar patterns to those reported here have been noted for other ixodid tick species. In *Rhipicephalus turanicus*, *Coxiella* sequences composed 90% of the total detected by bacterial 16S pyrosequencing in both males and females, but *Rickettsia* were relatively more abundant in infected males than in females (Lalzar *et al.* 2012). Males

also had greater bacterial richness than females, although there was no difference in diversity. In *Ixodes scapularis*, Moreno *et al.* used temporal temperature gradient gel electrophoresis of bacterial 16S amplicons to "finger print" the bacterial communities of ticks and found differences between nymphs, males, and females (Moreno *et al.* 2006). *Francisella* have been shown to increase in *Dermacentor variabilis* nymphs relative to larvae using 16S amplicon pyrosequencing (Hawlena *et al.* 2013). Studies of individual bacteria have also revealed sex and life stage biased prevalence. For example, *Candidatus* 'Midichloria mitochondrii' (hereafter "*M. mitochondrii*") is detected in only 40% of males but 100% of all other life stages of *Ixodes ricinus* (Lo *et al.* 2006). *Borrelia* and *Anaplasma* infection rates also differ between male and female *I. ricinus* (Halos *et al.* 2010).

I also identified other potentially important intracellular bacteria in *A. americanum*. Two OTUs of *M. mitochondrii* were found, which contained all three of the genotypes detected from *A. americanum* in my previous work (Williams-Newkirk *et al.* 2012). Operational taxonomic unit *Midichloria* 1 corresponds to genotypes A and B, and OTU *Midichloria* 2 is equivalent to genotype C. *Wolbachia* sequences were also identified in *A. americanum* from Georgia. *Wolbachia* have been previously reported in *A. americanum* by Zhang *et al.* (2011), but a direct comparison is not possible because the two studies used different regions of the *rrs* gene. *Wolbachia* sequences were detected in my unrarefied data from 11 nymphs and 2 females collected from the two Georgia sites (Panola Mountain nymphs: 4/30, females: 2/14; Sweetwater Creek nymphs: 7/36, females: 0/12). These infection rates were at least two times greater than the highest minimum infection rates previously observed in nymph and female populations in

Maryland (Zhang *et al.* 2011), which could be a result of the increased sensitivity of Next-Gen sequencing versus direct PCR, the dilution effect of pooling of nymphs in earlier work, or simply a statistical artifact of my smaller sample sizes. As in previous work (Zhang *et al.* 2011), I detected no *Wolbachia*-positive males, although my sample sizes for males were lower than those for females and nymphs. These *Wolbachia* have been previously typed as supergroup F (Zhang *et al.* 2011), which also contains *Wolbachia* that infect scorpions (Baldo *et al.* 2007), lice (Covacin & Barker 2007), filarial nematodes (Casiraghi *et al.* 2001), and other hosts (reviewed in Lefoulon *et al.* 2012). Interestingly, 11/13 of these *Wolbachia* positive ticks were also positive for genotype A of *M. mitochondrii*, which to date has only been found in Georgia (Williams-Newkirk *et al.* 2012). This co-infection was highly significant, despite the co-clustering of *M. mitochondrii* genotype A in OTU *Midichloria* 1 with genotype B ($p<0.0001$) (Figure 3.3). There are at least three potential mechanisms that may lead to this degree of co-infection, including (i) stable vertical transmission of both bacteria within one or more tick lineages, (ii) repeated, simultaneous horizontal transmission of both bacteria from a single source, or (iii) differential survival of *M. mitochondrii*-infected ticks when attacked by *Wolbachia*-bearing insects or nematodes. Further work will be needed to differentiate between these hypotheses.

The bacterial OTUs that I detected across all tick samples were very similar to those identified by Yuan (2010), who also used wild caught ticks and similar sequencing methods. My results also did not differ substantively from those of Clay *et al.* (2008) or Heise *et al.* (2010), except for the absence of *Arsenophonus* in my samples (Clay *et al.* 2008). However, my results were generally quite different from the previous work of

Menchaca *et al.* (2013), despite having used a similar fragment of the *rrs*. While both studies detected *Coxiella*, the ticks from the previous study contained a large proportion of taxa found rarely or not at all in my data, such as Clostridia, Caulobacteraceae, and Hyphomicrobiaceae. Whether these differences result from the different methods used by Menchaca *et al.* (Ion Torrent sequencing of nested PCR products), the use of laboratory raised ticks, or their exclusive feeding of ticks on a domestic chicken remains to be determined. However, some evidence for *D. variabilis* and *I. scapularis* ticks suggests that the vertebrate hosts have little effect on bacterial community structure (Hawlena *et al.* 2013), while some laboratory colonies of ticks have been observed to differ in some members of their bacterial community (Heise *et al.* 2010).

Most of the bacterial taxa identified in this study are broadly distributed in the environment with members found in soil, plants, and arthropods (Table 3.S3). Many were previously identified in similar deep sequencing studies of other tick species (Table 3.S3), but the low taxonomic resolution provided by short *rrs* fragments for these very diverse groups does not preclude similar surface contaminants on ticks collected on different continents. The enrichment of bacterial communities from the archival tick DNAs extracted without surface disinfection with potential environmental bacteria (Table 3.S4) suggests that they are indeed contaminants, but the confounding variables of geographic location, DNA extraction methods, and tick life stage prevent definitive analysis here. Previous comparisons of surface decontaminated versus untreated *A. americanum* adults supports Sphingobacteria as a surface contaminant, but not other high abundance OTUs (Menchaca *et al.* 2013). Several of these OTUs (Actinobacteria, *Pseudomonas*, Bacillales, Rhizobiales, and Burkholderiaceae) were also commonly found

in *A. americanum* samples collected in the Midwestern USA (Clay *et al.* 2008), Texas and Missouri (Yuan 2010), and laboratory colonies (Heise *et al.* 2010). Further work with bacterial species-specific methods should be undertaken to determine the topological relationship of these bacteria to ticks.

In summary, I have found a pattern of sex-specific bacterial community structure that was present irrespective of geographic origin of my *A. americanum* samples and significant differences in both the composition and structure of nymphal bacterial communities relative to adults. I also documented a novel association between *M. mitochondrii* and *Wolbachia* in ticks from Georgia. Given the innate limitations of the sequencing of *rrs* amplicon fragments, this metagenomic work has been effective as a discovery tool for future targeted studies. However, the extreme dominance of *R. amblyommii* and *Coxiella* in the *A. americanum* bacterial community makes it difficult to detect and analyze other community members in a statistically robust way. Future metagenomic studies would benefit from targeted depletion of the most prevalent and predominant bacteria to increase detection rates for other rarer but potentially important community members.

3.6 Figures

Figure 3.1 Comparisons of bacterial community alpha diversity by tick DNA set, life stage, and collection site. Box and whiskers plots represent the first and third quartiles (hinges), the median (bold line), and 1.5 times the interquartile range (error bars). Outlier points are plotted individually. Within each of three panels, n=131 (see Table 3.1) and groups labeled A are significantly different from groups labeled B at α=0.05 (Tukey's HSD). Group sample sizes are given in parentheses after the group name. For the tick DNA sample set comparison, the archival tick DNAs were older and processed under different extraction protocols than the Georgia samples.

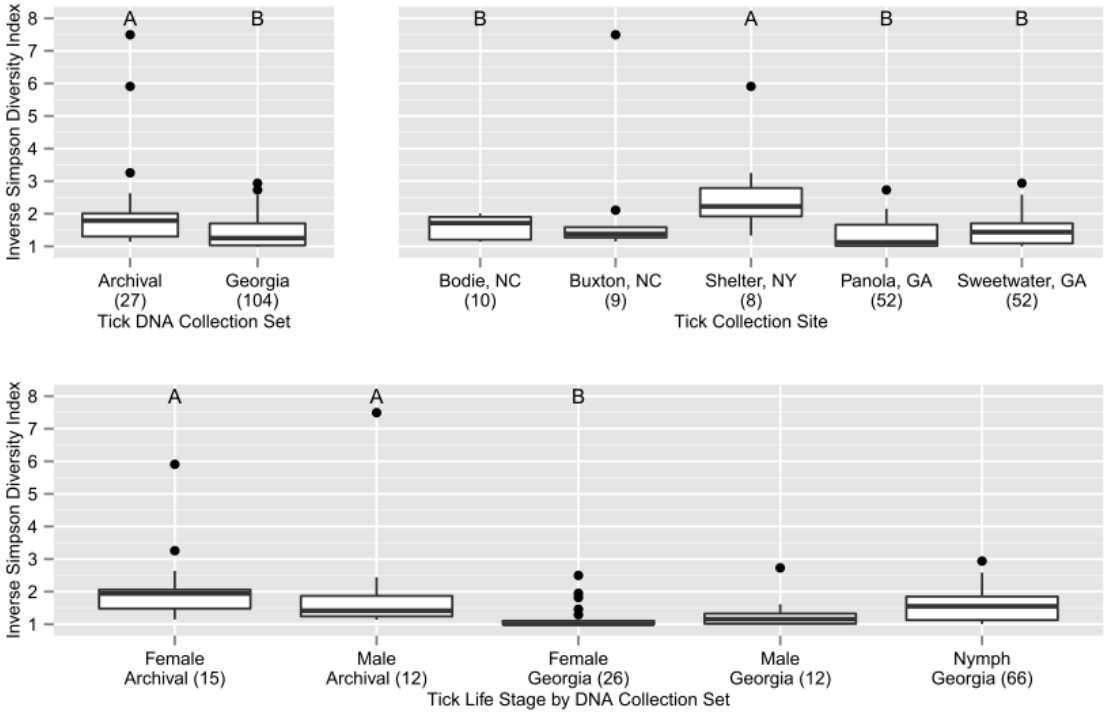


Figure 3.2 Bacterial operational taxonomic unit (OTU) sequence abundance detected in tick samples. OTUs present in greater than one third of all tick samples were considered broadly distributed. OTUs present at a mean abundance ≥1% of the community (i.e. 10 sequences) were considered locally abundant. Sequence abundance values are slightly offset to reveal overlapping OTUs.

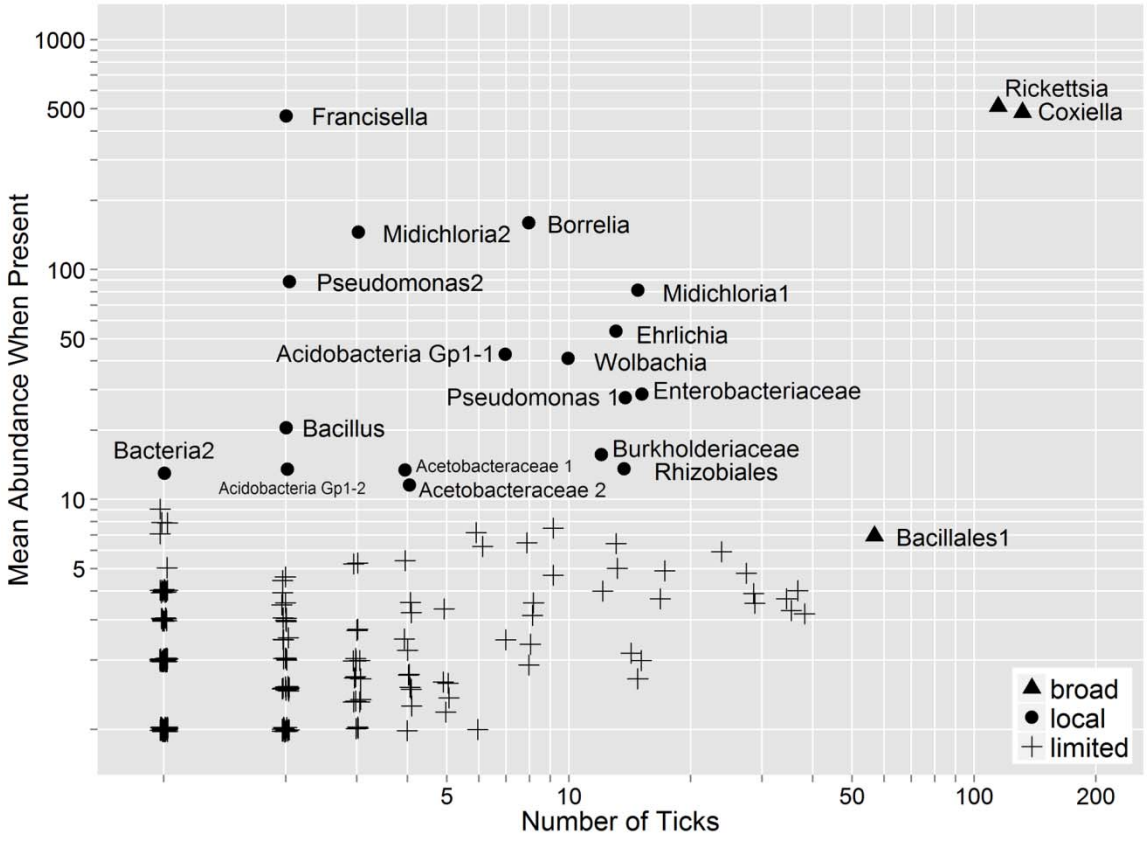


Figure 3.3 Significance of co-infection patterns between selected operational taxonomic units (OTUs). Only data from the Georgia tick DNA sample collection is shown. The upper triangle represents *p*-values of positive associations (codetection in the same sample) between OTUs and the lower triangle the *p*-values of negative associations. Black cells were not evaluated due to low expected co-occurrence (<1) between the OTUs (Veech 2013).

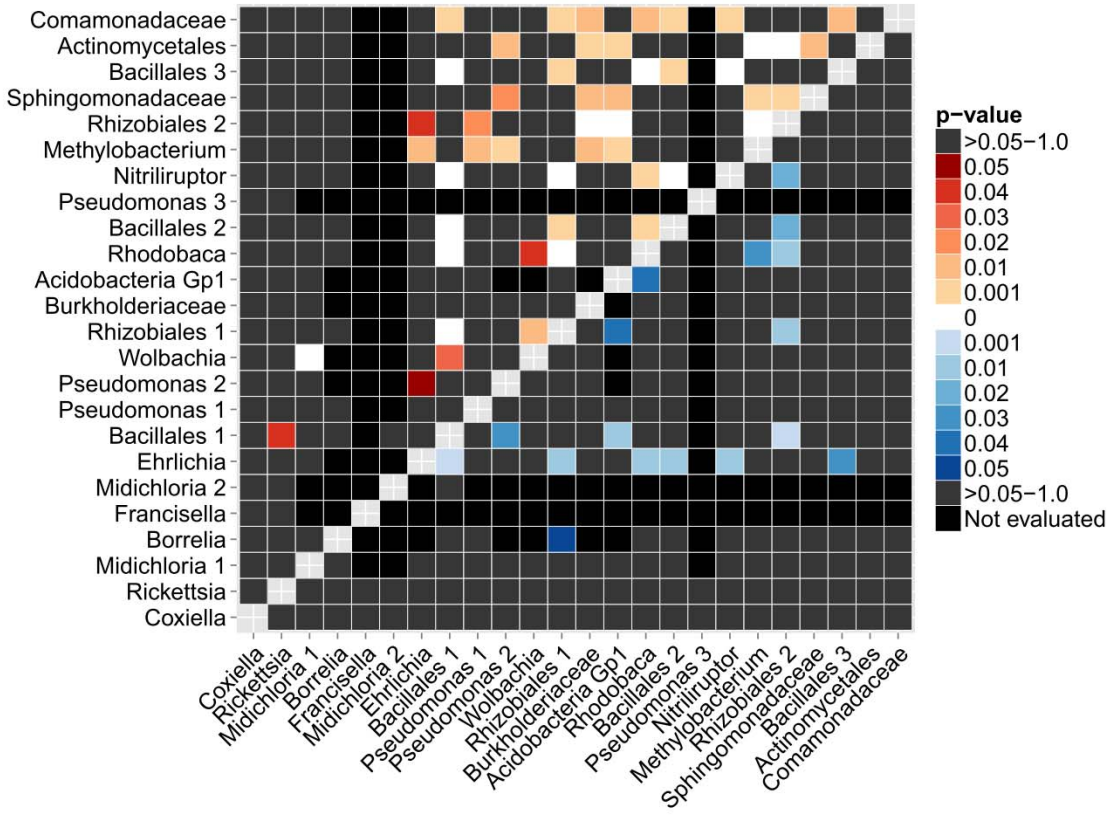


Figure 3.4 Variation in percentage of *Rickettsia* among total *Rickettsia* and *Coxiella* sequences by tick life stage. The number of unstandardized *Rickettsia* and *Coxiella* sequences from each tick sample (Georgia n=104, archival n=27) were summed and the percentage of *Rickettsia* sequences were calculated. Box and whiskers plots represent the first and third quartiles (hinges), the median (bold line), and 1.5 times the interquartile range (error bars). Outlier points are plotted individually. Within each DNA collection, life stages labeled with different letters are significantly different at α=0.05.

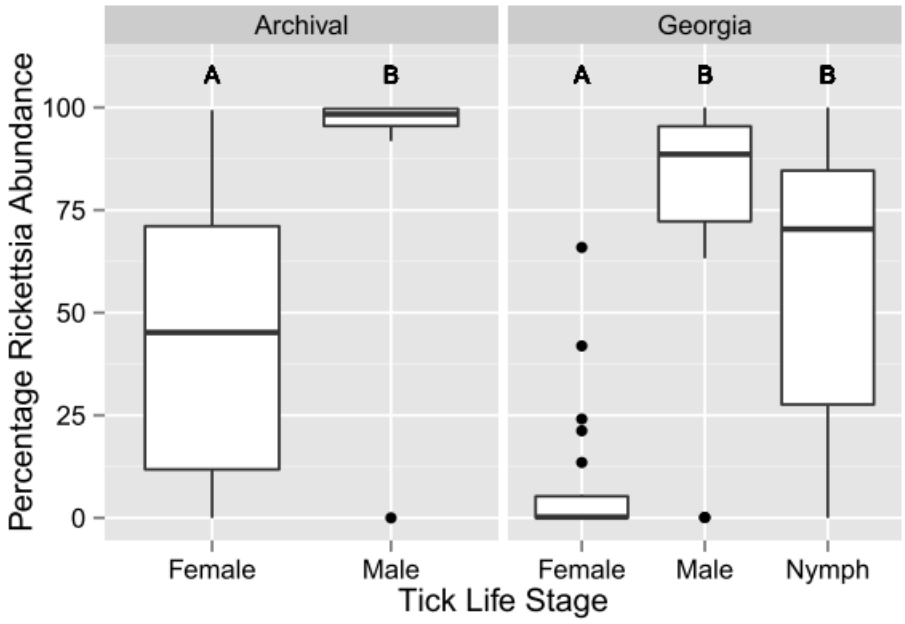


Figure 3.5 Non-metric multi-dimensional scaling (NMDS) of a Bray-Curtis distance matrix describing Georgian *Amblyomma americanum* bacterial communities. Each point symbolizes a single tick's community (n=104); some points may overlap completely. Point and ellipse colors indicate life stage; ellipses represent 95% confidence intervals around life stage centroids. Non-overlapping centroids are considered significantly different at α=0.05. R^2 values in the upper left corner of plots describe the amount of variation in the data set explained by the groupings. The stress value given is a measure of the disagreement between the rank order in the original data set and that in the NMDS (lower numbers indicate better agreement). Brown numbers indicate the species scores for select OTUs as follows: (1) *Coxiella*, (2) *Rickettsia*, (3) *Midichloria* 1, (4) *Borrelia*, (5) *Ehrlichia*, (6) Bacillales 1, (7) *Wolbachia*, (8) Rhizobiales 1, (9) *Rhodobaca*, (10) Bacillales 2, (11) *Nitriliruptor*, (12) Bacillales 3.

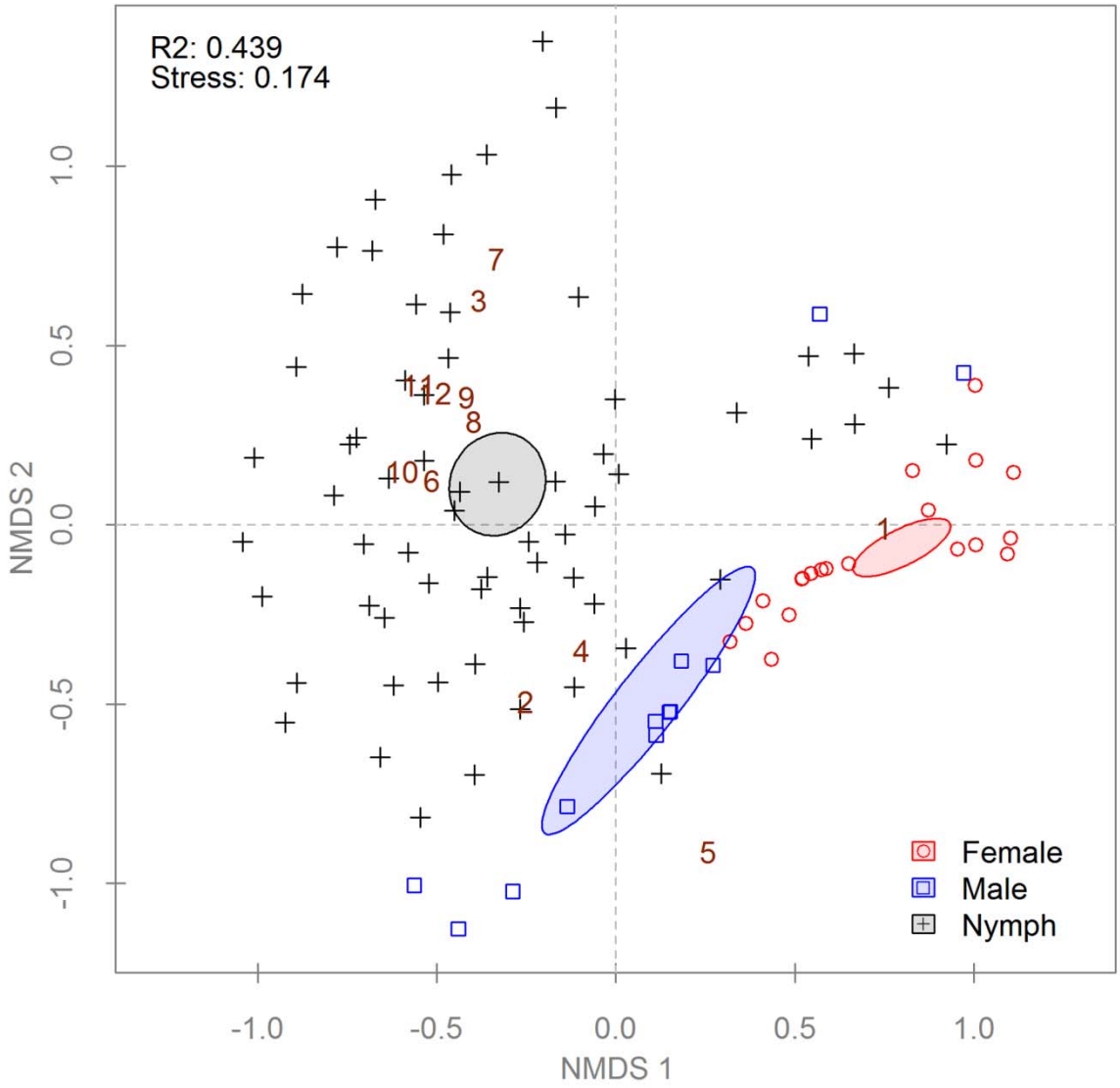

R2: 0.439
Stress: 0.174
NMDS 1
NMDS 2
-1.0
-0.5
0.0
0.5
1.0
1
2
3
4
5
6
7
8
9
10
11
12
Female
Male
Nymph

Figure 3.6 Non-metric multi-dimensional scaling (NMDS) of a Bray-Curtis distance matrix describing archival *Amblyomma americanum* bacterial communities. Each point symbolizes a single tick's community (n=27); some points may overlap completely. Point and ellipse colors indicate life stage; ellipses represent 95% confidence intervals around life stage centroids. Non-overlapping centroids are considered significantly different at α=0.05. R^2 values in the upper left corner of plots describe the amount of variation in the data set explained by the groupings. The stress value is a measure of the disagreement between the rank order in the original data set and that in the NMDS (lower numbers indicate better agreement). Numbers indicate the species scores for the plot as follows: (1) *Coxiella*, (2) *Rickettsia*, (3) *Midichloria* 1, (4) *Borrelia*, (5) *Francisella*, (6) *Midichloria* 2, (7) *Ehrlichia*, (8) Enterobacteriaceae, (9) *Pseudomonas* 1, (10) Burkholderiaceae, (11) Acidobacteria Gp1, (12) *Pseudomonas* 2, (13) Rhizobiales 2, (14) *Bacillus*, (15) Acetobacteraceae 1, (16) Acetobacteraceae 2, (17) Bacteria 1, (18) Bacteria 2.

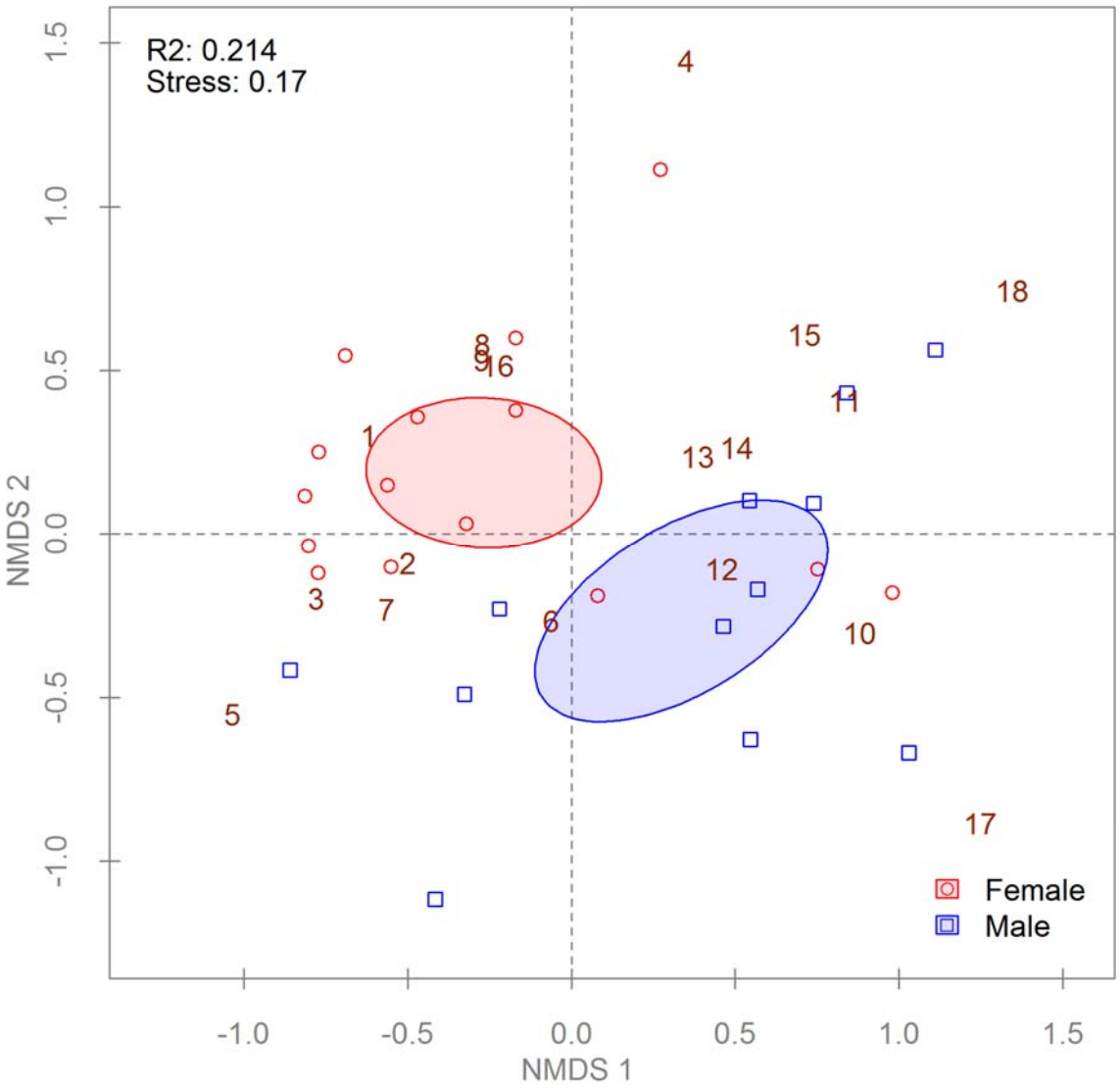

R2: 0.214
Stress: 0.17
NMDS 1
NMDS 2
-1.0
-0.5
0.0
0.5
1.0
1.5
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
Female
Male

Figure 3.S1 Comparison of raw 454 read length distributions between replicate libraries created with different polymerases. Each library was produced with the same set of DNA templates and barcoded primers. The polymerase varied between panels A and B/C, while B and C were produced from the same amplification products but pooled and sequenced independently. Panel A represents the library created with AccuPrime Taq, and panels B and C represent Platinum SuperMix replicates 1 and 2, respectively. The horizontal dashed line marks 100,000 on the log scaled y-axis on each plot.

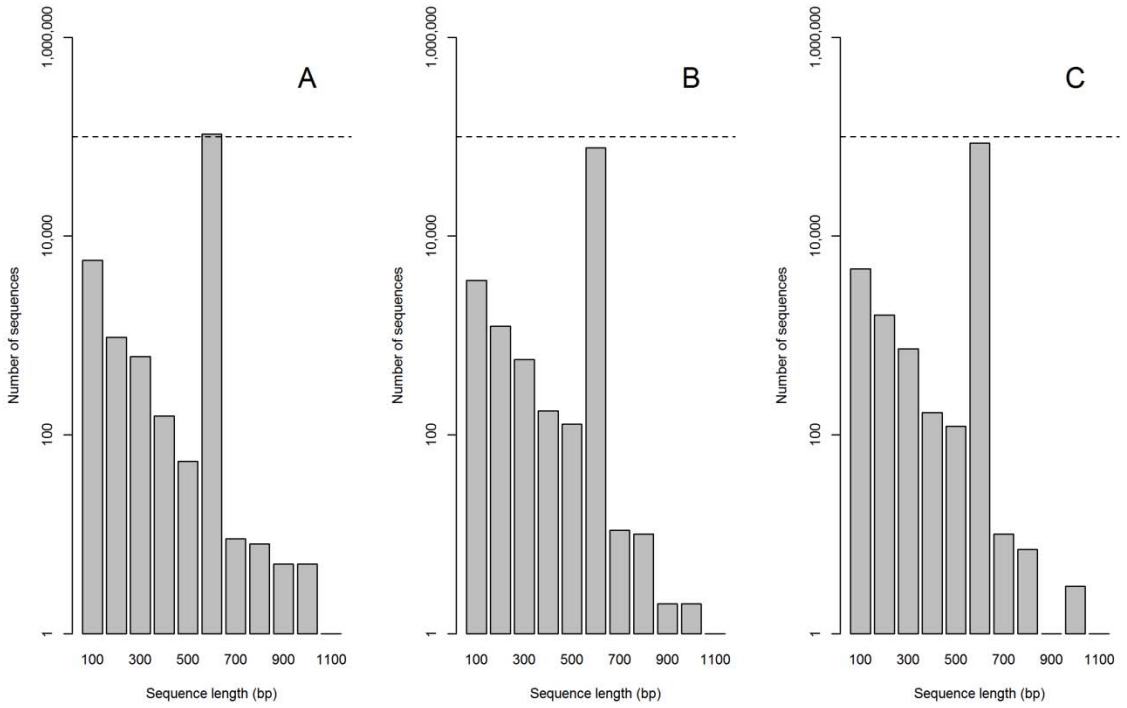


Figure 3.S2 Comparison of quality control filtering between replicate 454 libraries created with different polymerases. Each library was produced with the same set of DNA templates and barcoded primers. Libraries Platinum 1 and Platinum 2 where produced from the same amplification products but pooled and sequenced independently. The total number of raw sequences per library is given at the top of each bar. Note that the categories given in the key are in the same order as they appear in the plot; contaminants were present in all three libraries but accounted for <0.04% of each library.

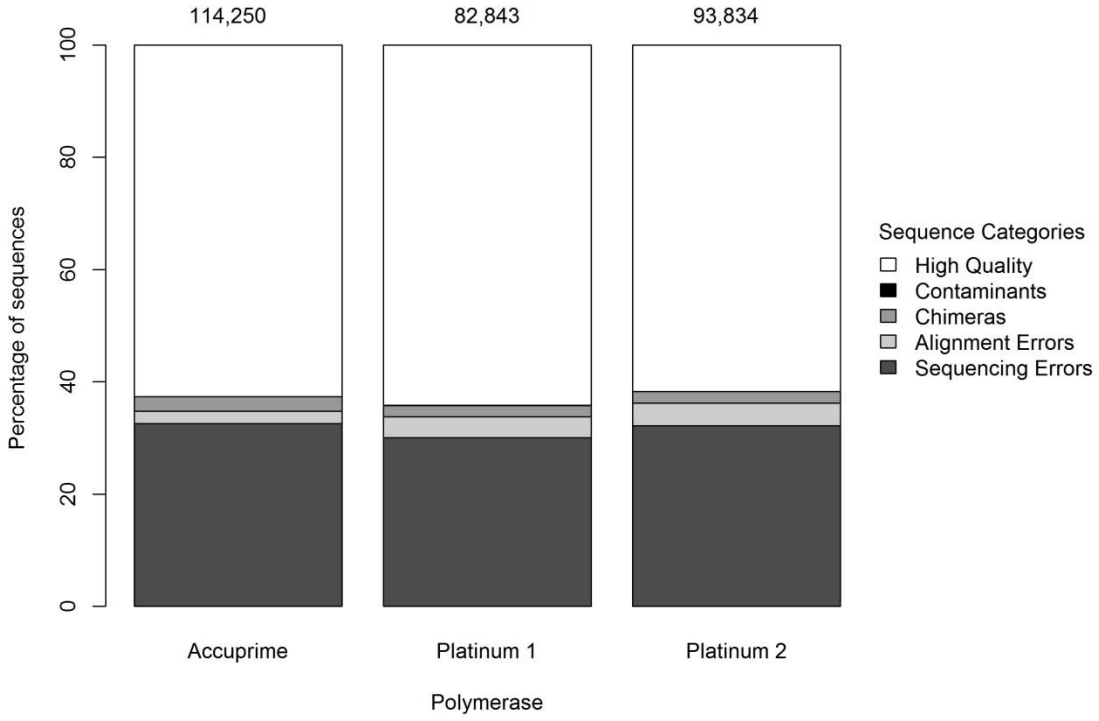


Figure 3.S3 Bacterial operational taxonomic unit rarefaction curve for *Amblyomma americanum* tick DNA sample sets. The shaded regions represent the conditional 95% confidence interval obtained from 1000 randomizations of the data.

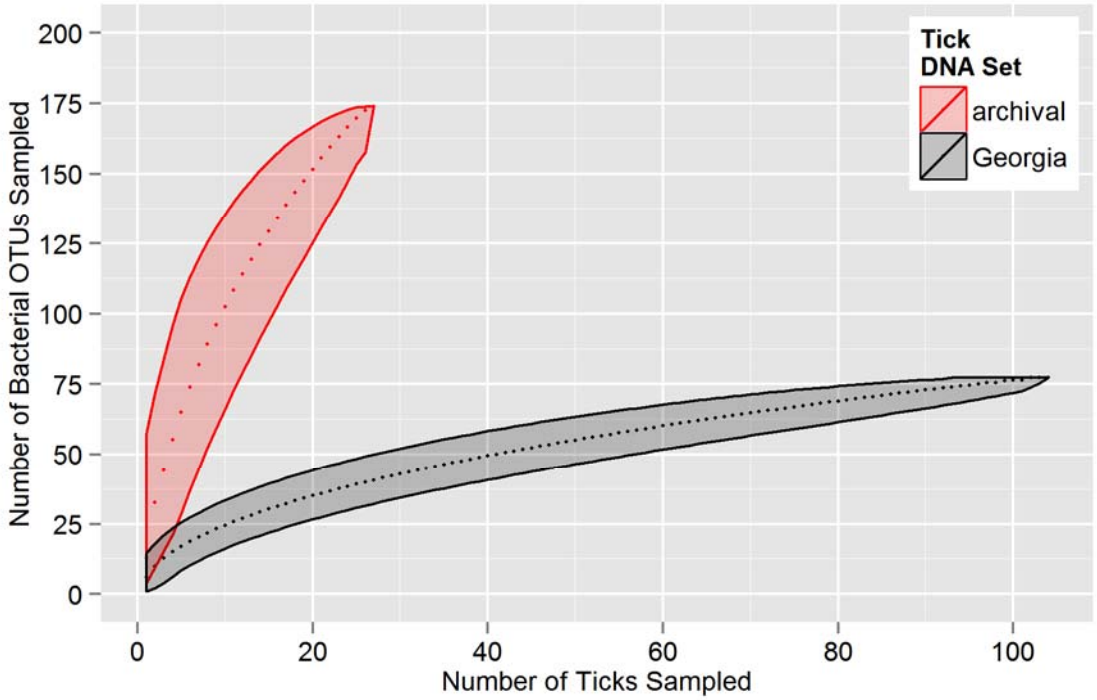


Figure 3.S4 Non-metric multi-dimensional scaling (NMDS) of a Bray-Curtis distance matrix describing *Amblyomma americanum* bacterial communities. Each of the 131 points symbolizes a single tick's community; some points may overlap completely. Point and ellipse colors (red=Georgia ticks, black=archival) indicate the tick DNA collection of origin; ellipses represent 95% confidence intervals around group centroids. Non-overlapping centroids are considered significantly different at $\alpha=0.05$. The R^2 value in the upper left corner of the plot describes the amount of variation in the data set explained by the groupings. The stress value is a measure of the disagreement between the rank order in the original data set and that in the NMDS (lower numbers indicate better agreement). Operational taxonomic units that differ in mean relative abundance between the groups are given in Table 3.S4.

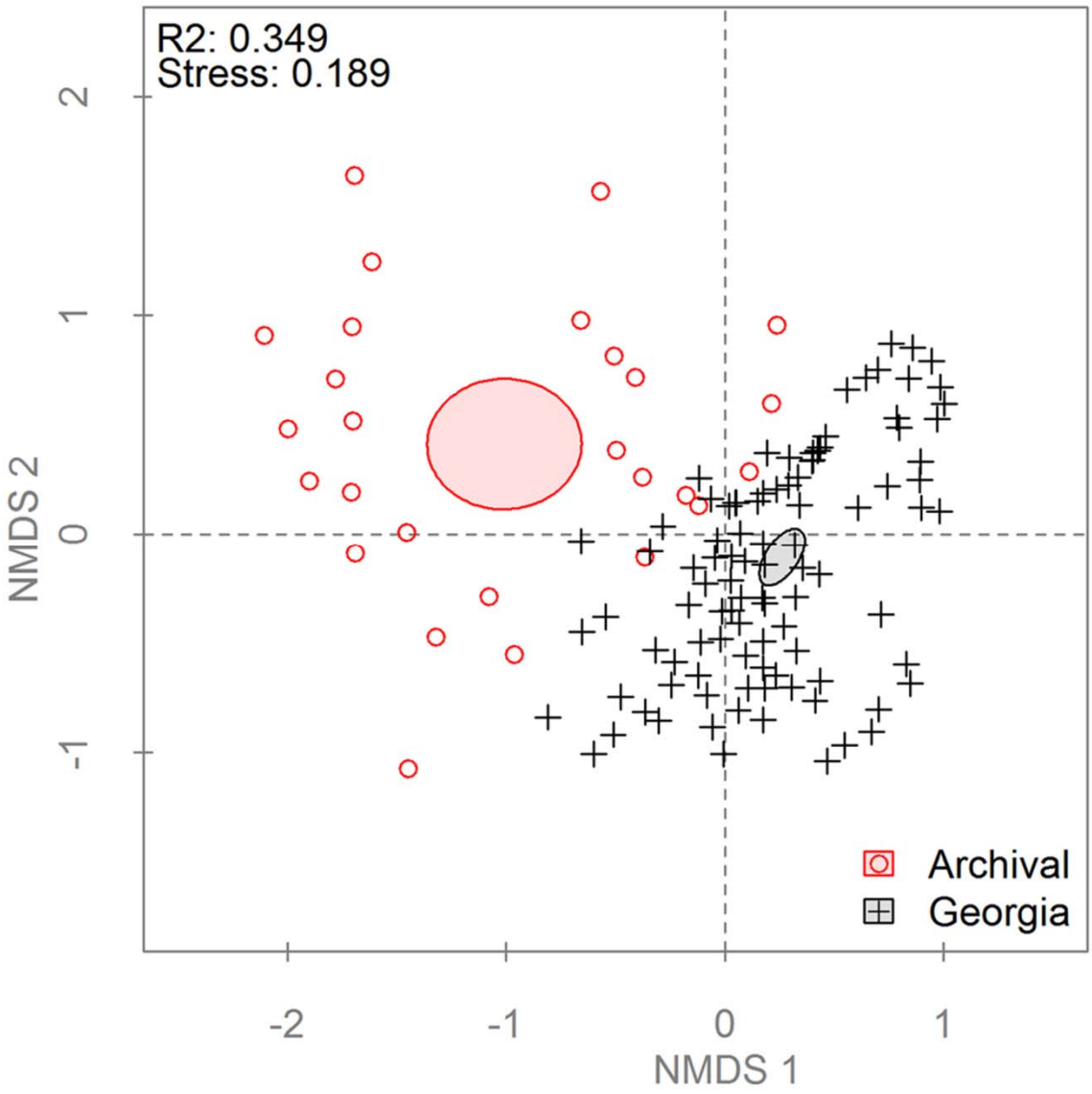

R2: 0.349
Stress: 0.189
NMDS 2
NMDS 1
2
1
0
-1
-2
-1
0
1
Archival
Georgia

Figure 3.S5 Non-metric multi-dimensional scaling (NMDS) of a Jaccard distance matrix describing Georgian *Amblyomma americanum* bacterial communities. Each point symbolizes a single tick's community (n=104); some points may overlap completely. Point and ellipse colors indicate life stage; ellipses represent 95% confidence intervals around life stage centroids. Non-overlapping centroids are considered significantly different at α=0.05. R^2 values in the upper left corner of plots describe the amount of variation in the data set explained by the groupings. The stress value is a measure of the disagreement between the rank order in the original data set and that in the NMDS (lower numbers indicate better agreement). Brown numbers indicate the species scores for select OTUs as follows: (1) *Coxiella*, (2) *Rickettsia*, (3) *Midichloria* 1, (4) *Borrelia*, (5) *Ehrlichia*, (6) Bacillales 1, (7) *Wolbachia*, (8) Rhizobiales 1, (9) *Rhodobaca*, (10) Bacillales 2, (11) *Nitriliruptor*, (12) Bacillales 3.

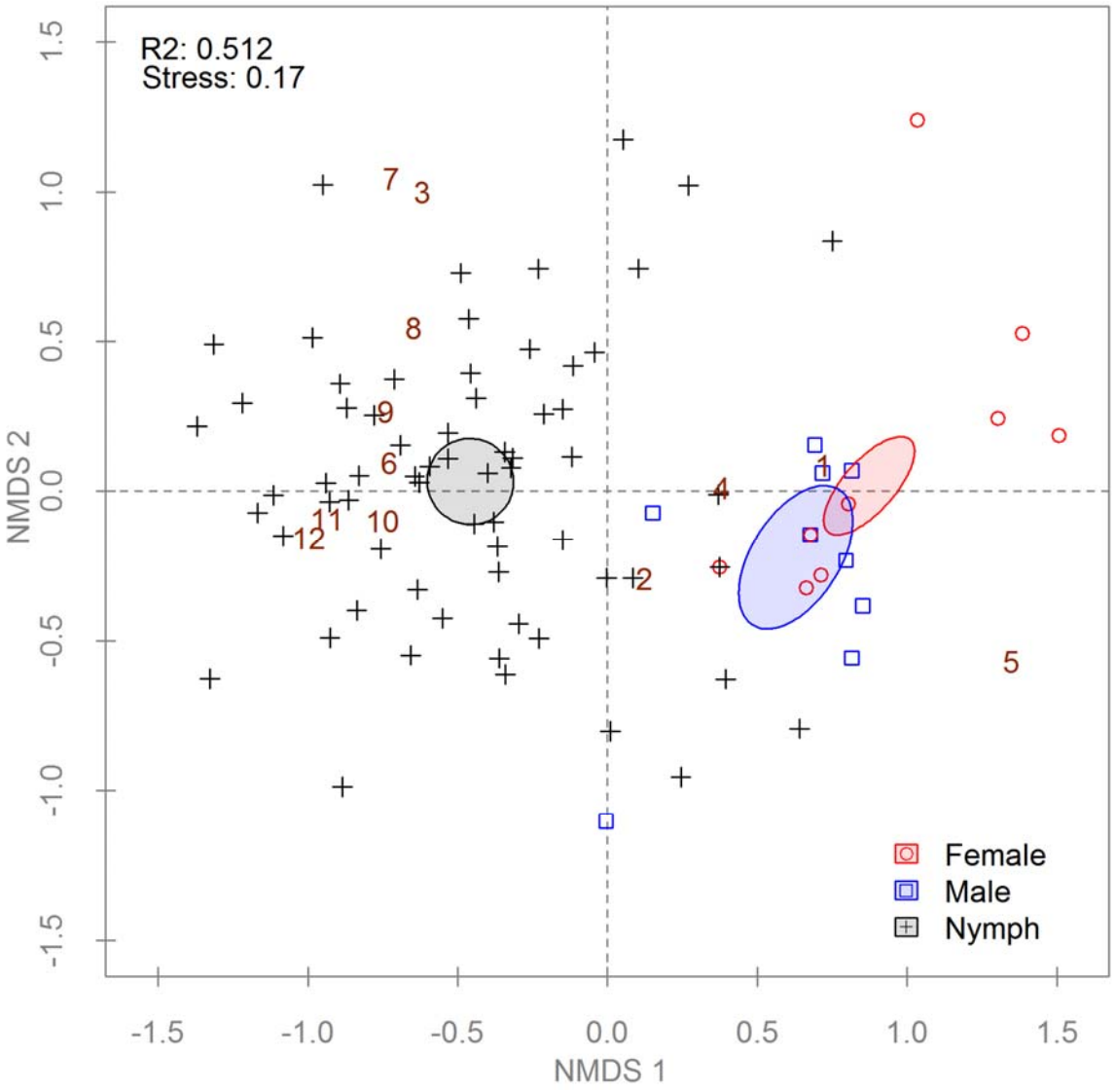

R2: 0.512
Stress: 0.17
NMDS 1
NMDS 2
-1.5
-1.0
-0.5
0.0
0.5
1.0
1.5
Female
Male
Nymph
1
2
3
4
5
6
7
8
9
10
11
12

Figure 3.S6 Mean relative abundance of operational taxonomic units (OTUs) from Georgian ticks by life stage. Significant differences ($p<0.05$) are labeled within rows by the letters that appear in cells. For example, within a row cells labeled A are significantly different from those labeled B or C but not cells labeled A. Empty cells do not differ from any group. Note the discontinuity between the white-black and yellow-red scales.

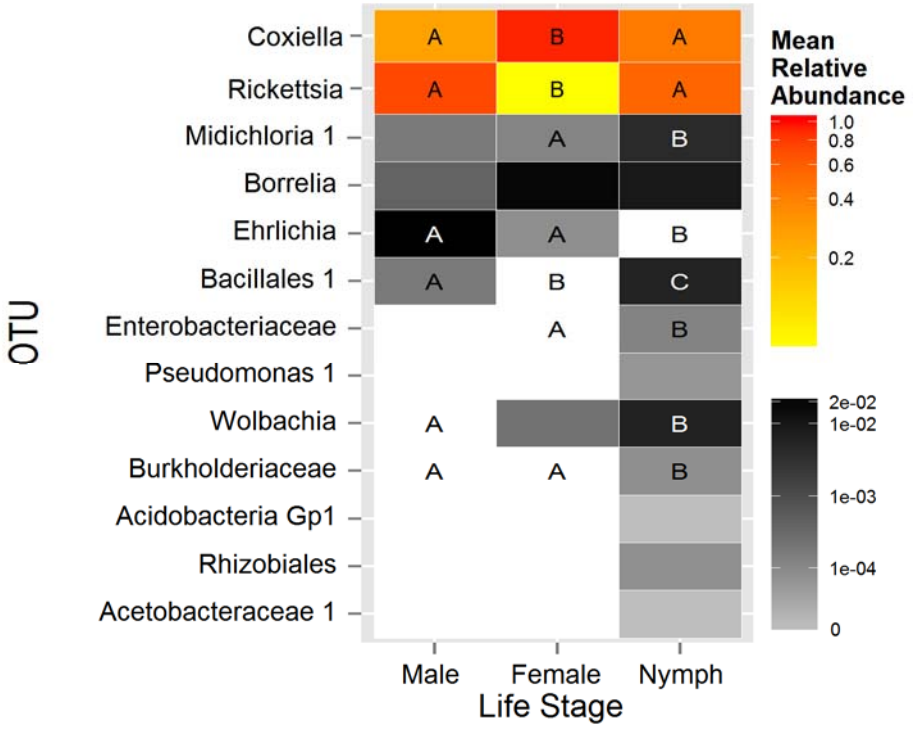
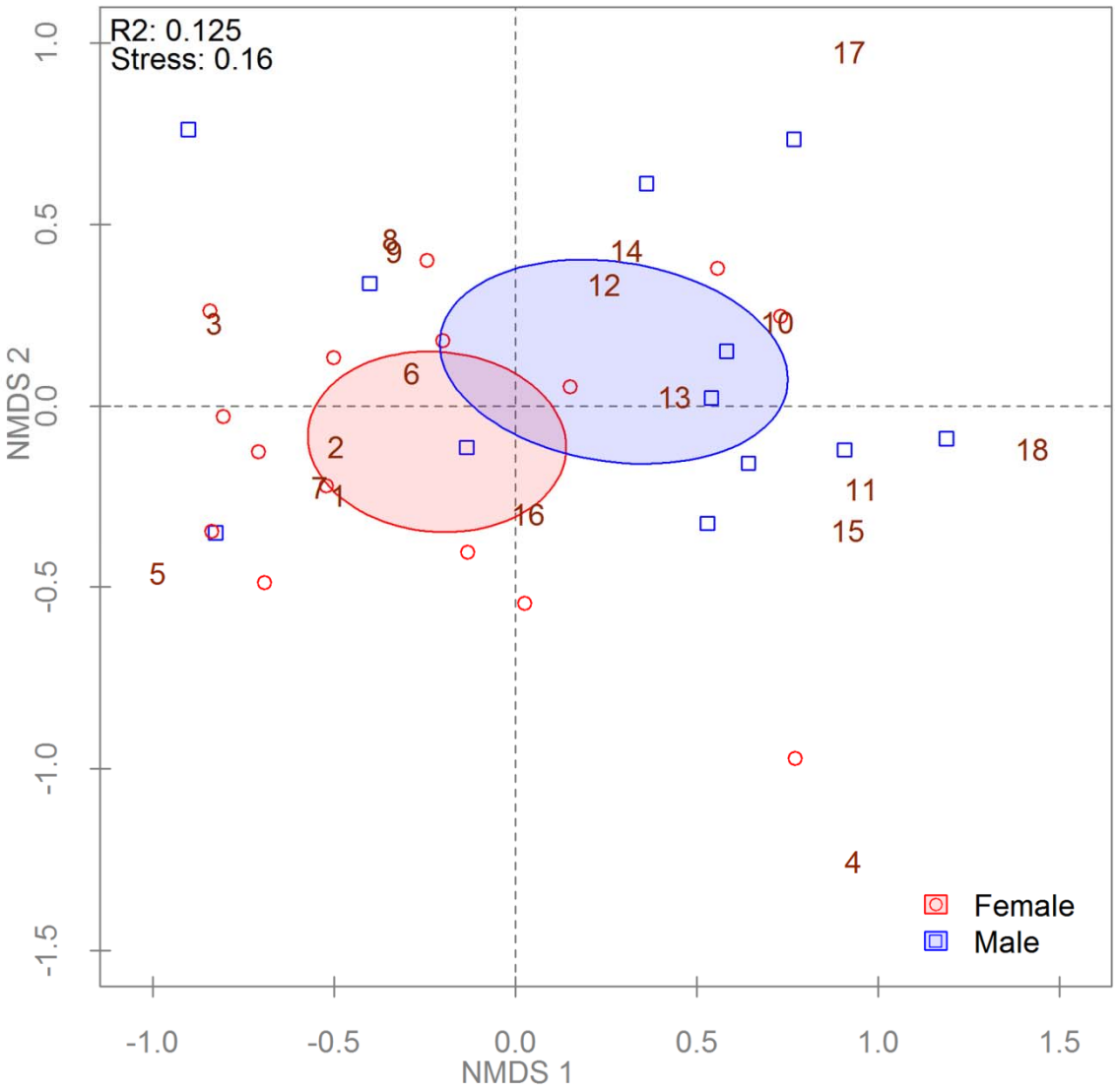


Figure 3.S7 Non-metric multi-dimensional scaling (NMDS) of a Jaccard distance matrix describing archival *Amblyomma americanum* bacterial communities. Each point symbolizes a single tick's community (n=27); some points may overlap completely. Point and ellipse colors indicate life stage; ellipses represent 95% confidence intervals around life stage centroids. Non-overlapping centroids are considered significantly different at α=0.05. R^2 values in the upper left corner of plots describe the amount of variation in the data set explained by the groupings. The stress value given is a measure of the disagreement between the rank order in the original data set and that in the NMDS (lower numbers indicate better agreement). Numbers indicate the species scores for the plot as follows: (1) *Coxiella*, (2) *Rickettsia*, (3) *Midichloria* 1, (4) *Borrelia*, (5) *Francisella*, (6) *Midichloria* 2, (7) *Ehrlichia*, (8) Enterobacteriaceae, (9) *Pseudomonas* 1, (10) Burkholderiaceae, (11) Acidobacteria Gp1, (12) *Pseudomonas* 2, (13) Rhizobiales 2, (14) *Bacillus*, (15) Acetobacteraceae 1, (16) Acetobacteraceae 2, (17) Bacteria 1, (18) Bacteria 2.


R2: 0.125
Stress: 0.16
NMDS 1
NMDS 2
-1.0
-0.5
0.0
0.5
1.0
1.5
-1.5
-1.0
-0.5
0.0
0.5
1.0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
Female
Male

3.7 Tables

Table 3.1 Life stage and geographic origin of *Amblyomma americanum* ticks used in bacterial community analyses.

Tick DNA Set	Collection Site (Year)	Nymph	Adult Male	Adult Female	Total
Archival	Bodie Island, NC (July 2002)	NT	5	5	10
Archival	Buxton Woods, NC (July 2002)	NT	4	5	9
Archival	Shelter Island, NY (July 2003)	NT	3	5	8
Archival Total		NT	12	15	27
Georgia	Panola Mountain, GA (2010-11)[§]	30	8	14	52
Georgia	Sweetwater Creek, GA (2010-11)*	36	4	12	52
Georgia Total		66	12	26	104

[§] Nymphs collected in July 2010; adults in July 2010 and May 2011.

*Nymphs collected in August 2010; adults in May 2011.

NT = Not tested

Table 3.2 Multivariate analysis of the effect of sex and life stage on the relatedness of *Amblyomma americanum* bacterial communities. Tick DNAs from the archival collection (n=27) were older, collected in a different geographical region, and processed using different protocols than tick DNAs from the Georgia collection (n=104).

			PERMANOVA			Dispersion Test	
Collection	Distance Metric	Metric Type	*F*	R^2	*p**	*F*	*p**
Georgia	Jaccard	Community composition	17.3	0.255	0.0020	12.5	0.0020
Georgia	Bray-Curtis	Community structure	24.9	0.330	0.0020	8.77	0.0020
Archival	Jaccard	Community composition	1.63	0.0611	0.026	0.533	0.96
Archival	Bray-Curtis	Community structure	8.28	0.249	0.0020	0.345	0.96

*Corrected for multiple comparisons using Holm's method; $\alpha = 0.05$.

Table 3.S1 Barcode sequences used to label eubacterial *rrs* gene variable regions 5-3 primers for sample multiplexing during pyrosequencing.

Barcode Name	Barcode Sequence
v2bBar8L	CACGC
v2bBar23L	CGCAAC
v2bBar174L	TGAAGC
v2bBar212L	TCACAC
v2bBar622L	ACGCGC
v2bBar72L	CCTCTC
v2bBar559L	AGACAC
v2bBar31L	CGACTC
v2bBar551L	AGCTTC
v2bBar1149L	AAGCCGC
v2bBar15L	CAAGAAC
v2bBar556L	AGTTGGC
v2bBar144L	TATCAAC
v2bBar575L	AGGCGGC
v2bBar48L	CGGTATC
v2bBar741L	ATACCAC
v2bBar228L	TCGCGGC
v2bBar807L	ATCTTAC
v2bBar1273L	AACCAGC
v2bBar441L	TTCGAGC
v2bBar1174L	AAGGTGC
v2bBar209L	TCTTGGC
v2bBar153L	TAATCTC
v2bBar213L	TCACCTC
v2bBar146L	TATTGAC
v2bBar554L	AGTCGAC
v2bBar646L	ACGGCTC
v2bBar158L	TGCGTTC

Barcode Name	Barcode Sequence
v2bBar207L	TCTCGAC
v2bBar601L	ACTCCTC
v2bBar419L	TTCATAC
v2bBar26L	CGTCGTC
v2bBar159L	TGCCGAAC
v2bBar147L	TATTCGTC
v2bBar141L	TAGGAATC
v2bBar119L	CCGGCCAC
v2bBar1267L	AACCTGGC
v2bBar637L	ACGAAGTC
v2bBar435L	TTCGTGGC
v2bBar433L	TTCGCGAC
v2bBar1156L	AAGAGTTC
v2bBar1173L	AAGGCCTC
v2bBar599L	ACTAATTC
v2bBar393L	TTGGAGGC
v2bBar350L	TTATCGGC
v2bBar1196L	AACTGTTC
v2bBar1031L	ATTCGTAC
v2bBar76L	CCTTCCGC
v2bBar1225L	AACGAGGC
v2bBar236L	TCGAGGAAC
v2bBar731L	ACCGGAAGC
v2bBar1250L	AACGGAGTC
v2bBar187L	TGGTTGGTC
v2bBar162L	TGTCCGGTC
v2bBar1292L	AACCGTGTC

Table 3.S2 454 sequencing statistics from Titanium FLX plates.

Plate ID	Plate Quarter ID	Polymerase*	Tick DNA Set	Experiment	Total No. DNAs	No. Raw Reads	No. High Quality Reads (% Raw)	No. Tick DNAs Retained[α]	No. Reads in Analysis[β]
1	A	AccuPrime	Archival	Life stage	50	114,250	85,465 (75%)	27	59,028
				Polymerase	50	114,250	71,552 (63%)	26	64,411
1	B	Platinum	Archival	Polymerase	50	82,843	53,175 (64%)	26	49,269
1	C	Platinum	Archival	Polymerase	50	93,834	57,912 (62%)	26	53,718
2	A	AccuPrime	Georgia	Life stage	50	82,087	55,560 (68%)	22	45,579
2	B	AccuPrime	Georgia	Life stage	50	71,821	47,664 (66%)	19	33,604
2	C	AccuPrime	Georgia	Life stage	50	64,743	36,245 (56%)	15	22,875
2	D	AccuPrime	Georgia	Life stage	50	60,428	44,004 (73%)	15	28,585
3	A	Platinum	Georgia	Life stage	50	276,596	177,476 (64%)	33	160,830

*Polymerases used include AccuPrime Taq High Fidelity and Platinum PCR SuperMix High Fidelity (Invitrogen, Carlsbad, California, USA).

α Discarded DNAs were inferior replicates of retained DNAs, were not sequenced to a depth of at least 1000 sequences, or belonged to groups that were insufficiently sampled to be included in the analysis.

β Total number of high quality sequences from retained ticks. Includes singleton OTUs.

Table 3.S3 Summary of previous reports of most abundant operational taxonomic units (OTUs) from *Amblyomma americanum*. The OTU number ranks the groups in order of decreasing abundance. In the classification column, the lowest taxon that all sequences within the OTU could be assigned to is provided first, followed by the class assignment in parentheses. Numbers are added to this column to differentiate OTUs classified to otherwise identical taxons; these are the same numbers used in the main paper. GenBank IDs are given for a representative sequence from each OTU, which was defined as the sequence with the smallest maximum distance from all other sequences in the OTU. The closest GenBank BLAST Hit was found by comparing the representative sequence against the GenBank database and reporting the hit with the highest identity. If there was no single best hit, this is indicated by the example designation before the GenBank ID number. The fifth column contains references for tick species from which these OTUs have previously been reported and is not exhaustive. Tick genus abbreviations are as follows: *A. Amblyomma, C. Carios, D. Dermacentor, Ha. Haemaphysalis, Hy. Hyalomma, I. Ixodes,* and *Rh. Rhipicephalus*.

OTU	Classification (Class)	GenBank ID	Closest GenBank BLAST Hit (GenBank ID, % identity)	Previous Reports from Ticks?	Notes
1	Genus *Coxiella* (Gammaproteobacteria)	KJ130495	*Coxiella* endosymbiont of *A. americanum* (AY939824.1, 99%)	*A. americanum* (Jasinskas *et al.* 2007)	Considered primary symbiont of *A. americanum* (Zhong *et al.* 2007). Related symbionts found in many other tick species (reviewed in Ahantarig *et al.* 2013).
2	Genus *Rickettsia* (Alphaproteobacteria)	KJ130496	*R. amblyommii* (NR_074471.1, 99%)	*A. americanum* (Burgdorfer *et al.* 1981) Other *Amblyomma*, eg. *A. cajennense* and *A. coelebs* (Labruna *et al.* 2004)	Many additional citations exist; some are given in the main article.

OTU	Classification (Class)	GenBank ID	Closest GenBank BLAST Hit (GenBank ID, % identity)	Previous Reports from Ticks?	Notes
4	Genus *Midichloria* 1 (Alphaproteobacteria)	KJ130497	*M. mitochondrii* strain AamerB (JQ678691.1, 100%)	*A. americanum* (Williams-Newkirk *et al.* 2012) *A. triste* (Venzal *et al.* 2008) *A. tuberculatum* (Epis *et al.* 2008) *D. andersoni* (Dergousoff & Chilton 2011) *Ha. punctate* (Epis *et al.* 2008) *Hy. marginatum* (Epis *et al.* 2008) *Hy. truncatum* (Epis *et al.* 2008) *I. ricinus* (Epis *et al.* 2008) *I. uriae* (Epis *et al.* 2008) *Rh. turanicus* (Epis *et al.* 2008) *Rh. bursa* (Epis *et al.* 2008)	A likely facultative symbiont (Pistone *et al.* 2012).

OTU	Classification (Class)	GenBank ID	Closest GenBank BLAST Hit (GenBank ID, % identity)	Previous Reports from Ticks?	Notes
5	Genus *Borrelia* (Spirochaetes)	KJ130498	*B. lonestari* (AY166715.1, 100%)	*A. americanum* (Barbour *et al.* 1996) *C. capensis* (Reeves *et al.* 2006)	*C. capensis*'s *Borrelia* has not been definitively identified as *B. lonestari*, but available evidence supports this identification.
6	Genus *Francisella* (Gammaproteobacteria)	KJ130499	*Francisella* spp. both pathogenic and symbiotic (e.g. JQ764629.1, 100%)	*A. americanum* (reviewed in Childs & Paddock 2003) *A. maculatum* (Scoles 2004) *Dermacentor* spp. (reviewed in Ahantarig *et al.* 2013) *Hy. marginatum* (Montagna *et al.* 2012)	Widespread symbiont of many tick species, but *A. americanum* is not known to carry a symbiotic form.

OTU	Classification (Class)	GenBank ID	Closest GenBank BLAST Hit (GenBank ID, % identity)	Previous Reports from Ticks?	Notes
7	Genus *Midichloria* 2 (Alphaproteobacteria)	KJ130500	*M. mitochondrii* strain AamerC (JQ678692.1, 100%)	*A. americanum* (Williams-Newkirk *et al.* 2012) *A. triste* (Venzal *et al.* 2008) *A. tuberculatum* (Epis *et al.* 2008) *D. andersoni* (Dergousoff & Chilton 2011) *Ha. punctate* (Epis *et al.* 2008) *Hy. marginatum* (Epis *et al.* 2008) *Hy. truncatum* (Epis *et al.* 2008) *I. ricinus* (Epis *et al.* 2008) *I. uriae* (Epis *et al.* 2008) *Rh. turanicus* (Epis *et al.* 2008) *Rh. bursa* (Epis *et al.* 2008)	A likely facultative symbiont (Pistone *et al.* 2012).

OTU	Classification (Class)	GenBank ID	Closest GenBank BLAST Hit (GenBank ID, % identity)	Previous Reports from Ticks?	Notes
8	Genus *Ehrlichia* (Alphaproteobacteria)	KJ130501	*Ehrlichia* spp. (e.g. U96436.1, 100%)	*A. americanum* (reviewed in Childs & Paddock 2003) *D. variabilis* (Johnson *et al.* 1998) *Rh. sanguineus* (Groves *et al.* 1975)	
9	Order Bacillales 1 (Bacilli)	KJ130502	*Bacillus* spp. (e.g. EU530599.1, 100%)	*A. americanum* (Heise *et al.* 2010; Yuan 2010) *A. testudinarium* (Nakao *et al.* 2013) *Ha. formosensis* (Nakao *et al.* 2013) *I. ovatus* (Nakao *et al.* 2013) *I. persulcatus* (Nakao *et al.* 2013) *I. ricinus* (Carpi *et al.* 2011; Nakao *et al.* 2013) *I. scapularis* (Yuan 2010) *Rh. microplus* (Andreotti *et al.* 2011)	Common environmental bacteria, but also known symbionts of other arthropods. Notably *Bacillus* protects against fungal pathogens in bees (reviewed in Kaltenpoth & Engl 2013).

OTU	Classification (Class)	GenBank ID	Closest GenBank BLAST Hit (GenBank ID, % identity)	Previous Reports from Ticks?	Notes
10	Family Enterobacteriaceae (Gammaproteobacteria)	KJ130503	Many genera; 60% of sequences classified to *Yersinia* (e.g. BX936398.1, 100%)	*A. americanum* (Clay *et al.* 2008; Heise *et al.* 2010; Yuan 2010) *I. scapularis* (Yuan 2010)	Many insect symbionts belong to this family (e.g. *Sodalis* (Dale & Maudlin 1999), *Wigglesworthia* (Pais *et al.* 2008), *Buchnera* (Douglas 1998)), but it also contains common and diverse environmental bacteria (Degelmann *et al.* 2009) and pathogens (Perry & Fetherston 1997).
11	Genus *Pseudomonas* 1 (Gammaproteobacteria)	KJ130504	*Pseudomonas* spp. (e.g. KF147045.1, 100%)	*A. americanum* (Clay *et al.* 2008; Heise *et al.* 2010) *I. ricinus* (Carpi *et al.* 2011) *Rh. microplus* (Andreotti *et al.* 2011)	Common environmental bacteria. Demonstrated *in vitro* lysis of vector-borne parasites by some species (Mercado & Colon-Whitt 1982)

OTU	Classification (Class)	GenBank ID	Closest GenBank BLAST Hit (GenBank ID, % identity)	Previous Reports from Ticks?	Notes
12	Genus *Wolbachia* (Alphaproteobacteria)	KJ130505	*Wolbachia* strains (e.g. KC522606.1, 99%)	*A. americanum* (Zhang *et al.* 2011) *I. ricinus* (Carpi *et al.* 2011; Plantard *et al.* 2012) *I. scapularis* (Benson *et al.* 2004) *Rh. microplus* (Andreotti *et al.* 2011) *Rh. sanguineus* (Inokuma *et al.* 2000)	Widespread symbiont of many arthropods known for reproductive manipulation of hosts and protective abilities (reviewed in Werren *et al.* 2008).
13	Order Rhizobiales 1 (Alphaproteobacteria)	KJ130506	Many genera; 52% of sequences classified to Bradyrhizobiaceae (e.g. GU125653.2, 99%)	*A. americanum* (Menchaca *et al.* 2013) *I. ricinus* (Carpi *et al.* 2011)	Common nitrogen-fixing plant symbionts (reviewed in Kneip *et al.* 2007), but also found in association with ants (van Borm *et al.* 2002; Kautz *et al.* 2013).
14	Family Burkholderiaceae (Betaproteobacteria)	KJ130507	Many genera; 78% of sequences classified to *Burkholderia* (e.g. KF059270.1, 100%)	*A. americanum* (Clay *et al.* 2008; Yuan 2010) *I. ricinus* (Carpi *et al.* 2011) *I. scapularis* (Yuan 2010; Hawlena *et al.* 2013)	*Burkholderia* exist as free-living (Salles *et al.* 2002), plant symbionts (Compant *et al.* 2008), arthropod symbionts (Kikuchi *et al.* 2012), and vertebrate pathogens (Valvano *et al.* 2005).

OTU	Classification (Class)	GenBank ID	Closest GenBank BLAST Hit (GenBank ID, % identity)	Previous Reports from Ticks?	Notes
15	Family "Gp1" *incertae sedis* (Acidobacteria)	KJ130508	Unidentified clones (e.g. HQ598755.1, 100%)	*A. variegatum* (Nakao *et al.* 2013)	"Gp1" bacteria are iron oxidizers previously found in environmental samples, like iron-rich ground water (Navarro-Noya *et al.* 2013).
16	Genus *Rhodobaca* (Alphaproteobacteria)	KJ130509	Unidentified clones (e.g. AB297418.1, 100%)	No	Genus containing anoxygenic purple nonsulfur bacteria usually associated with lake waters and soil (Milford *et al.* 2000). Never previously reported in association with ixodid ticks, but its closest phylogenetic relatives in the genus *Rhodobacter* have been (*I. scapularis* (Yuan 2010), *Rh. microplus* (Andreotti *et al.* 2011), and *I. ricinus* (Carpi *et al.* 2011)).

OTU	Classification (Class)	GenBank ID	Closest GenBank BLAST Hit (GenBank ID, % identity)	Previous Reports from Ticks?	Notes
17	Order Bacillales 2 (Bacilli)	KJ130510	Unidentified clones (e.g. JF968215.1, 100%)	*A. americanum* (Heise *et al.* 2010; Yuan 2010) *A. testudinarium* (Nakao *et al.* 2013) *Ha. formosensis* (Nakao *et al.* 2013) *I. ovatus* (Nakao *et al.* 2013) *I. persulcatus* (Nakao *et al.* 2013) *I. ricinus* (Carpi *et al.* 2011; Nakao *et al.* 2013) *I. scapularis* (Yuan 2010) *Rh. microplus* (Andreotti *et al.* 2011)	Common environmental bacteria, but also known symbionts of other arthropods. Notably *Bacillus* protects against fungal pathogens in bees (reviewed in Kaltenpoth & Engl 2013).
18	Genus *Pseudomonas* 2 (Gammaproteobacteria)	KJ130511	*Pseudomonas* spp. (e.g. GQ232466.1, 100%)	*A. americanum* (Clay *et al.* 2008; Heise *et al.* 2010) *I. ricinus* (Carpi *et al.* 2011) *Rh. microplus* (Andreotti *et al.* 2011)	Common environmental bacteria. Demonstrated in vitro lysis of vector-borne parasites by some species (Mercado & Colon-Whitt 1982).

OTU	Classification (Class)	GenBank ID	Closest GenBank BLAST Hit (GenBank ID, % identity)	Previous Reports from Ticks?	Notes
19	Genus *Nitriliruptor* (Actinobacteria)	KJ130512	*N. alkaliphilus* (NR_044203.1, 100%)	References to unclassified Actinobacteria: *A. americanum* (Heise *et al.* 2010; Yuan 2010) *A. variegatum* (Nakao *et al.* 2013) *D. variabilis* (Hawlena *et al.* 2013) *I. scapularis* (Carpi *et al.* 2011; Hawlena *et al.* 2013)	*Nitriliruptor* has been previously documented from soil samples (Neilson *et al.* 2012; Brown & Jumpponen 2013). Actinobacteria are common environmental bacteria and eukaryote-associates (reviewed in Ventura *et al.* 2007).
20	Genus *Methylobacterium* (Alphaproteobacteria)	KJ130513	*Methylobacterium* spp. (e.g. AB698724.1, 100%)	*A. americanum* (Heise *et al.* 2010) *I. ricinus* (Carpi *et al.* 2011)	Common environmental bacteria associated with soil and plants (e.g. Eller & Frenzel 2001). It is known from the human mouth (Anesti *et al.* 2005) and skin (Anesti *et al.* 2004) microbiome and may also be pathogenic (Lai *et al.* 2011).

OTU	Classification (Class)	GenBank ID	Closest GenBank BLAST Hit (GenBank ID, % identity)	Previous Reports from Ticks?	Notes
21	Order Rhizobiales 2 (Alphaproteobacteria)	KJ130514	Unidentified clones (e.g. JQ047240.1, 100%)	*A. americanum* (Clay *et al.* 2008) *D. variabilis* (Hawlena *et al.* 2013) *I. ricinus* (Carpi *et al.* 2011) *I. scapularis* (Hawlena *et al.* 2013) *Rh. microplus* (Andreotti *et al.* 2011)	Common nitrogen-fixing plant symbionts (reviewed in Kneip *et al.* 2007), but also found in association with ants (van Borm *et al.* 2002; Kautz *et al.* 2013).
22	Family Sphingomonadaceae (Alphaproteobacteria)	KJ130515	*Sphingomonas* spp.; 97% of sequences classified to *Sphingomonas* (e.g. KF542913.1, 100%)	*A. americanum* (Yuan 2010; Menchaca *et al.* 2013) *I. ricinus* (Carpi *et al.* 2011) *I. scapularis* (Yuan 2010)	*Sphingomonas* spp. are known arthropod associates, including mosquitoes (*Aedes aegypti*) (Terenius *et al.* 2012) and termites (*Macrotermes natalensis* and *Nasutitermes* sp.) (Aylward *et al.* 2013).

OTU	Classification (Class)	GenBank ID	Closest GenBank BLAST Hit (GenBank ID, % identity)	Previous Reports from Ticks?	Notes
23	Order Bacillales 3 (Bacilli)	KJ130516	Unidentified clones (e.g. AB637120.1, 100%)	*A. americanum* (Heise *et al.* 2010; Yuan 2010) *A. testudinarium* (Nakao *et al.* 2013) *Ha. formosensis* (Nakao *et al.* 2013) *I. ovatus* (Nakao *et al.* 2013) *I. persulcatus* (Nakao *et al.* 2013) *I. ricinus* (Carpi *et al.* 2011; Nakao *et al.* 2013) *I. scapularis* (Yuan 2010) *Rh. microplus* (Andreotti *et al.* 2011)	Common environmental bacteria, but also known symbionts of other arthropods. Notably *Bacillus* protects against fungal pathogens in bees (reviewed in Kaltenpoth & Engl 2013).
24	Order Actinomycetales (Actinobacteria)	KJ130517	*Mycobacterium* spp.; 99% of sequences classified to *Mycobacterium* (e.g. KF224994.1, 100%)	*A. americanum* (Yuan 2010) *D. variabilis* (Hawlena *et al.* 2013) *I. ricinus* (Carpi *et al.* 2011) *I. scapularis* (Yuan 2010)	Common environmental bacteria (soil, water). Also common pathogens of wildlife (birds, mammals) (Gronesova *et al.* 2008).

Table 3.S4 Metastats results comparing the differential abundance of bacterial OTUs between Amblyomma americanum DNA collections. Only results significant at q<0.05 and p<0.05 are given. Bootstrap values are given in parentheses for each level of the taxonomic assignment of each OTU.

OTU	No. Georgia Ticks Positive	No. Archival Ticks Positive	Total Sequence Abundance	Georgia Mean Relative Sequence Abundance	Archival Mean Relative Sequence Abundance	p	q	Known Tick Associate?	Taxonomy
Otu001	102	27	63180	0.526	0.291	0.005	0.021	Yes	Bacteria(100); Proteobacteria(100); Gammaproteobacteria(100); Legionellales(100); Coxiellaceae(100); Coxiella(100);
Otu007	0	3	398	0.000	0.014	0.009	0.036	Yes	Bacteria(100); Proteobacteria(100); Alphaproteobacteria(100); Rickettsiales(100); Rickettsiaceae(100); Midichloria(100);

OTU	No. Georgia Ticks Positive	No. Archival Ticks Positive	Total Sequence Abundance	Georgia Mean Relative Sequence Abundance	Archival Mean Relative Sequence Abundance	p	q	Known Tick Associate?	Taxonomy
Otu009	56	0	383	0.004	0.000	0.001	0.005	No	Bacteria(100); Firmicutes(100); Bacilli(100); Bacillales(100); Bacillaceae_1(67); unclassified(67);
Otu012	10	0	416	0.004	0.000	0.001	0.005	Yes	Bacteria(100); Proteobacteria(100); Alphaproteobacteria(100); Rickettsiales(100); Anaplasmataceae(100); Wolbachia(100);
Otu013	36	0	152	0.001	0.000	0.001	0.005	No	Bacteria(100); Proteobacteria(100); Alphaproteobacteria(100); Rhizobiales(100); Bradyrhizobiaceae(52); unclassified(52);

OTU	No. Georgia Ticks Positive	No. Archival Ticks Positive	Total Sequence Abundance	Georgia Mean Relative Sequence Abundance	Archival Mean Relative Sequence Abundance	p	q	Known Tick Associate?	Taxonomy
Otu016	39	0	133	0.001	0.000	0.001	0.005	No	Bacteria(100); Proteobacteria(100); Alphaproteobacteria(100); Rhodobacterales(100); Rhodobacteraceae(100); Rhodobaca(100);
Otu017	36	0	111	0.001	0.000	0.001	0.005	No	Bacteria(100); Firmicutes(100); Bacilli(100); Bacillales(100); unclassified(100); unclassified(100);
Otu019	35	0	130	0.001	0.000	0.001	0.005	No	Bacteria(100); Actinobacteria(100); Actinobacteria(100); Nitriliruptorales(100); Nitriliruptoraceae(100); Nitriliruptor(100);

OTU	No. Georgia Ticks Positive	No. Archival Ticks Positive	Total Sequence Abundance	Georgia Mean Relative Sequence Abundance	Archival Mean Relative Sequence Abundance	p	q	Known Tick Associate?	Taxonomy
Otu020	9	18	125	0.000	0.004	0.001	0.005	No	Bacteria(100); Proteobacteria(100); Alphaproteobacteria(100); Rhizobiales(100); Methylobacteriaceae(100); Methylobacterium(100);
Otu021	2	12	196	0.000	0.007	0.013	0.047	No	Bacteria(100); Proteobacteria(100); Alphaproteobacteria(100); Rhizobiales(100); unclassified(99); unclassified(99);

OTU	No. Georgia Ticks Positive	No. Archival Ticks Positive	Total Sequence Abundance	Georgia Mean Relative Sequence Abundance	Archival Mean Relative Sequence Abundance	*p*	*q*	Known Tick Associate?	Taxonomy
Otu023	29	0	110	0.001	0.000	0.001	0.005	No	Bacteria(100); Firmicutes(100); Bacilli(100); Bacillales(100); Sporolactobacillaceae (99); Sporolactobacillaceae _incertae_sedis(99);
Otu024	4	13	103	0.000	0.003	0.011	0.042	No	Bacteria(100); Actinobacteria(100); Actinobacteria(100); Actinomycetales(100); Mycobacteriaceae(99); Mycobacterium(99);

OTU	No. Georgia Ticks Positive	No. Archival Ticks Positive	Total Sequence Abundance	Georgia Mean Relative Sequence Abundance	Archival Mean Relative Sequence Abundance	*p*	*q*	Known Tick Associate?	Taxonomy
Otu028	1	8	46	0.000	0.002	0.001	0.005	No	Bacteria(100); Acidobacteria(100); Acidobacteria_Gp1(100); Acidobacteria_Gp1_order_incertae_sedis(100); Acidobacteria_Gp1_family_incertae_sedis(100); Gp1(100);
Otu030	0	9	76	0.000	0.003	0.001	0.005	No	Bacteria(100); Proteobacteria(100); Alphaproteobacteria(100); Rhizobiales(100); unclassified(93); unclassified(93);

OTU	No. Georgia Ticks Positive	No. Archival Ticks Positive	Total Sequence Abundance	Georgia Mean Relative Sequence Abundance	Archival Mean Relative Sequence Abundance	p	q	Known Tick Associate?	Taxonomy
Otu031	1	7	27	0.000	0.001	0.001	0.005	No	Bacteria(100); Proteobacteria(100); Alphaproteobacteria(100); Rhodospirillales(100); Acetobacteraceae(100); unclassified(100);
Otu032	0	13	70	0.000	0.002	0.001	0.005	No	Bacteria(100); Proteobacteria(100); Alphaproteobacteria(100); Rhizobiales(100); unclassified(100); unclassified(100);

OTU	No. Georgia Ticks Positive	No. Archival Ticks Positive	Total Sequence Abundance	Georgia Mean Relative Sequence Abundance	Archival Mean Relative Sequence Abundance	p	q	Known Tick Associate?	Taxonomy
Otu034	1	7	45	0.000	0.002	0.001	0.005	No	Bacteria(100); Proteobacteria(100); Alphaproteobacteria(100); Rhizobiales(100); Methylobacteriaceae(88); Methylobacterium(85);
Otu035	0	2	34	0.000	0.001	0.001	0.005	No	Bacteria(100); Firmicutes(100); Bacilli(100); Bacillales(100); Bacillaceae_1(100); Bacillus(100);
Otu036	0	4	15	0.000	0.001	0.001	0.005	No	Bacteria(100); Proteobacteria(100); Alphaproteobacteria(100); Rhodospirillales(100); Acetobacteraceae(100); unclassified;

OTU	No. Georgia Ticks Positive	No. Archival Ticks Positive	Total Sequence Abundance	Georgia Mean Relative Sequence Abundance	Archival Mean Relative Sequence Abundance	p	q	Known Tick Associate?	Taxonomy
Otu040	0	4	46	0.000	0.002	0.002	0.010	No	Bacteria(100); Proteobacteria(100); Alphaproteobacteria(100); Rhodospirillales(100); Acetobacteraceae(100); unclassified(100);
Otu044	0	7	23	0.000	0.001	0.001	0.005	No	Bacteria(100); Actinobacteria(100); Actinobacteria(100); Actinomycetales(100); Microbacteriaceae(100); unclassified(95);
Otu045	0	3	7	0.000	0.000	0.003	0.014	No	Bacteria(100); Proteobacteria(100); Alphaproteobacteria(100); Rhizobiales(100); Beijerinckiaceae(90); unclassified(83);

OTU	No. Georgia Ticks Positive	No. Archival Ticks Positive	Total Sequence Abundance	Georgia Mean Relative Sequence Abundance	Archival Mean Relative Sequence Abundance	p	q	Known Tick Associate?	Taxonomy
Otu046	0	4	23	0.000	0.001	0.002	0.010	No	Bacteria(100); Bacteroidetes(100); Sphingobacteria(100); Sphingobacteriales(100); Cytophagaceae(99); Hymenobacter(83);
Otu047	0	4	7	0.000	0.000	0.001	0.005	No	Bacteria(100); Acidobacteria(100); Acidobacteria_Gp1(100); Acidobacteria_Gp1_order_incertae_sedis(100); Acidobacteria_Gp1_family_incertae_sedis(100); Gp1(100);

OTU	No. Georgia Ticks Positive	No. Archival Ticks Positive	Total Sequence Abundance	Georgia Mean Relative Sequence Abundance	Archival Mean Relative Sequence Abundance	p	q	Known Tick Associate?	Taxonomy
Otu048	0	6	39	0.000	0.001	0.001	0.005	No	Bacteria(100); Proteobacteria(100); Alphaproteobacteria(100); Rhizobiales(100); unclassified(100); unclassified(100);
Otu050	8	0	19	0.000	0.000	0.001	0.005	No	Bacteria(100); Proteobacteria(100); Alphaproteobacteria(100); Rhizobiales(100); Phyllobacteriaceae(100); unclassified(73);
Otu051	0	5	19	0.000	0.001	0.001	0.005	No	Bacteria(100); Proteobacteria(100); Alphaproteobacteria(100); unclassified(99); unclassified(99); unclassified(99);

OTU	No. Georgia Ticks Positive	No. Archival Ticks Positive	Total Sequence Abundance	Georgia Mean Relative Sequence Abundance	Archival Mean Relative Sequence Abundance	p	q	Known Tick Associate?	Taxonomy
Otu053	0	4	13	0.000	0.000	0.002	0.010	No	Bacteria(100); Actinobacteria(100); Actinobacteria(100); Actinomycetales(100); Nocardiaceae(100); unclassified(98);
Otu055	0	1	9	0.000	0.000	0.011	0.042	No	Bacteria(100); Proteobacteria(100); Alphaproteobacteria(100); Caulobacterales(100); Caulobacteraceae(100); Phenylobacterium(72);

OTU	No. Georgia Ticks Positive	No. Archival Ticks Positive	Total Sequence Abundance	Georgia Mean Relative Sequence Abundance	Archival Mean Relative Sequence Abundance	p	q	Known Tick Associate?	Taxonomy
Otu056	0	2	28	0.000	0.001	0.003	0.014	No	Bacteria(100); Acidobacteria(100); Acidobacteria_Gp1(100); Acidobacteria_Gp1_order_incertae_sedis(100); Acidobacteria_Gp1_family_incertae_sedis(100); Gp1(100);
Otu058	0	3	20	0.000	0.001	0.001	0.005	No	Bacteria(100); Actinobacteria(100); Actinobacteria(100); Actinomycetales(100); unclassified(100); unclassified(100);
Otu059	6	0	8	0.000	0.000	0.004	0.018	Yes	Bacteria(100); Proteobacteria(100); Gammaproteobacteria(100); Legionellales(100); Coxiellaceae(100); Coxiella(100);

OTU	No. Georgia Ticks Positive	No. Archival Ticks Positive	Total Sequence Abundance	Georgia Mean Relative Sequence Abundance	Archival Mean Relative Sequence Abundance	p	q	Known Tick Associate?	Taxonomy
Otu064	0	4	4	0.000	0.000	0.001	0.005	No	Bacteria(100); Actinobacteria(100); Actinobacteria(100); Actinomycetales(100); Pseudonocardiaceae(71); unclassified(71);
Otu065	0	3	7	0.000	0.000	0.001	0.005	No	Bacteria(100); Proteobacteria(100); Gammaproteobacteria(100); Pseudomonadales(100); Moraxellaceae(100); Acinetobacter(100);

OTU	No. Georgia Ticks Positive	No. Archival Ticks Positive	Total Sequence Abundance	Georgia Mean Relative Sequence Abundance	Archival Mean Relative Sequence Abundance	p	q	Known Tick Associate?	Taxonomy
Otu066	0	2	10	0.000	0.000	0.004	0.018	No	Bacteria(100); Acidobacteria(100); Acidobacteria_Gp1(100); Acidobacteria_Gp1_order_incertae_sedis(100); Acidobacteria_Gp1_family_incertae_sedis(100); Gp1(100);
Otu067	0	2	2	0.000	0.000	0.001	0.005	No	Bacteria(100); Proteobacteria(100); Alphaproteobacteria(100); Rhodospirillales(100); Acetobacteraceae(100); unclassified(100);

OTU	No. Georgia Ticks Positive	No. Archival Ticks Positive	Total Sequence Abundance	Georgia Mean Relative Sequence Abundance	Archival Mean Relative Sequence Abundance	p	q	Known Tick Associate?	Taxonomy
Otu069	0	3	5	0.000	0.000	0.001	0.005	No	Bacteria(100); Bacteroidetes(100); Sphingobacteria(100); Sphingobacteriales(100); Sphingobacteriaceae(100); Pedobacter(100);
Otu070	0	2	10	0.000	0.000	0.007	0.029	No	Bacteria(100); Proteobacteria(100); Alphaproteobacteria(100); Rhizobiales(100); Rhizobiaceae(100); Rhizobium(100);
Otu075	0	4	9	0.000	0.000	0.001	0.005	No	Bacteria(100); Proteobacteria(100); Alphaproteobacteria(100); Rhizobiales(100); Rhizobiaceae(91); Rhizobium(87);

OTU	No. Georgia Ticks Positive	No. Archival Ticks Positive	Total Sequence Abundance	Georgia Mean Relative Sequence Abundance	Archival Mean Relative Sequence Abundance	p	q	Known Tick Associate?	Taxonomy
Otu076	0	3	6	0.000	0.000	0.001	0.005	No	Bacteria(100); unclassified(100); unclassified(100); unclassified(100); unclassified(100); unclassified(100);
Otu077	0	2	7	0.000	0.000	0.005	0.021	No	Bacteria(100); Bacteroidetes(100); Sphingobacteria(100); Sphingobacteriales(100); Cytophagaceae(100); Hymenobacter(100);

OTU	No. Georgia Ticks Positive	No. Archival Ticks Positive	Total Sequence Abundance	Georgia Mean Relative Sequence Abundance	Archival Mean Relative Sequence Abundance	p	q	Known Tick Associate?	Taxonomy
Otu078	0	2	6	0.000	0.000	0.003	0.014	No	Bacteria(100); Acidobacteria(100); Acidobacteria_Gp3(100); Acidobacteria_Gp3_order_incertae_sedis(100); Acidobacteria_Gp3_family_incertae_sedis(100); Gp3(100);
Otu079	0	3	10	0.000	0.000	0.003	0.014	No	Bacteria(100); Bacteroidetes(100); Sphingobacteria(100); Sphingobacteriales(100); Sphingobacteriaceae(100); Sphingobacterium(100);

OTU	No. Georgia Ticks Positive	No. Archival Ticks Positive	Total Sequence Abundance	Georgia Mean Relative Sequence Abundance	Archival Mean Relative Sequence Abundance	p	q	Known Tick Associate?	Taxonomy
Otu080	0	2	3	0.000	0.000	0.001	0.005	No	Bacteria(100); Bacteroidetes(100); Sphingobacteria(100); Sphingobacteriales(100); Chitinophagaceae(100); unclassified(100);
Otu084	0	5	5	0.000	0.000	0.005	0.021	No	Bacteria(100); Proteobacteria(100); Alphaproteobacteria(100); Rhizobiales(100); Rhizobiaceae(67); Rhizobium(56);
Otu088	0	4	5	0.000	0.000	0.001	0.005	No	Bacteria(100); Proteobacteria(100); Alphaproteobacteria(100); Rhizobiales(100); unclassified(100); unclassified(100);

OTU	No. Georgia Ticks Positive	No. Archival Ticks Positive	Total Sequence Abundance	Georgia Mean Relative Sequence Abundance	Archival Mean Relative Sequence Abundance	p	q	Known Tick Associate?	Taxonomy
Otu089	0	3	4	0.000	0.000	0.014	0.049	No	Bacteria(100); Proteobacteria(100); Alphaproteobacteria(100); Rhodospirillales(100); Acetobacteraceae(100); Roseomonas(100);
Otu092	0	2	6	0.000	0.000	0.001	0.005	No	Bacteria(100); Proteobacteria(100); Alphaproteobacteria(100); Rhizobiales(100); unclassified(54); unclassified(54);

OTU	No. Georgia Ticks Positive	No. Archival Ticks Positive	Total Sequence Abundance	Georgia Mean Relative Sequence Abundance	Archival Mean Relative Sequence Abundance	p	q	Known Tick Associate?	Taxonomy
Otu095	0	4	9	0.000	0.000	0.002	0.010	No	Bacteria(100); Proteobacteria(100); Alphaproteobacteria(100); Rhizobiales(100); Aurantimonadaceae(100); Aurantimonas(100);
Otu096	0	3	8	0.000	0.000	0.001	0.005	No	Bacteria(100); Actinobacteria(100); Actinobacteria(100); Actinomycetales(100); Corynebacteriaceae(100); Corynebacterium(100);

OTU	No. Georgia Ticks Positive	No. Archival Ticks Positive	Total Sequence Abundance	Georgia Mean Relative Sequence Abundance	Archival Mean Relative Sequence Abundance	p	q	Known Tick Associate?	Taxonomy
Otu098	0	1	10	0.000	0.000	0.010	0.039	No	Bacteria(100); Proteobacteria(100); Alphaproteobacteria(100); unclassified(65); unclassified(65); unclassified(65);
Otu101	0	2	4	0.000	0.000	0.006	0.025	No	Bacteria(100); Actinobacteria(100); Actinobacteria(100); Actinomycetales(100); Nocardiaceae(100); Rhodococcus(100);
Otu102	0	4	4	0.000	0.000	0.001	0.005	No	Bacteria(100); Actinobacteria(100); Actinobacteria(100); Actinomycetales(100); Propionibacteriaceae(100); Friedmanniella(79);

OTU	No. Georgia Ticks Positive	No. Archival Ticks Positive	Total Sequence Abundance	Georgia Mean Relative Sequence Abundance	Archival Mean Relative Sequence Abundance	p	q	Known Tick Associate?	Taxonomy
Otu104	0	2	3	0.000	0.000	0.001	0.005	No	Bacteria(100); Proteobacteria(100); Betaproteobacteria(100); Burkholderiales(100); unclassified(93); unclassified(93);
Otu108	0	3	6	0.000	0.000	0.001	0.005	No	Bacteria(100); Proteobacteria(100); Alphaproteobacteria(100); Sphingomonadales(100); Sphingomonadaceae(100); Sphingomonas(100);

OTU	No. Georgia Ticks Positive	No. Archival Ticks Positive	Total Sequence Abundance	Georgia Mean Relative Sequence Abundance	Archival Mean Relative Sequence Abundance	p	q	Known Tick Associate?	Taxonomy
Otu110	0	3	8	0.000	0.000	0.001	0.005	No	Bacteria(100); Bacteroidetes(100); Flavobacteria(100); Flavobacteriales(100); Flavobacteriaceae(100); Chryseobacterium(100);
Otu116	0	4	8	0.000	0.000	0.001	0.005	No	Bacteria(100); unclassified(100); unclassified(100); unclassified(100); unclassified(100); unclassified(100);
Otu117	0	2	6	0.000	0.000	0.001	0.005	No	Bacteria(100); Firmicutes(100); Clostridia(100); Clostridiales(100); Clostridiales_Incertae_Sedis_XI(100); Finegoldia(100);

OTU	No. Georgia Ticks Positive	No. Archival Ticks Positive	Total Sequence Abundance	Georgia Mean Relative Sequence Abundance	Archival Mean Relative Sequence Abundance	p	q	Known Tick Associate?	Taxonomy
Otu120	0	2	4	0.000	0.000	0.001	0.005	No	Bacteria(100); Proteobacteria(100); Alphaproteobacteria(100); Rhizobiales(100); Beijerinckiaceae(100); unclassified(100);
Otu121	0	2	3	0.000	0.000	0.001	0.005	No	Bacteria(100); Actinobacteria(100); Actinobacteria(100); Actinomycetales(100); Pseudonocardiaceae(90); Pseudonocardia(90);
Otu124	1	0	2	0.000	0.000	0.014	0.049	No	Bacteria(100); Actinobacteria(100); Actinobacteria(100); Actinomycetales(100); unclassified(100); unclassified(100);

OTU	No. Georgia Ticks Positive	No. Archival Ticks Positive	Total Sequence Abundance	Georgia Mean Relative Sequence Abundance	Archival Mean Relative Sequence Abundance	p	q	Known Tick Associate?	Taxonomy
Otu128	0	2	2	0.000	0.000	0.001	0.005	No	Bacteria(100); Proteobacteria(100); Alphaproteobacteria(100); Sphingomonadales(100); Sphingomonadaceae(100); Sphingomonas(100);
Otu130	0	2	3	0.000	0.000	0.001	0.005	No	Bacteria(100); Proteobacteria(100); Alphaproteobacteria(100); Sphingomonadales(100); Sphingomonadaceae(100); Sphingomonas(89);

OTU	No. Georgia Ticks Positive	No. Archival Ticks Positive	Total Sequence Abundance	Georgia Mean Relative Sequence Abundance	Archival Mean Relative Sequence Abundance	p	q	Known Tick Associate?	Taxonomy
Otu131	0	1	4	0.000	0.000	0.009	0.036	No	Bacteria(100); Proteobacteria(100); Alphaproteobacteria(100); Rhodospirillales(100); Acetobacteraceae(100); unclassified(100);
Otu141	0	2	6	0.000	0.000	0.004	0.018	No	Bacteria(100); Proteobacteria(100); Alphaproteobacteria(100); Rhodospirillales(100); unclassified(100); unclassified(100);
Otu147	0	1	4	0.000	0.000	0.001	0.005	No	Bacteria(100); Firmicutes(100); Bacilli(100); Lactobacillales(100); Enterococcaceae(86); Enterococcus(86);

OTU	No. Georgia Ticks Positive	No. Archival Ticks Positive	Total Sequence Abundance	Georgia Mean Relative Sequence Abundance	Archival Mean Relative Sequence Abundance	p	q	Known Tick Associate?	Taxonomy
Otu150	0	1	7	0.000	0.000	0.001	0.005	No	Bacteria(100); Proteobacteria(100); unclassified(72); unclassified(72); unclassified(72); unclassified(72);
Otu151	0	1	3	0.000	0.000	0.001	0.005	No	Bacteria(100); unclassified(100); unclassified(100); unclassified(100); unclassified(100); unclassified(100);
Otu152	0	2	3	0.000	0.000	0.001	0.005	No	Bacteria(100); Proteobacteria(100); Alphaproteobacteria(100); Rhodospirillales(86); Acetobacteraceae(72); unclassified(72);

OTU	No. Georgia Ticks Positive	No. Archival Ticks Positive	Total Sequence Abundance	Georgia Mean Relative Sequence Abundance	Archival Mean Relative Sequence Abundance	p	q	Known Tick Associate?	Taxonomy
Otu156	0	1	3	0.000	0.000	0.001	0.005	Yes	Bacteria(100); Proteobacteria(100); Alphaproteobacteria(100); Rickettsiales(100); Rickettsiaceae(100); unclassified(100);
Otu168	0	2	3	0.000	0.000	0.001	0.005	No	Bacteria(100); Proteobacteria(100); Alphaproteobacteria(100); unclassified(67); unclassified(67); unclassified(67);
Otu169	0	2	4	0.000	0.000	0.001	0.005	No	Bacteria(100); unclassified(100); unclassified(100); unclassified(100); unclassified(100); unclassified(100);

OTU	No. Georgia Ticks Positive	No. Archival Ticks Positive	Total Sequence Abundance	Georgia Mean Relative Sequence Abundance	Archival Mean Relative Sequence Abundance	p	q	Known Tick Associate?	Taxonomy
Otu184	0	2	3	0.000	0.000	0.001	0.005	No	Bacteria(100); Proteobacteria(100); Alphaproteobacteria(100); Sphingomonadales(100); Sphingomonadaceae(100); Sphingomonas(100);
Otu188	0	3	2	0.000	0.000	0.013	0.047	No	Bacteria(100); Proteobacteria(100); Alphaproteobacteria(100); Rhizobiales(100); Methylobacteriaceae(100); unclassified(61);

OTU	No. Georgia Ticks Positive	No. Archival Ticks Positive	Total Sequence Abundance	Georgia Mean Relative Sequence Abundance	Archival Mean Relative Sequence Abundance	*p*	*q*	Known Tick Associate?	Taxonomy
Otu195	0	1	3	0.000	0.000	0.001	0.005	No	Bacteria(100); Armatimonadetes(100); Armatimonadetes_gp5_class_incertae_sedis(100); Armatimonadetes_gp5_order_incetae_sedis(100); Armatimonadetes_gp5_family_incetae_sedis(100); Armatimonadetes_gp5(100);
Otu199	0	1	2	0.000	0.000	0.012	0.045	No	Bacteria(100); Bacteroidetes(100); unclassified(75); unclassified(75); unclassified(75); unclassified(75);

OTU	No. Georgia Ticks Positive	No. Archival Ticks Positive	Total Sequence Abundance	Georgia Mean Relative Sequence Abundance	Archival Mean Relative Sequence Abundance	p	q	Known Tick Associate?	Taxonomy
Otu222	0	2	3	0.000	0.000	0.001	0.005	No	Bacteria(100); unclassified(100); unclassified(100); unclassified(100); unclassified(100); unclassified(100);
Otu296	1	0	2	0.000	0.000	0.001	0.005	No	Bacteria(100); Bacteroidetes(100); Sphingobacteria(100); Sphingobacteriales(100); Sphingobacteriaceae(100); unclassified(100);
Otu328	0	1	2	0.000	0.000	0.014	0.049	No	Bacteria(100); Proteobacteria(100); Betaproteobacteria(100); Burkholderiales(100); Comamonadaceae(100); Schlegelella(100);

OTU	No. Georgia Ticks Positive	No. Archival Ticks Positive	Total Sequence Abundance	Georgia Mean Relative Sequence Abundance	Archival Mean Relative Sequence Abundance	p	q	Known Tick Associate?	Taxonomy
Otu339	1	0	2	0.000	0.000	0.001	0.005	No	Bacteria(100); unclassified(100); unclassified(100); unclassified(100); unclassified(100); unclassified(100);

3.8 References

Acinas SG, Sarma-Rupavtarm R, Klepac-Ceraj V, Polz MF (2005) PCR-induced sequence artifacts and bias: insights from comparison of two 16S rRNA clone libraries constructed from the same sample. *Applied and Environmental Microbiology*, **71**, 8966–8969.

Ahantarig A, Trinachartvanit W, Baimai V, Grubhoffer L (2013) Hard ticks and their bacterial endosymbionts (or would be pathogens). *Folia Microbiologica*, 1–10.

Anderson MJ (2001) A new method for non-parametric multivariate analysis of variance. *Austral Ecology*, **26**, 32–46.

Andreotti R, León AAP de, Dowd SE, Guerrero FD, Bendele KG, Scoles GA (2011) Assessment of bacterial diversity in the cattle tick *Rhipicephalus* (*Boophilus*) *microplus* through tag-encoded pyrosequencing. *BMC Microbiology*, **11**, 6.

Anesti V, McDonald IR, Ramaswamy M, Wade WG, Kelly DP, Wood AP (2005) Isolation and molecular detection of methylotrophic bacteria occurring in the human mouth. *Environmental Microbiology*, **7**, 1227–1238.

Anesti V, Vohra J, Goonetilleka S, McDonald IR, Sträubler B, Stackebrandt E, Kelly DP, Wood AP (2004) Molecular detection and isolation of facultatively methylotrophic bacteria, including *Methylobacterium podarium* sp. nov., from the human foot microflora. *Environmental Microbiology*, **6**, 820–830.

Apperson CS, Engber B, Nicholson WL, Mead DG, Engel J, Yabsley MJ, Dail K, Johnson J, Watson DW (2008) Tick-borne diseases in North Carolina: is "*Rickettsia amblyommii*" a possible cause of rickettsiosis reported as Rocky Mountain spotted fever? *Vector-Borne and Zoonotic Diseases*, **8**, 597–606.

Aylward FO, McDonald BR, Adams SM, Valenzuela A, Schmidt RA, Goodwin LA, Woyke T, Currie CR, Suen G, Poulsen M (2013) Comparison of 26 Sphingomonad genomes reveals diverse environmental adaptations and biodegradative capabilities. *Applied and Environmental Microbiology*, **79**, 3724–3733.

Baldo L, Prendini L, Corthals A, Werren JH (2007) *Wolbachia* are present in southern African scorpions and cluster with supergroup F. *Current Microbiology*, **55**, 367–373.

Barbour AG, Maupin GO, Teltow GJ, Carter CJ, Piesman J (1996) Identification of an uncultivable *Borrelia* species in the hard tick *Amblyomma americanum*: possible agent of a Lyme disease-like illness. *The Journal of Infectious Diseases*, **173**, 403–409.

Bartlett MS (1937) Properties of sufficiency and statistical tests. *Proceedings of the Royal Society of London Series A*, **160**, 268–282.

Bazzocchi C, Mariconti M, Sassera D, Rinaldi L, Martin E, Cringoli G, Urbanelli S, Genchi C, Bandi C, Epis S (2013) Molecular and serological evidence for the circulation of the tick symbiont *Midichloria* (Rickettsiales: Midichloriaceae) in different mammalian species. *Parasites and Vectors*, **6**, 1–7.

Benson MJ, Gawronski JD, Eveleigh DE, Benson DR (2004) Intracellular symbionts and other bacteria associated with deer ticks (*Ixodes scapularis*) from Nantucket and Wellfleet, Cape Cod, Massachusetts. *Applied Environmental Microbiology*, **70**, 616–620.

Bermúdez SE, Eremeeva ME, Karpathy SE, Samudio F, Zambrano ML, Zaldivar Y, Motta JA, Dasch GA (2009) Detection and identification of rickettsial agents in ticks from domestic mammals in eastern Panama. *Journal of Medical Entomology*, **46**, 856–861.

Berry D, Mahfoudh KB, Wagner M, Loy A (2011) Barcoded primers used in multiplex amplicon pyrosequencing bias amplification. *Applied and Environmental Microbiology*, **77**, 7846–7849.

Billeter SA, Blanton HL, Little SE, Levy MG, Breitschwerdt EB (2007) Detection of "*Rickettsia amblyommii*" in association with a tick bite rash. *Vector-Borne and Zoonotic Diseases*, **7**, 607–610.

Van Borm S, Buschinger A, Boomsma JJ, Billen J (2002) *Tetraponera* ants have gut symbionts related to nitrogen-fixing root-nodule bacteria. *Proceedings of the Royal Society B: Biological Sciences*, **269**, 2023–2027.

Bray JR, Curtis JT (1957) An ordination of the upland forest communities of southern Wisconsin. *Ecological Monographs*, **27**, 326–349.

Brown SP, Jumpponen A (2013) Contrasting primary successional trajectories of fungi and bacteria in retreating glacier soils. *Molecular Ecology*, **epub ahead of print**, 1–17.

Buller RS, Arens M, Hmiel SP, Paddock CD, Sumner JW, Rikihisa Y, Unver A, Gaudreault-Keener M, Manian FA, Liddell AM, Schmulewitz N, Storch GA (1999) *Ehrlichia ewingii*, a newly recognized agent of human ehrlichiosis. *New England Journal of Medicine*, **341**, 148–155.

Burgdorfer W, Hayes SF, Thomas L, Lancaster J (1981) A new spotted fever group rickettsia from the lone star tick, *Amblyomma americanum*. In: *Rickettsiae and Rickettsial Diseases* (eds Burgdorfer W, Anacker W), pp. 595–602. Academic Press, New York.

Carpi G, Cagnacci F, Wittekindt NE, Zhao F, Qi J, Tomsho LP, Drautz DI, Rizzoli A, Schuster SC (2011) Metagenomic profile of the bacterial communities associated with *Ixodes ricinus* ticks. *PLoS ONE*, **6**, e25604.

Casiraghi M, Favia G, Cancrini G, Bartoloni A, Bandi C (2001) Molecular identification of *Wolbachia* from the filarial nematode *Mansonella ozzardi*. *Parasitology Research*, **87**, 417–420.

Childs JE, Paddock CD (2003) The ascendancy of *Amblyomma americanum* as a vector of pathogens affecting humans in the United States. *Annual Review of Entomology*, **48**, 307–337.

Clay K, Fuqua C (2010) *The tick microbiome: diversity, distribution and influence of the internal microbial community for a blood-feeding disease vector*. Institute of Medicine, Washington, D. C.

Clay K, Klyachko O, Grindle N, Civitello D, Oleske D, Fuqua C (2008) Microbial communities and interactions in the lone star tick, *Amblyomma americanum*. *Molecular Ecology*, **17**, 4371–4381.

Compant S, Nowak J, Coenye T, Clément C, Ait Barka E (2008) Diversity and occurrence of *Burkholderia* spp. in the natural environment. *FEMS Microbiology Reviews*, **32**, 607–626.

Corrigan J (2012) Investigation of spotted fever group *Rickettsia* in dogs and ticks in northern California. Thesis (MS) Thesis. Humboldt State University.

Covacin C, Barker SC (2007) Supergroup F *Wolbachia* bacteria parasitise lice (Insecta: Phthiraptera). *Parasitology Research*, **100**, 479–485.

Dale C, Maudlin I (1999) *Sodalis* gen. nov. and *Sodalis glossinidius* sp. nov., a microaerophilic secondary endosymbiont of the tsetse fly *Glossina morsitans morsitans*. *International Journal of Systematic Bacteriology*, **49**, 267–275.

Dasch G, Kelly D, Richards A, Sanchez J, Rives C (1993) Western blotting analysis of sera from military personnel exhibiting serological reactivity to spotted fever group rickettsiae. *American Journal of Tropical Medicine and Hygiene*, **49**, 220.

Degelmann DM, Kolb S, Dumont M, Murrell JC, Drake HL (2009) Enterobacteriaceae facilitate the anaerobic degradation of glucose by a forest soil. *FEMS Microbiology Ecology*, **68**, 312–319.

Dergousoff SJ, Chilton NB (2011) Novel genotypes of *Anaplasma bovis*, "*Candidatus* Midichloria" sp. and *Ignatzschineria* sp. in the Rocky Mountain wood tick, *Dermacentor andersoni*. *Veterinary Microbiology*, **150**, 100–106.

Douglas AE (1998) Nutritional interactions in insect-microbial symbioses: aphids and their symbiotic bacteria *Buchnera*. *Annual Review of Entomology*, **43**, 17–37.

Eller G, Frenzel P (2001) Changes in activity and community structure of methane-oxidizing bacteria over the growth period of rice. *Applied and Environmental Microbiology*, **67**, 2395–2403.

Epis S, Sassera D, Beninati T, Lo N, Beati L, Piesman J, Rinaldi L, McCoy KD, Torina A, Sacchi L, Clementi E, Genchi M, Magnino S, Bandi C (2008) *Midichloria*

mitochondrii is widespread in hard ticks (Ixodidae) and resides in the mitochondria of phylogenetically diverse species. *Parasitology*, **135**, 485–494.

Esty WW (1986) The efficiency of Good's nonparametric coverage estimator. *Annals of Statistics*, **14**, 1257–1260.

Faith DP, Minchin PR, Belbin L (1987) Compositional dissimilarity as a robust measure of ecological distance. *Vegetatio*, **69**, 57–68.

Fritzen CM, Huang J, Westby K, Freye JD, Dunlap B, Yabsley MJ, Schardein M, Dunn JR, Jones TF, Moncayo AC (2011) Infection prevalences of common tick-borne pathogens in adult lone star ticks (*Amblyomma americanum*) and American dog ticks (*Dermacentor variabilis*) in Kentucky. *The American Journal of Tropical Medicine and Hygiene*, **85**, 718–723.

Ginsberg HS, Ewing CP, O'Connell AF, Bosler EM, Daley JG, Sayre MW (1991) Increased population densities of *Amblyomma americanum* (Acari: Ixodidae) on Long Island, New York. *The Journal of Parasitology*, **77**, 493–495.

Gladney WJ, Drummond RO (1970) Mating behavior and reproduction of the lone star tick, *Amblyomma americanum*. *Annals of the Entomological Society of America*, **63**, 1036–1039.

Gronesova P, Ficova M, Mizakova A, Kabat P, Trnka A, Betakova T (2008) Prevalence of avian influenza viruses, *Borrelia garinii*, *Mycobacterium avium*, and *Mycobacterium avium* subsp. *paratuberculosis* in waterfowl and terrestrial birds in Slovakia, 2006. *Avian Pathology*, **37**, 537–543.

Groves MG, Dennis GL, Amyx HL, Huxsoll DL (1975) Transmission of *Ehrlichia canis* to dogs by ticks (*Rhipicephalus sanguineus*). *American Journal of Veterinary Research*, **36**, 937–940.

Halos L, Bord S, Cotté V, Gasqui P, Abrial D, Barnouin J, Boulouis H-J, Vayssier-Taussat M, Vourc'h G (2010) Ecological factors characterizing the prevalence of bacterial tick-borne pathogens in *Ixodes ricinus* ticks in pastures and woodlands. *Applied and Environmental Microbiology*, **76**, 4413–4420.

Hartmann C, Lennartz K, Ibrahim H, Coz A, Kasper Y, Lenz C, Mathur D, Polidori M (2012) Application Note: Stable 16-year Storage of DNA Purified with the QIAamp® DNA Blood Mini Kit.

Hawlena H, Rynkiewicz E, Toh E, Alfred A, Durden LA, Hastriter MW, Nelson DE, Rong R, Munro D, Dong Q, Fuqua C, Clay K (2013) The arthropod, but not the vertebrate host or its environment, dictates bacterial community composition of fleas and ticks. *The ISME Journal*, **7**, 221–223.

Heise SR, Elshahed MS, Little SE (2010) Bacterial diversity in *Amblyomma americanum* (Acari: Ixodidae) with a focus on members of the genus *Rickettsia*. *Journal of Medical Entomology*, **47**, 258–268.

Hill TCJ, Walsh KA, Harris JA, Moffett BF (2003) Using ecological diversity measures with bacterial communities. *FEMS Microbiology Ecology*, **43**, 1–11.

[HMP] Jumpstart Consortium Human Microbiome Project Data Generation Working Group (2010) 16S 454 Sequencing Protocol HMP Consortium (version 4.2.2). Available: www.hmpdacc.org/doc/16S_Sequencing_SOP_4.2.2.pdf. Accessed 6 April 2014.

Holm S (1979) A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, **6**, 65–70.

Hong S, Bunge J, Leslin C, Jeon S, Epstein SS (2009) Polymerase chain reaction primers miss half of rRNA microbial diversity. *The ISME Journal*, **3**, 1365–1373.

Inokuma H, Raoult D, Brouqui P (2000) Detection of *Ehrlichia platys* DNA in Brown Dog Ticks (*Rhipicephalus sanguineus*) in Okinawa Island, Japan. *Journal of Clinical Microbiology*, **38**, 4219–4221.

Jaccard P (1908) Nouvelles recherches sur la distribution florale. *Société Vaudoise des Sciences Naturelles*, **44**, 223–270.

James AM, Liveris D, Wormser GP, Schwartz I, Montecalvo MA, Johnson BJB (2001) *Borrelia lonestari* infection after a bite by an *Amblyomma americanum* tick. *The Journal of Infectious Diseases*, **183**, 1810–1814.

Jasinskas A, Zhong J, Barbour AG (2007) Highly prevalent *Coxiella* sp. bacterium in the tick vector *Amblyomma americanum*. *Applied and Environmental Microbiology*, **73**, 334–336.

Jiang J, Yarina T, Miller MK, Stromdahl EY, Richards AL (2010) Molecular detection of *Rickettsia amblyommii* in *Amblyomma americanum* parasitizing humans. *Vector-Borne and Zoonotic Diseases*, **10**, 329–340.

Johnson EM, Ewing SA, Barker RW, Fox JC, Crow DW, Kocan KM (1998) Experimental transmission of *Ehrlichia canis* (Rickettsiales: Ehrlichieae) by *Dermacentor variabilis* (Acari: Ixodidae). *Veterinary Parasitology*, **74**, 277–288.

Jongejan F, Uilenberg G (2004) The global importance of ticks. *Parasitology*, **129**, S3–S14.

Kagemann J, Clay K (2013) Effects of infection by *Arsenophonus* and *Rickettsia* bacteria on the locomotive ability of the ticks *Amblyomma americanum*, *Dermacentor variabilis*, and *Ixodes scapularis*. *Journal of Medical Entomology*, **50**, 155–162.

Kaltenpoth M, Engl T (2013) Defensive microbial symbionts in Hymenoptera. *Functional Ecology*, **epub online ahead of print**, 1–13.

Kauserud H, Kumar S, Brysting AK, Nordén J, Carlsen T (2012) High consistency between replicate 454 pyrosequencing analyses of ectomycorrhizal plant root samples. *Mycorrhiza*, **22**, 309–315.

Kautz S, Rubin BER, Russell JA, Moreau CS (2013) Surveying the microbiome of ants: comparing 454 pyrosequencing with traditional methods to uncover bacterial diversity. *Applied and Environmental Microbiology*, **79**, 525–534.

Keirans JE, Lacombe EH (1998) First records of *Amblyomma americanum*, *Ixodes* (*Ixodes*) *dentatus*, and *Ixodes* (*Ceratixodes*) *uriae* (Acari: Ixodidae) from Maine. *The Journal of Parasitology*, **84**, 629–631.

Kikuchi Y, Hayatsu M, Hosokawa T, Nagayama A, Tago K, Fukatsu T (2012) Symbiont-mediated insecticide resistance. *Proceedings of the National Academy of Sciences*, **epub online ahead of print**, 1–5.

Kindt R, Coe R (2005) Chapter 10: analysis of ecological distance by ordination. In: *Tree diversity analysis: a manual and software for common statistical methods for ecological and biodiversity studies*, pp. 153–196. World Agroforestry Centre, Nairobi, Kenya.

Klyachko O, Stein BD, Grindle N, Clay K, Fuqua C (2007) Localization and visualization of a *Coxiella*-type symbiont within the lone star tick, *Amblyomma americanum*. *Applied and Environmental Microbiology*, **73**, 6584–6594.

Kneip C, Lockhart P, Voss C, Maier U-G (2007) Nitrogen fixation in eukaryotes - new models for symbiosis. *BMC Evolutionary Biology*, **7**, 55.

Kokernot RH, Calisher CH, Stannard LJ, Hayes J (1969) Arbovirus studies in the Ohio-Mississippi basin, 1964–1967 VII. Lone Star virus, a hitherto unknown agent isolated from the tick *Amblyomma americanum* (Linn.). *The American Journal of Tropical Medicine and Hygiene*, **18**, 789–795.

Labruna MB, Whitworth T, Bouyer DH, McBride J, Camargo LMA, Camargo EP, Popov V, Walker DH (2004) *Rickettsia bellii* and *Rickettsia amblyommii* in *Amblyomma* ticks from the State of Rondônia, Western Amazon, Brazil. *Journal of Medical Entomology*, **41**, 1073–1081.

Lai C-C, Cheng A, Liu W-L, Tan C-K, Huang Y-T, Chung K-P, Lee M-R, Hsueh P-R (2011) Infections caused by unusual *Methylobacterium* species▽. *Journal of Clinical Microbiology*, **49**, 3329–3331.

Lalzar I, Harrus S, Mumcuoglu KY, Gottlieb Y (2012) Composition and seasonal variation of *Rhipicephalus turanicus* and *Rhipicephalus sanguineus* bacterial communities. *Applied and Environmental Microbiology*, **78**, 4110–6.

Lefoulon E, Gavotte L, Junker K, Barbuto M, Uni S, Landmann F, Laaksonen S, Saari S, Nikander S, de Souza Lima S, Casiraghi M, Bain O, Martin C (2012) A new type F *Wolbachia* from Splendidofilariinae (Onchocercidae) supports the recent

emergence of this supergroup. *International Journal for Parasitology*, **42**, 1025–1036.

Lo N, Beninati T, Sassera D, Bouman EAP, Santagati S, Gern L, Sambri V, Masuzawa T, Gray JS, Jaenson TGT, Bouattour A, Kenny MJ, Guner ES, Kharitonenkov IG, Bitam I, Bandi C (2006) Widespread distribution and high prevalence of an alpha-proteobacterial symbiont in the tick *Ixodes ricinus*. *Environmental Microbiology*, **8**, 1280–1287.

Means RG, White DJ (1997) New distribution records of *Amblyomma americanum* (L.) (Acari : Ixodidae) in New York State. *Journal of Vector Ecology*, **22**, 133–145.

Menchaca AC, Visi DK, Strey OF, Teel PD, Kalinowski K, Allen MS, Williamson PC (2013) Preliminary assessment of microbiome changes following blood-feeding and survivorship in the *Amblyomma americanum* nymph-to-adult transition using semiconductor sequencing. *PLoS ONE*, **8**, e67129.

Mercado TI, Colon-Whitt A (1982) Lysis of *Trypanosoma cruzi* by *Pseudomonas fluorescens*. *Antimicrobial Agents and Chemotherapy*, **22**, 1051–1057.

Merten HA, Durden LA (2000) A state-by-state survey of ticks recorded from humans in the United States. *Journal of Vector Ecology*, **25**, 102–113.

Milford AD, Achenbach LA, Jung DO, Madigan MT (2000) *Rhodobaca bogoriensis* gen. nov. and sp. nov., an alkaliphilic purple nonsulfur bacterium from African Rift Valley soda lakes. *Archives of Microbiology*, **174**, 18–27.

Minchin PR (1987) An evaluation of the relative robustness of techniques for ecological ordination. *Vegetatio*, **69**, 89–107.

Mixson TR, Campbell SR, Gill JS, Ginsberg HS, Reichard MV, Schulze TL, Dasch GA (2006) Prevalence of *Ehrlichia*, *Borrelia*, and rickettsial agents in *Amblyomma americanum* (Acari: Ixodidae) collected from nine states. *Journal of Medical Entomology*, **43**, 1261–1268.

Moncayo AC, Cohen SB, Fritzen CM, Huang E, Yabsley MJ, Freye JD, Dunlap BG, Huang J, Mead DG, Jones TF, Dunn JR (2010) Absence of *Rickettsia rickettsii* and occurrence of other spotted fever group rickettsiae in ticks from Tennessee. *The American Journal of Tropical Medicine and Hygiene*, **83**, 653–657.

Montagna M, Chouaia B, Pella F, Mariconti M, Pistone D, Fasola M, Epis S (2012) Screening for bacterial DNA in the hard tick *Hyalomma marginatum* (Ixodidae) from Socotra Island (Yemen): detection of *Francisella*-like endosymbiont. *Journal of Entomological and Acarological Research*, **44**.

Moreno CX, Moy F, Daniels TJ, Godfrey HP, Cabello FC (2006) Molecular analysis of microbial communities identified in different developmental stages of *Ixodes scapularis* ticks from Westchester and Dutchess Counties, New York. *Environmental Microbiology*, **8**, 761–772.

Nakao R, Abe T, Nijhof AM, Yamamoto S, Jongejan F, Ikemura T, Sugimoto C (2013) A novel approach, based on BLSOMs (Batch Learning Self-Organizing Maps), to the microbiome analysis of ticks. *The ISME Journal*.

Navarro-Noya YE, Suárez-Arriaga MC, Rojas-Valdes A, Montoya-Ciriaco NM, Gómez-Acata S, Fernández-Luqueño F, Dendooven L (2013) Pyrosequencing analysis of the bacterial community in drinking water wells. *Microbial Ecology*, **66**, 19–29.

Neilson JW, Quade J, Ortiz M, Nelson WM, Legatzki A, Tian F, LaComb M, Betancourt JL, Wing RA, Soderlund CA, Maier RM (2012) Life at the hyperarid margin: novel bacterial diversity in arid soils of the Atacama Desert, Chile. *Extremophiles*, **16**, 553–566.

Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens MHH, Wagner H (2013) *vegan: Community Ecology Package*.

Pais R, Lohs C, Wu Y, Wang J, Aksoy S (2008) The obligate mutualist *Wigglesworthia glossinidia* influences reproduction, digestion, and immunity processes of its host, the tsetse fly. *Applied and Environmental Microbiology*, **74**, 5965–5974.

Pawitan Y, Michiels S, Koscielny S, Gusnanto A, Ploner A (2005) False discovery rate, sensitivity and sample size for microarray studies. *Bioinformatics*, **21**, 3017–3024.

Perry RD, Fetherston JD (1997) *Yersinia pestis*--etiologic agent of plague. *Clinical Microbiology Reviews*, **10**, 35–66.

Pilloni G, Granitsiotis MS, Engel M, Lueders T (2012) Testing the limits of 454 pyrotag sequencing: reproducibility, quantitative assessment and comparison to T-RFLP fingerprinting of aquifer microbes. *PLoS ONE*, **7**, e40467.

Pinto AJ, Raskin L (2012) PCR biases distort bacterial and archaeal community structure in pyrosequencing datasets. *PLoS ONE*, **7**, e43093.

Pistone D, Sacchi L, Lo N, Epis S, Pajoro M, Favia G, Mandrioli M, Bandi C, Sassera D (2012) *Candidatus* Midichloria mitochondrii: symbiont or parasite of tick mitochondrii? In: *Manipulative Tenants: Bacteria Associated With Arthropods*

Frontiers in Microbiology. (eds Zchori-Fein E, Bourtzis K), pp. 113–126. CRC Press, Boca Raton.

Plantard O, Bouju-Albert A, Malard M-A, Hermouet A, Capron G, Verheyden H (2012) Detection of *Wolbachia* in the tick *Ixodes ricinus* is due to the presence of the hymenoptera endoparasitoid *Ixodiphagus hookeri*. *PLoS ONE*, **7**, e30692.

Ponnusamy L, Gonzalez A, Treuren WV, Weiss S, Parobek CM, Juliano JJ, Knight R, Roe RM, Apperson CS, Meshnick SR (2014) Diversity of Rickettsiales in the microbiome of the lone star tick, *Amblyomma americanum*. *Applied and Environmental Microbiology*, **80**, 354–359.

R Core Team (2013) *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

Reeves W, Loftis A, Sanders F, Spinks M, Wills W, Denison A, Dasch G (2006) *Borrelia*, *Coxiella*, and *Rickettsia* in *Carios capensis* (Acari: Argasidae) from a brown pelican (*Pelecanus occidentalis*) rookery in South Carolina, USA. *Experimental and Applied Acarology*, **39**, 321–329.

Salles JF, De Souza FA, van Elsas JD (2002) Molecular method to assess the diversity of *Burkholderia* species in environmental samples. *Applied and environmental microbiology*, **68**, 1595–1603.

Savage HM, Godsey MS, Lambert A, Panella NA, Burkhalter KL, Harmon JR, Lash RR, Ashley DC, Nicholson WL (2013) First detection of Heartland virus (Bunyaviridae: Phlebovirus) from field collected arthropods. *The American Journal of Tropical Medicine and Hygiene*, **89**, 445–452.

Schloss PD (2009) A high-throughput DNA sequence aligner for microbial ecology studies. *PLoS ONE*, **4**, e8230.

Schloss PD (2013) Schloss SOP. *Mothur Wiki*. Available: /www.mothur.org/wiki/Schloss_SOP. Accessed 30 September 2013.

Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ, Weber CF (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology*, **75**, 7537–7541.

Scoles GA (2004) Phylogenetic analysis of the *Francisella*-like endosymbionts of *Dermacentor* ticks. *Journal of Medical Entomology*, **41**, 277–286.

Smith MP, Ponnusamy L, Jiang J, Ayyash LA, Richards AL, Apperson CS (2010) Bacterial pathogens in ixodid ticks from a Piedmont county in North Carolina: prevalence of rickettsial organisms. *Vector-Borne and Zoonotic Diseases*, **10**, 939–952.

Stromdahl EY, Vince MA, Billingsley PM, Dobbs NA, Williamson PC (2008) *Rickettsia amblyommii* infecting *Amblyomma americanum* larvae. *Vector-Borne and Zoonotic Diseases*, **8**, 15–24.

Terenius O, Lindh JM, Eriksson-Gonzales K, Bussière L, Laugen AT, Bergquist H, Titanji K, Faye I (2012) Midgut bacterial dynamics in *Aedes aegypti*. *FEMS Microbiology Ecology*, **80**, 556–565.

Valvano MA, Keith KE, Cardona ST (2005) Survival and persistence of opportunistic *Burkholderia* species in host cells. *Current Opinion in Microbiology*, **8**, 99–105.

Veech JA (2013) A probabilistic model for analysing species co-occurrence. *Global Ecology and Biogeography*, **22**, 252–260.

Ventura M, Canchaya C, Tauch A, Chandra G, Fitzgerald GF, Chater KF, Sinderen D van (2007) Genomics of Actinobacteria: tracing the evolutionary history of an ancient phylum. *Microbiology and Molecular Biology Reviews*, **71**, 495–548.

Venzal J, Estrada-Peña A, Portillo A, Mangold A, Castro O, de Souza C, Félix M, Pérez-Martínez L, Santibánez S, Oteo J (2008) Detection of alpha and gamma-proteobacteria in *Amblyomma triste* (Acari: Ixodidae) from Uruguay. *Experimental and Applied Acarology*, **44**, 49–56.

Werren JH, Baldo L, Clark ME (2008) *Wolbachia*: master manipulators of invertebrate biology. *Nature Reviews Microbiology*, **6**, 741–751.

White JR, Nagarajan N, Pop M (2009) Statistical methods for detecting differentially abundant features in clinical metagenomic samples. *PLoS Comput Biol*, **5**, e1000352.

Williams-Newkirk AJ, Rowe LA, Mixson-Hayden TR, Dasch GA (2012) Presence, genetic variability, and potential significance of "*Candidatus* Midichloria mitochondrii" in the lone star tick *Amblyomma americanum*. *Experimental and Applied Acarology*, **58**, 291–300.

Yuan D (2010) A metagenomic study of the tick midgut. Thesis (MS) Thesis. The University of Texas, Houston, Texas.

Zanetti AS, Pornwiroon W, Kearney MT, Macaluso KR (2008) Characterization of rickettsial infection in *Amblyomma americanum* (Acari: Ixodidae) by quantitative

real-time polymerase chain reaction. *Journal of Medical Entomology*, **45**, 267–275.

Zhang X, Norris DE, Rasgon JL (2011) Distribution and molecular characterization of *Wolbachia* endosymbionts and filarial nematodes in Maryland populations of the lone star tick (*Amblyomma americanum*). *FEMS Microbiology Ecology*, **77**, 50–56.

Zhang X, Ren X, Norris DE, Rasgon JL (2012) Distribution and infection frequency of "*Candidatus* Rickettsia amblyommii" in Maryland populations of the lone star tick (*Amblyomma americanum*) and culture in an *Anopheles gambiae* mosquito cell line. *Ticks and Tick-borne Diseases*, **3**, 38–42.

Zhong JM, Jasinskas A, Barbour AG (2007) Antibiotic treatment of the tick vector *Amblyomma americanum* reduced reproductive fitness. *PLoS One*, **2**, e405.

Zhou J, Wu L, Deng Y, Zhi X, Jiang Y-H, Tu Q, Xie J, Van Nostrand JD, He Z, Yang Y (2011) Reproducibility and quantitation of amplicon sequencing-based detection. *The ISME Journal*, **5**, 1303–1313.

CHAPTER 4

Sex and species phylogenies influence the bacterial community compositions of four sympatric North American tick species

4.1 Summary

Among ixodid ticks, important vectors of many human, livestock, and wildlife diseases worldwide, the bacterial communities of many species remain minimally described and the degree of interspecific exchange of non-vertebrate pathogens is largely unexplored. Variation in arthropod-associated bacterial communities may arise due to factors both intrinsic and extrinsic to the host, *e.g.* sex, species, diet, and environmental exposure. Additionally, bacterial community relationships may reflect the phylogeny of their hosts as a result of phylosymbiotic processes. I pyrosequenced 16S rRNA gene amplicons from 130 adult ticks from four sympatric species native to the eastern United States (*Amblyomma americanum*, *A. maculatum*, *Dermacentor variabilis*, and *Ixodes scapularis*) to determine the relatedness of their bacterial communities and to identify sex and species-specific assemblages. While bacterial alpha-diversity was similar across tick species, the patterns of bacterial operational taxonomic unit (OTU) occurrence and abundance were generally tick species specific. No bacterial OTU was common to all four tick species. Sex-specific communities were observed only within *A. americanum* and *I. scapularis*. Bacterial community relatedness quantified by weighted UniFrac was consistent with the host tick phylogeny generated from mitochondrial genes (Mantel test, r=0.67, p<0.05). However, the relationships derived from alternative metrics highlight

the possibility that other mechanisms, such as niche adaptation to specific hosts or environmental conditions, may also influence the membership and structure of tick bacterial communities. I also report the first observations of two genotypes of the potential endosymbiont, *Candidatus* 'Midichloria mitochondrii', in *A. maculatum*.

4.2 Introduction

Animals are dependent on their microbial communities for optimal health (McFall-Ngai *et al.* 2013). The beneficial effects of mutualistic bacteria on host survival, development, and fecundity are particularly well described (*e.g.* Montgomery & McFall-Ngai 1994; Oliver *et al.* 2003; Pais *et al.* 2008). Given their important functions, it is unsurprising that these bacterial communities are not a random subset of the available environmental bacteria (Duguma *et al.* 2013), but rather have a composition and structure that is specific to the host species (*e.g.* Jones *et al.* 2013). Describing the roles of and functions provided by microbes to their host and the methods by which the host acquires and maintains a compatible community is an active area of research.

Some of the broad mechanisms of microbial community assembly are understood, but the relative contributions of different processes and the ways in which each are effected within specific systems remain largely unknown. Diet is an important influence as specialized diets are associated with convergent gut bacterial communities in phylogenetically distant hosts (*e.g.* Colman *et al.* 2012). Sex and life stage have been correlated with intraspecific community variation (*e.g.* Brucker & Bordenstein 2012; Williams-Newkirk *et al.* 2014). Many of the most critical bacteria are vertically

transmitted by hosts and may be considered a portion of the heritable genetic variation passed from mother to offspring (Zilber-Rosenberg & Rosenberg 2008). Others are acquired horizontally, and it is the combined influence of the existing microbiota and the host's immune system that determines whether the invader colonizes or is cleared from the community and whether or not a dysbiotic phenotype emerges from the disturbance (Vavre & Kremer 2014). The heritability of the immune system and some of their microbiota can lead to closely related individuals having more similar microbial communities and ultimately may result in community relatedness reflecting the host's phylogeny, a pattern called phylosymbiosis (Brucker & Bordenstein 2013). Patterns consistent with phylosymbiosis have been observed across many different host lineages (reviewed in Vavre & Kremer 2014), but to my knowledge no study has examined ixodid ticks in this context.

Ixodid (hard) ticks feed exclusively on vertebrate blood during all post-egg live stages, resulting in a simplified bacterial community that is characteristic of many diet-restricted arthropods. Each tick species surveyed by Next Generation sequencing (NGS) to date has a bacterial community predominated by a few taxa (Yuan 2010; Andreotti *et al.* 2011; Carpi *et al.* 2011; Lalzar *et al.* 2012; Hawlena *et al.* 2013; Nakao *et al.* 2013; Menchaca *et al.* 2013; Zhang *et al.* 2014; Budachetri *et al.* 2014; Williams-Newkirk *et al.* 2014; Qiu *et al.* 2014). At least one of these members appears ubiquitous and is inferred (based on other diet-restricted systems, *e.g.* Douglas 1998; Sassera *et al.* 2013) to be a vertically transmitted obligate endosymbiont that supplies deficient nutrients. Only one published study has attempted to experimentally confirm an inferred obligate symbiotic relationship in ixodid ticks (Zhong *et al.* 2007). Exclusive hematophagy also greatly limits the

opportunities for acquisition of bacteria from the environment as ticks obtain water from the blood meal and water vapor absorption (McMullen *et al.* 1976) and feed in a manner that likely limits the introduction of environmental bacteria into the blood (Richter *et al.* 2013). Therefore the majority of horizontally acquired bacteria are expected to originate in the blood meal. These bacteria may originate from infected host blood or tissues, or the blood meal of one tick may be contaminated by the infectious salivary secretions of a second tick feeding in close physical proximity on the same host (*i.e.* co-feeding transmission) (Randolph *et al.* 1996).

Three ixodid tick species account for most of the human tick bites in the eastern United States: *Amblyomma americanum*, *Dermacentor variabilis*, and *Ixodes scapularis*. *Amblyomma maculatum* is most closely related to *A. americanum* and also bites humans opportunistically (Estrada-Peña & Jongejan 1999). All four species undergo three post-egg life stages and must find a vertebrate host and feed once during each life stage (Hooker *et al.* 1912). Two or more species occur sympatrically in most of the eastern US, and all four are common throughout much of the Southeast (Figure 4.S1). Birds and mammals are common hosts of these ticks, and they can be found co-feeding, but differences in host and habitat preferences and seasonal abundances exist (Guglielmone *et al.* 2014).

All four of these tick species transmit at least one confirmed human pathogen and are the target of extensive microbiological research (Estrada-Peña & Jongejan 1999). Both *D. variabilis* and *A. maculatum* are associated with an abundant *Francisella* endosymbiont (Scoles 2004), while *A. americanum* and *I. scapularis* are associated with a similarly copious *Coxiella* (Jasinskas *et al.* 2007) and *Rickettsia* (Noda *et al.* 1997) endosymbiont,

respectively. These symbiont lineages are discordant with the phylogeny of their tick hosts and show no evidence of phylosymbiosis, but it may be that the composition or structure of the complete bacterial communities may display different relationships. If so, a phylosymbiotic signature may indicate an important and heretofore unexplored functional role of these communities for these tick species. Additionally, my previous work identified sex-specific bacterial communities in *A. americanum* (Williams-Newkirk *et al.* 2014), a pattern which may or may not be common across the Ixodidae. To explore these questions, I pyrosequenced *rrs* gene amplicons from the bacterial communities of approximately 30 questing adult ticks from each species and examined their structure.

4.3 Materials and methods

Tick collection and DNA extraction

Adult ticks were collected by running a 1 m^2 flannel cloth over vegetation. Ticks were collected from five sites in four states (Table 4.1, Figure 4.S1); permission for collections was obtained from property managers prior to sampling. Ticks were identified morphologically using standard keys (Keirans & Litwak 1989). *Amblyomma americanum*, *D. variabilis*, and *I. scapularis* (Table 4.1) were placed live into 70% ethanol and stored at 4°C until this study. For DNA extraction, the ticks were sequentially washed in 10% sodium hypochlorite for 10 minutes and 70% ethanol for five minutes, followed by three sterile water rinses. DNA was extracted from individual ticks as described in Bermúdez *et al.* (2009) using the Promega (Madison, WI) Wizard SV 96 Genomic DNA Purification System. All *A. maculatum* were washed while live

sequentially in 2% Micro-Chem Plus (National Chemical Laboratories of PA, Inc., Philadelphia, PA), 10.5% sodium hypochlorite, 3% hydrogen peroxide, and sterile water and then longitudinally bisected with a sterile scalpel. DNA was immediately extracted from one half of each *A. maculatum* collected in 2005 (Table 4.1) using the QIAamp DNA Mini-Kit (Qiagen, Valencia, CA) as previously described (Nava *et al.* 2008). DNAs were stored at -80C until this study. One half of each surface disinfected *A. maculatum* collected in 2007 (Table 4.1) was placed in 0.5 mL of Eagle's minimum essential medium with Earle's salts (MEM) and held at -80C until this study. For DNA extraction, 1 mL of PBS was added to each tick half in MEM. The samples were gently mixed, centrifuged, and the supernatant discarded. DNA was extracted from the remaining pellet using the QIAamp DNA Mini-Kit.

454 library preparation and sequencing

Barcoded eubacterial *rrs* primers 357F and 926R targeting variable regions 5-3 were used to produce amplicons approximately 643 bp long from each tick DNA. Primer sequences, barcode sequences, and PCR conditions using Platinum PCR SuperMix High Fidelity (Invitrogen, Carlsbad, CA) are provided in Williams-Newkirk *et al.* (2014). Amplicons for sequencing plate A (Table 4.S1) were quantitated, purified, and pooled as previously described (Williams-Newkirk *et al.* 2014). Amplicons for sequencing plate B (Table 4.S1) were purified and normalized by applying 15 µL of each reaction to a SequalPrep Normalization Plate (Invitrogen) and following the manufacturer's overnight incubation protocol. Fifteen microliters of each normalized product were pooled, and the pooled products were concentrated to 12 µL using a Genomic DNA Clean & Concentrator

column (Zymo Research, Irvine, CA). Pools were sequenced on a Roche 454 GS-FLX sequencer as previously described (Williams-Newkirk *et al.* 2014).

Bioinformatic analysis

The quality control bioinformatic pipeline was modeled on the Schloss SOP pipeline (Schloss 2013) and executed in version 1.33.3 of the software mothur (Schloss *et al.* 2009); details are given in Williams-Newkirk *et al.* (2014). Trimmed, high-quality sequences were clustered into operational taxonomic units (OTUs) at 97% identity and classified as previously described (Williams-Newkirk *et al.* 2014). Singleton OTUs were removed from further analysis (Zhou *et al.* 2011). The most copious OTUs (>400 sequences) occurring in more than one tick species were analyzed using the supervised computational method Oligotyping (Eren *et al.* 2013) to identify subpopulations defined by SNP patterns and relative abundance. Table 4.S2 lists the versions of the Oligotyping software dependencies used in this analysis. OTUs were split if the analysis supported multiple genotypes (Table 4.S3). This combined approach allowed the more nuanced analysis of predominant community members without inflating the number of rare OTUs. Smaller OTUs were not analyzed by Oligotyping because OTUs containing 100-400 sequences did not produced well-defined (*i.e.* convergent) subtypes. Bacterial communities were rarefied to 1000 sequences per sample as appropriate before statistical analysis.

Bacterial diversity analyses

Good's community coverage estimate (Esty 1986), rarefaction curves, and the inverse Simpson alpha diversity index (1/D) (Williams 1964) were calculated in mothur using a

subsample size of 1000 sequences and 1000 iterations. Analysis of variance (ANOVA) and Tukey's honestly significant difference (HSD) post hoc tests were performed on the R statistical platform (version 3.1.1) (R Core Team 2013).

Bacterial community analyses

We computed the relatedness of rarefied tick bacterial communities using four metrics that addressed either community composition (presence/absence of OTUs) or community structure (OTU abundance) and treated OTUs as either independent characters or phylogenetically related (Table 4.2). All dissimilarity matrices were computed in mothur using a subsample size of 1000 sequences and the mean of 1000 iterations. The Jaccard (Jaccard 1908) and Bray-Curtis (Bray & Curtis 1957) matrices were calculated using an OTU table that included the subgroups identified by Oligotyping. The weighted (Lozupone *et al.* 2007) and unweighted (Lozupone & Knight 2005) UniFrac metrics were calculated using a relaxed neighbor-joining tree produced in the program Clearcut (Evans *et al.* 2006). The tree was based on an uncorrected pairwise distance matrix containing all unique sequences generated in mothur.

Sources of variation in the dissimilarity matrices were partitioned using a non-parametric permutational multivariate analysis of variance (PERMANOVA) as implemented in the R package "vegan" (version 2.0-10) (Oksanen *et al.* 2013) using the function adonis (permutions =1000). PERMANOVA assumes homogeneity of group variance and should be interpreted cautiously if this assumption is not met (Anderson 2001). I tested the compliance of each dissimilarity matrix with the vegan function betadisper (permutions =1000).

The results of the PERMANOVA were confirmed and interpreted using non-metric multi-dimensional scaling (NMDS) (Faith *et al.* 1987; Minchin 1987) ordination of each dissimilarity matrix. NMDS maximizes the rank-order correlation between the two-dimensional ordination and the original multidimensional dissimilarity matrix; individual tick bacterial communities that are closer together in ordination space are more similar than those that are more distant. The NMDS analysis was performed in vegan using the function metaMDS on the mothur-generated dissimilarity matrices directly with a maximum of 1000 random starts.

Phylosymbiosis analysis

We used the tick species phylogeny derived by Norris et al. (1999) from fragments of the mitochondrial 12S (~420 bp) and 16S (~460 bp) rRNA genes. Sequences were downloaded from GenBank, concatenated, and aligned in Geneious (version 6.16; www.geneious.com) using the Geneious aligner with default settings. A distance matrix was produced from the alignment using the Tamura-Nei genetic distance model (Tamura & Nei 1993), and a consensus tree was produced using majority rule (Margush & McMorris 1981) on 100 bootstrap replicates of trees generated by the unweighted pair group method with arithmetic mean (UPGMA). Genetic distance and tree calculations were performed in R using the packages "ape" (version 3.1-4) (Paradis *et al.* 2004) and "phangorn" (version 1.99-11) (Schliep 2011).

Two methods were used to derive dendrograms from each community relatedness dissimilarity matrix and to test for congruence to the tick mitochondrial gene-based phylogeny. The first approach was similar to that of Brucker and Bordenstein (2012).

Individual tick bacterial communities were pooled by species and the four dissimilarity matrices (Table 4.2) were calculated as described above. Dendrograms were produced from each matrix by UPGMA. Because matrix ordering can affect tree topology when ties occur in UPGMA (Backeljau *et al.* 1996), each matrix was systematically permuted to find all possible orderings and the majority rule method was used to find a single consensus tree for each dissimilarity matrix. Agreement between the mitochondrial and community dissimilarity trees was assessed visually.

Our second approach to comparing the mitochondrial and community dissimilarity matrices was to create a 4x4 species matrix from each 130x130 tick sample matrix by averaging the sample pairwise dissimilarities for each species pair. Nei's net number of differences between populations formula was applied to each 4x4 matrix to correct for the variability of bacterial communities within each species (equation 25, Nei & Li 1979). A Mantel test (Mantel 1967) was performed in vegan to assess the correlation between each species-level community dissimilarity matrix and the genetic distance matrix produced from the tick mitochondrial gene fragments. A consensus tree was produced for each dissimilarity matrix by permuting each matrix and applying the majority rule method. Scripts for tree calculations were written in R and in Python 2.7 using the package BioPython (version 1.64) (Cock *et al.* 2009).

Data accessibility

The full 454 sequencing data set is available in the NCBI Short Read Archive under study accession numbers PRJNA246422 (experiments SRX566183-215) and PRJNA271876.

4.4 Results

Library characteristics

More than 1000 high quality sequence reads were obtained from 130 of the 250 tick DNAs tested (Table 4.S1). From those 130 tick DNAs, I obtained a total of 514,777 high quality 16S rRNA gene sequence reads (Table 4.1) averaging 256.7 bp in length after trimming. These sequence reads comprised 207 OTUs when clustered at 97% identity, of which 78 contained a single sequence and were discarded. Oligotyping analysis split five of nine abundant OTUs into additional OTUs (Table 4.S3) for a total of 146 OTUs. Samples from each species produced similar numbers of sequence reads, but the predominant bacterial community members varied between species (Figure 4.1). The *A. americanum* samples were dominated by *Coxiella* (100% of samples, 89.4% of sequences) and *Rickettsia* (78.8% of samples, 9.5% of sequences), the *A. maculatum* by *Francisella* (100% of samples, 83.2% of sequences) and *Midichloria* (41.3% of samples, 10.9% of sequences), the *D. variabilis* by *Rickettsia* (18.8% of samples, 18.1% of sequences) and two *Francisella* OTUs (100% and 68.8% of samples, 73.6% and 3.3% of sequences, respectively), and the *I. scapularis* by *Rickettsia* (97.2% of samples, 88.1% of sequences) and *Borrelia* (50.0% of samples, 7.2% of sequences). Representative sequences from each dominant OTU were consistent with the predominant genotypes reported from each tick species in the literature (Table 4.S4).

Library purification method comparison

To ensure that the use of different amplicon purification methods for samples sequenced on plates A and B (Table 4.S1) did not bias results, the bacterial communities of *D.*

variabilis and *A. maculatum* from both plates were assessed for similarity in dispersion and tested by PERMANOVA for plate-related variation in bacterial communities. The sequencing plate did not account for a significant amount of the variation in the weighted UniFrac matrix for each species (*A. maculatum*: F=0.87, R^2=0.031, p=0.45; *D. variabilis*: F=1.12, R^2=0.036, p=0.37) and dispersions were similar (*A. maculatum*: F=1.07, p=0.32; *D. variabilis*: F=0.93, p=0.33). Unweighted UniFrac PERMANOVA results were equivocal due to significant differences in dispersion, but examination of the NMDS plot indicated that the bacterial communities of ticks sequenced on plate B (n=8) were a subset of the variation of the communities sequenced on plate A (n=24) (Figure 4.S2A). The unweighted UniFrac NMDS was similar for the *A. maculatum* samples once a single outlier from plate B that distorted the NMDS was excluded (Figure 4.S2B). Because there was no indication of systematic bias between the plates, the samples were pooled by tick species for all further analyses.

Novel community members

Candidatus 'Midichloria mitochondrii' (hereafter *M. mitochondrii*) was identified as a major component of the *A. maculatum* bacterial community in the ticks sampled. A single OTU predominated, and four minor OTUs were also present (Figure 4.1, Table 4.S3). The largest OTU (5001) was identified in five females and seven males with equal representation from both sample years. OTU5001 was identical to a genotype of *M. mitochondrii* previously amplified from *A. tuberculatum* (Table 4.S4). The second largest *Midichloria* OTU (5002) was found exclusively in one male and three female ticks collected in 2007. A single female produced >99% of the sequences, and the male's community also contained many sequences from OTU5001. OTU5002 was 99% identical

to 11 tick-derived (from seven tick species) and three horse (*Equus caballus*)-derived *M. mitochondrii* sequences in GenBank (Table 4.S4). I also detected a few (n=9) different sequences from one Ohio female *D. variabilis* and two female and one male *D. variabilis* from Georgia and in *A. americanum* (Williams-Newkirk *et al.* 2014) (Figure 4.1).

Bacterial richness and diversity

Rarefaction analysis indicated that the available bacterial OTU richness was well sampled for all tick species except *A. maculatum*, for which the slope of the curve did not approach an asymptote (Figure 4.S3). However, Good's coverage estimate showed similar levels of sampling completeness for all four species (*A. americanum* mean: 0.9995, range: 0.9986-1.0; *A. maculatum* mean: 0.9975, range: 0.9899-1.0; *D. variabilis* mean: 0.9989, range: 0.9953-1.0; *I. scapularis* mean: 0.9992, range: 0.9977-1.0). Alpha diversity did not vary between tick species (two-way ANOVA, $F(3, 122)=0.75$, $p=0.53$) or between the sexes within species ($F(3, 122)=0.58$, $p=0.63$), but males were more diverse than females when pooled across species ($F(1, 122)=7.25$, $p<0.01$) (Figure 4.2).

Bacterial community composition and structure

The composition of each bacterial community was generally species specific, with no OTUs appearing in all tick species and only four OTUs shared by three species (Figure 4.3). Of these four OTUs, only one contained more than 10 sequences (Figure 4.1, OTU2001-*Francisella*); it was found mostly in *A. maculatum* with fewer than 10 sequences found in *D. variabilis* and *I. scapularis* combined. In none of the shared OTUs containing more than 100 sequences were the sequences shared equally between tick species; the species producing fewer sequences always contained less than 10 sequences.

Dispersion tests were significant for all four of the bacterial community dissimilarity matrices (Table 4.3). PERMANOVA (Table 4.3) analysis of the dissimilarity matrices also supported a significant interaction between tick species and sex, while NMDS (Figures 4, S4-S6) supported significant interactions for only three matrices. However, sex added much less explanatory power to the models than did species (Table 4.3, R^2 values). Significant within species differences between the sexes were observed only in *A. americanum* and *I. scapularis* (Figures 4.4, 4.S5, 4.S6). Female *I. scapularis* contained a higher percentage of rickettsial endosymbiont (REIS) than males (two-way ANOVA, $F(1, 34)=4.94$, $p= 0.03$). Species were clearly delineated by all dissimilarity measures, although the clustering was much tighter in the OTU-based plots (Figures 4.4, 4.S4) than the phylogeny-based plots (Figures 4.S5, 4.S7).

Bacterial community relationships to host phylogeny

Measuring the relatedness of tick-associated bacterial communities pooled by species did not recover a consensus tree that recapitulated the mitochondrial gene phylogeny of the host tick species (Figure 4.5A) regardless of the metric of bacterial community dissimilarity used (Figures 4.5B, 4.S7A-C). However, the weighted UniFrac species-level matrix containing the means of pairwise sample comparisons for each species pair was significantly correlated with the mitochondrial gene distance matrix (Mantel test, 1000 permutations, $r=0.67$, $p<0.05$). All other matrix correlations were non-significant ($p>0.05$). The topology of the consensus tree produced from the weighted UniFrac matrix was identical to that of the mitochondrial gene tree (Figure 4.5C); the trees produced from the three other matrices varied in topology (Figure 4.S8D-F).

4.5 Discussion

Species-specific bacterial communities

Our analysis of the bacterial communities of adults of four tick species demonstrated the specificity of their microbiomes. Greater than 90% of the bacterial community variation captured by OTU-based metrics was explained by tick species. This is consistent with the findings of Hawlena *et al.* (2013), who collected co-feeding fleas and ticks from free-living hosts and found that the arthropod source explained most of the variation in their bacterial communities. In my study, the Bray-Curtis metric completely delineated tick species based on their bacterial communities (Figure 4.4), indicating that there is a combination of OTUs and their abundance that is characteristic for each of these species. The Jaccard metric supported the results of the Bray-Curtis and emphasized the limited overlap in membership between the four communities independent of abundance (Figures 4.3, 4.S4). Refinement in the detection of these patterns was made possible by the use of Oligotyping to split abundant OTUs; a strict 97% identity OTU definition would have obscured consistent differences in the genotypes of several predominant OTUs (*Francisella* and *Rickettsia*, among others; Table 4.S3) associated with each tick species. Narrower OTU definitions are frequently used in this type of analysis, but can have the undesirable effect of creating additional small, analytically uninformative OTUs which may reflect either real minor variants in the community or errors introduced by PCR or sequencing.

Phylogeny-based dissimilarity metrics also showed significant clustering of bacterial communities by tick species, but the clusters were not as well defined. Greater overlap

was observed on NMDS plots between species clusters for both the weighted (Figure 4.S5) and unweighted (Figure 4.S7) UniFrac metrics. The weighted UniFrac separated the species much better, analogous to the Bray-Curtis analysis but it was less efficient than the latter (Figures 4.4, 4.S5). The reduction in resolution between species in the phylogeny-based metrics versus the OTU-based metrics emphasized the sharing of broad taxa, but not specific genotypes, between species-specific communities.

Known and novel community members

This is this first report of *M. mitochondrii* in *A. maculatum* (48.3% prevalence), despite the inclusion of the species in the extensive survey of ixodid ticks performed by Epis *et al.* (2008). The previous study included only six females collected in Florida (site unspecified) and South Carolina; I do not know the proximity of the population I sampled to that of Epis *et al*. The predominant *M. mitochondrii* genotype detected was identical to that reported from *A. tuberculatum* (Epis *et al.* 2008). The two tick species have overlapping ranges and can be found in similar habitat types (Estrada-Peña *et al.* 2005; Ennen & Qualls 2011). The host ranges of the two species overlap primarily during the immature stages (Guglielmone *et al.* 2014), making horizontal transmission of the bacteria possible. Alternatively, each species may host a specific *M. mitochondrii* genotype which does not differ in the short region of the *rrs* sequenced here. The less abundant genotype detected in *A. maculatum* was similar to several strains of *M. mitochondrii* previously reported from *A. americanum* as well as other ticks and vertebrate hosts. Further work is needed to determine the prevalence of these strains in both inland and coastal *A. maculatum* populations.

We also detected a small number of *M. mitochondrii* sequences from *D. variabilis* collected in Ohio and Georgia, while Epis et al. (2008) did not find it in a small number of ticks from Florida. Additional surveys of the species with *M. mitochondrii*-specific assays will be necessary to confirm this finding.

A limited number of NGS surveys of whole bacterial communities were performed previously on these tick species. The literature on *A. americanum* bacterial communities has been summarized recently (Williams-Newkirk *et al.* 2014). *My* results with Florida collected *A. maculatum* were consistent at the phylum level with those of Budachetri *et al.* (2014) who pyrosequenced 16S rDNA (V1-V3) amplicons from the salivary gland and midgut-derived DNAs of several females field-collected in Mississippi. *Francisella*, *Propionibacterium*, *Rickettsia*, *Pseudomonas*, and *Corynebacterium* were found in both samples. Notably, I did not detect *Lactococcus*, *Raoultella*, or *Wolbachia* which were found commonly in the Mississippi ticks. In both studies the dominant *Rickettsia* sequences were identical to *R. parkeri*, a known emerging pathogen transmitted by *A. maculatum*. A second high abundance *Rickettsia* genotype was also detected here in one tick, which is likely either *Candidatus* 'Rickettsia andeanae' (Paddock *et al.* 2010) or the so-called rickettsial endosymbiont of *A. maculatum* (Budachetri *et al.* 2014) (possibly the same agent). Due to a paucity of publically available *rrs* sequence data for these two agents, I was unable to confirm its identity.

Two previous studies have used NGS to examine *I. scapularis* whole bacterial communities in both adults and nymphs (Yuan 2010; Hawlena *et al.* 2013), but their pooled data from multiple life stages does not permit a direct comparison to my adult data. As in these studies, I also found a predominance of *Rickettsia* consistent with the

known REIS. Similar to Hawlena *et al.*, I also identified Bacillales groups, although not at the high relative abundance (~39%) found previously; *Anaplasma* and *Borrelia* were more common in my samples. Some differences between my results and these studies may be due to differences in the region of the *rrs* amplified or real differences in community composition between tick populations. Interestingly, *Wolbachia* were found in 11% of my samples (two males and two females) but not in either previous NGS study, although a Sanger sequencing study had previously reported *Wolbachia* in one of four *I. scapularis* nymphs from Nantucket (Benson *et al.* 2004).

To my knowledge only Hawlena *et al.* (2013) have examined *D. variabilis* bacterial communities by NGS. My findings among the major community members differed only in that I found a small number of Bacillales sequences which may or may not correspond to the genus *Brevibacillus*, which constituted nearly one-third of the community previously identified. Whether this is a real difference between populations, an artifact of differing *rrs* region targets, or results from the inclusion of nymphs in the Hawlena *et al.* sample remains to be determined. I also found that co-infection of *D. variabilis* with two distinct *Francisella* 16S genotypes was common (22/32 samples). OTU2002 was present in all samples and always found in higher abundance when co-infection occurred with OTU2003. These genotypes were consistent with those previously identified (Goethert & Telford 2005). Among the *Rickettsia* OTUs identified from *D. variabilis* here, OTU1002 was the most prevalent (18.8%) and consistent with *R. bellii*, which is among the more common *Rickettsia* in *D. variabilis* (Pretzman *et al.* 1990). The OTU was found in the Ohio and Panola Mountain, Georgia populations only.

Sex-specific bacterial communities

Our previous work in *A. americanum* (Williams-Newkirk *et al.* 2014) found differences in the structure of adult male and female bacterial communities. Specifically, female communities were enriched for *Coxiella* while male communities tended to contain more *Rickettsia*. I hypothesized that sexually dimorphic bacterial communities may be common among ixodid ticks if obligate symbionts replicate in females to ensure transmission to eggs. However, sexual dimorphism was detected only in the *A. americanum* and *I. scapularis* samples. *Ixodes scapularis* sexes were similar to *A. americanum* sexes in that they differed only in the structure of their bacterial communities and not their membership when measured by OTU-based metrics (Figures 4.4 and 4.S4). This difference appears to be driven by an increase in the percentage of REIS present in females, which is consistent with the transmission to eggs hypothesis. Sexual dimorphism in the bacterial communities of *I. scapularis* was previously detected by Moreno *et al.* (2006) using temporal temperature gradient gel electrophoresis.

Sexual dimorphism was also observed in the unweighted UniFrac NMDS plot for both *A. americanum* and *I. scapularis* (Figure 4.S6). In both species males as a group contained more bacterial OTUs than females (Figure 4.1), but on average individual males did not have higher alpha diversity (Figure 4.2) or OTU richness (data not shown) than females. The greater variability of male bacterial community membership in these two species when measured by a phylogeny-based metric but not by an OTU-based metric (Figure 4.S4) suggests that an increased number of minor variants detected in males may be driving the difference in membership by unweighted UniFrac. The causal mechanism for the increased variability in males relative to females may be either a deregulation of male bacterial communities or inadequate detection of minor variants in females due to the

predominance of the primary endosymbiont. More precise, quantitative studies will be needed to address these questions.

It is possible that the presence or absence of sexual dimorphism in community structure is determined by the specific function(s) of the primary endosymbiont in a given tick species. For example, it may be that the *Coxiella* and REIS symbionts provide a selective advantage to *A. americanum* and *I. scapularis* primarily during the immature life stages or adult females during reproduction. Therefore, the females must retain higher density populations of their symbionts to ensure transmission to offspring, while there is no selection pressure on adult males to maintain large endosymbiont populations. Conversely, it may be that the *Francisella* symbionts of *A. maculatum* and *D. variabilis* provide a selective advantage to the functions of both males and females, necessitating similar population sizes in both sexes. An understanding of the functional relationship between ticks and their endosymbionts will be necessary to resolve these questions.

Phylosymbiosis

Patterns consistent with phylosymbiosis have been observed across many different host lineages. Ochman *et al.* (2010) were able to recreate the branching order of the mitochondrial gene tree for six great ape species from the relationships between each species' fecal bacterial communities. Brucker and Bordenstein (2012) were also able to recover the host phylogeny branching order among three species of *Nasonia* wasps using bacterial communities characterized by *rrs* gene cloning and Sanger sequencing. Additionally, nearly 80% of the variation in the gut bacterial community was correlated to the phylogenetic relationships of 25 turtle ant (*Cephalotes*) species (Sanders *et al.*

2014). Many other example studies could be cited (reviewed in Vavre & Kremer 2014), but to my knowledge no prior studies have examined ixodid ticks for phylosymbiosis.

Methods to assess datasets for phylosymbiosis have varied widely between studies. Brucker and Bordenstein (2013), who coined the term "phylosymbiosis", pooled the bacteria detected from all samples belonging to a variable group and calculated dissimilarity metrics between the pools. Hierarchical clustering was used to generate dendrograms, which were visually compared. I included this approach in my analysis for comparative purposes and found that with all dissimilarity metrics it consistently clustered *A. maculatum* and *D. variabilis* together; the relationships of *I. scapularis* and *A. americanum* varied depending on whether or not the metric was OTU or phylogeny-based (Figures 4.5 and 4.S8). While this method is easily understood and implemented, it does not take advantage of the statistical power provided by analyzing individual replicate samples. Furthermore, the bacterial communities of different individual hosts of a given species frequently encompass significant variation, which is likely overwhelmed by the dominant bacterial community members when data are pooled. Moreover, when using Brucker and Bordenstein's approach but without pooling by variable groups, I found that the intraspecific variation led to bushy, difficult to interpret dendrograms that could not be easily compared to the tick mitochondrial gene tree (data not shown).

Other methods of analysis have been used. Notably, Ochman et al.'s (2010) treatment of fecal bacterial OTUs as quantitative morphological traits benefits from the long history of this technique in taxonomy, but fails to capture the phylogenetic relationships of the bacteria. More recently, Sanders et al. (2014) used a novel method varying the width of

OTU bins and multiple measures of community relatedness to estimate quantitatively the amount of influence phylosymbiosis has on bacterial community assembly.

Our approach sought to account for intraspecific variation while reducing the complexity of the bacterial community dataset so that it could be readily compared to the mitochondrial gene matrix. I created 4x4 matrices comparing the bacterial communities of tick species from 130x130 tick sample matrices by taking the mean of the individual tick pairwise comparisons for each tick species pair and applying a correction for intraspecific variation. I preferred to assess the correlation between the dissimilarity matrices and the mitochondrial gene-derived matrix directly using a Mantel test because it assesses a specific hypothesis, as opposed to the comparison of dendrograms produced by cluster analysis, which is more suited to data exploration (Ramette 2007).

UniFrac metrics consider the shared branch length between two communities (Lozupone & Knight 2005; Lozupone *et al.* 2007), which may make them more effective at capturing the signal predicted by phylosymbiosis, *i.e.*, that more closely related hosts should have more closely related but not necessarily identical microbiota. Of the four community dissimilarity metrics used here, only the weighted UniFrac was significantly correlated with the mitochondrial gene distance (r=0.67). The weighted UniFrac appears to be capturing the similarly high abundances of alpha and gamma Proteobacteria in the two *Amblyomma* species, whereas the *I. scapularis* and *D. variabilis* communities are biased towards either the alpha or gamma Proteobacteria, respectively (Figure 4.1). The unweighted UniFrac metric did not produce the same pattern of community relatedness, supporting the conclusion that it was the abundance of taxa which drove the weighted UniFrac relationships.

None of the OTU-based metrics of community dissimilarity were significantly correlated with tick mitochondrial gene distance, which is unsurprising given the discreet and nearly equidistant clustering of bacterial communities by tick species (Figures 4.4 and 4.S4).

As was discussed by Sanders *et al*. (2014), the metrics and OTU definitions through which phylosymbiosis is detected may be used to suggest mechanisms driving signal emergence. While not directly addressed by Sanders *et al.*, host-bacterial community co-evolution is suggested if phylosymbiosis signals are found in phylogeny-based metrics while OTU-based metrics delineate species such that they are completely separated without differing degrees of relatedness. The pattern I report for these ixodid tick bacteriomes most closely resembles the last of these mechanisms, although the relatedness of the two *Amblyomma* species' bacterial communities appears dependent on similarity in alpha and gamma Proteobacteria abundance rather than a more recent divergence between their community members compared to the other tick species examined here. Bacterial abundance has been repeatedly shown to be regulated by the host's immune system in arthropods, notably in *Drosophila* (reviewed in Vavre & Kremer 2014). Dysregulation of the bacteriome by the immune system has also been suggested as a mechanism reducing hybrid vigor in *Nasonia* (Brucker & Bordenstein 2013).

The correlation between the bacterial weighted UniFrac matrix and the tick mitochondrial gene distance explained approximately 45% of the variation in the bacterial communities between these tick species. Identifying the variables which determine the remaining half will be an interesting future challenge. The most robustly supported clade across all of the metrics used here included *D. variabilis* and *A. maculatum*. While these species share

similar *Francisella* symbionts, the relationship was seen in both of the metrics which do not consider abundance (Figures 4.S4 and 4.S7), suggesting additional bacterial clades are involved. *Dermacentor variabilis* and *A. maculatum* are more commonly found in open habitats compared to *A. americanum* and *I. scapularis*, which prefer forested habitats (Hooker *et al.* 1912; Sonenshine & Levy 1972; Estrada-Peña & Jongejan 1999; Ennen & Qualls 2011). Whether the similarity in bacterial communities between these two species results from adaptation to shared environments or hosts remains to be determined. Host use by ticks is also likely to directly influence bacterial communities. Besides acting as reservoirs introducing bacteria into ticks, tick hosts may also exert negative influences on bacteria. For example, *Ixodes pacificus* nymphs fed on immunized western fence lizards (*Sceloporus occidentalis*) were cleared of *Borrelia burgdorferi* infection (Lane & Quistad 1998). In addition, tick age and environmental conditions have been shown to impact bacterial communities. Adult *A. americanum* bacterial communities become less diverse with age, and the relative ratios of bacterial community members are different when ticks are held under constant laboratory conditions compared to fluctuating field conditions (Menchaca *et al.* 2013).

The role of inheritance, either of bacterial symbionts or bacteria-selecting immune genes, leads phylosymbiosis theory to predict that the pattern of host bacterial community relatedness may be present during any host life stage. Brucker and Bordenstein (2012) found evidence of phylosymbiosis within the adults and pupae of three *Nasonia* species, but not the larvae. My previous work demonstrated a difference between the bacterial communities of *A. americanum* adults and nymphs in both membership and structure (Williams-Newkirk *et al.* 2014) and Hawlena et al. (2013) reported a significant increase

in the abundance of *Francisella* in *D. variabilis* nymphs compared to larvae, but to my knowledge no work has specifically compared the bacterial communities of the immature life stages of the four species examined here. Future work should examine the relatedness of the egg, larval, and nymphal bacterial communities of these ticks to determine if a composition consistent with phylosymbiosis is present.

4.6 Figures

Figure 4.1 Distribution of bacterial operational taxonomic units (OTU) by tick species and sex. The heat map provides the mean relative abundance value of each bacterial OTU in each tick group. Note the log scale of the color legend; white indicates the absence of an OTU. The dendrogram at left was created by UPGMA clustering of an uncorrected pairwise distance matrix based on a single representative sequence from each of the 124 OTUs remaining after each tick's bacterial community was standardized to 1000 sequences. The lowest available taxon to which all members of an OTU could be assigned is given at right with phylum and class groupings highlighted.

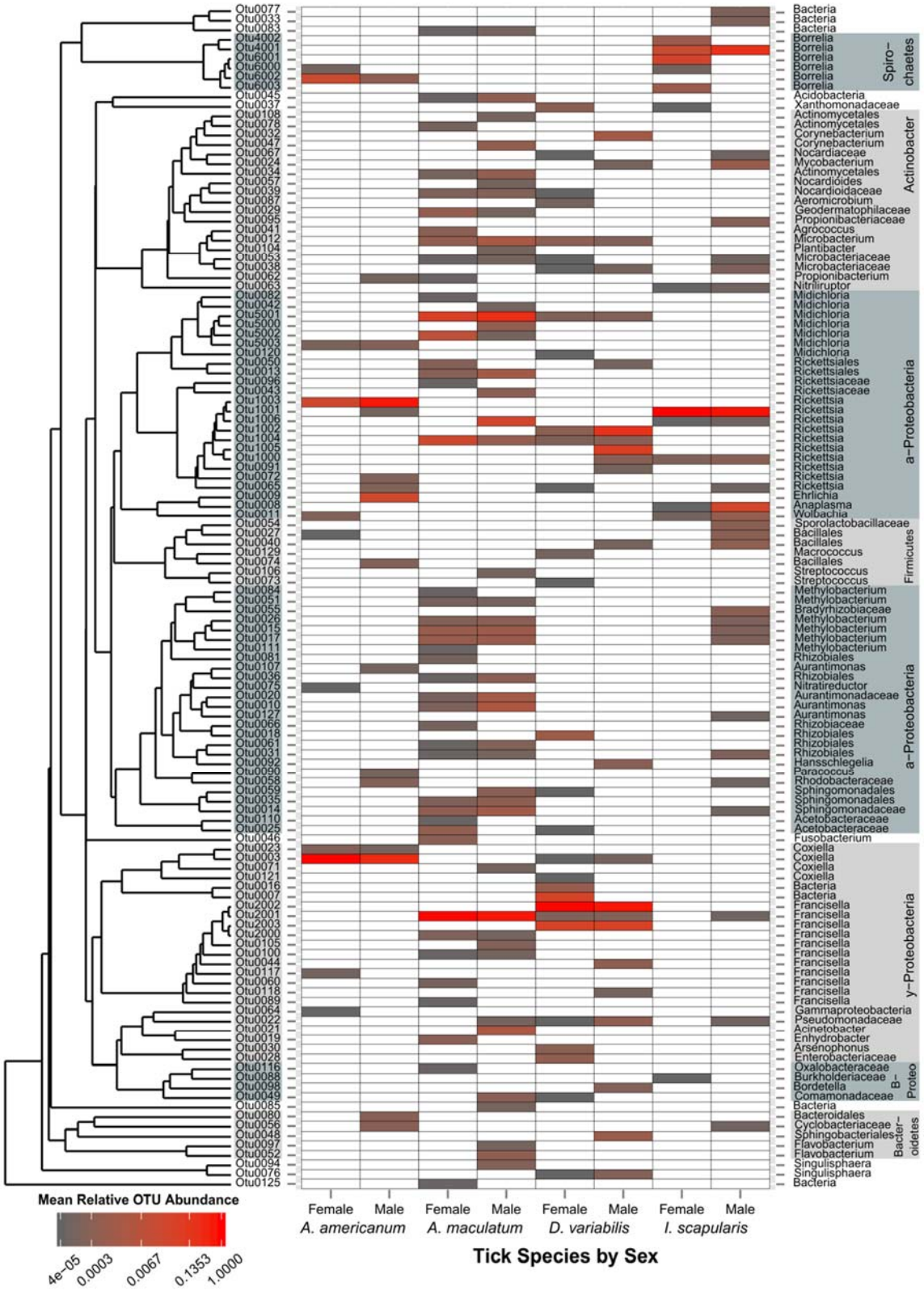

Mean Relative OTU Abundance
4e−05
0.0003
0.0067
0.1353
1.0000
Female
Male
A. americanum
A. maculatum
D. variabilis
I. scapularis
Tick Species by Sex
Spiro−chaetes
Actinobacter
a−Proteobacteria
Firmicutes
a−Proteobacteria
y−Proteobacteria
B−Proteo
Bacter−oidetes

Figure 4.2 Bacterial community alpha diversity by tick species and sex. Box and whiskers plots represent the first and third quartiles (hinges), the median (bold line), and 1.5 times the interquartile range (error bars). Outlier points are plotted individually. Within each of three panels, n=130 (see Table 4.1) and groups with an asterisk between them are significantly different at α=0.05. Group sample sizes are given in parentheses after the group name.

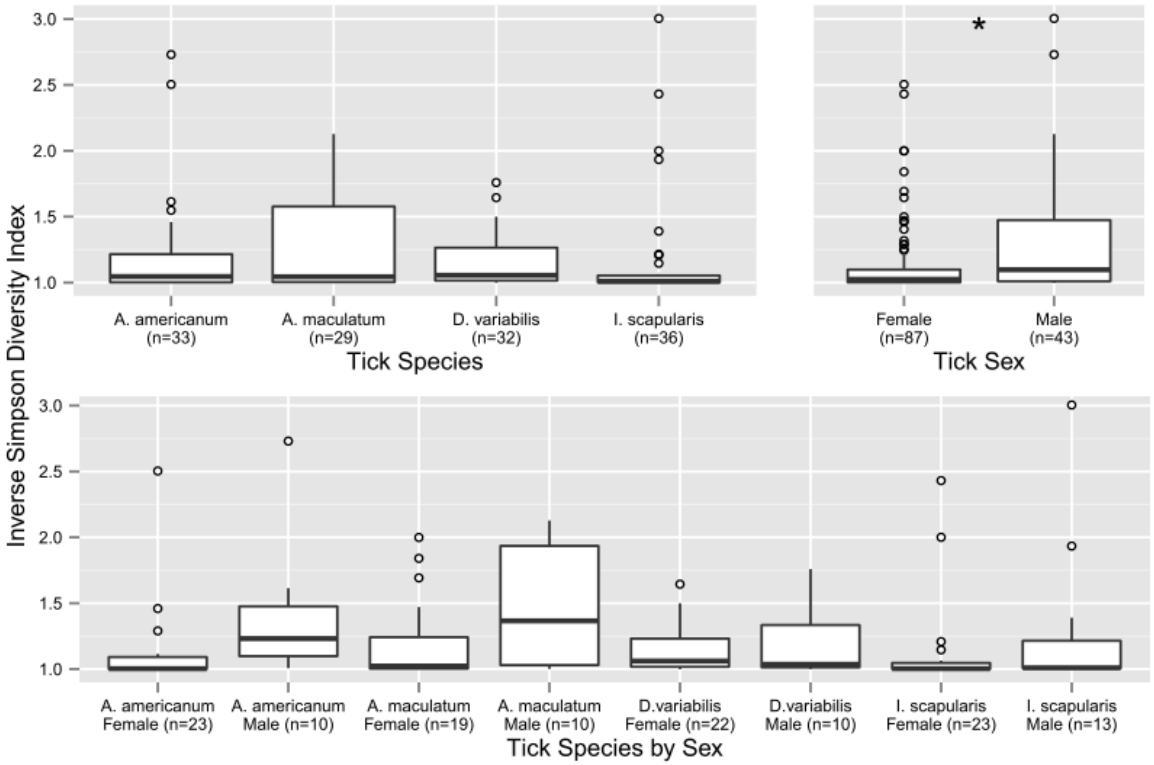


Figure 4.3 Venn diagram of bacterial OTU richness by tick species. Each tick's bacterial community was rarefied to 1000 sequences prior to comparison, for a total of 124 OTUs across all species. The number of ticks and the total number of OTUs found in each species is given under each species label.

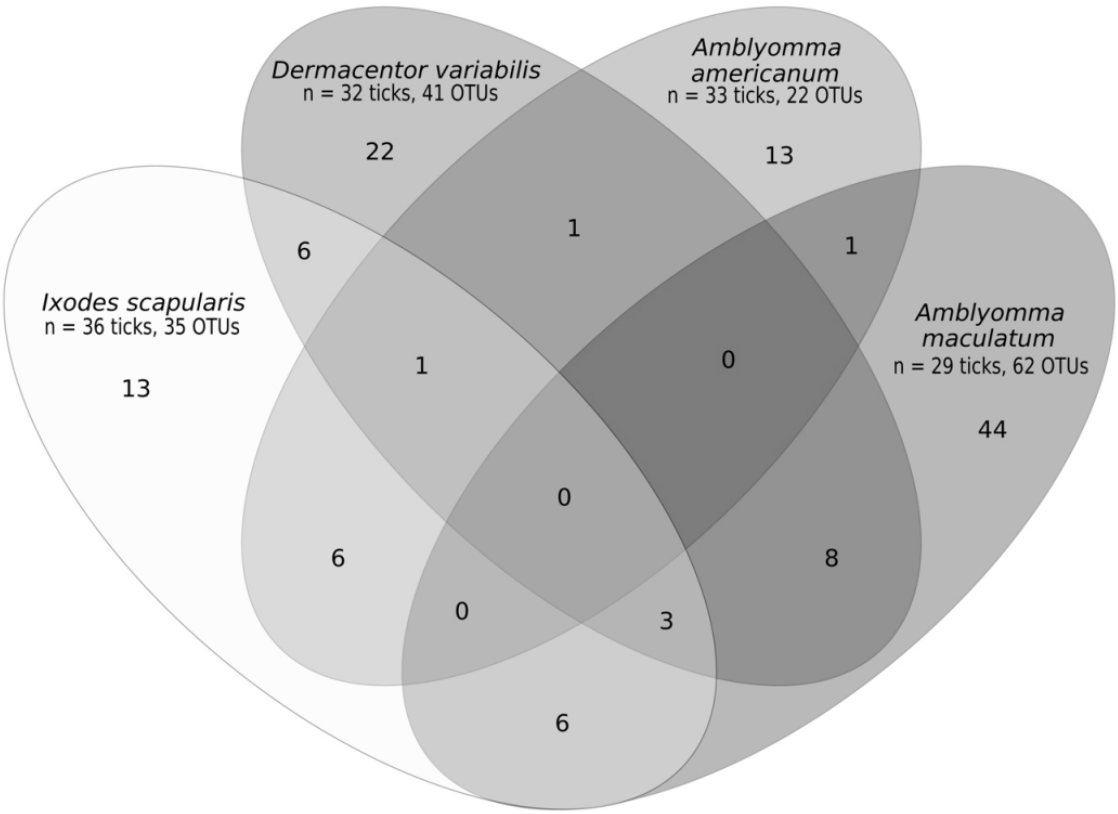


Figure 4.4 Non-metric multi-dimensional scaling (NMDS) of a Bray-Curtis dissimilarity matrix describing tick bacterial communities by species. Each point symbolizes a single tick's community (n=130); some points may overlap completely. Point and ellipse colors indicate species. Ellipses represent 95% confidence intervals around the centroids of each sex within species; non-overlapping centroids are considered significantly different at α=0.05. R^2 values in the upper left corner of plots describe the amount of variation in the data set explained by the groupings. The stress value given is a measure of the disagreement between the rank order in the original data set and that in the NMDS (lower numbers indicate better agreement).

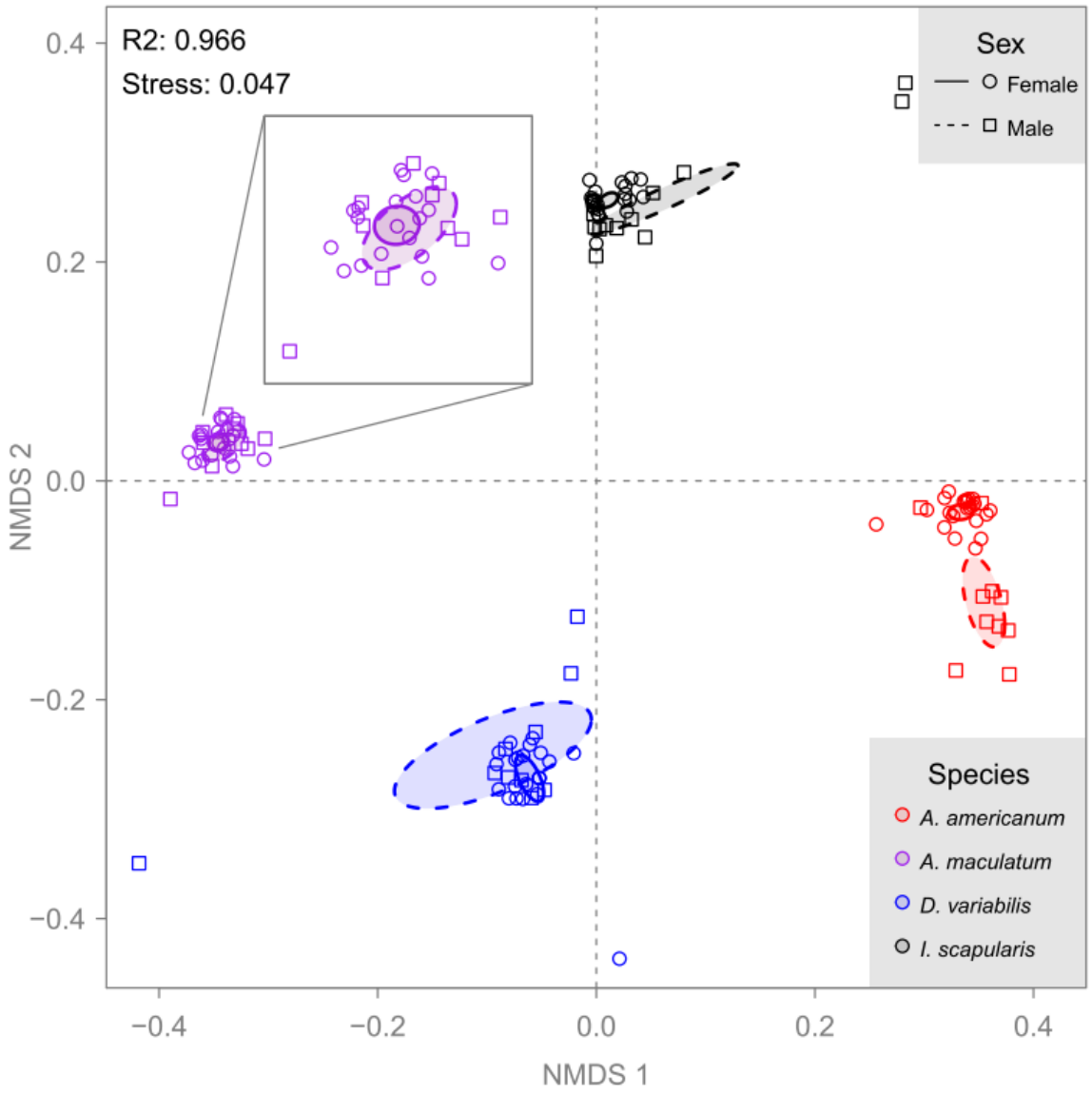

R2: 0.966
Stress: 0.047
Sex
Female
Male
Species
A. americanum
A. maculatum
D. variabilis
I. scapularis
NMDS 1
NMDS 2
0.4
0.2
0.0
−0.2
−0.4
−0.4
−0.2
0.0
0.2
0.4

Figure 4.5 Tick species relatedness dendrograms based on mitochondrial sequences and bacterial communities. All dendrograms are majority rule consensus trees summarizing bootstrapped (genetic distance) or permuted (community dissimilarity) unweighted pair group method with arithmetic mean (UPGMA) trees. Branch supports are percentages. A) Mitochondrial gene sequence tree after Norris *et al.* (1999). B) Bacterial community relatedness tree based on the weighted UniFrac dissimilarity metric; all samples for each species were pooled as in Brucker and Bordenstein (2012). C) Bacterial community relatedness tree based on the weighted UniFrac metric; mean pairwise sample dissimilarities were calculated for each tick species pair and corrected for within species variation.

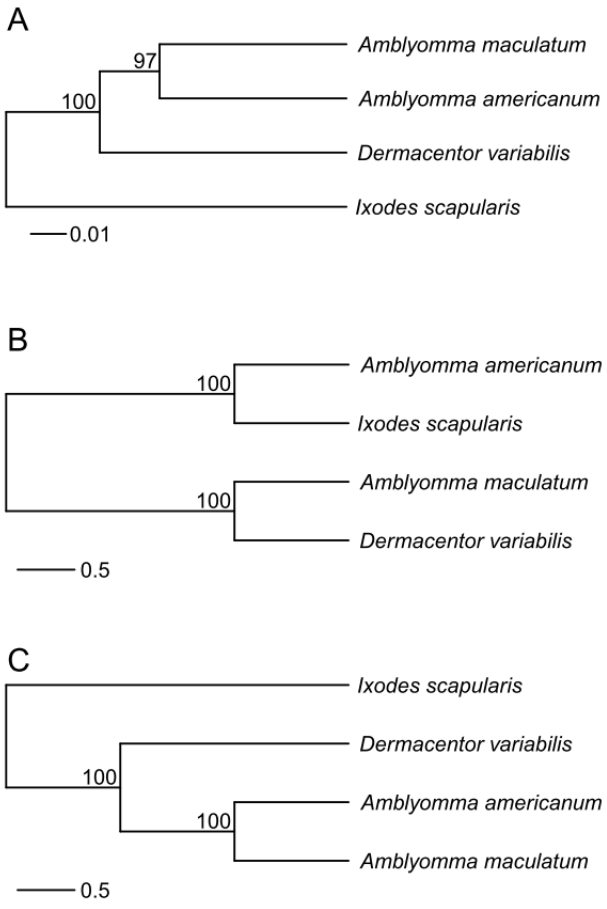


Figure 4.S1 Eastern United States tick species distribution map. Collection sites for this study are labeled as follows: 1 – Lake Dawson, Connecticut, 2 - Muskingum County, Ohio, 3 - Sweetwater Creek State Park, Georgia, 4 – Panola Mountain State Park, Georgia, 5 – Tate's Hell State Forest, Florida. Tick species are abbreviated in the legend as Aa – *Amblyomma americanum*, Am – *Amblyomma maculatum*, Dv – *Dermacentor variabilis*, and Is – *Ixodes scapularis*.

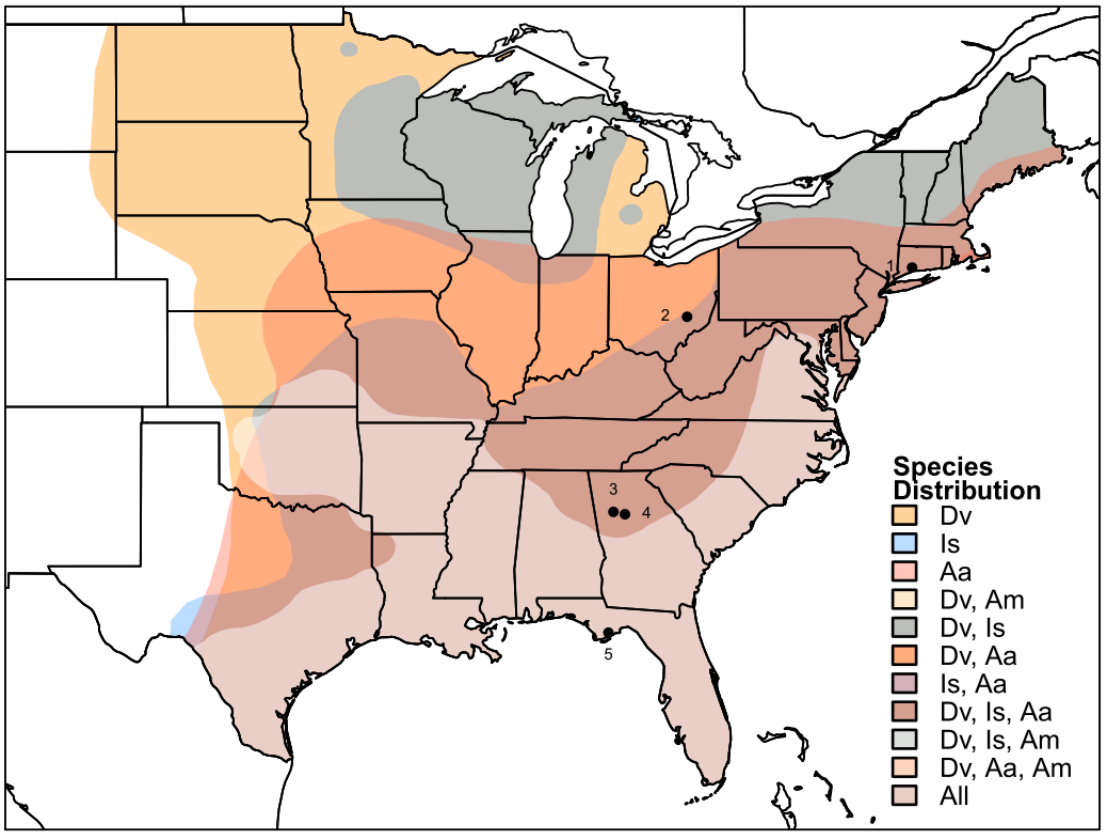
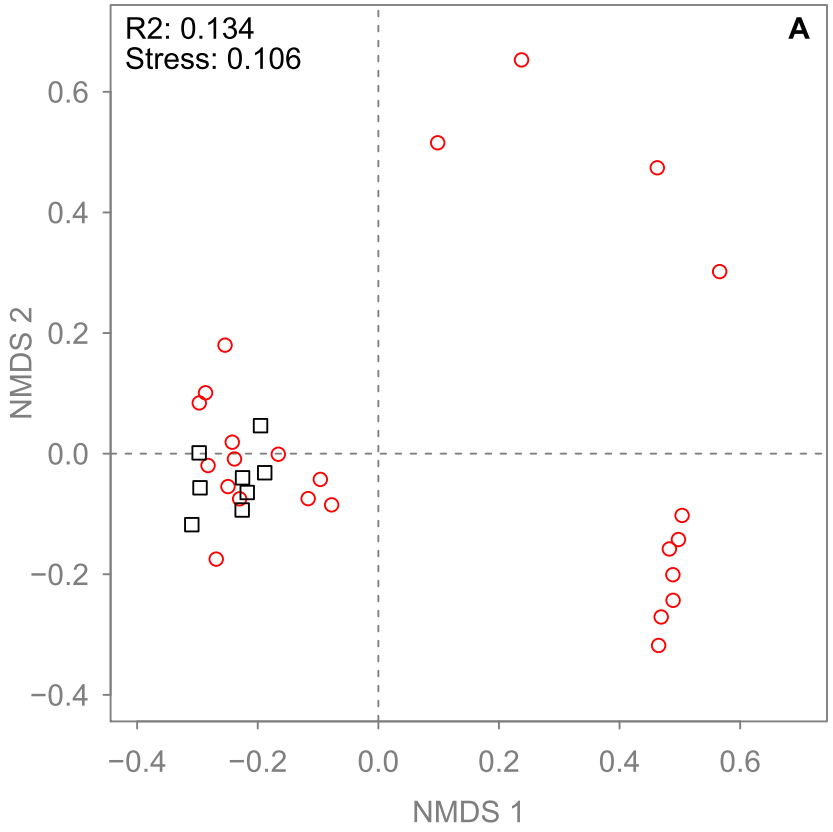


Figure 4.S2 Non-metric multi-dimensional scaling (NMDS) of unweighted UniFrac dissimilarity matrices describing tick bacterial communities by sequencing plate. Each point symbolizes a single tick's community; some points may overlap completely. Point colors and shapes indicate sequencing plate. R2 values in the upper left corner of plots describe the amount of variation in the data set explained by the groupings. The stress value given is a measure of the disagreement between the rank order in the original data set and that in the NMDS (lower numbers indicate better agreement). Panel A: *Dermacentor variabilis* (n=32), panel B: *Amblyomma maculatum* (n=29).


R2: 0.134
Stress: 0.106
A
NMDS 2
NMDS 1
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.4
−0.2
0.0
0.2
0.4
0.6

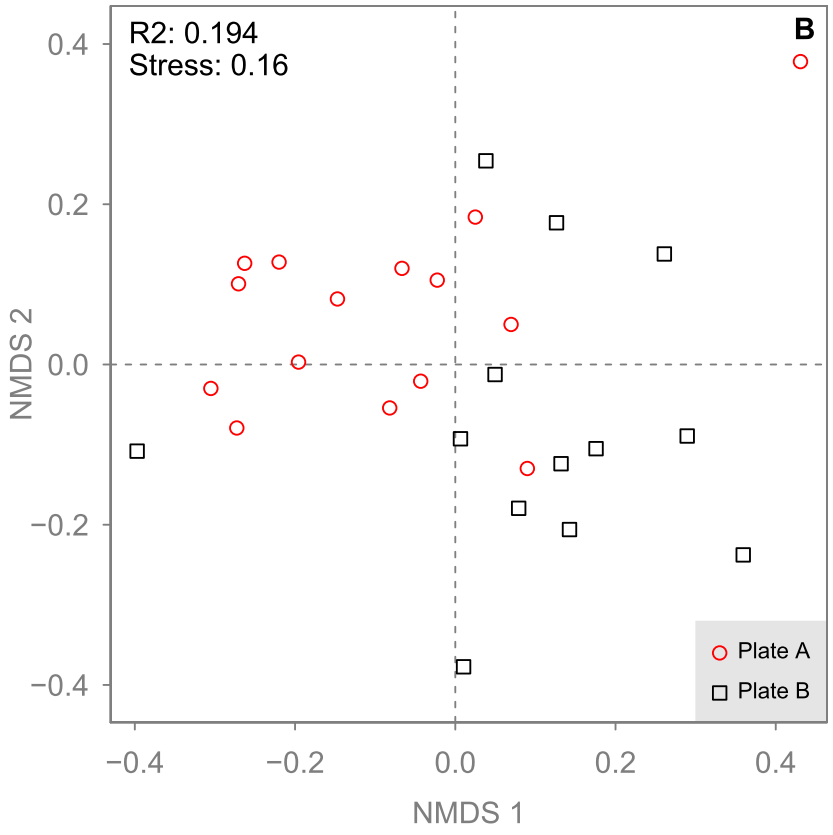

R2: 0.194
Stress: 0.16
B
NMDS 2
NMDS 1
0.4
0.2
0.0
−0.2
−0.4
−0.4
−0.2
0.0
0.2
0.4
Plate A
Plate B

Figure 4.S3 Bacterial community rarefaction curves for each tick species. The shaded regions represent the conditional 95% confidence interval obtained from 1000 randomizations of the data.

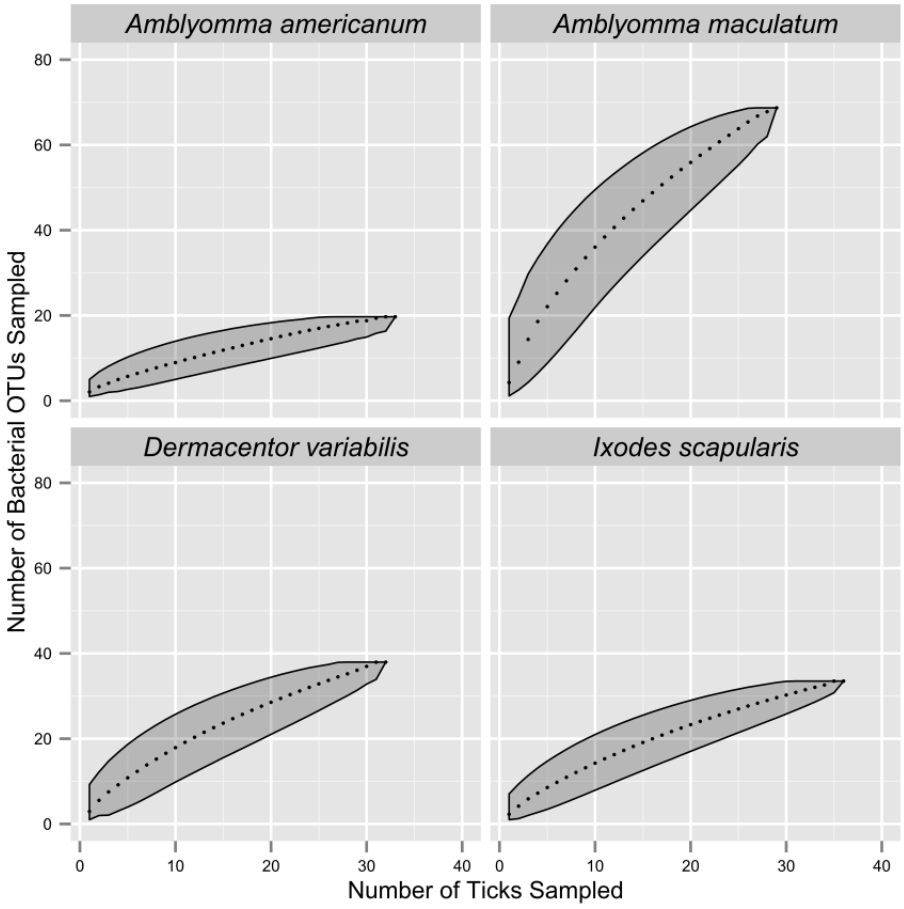


Figure 4.S4 Non-metric multi-dimensional scaling (NMDS) of a Jaccard dissimilarity matrix describing tick bacterial communities by species and sex. Each point symbolizes a single tick's community (n=130); some points may overlap completely. Point and ellipse colors indicate species, and point shapes and ellipse borders indicate sex. Ellipses represent 95% confidence intervals around the centroids of each sex within species; non-overlapping centroids are considered significantly different at $\alpha=0.05$. R^2 values in the upper left corner of plots describe the amount of variation in the data set explained by the groupings. The stress value given is a measure of the disagreement between the rank order in the original data set and that in the NMDS (lower numbers indicate better agreement).

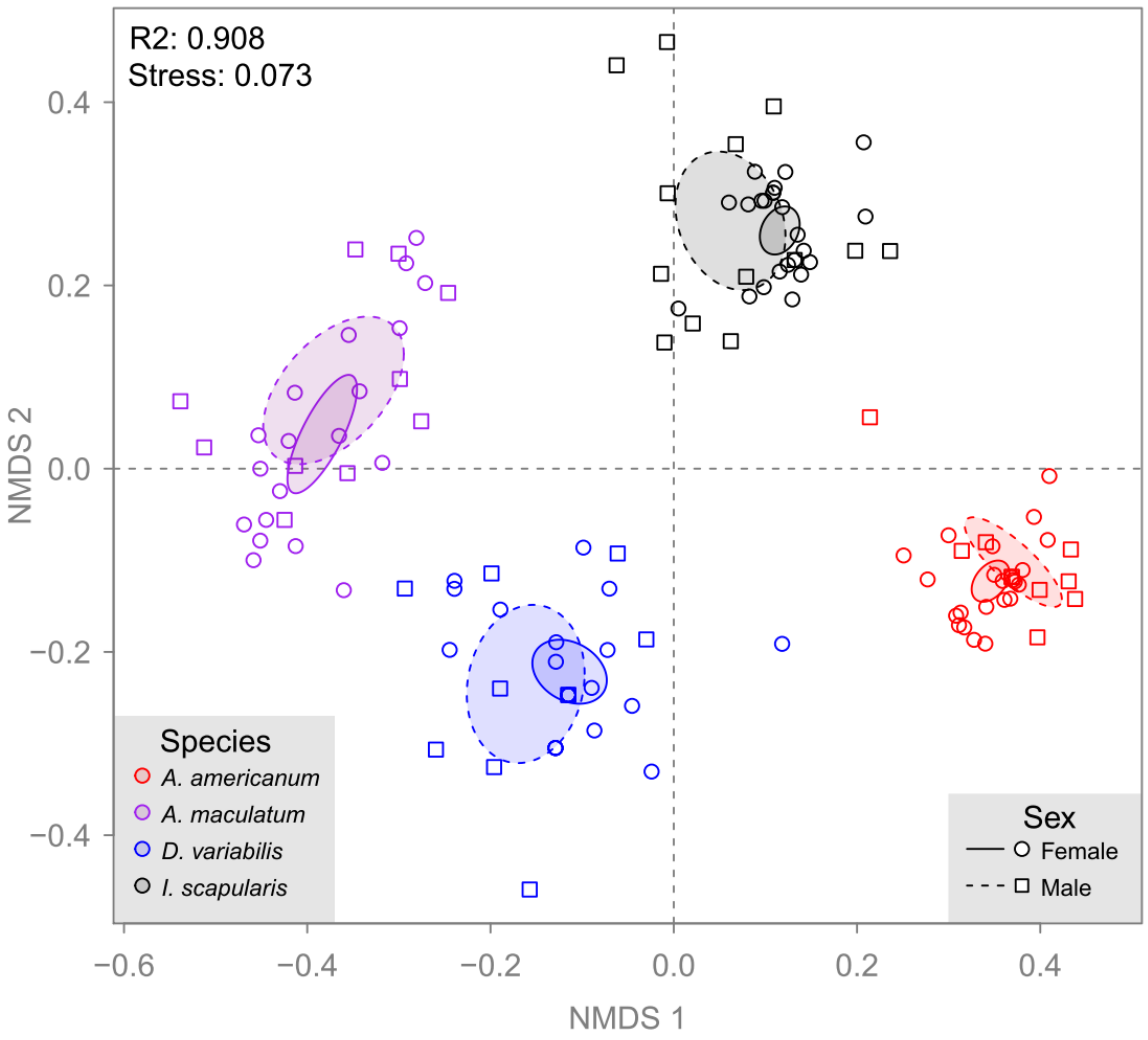

R2: 0.908
Stress: 0.073
NMDS 2
NMDS 1
0.4
0.2
0.0
−0.2
−0.4
−0.6
−0.4
−0.2
0.0
0.2
0.4
Species
A. americanum
A. maculatum
D. variabilis
I. scapularis
Sex
Female
Male

Figure 4.S5 Non-metric multi-dimensional scaling of a weighted UniFrac dissimilarity matrix describing tick bacterial communities by species and sex. See the Figure 4.S4 legend for additional details.

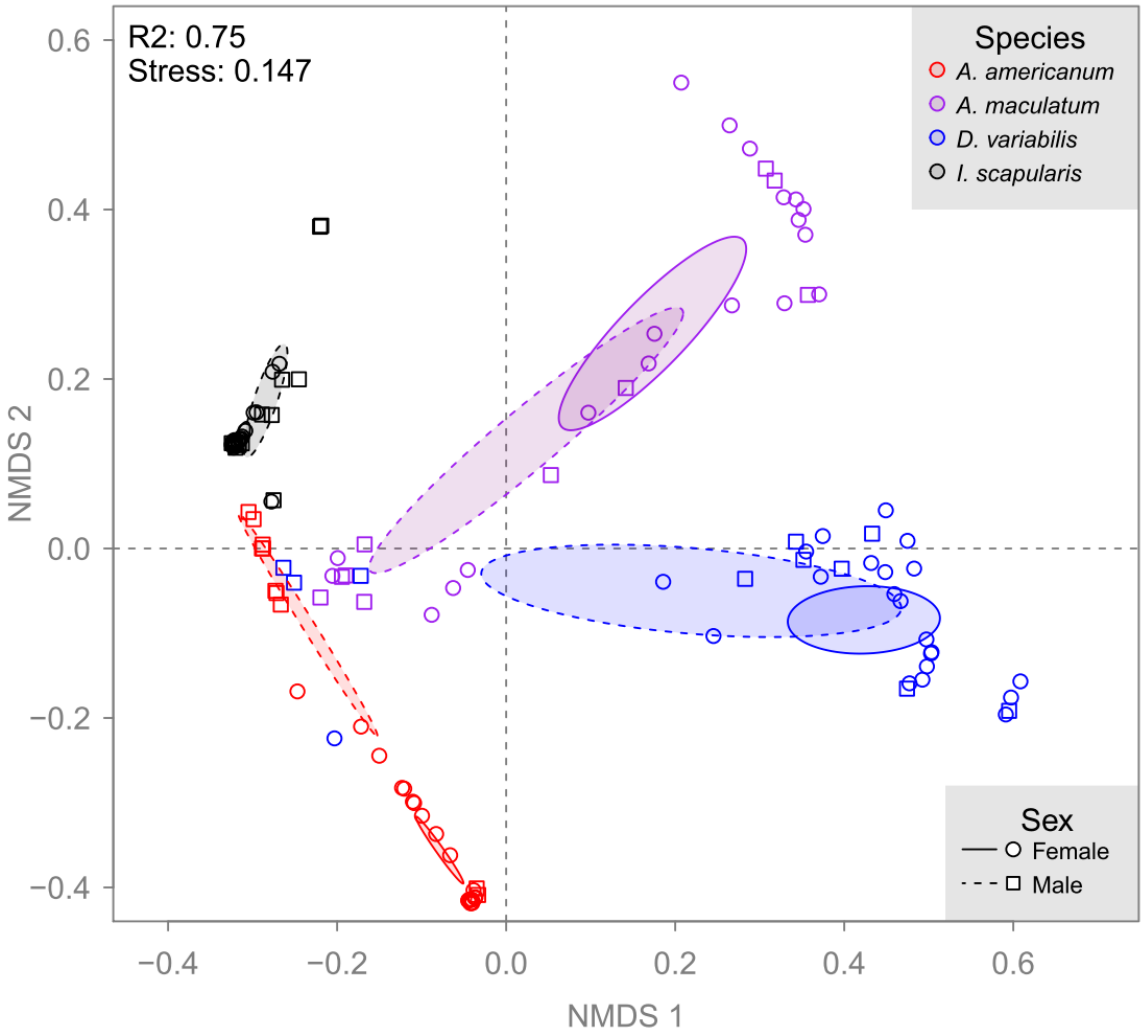


Figure 4.S6 Non-metric multi-dimensional scaling of an unweighted UniFrac dissimilarity matrix describing tick bacterial communities by species and sex. See the Figure 4.S4 legend for additional details.

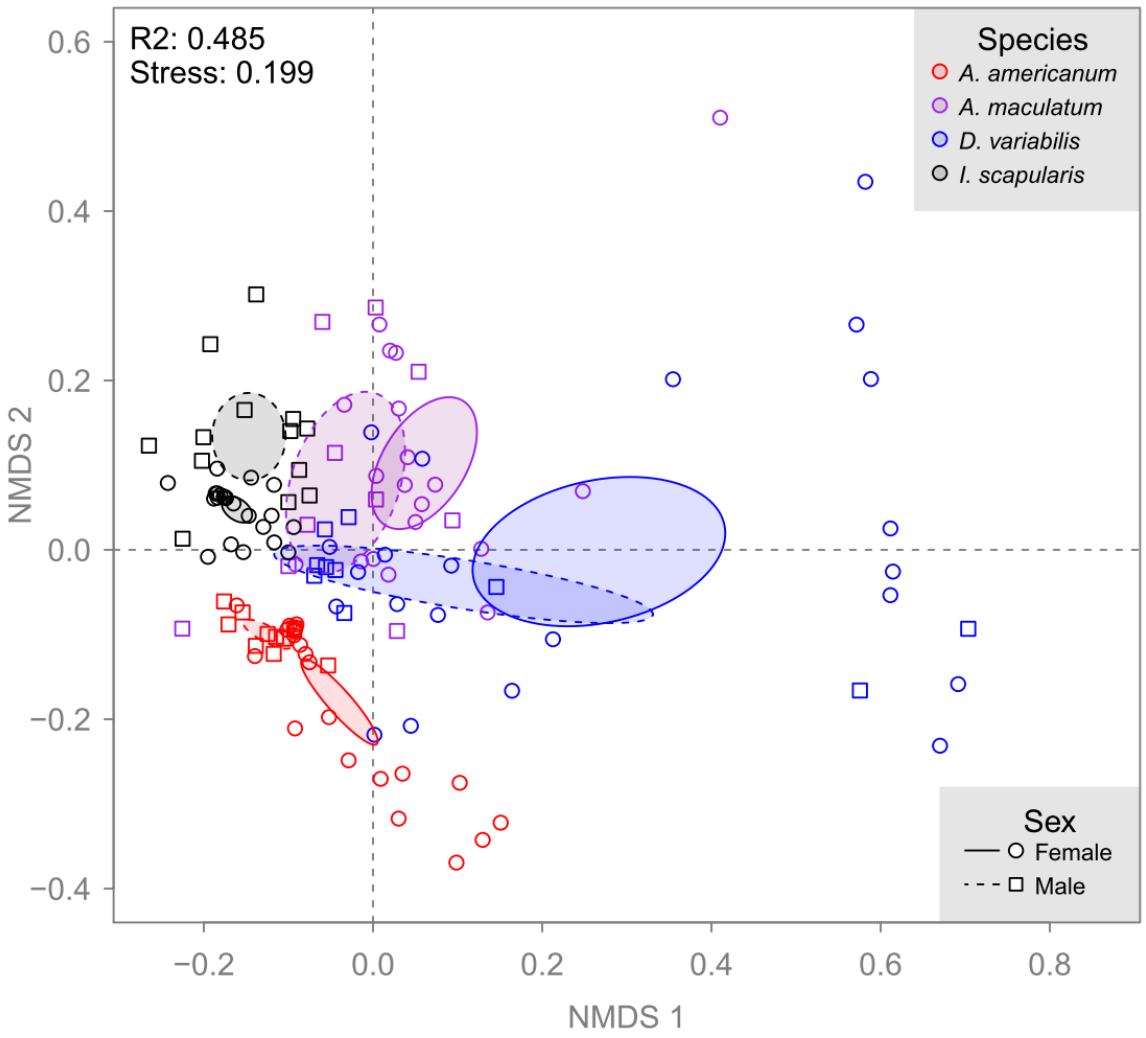


Figure 4.S7 Non-metric multi-dimensional scaling of an unweighted UniFrac dissimilarity matrix describing tick bacterial communities by species. See the Figure 4.S4 legend for additional details.

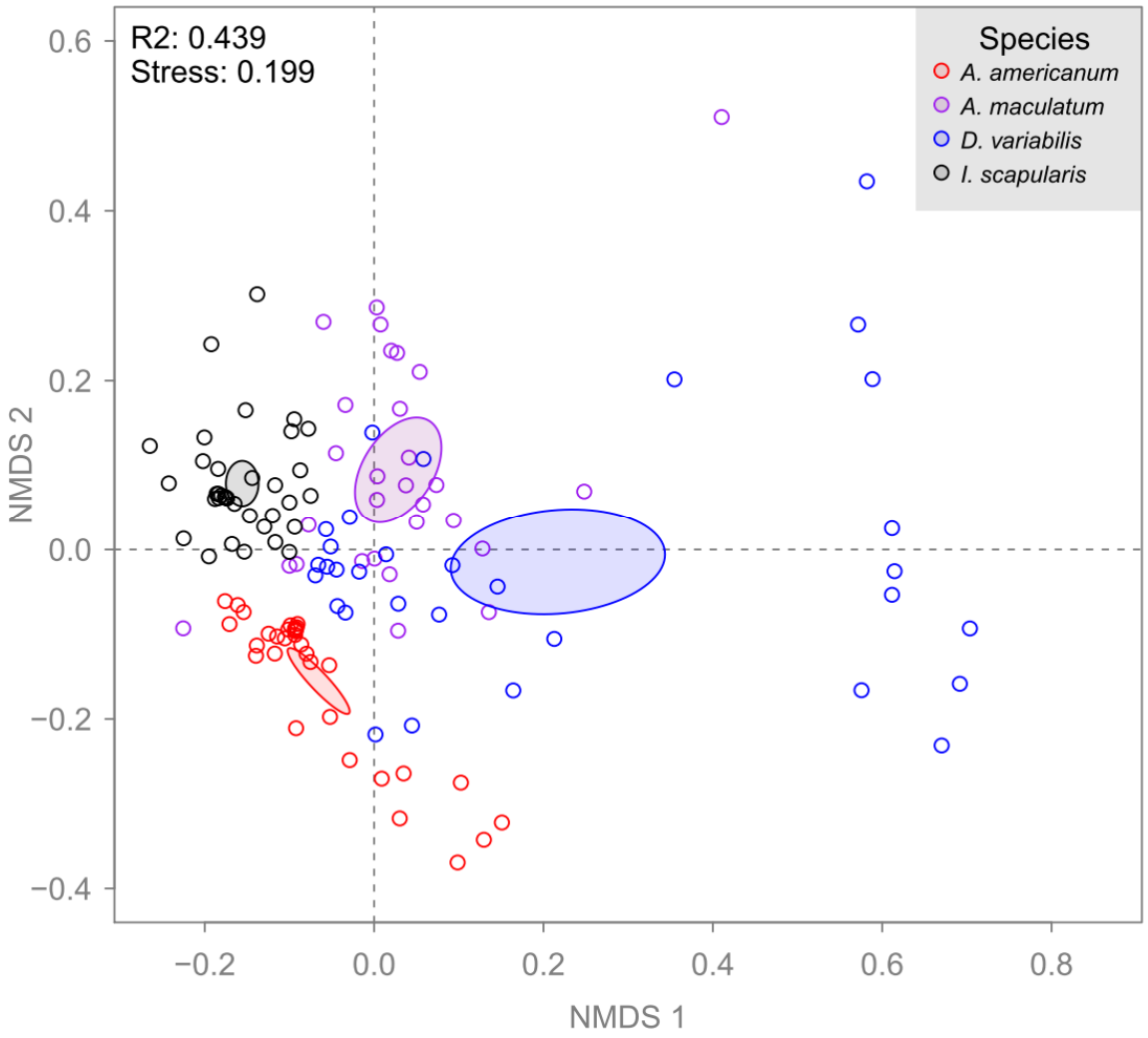
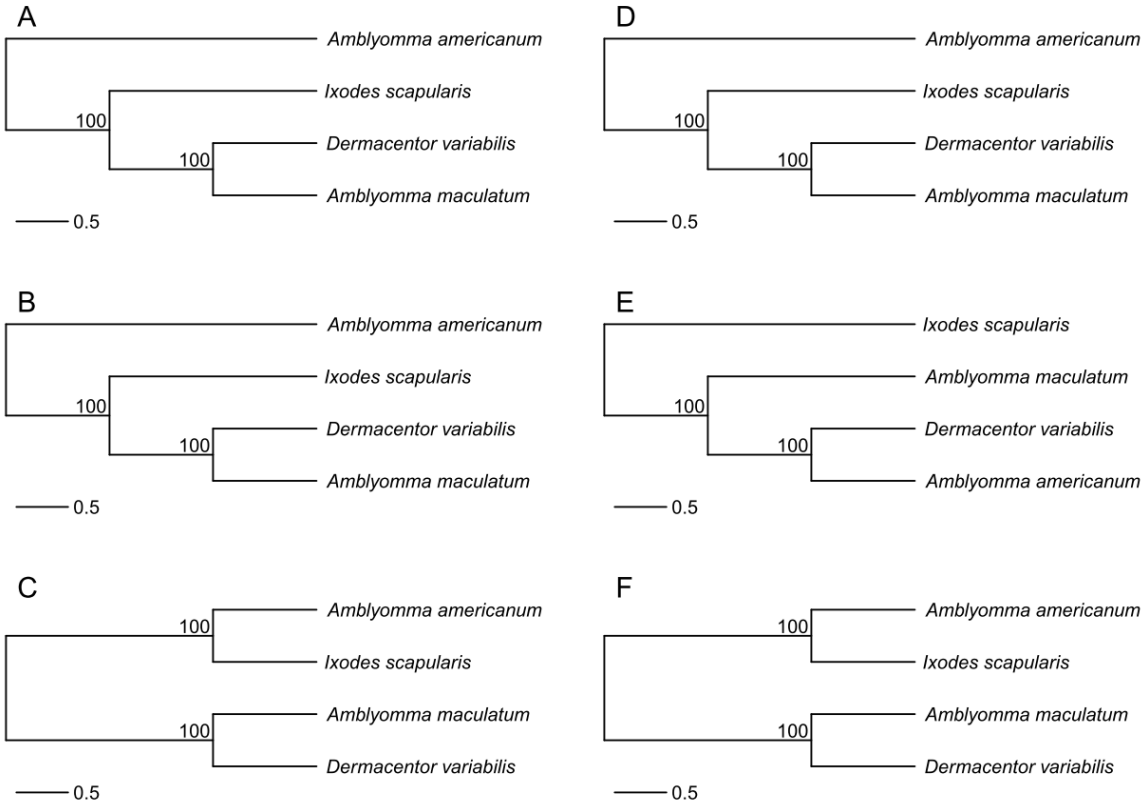


Figure 4.S8 Tick bacterial community relatedness dendrograms. All dendrograms are majority rule consensus trees summarizing bootstrapped unweighted pair group method with arithmetic mean (UPGMA) trees. Branch supports are percentages. Samples were pooled by species in panels A-C as in Brucker and Bordenstein (2012). In panels D-F community dissimilarities were obtained by averaging the pairwise sample dissimilarities for each tick species pair and applying a correction for within species variation. The community dissimilarities metrics used in each panel were as follows A and D: Jaccard, B and E: Bray-Curtis, C and F: unweighted UniFrac.



4.7 Tables

Table 4.1 Tick samples analyzed by location and date of origin, sex, and sequencing plate region.

Species	Collection Location	Collection Date	Females	Males	Plate*
Amblyomma americanum	Panola Mountain SP, GA	May 2011	11	6	A1
	Sweetwater Creek SP, GA	May 2011	12	4	A1
Amblyomma maculatum	Tate's Hell SF, FL	July – Aug 2005	12	4	A2, B1
	Tate's Hell SF, FL	July – Aug 2007	7	6	B1
Dermacentor variabilis	Panola Mountain SP, GA	June – Aug 2011	10	3	A3, B1
	Sweetwater Creek SP, GA	April – Aug 2011	6	3	A3, B1
	Muskingum County, OH	July 2011	5	4	A3, B1
	Tate's Hell SF, FL	July – Aug 2005	1	0	A2
Ixodes scapularis	Lake Dawson, CT	March 2013	23	13	A4

Abbreviations: SP=State Park, SF=State Forest, GA=Georgia, FL=Florida, OH=Ohio, CT=Connecticut

*See Table 4.S1 for sequencing plate region statistics.

Table 4.2 Classification of bacterial community dissimilarity metrics.

Metric type	OTU Relatedness	
	OTU-based	Phylogeny-based
Community membership	Jaccard	unweighted UniFrac
Community structure	Bray-Curtis	weighted UniFrac

Table 4.3 Multivariate analysis of the effect of species and sex on the relatedness of tick bacterial communities.

Dissimilarity Metric	Dispersion Test		PERMANOVA			
	F	*p*	Model Terms	*F*	R^2	*p*
Jaccard	8.83	0.001	Tick sex	2.36	0.010	0.008
			Tick species	35.4	0.448	0.001
			Interaction	2.19	0.028	0.001
Bray-Curtis	3.32	0.003	Tick sex	13.7	0.019	0.001
			Tick species	175.4	0.751	0.001
			Interaction	13.0	0.056	0.001
unweighted UniFrac	7.94	0.001	Tick sex	1.98	0.013	0.006
			Tick species	7.97	0.156	0.001
			Interaction	1.73	0.034	0.003
weighted UniFrac	13.5	0.001	Tick sex	10.0	0.032	0.001
			Tick species	54.9	0.517	0.001
			Interaction	7.19	0.068	0.001

Table 4.S1 454 sequencing statistics from Titanium FLX plates.

Plate ID	Plate Quarter ID	Tick Species	Total No. DNAs	No. Raw Reads	No. High Quality Reads (%Raw)	No. Tick DNAs Retained*	No. Reads in Analysis[§]
A	1	*Amblyomma americanum*	50	276,596	165,830 (60.0%)	33	160,826
A	2	*Amblyomma maculatum, Dermacentor variabilis*	50 (49, 1)	309,442	223,114 (72.1%)	16 (15, 1)	57,538 (50,641; 6,897)
A	3	*Dermacentor variabilis*	50	210,108	86,399 (41.1%)	23	77,332
A	4	*Ixodes scapularis*	50	294,565	163,422 (55.5%)	36	159,311
B	1	*Amblyomma maculatum, Dermacentor variabilis*	50 (25, 25)	229,414	77,174 (33.6%)	22 (14, 8)	59,770 (48,839; 10,931)

*Discarded DNAs were not sequenced to a depth of at least 1000 sequences or belonged to variable groups not included in this analysis.

[§]Total number of high quality sequences from retained ticks. Includes singleton OTUs.

Table 4.S2 Oligotyping software dependency information.

Environment	Package	Version
Python	Base	2.7.3
	Matplotlib	1.1.1rc
	BioPython	1.58
	SciPy	0.9.0
	PyCogent	1.5.3
	Django	1.6
R	Base	3.1.1
	vegan	2.0-10
	ggplot2	1.0.0
	gplots	2.14.1
	gtools	3.4.1
	reshape	0.8.5
	optparse	1.2.0
	pheatmap	0.7.7
	RColorBrewer	1.0-5

Table 4.S3 Oligotyping program settings and results for abundant operational taxonomic units (OTUs).

97% OTU	Classification	Settings				No. sequences	No. ticks	No. oligotypes	Derived OTUs
		-c	-s	-a	-M				
OTU0001	Rickettsia	6	1	1	0	178447	86	7	OTU1000-1006
OTU0002	Francisella	2	1	1	0	155910	63	4	OTU2000-2003
OTU0003	Coxiella	NA	NA	NA	NA	143782	36	1	NA
OTU0004	Borrelia	1	1	1	0	11631	19	3	OTU4000-4002
OTU0005	Midichloria	3	1	1	10	11299	21	4	OTU5000-5003
OTU0006	Borrelia	3	1	1	10	6914	11	4	OTU6000-6003
OTU0007	Proteobacteria	NA	NA	NA	NA	3437	4	1	NA
OTU0008	Anaplasma	NA	NA	NA	NA	1269	3	1	NA
OTU0009	Ehrlichia	NA	NA	NA	NA	481	4	1	NA

NA = Not applicable. Entropy analysis did not support the existance of >1 oligotype.

Table 4.S4 BLAST hits of representative sequences from bacterial operational taxonomic units.

Query	Subject Accession No.	Subject Title	Alignment Length	Expect Value	Bit Score	Raw Score	Percent Identical Matches	No. Identical Matches	No. Mis-matches	Total No. Gaps
Otu0003	AY939824	Coxiella endosymbiont of Amblyomma americanum 16S ribosomal RNA gene, partial sequence	256	4.00E-130	473	256	100.0	256	0	0
Otu0003	KJ130495	Uncultured Coxiella sp. clone GVPKRB303GJXPY 16S ribosomal RNA gene, partial sequence	245	1.00E-120	442	239	99.2	243	2	0
Otu0003	HM133590	Coxiella sp. US 16S ribosomal RNA gene, partial sequence	255	6.00E-119	436	236	97.7	249	5	1
Otu0003	JX846590	Uncultured Coxiella sp. clone AHOCCX 16S ribosomal RNA gene, partial sequence	254	1.00E-110	409	221	95.7	243	11	0
Otu0003	JX846589	Uncultured Coxiella sp. clone AHNCCX 16S ribosomal RNA gene, partial sequence	254	1.00E-110	409	221	95.7	243	11	0
Otu0003	EU143670	Uncultured Coxiella sp. clone 1357 16S ribosomal RNA gene, partial sequence	254	1.00E-110	409	221	95.7	243	11	0
Otu0003	EU143669	Uncultured Coxiella sp. clone 1358 16S ribosomal RNA gene, partial sequence	254	1.00E-110	409	221	95.7	243	11	0
Otu0003	KM079619	Coxiellaceae bacterium RFE03 16S ribosomal RNA gene,	254	3.00E-107	398	215	94.9	241	13	0

Query	Subject Accession No.	Subject Title	Alignment Length	Expect Value	Bit Score	Raw Score	Percent Identical Matches	No. Identical Matches	No. Mis-matches	Total No. Gaps
		partial sequence								
Otu0003	KM079617	Coxiellaceae bacterium CI01 16S ribosomal RNA gene, partial sequence	254	3.00E-107	398	215	94.9	241	13	0
Otu0003	KC776319	Uncultured Coxiella sp. clone XCP-1 16S ribosomal RNA gene, partial sequence	254	3.00E-107	398	215	94.9	241	13	0
Otu0007	EU817108	Uncultured bacterium clone Hg5a2E8 16S ribosomal RNA gene, partial sequence	264	8.00E-133	483	261	99.6	263	1	0
Otu0007	AB291637	Uncultured bacterium gene for 16S ribosomal RNA, partial sequence, clone: i10	266	6.00E-134	486	263	99.6	265	1	0
Otu0007	JQ855569	Uncultured bacterium clone l4-A9-E9-pl5.1 16S ribosomal RNA gene, partial sequence	266	6.00E-134	486	263	99.6	265	1	0
Otu0007	JQ855549	Uncultured bacterium clone l3-C1-G1-pl4 16S ribosomal RNA gene, partial sequence	266	6.00E-134	486	263	99.6	265	1	0
Otu0007	JX457300	Uncultured bacterium clone CR_229 16S ribosomal RNA gene, partial sequence	266	6.00E-134	486	263	99.6	265	1	0
Otu0007	JX457245	Uncultured bacterium clone CR_158 16S ribosomal RNA gene, partial sequence	266	6.00E-134	486	263	99.6	265	1	0
Otu0007	GU815103	Uncultured bacterium clone F3	266	6.00E-	486	263	99.6	265	1	0

Query	Subject Accession No.	Subject Title	Alignment Length	Expect Value	Bit Score	Raw Score	Percent Identical Matches	No. Identical Matches	No. Mis-matches	Total No. Gaps
		16S ribosomal RNA gene, partial sequence		134						
Otu0007	AB794406	Uncultured bacterium gene for 16S ribosomal RNA, partial sequence, clone: WsL02215	266	3.00E-132	481	260	99.3	264	2	0
Otu0007	JX406180	Candidatus Rickettsiella isopodorum strain JKI D244/2012 16S ribosomal RNA (rrs) gene, partial sequence	266	3.00E-132	481	260	99.3	264	2	0
Otu0007	HQ660943	Candidatus Rickettsiella isopodorum strain JKI A174/08 16S ribosomal RNA (rrs) gene, partial sequence	266	3.00E-132	481	260	99.3	264	2	0
Otu0008	KF805344	Anaplasma phagocytophilum clone gw1 16S ribosomal RNA gene, partial sequence	255	2.00E-129	472	255	100.0	255	0	0
Otu0008	KM259921	Uncultured Anaplasma sp. clone Belgium 16S ribosomal RNA gene, partial sequence	255	2.00E-129	472	255	100.0	255	0	0
Otu0008	HG916766	Anaplasma phagocytophilum partial 16S rRNA gene, isolate Ap-ha-CSF-23	255	2.00E-129	472	255	100.0	255	0	0
Otu0008	KF569910	Anaplasma phagocytophilum strain Hubei M16 16S ribosomal RNA gene, partial sequence	255	2.00E-129	472	255	100.0	255	0	0
Otu0008	KF569909	Anaplasma phagocytophilum	255	2.00E-	472	255	100.0	255	0	0

Query	Subject Accession No.	Subject Title	Alignment Length	Expect Value	Bit Score	Raw Score	Percent Identical Matches	No. Identical Matches	No. Mis-matches	Total No. Gaps
		strain Hubei E4 16S ribosomal RNA gene, partial sequence		129						
Otu0008	KF569908	Anaplasma phagocytophilum strain Hubei E1 16S ribosomal RNA gene, partial sequence	255	2.00E-129	472	255	100.0	255	0	0
Otu0008	CP006617	Anaplasma phagocytophilum str. JM, complete genome	255	2.00E-129	472	255	100.0	255	0	0
Otu0008	CP006616	Anaplasma phagocytophilum str. HZ2, complete genome	255	2.00E-129	472	255	100.0	255	0	0
Otu0008	KC800985	Anaplasma phagocytophilum isolate 9B13 16S ribosomal RNA gene, partial sequence	255	2.00E-129	472	255	100.0	255	0	0
Otu0008	CP006618	Anaplasma phagocytophilum str. Dog2 genome	255	2.00E-129	472	255	100.0	255	0	0
Otu0009	CP007480	Ehrlichia chaffeensis str. West Paces, complete genome	253	2.00E-128	468	253	100.0	253	0	0
Otu0009	CP007479	Ehrlichia chaffeensis str. Wakulla, complete genome	253	2.00E-128	468	253	100.0	253	0	0
Otu0009	CP007478	Ehrlichia chaffeensis str. Saint Vincent, complete genome	253	2.00E-128	468	253	100.0	253	0	0
Otu0009	CP007477	Ehrlichia chaffeensis str. Osceola, complete genome	253	2.00E-128	468	253	100.0	253	0	0
Otu0009	CP007476	Ehrlichia chaffeensis str. Liberty, complete genome	253	2.00E-128	468	253	100.0	253	0	0
Otu0009	CP007475	Ehrlichia chaffeensis str. Jax, complete genome	253	2.00E-128	468	253	100.0	253	0	0

Query	Subject Accession No.	Subject Title	Alignment Length	Expect Value	Bit Score	Raw Score	Percent Identical Matches	No. Identical Matches	No. Mis-matches	Total No. Gaps
Otu0009	CP007474	Ehrlichia sp. HF, complete genome	253	2.00E-128	468	253	100.0	253	0	0
Otu0009	CP007473	Ehrlichia chaffeensis str. Heartland, complete genome	253	2.00E-128	468	253	100.0	253	0	0
Otu0009	NR_074500	Ehrlichia chaffeensis strain Arkansas 16S ribosomal RNA gene, complete sequence	253	2.00E-128	468	253	100.0	253	0	0
Otu0009	JF952894	Ehrlichia chaffeensis 16S ribosomal RNA gene, partial sequence	253	2.00E-128	468	253	100.0	253	0	0
Otu1000	JQ805029	Uncultured bacterium clone Stegobium_mycetome_symbiont 16S ribosomal RNA gene, partial sequence	255	2.00E-124	455	246	98.8	252	3	0
Otu1000	AB702995	Rickettsia symbiont of Nephotettix cincticeps gene for 16S rRNA, partial sequence	255	2.00E-124	455	246	98.8	252	3	0
Otu1000	EU223247	Rickettsia endosymbiont of Lutzomyia apache isolate LA-1 16S ribosomal RNA gene, partial sequence	255	2.00E-124	455	246	98.8	252	3	0
Otu1000	FM955311	Rickettsia endosymbiont of Deronectes semirufus partial 16S rRNA gene	255	2.00E-124	455	246	98.8	252	3	0
Otu1000	HE583203	Rickettsia endosymbiont of Macrolophus sp. partial 16S	255	7.00E-123	449	243	98.4	251	4	0

Query	Subject Accession No.	Subject Title	Alignment Length	Expect Value	Bit Score	Raw Score	Percent Identical Matches	No. Identical Matches	No. Mis-matches	Total No. Gaps
		rRNA gene, clone MP1492-2								
Otu1000	FM177877	Rickettsia endosymbiont of Deronectes platynotus partial 16S rRNA gene, clone 10	255	7.00E-123	449	243	98.4	251	4	0
Otu1000	FM177876	Rickettsia endosymbiont of Deronectes platynotus partial 16S rRNA gene, clone 9	255	7.00E-123	449	243	98.4	251	4	0
Otu1000	FM177875	Rickettsia endosymbiont of Deronectes platynotus partial 16S rRNA gene, clone 8	255	7.00E-123	449	243	98.4	251	4	0
Otu1000	FM955312	Rickettsia endosymbiont of Deronectes delarouzei partial 16S rRNA gene	255	7.00E-123	449	243	98.4	251	4	0
Otu1000	FM955310	Rickettsia endosymbiont of Deronectes aubei partial 16S rRNA gene	255	7.00E-123	449	243	98.4	251	4	0
Otu1001	AM084233	Uncultured Rickettsia sp. partial 16S rRNA gene, TTGE band 6H3	252	7.00E-128	466	252	100.0	252	0	0
Otu1001	NR_115686	Rickettsia monacensis strain IrR/Munich 16S ribosomal RNA gene, partial sequence	255	2.00E-129	472	255	100.0	255	0	0
Otu1001	D84558	Rickettsia sp. (Ixodes scapularis symbiont) gene for 16S rRNA, partial sequence	255	2.00E-129	472	255	100.0	255	0	0
Otu1001	AF141908	Rickettsia sp. IRS4 16S	255	7.00E-	466	252	99.6	254	1	0

Query	Subject Accession No.	Subject Title	Alignment Length	Expect Value	Bit Score	Raw Score	Percent Identical Matches	No. Identical Matches	No. Mis-matches	Total No. Gaps
		ribosomal RNA gene, partial sequence		128						
Otu1001	AF141907	Rickettsia sp. IRS3 16S ribosomal RNA gene, partial sequence	255	7.00E-128	466	252	99.6	254	1	0
Otu1001	KJ410261	Rickettsia raoultii isolate BL029-2 16S ribosomal RNA gene, partial sequence	255	3.00E-126	460	249	99.2	253	2	0
Otu1001	KJ410260	Rickettsia raoultii isolate BL029-1 16S ribosomal RNA gene, partial sequence	255	3.00E-126	460	249	99.2	253	2	0
Otu1001	KJ410259	Rickettsia raoultii isolate TC250-11 16S ribosomal RNA gene, partial sequence	255	3.00E-126	460	249	99.2	253	2	0
Otu1001	KJ410258	Rickettsia raoultii isolate TC249-10 16S ribosomal RNA gene, partial sequence	255	3.00E-126	460	249	99.2	253	2	0
Otu1001	KF964552	Uncultured Rickettsia sp. isolate LF1-5 16S ribosomal RNA gene, partial sequence	255	3.00E-126	460	249	99.2	253	2	0
Otu1002	AB746410	Rickettsia endosymbiont of Curculio sp. 1 gene for 16S rRNA, partial sequence, isolate: P41_5	255	2.00E-129	472	255	100.0	255	0	0
Otu1002	AB746408	Rickettsia endosymbiont of Curculio kojimai gene for 16S	255	2.00E-129	472	255	100.0	255	0	0

Query	Subject Accession No.	Subject Title	Alignment Length	Expect Value	Bit Score	Raw Score	Percent Identical Matches	No. Identical Matches	No. Mis-matches	Total No. Gaps
		rRNA, partial sequence, isolate: P42_2								
Otu1002	KF598755	Uncultured bacterium clone TH-gut-46 16S ribosomal RNA gene, partial sequence	255	2.00E-129	472	255	100.0	255	0	0
Otu1002	KF646705	Rickettsia endosymbiont of Nesidiocoris tenuis isolate rbf1531r 16S ribosomal RNA gene, partial sequence	255	2.00E-129	472	255	100.0	255	0	0
Otu1002	AB772261	Rickettsia endosymbiont of Neophilaenus lineatus gene for 16S ribosomal RNA, partial sequence	255	2.00E-129	472	255	100.0	255	0	0
Otu1002	NR_074484	Rickettsia bellii strain RML369-C 16S ribosomal RNA gene, complete sequence	255	2.00E-129	472	255	100.0	255	0	0
Otu1002	JX943561	Rickettsia endosymbiont of Acyrthosiphon pisum strain 208 16S ribosomal RNA gene, partial sequence	255	2.00E-129	472	255	100.0	255	0	0
Otu1002	JX943562	Rickettsia endosymbiont of Acyrthosiphon pisum strain 141 16S ribosomal RNA gene, partial sequence	255	2.00E-129	472	255	100.0	255	0	0
Otu1002	JQ726774	Rickettsia endosymbiont of Nysius expressus clone	255	2.00E-129	472	255	100.0	255	0	0

Query	Subject Accession No.	Subject Title	Alignment Length	Expect Value	Bit Score	Raw Score	Percent Identical Matches	No. Identical Matches	No. Mis-matches	Total No. Gaps
		NeRckSpr 16S ribosomal RNA gene, partial sequence								
Otu1002	JQ726773	Rickettsia endosymbiont of Nysius sp. 2 clone Nsp2RckKmj 16S ribosomal RNA gene, partial sequence	255	2.00E-129	472	255	100.0	255	0	0
Otu1003	NR_074471	Candidatus Rickettsia amblyommii str. GAT-30V strain GAT-30V 16S ribosomal RNA, complete sequence	255	2.00E-129	472	255	100.0	255	0	0
Otu1003	CP003334	Candidatus Rickettsia amblyommii str. GAT-30V, complete genome	255	2.00E-129	472	255	100.0	255	0	0
Otu1003	U11012	Rickettsia amblyommii 16S rRNA gene, partial sequence	255	7.00E-128	466	252	99.6	254	1	0
Otu1003	KJ130496	Candidatus Rickettsia amblyommii clone GVPKRB301BS7D9 16S ribosomal RNA gene, partial sequence	256	7.00E-128	466	252	99.6	255	0	1
Otu1003	KJ410261	Rickettsia raoultii isolate BL029-2 16S ribosomal RNA gene, partial sequence	255	3.00E-126	460	249	99.2	253	2	0
Otu1003	KJ410260	Rickettsia raoultii isolate BL029-1 16S ribosomal RNA gene, partial sequence	255	3.00E-126	460	249	99.2	253	2	0

Query	Subject Accession No.	Subject Title	Alignment Length	Expect Value	Bit Score	Raw Score	Percent Identical Matches	No. Identical Matches	No. Mis-matches	Total No. Gaps
Otu1003	KJ410259	Rickettsia raoultii isolate TC250-11 16S ribosomal RNA gene, partial sequence	255	3.00E-126	460	249	99.2	253	2	0
Otu1003	KJ410258	Rickettsia raoultii isolate TC249-10 16S ribosomal RNA gene, partial sequence	255	3.00E-126	460	249	99.2	253	2	0
Otu1003	KF964552	Uncultured Rickettsia sp. isolate LF1-5 16S ribosomal RNA gene, partial sequence	255	3.00E-126	460	249	99.2	253	2	0
Otu1003	JX846591	Uncultured Rickettsia sp. clone AHNCRK 16S ribosomal RNA gene, partial sequence	255	3.00E-126	460	249	99.2	253	2	0
Otu1004	NR_074497	Rickettsia parkeri str. Portsmouth strain Portsmouth 16S ribosomal RNA, complete sequence	255	2.00E-129	472	255	100.0	255	0	0
Otu1004	CP003341	Rickettsia parkeri str. Portsmouth, complete genome	255	2.00E-129	472	255	100.0	255	0	0
Otu1004	NR_029156	Rickettsia parkeri strain Maculatum 20 16S ribosomal RNA gene, complete sequence	255	2.00E-129	472	255	100.0	255	0	0
Otu1004	L36673	Rickettsia parkeri 16S ribosomal RNA (16S rRNA) gene	255	2.00E-129	472	255	100.0	255	0	0
Otu1004	KJ410261	Rickettsia raoultii isolate BL029-2 16S ribosomal RNA gene, partial sequence	255	7.00E-128	466	252	99.6	254	1	0

Query	Subject Accession No.	Subject Title	Alignment Length	Expect Value	Bit Score	Raw Score	Percent Identical Matches	No. Identical Matches	No. Mis-matches	Total No. Gaps
Otu1004	KJ410260	Rickettsia raoultii isolate BL029-1 16S ribosomal RNA gene, partial sequence	255	7.00E-128	466	252	99.6	254	1	0
Otu1004	KJ410259	Rickettsia raoultii isolate TC250-11 16S ribosomal RNA gene, partial sequence	255	7.00E-128	466	252	99.6	254	1	0
Otu1004	KJ410258	Rickettsia raoultii isolate TC249-10 16S ribosomal RNA gene, partial sequence	255	7.00E-128	466	252	99.6	254	1	0
Otu1004	KF964552	Uncultured Rickettsia sp. isolate LF1-5 16S ribosomal RNA gene, partial sequence	255	7.00E-128	466	252	99.6	254	1	0
Otu1004	JX846591	Uncultured Rickettsia sp. clone AHNCRK 16S ribosomal RNA gene, partial sequence	255	7.00E-128	466	252	99.6	254	1	0
Otu1005	JQ805029	Uncultured bacterium clone Stegobium_mycetome_symbiont 16S ribosomal RNA gene, partial sequence	255	2.00E-129	472	255	100.0	255	0	0
Otu1005	AB702995	Rickettsia symbiont of Nephotettix cincticeps gene for 16S rRNA, partial sequence	255	2.00E-129	472	255	100.0	255	0	0
Otu1005	EU223247	Rickettsia endosymbiont of Lutzomyia apache isolate LA-1 16S ribosomal RNA gene, partial sequence	255	2.00E-129	472	255	100.0	255	0	0

Query	Subject Accession No.	Subject Title	Alignment Length	Expect Value	Bit Score	Raw Score	Percent Identical Matches	No. Identical Matches	No. Mis-matches	Total No. Gaps
Otu1005	FM955311	Rickettsia endosymbiont of Deronectes semirufus partial 16S rRNA gene	255	2.00E-129	472	255	100.0	255	0	0
Otu1005	HE583203	Rickettsia endosymbiont of Macrolophus sp. partial 16S rRNA gene, clone MP1492-2	255	7.00E-128	466	252	99.6	254	1	0
Otu1005	FM177877	Rickettsia endosymbiont of Deronectes platynotus partial 16S rRNA gene, clone 10	255	7.00E-128	466	252	99.6	254	1	0
Otu1005	FM177876	Rickettsia endosymbiont of Deronectes platynotus partial 16S rRNA gene, clone 9	255	7.00E-128	466	252	99.6	254	1	0
Otu1005	FM177875	Rickettsia endosymbiont of Deronectes platynotus partial 16S rRNA gene, clone 8	255	7.00E-128	466	252	99.6	254	1	0
Otu1005	FM955312	Rickettsia endosymbiont of Deronectes delarouzei partial 16S rRNA gene	255	7.00E-128	466	252	99.6	254	1	0
Otu1005	FM955310	Rickettsia endosymbiont of Deronectes aubei partial 16S rRNA gene	255	7.00E-128	466	252	99.6	254	1	0
Otu1006	KJ410262	Rickettsia slovaca isolate TC250-17 16S ribosomal RNA gene, partial sequence	255	3.00E-126	460	249	99.2	253	2	0
Otu1006	NR_103944	Rickettsia prowazekii str. NMRC Madrid E strain NMRC	255	3.00E-126	460	249	99.2	253	2	0

Query	Subject Accession No.	Subject Title	Alignment Length	Expect Value	Bit Score	Raw Score	Percent Identical Matches	No. Identical Matches	No. Mis-matches	Total No. Gaps
		Madrid E 16S ribosomal RNA, complete sequence								
Otu1006	CP004889	Rickettsia prowazekii str. Breinl, complete genome	255	3.00E-126	460	249	99.2	253	2	0
Otu1006	CP004888	Rickettsia prowazekii str. NMRC Madrid E, complete genome	255	3.00E-126	460	249	99.2	253	2	0
Otu1006	NR_074474	Rickettsia slovaca str. D-CWPP strain D-CWPP 16S ribosomal RNA, complete sequence	255	3.00E-126	460	249	99.2	253	2	0
Otu1006	NR_074462	Rickettsia slovaca 13-B strain 13-B 16S ribosomal RNA, complete sequence	255	3.00E-126	460	249	99.2	253	2	0
Otu1006	NR_074394	Rickettsia typhi strain Wilmington 16S ribosomal RNA gene, complete sequence	255	3.00E-126	460	249	99.2	253	2	0
Otu1006	JQ412125	Uncultured Rickettsia sp. clone APPV1 16S ribosomal RNA gene, partial sequence	255	3.00E-126	460	249	99.2	253	2	0
Otu1006	JQ412124	Uncultured Rickettsia sp. clone APOH8 16S ribosomal RNA gene, partial sequence	255	3.00E-126	460	249	99.2	253	2	0
Otu1006	JQ339355	Uncultured Rickettsia sp. clone APVS1 16S ribosomal RNA gene, partial sequence	255	3.00E-126	460	249	99.2	253	2	0
Otu2000	AY795978	Francisella sp. DVFSQ80_04	260	1.00E-	475	257	99.6	259	1	0

Query	Subject Accession No.	Subject Title	Alignment Length	Expect Value	Bit Score	Raw Score	Percent Identical Matches	No. Identical Matches	No. Mis-matches	Total No. Gaps
		16S ribosomal RNA gene, partial sequence		130						
Otu2000	AY795977	Francisella sp. DVFSQ83_04 16S ribosomal RNA gene, partial sequence	260	1.00E-130	475	257	99.6	259	1	0
Otu2000	AY795976	Francisella sp. DVFSQ81_04 16S ribosomal RNA gene, partial sequence	260	1.00E-130	475	257	99.6	259	1	0
Otu2000	JQ711250	Francisella sp. Belarus-450 16S ribosomal RNA gene, partial sequence	260	1.00E-125	459	248	98.5	256	4	0
Otu2000	KJ130499	Uncultured Francisella sp. clone GPZ3F3401BL6CT 16S ribosomal RNA gene, partial sequence	260	1.00E-125	459	248	98.5	256	4	0
Otu2000	CP007148	Francisella tularensis subsp. holarctica PHIT-FT049, complete genome	260	1.00E-125	459	248	98.5	256	4	0
Otu2000	CP007148	Francisella tularensis subsp. holarctica PHIT-FT049, complete genome	260	1.00E-125	459	248	98.5	256	4	0
Otu2000	CP007148	Francisella tularensis subsp. holarctica PHIT-FT049, complete genome	260	1.00E-125	459	248	98.5	256	4	0
Otu2000	JX101895	Francisella novicida strain LA114545 16S ribosomal RNA	260	1.00E-125	459	248	98.5	256	4	0

Query	Subject Accession No.	Subject Title	Alignment Length	Expect Value	Bit Score	Raw Score	Percent Identical Matches	No. Identical Matches	No. Mis-matches	Total No. Gaps
		gene, partial sequence								
Otu2000	JX101897	Francisella novicida strain LA114546 16S ribosomal RNA gene, partial sequence	260	1.00E-125	459	248	98.5	256	4	0
Otu2000	KF607100	Francisella tularensis subsp. holarctica isolate ad 16S ribosomal RNA gene, partial sequence	260	1.00E-125	459	248	98.5	256	4	0
Otu2000	JQ764629	Francisella endosymbiont of Dermacentor auratus clone SSPG3 16S ribosomal RNA gene, partial sequence	260	1.00E-125	459	248	98.5	256	4	0
Otu2001	JQ711250	Francisella sp. Belarus-450 16S ribosomal RNA gene, partial sequence	260	1.00E-130	475	257	99.6	259	1	0
Otu2001	KJ130499	Uncultured Francisella sp. clone GPZ3F3401BL6CT 16S ribosomal RNA gene, partial sequence	260	1.00E-130	475	257	99.6	259	1	0
Otu2001	CP007148	Francisella tularensis subsp. holarctica PHIT-FT049, complete genome	260	1.00E-130	475	257	99.6	259	1	0
Otu2001	CP007148	Francisella tularensis subsp. holarctica PHIT-FT049, complete genome	260	1.00E-130	475	257	99.6	259	1	0
Otu2001	CP007148	Francisella tularensis subsp.	260	1.00E-	475	257	99.6	259	1	0

Query	Subject Accession No.	Subject Title	Alignment Length	Expect Value	Bit Score	Raw Score	Percent Identical Matches	No. Identical Matches	No. Mis-matches	Total No. Gaps
		holarctica PHIT-FT049, complete genome		130						
Otu2001	JX101895	Francisella novicida strain LA114545 16S ribosomal RNA gene, partial sequence	260	1.00E-130	475	257	99.6	259	1	0
Otu2001	JX101897	Francisella novicida strain LA114546 16S ribosomal RNA gene, partial sequence	260	1.00E-130	475	257	99.6	259	1	0
Otu2001	KF607100	Francisella tularensis subsp. holarctica isolate ad 16S ribosomal RNA gene, partial sequence	260	1.00E-130	475	257	99.6	259	1	0
Otu2001	JQ764629	Francisella endosymbiont of Dermacentor auratus clone SSPG3 16S ribosomal RNA gene, partial sequence	260	1.00E-130	475	257	99.6	259	1	0
Otu2001	JQ764628	Francisella endosymbiont of Dermacentor auratus clone SSCM1 16S ribosomal RNA gene, partial sequence	260	1.00E-130	475	257	99.6	259	1	0
Otu2001	NR_07466 6	Francisella tularensis subsp. mediasiatica strain FSC147 16S ribosomal RNA gene, complete sequence	260	1.00E-130	475	257	99.6	259	1	0
Otu2001	NR_07466 5	Francisella tularensis subsp. novicida strain U112 16S	260	1.00E-130	475	257	99.6	259	1	0

Query	Subject Accession No.	Subject Title	Alignment Length	Expect Value	Bit Score	Raw Score	Percent Identical Matches	No. Identical Matches	No. Mis-matches	Total No. Gaps
		ribosomal RNA gene, complete sequence								
Otu2002	JQ711250	Francisella sp. Belarus-450 16S ribosomal RNA gene, partial sequence	260	3.00E-132	481	260	100.0	260	0	0
Otu2002	KJ130499	Uncultured Francisella sp. clone GPZ3F3401BL6CT 16S ribosomal RNA gene, partial sequence	260	3.00E-132	481	260	100.0	260	0	0
Otu2002	CP007148	Francisella tularensis subsp. holarctica PHIT-FT049, complete genome	260	3.00E-132	481	260	100.0	260	0	0
Otu2002	CP007148	Francisella tularensis subsp. holarctica PHIT-FT049, complete genome	260	3.00E-132	481	260	100.0	260	0	0
Otu2002	CP007148	Francisella tularensis subsp. holarctica PHIT-FT049, complete genome	260	3.00E-132	481	260	100.0	260	0	0
Otu2002	JX101895	Francisella novicida strain LA114545 16S ribosomal RNA gene, partial sequence	260	3.00E-132	481	260	100.0	260	0	0
Otu2002	JX101897	Francisella novicida strain LA114546 16S ribosomal RNA gene, partial sequence	260	3.00E-132	481	260	100.0	260	0	0
Otu2002	KF607100	Francisella tularensis subsp. holarctica isolate ad 16S	260	3.00E-132	481	260	100.0	260	0	0

Query	Subject Accession No.	Subject Title	Alignment Length	Expect Value	Bit Score	Raw Score	Percent Identical Matches	No. Identical Matches	No. Mis-matches	Total No. Gaps
		ribosomal RNA gene, partial sequence								
Otu2002	JQ764629	Francisella endosymbiont of Dermacentor auratus clone SSPG3 16S ribosomal RNA gene, partial sequence	260	3.00E-132	481	260	100.0	260	0	0
Otu2002	JQ764628	Francisella endosymbiont of Dermacentor auratus clone SSCM1 16S ribosomal RNA gene, partial sequence	260	3.00E-132	481	260	100.0	260	0	0
Otu2002	NR_074666	Francisella tularensis subsp. mediasiatica strain FSC147 16S ribosomal RNA gene, complete sequence	260	3.00E-132	481	260	100.0	260	0	0
Otu2002	NR_074665	Francisella tularensis subsp. novicida strain U112 16S ribosomal RNA gene, complete sequence	260	3.00E-132	481	260	100.0	260	0	0
Otu2003	AY795978	Francisella sp. DVFSQ80_04 16S ribosomal RNA gene, partial sequence	260	3.00E-132	481	260	100.0	260	0	0
Otu2003	AY795977	Francisella sp. DVFSQ83_04 16S ribosomal RNA gene, partial sequence	260	3.00E-132	481	260	100.0	260	0	0
Otu2003	AY795976	Francisella sp. DVFSQ81_04 16S ribosomal RNA gene,	260	3.00E-132	481	260	100.0	260	0	0

Query	Subject Accession No.	Subject Title	Alignment Length	Expect Value	Bit Score	Raw Score	Percent Identical Matches	No. Identical Matches	No. Mis-matches	Total No. Gaps
		partial sequence								
Otu2003	JQ711250	Francisella sp. Belarus-450 16S ribosomal RNA gene, partial sequence	260	3.00E-127	464	251	98.9	257	3	0
Otu2003	KJ130499	Uncultured Francisella sp. clone GPZ3F3401BL6CT 16S ribosomal RNA gene, partial sequence	260	3.00E-127	464	251	98.9	257	3	0
Otu2003	CP007148	Francisella tularensis subsp. holarctica PHIT-FT049, complete genome	260	3.00E-127	464	251	98.9	257	3	0
Otu2003	CP007148	Francisella tularensis subsp. holarctica PHIT-FT049, complete genome	260	3.00E-127	464	251	98.9	257	3	0
Otu2003	CP007148	Francisella tularensis subsp. holarctica PHIT-FT049, complete genome	260	3.00E-127	464	251	98.9	257	3	0
Otu2003	JX101895	Francisella novicida strain LA114545 16S ribosomal RNA gene, partial sequence	260	3.00E-127	464	251	98.9	257	3	0
Otu2003	JX101897	Francisella novicida strain LA114546 16S ribosomal RNA gene, partial sequence	260	3.00E-127	464	251	98.9	257	3	0
Otu2003	KF607100	Francisella tularensis subsp. holarctica isolate ad 16S ribosomal RNA gene, partial	260	3.00E-127	464	251	98.9	257	3	0

Query	Subject Accession No.	Subject Title	Alignment Length	Expect Value	Bit Score	Raw Score	Percent Identical Matches	No. Identical Matches	No. Mis-matches	Total No. Gaps
		sequence								
Otu2003	JQ764629	Francisella endosymbiont of Dermacentor auratus clone SSPG3 16S ribosomal RNA gene, partial sequence	260	3.00E-127	464	251	98.9	257	3	0
Otu4000	CP009656	Borrelia burgdorferi strain B31, complete genome	246	3.00E-121	444	240	99.2	244	2	0
Otu4000	KJ994333	Uncultured Borrelia sp. clone Villamonte 16S ribosomal RNA gene, partial sequence	246	3.00E-121	444	240	99.2	244	2	0
Otu4000	CP007564	Borrelia garinii SZ, complete genome	246	3.00E-121	444	240	99.2	244	2	0
Otu4000	CP005925	Borrelia burgdorferi CA382, complete genome	246	3.00E-121	444	240	99.2	244	2	0
Otu4000	NR_103929	Borrelia burgdorferi N40 strain N40 16S ribosomal RNA, complete sequence	246	3.00E-121	444	240	99.2	244	2	0
Otu4000	NR_102956	Borrelia bissettii strain DN127 16S ribosomal RNA gene, complete sequence	246	3.00E-121	444	240	99.2	244	2	0
Otu4000	NR_074854	Borrelia garinii PBi strain PBi 16S ribosomal RNA, complete sequence	246	3.00E-121	444	240	99.2	244	2	0
Otu4000	CP003866	Borrelia garinii NMJW1, complete genome	246	3.00E-121	444	240	99.2	244	2	0
Otu4000	JX082319	Uncultured Borrelia sp. clone J	246	3.00E-	444	240	99.2	244	2	0

Query	Subject Accession No.	Subject Title	Alignment Length	Expect Value	Bit Score	Raw Score	Percent Identical Matches	No. Identical Matches	No. Mis-matches	Total No. Gaps
		16S ribosomal RNA gene, partial sequence		121						
Otu4000	CP003151	Borrelia garinii BgVir chromosome linear, complete sequence	246	3.00E-121	444	240	99.2	244	2	0
Otu4001	CP009656	Borrelia burgdorferi strain B31, complete genome	251	3.00E-127	464	251	100.0	251	0	0
Otu4001	KJ994333	Uncultured Borrelia sp. clone Villamonte 16S ribosomal RNA gene, partial sequence	251	3.00E-127	464	251	100.0	251	0	0
Otu4001	CP007564	Borrelia garinii SZ, complete genome	251	3.00E-127	464	251	100.0	251	0	0
Otu4001	CP005925	Borrelia burgdorferi CA382, complete genome	251	3.00E-127	464	251	100.0	251	0	0
Otu4001	NR_103929	Borrelia burgdorferi N40 strain N40 16S ribosomal RNA, complete sequence	251	3.00E-127	464	251	100.0	251	0	0
Otu4001	NR_102956	Borrelia bissettii strain DN127 16S ribosomal RNA gene, complete sequence	251	3.00E-127	464	251	100.0	251	0	0
Otu4001	NR_074854	Borrelia garinii PBi strain PBi 16S ribosomal RNA, complete sequence	251	3.00E-127	464	251	100.0	251	0	0
Otu4001	CP003866	Borrelia garinii NMJW1, complete genome	251	3.00E-127	464	251	100.0	251	0	0
Otu4001	JX082319	Uncultured Borrelia sp. clone J	251	3.00E-	464	251	100.0	251	0	0

Query	Subject Accession No.	Subject Title	Alignment Length	Expect Value	Bit Score	Raw Score	Percent Identical Matches	No. Identical Matches	No. Mis-matches	Total No. Gaps
		16S ribosomal RNA gene, partial sequence		127						
Otu4001	CP003151	Borrelia garinii BgVir chromosome linear, complete sequence	251	3.00E-127	464	251	100.0	251	0	0
Otu4002	AB904793	Borrelia miyamotoi gene for 16S ribosomal RNA, partial sequence, strain: HT31	252	2.00E-119	438	237	98.0	247	5	0
Otu4002	AB900817	Borrelia miyamotoi gene for 16S ribosomal RNA, partial sequence, strain: Y13T1	252	2.00E-119	438	237	98.0	247	5	0
Otu4002	KM007554	Borrelia miyamotoi strain PoTiBmiy1 16S ribosomal RNA gene, partial sequence	252	2.00E-119	438	237	98.0	247	5	0
Otu4002	NR_121757	Borrelia miyamotoi 16S ribosomal RNA, complete sequence	252	2.00E-119	438	237	98.0	247	5	0
Otu4002	KJ412199	Borrelia miyamotoi clone rongeur215 16S ribosomal RNA gene, partial sequence	252	2.00E-119	438	237	98.0	247	5	0
Otu4002	KJ412198	Borrelia miyamotoi clone tik165 16S ribosomal RNA gene, partial sequence	252	2.00E-119	438	237	98.0	247	5	0
Otu4002	KJ412197	Borrelia miyamotoi clone tik84 16S ribosomal RNA gene, partial sequence	252	2.00E-119	438	237	98.0	247	5	0

Query	Subject Accession No.	Subject Title	Alignment Length	Expect Value	Bit Score	Raw Score	Percent Identical Matches	No. Identical Matches	No. Mis-matches	Total No. Gaps
Otu4002	KJ412196	Borrelia miyamotoi clone tik306 16S ribosomal RNA gene, partial sequence	252	2.00E-119	438	237	98.0	247	5	0
Otu4002	KJ412195	Borrelia miyamotoi clone tik322 16S ribosomal RNA gene, partial sequence	252	2.00E-119	438	237	98.0	247	5	0
Otu4002	KJ412194	Borrelia miyamotoi clone tik371 16S ribosomal RNA gene, partial sequence	252	2.00E-119	438	237	98.0	247	5	0
Otu5000	AM411593	Candidatus Midichloria sp. AtubES1 partial 16S rRNA gene, isolate AtubES1	256	2.00E-124	455	246	98.8	253	1	2
Otu5000	AM411594	Candidatus Midichloria sp. AtubES2 partial 16S rRNA gene, isolate AtubES2	256	2.00E-124	455	246	98.8	253	1	2
Otu5000	KJ130497	Candidatus Midichloria mitochondrii clone GPZ3F3401ANN8P 16S ribosomal RNA gene, partial sequence	256	7.00E-123	449	243	98.4	252	2	2
Otu5000	HF568840	Candidatus Midichloria mitochondrii partial 16S rRNA gene, strain OV2	256	7.00E-123	449	243	98.4	252	2	2
Otu5000	JQ678691	Candidatus Midichloria mitochondrii isolate AamerB 16S ribosomal RNA gene,	256	7.00E-123	449	243	98.4	252	2	2

Query	Subject Accession No.	Subject Title	Alignment Length	Expect Value	Bit Score	Raw Score	Percent Identical Matches	No. Identical Matches	No. Mis-matches	Total No. Gaps
		partial sequence								
Otu5000	AM411591	Candidatus Midichloria sp. HymarES2 partial 16S rRNA gene, isolate HymarES2	256	7.00E-123	449	243	98.4	252	2	2
Otu5000	AM411592	Candidatus Midichloria sp. HymarES3 partial 16S rRNA gene, isolate HymarES3	256	7.00E-123	449	243	98.4	252	2	2
Otu5000	AM411590	Candidatus Midichloria sp. HymarES1 partial 16S rRNA gene, isolate HymarES1	256	7.00E-123	449	243	98.4	252	2	2
Otu5000	AM181354	Candidatus Midichloria symbiont of Hyalomma marginatum partial 16S rRNA gene, strain 1	256	7.00E-123	449	243	98.4	252	2	2
Otu5000	DQ379965	Rickettsiales bacterium IE-B clone IE165 16S ribosomal RNA gene, partial sequence	256	7.00E-123	449	243	98.4	252	2	2
Otu5001	AM411593	Candidatus Midichloria sp. AtubES1 partial 16S rRNA gene, isolate AtubES1	254	6.00E-129	470	254	100.0	254	0	0
Otu5001	AM411594	Candidatus Midichloria sp. AtubES2 partial 16S rRNA gene, isolate AtubES2	254	6.00E-129	470	254	100.0	254	0	0
Otu5001	KJ130497	Candidatus Midichloria mitochondrii clone GPZ3F3401ANN8P 16S	254	3.00E-127	464	251	99.6	253	1	0

Query	Subject Accession No.	Subject Title	Alignment Length	Expect Value	Bit Score	Raw Score	Percent Identical Matches	No. Identical Matches	No. Mis-matches	Total No. Gaps
		ribosomal RNA gene, partial sequence								
Otu5001	HF568840	Candidatus Midichloria mitochondrii partial 16S rRNA gene, strain OV2	254	3.00E-127	464	251	99.6	253	1	0
Otu5001	JQ678691	Candidatus Midichloria mitochondrii isolate AamerB 16S ribosomal RNA gene, partial sequence	254	3.00E-127	464	251	99.6	253	1	0
Otu5001	AM411591	Candidatus Midichloria sp. HymarES2 partial 16S rRNA gene, isolate HymarES2	254	3.00E-127	464	251	99.6	253	1	0
Otu5001	AM411592	Candidatus Midichloria sp. HymarES3 partial 16S rRNA gene, isolate HymarES3	254	3.00E-127	464	251	99.6	253	1	0
Otu5001	AM411590	Candidatus Midichloria sp. HymarES1 partial 16S rRNA gene, isolate HymarES1	254	3.00E-127	464	251	99.6	253	1	0
Otu5001	AM181354	Candidatus Midichloria symbiont of Hyalomma marginatum partial 16S rRNA gene, strain 1	254	3.00E-127	464	251	99.6	253	1	0
Otu5001	DQ379965	Rickettsiales bacterium IE-B clone IE165 16S ribosomal RNA gene, partial sequence	254	3.00E-127	464	251	99.6	253	1	0
Otu5002	KJ130500	Candidatus Midichloria	254	6.00E-	453	245	98.8	251	3	0

Query	Subject Accession No.	Subject Title	Alignment Length	Expect Value	Bit Score	Raw Score	Percent Identical Matches	No. Identical Matches	No. Mis-matches	Total No. Gaps
		mitochondrii clone GPZ3F3401B2QI9 16S ribosomal RNA gene, partial sequence		124						
Otu5002	HF568843	Candidatus Midichloria mitochondrii partial 16S rRNA gene, strain NI1	254	6.00E-124	453	245	98.8	251	3	0
Otu5002	HF568842	Candidatus Midichloria mitochondrii partial 16S rRNA gene, strain MA8	254	6.00E-124	453	245	98.8	251	3	0
Otu5002	HF568841	Candidatus Midichloria mitochondrii partial 16S rRNA gene, strain MA7	254	6.00E-124	453	245	98.8	251	3	0
Otu5002	NR_074492	Candidatus Midichloria mitochondrii IricVA strain IricVA 16S ribosomal RNA, complete sequence	254	6.00E-124	453	245	98.8	251	3	0
Otu5002	JQ678692	Candidatus Midichloria mitochondrii isolate AamerC 16S ribosomal RNA gene, partial sequence	254	6.00E-124	453	245	98.8	251	3	0
Otu5002	CP002130	Candidatus Midichloria mitochondrii IricVA, complete genome	254	6.00E-124	453	245	98.8	251	3	0
Otu5002	FM992372	Candidatus Midichloria sp. Ixholo1 partial 16S rRNA gene,	254	6.00E-124	453	245	98.8	251	3	0

Query	Subject Accession No.	Subject Title	Alignment Length	Expect Value	Bit Score	Raw Score	Percent Identical Matches	No. Identical Matches	No. Mis-matches	Total No. Gaps
		isolate Ixholo1								
Otu5002	AM411600	Candidatus Midichloria sp. IuriES1 partial 16S rRNA gene, isolate IuriES1	254	6.00E-124	453	245	98.8	251	3	0
Otu5002	AM411599	Candidatus Midichloria sp. HapunES2 partial 16S rRNA gene, isolate HapunES2	254	6.00E-124	453	245	98.8	251	3	0
Otu5003	JQ678693	Candidatus Midichloria mitochondrii isolate AamerA 16S ribosomal RNA gene, partial sequence	254	6.00E-129	470	254	100.0	254	0	0
Otu5003	HF568843	Candidatus Midichloria mitochondrii partial 16S rRNA gene, strain NI1	254	3.00E-122	448	242	98.4	250	4	0
Otu5003	HF568840	Candidatus Midichloria mitochondrii partial 16S rRNA gene, strain OV2	254	1.00E-120	442	239	98.0	249	5	0
Otu5003	FM992373	Candidatus Midichloria sp. Ixholo2 partial 16S rRNA gene, isolate Ixholo2	254	1.00E-120	442	239	98.0	249	5	0
Otu5003	AM411591	Candidatus Midichloria sp. HymarES2 partial 16S rRNA gene, isolate HymarES2	254	1.00E-120	442	239	98.0	249	5	0
Otu5003	AM411592	Candidatus Midichloria sp. HymarES3 partial 16S rRNA gene, isolate HymarES3	254	1.00E-120	442	239	98.0	249	5	0

Query	Subject Accession No.	Subject Title	Alignment Length	Expect Value	Bit Score	Raw Score	Percent Identical Matches	No. Identical Matches	No. Mis-matches	Total No. Gaps
Otu5003	AM411590	Candidatus Midichloria sp. HymarES1 partial 16S rRNA gene, isolate HymarES1	254	1.00E-120	442	239	98.0	249	5	0
Otu5003	AM181354	Candidatus Midichloria symbiont of Hyalomma marginatum partial 16S rRNA gene, strain 1	254	1.00E-120	442	239	98.0	249	5	0
Otu5003	DQ379965	Rickettsiales bacterium IE-B clone IE165 16S ribosomal RNA gene, partial sequence	254	1.00E-120	442	239	98.0	249	5	0
Otu5003	DQ379964	Rickettsiales bacterium IE-B clone IE135 16S ribosomal RNA gene, partial sequence	254	1.00E-120	442	239	98.0	249	5	0
Otu6000	AB904793	Borrelia miyamotoi gene for 16S ribosomal RNA, partial sequence, strain: HT31	252	3.00E-126	460	249	99.6	251	1	0
Otu6000	AB900817	Borrelia miyamotoi gene for 16S ribosomal RNA, partial sequence, strain: Y13T1	252	3.00E-126	460	249	99.6	251	1	0
Otu6000	KM007554	Borrelia miyamotoi strain PoTiBmiy1 16S ribosomal RNA gene, partial sequence	252	3.00E-126	460	249	99.6	251	1	0
Otu6000	NR_121757	Borrelia miyamotoi 16S ribosomal RNA, complete sequence	252	3.00E-126	460	249	99.6	251	1	0
Otu6000	KJ412199	Borrelia miyamotoi clone	252	3.00E-	460	249	99.6	251	1	0

Query	Subject Accession No.	Subject Title	Alignment Length	Expect Value	Bit Score	Raw Score	Percent Identical Matches	No. Identical Matches	No. Mis-matches	Total No. Gaps
		rongeur215 16S ribosomal RNA gene, partial sequence		126						
Otu6000	KJ412198	Borrelia miyamotoi clone tik165 16S ribosomal RNA gene, partial sequence	252	3.00E-126	460	249	99.6	251	1	0
Otu6000	KJ412197	Borrelia miyamotoi clone tik84 16S ribosomal RNA gene, partial sequence	252	3.00E-126	460	249	99.6	251	1	0
Otu6000	KJ412196	Borrelia miyamotoi clone tik306 16S ribosomal RNA gene, partial sequence	252	3.00E-126	460	249	99.6	251	1	0
Otu6000	KJ412195	Borrelia miyamotoi clone tik322 16S ribosomal RNA gene, partial sequence	252	3.00E-126	460	249	99.6	251	1	0
Otu6000	KJ412194	Borrelia miyamotoi clone tik371 16S ribosomal RNA gene, partial sequence	252	3.00E-126	460	249	99.6	251	1	0
Otu6001	AB904793	Borrelia miyamotoi gene for 16S ribosomal RNA, partial sequence, strain: HT31	252	7.00E-128	466	252	100.0	252	0	0
Otu6001	AB900817	Borrelia miyamotoi gene for 16S ribosomal RNA, partial sequence, strain: Y13T1	252	7.00E-128	466	252	100.0	252	0	0
Otu6001	KM007554	Borrelia miyamotoi strain PoTiBmiy1 16S ribosomal RNA gene, partial sequence	252	7.00E-128	466	252	100.0	252	0	0

Query	Subject Accession No.	Subject Title	Alignment Length	Expect Value	Bit Score	Raw Score	Percent Identical Matches	No. Identical Matches	No. Mis-matches	Total No. Gaps
Otu6001	NR_121757	Borrelia miyamotoi 16S ribosomal RNA, complete sequence	252	7.00E-128	466	252	100.0	252	0	0
Otu6001	KJ412199	Borrelia miyamotoi clone rongeur215 16S ribosomal RNA gene, partial sequence	252	7.00E-128	466	252	100.0	252	0	0
Otu6001	KJ412198	Borrelia miyamotoi clone tik165 16S ribosomal RNA gene, partial sequence	252	7.00E-128	466	252	100.0	252	0	0
Otu6001	KJ412197	Borrelia miyamotoi clone tik84 16S ribosomal RNA gene, partial sequence	252	7.00E-128	466	252	100.0	252	0	0
Otu6001	KJ412196	Borrelia miyamotoi clone tik306 16S ribosomal RNA gene, partial sequence	252	7.00E-128	466	252	100.0	252	0	0
Otu6001	KJ412195	Borrelia miyamotoi clone tik322 16S ribosomal RNA gene, partial sequence	252	7.00E-128	466	252	100.0	252	0	0
Otu6001	KJ412194	Borrelia miyamotoi clone tik371 16S ribosomal RNA gene, partial sequence	252	7.00E-128	466	252	100.0	252	0	0
Otu6002	KJ130498	Uncultured Borrelia sp. clone GPZ3F3401AHN78 16S ribosomal RNA gene, partial sequence	252	7.00E-128	466	252	100.0	252	0	0
Otu6002	AY166715	Borrelia lonestari 16S ribosomal	252	7.00E-	466	252	100.0	252	0	0

Query	Subject Accession No.	Subject Title	Alignment Length	Expect Value	Bit Score	Raw Score	Percent Identical Matches	No. Identical Matches	No. Mis-matches	Total No. Gaps
		RNA gene, partial sequence		128						
Otu6002	AY442141	Borrelia lonestari isolate LS-1 16S ribosomal RNA gene, partial sequence	252	7.00E-128	466	252	100.0	252	0	0
Otu6002	AY682921	Borrelia lonestari isolate MO2002-V2 16S ribosomal RNA gene, partial sequence	252	7.00E-128	466	252	100.0	252	0	0
Otu6002	AY682920	Borrelia lonestari isolate MO2002-V1 16S ribosomal RNA gene, partial sequence	252	7.00E-128	466	252	100.0	252	0	0
Otu6002	NR_121718	Borrelia coriaceae strain Co53 16S ribosomal RNA gene, complete sequence	252	3.00E-126	460	249	99.6	251	1	0
Otu6002	KF569941	Borrelia theileri strain KAT 16S ribosomal RNA gene, partial sequence	252	3.00E-126	460	249	99.6	251	1	0
Otu6002	CP005745	Borrelia coriaceae Co53, complete genome	252	3.00E-126	460	249	99.6	251	1	0
Otu6002	CP005706	Borrelia hermsii YBT, complete genome	252	3.00E-126	460	249	99.6	251	1	0
Otu6002	NR_102957	Borrelia hermsii DAH strain DAH 16S ribosomal RNA, complete sequence	252	3.00E-126	460	249	99.6	251	1	0
Otu6003	AM182231	Borrelia spielmanii partial 16S rRNA gene, strain PJes	251	6.00E-124	453	245	99.2	249	2	0
Otu6003	AM182230	Borrelia spielmanii partial 16S	251	6.00E-	453	245	99.2	249	2	0

Query	Subject Accession No.	Subject Title	Alignment Length	Expect Value	Bit Score	Raw Score	Percent Identical Matches	No. Identical Matches	No. Mis-matches	Total No. Gaps
		rRNA gene, strain PAnz		124						
Otu6003	AY570512	Borrelia sp. Ir-5215 16S ribosomal RNA gene, partial sequence	251	6.00E-124	453	245	99.2	249	2	0
Otu6003	U44939	Borrelia burgdorferi 16S ribosomal RNA gene, partial sequence	251	6.00E-124	453	245	99.2	249	2	0
Otu6003	KF569941	Borrelia theileri strain KAT 16S ribosomal RNA gene, partial sequence	252	3.00E-122	448	242	98.8	249	2	1
Otu6003	CP005706	Borrelia hermsii YBT, complete genome	252	3.00E-122	448	242	98.8	249	2	1
Otu6003	NR_102957	Borrelia hermsii DAH strain DAH 16S ribosomal RNA, complete sequence	252	3.00E-122	448	242	98.8	249	2	1
Otu6003	CP009212	Borrelia afzelii Tom3107, complete genome	251	3.00E-122	448	242	98.8	248	3	0
Otu6003	JX219479	Borrelia turdi isolate Hongdo-4-1 16S ribosomal RNA gene, partial sequence	251	3.00E-122	448	242	98.8	248	3	0
Otu6003	JX219478	Uncultured Borrelia sp. clone Hongdo-4-4 16S ribosomal RNA gene, partial sequence	251	3.00E-122	448	242	98.8	248	3	0

4.8 References

Anderson MJ (2001) A new method for non-parametric multivariate analysis of variance. *Austral Ecology*, **26**, 32–46.

Andreotti R, León AAP de, Dowd SE, Guerrero FD, Bendele KG, Scoles GA (2011) Assessment of bacterial diversity in the cattle tick *Rhipicephalus* (*Boophilus*) *microplus* through tag-encoded pyrosequencing. *BMC Microbiology*, **11**, 6.

Backeljau T, de Bruyn L, de Wolf H, Jordaens K, van Dongen S, Winnepenninckx B (1996) Multiple UPGMA and neighbor-joining trees and the performance of some computer packages. *Molecular Biology and Evolution*, **13**, 309–313.

Benson MJ, Gawronski JD, Eveleigh DE, Benson DR (2004) Intracellular symbionts and other bacteria associated with deer ticks (*Ixodes scapularis*) from Nantucket and Wellfleet, Cape Cod, Massachusetts. *Applied Environmental Microbiology*, **70**, 616–620.

Bermúdez SE, Eremeeva ME, Karpathy SE, Samudio F, Zambrano ML, Zaldivar Y, Motta JA, Dasch GA (2009) Detection and identification of rickettsial agents in ticks from domestic mammals in eastern Panama. *Journal of Medical Entomology*, **46**, 856–861.

Bray JR, Curtis JT (1957) An ordination of the upland forest communities of southern Wisconsin. *Ecological Monographs*, **27**, 326–349.

Brucker RM, Bordenstein SR (2012) The roles of host evolutionary relationships (genus: *Nasonia*) and development in structuring microbial communities. *Evolution*, **66**, 349–362.

Brucker RM, Bordenstein SR (2013) The hologenomic basis of speciation: gut bacteria cause hybrid lethality in the genus *Nasonia*. *Science*, **341**, 667–669.

Budachetri K, Browning RE, Adamson SW, Dowd SE, Chao C-C, Ching W-M, Karim S (2014) An insight into the microbiome of the *Amblyomma maculatum* (Acari: Ixodidae). *Journal of Medical Entomology*, **51**, 119–129.

Carpi G, Cagnacci F, Wittekindt NE, Zhao F, Qi J, Tomsho LP, Drautz DI, Rizzoli A, Schuster SC (2011) Metagenomic profile of the bacterial communities associated with *Ixodes ricinus* ticks. *PLoS ONE*, **6**, e25604.

Cock PJA, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, Friedberg I, Hamelryck T, Kauff F, Wilczynski B, de Hoon MJL (2009) Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics*, **25**, 1422–1423.

Colman DR, Toolson EC, Takacs-Vesbach CD (2012) Do diet and taxonomy influence insect gut bacterial communities? *Molecular Ecology*, **21**, 5124–5137.

Douglas AE (1998) Nutritional interactions in insect-microbial symbioses: aphids and their symbiotic bacteria *Buchnera*. *Annual Review of Entomology*, **43**, 17–37.

Duguma D, Rugman-Jones P, Kaufman MG, Hall MW, Neufeld JD, Stouthamer R, Walton WE (2013) Bacterial communities associated with *Culex* mosquito larvae and two emergent aquatic plants of bioremediation importance. *PLoS ONE*, **8**, e72522.

Ennen JR, Qualls CP (2011) Distribution and habitat utilization of the gopher tortoise tick (*Amblyomma tuberculatum*) in southern Mississippi. *Journal of Parasitology*, **97**, 202–206.

Epis S, Sassera D, Beninati T, Lo N, Beati L, Piesman J, Rinaldi L, McCoy KD, Torina A, Sacchi L, Clementi E, Genchi M, Magnino S, Bandi C (2008) *Midichloria mitochondrii* is widespread in hard ticks (Ixodidae) and resides in the mitochondria of phylogenetically diverse species. *Parasitology*, **135**, 485–494.

Eren AM, Maignien L, Sul WJ, Murphy LG, Grim SL, Morrison HG, Sogin ML (2013) Oligotyping: differentiating between closely related microbial taxa using 16S rRNA gene data. *Methods in Ecology and Evolution*, **4**, 1111–1119.

Estrada-Peña A, Jongejan F (1999) Ticks feeding on humans: a review of records on human-biting Ixodoidea with special reference to pathogen transmission. *Experimental & Applied Acarology*, **23**, 685–715.

Estrada-Peña A, Venzal JM, Mangold AJ, Cafrune MM, Guglielmone AA (2005) The *Amblyomma maculatum* Koch, 1844 (Acari: Ixodidae: Amblyomminae) tick group: diagnostic characters, description of the larva of *A. parvitarsum* Neumann, 1901, 16S rDNA sequences, distribution and hosts. *Systematic Parasitology*, **60**, 99–112.

Esty WW (1986) The efficiency of Good's nonparametric coverage estimator. *Annals of Statistics*, **14**, 1257–1260.

Evans J, Sheneman L, Foster JA (2006) Relaxed neighbor-joining: a fast distance-based phylogenetic tree construction method. *Journal of Molecular Evolution*, **62**, 785–792.

Faith DP, Minchin PR, Belbin L (1987) Compositional dissimilarity as a robust measure of ecological distance. *Vegetatio*, **69**, 57–68.

Goethert HK, Telford SR (2005) A new *Francisella* (Beggiatiales: Francisellaceae) inquiline within *Dermacentor variabilis* say (Acari: Ixodidae). *Journal of Medical Entomology*, **42**, 502–505.

Guglielmone AA, Robbins RG, Apanaskevich DA, Petney TN, Estrada-Pena A, Horak IG (2014) *The Hard Ticks of the World (Acari: Ixodida: Ixodidae)*. Springer, London.

Hawlena H, Rynkiewicz E, Toh E, Alfred A, Durden LA, Hastriter MW, Nelson DE, Rong R, Munro D, Dong Q, Fuqua C, Clay K (2013) The arthropod, but not the vertebrate host or its environment, dictates bacterial community composition of fleas and ticks. *The ISME Journal*, **7**, 221–223.

Hooker WA, Bishopp FC, Wood HP, Hunter WD (1912) *The Life History and Bionomics of Some North American Ticks*. US Dept. of Agriculture, Bureau of Entomology, Washington DC.

Jaccard P (1908) Nouvelles recherches sur la distribution florale. *Société Vaudoise des Sciences Naturelles*, **44**, 223–270.

Jasinskas A, Zhong J, Barbour AG (2007) Highly prevalent *Coxiella* sp. bacterium in the tick vector *Amblyomma americanum*. *Applied and Environmental Microbiology*, **73**, 334–336.

Jones RT, Sanchez LG, Fierer N (2013) A cross-taxon analysis of insect-associated bacterial diversity. *PLoS ONE*, **8**, e61218.

Keirans JE, Litwak TR (1989) Pictorial key to the adults of hard ticks, family Ixodidae (Ixodida: Ixodoidea), east of the Mississippi River. *Journal of Medical Entomology*, **26**, 435–448.

Lalzar I, Harrus S, Mumcuoglu KY, Gottlieb Y (2012) Composition and seasonal variation of *Rhipicephalus turanicus* and *Rhipicephalus sanguineus* bacterial communities. *Applied and Environmental Microbiology*, **78**, 4110–6.

Lane R, Quistad G (1998) Borreliacidal factor in the blood of the western fence lizard (*Sceloporus occidentalis*). *Journal of Parasitology*, **84**, 29–34.

Lozupone CA, Hamady M, Kelley ST, Knight R (2007) Quantitative and qualitative β diversity measures lead to different insights into factors that structure microbial communities. *Applied and Environmental Microbiology*, **73**, 1576–1585.

Lozupone C, Knight R (2005) UniFrac: a new phylogenetic method for comparing microbial communities. *Applied and Environmental Microbiology*, **71**, 8228–8235.

Mantel N (1967) The detection of disease clustering and a generalized regression approach. *Cancer Research*, **27**, 209–220.

Margush T, McMorris FR (1981) Consensus n-trees. *Bulletin of Mathematical Biology*, **43**, 239–244.

McFall-Ngai M, Hadfield MG, Bosch TCG, Carey HV, Domazet-Lošo T, Douglas AE, Dubilier N, Eberl G, Fukami T, Gilbert SF, Hentschel U, King N, Kjelleberg S, Knoll AH, Kremer N, Mazmanian SK, Metcalf JL, Nealson K, Pierce NE, Rawls JF, Reid A, Ruby EG, Rumpho M, Sanders JG, Tautz D, Wernegreen JJ (2013) Animals in a bacterial world, a new imperative for the life sciences. *Proceedings of the National Academy of Sciences*, **110**, 3229–3236.

McMullen HL, Sauer JR, Burton RL (1976) Possible role in uptake of water vapour by ixodid tick salivary glands. *Journal of Insect Physiology*, **22**, 1281–1285.

Menchaca AC, Visi DK, Strey OF, Teel PD, Kalinowski K, Allen MS, Williamson PC (2013) Preliminary assessment of microbiome changes following blood-feeding and survivorship in the *Amblyomma americanum* nymph-to-adult transition using semiconductor sequencing. *PLoS ONE*, **8**, e67129.

Minchin PR (1987) An evaluation of the relative robustness of techniques for ecological ordination. *Vegetatio*, **69**, 89–107.

Montgomery MK, McFall-Ngai M (1994) Bacterial symbionts induce host organ morphogenesis during early postembryonic development of the squid *Euprymna scolopes*. *Development*, **120**, 1719–1729.

Moreno CX, Moy F, Daniels TJ, Godfrey HP, Cabello FC (2006) Molecular analysis of microbial communities identified in different developmental stages of *Ixodes scapularis* ticks from Westchester and Dutchess Counties, New York. *Environmental Microbiology*, **8**, 761–772.

Nakao R, Abe T, Nijhof AM, Yamamoto S, Jongejan F, Ikemura T, Sugimoto C (2013) A novel approach, based on BLSOMs (Batch Learning Self-Organizing Maps), to the microbiome analysis of ticks. *The ISME Journal*.

Nava S, Elshenawy Y, Eremeeva ME, Sumner JW, Mastropaolo M, Paddock CD (2008) *Rickettsia parkeri* in Argentina. *Emerging Infectious Diseases*, **14**, 1894–1897.

Nei M, Li WH (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences*, **76**, 5269–5273.

Noda H, Munderloh U, Kurtti T (1997) Endosymbionts of ticks and their relationship to *Wolbachia* spp. and tick-borne pathogens of humans and animals. *Applied and Environmental Microbiology*, **63**, 3926.

Norris DE, Klompen JSH, Black WC (1999) Comparison of the mitochondrial 12S and 16S ribosomal DNA genes in resolving phylogenetic relationships among hard ticks (Acari: Ixodidae). *Annals of the Entomological Society of America*, **92**, 117–129.

Ochman H, Worobey M, Kuo C-H, Ndjango J-BN, Peeters M, Hahn BH, Hugenholtz P (2010) Evolutionary relationships of wild hominids recapitulated by gut microbial communities. *PLoS Biology*, **8**, e1000546.

Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens MHH, Wagner H (2013) *vegan: Community Ecology Package*.

Oliver KM, Russell JA, Moran NA, Hunter MS (2003) Facultative bacterial symbionts in aphids confer resistance to parasitic wasps. *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 1803–1807.

Paddock CD, Fournier P-E, Sumner JW, Goddard J, Elshenawy Y, Metcalfe MG, Loftis AD, Varela-Stokes A (2010) Isolation of *Rickettsia parkeri* and identification of a novel spotted fever group *Rickettsia* sp. from Gulf Coast ticks (*Amblyomma maculatum*) in the United States. *Applied and Environmental Microbiology*, **76**, 2689–2696.

Pais R, Lohs C, Wu Y, Wang J, Aksoy S (2008) The obligate mutualist *Wigglesworthia glossinidia* influences reproduction, digestion, and immunity processes of its host, the tsetse fly. *Applied and Environmental Microbiology*, **74**, 5965–5974.

Paradis E, Claude J, Strimmer K (2004) APE: analyses of phylogenetics and evolution in R language. *Bioinformatics*, **20**, 289–290.

Pretzman C, Daugherty N, Poetter K, Ralph D (1990) The distribution and dynamics of *Rickettsia* in the tick population of Ohio. *Annals of the New York Academy of Sciences*, **590**, 227–236.

Qiu Y, Nakao R, Ohnuma A, Kawamori F, Sugimoto C (2014) Microbial population analysis of the salivary glands of ticks: a possible strategy for the surveillance of bacterial pathogens (LS van Overbeek, Ed,). *PLoS ONE*, **9**, e103961.

Ramette A (2007) Multivariate analyses in microbial ecology. *Fems Microbiology Ecology*, **62**, 142–160.

Randolph SE, Gern L, Nuttall PA (1996) Co-feeding ticks: epidemiological significance for tick-borne pathogen transmission. *Parasitology Today*, **12**, 472–479.

R Core Team (2013) *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

Richter D, Matuschka F-R, Spielman A, Mahadevan L (2013) How ticks get under your skin: insertion mechanics of the feeding apparatus of *Ixodes ricinus* ticks. *Proceedings of the Royal Society B: Biological Sciences*, **280**, 20131758.

Sanders JG, Powell S, Kronauer DJC, Vasconcelos HL, Frederickson ME, Pierce NE (2014) Stability and phylogenetic correlation in gut microbiota: lessons from ants and apes. *Molecular Ecology*, **23**, 1268–1283.

Sassera D, Epis S, Pajoro M, Bandi C (2013) Microbial symbiosis and the control of vector-borne pathogens in tsetse flies, human lice, and triatomine bugs. *Pathogens and Global Health*, **107**, 285–292.

Schliep KP (2011) phangorn: phylogenetic analysis in R. *Bioinformatics*, **27**, 592–593.

Schloss PD (2013) Schloss SOP. *Mothur Wiki*. Available at: www.mothur.org/wiki/Schloss_SOP. Accessed on 30 Sept 2013.

Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ, Weber CF (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology*, **75**, 7537–7541.

Scoles GA (2004) Phylogenetic analysis of the *Francisella*-like endosymbionts of *Dermacentor* ticks. *Journal of Medical Entomology*, **41**, 277–286.

Sonenshine D, Levy G (1972) Ecology of the American dog tick, *Dermacentor variabilis* in a study area in Virginia. 2. distribution in relation to vegetative types. *Annals of the Entomological Society of America*, **65**, 1175–1182.

Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution*, **10**, 512–526.

Vavre F, Kremer N (2014) Microbial impacts on insect evolutionary diversification: from patterns to mechanisms. *Current Opinion in Insect Science*, **in press**.

Williams CB (1964) *Patterns in the Balance of Nature*. Academic, London.

Williams-Newkirk AJ, Rowe LA, Mixson-Hayden TR, Dasch GA (2014) Characterization of the bacterial communities of life stages of free living lone star ticks (*Amblyomma americanum*). *PLoS ONE*, **9**, e102130.

Yuan D (2010) A metagenomic study of the tick midgut. Thesis (MS) Thesis. The University of Texas, Houston, Texas.

Zhang X-C, Yang Z-N, Lu B, Ma X-F, Zhang C-X, Xu H-J (2014) The composition and transmission of microbiome in hard tick, *Ixodes persulcatus*, during blood meal. *Ticks and Tick-borne Diseases*, **5**, 864–870.

Zhong JM, Jasinskas A, Barbour AG (2007) Antibiotic treatment of the tick vector *Amblyomma americanum* reduced reproductive fitness. *PLoS One*, **2**, e405.

Zhou J, Wu L, Deng Y, Zhi X, Jiang Y-H, Tu Q, Xie J, Van Nostrand JD, He Z, Yang Y (2011) Reproducibility and quantitation of amplicon sequencing-based detection. *The ISME Journal*, **5**, 1303–1313.

Zilber-Rosenberg I, Rosenberg E (2008) Role of microorganisms in the evolution of animals and plants: the hologenome theory of evolution. *FEMS Microbiology Reviews*, **32**, 723–735.

CHAPTER 5

The mitochondrial genome of *Amblyomma americanum*

5.1 Summary

Amblyomma americanum is an abundant tick in the southeastern, midwestern, and northeastern United States. It is a vector of multiple diseases, but limited genomic resources are available for it. I sequenced the complete mitochondrial genome of a single female *A. americanum* collected in Georgia using the Illumina platform. The consensus sequence was 14,709 bp long, and the mean coverage across the assembly was >12,000x. All genomic features were present and in the expected order for the Metastriata. Heteroplasmy rates were low compared to the most closely related tick for which data are available, *Amblyomma cajennense*. The phylogeny derived from the concatenated protein coding and rRNA genes from the 33 available tick mitochondrial genomes was consistent with those previously proposed for the Acari. This is the first complete mitochondrial sequence for *A. americanum*, which provides a useful reference for future studies of *A. americanum* population genetics and tick phylogeny.

5.2 Introduction

Genomic tools for ixodid (hard) ticks are severely limited in availability. Ixodid genome sizes are estimated to range from 1.04 – 7.1 Gb and are known to contain long stretches of repetitive DNA (Nene 2009), making them a challenge even to modern Next

Generation sequencing (NGS) and assembly methods. No closed ixodid genomes have been published, and only two species (*Ixodes scapularis* and *Rhipicephalus [Boophilus] microplus*) have substantial whole genome shotgun assembly data in GenBank. This lack of high quality, deep coverage reference sequence information significantly hinders research in areas of tick biology that are critical to the development of tick-borne disease control methods, including metabolic pathways (Institute of Medicine 2011), population dynamics and genetics (Shao & Barker 2007; Araya-Anchetta *et al.* 2015), tick-host interactions (McCoy *et al.* 2013) and tick-microbiome-pathogen interactions (Clay & Fuqua 2010).

While the availability of closed tick genomes may be some years in the future, the 14.5 – 15.0 kb tick mitochondrial (mt) genomes are well within the capabilities of current NGS technology. While many mitochondrial genes have a long history of use as population genetic markers (Shao & Barker 2007; Araya-Anchetta *et al.* 2015), fragments of the rRNA and cytochrome oxidase genes have been the predominant targets used in ixodid ticks (*e.g.* Norris *et al.* 1996; Mixson *et al.* 2006). To date, 21 complete ixodid mt genomes are currently available at GenBank, but only four of these are from North American species of ticks.

Amblyomma comprises the third largest genus of ixodid ticks (Kolonin 2009) with 129 species currently recognized (Species2000 2015), approximately half of which are found in the Americas (Kolonin 2009). Recent evidence from mt genome analysis supports the reassignment of *A. elaphense* and *A. sphenodonti*, two of the so-called "primitive" *Amblyomma*, to new genera (Burger *et al.* 2012, 2013). Many of the *Amblyomma* are among the most economically important ticks worldwide due to their role as vectors of

human and veterinary disease and the damage caused to livestock by the noxious wounds inflicted by the elongated chelicera of these species (Jongejan & Uilenberg 2004).

The lone star tick, *Amblyomma americanum*, is the most abundant and aggressive human-biting tick in the southeastern United States (Merten & Durden 2000). None of the three motile life stages of the tick are fastidious in their host choices. Mammals are preferred hosts, but birds and reptiles have also been documented (reviewed in Guglielmone *et al.* 2014). The broad host range of *A. americanum* positions it as a vector of human and veterinary zoonotic diseases, including but not limited to human ehrlichioses (Buller *et al.* 1999), a veterinary ehrlichiosis (Loftis *et al.* 2006), Heartland viral infection (Savage *et al.* 2013), rickettsiosis (Berrada *et al.* 2011), and tularemia (reviewed in Childs & Paddock 2003). *Amblyomma americanum* has expanded its geographical range in the last century into the northeastern and mid-western states (Cortinas & Spomer 2013, 2014; Springer *et al.* 2014), greatly increasing the human and domestic animal populations potentially exposed to the highly allergenic bite of this species and the diseases it transmits.

Using Illumina sequences obtained from DNA from a single female from central Georgia, I assembled and annotated the complete mt genome of *A. americanum*. A scaffold region in the assembly was closed and several other gene regions were confirmed by Sanger sequencing. I determined the phylogenetic relationship of this sequence to that of other tick mitochondrial genome sequences and examined patterns of codon bias and RNA folding in *A. americanum* and related tick species.

5.3 Materials and methods

Sample collection

Questing *A. americanum* were collected by running a 1 m^2 cloth over vegetation at Panola Mountain State Park near Atlanta, Georgia on 9 May 2011 and placed live into 70% ethanol. Collections were performed with permits from the Georgia Department of Natural Resources and the permission of park management. Ticks were identified to species using standard morphological keys (Keirans & Litwak 1989). Two adult females from the same 100 m transect were used in this study. Surface decontamination and DNA extraction followed Williams-Newkirk *et al.* (2014).

Illumina sequencing

The DNA of a single adult female was submitted to the Centers for Disease Control and Prevention Biotechnology Core Facility and a TruSeq Sample Preparation Kit (Illumina, Inc., San Diego, California) was used to prepare a 100 bp paired-end shotgun library with a target insert size of 300 bp. The library was sequenced on an Illumina HiSeq2500 using a HiSeq SBS kit.

Sanger sequencing

Amplicons for regions selected for Sanger sequencing were produced using the DNA extracted from the second female *A. americanum* as the template. Primers were designed using Primer3 (version 0.4.0) (Koressaar & Remm 2007; Untergasser *et al.* 2012) in Geneious (version 7.1.3) (Biomatters 2015); primer sequences, annealing temperatures, and the number of cycles are given in Table 5.1. Each 20 µL PCR contained 2 µL of

template (diluted 1:1000 in water), 10 µL Taq PCR Master Mix (Qiagen Inc., Valencia, California, USA), 6 µL water, and 1 µM each forward and reverse primers. The reactions were run in a Master Cycler Gradient thermocycler (Brinkmann Instruments, Westbury, New York, USA) at 94°C for 5 min, followed by 35-40 cycles of 94°C for 30 s, annealing for 30 s at 55-65°C , and 72°C for 1 min, with a final extension at 72°C for 10 min. The amplicons were visualized on a 1% agarose gel, and the bands were excised and purified using the Wizard SV Gel and PCR Clean-Up System (Promega, Madison, Wisconsin, USA). Purified amplicons were sequenced bi-directionally using the BigDye Terminator v3.1 Cycle Sequencing Kit (ABI, Foster City, California, USA) on an ABI 3130xl Genetic Analyzer. Trace files were trimmed and assembled in Geneious.

Sequence assembly and annotation

Raw Illumina read quality checks were performed on the samples using FastQC ("FastQC" 2014) and with in-house Perl scripts and Java programs. Primary analysis of the Illumina data was performed using CLC Bio Genomics Workbench (version 5.5.1) ("CLCbio" 2014) (hereafter CLC). The raw read files for each sample were trimmed with length (minimum: 50 bp) and quality score (0.02) filters. The trimmed reads were assembled into contigs with specific parameter settings (length fraction: 0.8; similarity fraction: 0.8; minimum contig length: 400 bp), and assembly statistics were parsed out in a table format using in-house scripts.

Contigs from the de novo assembly were analyzed in CLC (version 7.0.3) by BLAST against the *A. cajennense* mt genome (NC_020333.1). The two longest contigs that matched this reference sequence were 13,391 bp and 1,376 bp in length. When assembled

to the *A. cajennense* reference sequence in Geneious, the contigs were joined and found to cover the entire reference genome except for a 161 bp scaffold region corresponding to control region 1 (CR1). All ixodid mt genomes sequenced to date contain two control regions approximately 300 bp in length that appear to evolve in concert (Burger *et al.* 2012). Control region 2 (CR2) was present in the *A. americanum* mt assembly, but the depth of Illumina sequence coverage was approximately doubled in this region relative to the remaining sequence. I concluded that the reads for both control regions had assembled to CR2 due to their homology. Primers for Sanger sequencing of both CR1 and CR2 were designed to resolve the assembly in these regions (Table 5.1). Sanger sequencing was also used to verify the assembly in regions where Illumina contigs were manually joined and to obtain fragments of the ATP8, ND3, 12S, and 16S rRNA genes (Table 5.1). Sanger sequences, reference rRNA gene sequences from GenBank (L34313.1, U95849.1), and the consensus from the *de novo* Illumina read assembly were assembled into a single closed, circular sequence in Geneious. The Illumina reads were then remapped to the consensus to determine the final consensus sequence. All sequence data were deposited in the National Center for Biotechnology GenBank repository under accession numbers KP941755 (mitochondrial consensus sequence) and KP997023 – KP997030 (Sanger data).

Initial annotations were made using the MITOS WebServer (revision 605) (Bernt *et al.* 2013; "MITOS WebServer" 2015). Automated annotations were manually corrected by aligning the *A. americanum* consensus sequence to the *A. cajennense* reference sequence and translating CDS regions to ensure full ORFs were captured. Truncated stop codons were verified using Burger *et al.* (2012). Transfer RNA folding patterns provided by

MITOS were reviewed and atypical patterns were compared to known ixodid tick tRNA folding patterns in the mitochondrial transfer RNA database ("mitotRNAdb" 2009; Jühling *et al.* 2009) and to tRNA sequences extracted from the complete *Amblyomma sensu lato* mt genomes at GenBank and folded on the MITOS WebServer. The secondary structure of the 16S and 12S rRNAs were examined by comparison to the most closely related tick for which mt genome data was available, *A. cajennense*, using the LocARNA website (version 1.8.0) (Smith *et al.* 2010; Freiburg RNA Tools 2015) to perform simultaneous structure-based alignments and secondary structure predictions.

Heteroplasmy analysis

Rates of heteroplasmy in the mitochondrial assembly were determined following the methods of Xiong *et al.* (2013). Briefly, SNPs with a minimum frequency of 1.5% were identified in the assembly using the "Find SNPs / variations" tool in Geneious with the "Find SNPs only" option. The false positive rate for each site (P_f) was calculated according to the formula given in Xiong *et al.* (2013) from Storey and Tibshirani (2003) using an error rate of 0.5%. The false discovery rate (Q_f) for each site was calculated using the R package "qvalue" version 1.40.1 with $\pi_0 = 1$ because π_0 estimation fails under this method with a small number of P_f values (Dabney & Storey 2015). This is equivalent to the Benjamini and Hochberg (Benjamini & Hochberg 1995) method of Q_f estimation. Significance thresholds were set to $P_f < 1\%$ and $Q_f < 0.1\%$, and SNPs from sequencing error "hot spots" were excluded from analysis (Nakamura *et al.* 2011; Minoche *et al.* 2011; Xiong *et al.* 2013). All calculations were performed on the R statistical platform (version 3.1.2) (R Core Team 2013); the analysis scripts I used are available in supplemental materials (Figure 5.S1).

Genomic content and codon usage analysis

Three indices of codon bias were used to evaluate variation among the *Amblyomma s. l.* in codon usage. Relative synonymous codon usage (RSCU) is a measure of the deviation of synonymous codon usage from random (Sharp *et al.* 1986). Codons with a RSCU value of one are found at a rate that does not deviate from random, while values less than one and greater than one are indicative of codon usage frequency that is less than or greater than expected, respectively. Calculations were performed on concatenated genes of the 13 protein coding genes from the six available *Amblyomma s. l.* genomes using the software program CodonW version 1.4.4 (Peden 1999) as implemented on the Institut Pasteur Mobyle website (Institut Pasteur IT Center & Ressource Parisienne en Bioinformatique Structurale 2013). Truncated stop codons, which are believed to be completed by post-transcriptional poly-adenylation (Burger *et al.* 2012), were manually completed to maintain the correct reading frame in the concatenated sequences.

The effective number of codons (EN_c) index measures codon bias as the number of equally utilized codons required to reproduce the pattern observed across all amino acids in a gene (Wright 1990). The possible values of EN_c are 20 (extremely biased; one codon per amino acid) to 62 (unbiased). The EN_c index was calculated separately on each protein coding gene from the six *Amblyomma s. l.* mt genomes in CodonW. The ATP8 gene was excluded from this analysis because CodonW repeatedly failed to compute EN_c scores for three of six samples despite extensive error checking of the data. The EN_c values were compared between *Amblyomma s. l.* species using a single factor repeated measures ANOVA executed in R. Pairwise t-tests with the Holm (Holm 1979) correction for multiple comparisons were used as a post-hoc test.

The third measurement of codon bias applied to the *Amblyomma s. l.* mt genomes allowed us to quantify the pairwise differences between the species in relative codon usage. Adapting the methods of Suzuki et al. (2009) for quantifying codon bias between genes within a genome, the 13 protein coding genes for each species were concatenated and the number of times each codon was used per species was counted in CodonW. Next, the ratio of each codon's count to the most abundant synonymous codon's count was calculated for each species (x_{ac} in Suzuki *et al.* 2009). The pairwise dissimilarity between each species was then calculated as one minus the Pearson's product moment correlation coefficient between the vectors of codon x_{ac} values for each species (D_{ij} in Suzuki *et al.* 2009). A dendrogram visualizing the similarities in codon bias between species was produced by applying a complete linkage hierarchical clustering algorithm to the resulting distance matrix. All calculations were performed in R; analysis scripts are available in the supplemental materials (Figure 5.S2).

GC content and GC skew were calculated using a 50 bp sliding window. GC skew was calculated as (G-C)/(G+C).

Phylogenetic analysis

The mt genomes of the mite *Tetranychus malaysiensis* and the available argasid and ixodid tick species were downloaded from GenBank. The protein coding and rRNA genes were extracted from each genome using the annotations provided in GenBank. Each protein coding region was individually aligned using the "Align by translation" function in Geneious with default settings and the rRNA genes were aligned using the Geneious aligner with default settings. Aligned regions were concatenated and used to

produce 100 bootstrap maximum likelihood trees in GARLI 2.01 (Zwickl 2006) using the program default settings. A consensus tree was produced from the replicate trees by mapping the bootstrap values onto the tree with the best maximum-likelihood score with the SumTrees script in DendroPy (version 3.12.0) (Sukumaran & Holder 2010).

5.4 Results and discussion

Genome description

The mt genome of *A. americanum* was assembled from 1,790,436 Illumina paired end reads to form a circular consensus sequence of 14,709 bp (Figure 5.1). Six regions were confirmed by targeted Sanger sequencing. Mean Illumina read coverage across the genome was 12,144x. Coverage was the highest across the 16S rRNA gene (maximum 25,511x) and lowest at the 5' end of the NAD1 gene (minimum 2,173x) (Figure 5.1). All of the expected 13 protein coding regions, 2 rRNAs, 22 tRNAs, and 2 control regions were detected, and their order was consistent with that of the previously reported metastriate ticks (Black & Roehrdanz 1998; Shao *et al.* 2004; Burger *et al.* 2012, 2013, 2014; Xiong *et al.* 2013) (Figure 5.1). Total GC content across the genome was very low (23.2%) but similar to that of other known metastriate ticks (Black & Roehrdanz 1998; Shao *et al.* 2004; Burger *et al.* 2012, 2013, 2014; Xiong *et al.* 2013). The mean GC skew of the J strand was negative (-0.118) as is typical of metazoan mitochondria (Hassanin *et al.* 2005) (Figure 5.1).

RNA secondary structure

The folding patterns of the 22 *A. americanum* mt tRNAs were predicted and visually assessed for confirmation to the canonical metazoan "cloverleaf" secondary structure. Two tRNAs, cysteine and serine 1, were missing the D-arm (Figure 5.S3). Comparison to the tick tRNAs available in the mitochondrial transfer RNA database ("mitotRNAdb" 2009; Jühling *et al.* 2009) confirmed that the missing arm of the serine 1 tRNA is a common feature that is shared by all five of the species included in that database (*Carios capensis*, *Ixodes hexagonus*, *I. holocyclus*, *I. persulcatus*, and *I. uriae*) as well as the five *Amblyomma s. l.* species with whole mt genome data available at GenBank. However, examination of the cysteine tRNA from the mt tRNA database and the *Amblyomma s. l.* revealed a considerable diversity of secondary structures, including three species with both the D and T-arms missing (*A. cajennense, A. fimbriatum,* and *A. elaphense*), one species missing the D-arm (*A. sphenodonti*), one species missing the T-arm (*I. holocyclus*), and five species with three intact arms. Sanger sequencing using DNA extracted from a second *A. americanum* female tick from the same population confirmed the assembly of this region (Figure 5.1). These findings are consistent with an earlier study of arachnid tRNA evolution (Masta & Boore 2008) which found that the Parasitiformes (unlike some other arachnid groups, including the Acariformes) have the typical metazoan tRNA secondary structures with one to two deviations per mt genome.

We also performed structure-based alignments of the *A. americanum* and *A. cajennense* 12S (Figure 5.S4a) and 16S (Figure 5.S5a) rRNA genes because of their importance in tick identification and population genetics and because *A. cajennense* is the closest relative to *A. americanum* with a complete mitochondrial sequence. The consensus secondary structure for each alignment (Figure 5.S4b and 5.S5b) further demonstrates the

even distribution of polymorphisms within the two sequences and relative location of each within both molecules.

Heteroplasmy and other polymorphisms

Heteroplasmy is the presence of multiple mitochondrial genotypes within a single individual. Low levels of heteroplasmy are common in plants and animals (Kmiec *et al.* 2006) and have been well documented in many arthropod lineages (*e.g.* Nardi *et al.* 2001; Leeuwen *et al.* 2008; Magnacca & Brown 2010; Xiong *et al.* 2013). A recent examination of heteroplasmy rates in seven tick species from five genera (two argasid and three ixodid) reported broad variation in the number of heteroplasmic sites per species (six sites in each *Haemaphysalis* species and 166 sites in *A. cajennense*) (Xiong *et al.* 2013). Because no other *Amblyomma* species was tested in the previous paper and given my high depth of mitochondrial sequence coverage for *A. americanum*, I examined the present assembly to determine if high rates of heteroplasmy were characteristic of the genus. I identified six heteroplasmic sites in this *A. americanum* female which were located in the ND1, ND2, CYTB, and 16S genes (Table 5.2). This is low but within the range of previously assessed rates of heteroplasmy in ticks, indicating that either heteroplasmy rates are not genus specific or atypical individuals have been sampled from these species. Sampling of additional species and more individuals within each species will be necessary to determine if phylogenetic patterns of heteroplasmy exist. Of the four heteroplasmic sites in protein coding regions in *A. americanum*, only the SNP in the CYTB gene was nonsynonymous. Nonsynonymous heteroplasmic sites in the CYTB gene of the spider mite *Tetranychus urticae* have been linked with the rapid evolution of pesticide resistance (Leeuwen *et al.* 2008). *Amblyomma americanum* is a common pest of

livestock (Childs & Paddock 2003) and the evolution of pesticide resistance in this species would have significant economic impact. Further investigation of this CYTB polymorphism will be necessary to determine if it has any functional significance.

Six regions of the *A. americanum* mt genome were sequenced by Sanger methodology using the DNA of a second adult female collected at the same time as the female whose DNA was used to generate the Illumina sequences (Figure 5.1). Nine polymorphisms that were well supported by both sequence data sets were identified; of those, seven were in the 16S rRNA gene and the other two were in the ND2 and CYTB genes. All were SNPs except for two INDELs of three to four bases in the 16S rRNA gene. None of these polymorphisms overlapped with heteroplasmic sites. Comparison of the 33 16S rRNA haplotypes identified by Trout *et al.* (2010) in Arkansas to my consensus sequence indicates that the female used for Illumina sequencing is a novel haplotype. The available data suggest that the 16S rRNA gene may be the best single mt locus for population genetics studies (Mixson *et al.* 2006; Trout *et al.* 2010), but additional data for the other mt loci are needed before a definitive answer can be reached.

Phylogenetic analysis

A maximum-likelihood tree was constructed from the concatenated alignments of all of the protein coding and rRNA genes from the 32 tick mt genomes available at GenBank, plus the present *A. americanum* assembly and a spider mite (*Tetranychus malaysiensis*) mt genome selected as an outgroup (Figure 5.2). The two major families represented on the tree (Argasidae and Ixodidae) are monophyletic, as are the subfamilies of the Ixodidae. As was expected, *A. americanum* was most closely related to *A. cajennense* and

clustered with the other *Amblyomma sensu stricto*. The positioning of *A. sphenodonti* as a member of the genus *Haemaphysalis* is consistent with the findings of Burger *et al.* (2014). In contrast to Burger *et al.*'s work, my tree placed *A. elaphense* as a sister group to the *Bothriocroton* rather than as a sister group to the entire Metastriata. However, the support for *A. elaphense*'s position in my tree was weak (Figure 5.2). Genomic data from additional species will be necessary before the position of *A. elaphense* can be resolved.

Codon usage

Start codon usage was equitable among the mt protein coding genes in *A. americanum*; six genes used ATT and seven used ATG. Of note is the use of ATG by the ATP8 gene, which was not seen among the other Metastriata for which sequences are available and found only in *Ixodes hexagonus* among the Prostriata. TAA was the most common stop codon (seven protein coding genes), followed by the two truncated forms of TAA, T* (three genes) and TA* (two genes). The stop codon TAG was observed only once; it was used in the ND1 gene. Among the 32 additional tick species for which mt genomes were available, only two (*I. holocyclus* and *Nuttalliella namaqua*) also used TAG in the ND1 gene.

We assessed the effective number of codons (EN_c) (Wright 1990) used by the four *Amblyomma s. s.* and two other *Amblyomma (s. l.)* mt genomes in 12 of the 13 protein coding genes. There was a significant effect of species on the EN_c (repeated measures ANOVA, $F(5, 55) = 5.11$, $p = 6.45 \times 10^{-4}$), and post hoc tests indicated differences between *A. triguttatum* and *A. cajennense* ($p < 0.03$), *A. elaphense* and *A. fimbriatum* ($p <$

0.02), and *A. elaphense* and *A. cajennense* ($p < 0.002$). *Amblyomma cajennense* had the highest mean EN_c at 41.07 codons while *A. elaphense* had the lowest at 33.43 codons.

To determine the relative preference of each *Amblyomma s. l.* species for each synonymous codon, I calculated the relative synonymous codon usage (RSCU) index for the concatenated protein coding genes from each species. The patterns of codon bias were similar across species (Figure 5.S6). One threonine codon (ACG) was not used by *A. americanum*, and one alanine (GCG) and one leucine 1 (CTG) codon were missing in *A. elaphense*. All three missing codons were the least preferred among the *Amblyomma s. l.* examined and were also the most G-rich of the available codons for those amino acids (Figure 5.S6). This suggests that the loss of these codons may be due to the known mutation bias against G on the mt genome's J strand (Hassanin *et al.* 2005).

To test if codon usage bias reflected the phylogenetic relationships among the *Amblyomma s. l.*, I applied the codon usage dissimilarity measurement of Suzuki *et al.* (2009) to the concatenated protein coding genes of these species. Cluster analysis on the resulting distance matrix produced a dendrogram that was a poor representation of the phylogeny of these species, but did place *A. elaphense* as the outgroup (Figure 5.3). Burger *et al.* placed *A. elaphense* as a sister group to the Metastriata (Burger *et al.* 2012, 2013), suggesting that codon usage may be a useful marker of evolutionary relationships at the scale of family or subfamily but not below.

5.5 Conclusions

I obtained the first complete mt genome sequence for *A. americanum* and found that it has a typical metastriate tick mt organization. Phylogenetic analysis clustered *A. americanum* with the other members of the *Amblyomma s. s.*, confirming previous taxonomies (e.g. Norris *et al.* 1999). Heteroplasmy rates were low across the chromosome (six sites) and more similar to those observed in other tick genera than to the unusually high heteroplasmy rate previously reported for *A. cajennense* (Xiong *et al.* 2013). *Amblyomma americanum* has lost one threonine codon; all of the other three members of *Amblyomma s. s.* for which sequence data is available retain use of all codons. This genome sequence provides an important reference for future studies of *A. americanum* population genetics and contributes to ongoing efforts to fully resolve the phylogeny of the genus *Amblyomma*.

5.6 Figures

Figure 5.1 Gene order, assembly coverage, GC content, and GC skew of the *Amblyomma americanum* mitochondrial genome. Genes labeled on the outermost circle are coded on the major (J) strand, while those denoted on the second circle are found on the minor (N) strand. Transfer RNAs are indicated in red and labeled with the corresponding amino acid letter code. The gray lines indicate the regions which were sequenced by Sanger methodology. The black circle gives the Illumina sequence coverage for each position. GC content and GC skew are expressed as deviation from the mean in the blue and red circles, respectively.

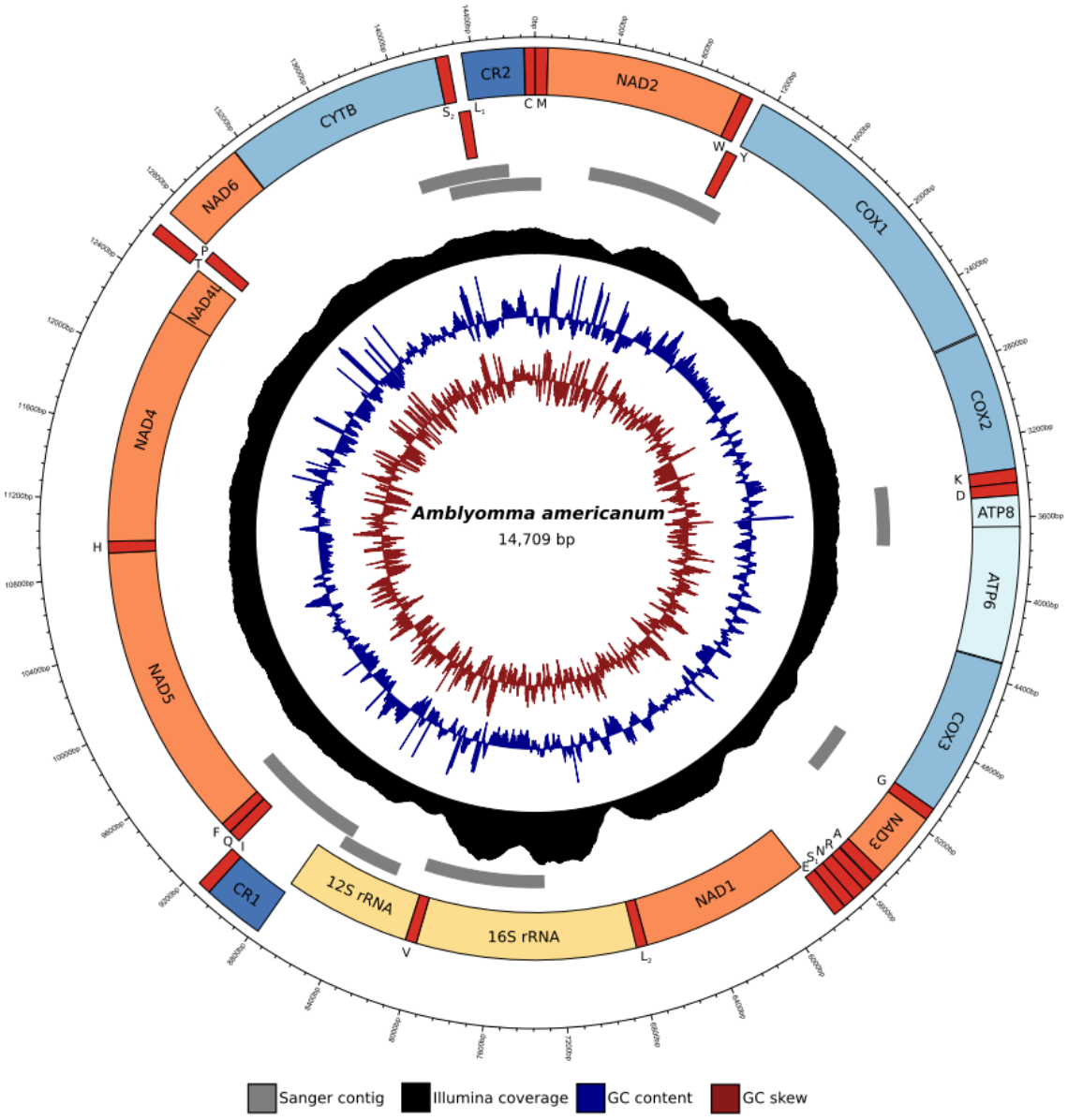

Amblyomma americanum
14,709 bp
CR2
NAD2
COX1
COX2
ATP8
ATP6
COX3
NAD3
NAD1
16S rRNA
12S rRNA
CR1
NAD5
NAD4
NAD4L
NAD6
CYTB
Sanger contig
Illumina coverage
GC content
GC skew

Figure 5.2 Mitochondrial genome maximum-likelihood tree showing phylogenetic relationships between tick species. The tree is based on concatenated alignments of rRNA and protein coding genes. Support values are percentages and were derived from 100 bootstrap replicates. GenBank accession numbers are given in parentheses after each species name. *Amblyomma americanum*, the novel mitochondrial sequence presented in this study, is bolded. Broken lines indicate that the branches are truncated; the total length of each truncated branch is given on the diagram.

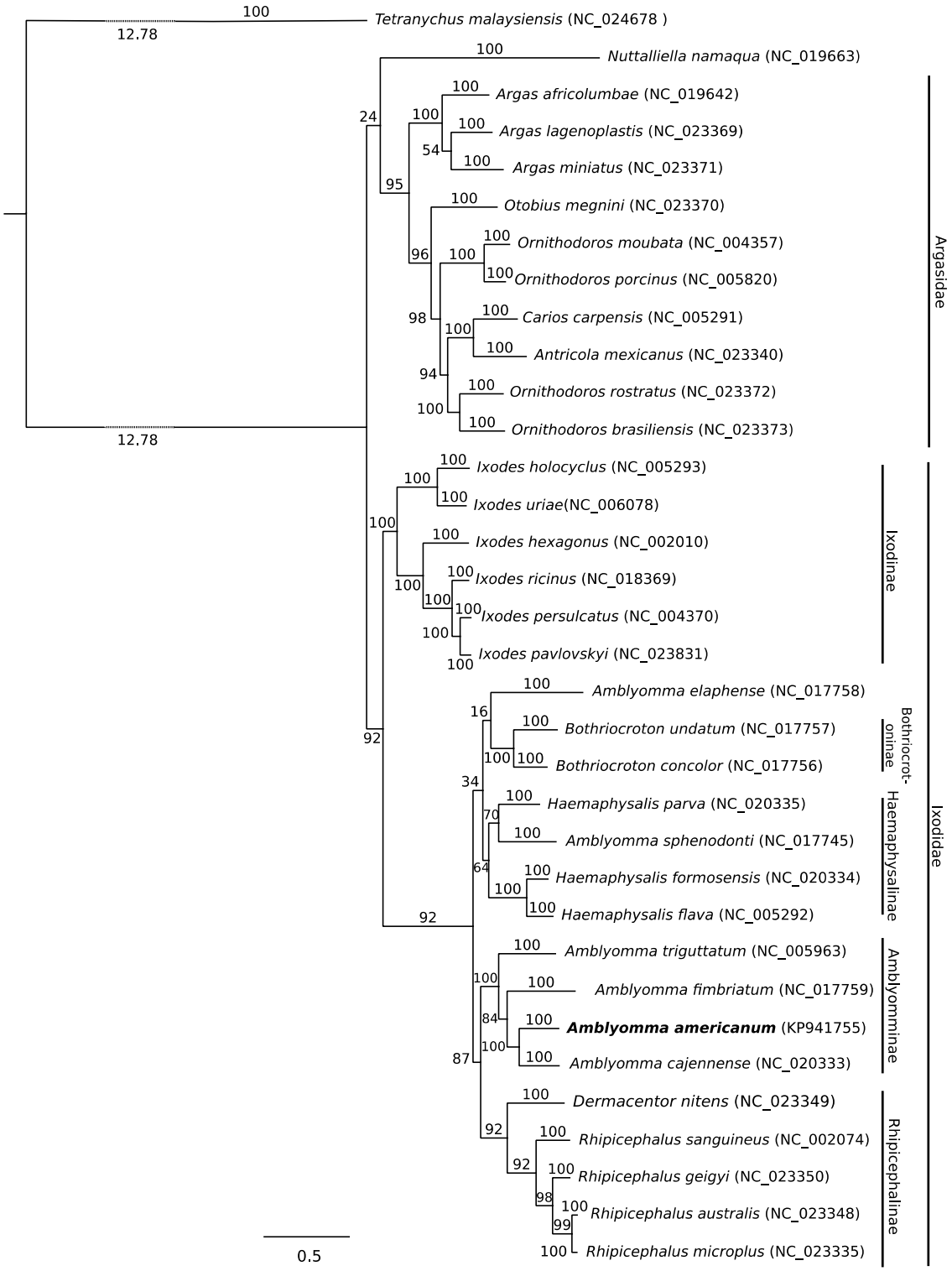

Tetranychus malaysiensis (NC_024678)
Nuttalliella namaqua (NC_019663)
Argas africolumbae (NC_019642)
Argas lagenoplastis (NC_023369)
Argas miniatus (NC_023371)
Otobius megnini (NC_023370)
Ornithodoros moubata (NC_004357)
Ornithodoros porcinus (NC_005820)
Carios carpensis (NC_005291)
Antricola mexicanus (NC_023340)
Ornithodoros rostratus (NC_023372)
Ornithodoros brasiliensis (NC_023373)
Ixodes holocyclus (NC_005293)
Ixodes uriae(NC_006078)
Ixodes hexagonus (NC_002010)
Ixodes ricinus (NC_018369)
Ixodes persulcatus (NC_004370)
Ixodes pavlovskyi (NC_023831)
Amblyomma elaphense (NC_017758)
Bothriocroton undatum (NC_017757)
Bothriocroton concolor (NC_017756)
Haemaphysalis parva (NC_020335)
Amblyomma sphenodonti (NC_017745)
Haemaphysalis formosensis (NC_020334)
Haemaphysalis flava (NC_005292)
Amblyomma triguttatum (NC_005963)
Amblyomma fimbriatum (NC_017759)
Amblyomma americanum (KP941755)
Amblyomma cajennense (NC_020333)
Dermacentor nitens (NC_023349)
Rhipicephalus sanguineus (NC_002074)
Rhipicephalus geigyi (NC_023350)
Rhipicephalus australis (NC_023348)
Rhipicephalus microplus (NC_023335)
12.78
12.78
0.5
Argasidae
Ixodidae
Ixodinae
Bothriocrot-oninae
Haemaphysalinae
Amblyomminae
Rhipicephalinae

Figure 5.3 Dendrogram representing the dissimilarity in mitochondrial codon bias between *Amblyomma sensu lato* species. The dendrogram was produced using complete linkage hierarchical clustering. Asterisks denote species which are awaiting reassignment to a different genus.

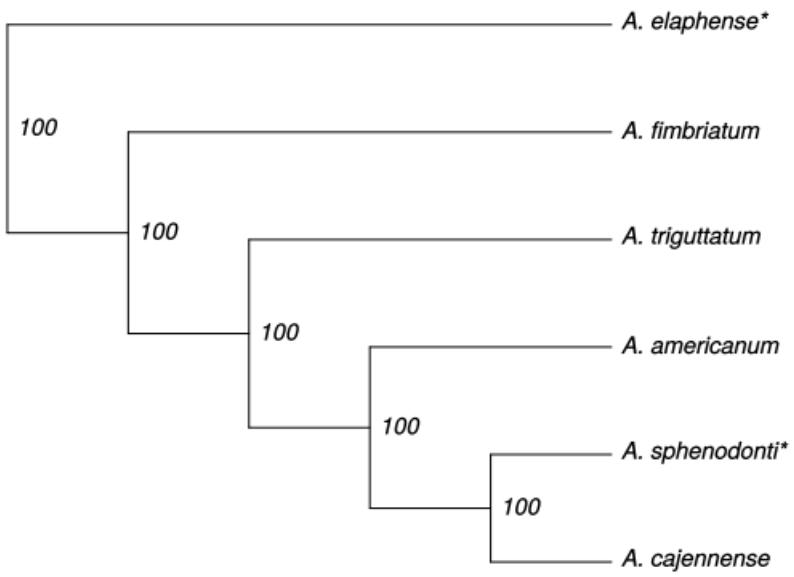


Figure 5.S1 Heteroplasmy analysis. Description and R code for the heteroplasmy analysis used in this manuscript.

Heteroplasmy analysis

This analysis was performed in Chapter 5: The mitochondrial genome of *Amblyomma americanum*. The goal of the analysis was to identify heteroplasmic sites in the mitochondrial genome of *A. americanum*. Additional details are available in the associated chapter. This analysis closely followed the methods described in Xiong *et al*. (2013).

All calculations were performed on the R statistical platform (version 3.1.2).

Calculating the false positive rate

The required R packages for this analysis were:

```
library(gmp)
library(dplyr)
library(qvalue)
```

`gmp` (version 0.5-12) provided the function `factorialZ` for calculating large factorials, `dplyr` (version 0.4.1) provided functions for data manipulation, and `qvalue` (version 1.40.0) provided the `qvalue` function for calculating false discovery rates.

To calculate the false positive rate (Pf), I wrote a function that used the formula for Pf from Storey and Tibshirani (2003).

```
fpos <- function(n, k, p) {
    stopifnot(is.numeric(c(n, k, p)))
    if (!isTRUE(all(c(n, k) == floor(c(n, k)))))
        stop("n, k must only contain integer values")
    cnk <- factorialZ(n)/(factorialZ(k) * factorialZ(n - k))
    pk <- p^k
    onep <- (1 - p)^(n - k)
    return(as.double(cnk * pk * onep))
}
```

The function `fpos` accepts three arguments: **n**, the sequence read coverage at a site; **k**, the number of times a given heteroplasmy is seen at the site; and **p**, the sequencing error rate for the platform used. For the Illumina platform, `p = 0.005`.

Next, I read in our partial data table containing the n and k values for each putative heteroplasmic site.

```
Aamer <- read.delim("/home/igy7/git-repos/falsepos/R/Table 1 SNP-data.c
sv",
    sep = ",", header = TRUE)
Aamer
```

```
##   Position Gene Reference Variant Coverage Count  Frequency        E
ffect
## 1      698  ND2         C       T     9754   740 0.07586631 I(ATC)>I
(ATT)
## 2     6598  ND1         G       T     9808   170 0.01733279 G(GGC)>G
(GGA)
## 3     6622  ND1         T       C    10014   181 0.01807470 G(GGA)>G
(GGG)
## 4     7978  16S         A       T     4527    84 0.01855533
## 5     8008  16S         T       A     5141    89 0.01731181
## 6    14067 CYTB         T       C     2203    35 0.01588743 F(TTT)>L
(CTT)
##    p
## 1 NA
## 2 NA
## 3 NA
## 4 NA
## 5 NA
## 6 NA
```

Because the P value column was blank, R filled it with NAs. I removed the entire column and replaced it with the column Pf which contains the Pf values for each site calculated using the `fpos` function.

```
Aamer$p <- NULL
Aamer <- mutate(Aamer, Pf = fpos(Coverage, Count, 0.005))
Aamer
```

```
##   Position Gene Reference Variant Coverage Count  Frequency        E
ffect
## 1      698  ND2         C       T     9754   740 0.07586631 I(ATC)>I
(ATT)
## 2     6598  ND1         G       T     9808   170 0.01733279 G(GGC)>G
(GGA)
## 3     6622  ND1         T       C    10014   181 0.01807470 G(GGA)>G
(GGG)
## 4     7978  16S         A       T     4527    84 0.01855533
## 5     8008  16S         T       A     5141    89 0.01731181
## 6    14067 CYTB         T       C     2203    35 0.01588743 F(TTT)>L
(CTT)
##             Pf
## 1 0.000000e+00
## 2 0.000000e+00
## 3 0.000000e+00
## 4 1.873370e-23
```

```
## 5 8.777128e-23
## 6 4.147151e-09
```

Calculating the false discovery rate

I calculated the false discovery rate (Qf) for each site following Storey (2002) and appended those values to the column Qf on our data table. This table appears in the manuscript as Table 5.2.

```
qobj <- qvalue(Aamer$Pf, lambda = 0, fdr.level = 0.05)
Aamer <- mutate(Aamer, Qf = qobj$qvalues)
Aamer
```

```
##   Position Gene Reference Variant Coverage Count  Frequency        E
ffect
## 1      698  ND2         C       T     9754   740 0.07586631 I(ATC)>I
(ATT)
## 2     6598  ND1         G       T     9808   170 0.01733279 G(GGC)>G
(GGA)
## 3     6622  ND1         T       C    10014   181 0.01807470 G(GGA)>G
(GGG)
## 4     7978  16S         A       T     4527    84 0.01855533
## 5     8008  16S         T       A     5141    89 0.01731181
## 6    14067 CYTB         T       C     2203    35 0.01588743 F(TTT)>L
(CTT)
##              Pf           Qf
## 1 0.000000e+00 0.000000e+00
## 2 0.000000e+00 0.000000e+00
## 3 0.000000e+00 0.000000e+00
## 4 1.873370e-23 2.810054e-23
## 5 8.777128e-23 1.053255e-22
## 6 4.147151e-09 4.147151e-09
```

Note that the normal usage of qvalue does not specify lambda and allows the software to estimate pi0. This fails with a small number of *p* values or "pathological" situations (*i.e.* clustered low *p* values). Setting lambda to 0 makes pi0 equal to 1 and is a conservative case of the Storey (2002) methodology, equivalent to the method of Benjamini and Hochberg (1995).

References

Benjamini and Hochberg (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. Series B 57, 289–300.

Storey (2002) A direct approach to false discovery rates. J. R. Stat. Soc. Series B 64: 479–498. doi: 10.1111/1467-9868.00346

Storey and Tibshirani (2003) Statistical significance for genome-wide studies. Proc. Nati. Acad. Sci. U. S. A. 100: 9440–9445. doi: 10.1073/pnas.1530509100

Xiong *et al.* (2013) Heteroplasmy in the mitochondrial genomes of human lice and ticks revealed by high throughput sequencing. PLoS ONE 8(9): e73329. doi: 10.1371/journal.pone.0073329

Figure 5.S2 Pairwise dissimilarity in synonymous codon usage. Description and R code for the calculation of whole mt genome pairwise dissimilarity in synonymous codon usage and cluster analysis used in this manuscript.

Pairwise dissimilarity in synonymous codon usage

This analysis was performed in Chapter 5: The mitochondrial genome of *Amblyomma americanum*. The goal of the analysis was determine how the mitochondrial genomes of the members of the genus *Amblyomma sensu lato* are related in terms of their codon usage. I quantified the "distance" between mitochondrial genomes using the method of Suzuki *et al*. (2009) and then performed a cluster analysis on the resulting distance matrix.

All calculations were performed on the R statistical platform (version 3.1.2).

Calculating the ratio xac

The required R packages for this analysis were:

```
library(dplyr)
library(reshape2)
library(ape)
library(combinat)
library(phangorn)
```

`dplyr` (version 0.4.1) and `reshape2` (version 1.4.1) were used for data manipulation, while `ape` (version 3.2) and `phangorn` (version 1.99-12) were used for dendrogram plotting. `combinat` (version 0.0-8) was used to generate all possible orderings of tick names.

First, I imported a table containing the number of times each codon was observed in the concatenated mitochondrial protein coding genes from six *Amblyomma s. l.* species. Because many stop codons were truncated in the original sequence and had to be manually completed to maintain the correct reading frame in the concatenated sequences, the stop codons (abbreviated TER in the table) were removed from the analysis.

```
alldata <- read.delim("seqfile.fasta.bulk.txt", sep = ",", header = TRU
E)
alldata <- arrange(alldata, Species, AA, Observed)
alldata <- filter(alldata, AA != "TER")
alldata <- droplevels(alldata)
```

The table was then formatted as below (first 6 rows shown):

```
##   Species  AA Codon Observed RSCU
## 1   Aamer Ala   GCG        3 0.11
## 2   Aamer Ala   GCC       15 0.57
## 3   Aamer Ala   GCA       41 1.55
## 4   Aamer Ala   GCT       47 1.77
## 5   Aamer Arg   CGC        1 0.10
## 6   Aamer Arg   CGG        6 0.57
```

Next, I extracted the data for each tick species into its own data frame using the function `mng`.

```
mng <- function(df, specs) {
    mydf <- filter(df, Species == specs)
    mydf <- select(mydf, AA, Observed)
    return(mydf)
}

Aamer <- mng(alldata, "Aamer")
Acaj <- mng(alldata, "NC_20333")
Aelap <- mng(alldata, "NC_17758")
Asphen <- mng(alldata, "NC_17745")
Atrig <- mng(alldata, "NC_5963")
Afim <- mng(alldata, "NC_17759")
```

To calculate the ratio xac from Suzuki *et al.* (2009), I applied the function `xac` to each tick species's data frame. This created a vector of xac values for each tick species (one value per codon).

```
xac <- function(df) {
    x <- c()
    for (aa in levels(df$AA)) {
        curraa <- filter(df, AA == aa)
        for (cdn in curraa$Observed) {
            x <- c(x, cdn/max(curraa$Observed))
        }
    }
    return(x)
}

Aamer_xac <- xac(Aamer)
Acaj_xac <- xac(Acaj)
Aelap_xac <- xac(Aelap)
Asphen_xac <- xac(Asphen)
Atrig_xac <- xac(Atrig)
Afim_xac <- xac(Afim)

# Example vector of xac values
Aamer_xac
```

```
##  [1] 0.06382979 0.31914894 0.87234043 1.00000000 0.04347826 0.260869
57
##  [7] 0.52173913 1.00000000 0.23684211 1.00000000 0.36363636 1.000000
00
## [13] 0.47826087 1.00000000 0.06250000 1.00000000 0.13513514 1.000000
00
## [19] 0.19277108 0.30120482 0.50602410 1.00000000 0.24000000 1.000000
00
## [25] 0.06605923 1.00000000 0.10588235 0.10588235 0.45882353 1.000000
00
## [31] 0.12307692 1.00000000 0.08661417 1.00000000 0.14396887 1.000000
00
## [37] 0.12359551 1.00000000 0.05357143 0.35714286 0.76785714 1.000000
00
## [43] 0.04109589 0.06849315 0.45205479 1.00000000 0.03937008 0.125984
25
## [49] 0.77165354 1.00000000 0.00000000 0.16883117 0.85714286 1.000000
00
## [55] 0.22388060 1.00000000 0.34782609 1.00000000 0.10294118 0.220588
24
## [61] 0.88235294 1.00000000
```

Calculating the pairwise dissimilarity between species

To calculate the pairwise dissimilarity between each species pair (`Dij` in Suzuki *et al.* (2009)), I first set up four vectors. `xac_vec` to contain the names of each of the variables holding the xac values for each species, `Species1` to hold the name of the xac vector for the first species in each pair, `Species2` to hold the name of the xac vector for the second species in each pair, and `Dissim` to hold one minus the Pearson correlation coefficient for each pair.

```
xac_vec <- c("Aamer_xac", "Acaj_xac", "Aelap_xac", "Asphen_xac", "Atrig
_xac",
    "Afim_xac")
Species1 <- c()
Species2 <- c()
Dissim <- c()
```

Next, I used a set of nested for loops to iterate over the vector of xac variable names. This created all possible species pairs and calculated their pairwise dissimilarities.

```
for (m in xac_vec) {
    for (n in xac_vec) {
        Species1 <- c(Species1, m)
        Species2 <- c(Species2, n)
        Dissim <- c(Dissim, (1 - as.numeric(cor.test(get(m), get(n), me
thod = "pearson")$estimate)))
    }
}
```

I placed all of these values in a table to view them.

```
dist_tab <- data.frame(Species1, Species2, Dissim)
dcast(dist_tab, Species1 ~ Species2)

## Using Dissim as value column: use value.var to override.

##     Species1  Aamer_xac   Acaj_xac  Aelap_xac   Afim_xac   Asphen_xa
c
## 1  Aamer_xac 0.00000000 0.01580036 0.02530334 0.02383185 1.722315e-0
2
## 2   Acaj_xac 0.01580036 0.00000000 0.03109592 0.04108010 1.035313e-0
2
## 3  Aelap_xac 0.02530334 0.03109592 0.00000000 0.04531353 2.936396e-0
2
## 4   Afim_xac 0.02383185 0.04108010 0.04531353 0.00000000 3.919674e-0
2
## 5 Asphen_xac 0.01722315 0.01035313 0.02936396 0.03919674 1.110223e-1
6
## 6  Atrig_xac 0.02563353 0.01667779 0.04187483 0.03554769 2.243910e-0
2
##    Atrig_xac
## 1 0.02563353
## 2 0.01667779
## 3 0.04187483
## 4 0.03554769
## 5 0.02243910
## 6 0.00000000
```

Everything appeared as it should be in this square distance matrix except for the *A. sphenodonti* x *A. sphenodonti* comparison. Since I was comparing the tick to itself, the value should be zero, but there was a floating point rounding error causing a very small, non-zero value. I verified that this was not the result of a programming error by running the `cor.test` function on the *A. sphenodonti* x *A. sphenodonti* comparison directly and verifying that the same value was returned.

```
1 - as.numeric(cor.test(get("Asphen_xac"), get("Asphen_xac"), method =
"pearson")$estimate)

## [1] 1.110223e-16
```

I manually corrected this error by setting the *A. sphenodonti* x *A. sphenodonti* comparison to 0. For more information about floating point rounding error, see Why Don't My Numbers Add Up? for a plain language explanation or What Every Programmer Should Know About Floating-Point Arithmetic for a technical description of the problem and its ubiquity in computer mathematics. I also corrected the row and column labels.

```
dist_mat <- matrix(Dissim, nrow = 6)
dist_mat[4, 4] <- 0
dimnames(dist_mat) <- list(c("A. americanum", "A. cajennense", "A. elap
```

```
hense*",
    "A. sphenodonti*", "A. triguttatum", "A. fimbriatum"), c("A. americ
anum",
    "A. cajennense", "A. elaphense*", "A. sphenodonti*", "A. triguttatu
m", "A. fimbriatum"))
dist_mat
```

```
##                 A. americanum A. cajennense A. elaphense* A. sphenod
onti*
## A. americanum      0.00000000    0.01580036    0.02530334      0.017
22315
## A. cajennense      0.01580036    0.00000000    0.03109592      0.010
35313
## A. elaphense*      0.02530334    0.03109592    0.00000000      0.029
36396
## A. sphenodonti*    0.01722315    0.01035313    0.02936396      0.000
00000
## A. triguttatum     0.02563353    0.01667779    0.04187483      0.022
43910
## A. fimbriatum      0.02383185    0.04108010    0.04531353      0.039
19674
##                 A. triguttatum A. fimbriatum
## A. americanum       0.02563353    0.02383185
## A. cajennense       0.01667779    0.04108010
## A. elaphense*       0.04187483    0.04531353
## A. sphenodonti*     0.02243910    0.03919674
## A. triguttatum      0.00000000    0.03554769
## A. fimbriatum       0.03554769    0.00000000
```

Dendrogram creation by hierarchical clustering

I used complete linkage hierarchical cluster analysis of the pairwise distance matrix to visualize the relationships between ticks in codon usage bias. However, cluster methods frequently produce different dendrograms depending on the order of the distance matrix. To account for order bias, I systematically permuted the rows and columns in the distance matrix and repeated the cluster analysis for each matrix permutation. A majority rule consensus tree was derived from the collection of trees.

First, I created two functions. `matperm` accepts an input matrix and a vector of possible row and column orders and outputs a list of permuted matrices. `list_trees` accepts a list of matrices and calculates a complete linkage dendrogram for each one. The output is a list of trees.

```
matperm <- function(m, ord) {
    mats <- list(NULL)
    for (num in 1:length(ord)) {
        mats[[num]] <- m[ord[[num]], ord[[num]]]
    }
    return(mats)
}

list_trees <- function(listm) {
    trees <- list(NULL)
    for (num in 1:length(listm)) {
        trees[[num]] <- as.phylo(hclust(as.dist(listm[[num]]), method =
"complete"))
    }
    class(trees) <- "multiPhylo"
    return(trees)
}
```

Next, I created a vector containing all of the possible orderings of the row and column names for the matrix using the function `permn`. I then passed the vector of row and column orders and the distance matrix to `matperm` to perform the distance matrix permutations. The output of `matperm` was passed to `list_trees` to create a dendrogram from each matrix permutation.

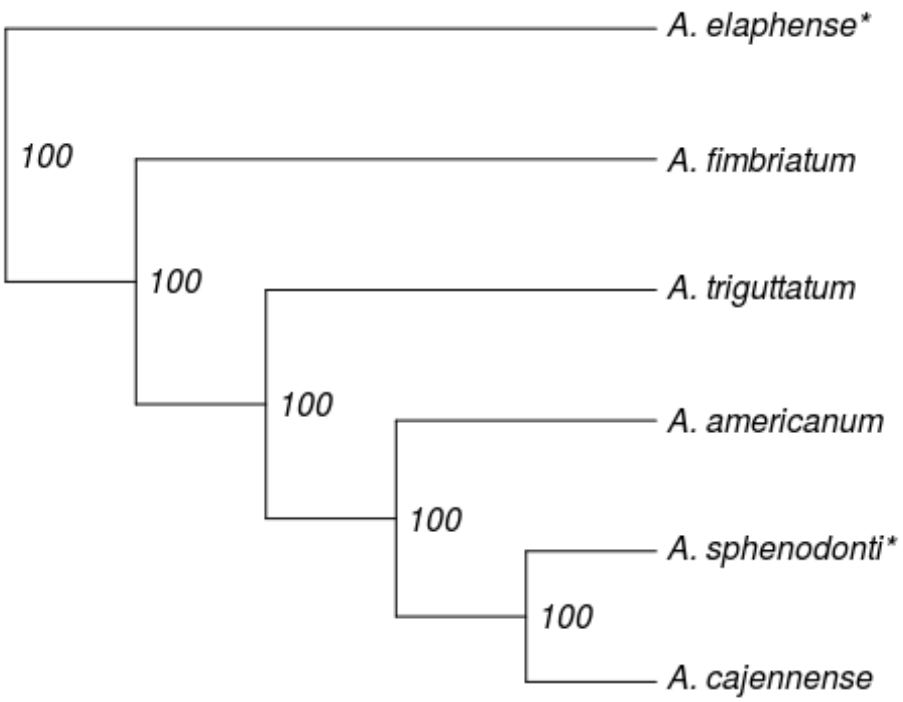
```
orders <- permn(row.names(dist_mat), fun = NULL)
mymtrx_perms <- matperm(dist_mat, orders)
mymtrx_trees <- list_trees(mymtrx_perms)
```

Finally, I used `consensus` to create the majority rule tree from the `list_trees` output. The function `plotBS` was used to determine node support values for the consensus tree using the trees derived from the permuted distance matrices.

```
mymtrx_cons <- consensus(mymtrx_trees, p = 0.5)
mymtrx_BS <- plotBS(mymtrx_cons, mymtrx_trees, type = "phylogram")
```

The final `plot` command generated Figure 5.3 in the chapter. Recall that these are not "real" bootstrap support values, but rather represent the percentage of dendrograms derived from matrix permutations that contain each node.

```
plot(mymtrx_BS, label.offset = 0.09, y.lim = c(0.5, 6), show.node.label
= TRUE,
    no.margin = TRUE)
```



References

Suzuki *et al.* (2009) Measure of synonymous codon usage diversity among genes in bacteria. BMC Bioinformatics 10: 167. doi:10.1186/1471-2105-10-167.

Figure 5.S3 *Amblyomma americanum* mitochondrial tRNA folding patterns predicted by the MITOS webserver (mitos.bioinf.uni-leipzig.de/index.py). Anticodons are given in parentheses beneath each tRNA name.

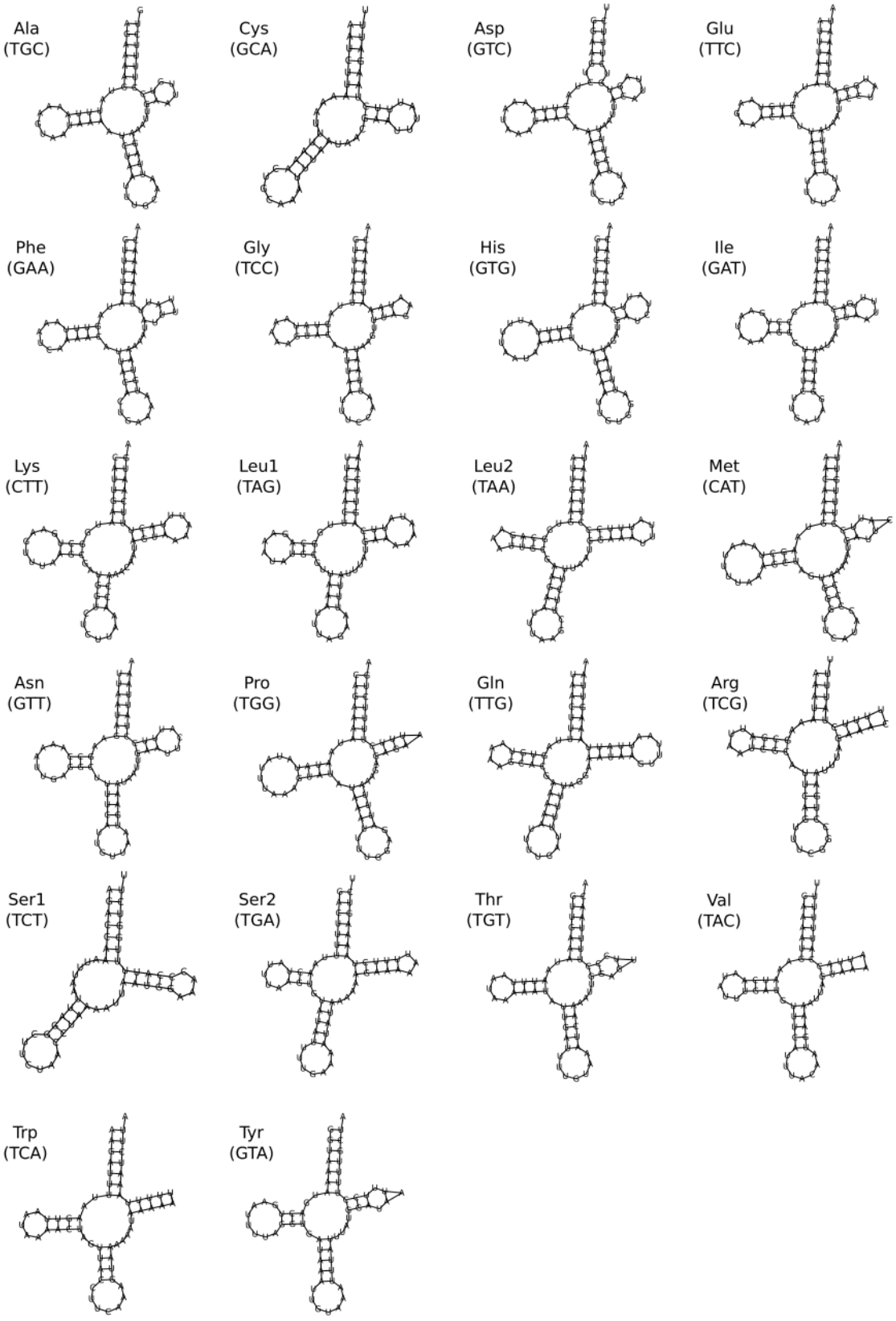
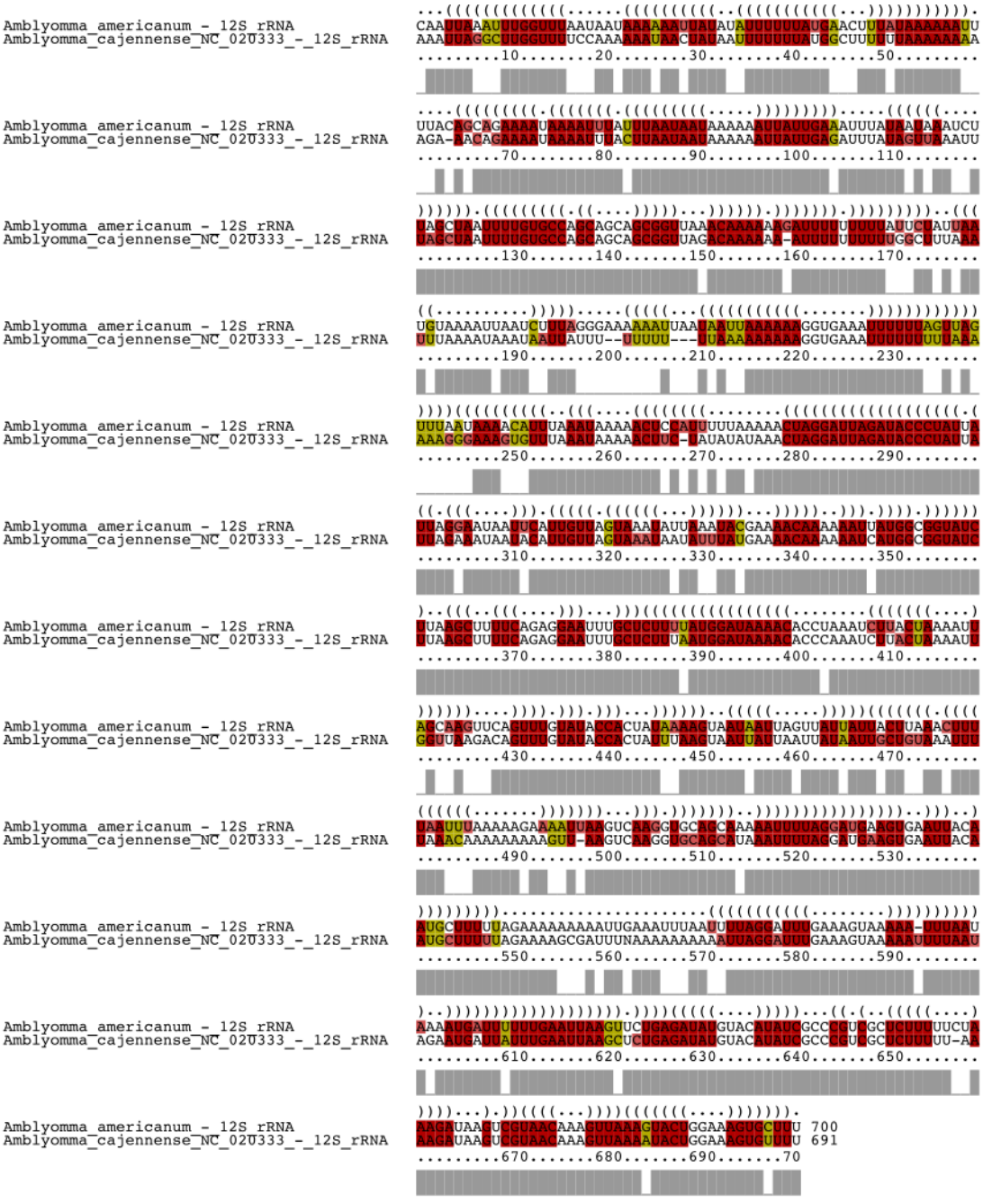
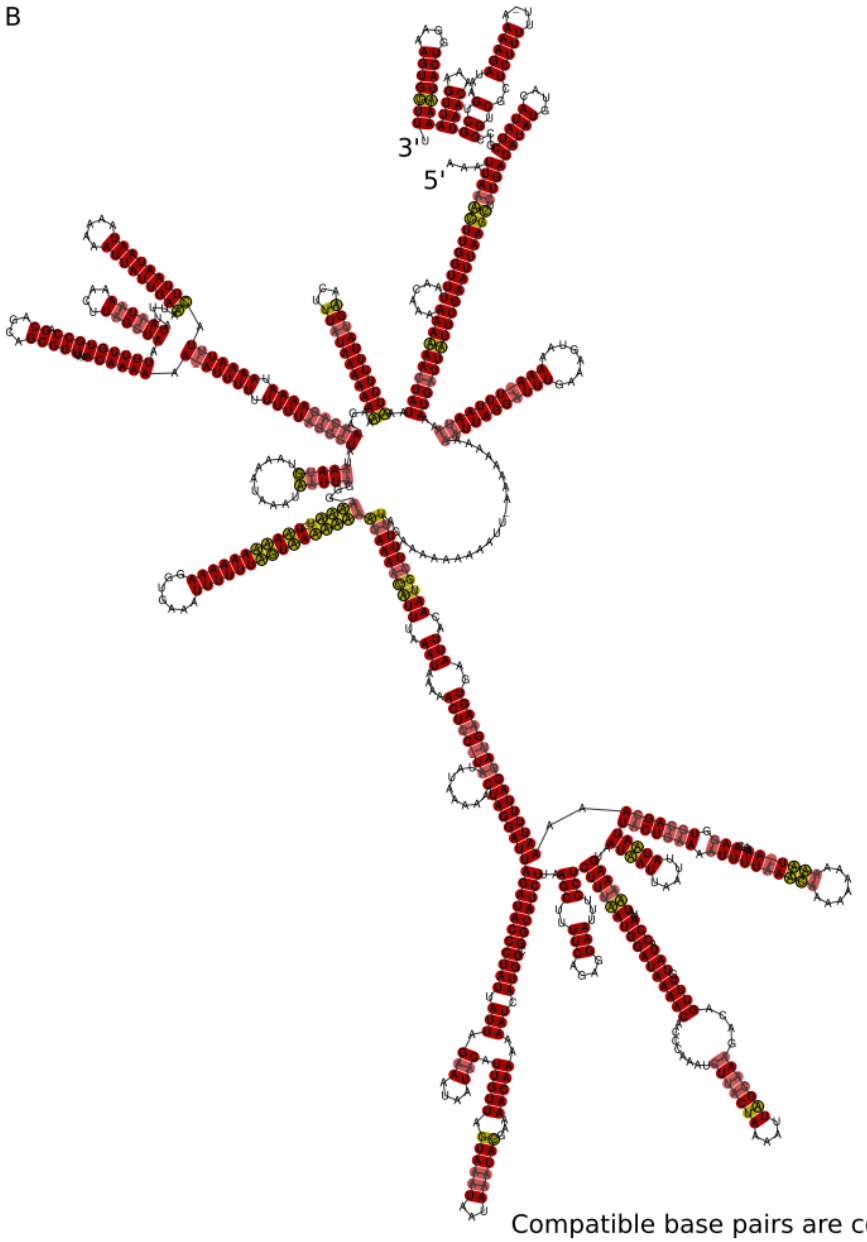

Ala
(TGC)
Cys
(GCA)
Asp
(GTC)
Glu
(TTC)
Phe
(GAA)
Gly
(TCC)
His
(GTG)
Ile
(GAT)
Lys
(CTT)
Leu1
(TAG)
Leu2
(TAA)
Met
(CAT)
Asn
(GTT)
Pro
(TGG)
Gln
(TTG)
Arg
(TCG)
Ser1
(TCT)
Ser2
(TGA)
Thr
(TGT)
Val
(TAC)
Trp
(TCA)
Tyr
(GTA)

Figure 5.S4 Mitochondrial 12S rRNA alignment and consensus folding structure for *Amblyomma americanum* and *A. cajennense*. (A) Structure-based alignment from the LocARNA website (rna.informatik.uni-freiburg.de/LocARNA). Corresponding bracket pairs indicate base pairing. (B) Consensus secondary structure for the *A. americanum* and *A. cajennense* 12S rRNA. The 5' and 3' ends of the molecule are indicated on the diagram. The color legend applies to both figures and is given at the bottom as it appears on the LocARNA website.

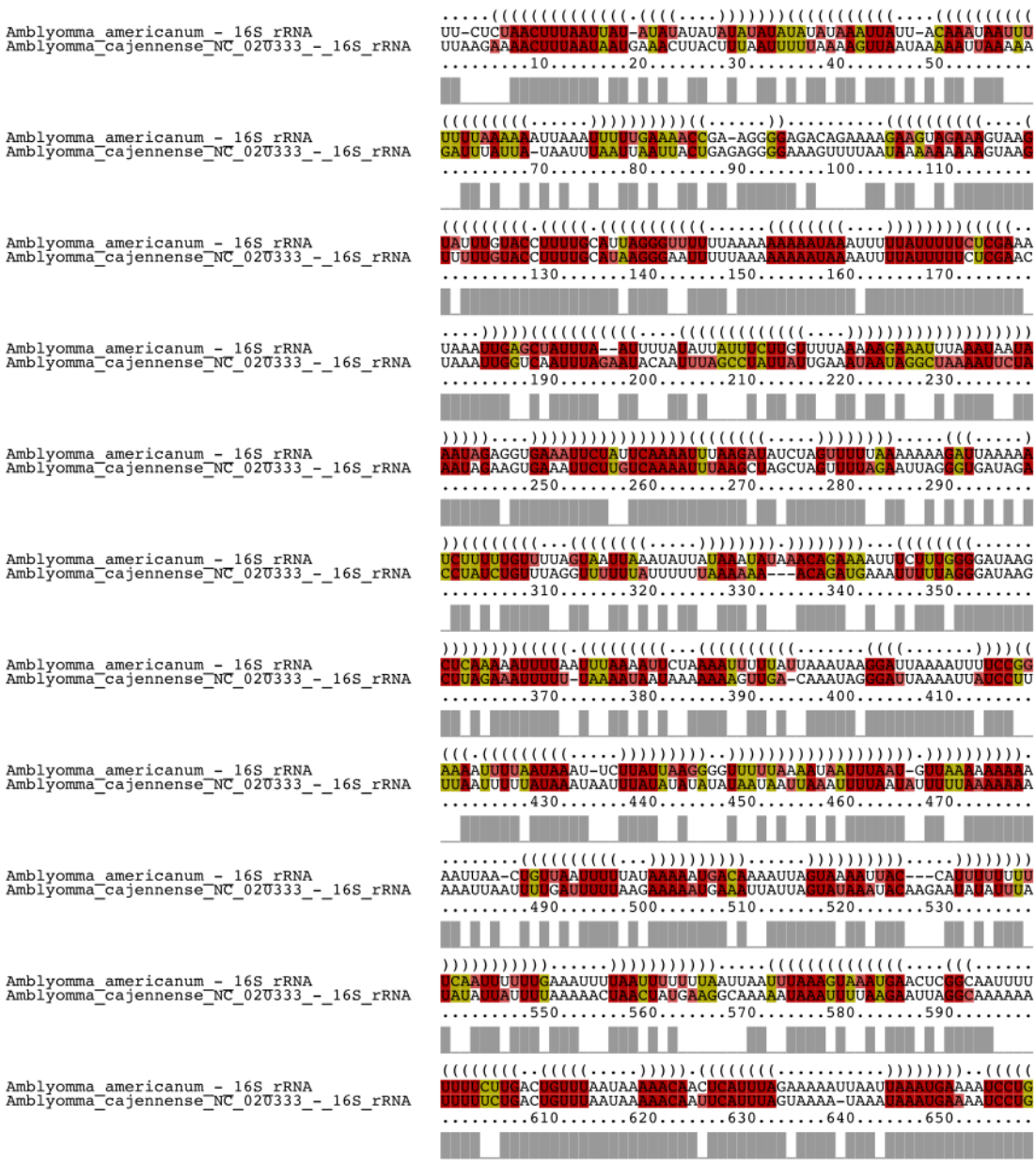

Amblyomma_americanum_-_12S_rRNA
Amblyomma_cajennense_NC_020333_-_12S_rRNA
A

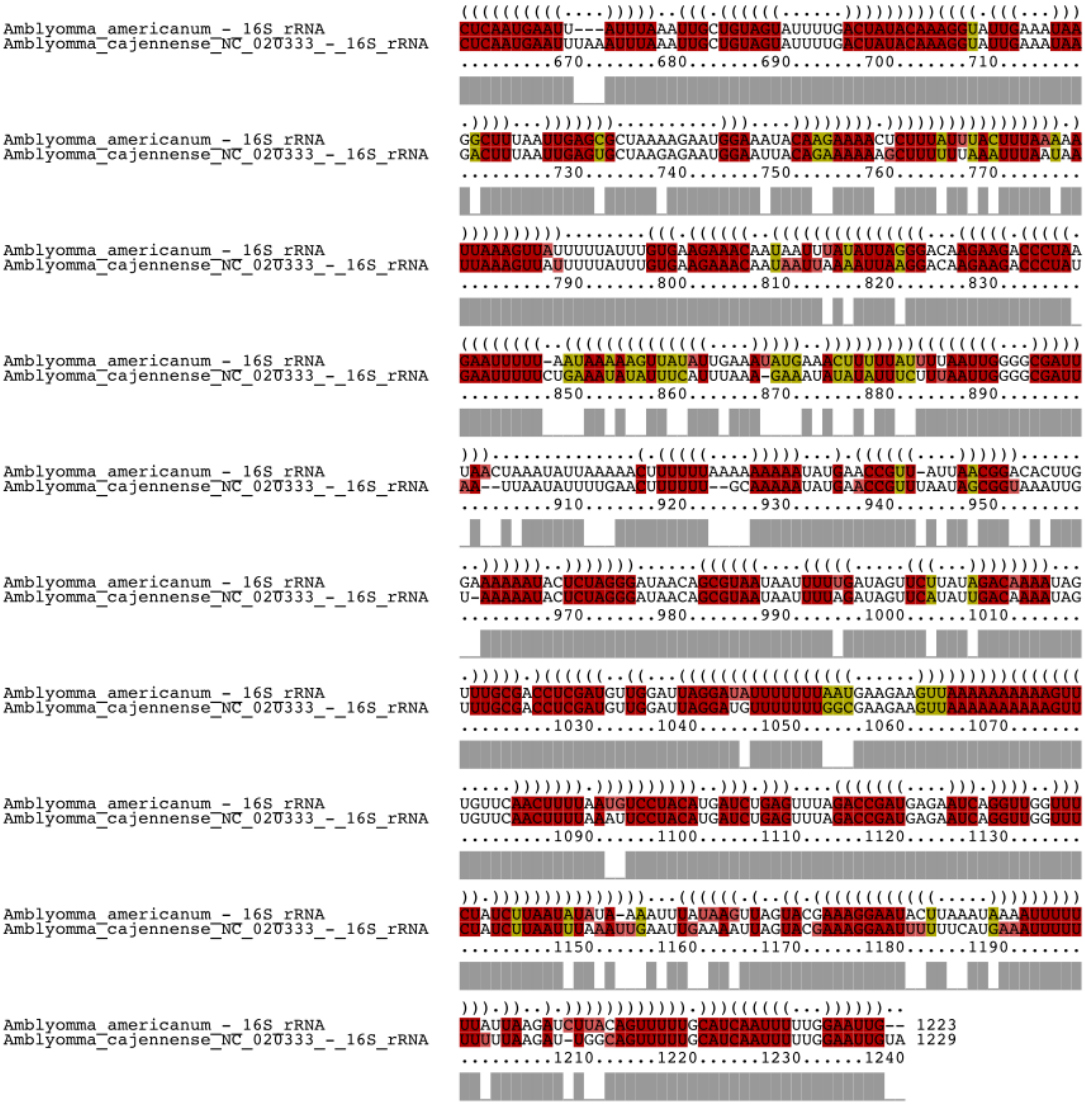


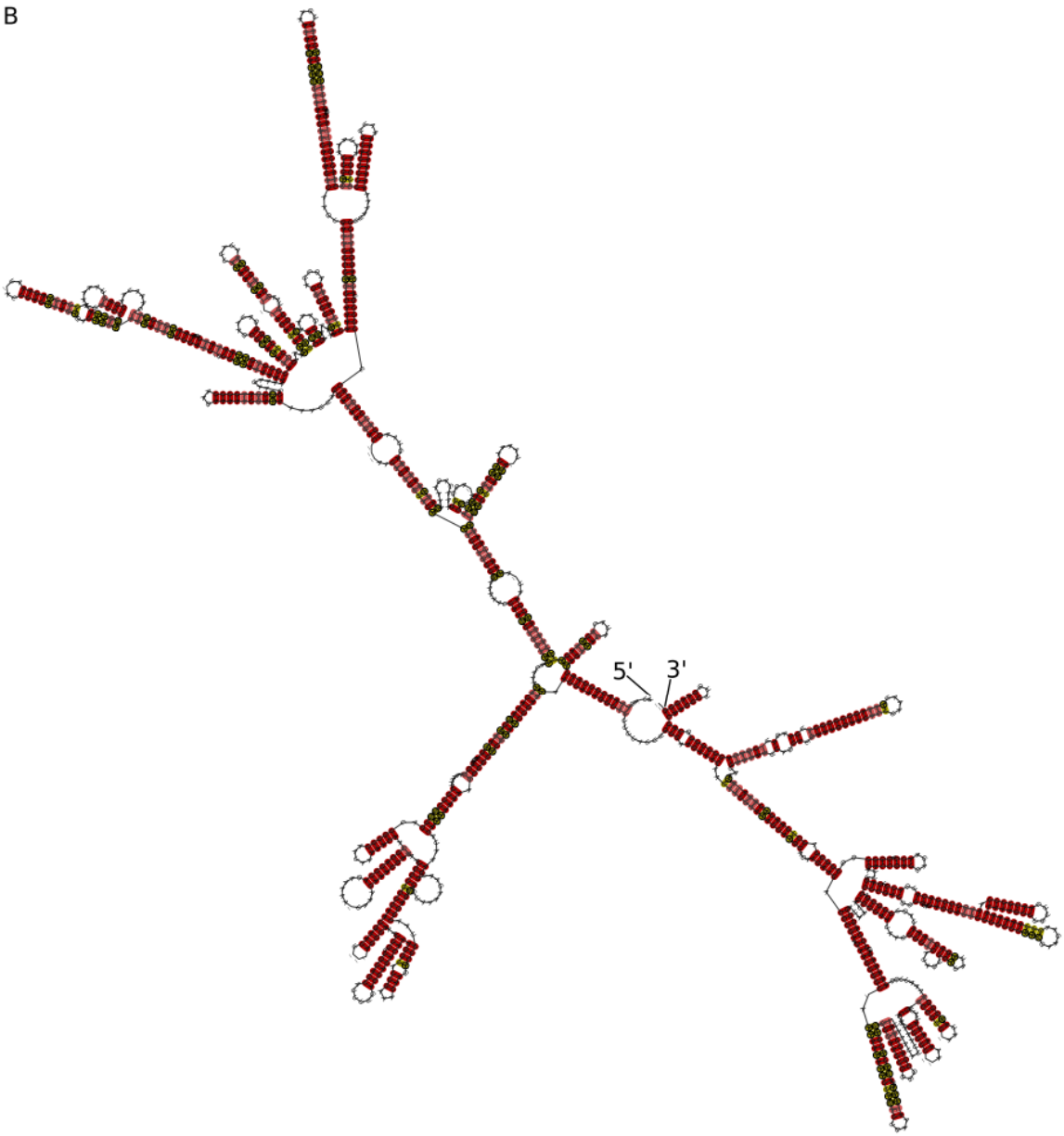
Compatible base pairs are colored, where the hue shows the number of different types C-G, G-C, A-U, U-A, G-U or U-G of compatible base pairs in the corresponding columns. In this way the hue shows sequence conservation of the base pair. The saturation decreases with the number of incompatible base pairs. Thus, it indicates the structural conservation of the base pair. Excerpted from http://rna.informatik.uni-freiburg.de/Help.jsp?toolName=LocARNA#output.

Figure 5.S5 Mitochondrial 16S rRNA alignment and consensus folding structure for *Amblyomma americanum* and *A. cajennense*. (A) Structure-based alignment from the LocARNA website (rna.informatik.uni-freiburg.de/LocARNA). Corresponding bracket pairs indicate base pairing. (B) Consensus secondary structure for the *A. americanum* and *A. cajennense* 16S rRNA. The 5' and 3' ends of the molecule are indicated on the diagram. The color legend applies to both figures and is given at the bottom as it appears on the LocARNA website.

A

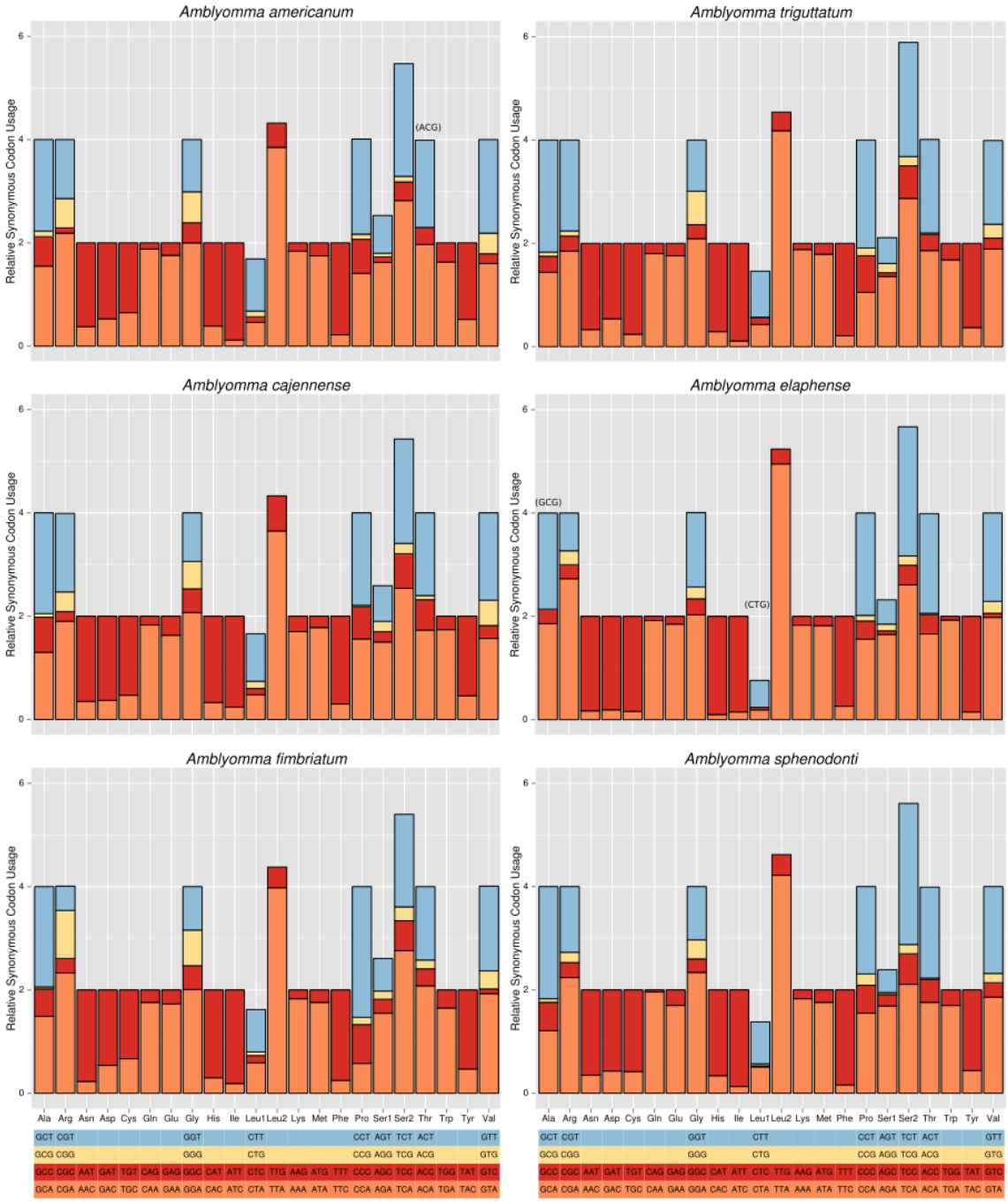

Amblyomma_americanum_-_16S_rRNA
Amblyomma_cajennense_NC_020333_-_16S_rRNA


Amblyomma_americanum_-_16S_rRNA
Amblyomma_cajennense_NC_020333_-_16S_rRNA



Compatible base pairs are colored, where the hue shows the number of different types C-G, G-C, A-U, U-A, G-U or U-G of compatible base pairs in the corresponding columns. In this way the hue shows sequence conservation of the base pair. The saturation decreases with the number of incompatible base pairs. Thus, it indicates the structural conservation of the base pair. Excerpted from http://rna.informatik.uni-freiburg.de/Help.jsp?toolName=LocARNA#output.

Figure 5.S6 Relative synonymous codon usage (RSCU) values for the mitochondrial protein coding genes of the *Amblyomma*. Codon families are given on the x-axis; absent codons are given at the top of each bar.



5.7 Tables

Table 5.1 Sanger sequencing primers used to confirm the complete *Amblyomma americanum* mitochondrial genome sequence assembly. If the primer pairs do not amplify the entire gene, the targeted region is indicated after the gene name.

Target	Annealing Temperature (°C)	No. PCR Cycles	Primer Name	Primer Sequence
CR1	65	35	AamerMT_8623F	TTT GTT TAA CCG CTG CTG CTG GC
			AamerMT_9395R	CCC CAA GGC CTT GAA AGA AGA T
CR2	60	35	AamerMT_14141F	ACC CAT CGA ATA CCC CTA TGA
			AamerMT_40R	GGT ATG AAC CCA CTA GCT TT
CYTB-3'*	65	35	AamerMT_13961F	ACG TTC AAT CCC TAA CAA ACT AGG TGG
			AamerMT_14538R	AGG AAT TTC AAC TTG GCA GAA ATA TGC
ND2-3'*	65	35	AamerMT_356F	TCC TTT CCA TTT TTG ACT TAC CTC TCT
			AamerMT_1231R	CCC GCT ATT CCG GCT CAC CTT C
12S-5'	55	40	AamerMT_8259F	AAT TTT TGC TGC ACC TTG ACT T
			AamerMT_8657R	AGC TAA TTT TGT GCC AGC AGC
16S-5'	55	40	AamerMT_7295F	GCG CTC AAT TAA AGC CTT ATT TCA A
			AamerMT_8093R	ACT GGA AAG TGC TTT AGA AAA TGA
ATP8	55	40	AamerMT_3375F	GGC TGA AGT TTA AGC GAT GGT C
			AamerMT_3761R	AGT TGA TCA GAA AGG CAA AGG A
ND3-5'	55	40	AamerMT_5014F	GTT TGA GGC CTC TGC TTG GT
			AamerMT_5309R	TTC AAA GAA AAG GGC ACA CGA G

* Primer pairs used to confirm Illumina contig joins

Table 5.2 Heteroplasmic sites and their effect in *Amblyomma americanum*.

Consensus Position	Gene	Ref. Nuc.	Variant Nuc.	Assembly Coverage	Variant Count	Variant Freq.	Variant Effect*	P_f	Q_f
698	ND2	C	T	9,754	740	0.076	I (ATC) > I (ATT)	< 1.00E-6	< 1.00E-6
6598	ND1	G	T	9,808	170	0.017	G (GGC) > G (GGA)	< 1.00E-6	< 1.00E-6
6622	ND1	T	C	10,014	181	0.018	G (GGA) > G (GGG)	< 1.00E-6	< 1.00E-6
7978	16S	A	T	4,527	84	0.019		1.87E-23	2.81E-23
8008	16S	T	A	5,141	89	0.017		8.78E-23	1.05E-22
14067	CYTB	T	C	2,203	35	0.016	F (TTT) > L (CTT)	4.15E-09	4.15E-09

* Consensus amino acid (consensus codon) > variant amino acid (variant codon). Abbreviations: Ref. = Reference, Nucleo. = Nucleotide, Freq. = Frequency.

5.8 References

Araya-Anchetta A, Busch JD, Scoles GA, Wagner DM (2015) Thirty years of tick population genetics: a comprehensive review. *Infection, Genetics and Evolution*, **29**, 164–179.

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, **57**, 289–300.

Bernt M, Donath A, Jühling F, Externbrink F, Florentz C, Fritzsch G, Pütz J, Middendorf M, Stadler PF (2013) MITOS: Improved *de novo* metazoan mitochondrial genome annotation. *Molecular Phylogenetics and Evolution*, **69**, 313–319.

Berrada ZL, Goethert HK, Cunningham J, Telford SR (2011) *Rickettsia rickettsii* (Rickettsiales: Rickettsiaceae) in *Amblyomma americanum* (Acari: Ixodidae) from Kansas. *Journal of Medical Entomology*, **48**, 461–467.

Biomatters (2015) Geneious. *Geneious*. Available at: www.geneious.com.

Black WC, Roehrdanz RL (1998) Mitochondrial gene order is not conserved in arthropods: prostriate and metastriate tick mitochondrial genomes. *Molecular Biology and Evolution*, **15**, 1772–1785.

Buller RS, Arens M, Hmiel SP, Paddock CD, Sumner JW, Rikihisa Y, Unver A, Gaudreault-Keener M, Manian FA, Liddell AM, Schmulewitz N, Storch GA (1999) *Ehrlichia ewingii*, a newly recognized agent of human ehrlichiosis. *New England Journal of Medicine*, **341**, 148–155.

Burger TD, Shao R, Barker SC (2013) Phylogenetic analysis of the mitochondrial genomes and nuclear rRNA genes of ticks reveals a deep phylogenetic structure

within the genus *Haemaphysalis* and further elucidates the polyphyly of the genus *Amblyomma* with respect to *Amblyomma sphenodonti* and *Amblyomma elaphense*. *Ticks and Tick-borne Diseases*, **4**, 265–274.

Burger TD, Shao R, Barker SC (2014) Phylogenetic analysis of mitochondrial genome sequences indicates that the cattle tick, *Rhipicephalus* (*Boophilus*) *microplus*, contains a cryptic species. *Molecular Phylogenetics and Evolution*, **76**, 241–253.

Burger TD, Shao R, Beati L, Miller H, Barker SC (2012) Phylogenetic analysis of ticks (Acari: Ixodida) using mitochondrial genomes and nuclear rRNA genes indicates that the genus *Amblyomma* is polyphyletic. *Molecular Phylogenetics and Evolution*, **64**, 45–55.

Childs JE, Paddock CD (2003) The ascendancy of *Amblyomma americanum* as a vector of pathogens affecting humans in the United States. *Annual Review of Entomology*, **48**, 307–337.

Clay K, Fuqua C (2010) *The tick microbiome: diversity, distribution and influence of the internal microbial community for a blood-feeding disease vector*. Institute of Medicine, Washington, D. C.

CLCbio (2014) *CLCbio, a QIAGEN Company*. Available at: www.clcbio.com.

Cortinas R, Spomer S (2013) Lone star tick (Acari: Ixodidae) occurrence in Nebraska: historical and current perspectives. *Journal of Medical Entomology*, **50**, 244–251.

Cortinas R, Spomer SM (2014) Occurrence and county-level distribution of ticks (Acari: Ixodoidea) in Nebraska using passive surveillance. *Journal of Medical Entomology*, **51**, 352–359.

Dabney A, Storey JD (2015) *qvalue: Q-value estimation for false discovery rate control.*

FastQC (2014) *Babraham Bioinformatics*. Available at: www.bioinformatics.bbsrc.ac.uk/projects/fastqc.

Freiburg RNA Tools (2015) LocARNA - Alignment and Folding. Available at: rna.informatik.uni-freiburg.de/LocARNA/.

Guglielmone AA, Robbins RG, Apanaskevich DA, Petney TN, Estrada-Pena A, Horak IG (2014) *The Hard Ticks of the World (Acari: Ixodida: Ixodidae)*. Springer, London.

Hassanin A, Léger N, Deutsch J (2005) Evidence for multiple reversals of asymmetric mutational constraints during the evolution of the mitochondrial genome of Metazoa, and consequences for phylogenetic inferences. *Systematic Biology*, **54**, 277–298.

Holm S (1979) A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, **6**, 65–70.

Institute of Medicine (2011) *Critical Needs and Gaps in Understanding Prevention, Amelioration, and Resolution of Lyme and Other Tick-Borne Diseases: The Short-Term and Long-Term Outcomes: Workshop Report*. The National Academies Press, Washington, D. C.

Institut Pasteur IT Center, Ressource Parisienne en Bioinformatique Structurale (2013) Mobyle 1.5. *Mobyle @Pasteur*. Available at: mobyle.pasteur.fr.

Jongejan F, Uilenberg G (2004) The global importance of ticks. *Parasitology*, **129**, S3–S14.

Jühling F, Mörl M, Hartmann RK, Sprinzl M, Stadler PF, Pütz J (2009) tRNAdb 2009: compilation of tRNA sequences and tRNA genes. *Nucleic Acids Research*, **37**, D159–D162.

Keirans JE, Litwak TR (1989) Pictorial key to the adults of hard ticks, family Ixodidae (Ixodida: Ixodoidea), east of the Mississippi River. *Journal of Medical Entomology*, **26**, 435–448.

Kmiec B, Woloszynska M, Janska H (2006) Heteroplasmy as a common state of mitochondrial genetic information in plants and animals. *Current Genetics*, **50**, 149–159.

Kolonin GV (2009) *Fauna of the Ixodid Ticks of the World (Acari, Ixodidae)*. Moscow. Available at: www.kolonin.org/1.html.

Koressaar T, Remm M (2007) Enhancements and modifications of primer design program Primer3. *Bioinformatics*, **23**, 1289–1291.

Leeuwen TV, Vanholme B, Pottelberge SV, Nieuwenhuyse PV, Nauen R, Tirry L, Denholm I (2008) Mitochondrial heteroplasmy and the evolution of insecticide resistance: non-Mendelian inheritance in action. *Proceedings of the National Academy of Sciences*, **105**, 5980–5985.

Loftis AD, Reeves WK, Spurlock JP, Mahan SM, Troughton DR, Dasch GA, Levin ML (2006) Infection of a goat with a tick-transmitted *Ehrlichia* from Georgia, U.S.A., that is closely related to *Ehrlichia ruminantium*. *Journal of Vector Ecology*, **31**, 213–223.

Magnacca KN, Brown MJF (2010) Tissue segregation of mitochondrial haplotypes in heteroplasmic Hawaiian bees: implications for DNA barcoding. *Molecular Ecology Resources*, **10**, 60–68.

Masta SE, Boore JL (2008) Parallel evolution of truncated transfer RNA genes in arachnid mitochondrial genomes. *Molecular Biology and Evolution*, **25**, 949–959.

McCoy KD, Léger E, Dietrich M (2013) Host specialization in ticks and transmission of tick-borne diseases: a review. *Frontiers in Cellular and Infection Microbiology*, **3**.

Merten HA, Durden LA (2000) A state-by-state survey of ticks recorded from humans in the United States. *Journal of Vector Ecology*, **25**, 102–113.

Minoche AE, Dohm JC, Himmelbauer H (2011) Evaluation of genomic high-throughput sequencing data generated on Illumina HiSeq and Genome Analyzer systems. *Genome Biology*, **12**, R112.

MITOS WebServer (2015). Available at: mitos.bioinf.uni-leipzig.de.

mitotRNAdb (2009). Available at: mttrna.bioinf.uni-leipzig.de/mtDataOutput/.

Mixson TR, Lydy SL, Dasch GA, Real LA (2006) Inferring the population structure and demographic history of the tick, *Amblyomma americanum* Linnaeus. *Journal of Vector Ecology*, **31**, 181–192.

Nakamura K, Oshima T, Morimoto T, Ikeda S, Yoshikawa H, Shiwa Y, Ishikawa S, Linak MC, Hirai A, Takahashi H, Altaf-Ul-Amin M, Ogasawara N, Kanaya S (2011) Sequence-specific error profile of Illumina sequencers. *Nucleic Acids Research*, **39**, e90.

Nardi F, Carapelli A, Fanciulli PP, Dallai R, Frati F (2001) The complete mitochondrial DNA sequence of the basal hexapod *Tetrodontophora bielanensis*: evidence for

heteroplasmy and tRNA translocations. *Molecular Biology and Evolution*, **18**, 1293–1304.

Nene V (2009) Tick genomics--coming of age. *Frontiers in bioscience (Landmark edition)*, **14**, 2666–2673.

Norris DE, Klompen JSH, Black WC (1999) Comparison of the mitochondrial 12S and 16S ribosomal DNA genes in resolving phylogenetic relationships among hard ticks (Acari: Ixodidae). *Annals of the Entomological Society of America*, **92**, 117–129.

Norris DE, Klompen JSH, Keirans JE, Black WC (1996) Population genetics of *Ixodes scapularis* (Acari: Ixodidae) based on mitochondrial 16S and 12S genes. *Journal of Medical Entomology*, **33**, 78–89.

Peden JF (1999) Analysis of Codon Usage. Ph.D. Thesis. University of Nottingham.

R Core Team (2013) *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

Savage HM, Godsey MS, Lambert A, Panella NA, Burkhalter KL, Harmon JR, Lash RR, Ashley DC, Nicholson WL (2013) First detection of Heartland virus (Bunyaviridae: Phlebovirus) from field collected arthropods. *The American Journal of Tropical Medicine and Hygiene*, **89**, 445–452.

Shao R, Aoki Y, Mitani H, Tabuchi N, Barker SC, Fukunaga M (2004) The mitochondrial genomes of soft ticks have an arrangement of genes that has remained unchanged for over 400 million years. *Insect Molecular Biology*, **13**, 219–224.

Shao R, Barker SC (2007) Mitochondrial genomes of parasitic arthropods: implications for studies of population genetics and evolution. *Parasitology*, **134**, 153–167.

Sharp PM, Tuohy TMF, Mosurski KR (1986) Codon usage in yeast: cluster analysis clearly differentiates highly and lowly expressed genes. *Nucleic Acids Research*, **14**, 5125–5143.

Smith C, Heyne S, Richter AS, Will S, Backofen R (2010) Freiburg RNA Tools: a web server integrating IntaRNA, ExpaRNA and LocARNA. *Nucleic Acids Research*, **38**, W373–W377.

Species2000 (2015) Species2000. *Species2000*. Available at: www.sp2000.org/sp2kwebsite/index.php.

Springer YP, Eisen L, Beati L, James AM, Eisen RJ (2014) Spatial distribution of counties in the continental United States with records of occurrence of *Amblyomma americanum* (Ixodida: Ixodidae). *Journal of Medical Entomology*, **51**, 342–351.

Storey JD, Tibshirani R (2003) Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 9440–9445.

Sukumaran J, Holder MT (2010) DendroPy: a Python library for phylogenetic computing. *Bioinformatics*, **26**, 1569–1571.

Suzuki H, Saito R, Tomita M (2009) Measure of synonymous codon usage diversity among genes in bacteria. *BMC Bioinformatics*, **10**, 167.

Trout RT, Steelman CD, Szalanski AL (2010) Population genetics of *Amblyomma americanum* (Acari; Ixodidae) collected from Arkansas. *Journal of Medical Entomology*, **47**, 152–161.

Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG (2012) Primer3—new capabilities and interfaces. *Nucleic Acids Research*, **40**, e115.

Williams-Newkirk AJ, Rowe LA, Mixson-Hayden TR, Dasch GA (2014) Characterization of the bacterial communities of life stages of free living lone star ticks (*Amblyomma americanum*). *PLoS ONE*, **9**, e102130.

Wright F (1990) The "effective number of codons" used in a gene. *Gene*, **87**, 23–29.

Xiong H, Barker SC, Burger TD, Raoult D, Shao R (2013) Heteroplasmy in the mitochondrial genomes of human lice and ticks revealed by high throughput sequencing. *PLoS ONE*, **8**, e73329.

Zwickl DJ (2006) Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. Ph.D. Thesis. The University of Texas at Austin, Austin, Texas, USA.

CHAPTER 6

Strengths, limitations, and conclusions

As climate change continues to alter the range of vectors worldwide, the distribution and prevalence of their associated diseases will also change. Without an adequate understanding of the many variables that determine the ability of a vector to transmit pathogens, accurate predictive models cannot be created. The results presented in this dissertation comprise a systematic investigation into a neglected aspect of the tick *Amblyomma americanum*'s vector capacity – its bacterial community. Both the findings and the limitations of these studies provide a foundation for future work on this species and other ixodid ticks.

6.1 Key findings and strengths

Discovery of novel potential symbionts in Amblyomma *species*

In my surveys of the bacterial diversity associated with ticks, I detected three novel genotypes of *Candidatus* 'Midichloria mitochondrii' (hereafter *M. mitochondrii*) in *A. americanum* (Chapter 2, Williams-Newkirk *et al.* 2012) and two novel genotypes of *M. mitochondrii* in *A. maculatum* (Chapter 4, Williams-Newkirk *et al.* 2015b). The functional significance of *M. mitochondrii* in both of these species remains to be determined, but evidence from the best studied tick-*M. mitochondrii* system (*Ixodes ricinus*) to date suggests that the bacterium is a facultative symbiont which facilitates tick

survival under natural but not laboratory conditions (Lo *et al.* 2006; Pistone *et al.* 2012). A previous conventional targeted PCR survey of ticks for this bacterium (Epis *et al.* 2008) did not detect it in either of these species despite its relatively high prevalence in those tick populations where it was found (11-48%). Next Generation sequencing is an important tool for the discovery of novel bacteria as it does not require *a priori* knowledge of the sequence or prevalence of the target, and it avoids potential false negative results due to target gene-specific primer mismatches since the 16S gene is probably the most widely characterized bacterial gene. Its conserved regions are well-known and extensive databases exist (DeSantis *et al.* 2006; Cole *et al.* 2009; Quast *et al.* 2013).

Sex-specific bacterial communities are found in disparate populations of A. americanum

I analyzed *A. americanum* from five populations in three eastern states and found the same pattern of sexual dimorphism in bacterial communities despite differences in sample handling and tick storage (Chapter 3, Williams-Newkirk *et al.* 2014). Community dissimilarity results were linked to a shift in the relative proportions of the dominant members of the tick bacterial community, *Coxiella* and *Candidatus* 'Rickettsia amblyommii' (hereafter *R. amblyommii*) in males and females. I also found that the bacterial communities of nymphal ticks spanned the range of variation seen in males and females, suggesting that sex-specific bacterial communities may already be forming in unfed nymphal ticks despite their otherwise distinct community composition. However, this hypothesis cannot be tested until molecular sexing or more convenient cytological methods are developed for immature ticks.

Sexually dimorphic bacterial communities are not a fixed trait of ixodid ticks

Previous studies of the whole bacterial communities of ixodid ticks by other investigators (Andreotti *et al.* 2011; Carpi *et al.* 2011) did not have sufficiently large sample sizes to examine the differences between the sexes with any statistical validity or did not provide these analyses in their findings (Lalzar *et al.* 2012; Nakao *et al.* 2013; Ponnusamy *et al.* 2014; Budachetri *et al.* 2014). It was unclear from my findings with *A. americanum* whether or not the sexual and stage differences were a generalized phenomenon in other ticks. Consequently, I examined the bacterial communities of males and females from three additional sympatric tick species with varying degrees of phylogenetic relatedness to *A. americanum* but found evidence of sexual dimorphism only in *I. scapularis* (Chapter 4, Williams-Newkirk *et al.* 2015b). As in *A. americanum*, the *I. scapularis* sexual differences appeared to be driven by an increase in the proportion of the primary symbiont in females. However, it is notable that while *A. americanum* and *I. scapularis* share similar habitat preferences (forest), they have different primary symbionts (*Coxiella* versus *Rickettsia*). In contrast, while *D. variabilis* and *A. maculatum* share both a habitat preference (fields and open woodland but with differences in humidity) and a third type of primary symbiont (related *Francisella*-like endosymbionts), neither showed clear evidence of sexual dimorphism. These observations suggest that the population dynamics of ixodid tick primary symbionts differ between species, a finding which further suggests potential significant differences in their functional roles in their respective tick hosts.

Sympatric tick species share few bacteria

In my comparison of the bacterial communities of four sympatric tick species, I demonstrated that while these species share habitats and are known to co-feed on many of the same hosts, they share very few of their bacterial community members (Chapter 4, Williams-Newkirk *et al.* 2015b). I know of no previous study which addressed this question using whole community data. These findings suggest that ticks are not as open to colonization by introduced bacteria as has been previously hypothesized and likely exert substantial control over their microbiome compositions.

Phylosymbiosis plays a modest role in structuring tick bacterial communities

Ixodid ticks have long, complex life cycles which expose their bacterial communities to a wide variety of environmental conditions and diverse vertebrate hosts which may directly or indirectly affect their community composition and structure. I examined for the first time the effect of tick phylogeny on their associated bacterial communities (Chapter 4, Williams-Newkirk *et al.* 2015b), a pattern recently termed phylosymbiosis (Brucker & Bordenstein 2013). Phylosymbiosis makes no assumptions about the specific evolutionary mechanism causing host and bacterial community relationships to mirror each other, although co-evolution of communities and their hosts is one possible mechanism. I quantified the dissimilarity between communities using four different metrics and found only one (weighted UniFrac) which produced a relationship consistent with that seen in the tick mitochondrial genes. Further examination of the data indicated that the two *Amblyomma* species had more similar bacterial communities due to their high prevalence of both alpha and gamma proteobacteria in their communities, whereas

the other species' communities were skewed towards one taxa or the other. The quantitative and phylogenetically distant relationship between the bacterial communities of sympatric species from the same genus suggests that phylosymbiosis was unlikely to be the dominant force which determined the structure of these contemporary communities.

Resolution of conflicting descriptions of the A. americanum *microbiome found in the literature*

At the time of publication of Chapter 3 (Williams-Newkirk *et al.* 2014) two Sanger sequencing studies (Clay *et al.* 2008; Heise *et al.* 2010) and three NGS studies (Yuan 2010; Menchaca *et al.* 2013; Ponnusamy *et al.* 2014) of the *A. americanum* bacterial community were available. While the two Sanger studies identified *Coxiella* and *Rickettsia* as the predominant community members, all three of the earlier NGS studies reported bacterial communities with some aberrant feature compared to these historical results. It may be that the use of ticks derived from laboratory colonies (Menchaca *et al.* 2013) or the very small samples size (Ponnusamy *et al.* 2014) skewed their results relative to previous work on wild ticks. Similarly, the early work of Yuan appears to have had significant methodological issues (*i.e.* potential sample cross-contamination) and has not been published in the peer-reviewed literature. However, the work in the present dissertation provided a sufficiently large sample size of wild ticks to confirm the basic conclusions from PCR assays and Sanger sequencing studies on amplicon clones while greatly expanding our knowledge about minority community members that were not detected in previous studies.

Creation of a reference mitochondrial genome for A. americanum

Genomic tools for ticks are extremely limited and no closed sequences are yet publically available due to the large size and high repeat content of tick genomes (Nene 2009). This makes population genetics and other studies requiring genetic markers difficult. Mitochondrial genes have a long history of use as genetic markers, but little sequence data was available for *A. americanum* except for fragments of the rRNA genes. I sequenced and assembled the whole mitochondrial genome of a single female *A. americanum* collected in Georgia (Chapter 5, Williams-Newkirk *et al.* 2015a) to provide a basis for future development of genetic markers for this species. Furthermore, with four whole mitochondrial genome sequences available for *Amblyomma* from the Americas and Australia, it will now be possible to design conserved primers which may be useful in the *de novo* sequencing of other genus members and phylogenetic analyses.

6.2 Limitations

All studies necessarily compromise on some aspects of study design to achieve other goals. While there were numerous significant findings in my research, my decision to focus on free living ticks due to known shifts in bacterial community structure and composition in laboratory colonies (Heise *et al.* 2010) potentially inhibited my ability to control and evaluate several important variables which may need further investigation. It is not my intention to discount either approach, but rather to emphasize the limitations inherent in both and to highlight areas where future laboratory investigations may complement findings from this natural system.

Lack of control for seasonal variation

The life stages of each tick species vary seasonally in abundance, and there is no concordance between species in the timing of a life stage's peak (Figure 6.1). Work in another laboratory has also demonstrated environmental and age related changes in tick bacterial communities (Ponnusamy *et al.* 2014). All of the ticks analyzed in my dissertation were wild collected and most were sampled during or near the peak of their species and life stage abundance. This was not optimal when attempting to compare bacterial communities between species, but it does reflect the reality of the cycles of these animals in their natural habitat. Future laboratory studies may examine the influence of seasonality on the tick microbiome in more detail.

Samples of sympatric tick species were collected from different sites

Samples of ticks analyzed in Chapter 4 (Williams-Newkirk *et al.* 2015b) were drawn from collections made by several collaborators in addition to those I personally collected. While the species of ticks are sympatric in parts of their respective ranges (Chapter 4, Figure 4.S1), not all samples were collected from exactly the same sites and not all species had been previously recorded at all sites. It is possible that differences in environmental exposures and host use could have changed the composition of some of the bacterial communities between sites. While not explicitly analyzed in the results presented due to a lack of statistical power, analysis of *D. variabilis* and *A. americanum* data from multiple sites did not indicate a strong effect of site on bacterial community composition or structure. Regardless, future work could utilize a "common garden" experimental design which holds ticks under similar environmental conditions and uses

the same host species for feeding to explore the impact of geographic variation and feeding exposure more explicitly.

Differences in sample handling

DNA extraction methods (Willner *et al.* 2012; Rubin *et al.* 2014) and library preparation protocols (Acinas *et al.* 2005; Berry *et al.* 2011) have been shown by others to affect the results of bacterial community studies in other systems. In Chapters 3 (Williams-Newkirk *et al.* 2014) and 4 (Williams-Newkirk *et al.* 2015b) I analyzed samples which had been extracted using two different protocols. In the *A. americanum* data set in Chapter 3 I found that there were significant differences in the sample sets (archival and Georgia) handled under two different protocols and performed a separate analysis for each data set. In Chapter 4, all of the *A. maculatum* DNAs were extracted by a collaborator using a different protocol than that applied to the other three species in the analysis. With no *A. maculatum* samples prepared using my standard protocol I was unable to determine if this extraction difference would cause a significant bias in the results. The similarity in distance between species clusters on NMDS plots suggested that it did not cause a difference, but the possibility cannot be excluded categorically with the present data. Future studies could minimally compare *A. maculatum* data obtained from both methods to determine if the conclusions would have changed without this potentially confounding variable.

Use of 454 sequencing technology

Sequencing technology evolved rapidly during the course of this dissertation. At its inception the Roche 454 platform was the leading technology for sequencing 16S

amplicons (*e.g.* Yildirim *et al.* 2010; Koenig *et al.* 2011; Ravel *et al.* 2011), but the Illumina MiSeq platform is currently favored. Illumina now provides higher throughput, higher quality, and lower per base cost than the last available 454 chemistry (reviewed in Nelson *et al.* 2014). These advancements could have allowed larger sample sizes and greater depth of coverage for the studies executed in the present dissertation. This may have revealed additional rare members of the tick microbiome. However, due to the abundance of the *Coxiella* and other primary symbionts present in these tick species, it is unlikely that a change in platforms would have significantly altered the conclusions presented here. Several individual ticks were deeply sequenced (~14,000x) on the 454 in the course of these studies with no increase in detected diversity (data not shown). Until methods can be developed to reduce the amount of primary symbiont DNA in these samples, it is unlikely that significant advances will be made in the description of the tick rare microbiome regardless of sequencing platform.

6.3 Future directions

This dissertation was conducted at a critical point of transition in tick microbiome research. Technologies which became available in the last decade have transformed the scope of studies and greatly reduced the effort and cost associated with their execution. I chose to focus on a single tick species with great ecological and public health importance in the southeastern United States and have successfully described its bacterial community both in terms of intraspecific variation and its relationship to other species.

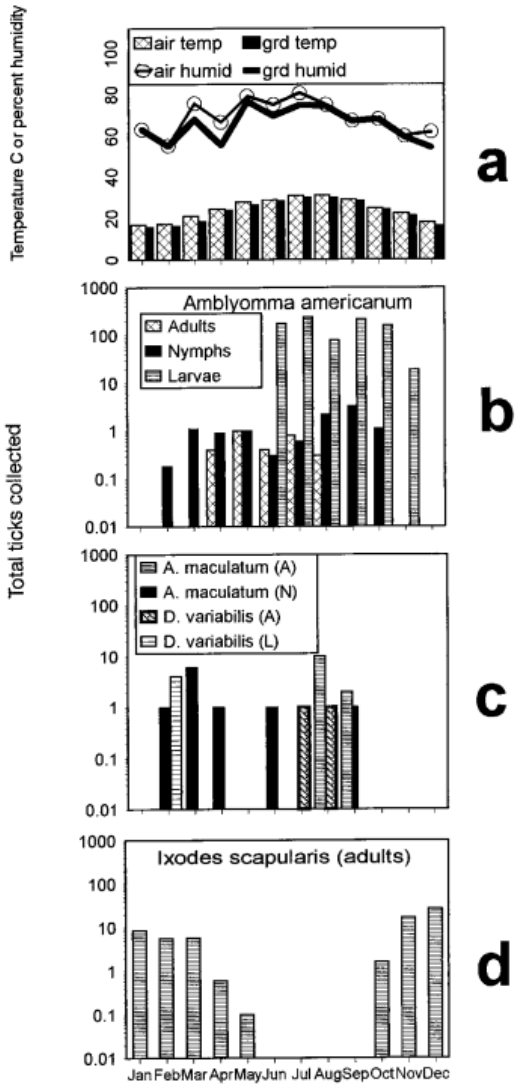
By using 16S amplicon sequencing I was able to rapidly characterize many members of the tick bacterial community, identifying novel members and assessing possible interactions. The bacteria of greatest abundance and / or potential impact may now be targeted for whole genome sequencing. With additional genetic data, it will be possible to map the metabolic pathways present within these bacteria to enable hypothesis development concerning their contribution to the tick holobiont. These hypotheses may then be tested through cloning of genes into vector cells or in situ using RNAi to knockout pathways of interest.

The reference mitochondrial sequence I have made available for *A. americanum* will be an important tool for all future population genetics studies of this species. The continued range expansion of *A. americanum* may now be more accurately monitored and the relationship between tick genetic variation and microbiome variation can be explored. Furthermore, as additional mitochondrial genomes become available, any conflicts within the phylogeny of the *Amblyomma* (currently believed to be polyphyletic) can be reconciled.

In conclusion, the long, complex life cycles of ixodid ticks provide opportunities for many different variables to affect their bacterial communities. My research has focused on species, life stage, sex, and phylogenetic relationship in natural, out breeding populations of ticks. The data I provided will form a foundation for many future studies in tick and TBD ecology and the approaches I have taken to investigating the incompletely described community of *A. americanum* may serve as a model for other ticks and for other hematophagous vector species.

6.4 Figures

Figure 6.1 Temporal variation in the abundance of tick species and life stages in the environmental. (a) Recorded average monthly air, ground temperature (8C) and humidity (%) during time of sampling (February 1994 through February 1996), (b) monthly total number *A. americanum* adults, nymphs, and larvae, (c) monthly total *A. maculatum* adults and nymphs, and *D. variabilis* adults and larvae, (d) monthly total *I. scapularis* adults collected from drags in St. Andrew's Recreational Area, Panama City, FL. Figure and description from Cilek and Olson (2000).



6.5 References

Acinas SG, Sarma-Rupavtarm R, Klepac-Ceraj V, Polz MF (2005) PCR-induced sequence artifacts and bias: insights from comparison of two 16S rRNA clone libraries constructed from the same sample. *Applied and Environmental Microbiology*, **71**, 8966–8969.

Andreotti R, León AAP de, Dowd SE, Guerrero FD, Bendele KG, Scoles GA (2011) Assessment of bacterial diversity in the cattle tick *Rhipicephalus* (*Boophilus*) *microplus* through tag-encoded pyrosequencing. *BMC Microbiology*, **11**, 6.

Berry D, Mahfoudh KB, Wagner M, Loy A (2011) Barcoded primers used in multiplex amplicon pyrosequencing bias amplification. *Applied and Environmental Microbiology*, **77**, 7846–7849.

Brucker RM, Bordenstein SR (2013) The hologenomic basis of speciation: gut bacteria cause hybrid lethality in the genus *Nasonia*. *Science*, **341**, 667–669.

Budachetri K, Browning RE, Adamson SW, Dowd SE, Chao C-C, Ching W-M, Karim S (2014) An insight into the microbiome of the *Amblyomma maculatum* (Acari: Ixodidae). *Journal of Medical Entomology*, **51**, 119–129.

Carpi G, Cagnacci F, Wittekindt NE, Zhao F, Qi J, Tomsho LP, Drautz DI, Rizzoli A, Schuster SC (2011) Metagenomic profile of the bacterial communities associated with *Ixodes ricinus* ticks. *PLoS ONE*, **6**, e25604.

Cilek JE, Olson MA (2000) Seasonal distribution and abundance of ticks (Acari: Ixodidae) in Northwestern Florida. *Journal of Medical Entomology*, **37**, 439–444.

Clay K, Klyachko O, Grindle N, Civitello D, Oleske D, Fuqua C (2008) Microbial communities and interactions in the lone star tick, *Amblyomma americanum*. *Molecular Ecology*, **17**, 4371–4381.

Cole JR, Wang Q, Cardenas E, Fish J, Chai B, Farris RJ, Kulam-Syed-Mohideen AS, McGarrell DM, Marsh T, Garrity GM, Tiedje JM (2009) The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Res.*, **37**, D141–D145.

DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, Huber T, Dalevi D, Hu P, Andersen GL (2006) Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Applied and Environmental Microbiology*, **72**, 5069–5072.

Epis S, Sassera D, Beninati T, Lo N, Beati L, Piesman J, Rinaldi L, McCoy KD, Torina A, Sacchi L, Clementi E, Genchi M, Magnino S, Bandi C (2008) *Midichloria mitochondrii* is widespread in hard ticks (Ixodidae) and resides in the mitochondria of phylogenetically diverse species. *Parasitology*, **135**, 485–494.

Heise SR, Elshahed MS, Little SE (2010) Bacterial diversity in *Amblyomma americanum* (Acari: Ixodidae) with a focus on members of the genus *Rickettsia*. *Journal of Medical Entomology*, **47**, 258–268.

Koenig JE, Spor A, Scalfone N, Fricker AD, Stombaugh J, Knight R, Angenent LT, Ley RE (2011) Succession of microbial consortia in the developing infant gut microbiome. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 4578–4585.

Lalzar I, Harrus S, Mumcuoglu KY, Gottlieb Y (2012) Composition and seasonal variation of *Rhipicephalus turanicus* and *Rhipicephalus sanguineus* bacterial communities. *Applied and Environmental Microbiology*, **78**, 4110–6.

Lo N, Beninati T, Sassera D, Bouman EAP, Santagati S, Gern L, Sambri V, Masuzawa T, Gray JS, Jaenson TGT, Bouattour A, Kenny MJ, Guner ES, Kharitonenkov IG, Bitam I, Bandi C (2006) Widespread distribution and high prevalence of an alpha-proteobacterial symbiont in the tick *Ixodes ricinus*. *Environmental Microbiology*, **8**, 1280–1287.

Menchaca AC, Visi DK, Strey OF, Teel PD, Kalinowski K, Allen MS, Williamson PC (2013) Preliminary assessment of microbiome changes following blood-feeding and survivorship in the *Amblyomma americanum* nymph-to-adult transition using semiconductor sequencing. *PLoS ONE*, **8**, e67129.

Nakao R, Abe T, Nijhof AM, Yamamoto S, Jongejan F, Ikemura T, Sugimoto C (2013) A novel approach, based on BLSOMs (Batch Learning Self-Organizing Maps), to the microbiome analysis of ticks. *The ISME Journal*.

Nelson MC, Morrison HG, Benjamino J, Grim SL, Graf J (2014) Analysis, optimization and verification of Illumina-generated 16S rRNA gene amplicon surveys. *PLoS ONE*, **9**, e94249.

Nene V (2009) Tick genomics--coming of age. *Frontiers in bioscience (Landmark edition)*, **14**, 2666–2673.

Pistone D, Sacchi L, Lo N, Epis S, Pajoro M, Favia G, Mandrioli M, Bandi C, Sassera D (2012) *Candidatus* Midichloria mitochondrii: symbiont or parasite of tick mitochondrii? In: *Manipulative Tenants: Bacteria Associated With*

Arthropods Frontiers in Microbiology. (eds Zchori-Fein E, Bourtzis K), pp. 113–126. CRC Press, Boca Raton.

Ponnusamy L, Gonzalez A, Treuren WV, Weiss S, Parobek CM, Juliano JJ, Knight R, Roe RM, Apperson CS, Meshnick SR (2014) Diversity of Rickettsiales in the microbiome of the lone star tick, *Amblyomma americanum*. *Applied and Environmental Microbiology*, **80**, 354–359.

Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research*, **41**, D590–D596.

Ravel J, Gajer P, Abdo Z, Schneider GM, Koenig SSK, McCulle SL, Karlebach S, Gorle R, Russell J, Tacket CO, Brotman RM, Davis CC, Ault K, Peralta L, Forney LJ (2011) Vaginal microbiome of reproductive-age women. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 4680–4687.

Rubin BER, Sanders JG, Hampton-Marcell J, Owens SM, Gilbert JA, Moreau CS (2014) DNA extraction protocols cause differences in 16S rRNA amplicon sequencing efficiency but not in community profile composition or structure. *Microbiology Open*, **3**, 910–921.

Williams-Newkirk AJ, Burroughs M, Changayile SS, Dasch GA (2015a) The mitochondrial genome of *Amblyomma americanum*. *Ticks and Tick-borne Diseases*, **in preparation**.

Williams-Newkirk AJ, Rowe LA, Mixson-Hayden TR, Dasch GA (2012) Presence, genetic variability, and potential significance of "*Candidatus* Midichloria

mitochondrii" in the lone star tick *Amblyomma americanum*. *Experimental and Applied Acarology*, **58**, 291–300.

Williams-Newkirk AJ, Rowe LA, Mixson-Hayden TR, Dasch GA (2014) Characterization of the bacterial communities of life stages of free living lone star ticks (*Amblyomma americanum*). *PLoS ONE*, **9**, e102130.

Williams-Newkirk AJ, Rowe LA, Sims S, Paddock C, Dasch GA (2015b) Sex and species phylogenies influence the bacterial communities of four sympatric North American tick species. *BMC Microbiome*, **in review**.

Willner D, Daly J, Whiley D, Grimwood K, Wainwright CE, Hugenholtz P (2012) Comparison of DNA extraction methods for microbial community profiling with an application to pediatric bronchoalveolar lavage samples. *PLoS ONE*, **7**, e34605.

Yildirim S, Yeoman CJ, Sipos M, Torralba M, Wilson BA, Goldberg TL, Stumpf RM, Leigh SR, White BA, Nelson KE (2010) Characterization of the fecal microbiome from non-human wild primates reveals species specific microbial communities. *PLoS ONE*, **5**, e13963.

Yuan D (2010) A metagenomic study of the tick midgut. Thesis (MS) Thesis. The University of Texas, Houston, Texas.