

The Texas Medical Center Library

DigitalCommons@TMC

The University of Texas MD Anderson Cancer
Center UTHealth Graduate School of
Biomedical Sciences Dissertations and Theses
(Open Access)

The University of Texas MD Anderson Cancer
Center UTHealth Graduate School of
Biomedical Sciences

8-2010

A METAGENOMIC STUDY OF THE TICK MIDGUT

Daniel T. Yuan

Follow this and additional works at: https://digitalcommons.library.tmc.edu/utgsbs_dissertations



Part of the [Biology Commons](#), [Laboratory and Basic Science Research Commons](#), and the [Pathogenic Microbiology Commons](#)

Recommended Citation

Yuan, Daniel T., "A METAGENOMIC STUDY OF THE TICK MIDGUT" (2010). *The University of Texas MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences Dissertations and Theses (Open Access)*. 85.

https://digitalcommons.library.tmc.edu/utgsbs_dissertations/85

This Thesis (MS) is brought to you for free and open access by the The University of Texas MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences at DigitalCommons@TMC. It has been accepted for inclusion in The University of Texas MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences Dissertations and Theses (Open Access) by an authorized administrator of DigitalCommons@TMC. For more information, please contact digitalcommons@library.tmc.edu.



A METAGENOMIC STUDY OF THE TICK MIDGUT

by

Daniel Tau Yuan, B.S.

APPROVED:

Steven J. Norris, Ph.D.

George E. Fox, Ph.D.

Angel M. Paredes, Ph.D.

Joseph F. Petrosino, Ph.D.

Hung Ton-That, Ph.D.

APPROVED:

Dean, The University of Texas
Graduate School of Biomedical Sciences at Houston

A METAGENOMIC STUDY OF THE TICK MIDGUT

A

THESIS

Presented to the Faculty of

The University of Texas

Health Science Center at Houston

and

The University of Texas

M. D. Anderson Cancer Center

Graduate School of Biomedical Sciences

in Partial Fulfillment

of the Requirements

for the Degree of

MASTER OF SCIENCE

by

Daniel T Yuan, B.S.

Houston, Texas

August 2010

Dedication

I dedicate this thesis to my family and friends. A man is only as good as the people who surround him. I am eternally grateful for their love and encouragement.

Acknowledgments

I would first like to thank both Dr. Steven J. Norris, Ph.D. and Dr. Tao Lin, D.V.M. for providing me with this extraordinary opportunity. This project would not have been possible without their knowledge and guidance. Thank you for not only teaching me technical skills, but also life skills that I will carry on as I enter the next phase of my life. I would also like to thank all of the professors who have served on my committee: Drs. George E. Fox, Angel M. Paredes, Joseph F. Petrosino, Ann-Bin Shyu, and Hung Ton-That. I am grateful for your constructive comments and guidance throughout this project. Every member of the Norris lab has somehow contributed to my success. I would like to thank them all. I have appreciated all of your help and have enjoyed building great working relationships with each of you.

I would like to give special appreciation to the late Dr. Edwin J. Masters. This project would simply not exist had it not been for his work in characterizing STARI. A special thanks goes to Dr. Pete Teel and Otto Strey of Texas A&M. Their knowledge of the tick is only surpassed by their willingness and diligence to provide us with hundreds of tick samples. Likewise, Dr. Ira Schwartz of New York Medical College provided us with New York ticks as well as a tremendous insight. Finally, Dr. Scot E. Dowd of Research and Testing Laboratories conducted our second 454 sequencing run. His desire to work with us and provide his own knowledge and resources was greatly appreciated.

Finally, my thanks to the Graduate School. Deans George Stancel, Thomas Goka, and Victoria Knutson and the staff have been tremendously supportive and keep the environment here a wonderful place for students to learn and grow.

A Metagenomic Study of the Tick Midgut

Daniel Yuan, B.S.

Supervisory Professor : Steven J. Norris, Ph.D.

Southern tick–associated rash illness (STARI) or Master’s disease is a Lyme-like illness that occurs following bites by *Amblyomma americanum*, the lone-star tick. Clinical symptoms include a bull’s eye rash similar to the erythema migrans lesions of Lyme disease, as well as fever and joint pains. Lyme disease is caused by *Borrelia burgdorferi* and related spirochetes. However, *B. burgdorferi* has not been detected in STARI patients, or in ticks in the South Central U.S. The causative agent of STARI has not been identified, although it was once thought to be caused by another *Borrelia* species, *Borrelia lonestari*. Furthermore, while adult *A. americanum* have up to a 5.6% *Borrelia lonestari* infection rate, the prevalence of all *Borrelia* species in Texas ticks as a whole is not known. Previous studies indicate that 6%-30% of Northern *Ixodes scapularis* ticks are infected by *Borrelia burgdorferi* while only 10% of Northern *A. americanum* and *I. scapularis* ticks are infected by *Borrelia* species. The first specific aim of this project was to determine the bacterial community that inhabits the midgut of Texas and Northeastern ticks by using high throughput metagenomic sequencing to sequence bacterial 16S rDNA. Through the use of massively parallel 454 sequencing, we were able to individually sequence hundreds of thousands of 16S rDNA regions of the bacterial flora from 133 ticks from the New York, Missouri and Texas. The presence of previously confirmed endosymbionts, specifically the *Rickettsia* spp. and *Coxiella* spp., that are commonly found in ticks were confirmed, as well as some highly prevalent genera that were previously undocumented. Furthermore, multiple pathogenic genera sequences were often found in the same tick, suggesting the possibility of co-infection of multiple pathogenic species. The second

specific aim was to use *Borrelia* specific primers to screen 344 individual ticks from Missouri, Texas and the Northeast to determine the prevalence of *Borrelia* species in ticks. To screen for *Borrelia* species, two housekeeping genes, *uvrA* and *recG*, were selected as well as the 16S-23S rDNA intergenic spacer. Ticks from Missouri, Texas and New York were screened. None of the Missouri or Texas ticks tested positive for *Borrelia* spp. The rate of *I. scapularis* infection by *B. burgdorferi* is dependent on tick feeding activity as well as reservoir availability. *B. burgdorferi* is endemic in the Northeast, sometimes reported as highly present in over 50% of all *I. scapularis* ticks. 11.6% of all New York ticks were positive for a species of *Borrelia*, however only 6.9% of all New York ticks were positive for *B. burgdorferi*. Despite being significantly lower than 50%, the results still fall in line with previous reports of about the prevalence of *B. burgdorferi*. 1.5% of all Texas ticks were positive for a *Borrelia* species, specifically *B. lonestari*. While this study was unable to identify the causative agent for STARI, 454 sequencing was able to provide a tremendous insight into the bacterial flora and possible pathogenic species of both the *I. scapularis* and the *A. americanum* tick.

TABLE OF CONTENTS

Title page	ii
Dedication	iii
Acknowledgements	iv
Abstract	v
LIST OF FIGURES	ix
LIST OF TABLES.....	x
CHAPTER 1 :INTRODUCTION	1
TICKS AND ASSOCIATED HUMAN PATHOGENS.....	2
LYME DISEASE AND S.T.A.R.I.	4
<i>BORRELIA BURGDORFERI</i> GENOME OVERVIEW.....	6
16S rDNA.....	7
454 SEQUENCING.	8
SIGNIFICANCE.	9
CHAPTER 2 : METHODS AND MATERIALS	17
TICK GENOMIC DNA EXTRACTION.	18
FULL LENGTH 16S rDNA PCR AND SEQUENCING.....	19
ELIMINATION OF THE DOMINANT <i>COXIELLA</i> ENDOSYMBIONT SPECIES	20
454 PILOT RUN.	23
SECOND 454 RUN AT THE RESEARCH AND TESTING LABORATORY.....	24
<i>BORRELIA</i> SEQUENCE DETECTION.	26
CHAPTER 3 : RESULTS	32

FULL LENGTH 16S rDNA CLONING	33
ELIMINATION OF DOMINANT COXIELLA ENDOSYMBIONT SPECIES	33
454 PILOT RUN AT BAYLOR COLLEGE OF MEDICINE	57
454 SEQUENCING RESULTS FROM RESEARCH AND TESTING LABORATORY	58
BORRELIA DETECTION.....	35
CHAPTER 4 : DISCUSSION.....	87
FULL LENGTH 16S rDNA CLONING.....	88
BORRELIA SPP. DETECTION USING BORRELIA SPECIFIC PRIMERS	89
454 SEQUENCING RESULTS.....	92
CHAPTER 5 : REFERENCES.....	99
VITA	103

LIST OF FIGURES

FIGURE 1. MAP SHOWING THE COLLECTION POINTS USED TO COLLECT TEXAS TICKS.....	14
FIGURE 2. <i>RECG</i> PHYLOGENIC TREE.....	48
FIGURE 3. <i>UVRA</i> PHYLOGENETIC TREE.....	50
FIGURE 4. 16S -23S INTERGENIC SPACER PHYLOGENETIC TREE	52
FIGURE 5. AVERAGE PERCENTAGE OF SEQUENCES, ALL <i>I. SCAPULARIS</i> <i>SPECIMENS</i>	73
FIGURE 6. AVERAGE PERCENTAGE OF SEQUENCES, ALL <i>A. AMERICANUM</i> <i>SPECIMENS</i>	75
FIGURE 7. STACKED BAR GRAPHS OF THE TOP FIVE GENERA, IN TERMS OF AVERAGE PERCENT SEQUENCE PER TICK.....	77
FIGURE 8. DUAL DENDROGRAM AND HEAT MAP INDICATING THE RELATEDNESS OF BACTERIAL CONTENT.....	83

LIST OF TABLES

TABLE 1. TICK COLLECTION INFORMATION.....	18
TABLE 2. LIST OF THE <i>BORRELIA</i> SPP. SPECIFIC PRIMERS.....	29
TABLE 3. <i>BORRELIA</i> I SPECIFIC PRIMER RESULTS.....	39
TABLE 4. <i>BORRELIA BURGDORFERI</i> STRAINS AND THEIR GEOGRAPHIC SOURCES.....	42
TABLE 5. <i>RECG</i> ALLELE TYPES OF <i>BORRELIA BURGDORFERI</i>	44
TABLE 6. <i>UVRA</i> ALLELE TYPES OF <i>BORRELIA BURGDORFERI</i>	46
TABLE 7. <i>BORRELIA</i> SPECIFIC QPCR RESULTS.....	56
TABLE 8. THE 20 BACTERIAL GENERA/SPECIES MOST CONSISTENTLY DETECTED IN A TICK SPECIES OR GEOGRAPHICAL GROUP.....	61
TABLE 9. DIFFERENCES IN PREVALENCE BETWEEN ALL <i>I. SCAPULARIS</i> TICKS AND ALL <i>A. AMERICANUM</i> TICKS; BETWEEN NEW YORK AND TEXAS <i>I.</i> <i>SCAPULARIS</i> TICKS; AND BETWEEN TEXAS AND MISSOURI <i>A. AMERICANUM</i> TICKS.	67
TABLE 10. AVERAGE PERCENTAGE EACH GENERA OR SPECIES REPRESENTED IN THE DIFFERENT GROUPS OF TICKS.....	80
TABLE 11. NUMBER OF BACTERIAL GENERA FOUND IN EACH TICK.....	86

CHAPTER 1 : INTRODUCTION

Ticks and associated human pathogens

Ticks are second to mosquitoes as arthropod vectors of human disease [1]. The majority of these diseases are zoonothroponoses, which are infections of animals that can be transmitted to humans, who are dead end hosts. Diseases transmitted by ticks include Lyme disease (also known as Lyme borreliosis), tularemia, relapsing fever, Rocky Mountain Spotted fever, erlichiosis, anaplasmosis, babesiosis, tick-borne encephalitis, Crimean-Congo hemorrhagic fever and Southern tick-associated rash illness (STARI) [1-3]. In the United States, Lyme disease is the most common tick-borne human pathogen, with 28,921 confirmed cases in 2008 compared to the less than 1200 cases of Rocky Mountain spotted fever annually [1, 4, 5]. Although the causative agent of Lyme disease, *Borrelia burgdorferi*, has yet to be confirmed to exist in Texas ticks, Lyme disease was the highest reported zoonotic disease in Texas in 2009 with 276 reported cases [6].

To better understand tick-borne human pathogens, it is important to be familiar with the vector itself. Ticks fall into two broad families; the Ixodidae, also known as hard-ticks, and the Argasidae, or soft ticks. Hard ticks are so called due to the presence of a hard plate covering the dorsal body surface known as the scutum, while soft ticks lack the scutum [3]. Of the Ixodidae ticks, six genera (*Ixodes*, *Dermacentor*, *Ambylomma*, *Haemaphysalis* and *Rhiicephalus*) are known to transmit human disease. Only the *Ornithodoros* genus of the Argasidae tick family is known to vector human pathogens [1, 3]. Both Ixodidae and Argasidae ticks go through a three stage life cycle : larva, nymph and then adult. All ticks feed on blood during one or all stages of their life [1]. Generally, ticks are long lived and can survive over a year without feeding[1]. Unfed adult ticks range in size from 2 mm to 20 mm, however when engorged, females may increase 100 times in weight and increase to more than 25 mm in length [3]. The physiology of the tick is divided into three major regions: capitulum,

body and legs. The anterior end of the capitulum contains all the mouth parts, including the chelicerae are needed to cut into the host's skin during feeding. The cuticle of the tick body is important to protect the tick from desiccation and injury while being flexible enough to expand during feeding. A large sac like structure in the tick body, known as the mid gut, contains numerous diverticula used to digest blood cells after a blood meal. Although ticks have an innate immune system, it is here in the midgut that infectious microbes such as *Borrelia burgdorferi* elude the tick immune system. There are four pairs of walking legs in adult and nymph ticks but only three pair walking legs on tick larva. The Haller's organ, a unique sensory pit which is present on all tick life stages, is found on the dorsal surface of the tarsus of the first leg [1]. The Haller's organ is able to detect odors and chemicals on the host skin as well as air currents and temperature changes through the use of multipurpose sensilla.

Environmental changes as well as the need to feed cause ticks to engage in host seeking behavior. There are general two strategies employed by ticks to seek their hosts : ambush strategy or the hunter strategy. The strategy of ambushing ticks varies depending on life stage. For example, adult ticks climb up vegetation to wait for hosts while tick nymphs tend to remain closer to the ground in hopes of encountering smaller hosts[1]. Ambushing ticks engage in a behavior known as questing, where alert but patient ticks wait for a host to pass by while responding to environmental changes such as vibrations, humidity, odors and other stimuli indicating the presence of a host [1].

Ticks that hunt are generally prefer larger mammals and have an array of abilities to detect preferred hosts. For example, *Rhipicephalus microplus*, commonly known as the cattle tick, responds quickly to odors derived from cattle skin but less so to human odors or dry air [1]. Hunting strategies include the ability to distinguish between dark shapes against bright backgrounds or sensitivity to high carbon dioxide concentrations to determine the location of a

host [1]. It is also believed that tick pheromones or other host odors detected by the Haller's organ can provide directional information to aid hunting ticks towards their host. When excited by host odors and other stimuli, some ticks that hunt are known to cover up to three meters to attack a potential host [1]. Low-volatility compounds, such as butyric acid or urea, are believed to be detectable by ticks and can provide selective information to determine suitable hosts[1]. Other physical stimuli such as feathers or hair, combined with chemical stimuli, must be present in order for a tick to feed on a particular host.

Lyme disease and STARI

In the United States, Lyme disease, also known as Lyme borreliosis, is strictly defined as a microbiological infection caused by the spirochete *Borrelia burgdorferi*. *Borrelia burgdorferi* is mainly transmitted by *Ixodes scapularis* tick in the eastern United States or by *Ixodes pacificus* in the western United States[4]. In 2008, 49 of the 50 states had at least one confirmed case of Lyme disease, with New York reporting the most confirmed cases at 5,741 cases [5]. However, in Texas and the South as a whole, Lyme disease is quite rare and instead, a Lyme-like illness has been observed [7]. Southern tick-associated rash illness (STARI) or Master's disease is a Lyme-like illness that occurs following the bites of *Amblyomma americanum*, the lone-star tick. Clinical symptoms include small annular lesions similar to erythema migrans or bull's eye rash of Lyme disease as well as fever and joint pains [7]. While both STARI and Lyme disease cause a bull's eye rash, STARI is less likely than Lyme disease to present other symptoms such as dizziness, headache or neck stiffness [7]. Furthermore, persistent or late stage Lyme borreliosis, which can lead to cardiac involvement, neurological damage or arthritis, does not appear to occur in STARI patients [7].

Although some symptoms are similar, STARI does not match the microbial definition of Lyme disease. Previous publications indicate that the STARI vector, *Amblyomma americanum*, is an incompetent vector of *Borrelia burgdorferi* due to the bactericidal effects of phospholipase A(2) present in the tick's saliva [8, 9]. In these studies, *B. burgdorferi* was cultured in the presence of saliva from *A. americanum* and *I. scapularis*, as well as a control containing the saliva induction chemical pilocarpine [8]. After 24 hours and 48 hours, spirochetes from each group were examined using a fluorescence method to quantify spirochete viability [8]. Spirochetes exposed to *A. americanum* saliva showed a significantly reduced average number of live spirochetes both at the 24 hour and 48 hour time points when compared to *I. scapularis* saliva samples or the pilocarpine control [8]. Additional evidence that STARI is not caused by *B. burgdorferi* is that the host seeking behavior of the Lyme disease vector, *I. scapularis*, does not correspond to the time which STARI is usually reported [7]. Furthermore, the appearance of the erythema migrans-like rash does correlate with an increased incidence of *A. americanum* tick bites [4]. Finally both PCR analysis of 31 STARI skin biopsies and *B. burgdorferi* serological testing (ELISAs and western blot) of 25 Missouri STARI patients indicated an absence of *B. burgdorferi* [7].

Borrelia lonestari has been identified as a species of *Borrelia* that infects the *A. americanum* tick. Although little is known about *Borrelia lonestari*, it is believed that *B. lonestari* is a member of Relapsing Fever group *Borrelia* due to analysis of the 16S rRNA sequence, flagellin genes and the presence of a *glpQ* ortholog, a gene only seen in Relapsing Fever group *Borrelia* [10]. As a whole, *Borrelia lonestari* has been reported to be found in 0.2%- 5.6% of southern *A. americanum* ticks and was once thought to be the causative agent of STARI however, the infection rate of *B. lonestari* can vary depending on location. A study of the *A. americanum* tick in Mississippi found that 3% of tested samples were positive for *B.*

lonestari. The prevalence of *Borrelia lonestari* can vary even within the same state. One report of Kentucky ticks indicates a 1.5% infection rate of *A. americanum* ticks by *B. lonestari* while another report, a two year study conducted between 2005-2007, found no evidence of *B. lonestari* infection in ticks collected by forty-two participants despite 14 participants reporting STARI like symptoms [39]. However, the prevalence of *B. lonestari* in Texas ticks is not known, nor is the exact role of *B. lonestari*. A previous report of a STARI skin biopsy as well as the attached *A. americanum* tick revealed the presence of *B. lonestari* [7]. However, subsequent studies of other STARI patients found no evidence of *B. lonestari*, making *B. lonestari* unlikely to be the causative agent [7, 11]. Attempts to culture a causative agent or agents of STARI using both media and in vivo culture techniques such as immunodeficient mice and chicken embryos have been unsuccessful [7].

***Borrelia burgdorferi* genome overview**

16S rDNA sequence analysis categorizes *Borrelia* spp. into two major groups: agents of Lyme borreliosis, such as *B. burgdorferi*, and relapsing fever agents such as *Borrelia hermsii*. The genome of *B. burgdorferi* consists of one linear chromosome that is just under 1 Mb with 21 extrachromosomal genetic elements, also known as plasmids, consisting of 9 circular and 12 linear plasmids [12, 13]. These 21 plasmids constitute 40% of total DNA. Strains of Lyme disease borreliae all have closely related linear chromosomes, but can differ in plasmid content and conformation. *B. burgdorferi* B31 has been shown to rapidly lose plasmids during in vitro culture, including two linear plasmids required for infectivity, lp25 and lp28-1 [13, 14]. The *B. burgdorferi* genome is predicted to encode 127 lipoproteins, more than any other characterized bacterial genome of similar size [4, 15]. These lipoproteins include both outer surface protein OspA, which mediates adhesion to tick gut epithelial cells, and OspC, which is involved in early colonization of mammalian hosts and tick salivary gland

invasion [16, 17]. Unlike typical gram-negative bacteria, *B. burgdorferi* does not have outer membrane lipopolysaccharide, nor does it appear to require iron [4, 18]. Stranger still, *B. burgdorferi* does not appear to express any toxins but instead appears to cause disease through adherence, dissemination, and evasion of immune clearance, leading to long-term infection and the induction of host inflammatory responses [4].

16S rDNA

The 16S ribosomal RNA is a crucial component of the 30S ribosomal subunit and is used by prokaryotes for the translation of RNA into proteins. The 3' end of the 16S ribosomal RNA contains the anti-Shine-Delgarno sequence which binds to the ribosome binding site at the 5' end of mRNA. Through interactions with the 23S rRNA of the 50S subunit, the 16S rRNA aids in binding the two ribosomal subunits to create the complete bacterial ribosome. Although 16S rDNA is highly conserved among most bacteria, there are nine hypervariable regions that demonstrate considerable sequence diversity and can be used for species identification [19]. Because of the significant sequence diversity, full length 16S rDNA sequencing has been used for over two decades to examine bacterial biodiversity without the need for culture or isolation.

A 2007 publication describes the hypervariable regions of 113 different bacterial 16S rDNA sequences, including those of blood borne pathogens, CDC select agents and environmental flora [19]. Nine hypervariable regions, known as V1 through V9, have been described. These regions are located at nucleotides 69-99, 137-242, 433-497, 576-682, 822-879, 986-1043, 1117-1173, 1243-1294 and 1435-1465 for V1 to V9 respectively according to the *E. coli* system of nomenclature [19]. Through the identification of these nine hypervariable regions, it is possible to differentiate bacterial species by sequencing only a portion of the 16S rDNA or perhaps a single hypervariable region, rather than the full length

16S rDNA [19]. Recently, it has been demonstrated that even short reads (230 basepairs or less) of region V3 and V6 are nearly as useful to analyze bacterial communities and can cost significantly less per read than traditional full length 16S rDNA sequencing [20, 21].

454 Sequencing

The dideoxynucleotide chain termination method developed by Sanger et al. has dominated DNA sequencing since it was first introduced in 1977 [22]. Over time, Sanger sequencing has improved in accuracy and read length; however, the relatively high cost limits its usefulness for large scale projects such as comparative sequencing of multiple microbial genomes. Now, massively parallel sequencing techniques are available, including the 454 Titanium platform. Unlike traditional sequencing techniques, 454 sequencing generates over 1 million reads per 454 run. With the current average read length of 400 bp, one sequencing reaction can generate 400-600 billion base reads per run [23]. Unfortunately, with an accuracy of 99%, 454 sequencing is not nearly as accurate as Sanger sequencing. However, unlike Sanger sequencing, which gives a single read per template, 454 sequencing sequences a template hundreds or perhaps thousands of times to create a depth of coverage not possible in traditional sequencing. Furthermore, 454 sequencing allows high throughput, parallel analysis of a mixture of DNA templates.

Shotgun 454 sequencing begins with the addition of special adapters to randomly fragmented template DNA. These adapters allow single stranded DNA to be attached onto specifically designed DNA Capture beads under conditions that allow one single strand fragment to bind to one bead. Beads are then transferred into an oil/water emulsion to create a microreactor consisting of one bead surrounded by all the reagents needed for a special PCR reaction known as an emulsion PCR. The emulsion PCR, or emPCR, covers each DNA capture bead with ~10 million copies of the original single strand fragment. The emPCR beads are then

loaded on to a special plate (called a PicoTiterPlate) that contains wells that allow only one bead per well. To perform the sequencing reaction, sequencing by synthesis occurs in which the complementary DNA strand is synthesized. Nucleotides are flowed sequentially in a fixed order across the PicoTiterPlate. DNA polymerase incorporates nucleotides into the complementary strand which causes the release of a pyrophosphate. In the presence of adenosine 5' phosphosulfate, ATP sulfurylase converts pyrophosphate into ATP which is used by luciferase to convert luciferin to oxyluciferin [24]. This conversion generates a light signal that can be detected by fiber optic and high density CCD technology. Signal strength is proportional to the number of nucleotides incorporated. For example, if the incorporation of a dATP results in signal strength of X, then the incorporation of two dATPs gives a signal strength of 2X [24]. Unincorporated nucleotides and ATP are degraded by apyrase, allowing the reaction to start again with a different nucleotide.

A variation of 454 sequencing has been developed that permits high throughput analysis of the bacterial population present in a sample (the microbiome). In this method, regions of 16S rDNA are amplified and purified before emPCR using a PCR amplification technique that simultaneously adds 454 sequencing adapters and specifically targets regions of bacterial 16S rDNA. The 454 sequencing reaction proceeds as normal with the attachment of only 16S rDNA sequences to DNA capture beads via 454 sequencing adapters.

Significance

An important step to investigating the existence of Lyme disease in Texas or a causative agent of STARI in Texas is to detect and characterize the causative agent from ticks. The overall long-term goal of our investigation is to identify possible causative agents of STARI in the Southern United States. A metagenomic survey of organisms present in the midguts of

Ixodes and *Amblyomma* ticks from the Southern states would define the flora present in the tick midgut and aid in the identification of the causative agent of STARI as well as other possible unrecognized pathogens. Furthermore, understanding the prevalence of *Borrelia* species in Texas using *Borrelia* specific primers would be an important step in determining the presence or absence of Lyme disease in Texas. Therefore, my specific aims were to:

- 1) Use *Borrelia* specific primers to screen 400 individual ticks from Missouri, Texas and the Northeast to determine the identity and prevalence of *Borrelia* species in these tick populations.
- 2) Determine the bacterial community that inhabits the midgut of Texas and Northeastern ticks by using high throughput metagenomic sequencing and identifying the bacterial 16S rDNAs present.

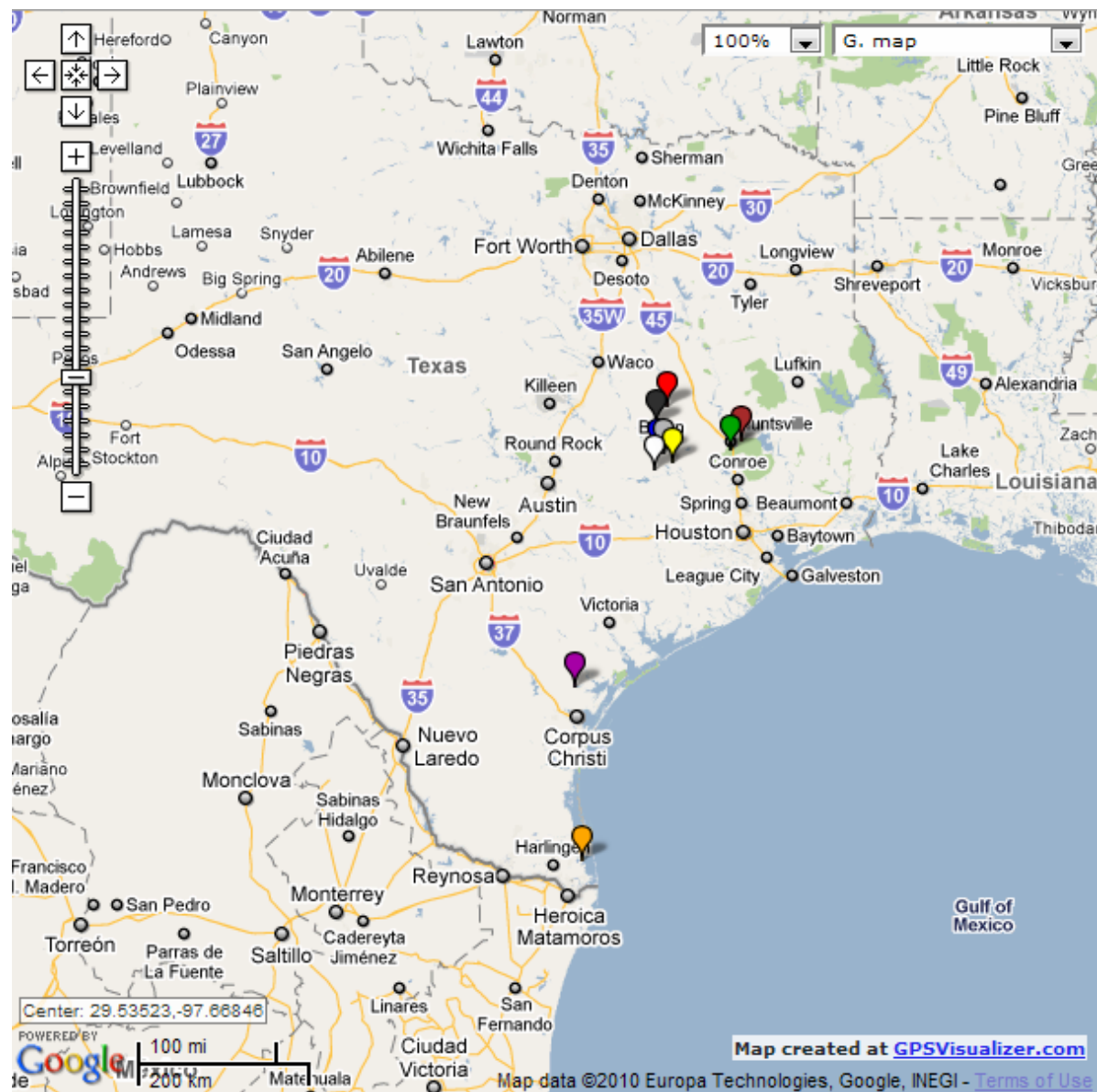
CHAPTER 2: MATERIALS AND METHODS

Collection and Decontamination of Ticks

Otto Strey, a senior research associate in Dr. Pete D. Teal's laboratory at Texas A&M, Dr. Edwin Masters from Missouri and Dr. Ira Schwartz from New York provided us with tick samples. Of these ticks, 732 were *Amblyomma americanum* ticks, including 414 adult ticks and 316 nymphal ticks. 115 ticks were *Amblyomma cajennense* ticks of which 35 were adult ticks and 80 were tick nymphs. One tick was an adult *Amblyomma maculatum* ticks. 143 ticks were *Ixodes scapularis* ticks of which 97 were adults and 46 are nymphs. 405 ticks were adult *Amblyomma imitator* ticks. In total, 1396 ticks were collected in the past two years.

Ticks were collected by dragging 1 m X 1 m flannel sheets across the ground, capturing and immobilizing the ticks in the flannel fibers allowing the tick to be individually collected and processed. Ticks were then stored in a glass dessicator with a solution of saturated potassium sulfate and water to maintain a high level of relative humidity. All ticks were processed in a sterile, ventilated hood to prevent environmental contamination. To prevent contamination from environmental bacterial species, all ticks were surface sterilized in successive three minutes washes, first in 3% hydrogen peroxide, followed by 95% ethanol, 1M sodium hypochlorite and finally washed and preserved in PBS. Ticks were vertically divided with one half used for DNA extraction. In certain circumstances, one tick half would once again be halved leaving only one quarter of the total tick for DNA. If tick genomic DNA could not be immediately extracted, tick halves were stored in a -20°C freezer.

Figure 1. Map showing the collection points used by Otto Strey to collect Texas ticks. Map created using GPSvisualizer.com with GPS coordinates provided by Mr. Strey. The legend provides exact GPS coordinates for each site.



Name	Latitude	Longitude	Color
Brazos Co. Private Residence	N30 34.614	W96 24.800	
Camp Creek	N31 03.780	W96 18.667	
Hallmark Ranch	N30 52.705	W96 26.608	
McFarland Ranch	N30 29.655	W96 14.119	
Huntsville State Park	N30 37.260	W95 32.076	
Laguna Atascosa Refuge	N26 13.735	W97 20.871	
Sam Houston National Forest	N30 42.439	W95 24.355	
Snook Ranch	N30 24.36	W96 27.40	
TAMU Range Area - East	N30 34.4143	W96 21.66	
Welder Wildlife Refuge	N28 07.112	W97 25.119	

Table 1. A table of tick collection information. The location, species, life stage and collector are indicated.

Tick Species	Location	Adults	Nymphs	Larva	collector
<i>A. americanum</i>	Camp Creek, Robertson Co. TX	200	113		Otto Strey
<i>A. americanum</i>	Dr. Master' s Tree Farm, MO	36			Dr. Ed Masters
<i>A. americanum</i>	Hallmark Ranch in Robertson Co.	178	190	1	Otto Strey
<i>A. americanum</i>	Laguna Atascosa National Wildlife Refuge		13	1	Otto Strey
<i>A. cajennense</i>	Laguna Atascosa National Wildlife Refuge	12			Otto Strey
<i>A. cajennense</i>	Welder Wildlife Refuge	23	80		Otto Strey
<i>A. imitator</i>	Laguna Atascosa National Wildlife Refuge	405			Otto Strey
<i>A. maculatum</i>	McFarland Ranch, Brazo &/or Washington Co	1			Otto Strey
<i>I. scapularis</i>	Brazos County Private Residence	7			Otto Strey
<i>I. scapularis</i>	Camp Creek, Robertson Co. TX	6			Otto Strey
<i>I. scapularis</i>	Fordham University, NY		43		Dr. Ira Schwartz
<i>I. scapularis</i>	McFarland Ranch, Brazo &/or Washington Co	44			Otto Strey
<i>I. scapularis</i>	Sam Houston National Forest Park, TX	35	1		Otto Strey
<i>I. scapularis</i>	Snook Ranch, TX	4	2		Otto Strey
<i>I. scapularis</i>	TAMU Range Area	1			Otto Strey
	Total <i>A. cajennense</i>	11	2	0	
	Total <i>A. americanum</i>	287	113	2	
	Total <i>I. scapularis</i>	97	46	0	
	Total <i>A. maculatum</i>	1			
	Total <i>A. imitator</i>	405			
	Total : Each life stage	952	442	2	
	Total Ticks	1396			

Tick Genomic DNA extraction

When possible, fresh tick halves were processed immediately using the Fast Prep 24 machine. The Fast Prep 24 machine homogenizes samples with multidirectional, simultaneous impaction using a ¼ inch ceramic bead and irregularly shaped garnet particles. Individual tick samples were suspended in the CLS-TC lysis solution provided by the Fast Prep 24 kit. Samples underwent two rounds of the multidirectional, simultaneous impaction for 40 seconds each round at 6.0 meters/second. Samples were placed on ice for two minutes between each round of impaction. Homogenized samples were then bound to a spin column and eluted using 100 µl of ultra pure nuclease free water. The elution step was repeated to obtain the maximum amount of genomic DNA.

Many of the samples were not processed immediately and instead were stored in a -20°C freezer. While in the -20°C freezer, much of the soft tissue of the *Amblyomma americanum* ticks desiccated. In order to extract any remaining genomic DNA, I used a modified version of the Fast Prep 24 protocol. Samples underwent the four rounds of the multidirectional, simultaneous impaction using the Fastprep machine for 40 seconds each round at 6.0 m/s. Samples were then incubated overnight at 4°C. The next morning, samples were bound to a spin column and eluted using 100 µl of ultra nuclease free pure water.

The ticks provided by Dr. Schwartz were over a decade old and preserved in 70% ethanol. To obtain genomic DNA from the ticks, Dr. Schwartz's genomic DNA extraction protocol was followed. The protocol is a modified version of Qiagen's DNeasy Blood and Tissue kit. Ticks were removed from the 70% ethanol and dried on filter paper for two minutes. 180 µl of Buffer ATL and 20 µl of Proteinase K were added to each tick sample. Samples were cut into quarters using a 20 g needle and then crushed using a wide bore 200 µl

tip. Samples then were incubated at 56°C overnight on a heating block. The next morning, 400 µl of a Buffer AL was added to each sample followed by elution through the DNeasy Minispin column. Successive washes with 500 µl of Buffer AW1 and Buffer AW2 followed. The final product was eluted in 100 µl of ultrapure water. The elution step was repeated to obtain maximum DNA yield. For each round of genomic DNA extraction, a blank control was used to test for potential contamination in the genomic DNA extraction kit solutions.

In all, 434 adult and nymph ticks were processed by using the DNeasy Blood and Tissue Kit or the Fastprep 24 kit. Of those, 5 µl of each tick genomic DNA extract was loaded on to 1.5 % agarose gel to be visualized. For 281 of the samples, DNA was clearly visible by ethidium bromide staining, but the quality of genomic DNA varied. The remaining 153 DNA samples could not be visualized using ethidium bromide staining. All 434 samples were used as template DNA with 16S rDNA degenerate primers 5035 and 5036 to determine the ability of each sample for downstream analysis. Of the 434 samples, 344 samples gave strong bands and were considered useful for further analysis.

Full length 16S rDNA PCR and Sequencing.

Full length bacterial 16S rRNA gene amplification was performed using 5035 (5'-AG AGT TTG ATY MTG GCT CAG-3') and 5036 (5'-AA GGA GGT GWT CCA RCC GCA-3') degenerate primers. Because of the high degree of sequence homology that exists between bacterial 16S rRNA genes, Phusion[®] (Finnzymes), a high fidelity DNA polymerase, was used to amplify our targets. Each 25 µl reaction consisted of 17.25 µl of nuclease free water, 5 µl of Phusion 5X buffer, 0.5 µl of 10 mM dNTPs, 0.5 µl of each 10 mM primer working stock, 1 unit of Phusion DNA Polymerase and 1 µl of tick genomic DNA extract. All PCR reactions were performed on the Eppendorf EP Gradient S thermocycler. After an initial 30 second

denaturation at 98°C, the PCR proceeded for 35 cycles as follows: (i) 98°C, 10 seconds; (ii) 55°C, 30 seconds; (iii) 72°C, 1 min; and (iv) a final extension at 72°C for 10 min. PCR products were analyzed on a 1.5% agarose gel for 30 minutes at 100 volts. Full length amplicons, between 1.5 kb to 1.6 kb, were gel purified and extracted using the QiaQuick Gel Purification Kit and eluted with nuclease free water to a final volume of 50 µl. In order to TA clone the 16S rDNA amplicons, an A-tailing reaction was performed. The 10 µl A-tailing reactions used 1 unit of DNAzyme II (Finnzymes), 8.5 µl of gel purified sample, 0.25 µl of 5mM dATP and 1 µl of 10X PCR buffer. Samples were incubated at 72°C for 20 minutes, after which 2 µl of fresh A-tailed product was directly recombined into the PCR 2.1 plasmid and then transformed into chemically competent TOP10 *E. coli* cells (Invitrogen).

The transformed TOP10 cells were then plated on LB/ carbenicillin plates with 50 µl of 40 µg/ml X-gal for a blue/white screen. White or near white clones were cultured in 96 well deep well plates with 2 ml of LB liquid media with 50 µg/ml of carbenicillin for selection. Plasmids were extracted using the Direct Prep96 kit which allowed for simultaneous processing of a 96 well deep well plate cultures using a vacuum manifold. Inserts were verified after a two hour *EcoR1* restriction digestion. Positive clones were then sent to the University of Washington to be sequenced using M13 forward primers.

Elimination of the dominant *Coxiella* endosymbiont species 16S rDNA sequences

Our initial full length 16S rDNA sequencing results of the midgut of the *Amblyomma americanum* was dominated by the highly prevalent *Coxiella* endosymbiont. Two methods were used in an attempt to remove the *Coxiella* endosymbiont and enrich our downstream PCR reactions : Suicide Polymerase Endonuclease Restriction digest (SuPER digest)[25] and a simple digest of the endosymbiont using *Pst*I.

For the SuPER reactions, forward and reverse oligonucleotides were designed that specifically hybridize to a region of the *Coxiella* endosymbiont 16S rDNA sequence, in this case CE 852F (5'- CTG TTA GAA AAC TTG TTT TCT G -3') and CE 852R (5'- CAG AAA ACA AGT TTT CTA ACA G -3'). Although the two primers were reverse complements of each other, the initial primer annealing tests along with the paper describing the technique indicated that reverse complements would not prevent a successful reaction. The 22 µl reaction begins with 5 µl of Phusion 5X High Fidelity Buffer, 1 µl of 100 pmol primer CE 852 F, 1 µl of 100 pmol primer CE 852 R, 2 µl tick genomic DNA, 1 unit of Phusion High Fidelity Taq. SuPER reactions began by denaturing all DNA for 3 minutes at 95°C. The thermal cycler was paused to allow 10 units of *Tsp509I* and 2 µl of 10 mM dNTPs, prewarmed to 65°C, to be added to the reaction. The restriction enzyme *Tsp509I* is fully functional at 65°C and only digests double stranded DNA TTAA sites. The reaction was resumed allowing CE 852F and CE 852R to specifically anneal to the *Coxiella* endosymbiont, priming extension of the target DNA by the Phusion High Fidelity polymerase to create double stranded DNA. By designing primers specific to the *Coxiella* endosymbiont, only the *Coxiella* endosymbiont 16S rDNA should be extended (made double-stranded) by the polymerase and be available for *Tsp509I* digestion. Unlike a PCR reaction, which has multiple cycles of annealing and denaturing, SuPER has only one denaturing and annealing step which allowed all DNA not targeted by our primers (i.e. all non *Coxiella* endosymbiont DNA) to remain single-stranded and avoid *Tsp509I* digestion. The reaction proceeded at 68°C for one hour, followed by 30 minutes at 95°C to inactivate *Tsp509I*. The reaction was once again paused to add 4.8 µl of a 20 mg/ml Proteinase K solution then incubated for another 30 minutes at 58°C followed by 95°C for 10 minutes to completely inactivate the Proteinase K. Reactions remained at 10°C overnight to

allow all complementary strands to reanneal. SuPER reactions were then purified using the QIAGEN PCR purification kit and eluted in 40 µl of water.

Two controls were used in each set of SuPER digestions : a *Coxiella* endosymbiont 16S rDNA clone and *Borrelia burgdorferi* B31 genomic DNA. A successful SuPER digestion would eliminate the *Coxiella* endosymbiont 16S rDNA clone while leaving the *B.burgdorferi* 16S rDNA undigested. In order to determine the success of the SuPER digestion, all purified SuPER digestion samples, including positive and negative controls, underwent a PCR amplification using 5035 and 5036 16S rDNA degenerate primers. The 50 µl reaction that consists of 36.5 µl of purified SuPER reaction product, 10 µl of Phusion 5X Buffer, 1 µl of 10 mM dNTPs, 1 µl of each 10 mM primer and 2 units of Phusion High Fidelity Taq. Reactions were analyzed on 1.5% agarose gels.

We also attempted to find a unique restriction site in the *Coxiella* endosymbiont 16 S rRNA gene that would allow us to simply digest the endosymbiont DNA and prevent it from being amplified in downstream PCR reactions. A *Pst*I (CTGCAG) site was found 512 basepairs into the full length 16 S rRNA gene of the *Coxiella* endosymbiont that was not present in the next 20 related species. *Pst*I sites on other bacterial species were screened using a FASTA file containing 16S rDNA sequences from over a thousand bacteria species, and although *Pst*I sites exist in other bacterial 16S rRNA genes, they were bacterial species that were unlikely to be found in ticks. We believed that we could enrich our PCR reaction by simply digesting our template DNA for one hour at 37 °C in a 20 µl reaction consisting of 20 units of *Pst*I, 2 µl of NEB Buffer 3 and 16 µl of template DNA. Each digestion would be purified with the Qiagen PCR Purification kit and used as template for a PCR reaction that amplified the full length 16 S rRNA gene.

454 Pilot Run

The initial 454 pilot run consisted of only Missouri *Amblyomma americanum* ticks, and was performed at the Baylor College of Medicine Human Genome Center. Each sample was amplified using one of three sets of primers. Each set of primers targets a cluster of hypervariable regions in the 16S ribosomal gene and added a barcode to identify which hypervariable regions were amplified. Hypervariable region V1 to hypervariable region V3 were amplified using 454 Bact 8F (5'- CCT ATC CCC TGT GTG CCT TGG CAG TCT CAG CCT ACG AGA GTT TGA TCC TGG CTC AG -3') and V1V3R -XLR_534R_v2bBar212L (5' - CCA TCT CAT CCC TGC GTG TCT CCG ACT CAG TCA CAC ATT ACC GCG GCT GCT GG - 3'). Hypervariable region V3 to hypervariable region V5 were targeted using V3V5F - AG4788_XLRB_357F (5'-CTA TCC CCT GTG TGC CTT GG CAG TCT CAG CCT ACG GGA GGC AGC AG - 3') and V3V5R- XLR_926R_v2bBar8L(5' - CCA TCT CAT CCC TGC GTG TCT CCG ACT CAG CAC GCC CGT CAA TTC MTT TRA GT- 3'). Finally, hypervariable region V6 to hypervariable region V9 were amplified using V6V9F- AG4790_XLRB_U968f (5'- CCT ATC CCC TGT GTG CCT TGG CAG TCT CAG AAC GCG AAG AAC CTT AC - 3') and V6V9R - XLR_1492R_v2bBar23L (5' - CCA TCT CAT CCC TGC GTG TCT CCG ACT CAG CGC AAC TAC GGY TAC CTT GTT AYG ACT T - 3'). Primers consisted of a 29 base pair adapter and the four to six base pair barcode upstream of the priming region to allow for sequence identification.

Each Missouri tick genomic DNA sample was used as template DNA for amplification using all three primer sets. Each cluster of variable regions was amplified in a 25 µl reaction consisting of 18.3 µl of nuclease free water, 5 µl of PlatinumPFX 10X buffer, 2.5 µl of 10mM dNTPs, 0.1 µl of 10mM primer working stocks, 1 unit (0.2 µl) of PlatinumPFX Taq, 0.5 µl of 50 mM MgSO₄ and 1 µl of tick genomic DNA extract. All PCR reactions were performed on

the Eppendorf EP Gradient S thermocycler. After an initial two minute denaturation at 94°C, the PCR proceeded for 35 cycles as follows: (i) 9 °C, 15 seconds; (ii) 52.6°C, 30 seconds; (iii) 68°C, 30 seconds; and (iv) a final extension at 68°C for 10 min.

5 µl of the PCR products were analyzed on a 1.5% agarose gel for 30 minutes at 100 volts. If a single amplicon was present, the product was purified following Qiagen's Qiaquick PCR Purification Kit protocol. If multiple bands were seen, then PCR reactions were gel purified using Qiagen's Qiaquick Gel Purification Kit protocol.

Second 454 Run at the Research and Testing Laboratory

DNA prepared from ticks from Missouri, New York, and Texas were used for a second 454 sequencing performed at the Research and Testing Laboratories (RTL) in Lubbock, Texas. Extracts from 133 ticks prepared in our laboratory in Houston were selected for this analysis, and are described in greater detail below. As indicated previously, the quality of the tick genomic DNA as determined by agarose gel electrophoresis varied substantially in terms of both quantity and quality. Some samples appeared to be highly fragmented while others contained only tight, high molecular weight bands; the intensity of ethidium bromide staining also varied substantially. Therefore, preliminary PCR reactions using the 16S rDNA degenerate primers 5035 and 5036 were utilized to provide an indication of the suitability of each genomic DNA preparation for further analysis. Each 25 µl reaction consisted of 17.25 µl of nuclease free water, 5 µl of Phusion 5X buffer, 0.5 µl of 10mM dNTPs, 0.5 µl of 10 mM primer working stocks, 1 unit of Phusion Taq polymerase and 1 µl of tick genomic DNA extract. All PCR reactions were performed on a Eppendorf EP Gradient S thermocycler. After an initial 30 second denaturation at 98°C, the PCR proceeded for 35 cycles as follows: (i) 98°C, 10 seconds; (ii) 55°C, 30 seconds; (iii) 72°C, 1 min; and (iv) a final extension at 72°C for 10

min. PCR products were analyzed on a 1.5% agarose gel for 30 minutes at 100 volts. 125 of the 133 ticks gave strong PCR results. An additional 434 tick genomic DNA samples were tested using the 16S degenerate primers, of which 199 gave positive PCR products, bringing the total number of acceptable genomic DNA samples to 332.

A total of 147 tick samples were sent for individual 16S rDNA PCR amplification and sequencing by 454 technology at the RTL. These samples included the 133 tick gDNA extracts prepared at UT Medical School, consisting of DNA preparations from 43 *Ixodes scapularis* nymphs from Fordham University in New York, 42 *Amblyomma americanum* adults from Missouri (collected by Dr. Edwin Masters), and 48 ticks collected by Mr. Otto Strey and Dr. Pete Teel (Texas A&M University) in Texas, including 26 adult *I. scapularis* and 22 adult *A. americanum* ticks. Most of these samples were extracted from fresh tick specimens and represent extracts from half or a quarter of a whole tick. In addition to these DNA preparations, 14 frozen tick halves from 7 *Amblyomma cajennense* adults and 7 *A. americanum* nymphs collected in Texas were sent to the RTL. DNA from these frozen specimens was prepared at the RTL by the following method.

An initial amplification utilized the 8F bacterial 16S rDNA degenerate primer set, resulting in an amplicon that includes hypervariable regions 1, 2, and 3 of the 16S rDNA. 454 sequencing on the Titanium platform began by normalizing genomic DNA samples to 100 ng/μl [26]. 100 ng of genomic DNA was used in a 50 μl one step polymerase reaction that uses a mixture of Hot Start and HotStar high fidelity Taq polymerases to provide hybrid products consisting of adaptor sequences bound to the target bacterial 16S region [26, 27]. This reaction adds the unique barcodes used to identify each sample as well as Linker A and Linker B sequences, which are needed for attachment to the 454 beads and for priming the sequencing reaction. All amplicon products from different tick samples were combined and

purified using Agencourt Ampure beads [26]. Purified PCR products were then ready for 454 sequencing which begins with attachment to DNA capture beads, emPCR, placement into PicoTiter plates, and finally pyrosequencing [26]. The unique barcode sequences used for each sample permitted pooling of amplicons from up to 40 samples per 454 quadrant. Of the 147 samples sent to RTL, 129 samples were processed using 454 sequencing, generating 553,804 total reads.

Sequences were matched at the genus level and were documented in a table containing the percentage each bacterial genus represented in individual ticks. The actual sequences from the 454 run were organized into a large FASTA file. Data from table was organized into geographical groups as well as tick species. From the table, two important sets of data were created: genera prevalence among ticks and average genera percentage among ticks. To best illustrate the average percentage a bacterial genera represented for a group of ticks, stacked bar charts were created. Genera that represented less than 1% of total sequences for a group of ticks were grouped together to create one segment in the stacked bar chart. The remaining genera were assigned their own segment in the chart. For certain bacterial genera, the percentage of total sequences a genus represented varied substantially. For example, *Coxiella* spp. could be as high as 99.6 % of all sequences one tick but be completely absent in another tick.

***Borrelia* sequence detection**

In 2008, a novel multilocus sequencing typing (MLST) scheme was developed by a group of researchers in Europe to aid in identifying *Borrelia burgdorferi* strains [25]. Sequence information from the eight chromosomal housekeeping genes (*clpA*, *clpX*, *nifS*, *pepX*, *pyrG*, *recG*, *rplB*, and *uvrA*), the 16S rDNA – 23S rDNA intergenic spacer and the *ospC* gene provides sufficient information to characterize different genotypic lineages of *B. burgdorferi* and permit evolutionary and phylogenetic analysis. All reactions were either nested or semi-

nested reactions that involved two sets of primers. Although these primers were developed to amplify the corresponding genes of all Lyme disease agents, many of the *B. burgdorferi* sequences are not well conserved in other *Borrelia* species, such as *B. lonestari* and relapsing fever group *Borrelia* such as *B. hermsii*. To determine their ability to amplify all known *Borrelia* species, each primer set described in the MLST paper was compared to all known *Borrelia* sequences using the NCBI database and the BLAST utility. Of the primers described in the MLST study, only the *uvrA* and *recG* primers were 100% identical to the corresponding gene regions of the relapsing fever *Borrelia* species such as *B. hermsii*. The *Borrelia*-specific primer sets used in this study are listed in Table 2. Each set includes an initial forward and reverse primer, and a nested or semi-nested primer set that amplifies a region within the first product.

Table 2. List of the *Borrelia* spp. specific primers. The F and R following the primer indicates the forward and reverse primer for the primary PCR. The Fn and Rn indicate forward nested and reverse nested primers for the secondary PCR.

	16S - 23S Primers
IGS F	5'- GTA TGT TTA GTG AGG GGG GTG -3'
IGS R	5'- GGA TCA TAG CTC AGG TGG TTA G -3'
IGS Fn	5'- AGG GGG GTG AAG TCG TAA CAA G -3'
IGS Rn	5'- GTC TGA TAA ACC TGA GGT CGG A -3'
	<i>uvrA</i> Primers
uvrA F	5'- GCT TAA ATT TTT AAT TGA TGT TGG - 3'
uvrA R	5'- CCT ATT GGT TTT TGA TTT ATT TG - 3'
uvrA Fn	5' - GAA ATT TTA AAG GAA ATT AAA AGT AG - 3'
uvrA Rn	5' - CAA GGA ACA AAA ACA TCT GG - 3'
	<i>recG</i> Primers
recG F	5'- CTT TAA TTG AAG CTG GAT ATC -3'
recG R	5'- CAA GTT GCA TTT GGA CAA TC -3'
recG Rn	5' -GAA AGT CCA AAA CGC TCA G -3'

Initially, a gradient PCR using purified *Borrelia burgdorferi* B31 genomic DNA was used to determine each primer set's optimal annealing temperature. All PCR reactions were performed on the Eppendorf EP Gradient S thermocycler. Each 25 µl reaction consisted of 17.25 µl of nuclease free water, 5 µl of Phusion 5X buffer, 0.5 µl of 10mM dNTPs, 0.5 µl of 10mM primer working stocks, 1 unit of Phusion taq and 1 µl of template DNA (tick genomic DNA for the primary reaction, primary PCR product for the nested/seminested amplification). *uvrA* gene amplifications used the following conditions for the primary PCR reaction : After an initial 30 second denaturation at 98°C, the PCR proceeds for 8 cycles as follows: (i) 98°C, 10 seconds; (ii) 58°C, 30 seconds (decreasing by 1°C every successive cycle); (iii) 72°C, 15 seconds. 27 additional cycles followed the first 8 touchdown cycles as follows: (iv) 98°C, 10 seconds; (v) 50°C, 30 seconds; (vi) 72°C, 15 seconds; (vii) 72°C, 5 minutes and finally (viii) 10°C hold. The nested PCR reaction proceeded as follows : An initial 30 second denaturation at 98°C, the PCR proceeds for 35 cycles as follows: (i) 98°C, 10 seconds; (ii) 52°C, 30 seconds; (iii) 72°C, 15 seconds; and (iv) a final extension at 72°C for 10 min.

The 16S-23S intergenic spacer used separate PCR conditions. After an initial 30 second denaturation at 98°C, the PCR proceeds for 35 cycles as follows: (i) 98°C, 10 seconds; (ii) 56.7°C, 30 seconds; (iii) 72°C, 15 seconds; and (iv) a final extension at 72°C for 10 min. The nested PCR reaction conditions were an initial 30 second denaturation at 98°C, followed by 35 cycles of (i) 98°C, 10 seconds; (ii) 60°C, 30 seconds; (iii) 72°C, 15 seconds; and (iv) a final extension at 72 °C for 10 min.

Finally, *recG* gene amplification used one set of PCR conditions for both the primary and nested PCR reactions. After an initial 30 second denaturation at 98°C, the PCR proceeds for 35 cycles as follows: (i) 98°C, 10 seconds; (ii) 60°C, 15 seconds; (iii) 72°C, 15 seconds ; and (iv) a final extension at 72°C for 10 min.

Further PCR experiments were conducted to determine the sensitivity of the primer sets. *Borrelia burgdorferi* 5A4 clone were cultured in 2 ml of BSK media until a density of $\sim 1 \times 10^8$ organisms per ml was achieved. One ml of culture was centrifuged and resuspended in PBS + Mg^{2+} to achieve a final concentration of 1×10^9 cells per ml, which was then serially diluted 1:10 down to a concentration of 1 organism per ml. Diluted samples were then used as template for PCR reactions. Each reaction consisted of 17.25 μ l of nuclease free water, 5 μ l of Phusion 5X buffer, 0.5 μ l of 10mM dNTPs, 0.5 μ l of 10mM primer working stocks, 1 unit of Phusion Taq polymerase and 1 μ l of 5A4 culture.

Genomic DNA from a tick free of *Borrelia* species was combined with 5A4 organisms to ensure specificity and determine whether tick genomic DNA would interfere in *Borrelia* detection or lead to nonspecific binding by our primer sets. Each reaction consisted of 16.25 μ l of nuclease free water, 5 μ l of Phusion 5X buffer, 0.5 μ l of 10mM dNTPs, 0.5 μ l of 10mM primer working stocks, 1 unit of Phusion Taq polymerase, 1 μ l of tick genomic DNA and 1 μ l of the serially diluted 5A4 culture. The same tests were conducted using a live *B. lonestari* culture to ensure each primer set would be able to amplify relapsing fever group *Borrelia* species.

All positive samples were directly sequenced using their respective nested forward primers; in all cases, the sequences indicated that amplicon was derived from a member of the genus *Borrelia*. Samples were considered positive for *Borrelia* species if they yielded PCR products for two of the three primer sets. All 16S – 23S intergenic spacer, *uvrA* and *recG* sequences were aligned using the ClustalW algorithm of Bioedit 7.0.9. Included in each alignment were the 63 North American and European *B. burgdorferi* strains used in the MLST paper as well as additional *Borrelia* species, including other Lyme disease *Borrelia* and relapsing fever group *Borrelia* [25]. The 63 *B. burgdorferi* strains used in the MLST paper yielded 18

uvrA alleles, named *uvrA1* to *uvrA18*, and 17 *recG* alleles, named *recG1* to *recG17*. Trees for each gene set were generated using the neighbor-joining routine using 1000 bootstrap iterations using Bioedit 7.0.9.

CHAPTER 3 : RESULTS

Full length 16S rDNA cloning

Full length 16S rDNA amplification and cloning is a very simple and cost effective method for identifying microbial communities. In early studies of the bacterial content of ticks, whole tick genomic DNA preparations were utilized as template for amplification using the degenerate primers 5035 and 5036, which nonspecifically targeted prokaryotic 16S rDNA sequences to generate a pool of prokaryotic 16S rDNA amplicons. To differentiate the amplicons, each PCR reaction was cloned and sequenced to determine the bacterial content of the individual ticks. Ten *Amblyomma americanum* tick samples, all from Missouri, were used as template DNA for 16S rDNA amplification. The resulting amplicons were then cloned into pCR2.1 to create 542 clones, and the DNA inserts were sequenced. Of those clones, the vast majority of 16 S rDNA sequences were the highly prevalent *Coxiella* endosymbiont species of the *Amblyomma americanum* tick [28]. The *Coxiella* endosymbiont accounted for 527 of the 542 clones, or 98.3% of all clones. Of the remaining clones, two clones matched an uncultured *Propionibacteriaceae* species. One clone most closely matched the 16S rDNA of an unknown and uncultured bacteria species. Three sequences failed to return a significant match when subjected to a BLAST search. Nine sequences had partial matches with 16S rDNA sequences in GenBank; however, sequence homology existed only at the 5' end of these sequences and matched with 16S rDNA degenerate primer forward primer, 5035.

Elimination of dominant *Coxiella* endosymbiont species

16S rDNA cloning revealed a dominant *Coxiella* endosymbiont sequence in the full length 16S rDNA cloning experiments. In order to more effectively study the bacterial content of ticks, the *Coxiella* endosymbiont sequences needed to be eliminated or reduced before amplification of 16S rDNA. However, attempts to eliminate the *Coxiella* endosymbiont species or enrich our full length 16S rRNA gene PCR reaction were unsuccessful. We used *Borrelia*

burgdorferi genomic DNA as a positive control and a *Coxiella* endosymbiont 16S rRNA gene clone in plasmid PCR 2.1 as our negative control for the SuPER reactions. These controls were needed to prove specificity of our primers to the endosymbiont as well as success or failure of our SuPER digestion. PCR amplification using degenerate 16S rDNA primers 5035 and 5036 with both samples gave a 1.5 kb amplicon. In theory, SuPER digestion would specifically digest the *Coxiella* endosymbiont 16S rDNA sequence through the use of *Coxiella* endosymbiont specific primers. After an initial denaturization, primers CE852F and CE852R would specifically prime the *Coxiella* endosymbiont 16S rDNA gene, allowing DNA polymerase to extend the *Coxiella* endosymbiont 16S rDNA gene to create dsDNA. Double stranded *Coxiella* endosymbiont 16S rDNA would then be digested using *TSP509I*, leaving non *Coxiella* endosymbiont species undigested and free to reanneal. PCR amplification using degenerate 16S rDNA primers 5035 and 5036 followed. In terms of the controls, PCR amplification of the *Coxiella* endosymbiont 16S rRNA control gave a negative PCR reaction and SuPER treated *Borrelia burgdorferi* positive results. The controls consistently gave us the correct results in 16S rDNA amplification following SuPER digestion. However, once tick genomic DNA samples were used in the SuPER reaction, we were unable to eliminate the *Coxiella* endosymbiont.

A unique *PstI* digestion site was found in the *Coxiella* endosymbiont 16S rDNA gene. It was believed that the *PstI* restriction site would digest *Coxiella* endosymbiont sequences and leave non-*Coxiella* sequences available for 16S rDNA amplification. Initial tests were performed using the same *Coxiella* endosymbiont clone used in the SuPER reaction. Digestion was also unsuccessful as it appeared that the digestion was unable to eliminate the *Coxiella* endosymbiont at a concentration higher than 16 ng/μl. Increasing the concentration of *PstI* to 10 units/ μl per reaction or digesting the samples overnight did not completely digest the

Coxiella endosymbiont sequence. Initially it was thought that our *Pst*I enzyme had lost its effectiveness over time. However, a complete digestion of 500 ng of a 4 kb plasmid with a *Pst*I site demonstrated the effectiveness of *Pst*I. It was suggested that our plasmid did not contain the *Coxiella* endosymbiont 16S rDNA insert or perhaps we had contaminated the sample used in our experiments. The clone was regrown in liquid LB + Carb medium, prepped and resequenced from both ends using M13 forward and M13 reverse primers. The sequence obtained was not only clean and free of stray peaks, but clearly indicated the existence of a *Pst*I site at the 512 basepair position. Reactions were independently repeated by two other members of the lab with identical results.

***Borrelia* detection**

An important step to investigating the existence of Lyme disease in Texas and causative agent of STARI in Texas is to detect and characterize the causative agent from ticks. The prevalence of *Borrelia* species, specifically *Borrelia burgdorferi*, in Texas and South Central United States ticks is still not known nor has the causative agent for STARI been identified. The use of *Borrelia* specific primers is the best way to detect and discover the prevalence of *Borrelia* species in this region.

Being a Lyme disease lab, *Borrelia burgdorferi* contamination was a concern. It was determined that a positive result if the sequence was not *Borrelia burgdorferi* B31. Initial PCR reactions that used purified *B. burgdorferi* 5A4 clones as well as *B. lonestari* clones revealed the 16S-23S intergenic spacer (IGS) primers amplifies a 412 bp portion of the *B. lonestari* IGS spacer while amplifying a 945 bp fragment of the *B. burgdorferi* IGS spacer. *RecG* primers amplified a~ 740 bp fragment while the *uvrA* primer set amplified a ~680 base pair fragment.

Three hundred forty-four samples were screened using the 16S – 23S rDNA intergenic spacer primers (IGS), *recG* primers and *uvrA* primers. Of these samples, eight were true positives; five New York *I. scapularis* nymphs and three Texas *A. americanum* adults. The five positive NY ticks represents 11.6 % of total New York ticks while the three positive *A. americanum* ticks represents 1.5% of all Texas *A. americanum* ticks. None of the Missouri ticks were positive for *Borellia* spp.

Of the 199 Texas samples screened using the *uvrA*, 16S-23S intergenic spacer (IGS) and *recG* primer sets, three yielded positive results [25, 29]. IGS amplicons from Texas samples AN3-A and HR45-A, both *A. americanum* adults, were identical and matched one of the three IGS genotypes for *B. lonestari* described in previous papers. *RecG* sequences from AN3-A and HR45-A confirmed the presence of *B. lonestari*. Furthermore, this IGS sequence in these specimens was not identical to the IGS genotype for the *B. lonestari* culture used in our initial primer tests, which ruled out possible laboratory contamination. The third sample, CC71A, gave positive results using all three primer sets. The CC71A IGS sequence was identical to our laboratory *B. lonestari* strain, but both the *recG* and *uvrA* amplicons were a different genotype than that of our laboratory train [29].

Of the 5 positive samples from New York, there were a few relapsing fever group (RFG) *Borrelia* sequences as well as possible *B. burgdorferi* sequences. Both the 16S – 23S intergenic spacer sequences and the *recG* sequences obtained from tick sample NY07 and tick sample NY20 appear to belong to the relapsing fever group of *Borrelia*. In fact, NY07's IGS sequence matches the IGS sequence of *Borrelia miyamotoi*, a common relapsing fever group *Borrelia* of the Northeast United States [30]. Neither NY07 nor NY20 gave a positive PCR result using *uvrA* primers. Both the *recG* and the *uvrA* sequences from NY08 indicated a *B. burgdorferi* sequence that appeared to group with *Borrelia burgdorferi* strains from human

patients first collected in Westchester, New York. Strangely, NY08's IGS sequence grouped more closely with the relapsing fever group *Borrelia* and the European Lyme disease agent *Borrelia afzelii* than with American *Borrelia burgdorferi* strains. It is possible that NY08 was co-infected by two strains of species of *Borrelia*, one from the Lyme disease group and one from the relapsing fever group.

Like NY08's *uvrA* sequence, NY16's *recG*, *uvrA* and IGS sequences appeared to be similar to *B. burgdorferi* sequences first isolated from human patients from Westchester County. Although there was not a *uvrA* amplicon from NY27, both of NY27's *recG* and intergenic spacer sequences grouped with relapsing fever group *Borrelia*. The samples sent to RTL were also screened using all three sets of primers. The same conditions for a positive result applied, where amplicons from two of the three sets of primers was considered a positive result. Only NY08 and NY16 gave positive results using the *uvrA* primer set. These positive samples, combined with sequences from the MLST paper and other *Borrelia* species were used to create phylogenetic trees.

Table 3. Table indicating the location, number of samples per species and the number of positive PCR results using *Borrelia* specific housekeeping genes. Top row indicates which individual housekeeping genes or which combination of housekeeping genes gave a positive result for an individual samples. qPCR results conducted at RTL were included for comparison.

Location	Species	Total samples	Number of specimens yielding positive PCR results for one or more primer set							qPCR
			only IGS	only recG	only uvrA	both IGS & recG	both IGS & uvrA	both uvrA & recG	All primers	
New York										
	<i>I. scapularis</i>	43				3		1	1	6
Texas										
	<i>A. americanum</i>	186	2						1	5
	<i>A. cajenense</i>	21								
	<i>I. scapularis</i>	32								3
	<i>A. imitator</i>	20								
Missouri										
	<i>A. americanum</i>	42								5
	Total	344	2	0	0	3	0	1	2	19

Table 4. A list of the *B. burgdorferi* strain and their geographical source. These strains were used in this study to create the phylogenetic trees.

<i>B. burgdorferi</i> strain	source	location		<i>B. burgdorferi</i> strain	source	location
B31	<i>I. scapularis</i>	Shelter Island, NY		15912UT	<i>I. scapularis</i>	Cumberland County, ME
297	human patient	Connecticut		519014UT	<i>I. scapularis</i>	Van Buren County, MI
JD1	<i>I. scapularis</i>	Ipswich, MA		48102UT	<i>I. scapularis</i>	Hubbard County, MN
N40	<i>I. scapularis</i>	Westchester County, NY		498801UT	<i>I. scapularis</i>	Suffolk County, NY
BL206	human patient	Westchester County, NY		114311UT	<i>I. scapularis</i>	Cecil County, MD
B515	human patient	Westchester County, NY		Ca4	<i>I. pacificus</i>	California
B504	human patient	Westchester County, NY		Ca5	<i>I. pacificus</i>	California
B509	human patient	Westchester County, NY		Ca6	<i>I. pacificus</i>	California
MR623	human patient	Westchester County, NY		Ca92–0953	human patient	California
B373	human patient	Westchester County, NY		Ca92–1096	human patient	California
B156	human patient	Westchester County, NY		Ca92–1337	human patient	California
MR661	human patient	Westchester County, NY		CaWTB27	<i>I. pacificus</i>	California
MR654	human patient	Westchester County, NY		CaWTB32	<i>I. pacificus</i>	California
BL538	human patient	Westchester County, NY		IPT2	<i>I. ricinus</i>	Alsace, France
BL515	human patient	Westchester County, NY		IPT19	<i>I. ricinus</i>	Alsace, France
BL522	human patient	Westchester County, NY		IPT23	<i>I. ricinus</i>	Alsace, France
B356	human patient	Westchester County, NY		IPT39	<i>I. ricinus</i>	Alsace, France
MR616	human patient	Westchester County, NY		IPT58	<i>I. ricinus</i>	Alsace, France
B331	human patient	Westchester County, NY		IPT69	<i>I. ricinus</i>	Alsace, France
B361	human patient	Westchester County, NY		IPT135	<i>I. ricinus</i>	Auvergne, France
B500	human patient	Westchester County, NY		IPT137	<i>I. ricinus</i>	Alsace, France
B485	human patient	Westchester County, NY		IPT190	<i>I. ricinus</i>	Normandy, France
MR607	human patient	Westchester County, NY		IPT191	<i>I. ricinus</i>	Normandy, France
MR662	human patient	Westchester County, NY		IPT193	<i>I. ricinus</i>	Normandy, France
B418	human patient	Westchester County, NY		IPT198	<i>I. ricinus</i>	Normandy, France
MR640	human patient	Westchester County, NY		NE49	<i>I. ricinus</i>	Switzerland
B348	human patient	Westchester County, NY		Z41293	<i>I. ricinus</i>	Germany
47703UT	<i>I. scapularis</i>	Cass County, MN		Z41493	<i>I. ricinus</i>	Germany
51405UT	<i>I. scapularis</i>	Cass County, MN		21509LT	<i>I. ricinus</i>	Babite, Latvia
16812UT	<i>I. scapularis</i>	Cumberland County, ME		20111LT	<i>I. ricinus</i>	Babite, Latvia

15903UT	<i>I. scapularis</i>	Cumberland County, ME		22521LT	<i>I. ricinus</i>	Babite, Latvia
15506UT	<i>I. scapularis</i>	Cumberland County, ME		20604LT	<i>I. ricinus</i>	Kumeri, Latvia

Table 5: A list of the *recG* allele types and the *B. burgdoferi* strains that are associated with each type.

<i>recG</i> allele Types	<i>B.burgdorferi</i> strains
1	B31, BL206, B515, 16812UT, Ca4, Ca5, Ca6, JD1, BL538, BL515, 114311UT, JD1, BL538, BL515, 114311UT, IPT137, 51405UT
2	B500, B331, B361
3	48102UT, B418, N40, MR640, B348, 15903UT
4	BL522, B356, Ca92-1337
5	B156
6	MR616, 15912UT, Ca92-1337, MR661, MR654, 15506UT, B509, 297, B504, 498801UT
7	MR623, B373
8	Ca92-1096
9	NE49, IPT193, IPT198
10	IPT19, 21509LT
11	IPT39, IPT58
12	IPT135
13	Z41293
14	Z41493
15	47703UT
16	519014UT
17	20111LT

Table 6. A list of the *uvrA* allele types and the *B. burgdoferi* strains that are associated with each type.

<i>uvrA</i> Strain Type	<i>B.burgdorferi</i> strains
1	B31,BI206, B515, 16812UT
2	B500, B361, 47703UT
3	B418, N40, MR640, B348, 15903UT
4	Ca92-1337, BL522, B356
5	B485, MR607, MR662
6	B156, MR661, MR654, 15506UT, Ca WTB32
7	297, B504, 498801UT, B509
8	MR623, B373
9	JD1, BL538, BL515, 114311UT
10	MR616, 15912UT, IPT2, IPT23, IPT69, IPT191, IPT190, 22521LT, IPT135, IPT2, IPT23, IPT69, IPT191, IPT190, 22521LT,IPT137
11	Ca4, Ca5, Ca6
12	Ca92-0953
13	Ca92-1096
14	CaWTB27
15	IPT193, IPT198,NE49
16	20111LT, IPT39, IPT58
17	Z41293
18	Z41493

Figure 2. *recG* phylogenic tree with positive samples from this study (NY08, NY16, CC71A, NY20, NY07, and NY27), the MLST paper *B. burgdorferi* *recG* allele types, and other *Borrelia* spp. *recG* sequences. Our samples are indicated with red arrows.

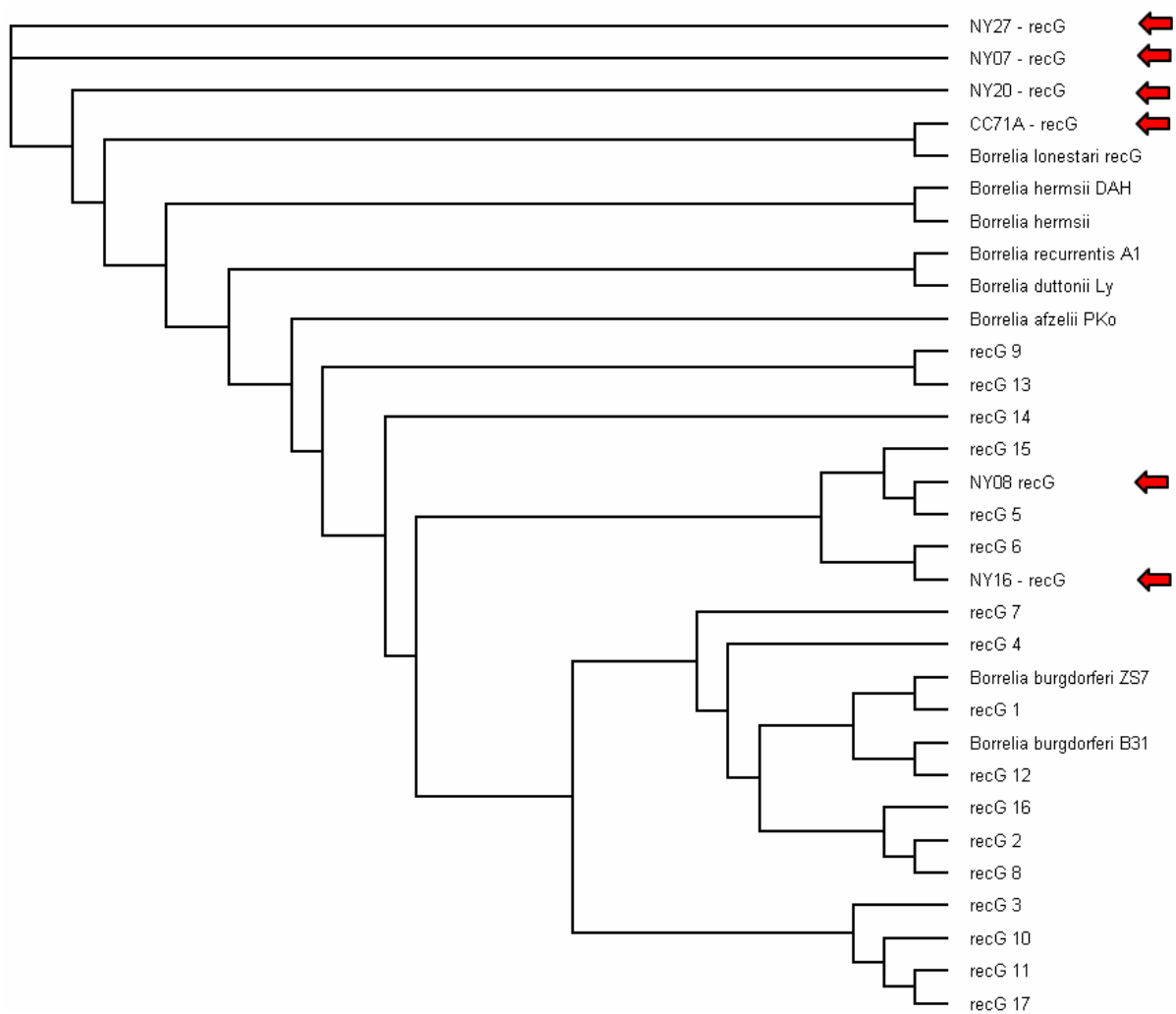


Figure 3. *uvrA* phylogenetic tree with positive samples from this study (NY16, NY08, and CC71A), the MLST paper *uvrA* allele types, as well as other *Borrelia* spp. *uvrA* sequences.

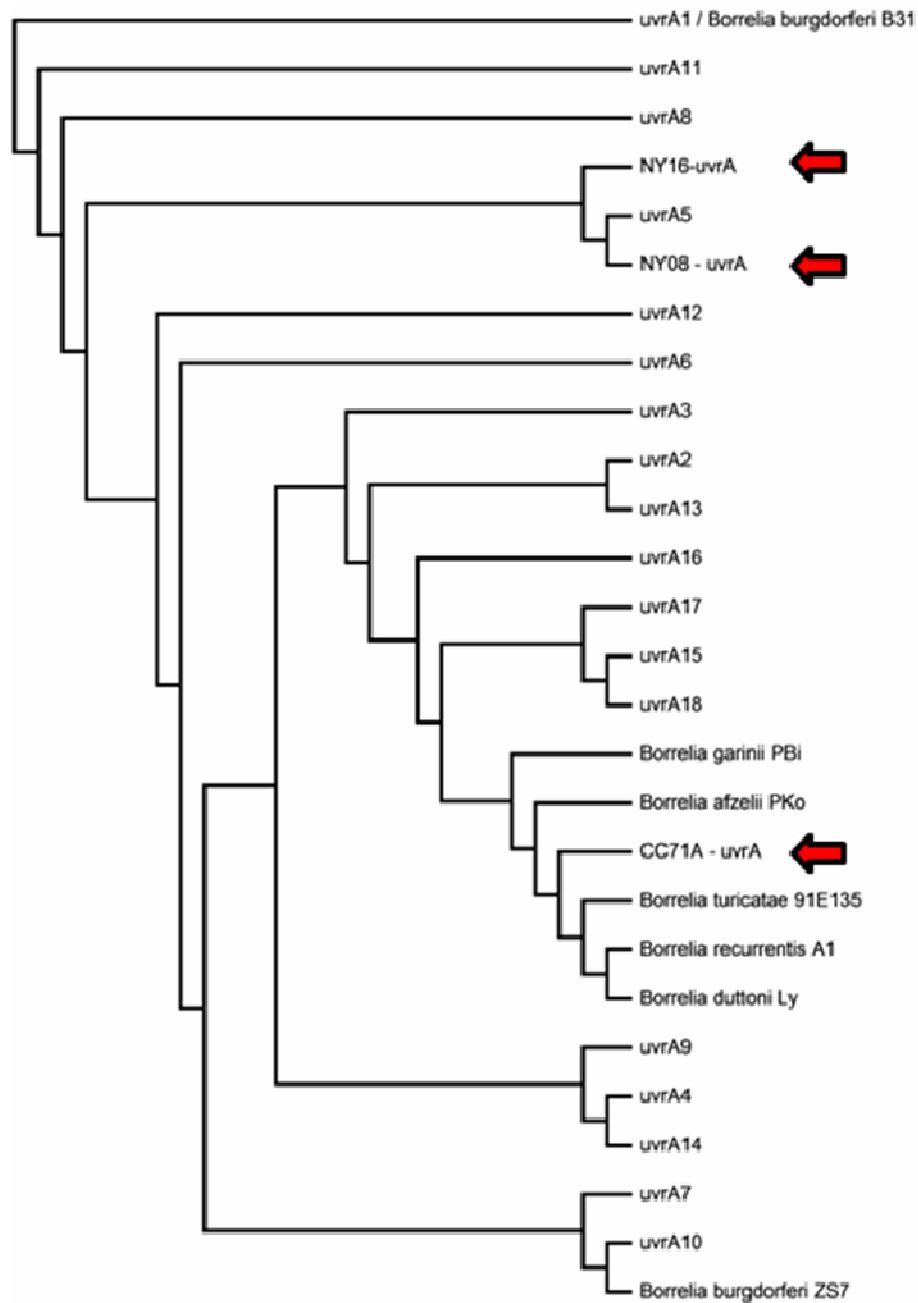
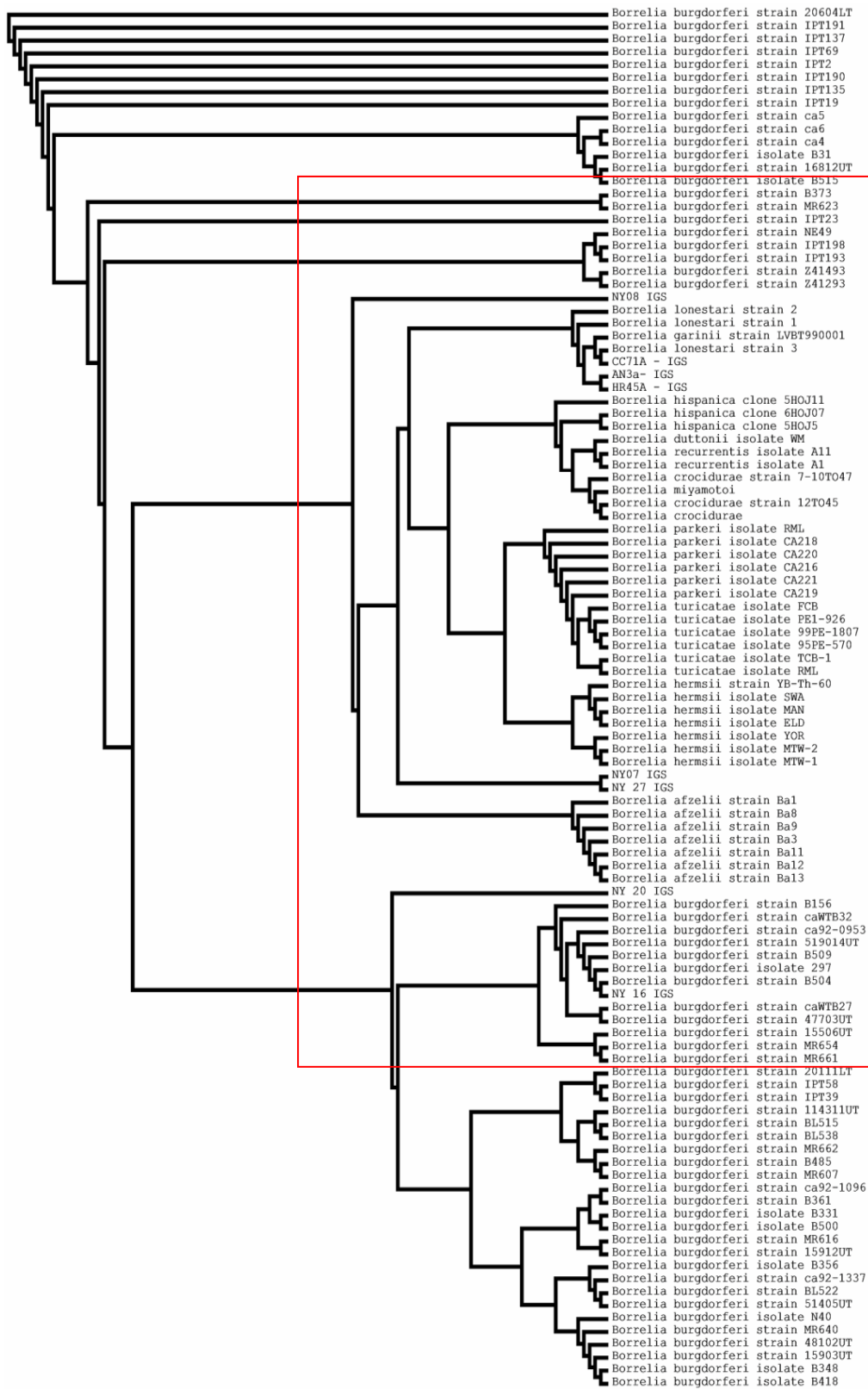
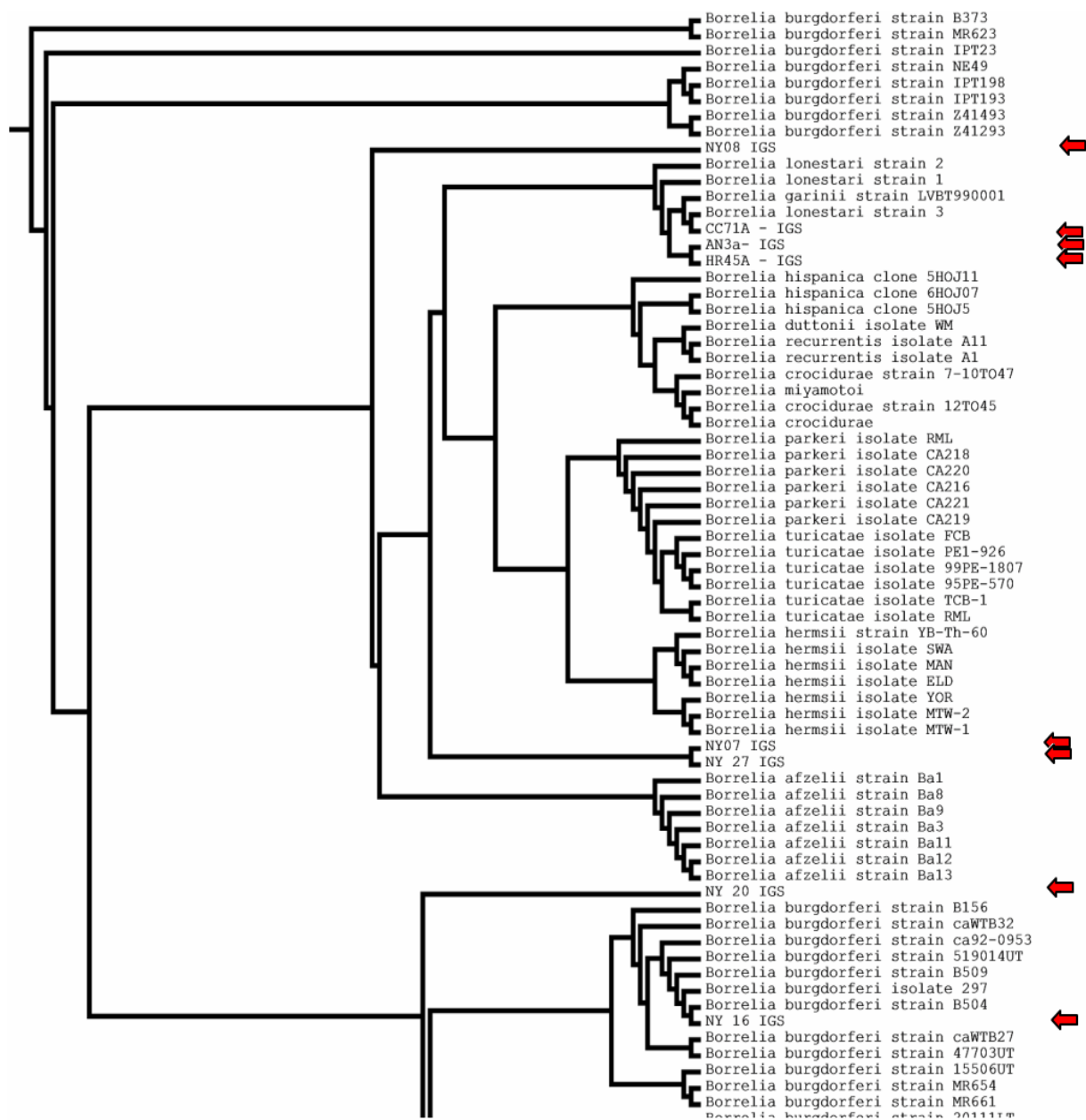


Figure 4. 16S -23S Intergenic Spacer phylogenetic tree. Among *B. burgdorferi* strains, the 16S-23S Intergenic Spacer is far more variable than *uvrA* or *recG* genes. Therefore, no allele groups were found in the MLST paper. Multiple allele types exist for other *Borrelia* species and have been included in this phylogenetic tree. The whole tree was shown, then part of the tree was shown to better illustrate where our samples grouped.





In addition to the 454 Sequencing, RTL conducted a qPCR test using *Borrelia* specific 16S rDNA primers in order to detect the presence or absence of *Borrelia* species (Table 7). Positive samples were considered to be before 35 cycles. 19 tick genomic DNA samples were considered positive with an addition two samples considered late positive (detection after >34 cycles); the late positive specimens are not included in Table 7.

There were discrepancies between the samples that were tested using *Borrelia* spp. specific primers and the positive results from the qPCR reaction conducted at RTL. qPCR at RTL gave 19 positive results for the possible 147 samples, or 12.9% of 147 tested samples. Our PCR tests of these 147 samples using *Borrelia* IGS, *uvrA* and *recG* sequences gave us eight positive samples. Seven of these specimens were also positive in the qPCR results. The remaining 12 specimens that were positive by the RTL qPCR were negative for all three primer sets in our PCR studies. It is unclear as to why this occurred. The primers used by RTL were 16S rDNA primers, which might have amplified species very similar to *Borrelia*; alternatively, the primers used in the qPCR reaction may be able to detect a yet unknown species of *Borrelia* that were not detected in our studies. Lastly, the incongruence may simply reflect differences in sensitivity.

Table 7. *Borrelia* species 16S rDNA qPCR. B.burg PC is *Borrelia burgdorferi* positive control. Less than 35 cycles was considered a positive result, however, positive results close to 35 were considered questionable and labeled as "Late Positive".

Sample	Cycle Positive	Result
MO 87	33.41	Positive
NY16	28.04	Positive
NY22	32.97	Positive
SR01 A	33.87	Positive
SR01 D	33.74	Positive
NY07	28.23	Positive
NY08	29.24	Positive
NY09	30.02	Positive
MO 84	31.67	Positive
MO 85	33.39	Positive
HR 86B	32.73	Positive
McDR 22	34.59	Late Positive
NY 20	30.72	Positive
MO 78	28.77	Positive
HR 72	33.1	Positive
HR 74	32.71	Positive
HR 75	33.85	Positive
HR 81C	32.38	Positive
HR 84	32.78	Positive
MO 94	34.56	Late Positive
MO 97	33.85	Positive
B burg PC	29.99	Positive

454 pilot run at Baylor College of Medicine

Elimination of *Coxiella* endosymbiont sequences proved difficult and time consuming. 454 massively parallel sequencing was seen as an alternative to normal 16S rDNA amplification and cloning. In the normal 16S rDNA amplicon cloning reaction, roughly 2% of the clones were non-*Coxiella* endosymbiont sequences. One 454 sequencing reaction generates up to 1 million reads. Even if 98% of the sequences were *Coxiella* sequences, 454 could generate 20,000 non-*Coxiella* reads. Rather than try to eliminate the *Coxiella* endosymbiont sequences in hope of finding unique sequences, 454 sequencing would attempt to sequence all bacterial 16S rDNA sequences present in a tick.

The initial 454 pilot run used only Missouri ticks as template DNA. From the 36 Missouri tick genomic DNA extractions, 55 amplicons were created. Samples were grouped and tagged by hypervariable regions which created three sample sets; V1-V3 samples, V3-V5 samples and V6-V9 samples. Before sending the samples to be sequenced by Baylor College of Medicine, five samples were Sanger sequenced to ensure bacterial sequences were present. All five sequences were either the highly prevalent *Coxiella* endosymbiont or the prevalent *Rickettsia* endosymbiont species found in *Amblyomma americanum* ticks. The 55 amplicons were sent to Baylor College of Medicine for 454 sequencing but unfortunately, 454 sequencing gave less than optimal results. Although 454 sequencing is capable of generating 1 million reads at 400 to 500 basepairs per read, our samples did not achieve the read length of 400 bp and instead were much shorter. The 454 sequencing reaction was repeated but the quality of the results did not improve.

454 Sequencing Results from Research and Testing Laboratory

In pursuit of the 454 sequencing approach, a set of tick samples were sent to Dr. Scot Dowd at the Research and Testing Laboratory (RTL) in Lubbock, Texas. In this case, samples were amplified at RTL using degenerate primers specific for the V1 to V3 regions of bacterial 16S rDNA. In addition, the primer sets contained barcode sequences that identified the rDNA sequences associated with each tick specimen. DNA prepared from 133 ticks from Missouri, New York, and Texas was used in this sequencing run. In addition to the 133 genomic DNA samples, 14 tick halves were also sent to have their genomic DNA extracted at RTL before 454 sequencing, bringing the total number of genomic DNA samples to 147. Of the 147 genomic DNA samples, 129 were selected for 454 sequencing.

The 129 samples created a total of 553,804 reads; an average of 4,293 reads per tick. Table 6 illustrates the average number of genera per group of ticks. On average, ticks processed at UT-Houston averaged 30.3 genera per tick, which was much lower than the samples processed at RTL, which averaged 133 genera per tick. Missouri *A. americanum* ticks, average had an overall higher number of genera per tick, with 38 genera per tick, than their Texas counterparts which had 19.5 genera per tick. The same is true of New York *I. scapularis*, averaging 36.7 genera per tick, to the 20 genera per tick seen in their Texas counter parts. In all 803 genera and species of bacteria were found, of which 208 genera and species were unique to ticks processed at RTL and 251 genera and species unique to tick samples processed at UT-Houston.

For the second 454 run, a total of 89 *Borrelia* spp. sequences came from three samples: 83 sequences from *I. scapularis* sample NY-07, four sequences from Missouri *A. americanum* sample MO-78 and one sequence from sample Missouri *A. americanum* sample MO-94. A BLAST search indicated that 88 of the 89 sequences were most related to *Borrelia burgdorferi*

but were not the common laboratory strain B31 [15]. The remaining sequence, from sample MO-94, appeared to be *B. burgdorferi* strain B31 contamination as it aligns perfectly with B31 once 454 sequencing errors were accounted for and corrected.

Propionibacterium spp. (present in 74.42% of tick samples), *Coxiella* spp. (70.54%), *Rickettsia* spp. (68.99%), *Acinetobacter* spp. (62.02%) and *Pseudomonas* spp. (59.69%) were the most prevalent genera among all ticks (Table 2). When comparing the results obtained with individual ticks, however, there was remarkable variation in the prevalence of the bacterial genera found in individual tick. For example, *Rickettsia* accounted for 99.3% of the sequenced amplicons for an individual tick while being absent in another sample. Similarly, the percentage of *Coxiella* sequences in individual tick samples would range from 0% to 99.2% (Figure 2). *Propionibacterium* sequences accounted for as little as 0% or as high as 25.3% of the sequences in individual tick samples. Furthermore, *Propionibacterium* sequences represented less than 2% of all sequences in 68 of the 96 tick samples that tested positive for *Propionibacterium*. *Pseudomonas* accounted for 0% to 29% of total sequences in individual ticks. *Acinetobacter* represented 0% to 19% of total sequences in individual samples. On average, 26.3% of total sequences in all ticks were represented by genera that, individually, accounted for less than 1% of total sequences.

The sequencing data indicated a difference in the microflora of different tick species. 14 genera or species, most notably *Brevibacterium* spp., *Rickettsia* spp. and *Flavobacterium* spp. were more prevalent in the *I. scapularis* tick samples than compared to *A. americanum* tick samples (Table 2). Thirty-seven genera or species were more prevalent in all *A. americanum* than in all *I. scapularis* samples. For example, *Coxiella* spp., *Devosia* spp., *Microbacterium* spp., and *Exiguobacterium* spp. were far more prevalent in *A. americanum* than in *I. scapularis* ticks. Five genera (*Lachnobacterium*, *Sporichthya*, *Micrococcus*,

Geobacter and *Stella*) and one species (*Fluviicola taffensis*) were found exclusively in *A. americanum* (Table 2).

These data also indicated a significant difference in the prevalence of genera and species of tick species collected from different geographical locations. Forty-six genera and species showed a significant difference in prevalence between *I. scapularis* samples found in New York versus Texas. *Brevibacterium* spp., *Flavobacterium* spp., *Escherichia coli* and *Tsukamurella* spp. were far more prevalent in New York *I. scapularis* ticks than in *I. scapularis* from Texas (Table 3). As an example, *Tsukamurella* spp. were present in over 50% of New York *I. scapularis* ticks but were not detected in Texas *I. scapularis* ticks.

In the *A. americanum* samples, 39 genera and species were more prevalent in Missouri *A. americanum* specimens than in Texas *A. americanum* (Table 3). Conversely, 83 genera and species were more prevalent in Texas *A. americanum* than Missouri *A. americanum* (Table 3). *Fluviicola* spp., *Sporichthya* spp., *Candidatus Planktoluna*, *Candidatus Rhodoluna* and *Kineosporia* spp. were present in 25% of Texas *A. americanum* ticks but completely absent in the Missouri samples. Furthermore, *Cryocola* spp., *Dermabacter* spp., *Subdoligranulum* spp. and *Anaeromyxobacter* spp. were found in over 20% of Texas *A. americanum* samples but absent in the Missouri *A. americanum* samples (Table 4). The opposite was true of *Actinomyces* spp. and *Tepidimonas* spp., which were present in over 20% of Missouri *A. americanum* ticks but not found in the Texas *A. americanum* samples (Table 4).

Table 8. The 20 bacterial genera/species most consistently detected in a tick species or geographical group. These tables indicate the presence of a particular genera or species, not its prevalence within each tick group. Note that the top 20 bacterial genera in the *A. cajennesse* ticks were present in all *A. cajennesse* samples.

Genus/Species	Percent of all ticks containing each bacterial group		Genus/Species	Percent of <i>I. scapularis</i> containing each bacterial group
<i>Propionibacterium</i> spp.	74.42		<i>Rickettsia</i> spp.	80.00
<i>Coxiella</i> spp.	70.54		<i>Propionibacterium</i> spp.	75.00
<i>Rickettsia</i> spp.	68.99		<i>Stenotrophomonas</i> spp.	68.33
<i>Acinetobacter</i> spp.	62.02		<i>Acinetobacter</i> spp.	63.33
<i>Pseudomonas</i> spp.	59.69		<i>Pseudomonas</i> spp.	61.67
<i>Stenotrophomonas</i> spp.	56.59		<i>Acidovorax</i> spp.	55.00
<i>Acidovorax</i> spp.	55.04		<i>Coxiella</i> spp.	51.67
<i>Mycobacterium</i> spp.	48.84		<i>Brevibacterium</i> spp.	46.67
<i>Staphylococcus</i> spp.	46.51		<i>Escherichia coli</i>	46.67
<i>Escherichia coli</i>	45.74		<i>Flavobacterium</i> spp.	46.67
<i>Bacillus</i> spp.	44.19		<i>Bacillus</i> spp.	45.00
<i>Sphingomonas</i> spp.	42.64		<i>Hydrogenophilus</i> spp.	45.00
<i>Corynebacterium</i> spp.	42.64		<i>Mycobacterium</i> spp.	43.33
<i>Diaphorobacter</i> spp.	37.21		<i>Diaphorobacter</i> spp.	43.33
<i>Flavobacterium</i> spp.	36.43		<i>Derxia</i> spp.	40.00
<i>Chryseobacterium</i> spp.	35.66		<i>Corynebacterium</i> spp.	40.00
<i>Streptococcus</i> spp.	35.66		<i>Brevundimonas</i> spp.	38.33
<i>Brevundimonas</i> spp.	34.88		<i>Shigella</i> spp.	35.00
<i>Bradyrhizobium</i> spp.	33.33		<i>Staphylococcus</i> spp.	35.00
<i>Shigella</i> spp.	33.33		<i>Serratia</i> spp.	35.00

Genus/Species	Percent of <i>I. scapularis</i> in New York containing each bacterial group		Genus/Species	Percent of <i>I. scapularis</i> in Texas containing each bacterial group
<i>Stenotrophomonas</i> spp.	85.29		<i>Rickettsia</i> spp.	76.92
<i>Rickettsia</i> spp.	82.35		<i>Propionibacterium</i> spp.	65.38
<i>Propionibacterium</i> spp.	82.35		<i>Mycobacterium</i> spp.	61.54
<i>Brevibacterium</i> spp.	76.47		<i>Pseudomonas</i> spp.	53.85
<i>Acinetobacter</i> spp.	76.47		<i>Coxiella</i> spp.	46.15
<i>Flavobacterium</i> spp.	73.53		<i>Acinetobacter</i> spp.	46.15
<i>Escherichia coli</i>	70.59		<i>Stenotrophomonas</i> spp.	46.15
<i>Acidovorax</i> spp.	70.59		<i>Hydrogenophilus</i> spp.	42.31
<i>Pseudomonas</i> spp.	67.65		<i>Diaphorobacter</i> spp.	38.46
<i>Bacillus</i> spp.	58.82		<i>Corynebacterium</i> spp.	38.46
<i>Coxiella</i> spp.	55.88		<i>Bradyrhizobium</i> spp.	34.62
<i>Tsukamurella</i> spp.	52.94		<i>Acidovorax</i> spp.	34.62
<i>Shigella</i> spp.	52.94		<i>Achromobacter</i> spp.	30.77
<i>Derxia</i> spp.	50.00		<i>Staphylococcus</i> spp.	30.77
<i>Hydrogenophilus</i> spp.	47.06		<i>Derxia</i> spp.	26.92
<i>Escherichia</i> spp.	47.06		<i>Bacillus</i> spp.	26.92
<i>Diaphorobacter</i> spp.	47.06		<i>Brevundimonas</i> spp.	26.92
<i>Brevundimonas</i> spp.	47.06		<i>Sphingopyxis</i> spp.	26.92
<i>Sphingomonas</i> spp.	41.18		<i>Serratia</i> spp.	26.92
<i>Serratia</i> spp.	41.18		<i>Mitsuaria</i> spp.	26.92

Genus/Species	Percent of all <i>A. americanum</i> containing each bacterial group		Genus/Species	Percent of <i>A. americanum</i> in Missouri containing each bacterial group
<i>Coxiella</i> spp.	88.71		<i>Coxiella</i> spp.	88.24
<i>Propionibacterium</i> spp.	75.81		<i>Propionibacterium</i> spp.	88.24
<i>Rickettsia</i> spp.	58.06		<i>Rickettsia</i> spp.	64.71
<i>Acinetobacter</i> spp.	58.06		<i>Acinetobacter</i> spp.	64.71
<i>Mycobacterium</i> spp.	56.45		<i>Sphingomonas</i> spp.	61.76
<i>Pseudomonas</i> spp.	53.23		<i>Staphylococcus</i> spp.	58.82
<i>Staphylococcus</i> spp.	51.61		<i>Acidovorax</i> spp.	55.88
<i>Sphingomonas</i> spp.	50.00		<i>Mycobacterium</i> spp.	52.94
<i>Acidovorax</i> spp.	50.00		<i>Pseudomonas</i> spp.	52.94
<i>Stenotrophomonas</i> spp.	43.55		<i>Streptococcus</i> spp.	52.94
<i>Escherichia coli</i>	43.55		<i>Chryseobacterium</i> spp.	47.06
<i>Corynebacterium</i> spp.	43.55		<i>Escherichia coli</i>	47.06
<i>Chryseobacterium</i> spp.	40.32		<i>Bacillus</i> spp.	47.06
<i>Bacillus</i> spp.	40.32		<i>Prevotella</i> spp.	47.06
<i>Streptococcus</i> spp.	40.32		<i>Stenotrophomonas</i> spp.	44.12
<i>Prevotella</i> spp.	35.48		<i>Corynebacterium</i> spp.	44.12
<i>Microbacterium</i> spp.	33.87		<i>Microbacterium</i> spp.	41.18
<i>Bradyrhizobium</i> spp.	32.26		<i>Novosphingobium</i> spp.	41.18
<i>Roseburia</i> spp.	32.26		<i>Rikenella</i> spp.	38.24
<i>Diaphorobacter</i> spp.	32.26		<i>Exiguobacterium</i> spp.	35.29

Genus/Species	Percent of <i>A. americanum</i> in Texas containing each bacterial group		Species/Genus	Percent of all <i>A.cajennense</i> containing each bacterial group
<i>Coxiella</i> spp.	89.29		<i>Dermabacter</i> spp.	100.00
<i>Mycobacterium</i> spp.	60.71		<i>Faecalibacterium</i> spp.	100.00
<i>Propionibacterium</i> spp.	60.71		<i>Zoogloea</i> spp.	100.00
<i>Pseudomonas</i> spp.	53.57		<i>Clostridium</i> spp.	100.00
<i>Rickettsia</i> spp.	50.00		<i>Fluviicola</i> spp.	100.00
<i>Acinetobacter</i> spp.	50.00		<i>Chryseobacterium</i> spp.	100.00
<i>Roseburia</i> spp.	46.43		<i>Flavobacterium</i> spp.	100.00
<i>Stenotrophomonas</i> spp.	42.86		<i>Sporichthya</i> spp.	100.00
<i>Staphylococcus</i> spp.	42.86		<i>Staphylococcus</i> spp.	100.00
<i>Acidovorax</i> spp.	42.86		<i>Kineosporia</i> spp.	100.00
<i>Sphingopyxis</i> spp.	42.86		<i>Arcicella</i> spp.	100.00
<i>Corynebacterium</i> spp.	42.86		<i>Acidovorax</i> spp.	100.00
<i>Escherichia coli</i>	39.29		<i>Pseudomonas</i> spp.	100.00
<i>Faecalibacterium</i> spp.	35.71		<i>Curvibacter</i> spp.	100.00
<i>Sphingomonas</i> spp.	35.71		<i>Candidatus Planktoluna</i>	100.00
<i>Ruminococcus</i> spp.	35.71		<i>Leptothrix</i> spp.	100.00
<i>Bacteroides</i> spp.	35.71		<i>Fluviicola taffensis</i>	100.00
<i>Herminiimonas</i> spp.	32.14		<i>Candidatus Rhodoluna</i>	100.00
<i>Bradyrhizobium</i> spp.	32.14		<i>Aquabacterium</i> spp.	100.00
<i>Chryseobacterium</i> spp.	32.14		<i>Terrimonas</i> spp.	100.00

Table 9. Three tables indicating differences in prevalence between all *I. scapularis* ticks and all *A. americanum* ticks; between New York and Texas *I. scapularis* ticks; and between Texas and Missouri *A. americanum* ticks. Cut off for significant difference was set at a 10% difference between two groups.

Difference of >10% in Genus or Species Prevalence of <i>I. scapularis</i> from New York and Texas			
Species/Genus	Percent of <i>I. scapularis</i> in New York	Percent of <i>I. scapularis</i> in Texas	Difference
<i>Brevibacterium</i> spp.	76.47	7.69	68.78
<i>Flavobacterium</i> spp.	73.53	11.54	61.99
<i>Escherichia coli</i>	70.59	15.38	55.20
<i>Tsukamurella</i> spp.	52.94	0.00	52.94
<i>Shigella</i> spp.	52.94	11.54	41.40
<i>Stenotrophomonas</i> spp.	85.29	46.15	39.14
<i>Acidovorax</i> spp.	70.59	34.62	35.97
<i>Phenylobacterium</i> spp.	41.18	7.69	33.48
<i>Bacillus</i> spp.	58.82	26.92	31.90
<i>Acinetobacter</i> spp.	76.47	46.15	30.32
<i>Enterobacter</i> spp.	35.29	7.69	27.60
<i>Firmicutes</i> oral	35.29	7.69	27.60
<i>Rubrivivax</i> spp.	29.41	3.85	25.57
<i>Lysobacter</i> spp.	35.29	11.54	23.76
<i>Deroxia</i> spp.	50.00	26.92	23.08
<i>Burkholderia</i> spp.	29.41	7.69	21.72
<i>Brevundimonas</i> spp.	47.06	26.92	20.14
<i>Arcicella</i> spp.	23.53	3.85	19.68
<i>Janthinobacterium</i> spp.	29.41	11.54	17.87
<i>Methylobacterium</i> spp.	29.41	11.54	17.87
<i>Anaerococcus</i> spp.	17.65	0.00	17.65
<i>Propionibacterium</i> spp.	82.35	65.38	16.97
<i>Tepidimonas</i> spp.	23.53	7.69	15.84
<i>Flavobacterium psychrophilum</i>	14.71	0.00	14.71
<i>Brevibacterium antiquum</i>	14.71	0.00	14.71
<i>Rhodobacter</i> spp.	14.71	0.00	14.71
<i>Erwinia</i> spp.	14.71	0.00	14.71
<i>Rheinheimera</i> spp.	14.71	0.00	14.71
<i>Serratia</i> spp.	41.18	26.92	14.25
<i>Chryseobacterium</i> spp.	29.41	15.38	14.03
<i>Pseudomonas</i> spp.	67.65	53.85	13.80
<i>Rhodoferrax</i> spp.	17.65	3.85	13.80
<i>Papillibacter</i> spp.	23.53	11.54	11.99
<i>Terrimonas</i> spp.	23.53	11.54	11.99
<i>Mycobacterium avium</i>	11.76	0.00	11.76
<i>Staphylococcus epidermidis</i>	11.76	0.00	11.76
<i>Methylophilus</i> spp.	11.76	0.00	11.76
<i>Sphingopyxis</i> spp.	38.24	26.92	11.31
<i>Enhydrobacter</i> spp.	26.47	15.38	11.09
<i>Niastella</i> spp.	26.47	15.38	11.09
<i>Dorea</i> spp.	14.71	3.85	10.86
<i>Pantoea</i> spp.	14.71	3.85	10.86
<i>Curtobacterium</i> spp.	0.00	11.54	-11.54
<i>Rickettsiella</i> spp.	0.00	15.38	-15.38
<i>Mycobacterium</i> spp.	29.41	61.54	-32.13

Difference of >10 in Genus/Species Prevalence in *A. americanum* from Texas and Missouri, ctd.

Species/Genus	Percent of <i>A. americanum</i> in Missouri	Percent of <i>A. americanum</i> in Texas	Difference
<i>Novosphingobium</i> spp.	41.18	3.57	37.61
<i>Paracoccus</i> spp.	32.35	3.57	28.78
<i>Nocardioides</i> spp.	35.29	7.14	28.15
<i>Streptococcus</i> spp.	52.94	25.00	27.94
<i>Propionibacterium</i> spp.	88.24	60.71	27.52
<i>Sphingomonas</i> spp.	61.76	35.71	26.05
<i>Prevotella</i> spp.	47.06	21.43	25.63
<i>Exiguobacterium</i> spp.	35.29	10.71	24.58
<i>Rikenella</i> spp.	38.24	14.29	23.95
<i>Actinomyces</i> spp.	23.53	0.00	23.53
<i>Tepidimonas</i> spp.	20.59	0.00	20.59
<i>Sphingobium</i> spp.	23.53	3.57	19.96
<i>Methylobacterium</i> spp.	29.41	10.71	18.70
<i>Niastella</i> spp.	35.29	17.86	17.44
<i>Microbacterium</i> spp.	41.18	25.00	16.18
<i>Staphylococcus</i> spp.	58.82	42.86	15.97
<i>Chryseobacterium</i> spp.	47.06	32.14	14.92
<i>Bacillus</i> spp.	47.06	32.14	14.92
<i>Rickettsia</i> spp.	64.71	50.00	14.71
<i>Acinetobacter</i> spp.	64.71	50.00	14.71
<i>Haemophilus</i> spp.	14.71	0.00	14.71
<i>Planococcus</i> spp.	14.71	0.00	14.71
<i>Aquabacterium</i> spp.	32.35	17.86	14.50
<i>Acidovorax</i> spp.	55.88	42.86	13.03
<i>Lysobacter</i> spp.	23.53	10.71	12.82
<i>Derxia</i> spp.	23.53	10.71	12.82
<i>Enhydrobacter</i> spp.	26.47	14.29	12.18
<i>Terrimonas</i> spp.	26.47	14.29	12.18
<i>Atopostipes suicloacalis</i>	11.76	0.00	11.76
<i>Agrococcus lahaulensis</i>	11.76	0.00	11.76
<i>Hallella</i> spp.	11.76	0.00	11.76
<i>Leptotrichia</i> spp.	11.76	0.00	11.76
<i>Rothia</i> spp.	11.76	0.00	11.76
<i>Dietzia</i> spp.	14.71	3.57	11.13
<i>Veillonella</i> spp.	14.71	3.57	11.13
<i>Deinococcus</i> spp.	17.65	7.14	10.50
<i>Agrococcus</i> spp.	17.65	7.14	10.50

Difference of >10 in Genus/Species Prevalence in *A. americanum* from Texas and Missouri

Species/Genus	Percent of <i>A. americanum</i> in Missouri	Percent of <i>A. americanum</i> in Texas	Difference
<i>Shigella</i> spp.	35.29	25.00	10.29
<i>Devosia</i> spp.	35.29	25.00	10.29
<i>Stigonema ocellatum</i>	0.00	10.71	-10.71
<i>Fusibacter</i> spp.	0.00	10.71	-10.71
<i>Lactococcus lactis</i>	0.00	10.71	-10.71
<i>Klebsiella pneumoniae</i>	0.00	10.71	-10.71
<i>Alcaligenes</i> spp.	0.00	10.71	-10.71
<i>Enterococcus phoeniculicola</i>	0.00	10.71	-10.71
<i>Nevskia</i> spp.	0.00	10.71	-10.71
<i>Alcanivorax</i> spp.	0.00	10.71	-10.71
<i>Actinoplanes</i> spp.	0.00	10.71	-10.71
<i>Phormidiaceae cyanobacterium</i>	0.00	10.71	-10.71
<i>Clostridium bolteae</i>	0.00	10.71	-10.71
<i>Sanguibacter</i> spp.	0.00	10.71	-10.71
<i>Geitlerinema</i> spp.	0.00	10.71	-10.71
<i>Macrococcus</i> spp.	0.00	10.71	-10.71
<i>Haliscomenobacter</i> spp.	0.00	10.71	-10.71
<i>Cystobacter</i> spp.	0.00	10.71	-10.71
<i>Marinilabilia</i> spp.	0.00	10.71	-10.71
<i>Legionella</i> spp.	2.94	14.29	-11.34
<i>Hoeflea</i> spp.	2.94	14.29	-11.34
<i>Cetobacterium</i> spp.	2.94	14.29	-11.34
<i>Bdellovibrio</i> spp.	2.94	14.29	-11.34
<i>Turicibacter</i> spp.	2.94	14.29	-11.34
<i>Eubacterium hallii</i>	2.94	14.29	-11.34
<i>Pelomonas</i> spp.	5.88	17.86	-11.97
<i>Peptoniphilus</i> spp.	5.88	17.86	-11.97
<i>Geobacter</i> spp.	5.88	17.86	-11.97
<i>Bosea</i> spp.	5.88	17.86	-11.97
<i>Streptacidiphilus</i> spp.	5.88	17.86	-11.97
<i>Janthinobacterium</i> spp.	8.82	21.43	-12.61
<i>Curvibacter</i> spp.	8.82	21.43	-12.61
<i>Hyphomicrobium</i> spp.	11.76	25.00	-13.24
<i>Halochromatium</i> spp.	0.00	14.29	-14.29
<i>Streptococcus agalactiae</i>	0.00	14.29	-14.29
<i>Clostridium bartlettii</i>	0.00	14.29	-14.29
<i>Halomicronema</i> spp.	0.00	14.29	-14.29
<i>Rubrobacter</i> spp.	0.00	14.29	-14.29
<i>Cryobacterium</i> spp.	0.00	14.29	-14.29
<i>Orientia</i> spp.	0.00	14.29	-14.29
<i>Lactococcus</i> spp.	0.00	14.29	-14.29
<i>Edwardsiella</i> spp.	0.00	14.29	-14.29
<i>Magnetospirillum</i> spp.	0.00	14.29	-14.29

Difference of >10 in Genus/Species Prevalence in *A. americanum* from Texas and Missouri

Species/Genus	Percent of <i>A. americanum</i> in Missouri	Percent of <i>A. americanum</i> in Texas	Difference
<i>Anaerostipes</i> spp.	0.00	14.29	-14.29
<i>Cyanothece</i> spp.	0.00	14.29	-14.29
<i>Herminiimonas</i> spp.	17.65	32.14	-14.50
<i>Cetobacterium somerae</i>	2.94	17.86	-14.92
<i>Streptomyces</i> spp.	5.88	21.43	-15.55
<i>Polynucleobacter</i> spp.	5.88	21.43	-15.55
<i>Levilinea</i> spp.	5.88	21.43	-15.55
<i>Zoogloea</i> spp.	8.82	25.00	-16.18
<i>Arthrobacter</i> spp.	8.82	25.00	-16.18
<i>Rhizobium</i> spp.	11.76	28.57	-16.81
<i>Adhaeribacter</i> spp.	0.00	17.86	-17.86
<i>Byssovorax</i> spp.	0.00	17.86	-17.86
<i>Opitutus</i> spp.	0.00	17.86	-17.86
<i>Klebsiella</i> spp.	0.00	17.86	-17.86
<i>Methylophilus</i> spp.	0.00	17.86	-17.86
<i>Anaerococcus vaginalis</i>	0.00	17.86	-17.86
<i>Tannerella</i> spp.	0.00	17.86	-17.86
<i>Bacteroides</i> spp.	17.65	35.71	-18.07
<i>Dorea</i> spp.	2.94	21.43	-18.49
<i>Burkholderia</i> spp.	2.94	21.43	-18.49
<i>Parabacteroides</i> spp.	2.94	21.43	-18.49
<i>Conexibacter</i> spp.	5.88	25.00	-19.12
<i>Cryocola</i> spp.	0.00	21.43	-21.43
<i>Dermabacter</i> spp.	0.00	21.43	-21.43
<i>Subdoligranulum</i> spp.	0.00	21.43	-21.43
<i>Anaeromyxobacter</i> spp.	0.00	21.43	-21.43
<i>Arcicella</i> spp.	2.94	25.00	-22.06
<i>Acidimicrobium</i> spp.	2.94	25.00	-22.06
<i>Pedobacter</i> spp.	5.88	28.57	-22.69
<i>Delftia</i> spp.	5.88	28.57	-22.69
<i>Flavobacterium</i> spp.	8.82	32.14	-23.32
<i>Ruminococcus</i> spp.	11.76	35.71	-23.95
<i>Fluviicola</i> spp.	0.00	25.00	-25.00
<i>Sporichthya</i> spp.	0.00	25.00	-25.00
<i>Candidatus Planktoluna</i>	0.00	25.00	-25.00
<i>Fluviicola taffensis</i>	0.00	25.00	-25.00
<i>Candidatus Rhodoluna</i>	0.00	25.00	-25.00
<i>Sphingopyxis</i> spp.	17.65	42.86	-25.21
<i>Roseburia</i> spp.	20.59	46.43	-25.84
<i>Staphylococcus aureus</i>	5.88	32.14	-26.26
<i>Kineosporia</i> spp.	0.00	28.57	-28.57
<i>Faecalibacterium</i> spp.	5.88	35.71	-29.83

Difference of >10 in Genus/Species Prevalence between <i>I.scapularis</i> and <i>A.americanum</i> Ticks			
Species/Genus	Percent of all <i>I. scapularis</i>	Percent of all <i>A. americanum</i>	Difference
<i>Brevibacterium</i> spp.	46.67	9.68	36.99
<i>Flavobacterium</i> spp.	46.67	19.35	27.31
<i>Hydrogenophilus</i> spp.	45.00	19.35	25.65
<i>Tsukamurella</i> spp.	30.00	4.84	25.16
<i>Stenotrophomonas</i> spp.	68.33	43.55	24.78
<i>Derrxia</i> spp.	40.00	17.74	22.26
<i>Rickettsia</i> spp.	80.00	58.06	21.94
<i>Serratia</i> spp.	35.00	16.13	18.87
<i>Achromobacter</i> spp.	30.00	12.90	17.10
<i>Firmicutes</i> oral	23.33	8.06	15.27
<i>Mitsuaria</i> spp.	31.67	19.35	12.31
<i>Serratia marcescens</i>	21.67	9.68	11.99
<i>Diaphorobacter</i> spp.	43.33	32.26	11.08
<i>Rubrivivax</i> spp.	18.33	8.06	10.27
<i>Papillibacter</i> spp.	18.33	29.03	-10.70
<i>Faecalibacterium</i> spp.	8.33	19.35	-11.02
<i>Staphylococcus epidermidis</i>	6.67	17.74	-11.08
<i>Staphylococcus aureus</i>	6.67	17.74	-11.08
<i>Streptomyces</i> spp.	1.67	12.90	-11.24
<i>Kineosporia</i> spp.	1.67	12.90	-11.24
<i>Levilinea</i> spp.	1.67	12.90	-11.24
<i>Acidimicrobium</i> spp.	1.67	12.90	-11.24
<i>Rhodoplanes</i> spp.	1.67	12.90	-11.24
<i>Lachnobacterium</i> spp.	0.00	11.29	-11.29
<i>Sporichthya</i> spp.	0.00	11.29	-11.29
<i>Micrococcus</i> spp.	0.00	11.29	-11.29
<i>Fluviicola taffensis</i>	0.00	11.29	-11.29
<i>Geobacter</i> spp.	0.00	11.29	-11.29
<i>Bacteroides</i> spp.	13.33	25.81	-12.47
<i>Aquabacterium</i> spp.	13.33	25.81	-12.47
<i>Sporobacter</i> spp.	6.67	19.35	-12.69
<i>Hyphomicrobium</i> spp.	5.00	17.74	-12.74
<i>Stella</i> spp.	0.00	12.90	-12.90
<i>Mycobacterium</i> spp.	43.33	56.45	-13.12
<i>Prevotella</i> spp.	21.67	35.48	-13.82
<i>Rhizobium</i> spp.	5.00	19.35	-14.35
<i>Enterococcus</i> spp.	5.00	19.35	-14.35
<i>Zoogloea</i> spp.	1.67	16.13	-14.46
<i>Conexibacter</i> spp.	0.00	14.52	-14.52
<i>Streptococcus</i> spp.	25.00	40.32	-15.32
<i>Staphylococcus</i> spp.	35.00	51.61	-16.61
<i>Chryseobacterium</i> spp.	23.33	40.32	-16.99
<i>Novosphingobium</i> spp.	6.67	24.19	-17.53
<i>Nocardioide</i> s spp.	5.00	22.58	-17.58
<i>Lactobacillus</i> spp.	6.67	25.81	-19.14

Difference of >10 in Genus/Species Prevalence between <i>I.scapularis</i> and <i>A.americanum</i> Ticks			
Species/Genus	Percent of all <i>I. scapularis</i>	Percent of all <i>A. americanum</i>	Difference
<i>Sphingomonas</i> spp.	30.00	50.00	-20.00
<i>Roseburia</i> spp.	10.00	32.26	-22.26
<i>Exiguobacterium</i> spp.	1.67	24.19	-22.53
<i>Microbacterium</i> spp.	8.33	33.87	-25.54
<i>Devosia</i> spp.	5.00	30.65	-25.65
<i>Coxiella</i> spp.	51.67	88.71	-37.04

Figure 5. The percentage of sequences of all bacterial genera in all *I. scapularis* ticks were averaged. Genera that represented, on average, less than 1% of total sequences were pooled together to create one wedge with the other genera given their own wedge. Again, Genera that account for less than 1% of total sequences was listed first, with the remaining genera listed from highest average percentage to lowest average percentage

Average Percentage of Sequences, All *Ixodes scapularis* specimens

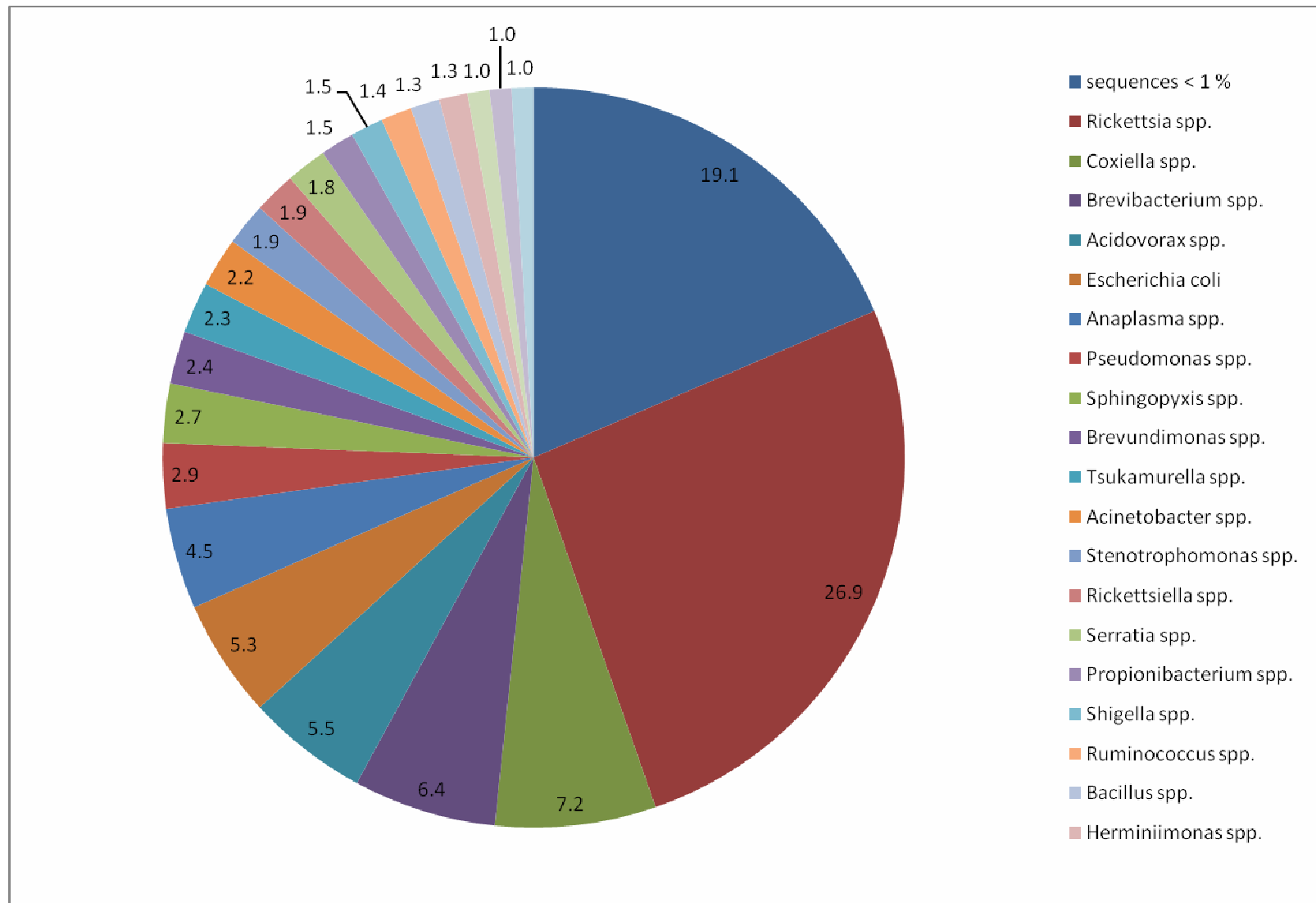


Figure 6. The percentage of sequences of all bacterial genera in all *A. americanum* ticks were averaged. Genera that represented, on average, less than 1% of total sequences were pooled together to create one wedge with the other genera given their own wedge. *Coxiella* spp. made up, on average, over 40% of total sequences per tick. *Rickettsia* spp. also made up a large average percentage of sequences at nearly 20% of total sequences.

Average Percentage of Sequences, All *A. americanum* specimens

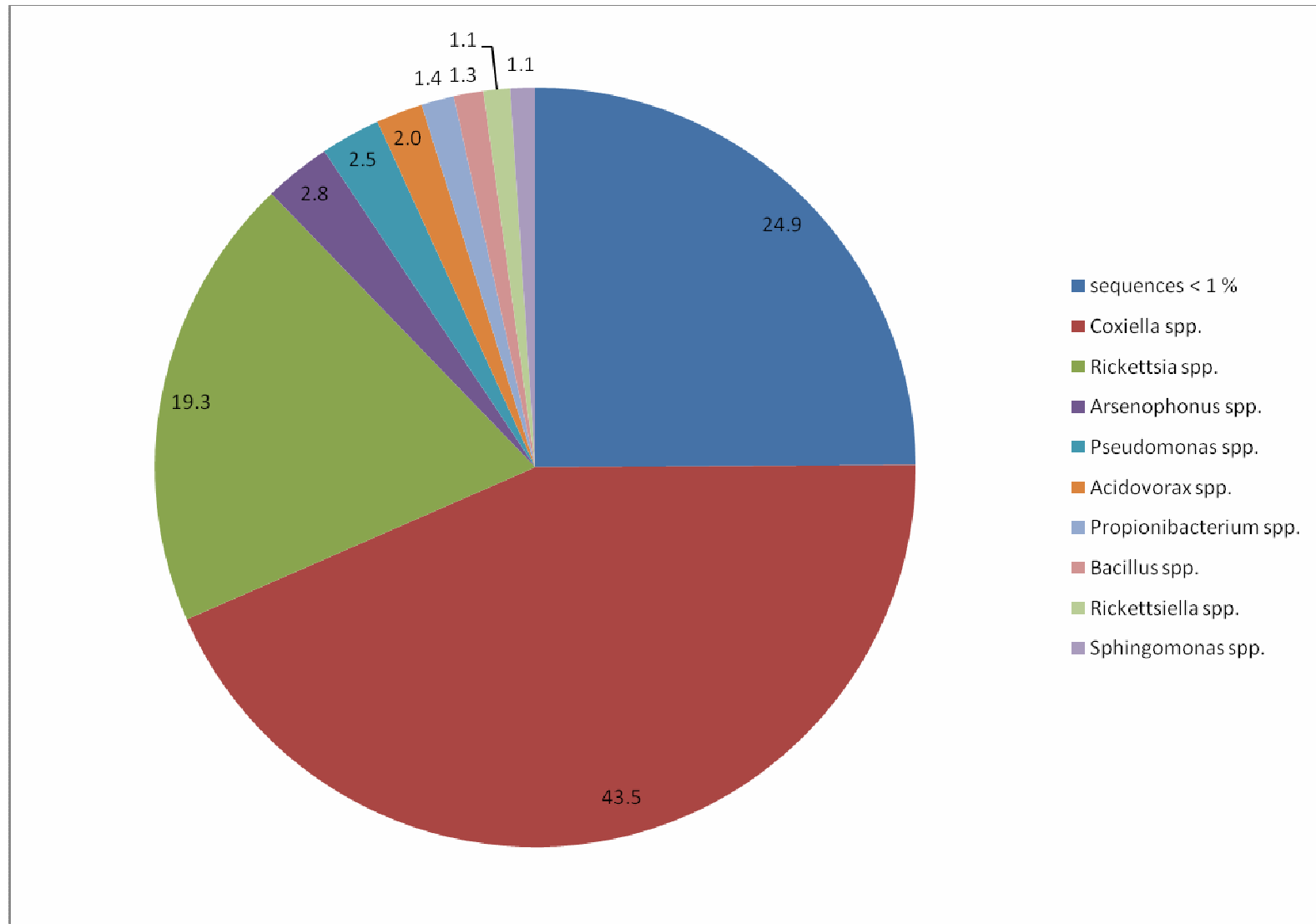


Figure 7. Top five genera in tick groups, in terms of average percent sequence per group. Ticks were grouped together by species and location. The top five genera, in terms of average percent sequence per tick, were determined and charted on a clustered bar graph. Percentages were rounded off to the 10th. The remaining genera were combined and charted as one segment on each bar.

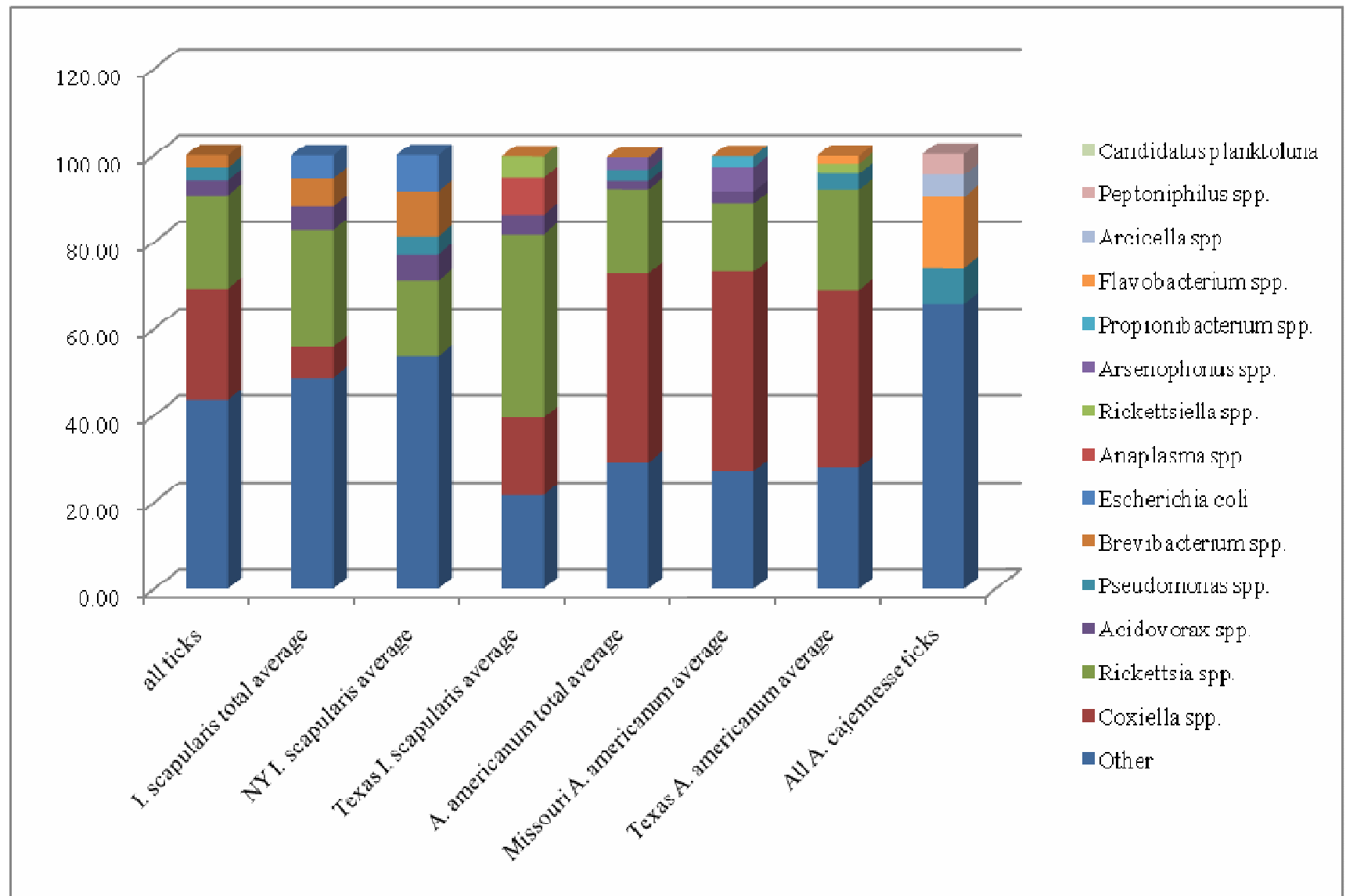
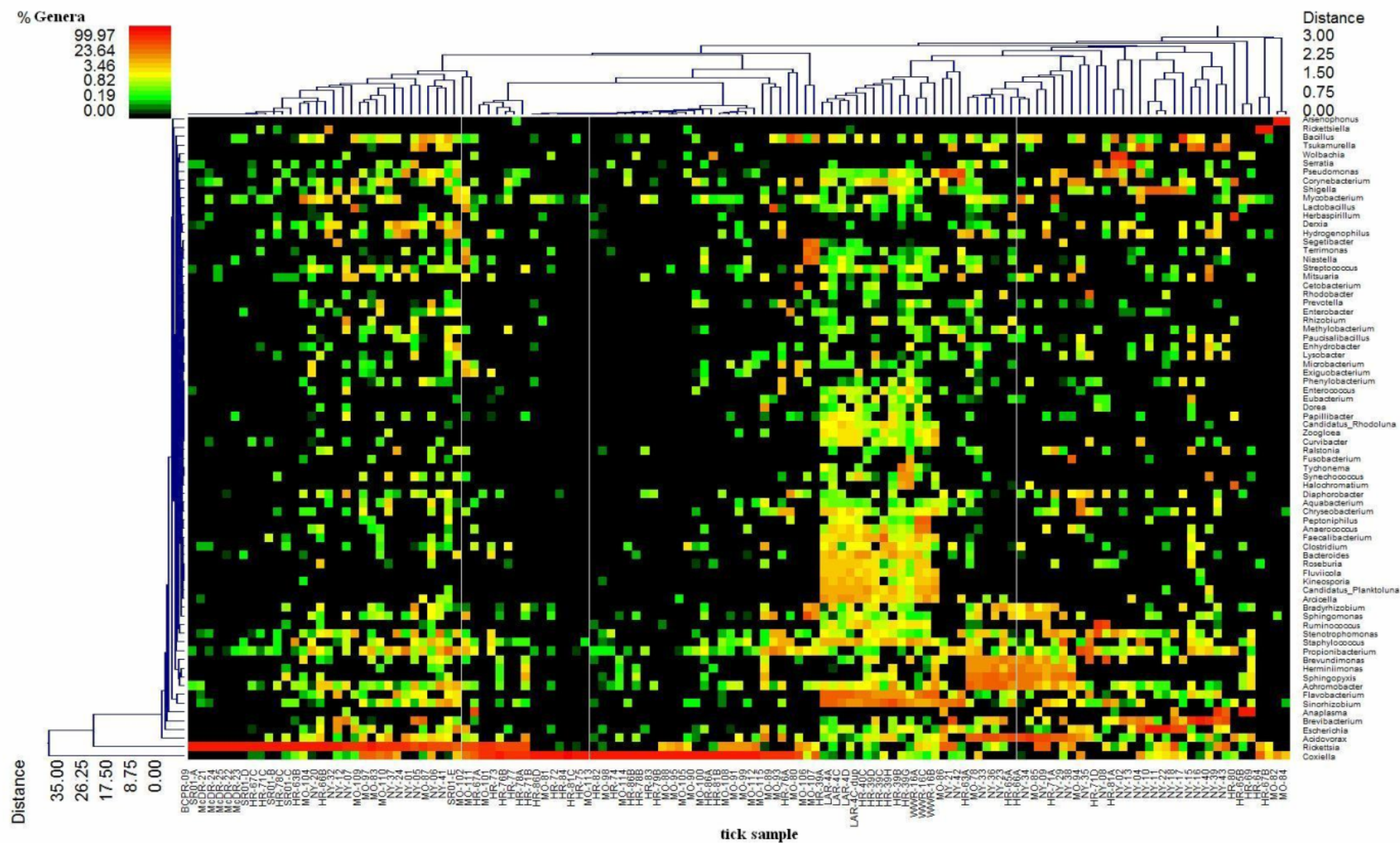


Table 10. Actual average percentage each genera or species represented in the different groups of ticks seen in Figure 1.

Genera	all ticks	<i>I. scapularis</i> total average	NY <i>I. scapularis</i> average	Texas <i>I. scapularis</i> average	<i>A. americanum</i> total Average	Missouri <i>A. americanum</i> average	Texas <i>A. americanum</i> average	All <i>A. cajuputi</i> average
Other	43.54	48.57	53.64	21.62	29.23	27.11	25.02	6.00
<i>Coxiella</i> spp.	25.70	7.22	0.00	17.87	43.54	46.05	40.97	0.00
<i>Rickettsia</i> spp.	21.56	26.90	17.42	42.26	19.34	15.71	23.08	0.00
<i>Acidovorax</i> spp.	3.39	5.47	6.05	4.52	1.98	2.82	0.00	0.00
<i>Pseudomonas</i> spp.	3.03	0.00	4.20	0.00	2.55	0.00	3.71	8.00
<i>Brevibacterium</i> spp.	2.75	6.43	10.39	0.00	0.00	0.00	0.00	0.00
<i>Escherichia coli</i>	2.5	5.31	8.28	0.00	0.00	0.00	0.00	0.00
<i>Anaplasma</i> spp.	2.3	0.00	0.00	8.52	0.00	0.00	0.00	0.00
<i>Rickettsiella</i> spp.	1.4	0.00	0.00	4.92	0.00	0.00	2.29	0.00
<i>Arsenophonus</i> spp.	1.5	0.00	0.00	0.00	2.85	5.60	0.00	0.00
<i>Opionibacterium</i> spp.	1.4	0.00	0.00	0.00	0.00	2.45	0.00	0.00
<i>Flavobacterium</i> spp.	1.7	0.00	0.00	0.00	0.00	0.00	1.77	1.00
<i>Arcicella</i> spp.	0.5	0.00	0.00	0.00	0.00	0.00	0.00	5.00
<i>Peptoniphilus</i> spp.	0.3	0.00	0.00	0.00	0.00	0.00	0.00	4.00
<i>Candidatus planktoluna</i>	0.4	0.00	0.00	0.00	0.00	0.00	0.00	0.00

A dual dendrogram (Figure 3) generated by Dr. Dowd grouped the 129 samples sequenced at RTL by similarity according to percentage of bacterial genera present in each sample. The X-axis indicates tick samples, while the Y-axis indicates bacteria genera. The percentage of total sequences a genus represents is indicated by color, with dark green or green representing less than 1% of total sequences and red presenting up to 99.97% of total sequences. The brackets at the top of the dendrogram indicate the metagenomic difference between individual tick samples: the closer the brackets, the more metagenomically similar. Samples appeared to group together relatively well, however, many ticks were grouped according to whether the sequences were dominated by *Rickettsia* spp. sequences or by *Coxiella* spp. sequences. Trends beyond *Coxiella* spp. or *Rickettsia* spp. dominance were seen. For example, *Ixodes* samples grouped according to geographical location, with 25 of the 35 New York samples grouped together with the other 10 samples grouped with the Texas *Ixodes* samples. For reasons unknown, the two groups of *Ixodes* ticks were the least similar, as they are on opposite sides of the dendrogram. The *Amblyomma americanum* samples loosely grouped together, with little to no visible differences between Missouri *A. americanum* and Texas *A. americanum*. The *A. americanum* ticks that appeared more metagenomically similar to the *Ixodes* tick samples were grouped because of the prevalence of *Rickettsia* spp. sequences.

Figure 8. Dual dendrogram and heat map indicating the relatedness of bacterial content and the proportion of sequences from each bacterial group present in the 133 ticks examined in the RTL analysis. Percentage of total sequences are represented by color, with dark green representing 1% or less of total sequences and bright red representing up to 99.97% of total sequences. Brackets above the dendrogram indicate metagenomic similarity; the closer the brackets, the more metagenomically similar. The brackets to the left of the dendrogram indicates phylogenetic similarity between bacterial genera. The tick names at the bottom of the dendrogram helped to indicate geographic location; NY ticks were New York *I. scapularis* ticks, MO ticks were Missouri *A. americanum* ticks with the remaining ticks coming from Texas. Both the samples from McDonald's ranch and the engorged *Ixodes* tick samples grouped extremely well and were the tightest clustered groups.



Of the frozen tick halves that were processed at RTL before being individually sequenced, seven *A. americanum* and seven *A. cajennense* samples, clustered extremely well despite being different species collected from different locations. Although each of these samples had a relatively high percentage (up to 30%) of *Flavobacterium* spp. and *Pseudomonas* spp. sequences, they lacked a dominant genus, such as *Rickettsia* spp. or *Coxiella* spp. which accounted for over 90% of sequences in many of the samples processed at UT-Houston. Furthermore, these samples appeared to be positive for many more genera and species when compared to samples that were processed at UT-Houston. RTL prepared genomic DNA was, on average, positive for 133 different genera and species while the samples prepared at UT-Houston had an average of 30 positive samples (Table 3). The fourteen frozen samples prepared at RTL had 208 unique genera or species not present in the other 115 samples that were prepared at UT-Houston. The 115 samples prepared at UT-Houston had 251 unique genera or species.

Table 11. The number of bacterial genera found in each tick was averaged and grouped by tick species and location of genomic DNA extraction. Results were compared and genera unique to RTL extracted tick genomic DNA and UT-Houston extracted tick genomic DNA were noted.

Average Number of Genera	Genera
All RTL processed ticks	133.07
RTL processed <i>A. cajenense</i> ticks	119.71
RTL processed <i>A. americanum</i> ticks	146.43
All UT-Houston processed ticks	30.30
UT-Houston processed in all <i>A. americanum</i> ticks	31.15
UT-Houston processed Texas <i>A. americanum</i> ticks	19.52
UT-Houston processed Missouri <i>A. americanum</i> ticks	38.32
UT-Houston processed in all <i>I. scapularis</i> ticks	29.53
UT-Houston processed Texas <i>I. scapularis</i> ticks	20.15
UT-Houston processed New York <i>I. scapularis</i> ticks	36.71
Genera exclusive to all ticks processed at RTL	208.00
Genera exclusive to all ticks processed at UT- Houston	234.00

CHAPTER 4 : DISCUSSION

This study generated a tremendous amount of data on the microflora present in different tick species, as well as the prevalence of *Borrelia* organisms in ticks from Texas, Missouri, and New York. The discussion is focused on the aspects of this information that are considered most important and novel.

Full Length 16S rDNA cloning

The highly prevalent *Coxiella* endosymbiont was first described in 2007, while a subsequent paper published in 2008 details the prevalence of a *Rickettsia* endosymbiont in adult *Amblyomma americanum* [28, 31]. Both papers found the *Coxiella* endosymbiont present in 100% of their tick samples while the *Rickettsia* endosymbiont is present in 40%-60% of all *Amblyomma* ticks [28, 31]. It also appears as if an *Arsenophonus* endosymbiont also exists in *A. americanum* ticks however, when BLAST searched, none of our clones matched with any *Rickettsia* or *Arsenophonus* species despite previous tests showing that our primers would amplify these genera[28]. Due to the dominance of the *Coxiella* endosymbiont sequences, microbes in low abundance were extremely difficult to detect.

These results do not indicate a lack of diversity in the microflora of the tick midgut, but only that the cloning and sequencing of individual 16S rDNA amplicons was insufficient to detect this bacterial diversity. The process of amplification followed by cloning lends itself to bias. Any dominant species, in this case the *Coxiella* endosymbiont, would be more likely to be amplified than minor species. The problem of bias would be further exacerbated in the cloning process, which could lead to the overwhelming number of *Coxiella* 16S rDNA clones.

The nature of the relationship between *Coxiella* endosymbionts and host ticks is unknown. Based on the isolation of specific *A. americanum* tissue types for bacterial 16S rDNA analysis as well fluorescence in situ hybridization, *Coxiella* endosymbionts tend to be

localized, but not limited, to the salivary glands and ovarian tissues of the tick [32]. Colonization in the ovarian tissues is necessary for vertical transmission of the *Coxiella* endosymbiont and may benefit both the bacterium and the tick host. Treatment of *A. americanum* with the antibiotic rifampin and tetracycline in order to eliminate or reduce the *Coxiella* and *Rickettsia* endosymbionts reduces the average weight of *A. americanum* ticks as well as lowering the proportion of ticks hatched from engorged females [33]. Based on this finding, the *Coxiella* endosymbiont could be involved with providing essential nutrients to ticks and possibly to tick oocytes. It is unclear when the *Coxiella* endosymbiont colonizes each organ; however, it is theorized that the endosymbiont colonizes the ovary after the final blood meal [32].

***Borrelia* spp. detection using *Borrelia* specific primers**

Three methods for *Borrelia* spp. detection were used in this study : qPCR, 454 sequencing analysis and the use of *Borrelia* species specific primers. Of the ticks samples used in this study, only the *Ixodes scapularis* tick is a known vector of *Borrelia burgdorferi* [4]. Five New York ticks, or 11.6 % of total New York ticks, were positive for a *Borrelia* species when using *Borrelia* specific primers. Of these five New York ticks, only NY07 was positive for *Borrelia* spp. in both the 454 sequencing run and when used as template DNA for the amplification of *Borrelia* specific housekeeping genes. Our 454 sequencing reaction was able to detect organisms that were present in ticks at a level greater than 10^3 bacteria per tick. However, it is estimated that *I. scapularis* nymphs are infected with 200-2200 *Borrelia* spirochetes per tick, meaning that our 454 sequencing reaction might not be able to detect the presence of *Borrelia* spp. in *I. scapularis* ticks. Because of the relatively low sensitivity of 454 sequencing, it is possible that NY07 carried a very high *Borrelia* load. Of the five positive

ticks, three samples or 6.9% of all New York ticks, appeared to have a strain of *Borrelia burgdorferi*. This is far less than previous studies of *B. burgdorferi* prevalence in New York, which found at least 20% of *I. scapularis* nymphs ticks are infected with *B. burgdorferi*. Co-infection does occur in *I. scapularis* ticks and it appears that, in New York, these strains of *B. miyamotoi* do not infect ticks alone, but co-infect ticks along with another *Borrelia* species at a ratio of 1:10 [30, 34]. Sample NY16 is consistent with this observation, as we obtained both *B. miyamotoi* and *B. burgdorferi* sequences from a single tick. However, three of our tick samples were infected with a non-*B. miyamotoi* relapsing fever group *Borrelia*. In this case, the likely *Borrelia* species appears to be the same unculturable relapsing fever group *Borrelia* described in a previous publication that infects 2.5% of New York *I. scapularis* ticks [35].

Although many other studies found the *B. burgdorferi* infection rate of New York *I. scapularis* nymphs to be 20% or higher [34, 35], our study indicated a lower infection rate of 6.9%. Several factors could explain the difference in infection rate detected. *B. burgdorferi* infection rate of *I. scapularis* ticks depends heavily on host/reservoir activity as well as the environmental factors that directly affect tick activity [36]. *B. burgdorferi* is not vertically transmitted and must be acquired from a mammalian reservoir; therefore host availability is an important factor in *B. burgdorferi* prevalence [37]. Furthermore, some strains of *Borrelia burgdorferi* do not persist in reservoir hosts, creating a short window for a tick to become infected [38]. The seasonal activity of tick nymphs plays a critical role in the infection of *I. scapularis* by *B. burgdorferi*. Questing nymphs are most active during the late summer and early fall months, when potential hosts are readily available and tend to enter diapause during the colder months. All these factors lead to short period of time in which a tick can be infected. The exact date of collection for the New York *I. scapularis* ticks was not known, however, if they were collected in the winter or early spring a lower *B. burgdorferi* infection rate could be

possible. In fact, one study observed as low as 3.2% *B. burgdorferi* infection rate in New York *I. scapularis* which is believed to be caused by collection of the ticks in the winter [39].

This and prior studies indicate that colonization of *A. americanum* ticks by *Borrelia* spp. occurs at relatively low rates. The distribution of the *A. americanum* tick starts in the South Central United States and extends as far Northeast as Maryland. Previous studies report that Kentucky and Virginia *A. americanum* are infected with *B. lonestari* at rates of 5.3% and 1.3%, respectively [39]. In the south, *Borrelia lonestari* was found to infect 3% of the *A. americanum* ticks in Mississippi while 1.03% infection rate by *Borrelia* species was seen in Texas{Bacon, 2003 #94, Rawlings, 1994 #105}. In our studies, three of 199 Texas *A. americanum* ticks examined (1.5 %) was positive for a *Borrelia* species, specifically *B. lonestari*. None of the 36 Missouri *A. americanum* ticks were positive for *Borrelia* species despite reports that up to 5.6% of Missouri ticks are infected with *B. lonestari* [40]. The same concerns of tick seasonal activity, host/reservoir availability and environmental factors could explain the lack of *Borrelia* infection in the Missouri samples. However, with such a low reported incidence of *Borrelia* species infection, sample size could play a factor. 42 Missouri *A. americanum* might simply be too few to accurately determine the incidence of *Borrelia* spp. infection. Our incidence of 1.5% *Borrelia* spp. infection in Texas ticks falls in line with previous reports that found 1.03% of *A. americanum* adults were infected with a species of *Borrelia* [41]. Furthermore, it falls in line with previous reports of *Borrelia* infection of the other Gulf coast states. None of the Texas *I. scapularis* ticks appeared to be infected with any *Borrelia* species, a finding that also agrees with previous studies [41].

Quantitative PCR reactions were performed at RTL using *Borrelia* species-specific 16S rDNA primers of which 19 samples were positive for *Borrelia* species. Seven of the 19 qPCR positive samples were also positive when using *recG* primers, *uvrA* primers and/or 16S-23S

intergenic spacer primers. It is possible that the qPCR primers, which contained a few degenerate bases, did not specifically amplify *Borrelia* species. The addition of the degenerate bases may have decreased specificity. An examination of the 454 sequencing data from the 19 qPCR positive ticks might reveal a genus common to all 19 ticks, however, there was no bacterial genus common to all 19 samples. It must be noted that because the qPCR targeted a specific subset of bacterial 16S rDNA sequences, that it would be more sensitive than the 454 sequencing reaction, which nonspecifically amplified all bacterial 16S rDNA.

454 Sequencing results

As the major vector of Lyme disease in the northeastern part of the United States, multiple studies have investigated the bacterial diversity of the *I. scapularis* tick. However, the work reported in this thesis represents the first attempt to characterize the microbiome of ticks using high throughput sequencing techniques. Due to the limitations of 454 sequencing, the data is only accurate to the genus level. On average, nearly 37 genera per tick were found in New York *I. scapularis* ticks and 20 genera per tick were found in Texas *I. scapularis* ticks, demonstrating a diversity not seen in previous studies that used traditional 16S rDNA sequencing.

When characterizing host/bacteria relationships in ticks, bacterial endosymbionts will inevitably predominate. The results from Texas and New York *I. scapularis* ticks were of particular interest, as this species of tick is a major vector the Lyme Disease in the eastern United States. In our study, the most prevalent genera, *Rickettsia*, was found in 82% of NY *I. scapularis* ticks, and also accounted for the highest average percentage of sequences in individual ticks (Fig. 2, Table 5). The prevalence of a possible *Rickettsia* endosymbiont in *I. scapularis* ticks is well documented, and previous studies suggest transovarial transmission rather than constant reinfection from a reservoir host [42]. The exact proportion of *I. scapularis*

ticks containing *Rickettsia* spp. was not known precisely, with estimates ranging from as low as 40% to as high as 100% for New York *I. scapularis* ticks. Our data, at 82%, lie within this range [42, 43].

Likewise, previous studies of *A. americanum* nymphs found four known genera to be vertically transmitted endosymbionts of *A. americanum* ticks : *Coxiella* endosymbiont (ranging from 89%-100% of ticks), *Rickettsia* endosymbiont (ranging from 45%-65% of ticks), *Brevibacterium* endosymbiont (absent from ticks or in 11.7% of ticks) and *Arsenophonus* endosymbiont (depending on geographic location, are absent from ticks or in 90% of ticks) [28, 32, 44]. All four genera were detected in our *A. americanum* samples. For these four bacteria, both geographic groups of the *A. americanum* fell well within the reported range with very little variation in prevalence with the exception of *Brevibacterium* spp. which was nearly 10% more prevalent in Missouri ticks. Both the *Rickettsia* endosymbiont and *Coxiella* endosymbiont are considered to be primary endosymbionts, and both are believed to be essential for nutrient metabolism. The *Coxiella* endosymbiont is believed to be a required endosymbiont for *Amblyomma* ticks; however, over 10% the *A. americanum* ticks lacked detectable *Coxiella* sequences in our study. As stated earlier, colonization in the ovarian tissues is necessary for vertical transmission of the *Coxiella* endosymbiont and since male ticks lack ovaries, their *Coxiella* infection load might be too low for detection. At such low prevalence in Texas *A. americanum* adult tick, it does not appear that the *Arsenophonus* species nor the *Brevibacterium* spp. is necessary at this life stage in the tick.

Mutualistic bacteria species, particularly in the tick midgut, and commensals are also points of interest in this study. Common anaerobic bacteria genera, including *Bacteroides* spp., *Prevotella* spp., *Fusobacterium* spp., *Ruminococcus* spp. and *Faecalibacterium* spp., were present in both the *I. scapularis* and *A. americanum* sequencing results. These genera

represented a higher percentage of total sequences in *A. americanum* samples than in *I. scapularis* samples. However, this difference is not easily interpreted, because our *A. americanum* population contained a higher proportion of adults than the *I. scapularis* group.

In the case of New York *I. scapularis* ticks, numerous genera were present. One possibility is that the samples were contaminated during a lengthy storage period in 70% ethanol (>10 years) or were inadequately surface sterilized before sequencing. *E. coli*, the most well known mammalian intestinal bacterial species, was almost exclusively found in the New York *I. scapularis* ticks at a prevalence of 70%. *Sphingomonas* spp., some of which are known to form biofilms, as well as *Stenotrophomonas* spp., *Flavobacterium* spp. and *Acidovorax* spp. were at least two times more prevalent in New York *I. scapularis* ticks than their Texas counterparts. Despite accounting for a relatively large proportion of sequences in New York ticks, none of these species have been previously reported to be tick-associated. However, it could be speculated that these genera, with the exception of *E. coli*, are actually part of the normal surface microflora of the tick, much like *Staphylococcus* spp. and *Propionibacterium* spp. are commonly found on human skin.

Ticks are an incredibly important vector of disease and understanding the prevalence of these bacterial species is of great usefulness to human health. In addition to Lyme disease, ticks are known to vector babesiosis, bartonellosis and erlichiosis. Furthermore, Lyme disease patients are occasionally infected by multiple pathogens. It is possible that 454 sequencing may result in the identification of new tick-transmitted bacterial pathogens, either within genera already known to be tick-associated (e.g. *Borrelia* and *Ehrlichia*) or within other genera not considered previously to be tick-borne pathogens.

As discussed earlier, 454 sequencing resulted in three samples that were positive for *Borrelia* spp. of which only one was positive for a non- *Borrelia burgdorferi* B31 sequence.

However, the use of *Borrelia* specific primers showed eight ticks positive for *Borrelia* species. The low reported *Borrelia burgdorferi* burden in individual ticks could explain the lack of *Borrelia* spp. sequences in our 454 sequencing results. Had we increased the depth of coverage of our 454 reaction by using less samples per run, we would have likely seen more *Borrelia* sequences. However, in this study, the use of *Borrelia* specific primers appeared to be a far better method of *Borrelia* detection over this particular 454 sequencing run.

Outside of the *Borrelia* spp. sequences, three other non-*Borrelia* pathogenic genera were found in *I. scapularis* samples. Of these three non-*Borrelia* pathogenic genera, two of the genera are present in ten *I. scapularis* ticks, and three genera are present in two *I. scapularis* ticks. The most prevalent genus, *Mycobacterium*, was present in 43.33% of all *I. scapularis* ticks and over 50% of all *A. americanum* ticks. Of these ticks, 6.6% of all *I. scapularis* are infected with a very small numbers of *M. avium*. None of the *A. americanum* ticks appeared to be infected with *M. avium*. New evidence suggests *Mycobacterium bovis* still exists in free roaming white tail deer despite being eradicated from United States cattle [45]. Considering that both *I. scapularis* and *A. americanum* are commonly found feeding on white tail deer, it is possible that the tick has acquired *M. bovis* from infected hosts [46]. However, the *Mycobacterium* spp. category represents a very low percentage of all sequences in both species of ticks. It is difficult to know if the *Mycobacterium* spp. are from a pathogenic species or commensal species.

Anaplasma spp. sequences was found in 11.7% of all *I. scapularis* ticks and none of the *A. americanum* ticks. Previous reports of the prevalence of *Anaplasma phagocytophilum*, the causative agent of anaplasmosis, in the northeast United States depends on geographical location. Infection rate varies from 1.9% in northwestern Pennsylvania to 39.8% infection rate in southeastern Pennsylvania [47]. It is possible that some the *Anaplasma* spp. sequences

detected in our study were due to *Anaplasma phagocytophilum* infection of *I. scapularis* ticks, but this interpretation would require further investigation using species-specific primers and/or sequencing.

Serratia marcescens, a gram negative bacterium responsible for nosocomial infections in the United States, was present in less than a quarter of all *Ixodes scapularis* samples but absent in *A. americanum* samples. The protozoal pathogen *Babesia microti*, the causative agent of babesiosis, is commonly vectored by *Ixodes* ticks but considered rare in Texas. A clinical assay to detect the causative agent for babesiosis was performed at RTL using the same samples sent to RTL for 454 sequencing (data not shown). 33 of the samples tested positive for *B. microti*, three of which were Texas *I. scapularis* ticks, 21 were *A. americanum* ticks Texas and nine were *A. americanum* ticks from Missouri. There are no known publications describing the presence of *Babesia microti* in *A. americanum* ticks. However, at an infection rate of 43% of the Texas *A. americanum* ticks tested, *Babesia microti* could demand more attention to determine potential health risks posed by *A. americanum* ticks.

A. americanum ticks examined in this study yielded sequences for two additional known pathogenic genera: *Francisella* spp., whose species can cause tularemia, and *Ehrlichia* spp., whose species can cause ehrlichiosis in canines and humans. However, *Bartonella* sequences were detected in 7.2% of Texas *A. americanum* ticks. *A. americanum* is a competent vector for seven *Bartonella* spp., including *B. henselae* which causes cat-scratch disease, as well as a related organism, *Peliosis hepatis*, in immunocompromised patients. It is not known if the *Bartonella* sequences seen in the Texas *A. americanum* ticks is from a pathogenic species.

Finally, the frozen tick halves prepared at RTL were not considered in the final analysis of ticks and their associated genera. The seven *A. americanum* and seven *A. cajennense* samples

clustered extremely well in the dual dendrogram despite being different species collected from different locations. This pattern was not seen with the DNA samples from *I. scapularis* and *A. americanum* ticks that were extracted in Houston. Each of the RTL-prepared samples had a relatively high percentage (up to 30%) of *Flavobacterium* spp. and *Pseudomonas* spp. sequences; however, the *A. americanum* ticks lacked *Rickettsia* spp. or *Coxiella* spp., which accounted for over 90% of sequences in many of the Houston-extracted samples. Furthermore, the RTL-prepared samples appeared to be positive for many more genera and species when compared to samples that were processed at UT-Houston. The RTL prepared genomic DNA was, on average, positive for 133 different genera and species while the samples prepared at UT-Houston had an average of 30 positive samples. The fourteen frozen samples prepared at RTL had 208 unique genera or species not present in the other 115 samples that were prepared at UT-Houston. The 115 samples prepared at UT-Houston had 251 unique genera or species. The fourteen genomic DNA samples prepared at RTL averaged 4,307 reads per tick while the other 115 samples prepared at UT-Houston averaged 4291 reads per tick, a difference of only 16 reads per tick. The difference between the two sample sets is quite low and most likely does not account for the high number of unique genera and species found in the RTL prepared ticks. It should be noted that the tick genomic DNA extraction techniques used at UT-Houston were different than methods used at RTL. Furthermore, genomic DNA extracted at RTL was sheared to give a more robust initial PCR amplification. Because of these rather large differences, it is difficult to draw any conclusions about the data collected from the samples whose genomic DNA was extracted at RTL.

Despite our best efforts, the causative agent of STARI is still unknown. Although a still unknown *Borrelia* species was believed to be the causative agent of STARI, no new *Borrelia* species were detected by either of our two methods. The results seen from the *Borrelia* specific

primer screen lies within previously reported infection rate for both *A. americanum* and *I. scapularis* ticks. However, a tremendous amount of information was obtained from the 454 sequencing results and documented a high degree of diversity unseen in previous literature. Pathogenic bacteria species are always of particular interest when it comes to ticks. The detection of *Mycobacteria* species combined with this study's documentation of the prevalence of *Babesia microti* in *A. americanum* places a new importance on *A. americanum* as a possible vector of human disease. The characterization of the microbial communities of both the *I. scapularis* and *A. americanum* tick is crucial to fully understand both tick species. We have identified numerous apparent commensals and possible endosymbionts in both species of ticks. As expected, the microbial communities of the *Ixodes* and *Amblyomma* ticks examined in this study exhibited tremendous variation, both within and between the two tick species. In addition to tick genus and species, geographic location, life stage, sex and environmental factors appear to play an important role in determining the microflora present. However, even when all the previous factors are equal, variability still exists in the microbial sequences obtained from individual ticks of the same species collected at the same location. In the end, these results only represent an initial picture of the microbial diversity of the tick. Subsequent studies need to be conducted to determine the roles these genera play in the fitness of ticks as well as their effects on human health.

This study does not bring us any closer to discovering the causative agent of STARI. Only *B. lonestari* was detected in either Texas *I. scapularis* ticks or Texas and as *B. lonestari* has been ruled out as the causative agent of STARI, the causative agent of STARI is still not known. With the evidence from previous studies as well as the results seen in this study, it could be reasoned, that STARI is not caused by a *Borrelia* species. However, a rare but unknown species of *Borrelia* might still exist that could be the causative agent of STARI.

CHAPTER 5 : REFERENCES

1. David T. Dennis, J.F.P., *Chapter 1 - Overview of Tick-Borne Infections of Humans* in *Tick-Borne Disease of Humans*, D.T.D. Jesse L Goodman, Daniel E. Sonenshine, Editor. 2005. ASM Press: Washington, D.C.
2. Childs, J.E. and C.D. Paddock, *The ascendancy of Amblyomma americanum as a vector of pathogens affecting humans in the United States*. Annu Rev Entomol, 2003. **48**: p. 307-37.
3. Sonenshine, D.E., *Chapter 2 - The Biology of Tick Vectors of Human Disease*, in *Tick Borne Disease of Humans*, D.T.D. Jesse L Goodman, Daniel E. Sonenshine, Editor. 2005, ASM Press: Washington, D.C. p. 13-64
4. Allen C. Steere, J.B., Lisa Glickstein, *Chapter 11 - Lyme Borreliosis* in *Tick-Borne Disease of Humans*, D.T.D. Jesse L. Goodman, Daniel E. Sonenshine, Editor. 2005, ASM Press: Washington, DC. p. 176-206.
5. *Reported Lyme disease cases by state, 1999-2008*. 2009.
6. Unit, T.D.o.S.H.S.-I.D.C., *Human Cases of Reportable Zoonotic Disease in Texas, 1990 to present*. 2010.
7. Masters, E.J., C.N. Grigery, and R.W. Masters, *STARI, or Masters disease: Lone Star tick-vectored Lyme-like illness*. Infect Dis Clin North Am, 2008. **22**(2): p. 361-76, viii.
8. Ledin, K.E., N.S. Zeidner, J.M. Ribeiro, B.J. Biggerstaff, M.C. Dolan, G. Dietrich, L. Vredevoe, and J. Piesman, *Borreliacidal activity of saliva of the tick Amblyomma americanum*. Med Vet Entomol, 2005. **19**(1): p. 90-5.

9. Zeidner, N., A. Ullmann, C. Sackal, M. Dolan, G. Dietrich, J. Piesman, and D. Champagne, *A borreliacidal factor in Amblyomma americanum saliva is associated with phospholipase A2 activity*. Exp Parasitol, 2009. **121**(4): p. 370-5.
10. Bacon, R.M., M.A. Pilgard, B.J. Johnson, S.J. Raffel, and T.G. Schwan, *Glycerophosphodiester phosphodiesterase gene (glpQ) of Borrelia lonestari identified as a target for differentiating Borrelia species associated with hard ticks (Acari:Ixodidae)*. J Clin Microbiol, 2004. **42**(5): p. 2326-8.
11. Barbour, A.G., G.O. Maupin, G.J. Teltow, C.J. Carter, and J. Piesman, *Identification of an uncultivable Borrelia species in the hard tick Amblyomma americanum: possible agent of a Lyme disease-like illness*. J Infect Dis, 1996. **173**(2): p. 403-9.
12. Iyer, R., O. Kalu, J. Purser, S. Norris, B. Stevenson, and I. Schwartz, *Linear and circular plasmid content in Borrelia burgdorferi clinical isolates*. Infect Immun, 2003. **71**(7): p. 3699-706.
13. Casjens, S., N. Palmer, R. van Vugt, W.M. Huang, B. Stevenson, P. Rosa, R. Lathigra, G. Sutton, J. Peterson, R.J. Dodson, D. Haft, E. Hickey, M. Gwinn, O. White, and C.M. Fraser, *A bacterial genome in flux: the twelve linear and nine circular extrachromosomal DNAs in an infectious isolate of the Lyme disease spirochete Borrelia burgdorferi*. Mol Microbiol, 2000. **35**(3): p. 490-516.
14. Labandeira-Rey, M. and J.T. Skare, *Decreased infectivity in Borrelia burgdorferi strain B31 is associated with loss of linear plasmid 25 or 28-1*. Infect Immun, 2001. **69**(1): p. 446-55.

15. Fraser, C.M., S. Casjens, W.M. Huang, G.G. Sutton, R. Clayton, R. Lathigra, O. White, K.A. Ketchum, R. Dodson, E.K. Hickey, M. Gwinn, B. Dougherty, J.F. Tomb, R.D. Fleischmann, D. Richardson, J. Peterson, A.R. Kerlavage, J. Quackenbush, S. Salzberg, M. Hanson, R. van Vugt, N. Palmer, M.D. Adams, J. Gocayne, J. Weidman, T. Utterback, L. Watthey, L. McDonald, P. Artiach, C. Bowman, S. Garland, C. Fuji, M.D. Cotton, K. Horst, K. Roberts, B. Hatch, H.O. Smith, and J.C. Venter, *Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi*. Nature, 1997. **390**(6660): p. 580-6.
16. Pal, U., X. Yang, M. Chen, L.K. Bockenstedt, J.F. Anderson, R.A. Flavell, M.V. Norgard, and E. Fikrig, *OspC facilitates Borrelia burgdorferi invasion of Ixodes scapularis salivary glands*. J Clin Invest, 2004. **113**(2): p. 220-30.
17. Yang, X.F., U. Pal, S.M. Alani, E. Fikrig, and M.V. Norgard, *Essential role for OspA/B in the life cycle of the Lyme disease spirochete*. J Exp Med, 2004. **199**(5): p. 641-8.
18. Posey, J.E. and F.C. Gherardini, *Lack of a role for iron in the Lyme disease pathogen*. Science, 2000. **288**(5471): p. 1651-3.
19. Chakravorty, S., D. Helb, M. Burday, N. Connell, and D. Alland, *A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria*. J Microbiol Methods, 2007. **69**(2): p. 330-9.
20. Dethlefsen, L., S. Huse, M.L. Sogin, and D.A. Relman, *The pervasive effects of an antibiotic on the human gut microbiota, as revealed by deep 16S rRNA sequencing*. PLoS Biol, 2008. **6**(11): p. e280.

21. Huse, S.M., L. Dethlefsen, J.A. Huber, D. Mark Welch, D.A. Relman, and M.L. Sogin, *Exploring microbial diversity and taxonomy using SSU rRNA hypervariable tag sequencing*. PLoS Genet, 2008. **4**(11): p. e1000255.
22. Sanger, F., S. Nicklen, and A.R. Coulson, *DNA sequencing with chain-terminating inhibitors*. Proc Natl Acad Sci U S A, 1977. **74**(12): p. 5463-7.
23. Rogers, Y.H. and J.C. Venter, *Genomics: massively parallel sequencing*. Nature, 2005. **437**(7057): p. 326-7.
24. Ronaghi, M., M. Uhlen, and P. Nyren, *A sequencing method based on real-time pyrophosphate*. Science, 1998. **281**(5375): p. 363, 365.
25. Margos, G., A.G. Gatewood, D.M. Aanensen, K. Hanincova, D. Terekhova, S.A. Vollmer, M. Cornet, J. Piesman, M. Donaghy, A. Bormane, M.A. Hurn, E.J. Feil, D. Fish, S. Casjens, G.P. Wormser, I. Schwartz, and K. Kurtenbach, *MLST of housekeeping genes captures geographic population structure and suggests a European origin of Borrelia burgdorferi*. Proc Natl Acad Sci U S A, 2008. **105**(25): p. 8730-5.
26. Dowd, S.E., T.R. Callaway, R.D. Wolcott, Y. Sun, T. McKeehan, R.G. Hagevoort, and T.S. Edrington, *Evaluation of the bacterial diversity in the feces of cattle using 16S rDNA bacterial tag-encoded FLX amplicon pyrosequencing (bTEFAP)*. BMC Microbiol, 2008. **8**: p. 125.
27. Pitta, D.W., W.E. Pinchak, S.E. Dowd, J. Osterstock, V. Gontcharova, E. Youn, K. Dorton, I. Yoon, B.R. Min, J.D. Fulford, T.A. Wickersham, and D.P. Malinowski,

Rumen Bacterial Diversity Dynamics Associated with Changing from Bermudagrass Hay to Grazed Winter Wheat Diets. Microb Ecol, 2009.

28. Clay, K., O. Klyachko, N. Grindle, D. Civitello, D. Oleske, and C. Fuqua, *Microbial communities and interactions in the lone star tick, Amblyomma americanum.* Mol Ecol, 2008.
29. Bunikis, J., J. Tsao, U. Garpmo, J. Berglund, D. Fish, and A.G. Barbour, *Typing of Borrelia relapsing fever group strains.* Emerg Infect Dis, 2004. **10**(9): p. 1661-4.
30. Tokarz, R., K. Jain, A. Bennett, T. Briese, and W.I. Lipkin, *Assessment of polymicrobial infections in ticks in New York state.* Vector Borne Zoonotic Dis. **10**(3): p. 217-21.
31. Jasinskas, A., J. Zhong, and A.G. Barbour, *Highly prevalent Coxiella sp. bacterium in the tick vector Amblyomma americanum.* Appl Environ Microbiol, 2007. **73**(1): p. 334-6.
32. Klyachko, O., B.D. Stein, N. Grindle, K. Clay, and C. Fuqua, *Localization and visualization of a coxiella-type symbiont within the lone star tick, Amblyomma americanum.* Appl Environ Microbiol, 2007. **73**(20): p. 6584-94.
33. Zhong, J., A. Jasinskas, and A.G. Barbour, *Antibiotic treatment of the tick vector Amblyomma americanum reduced reproductive fitness.* PLoS One, 2007. **2**(5): p. e405.
34. Barbour, A.G., J. Bunikis, B. Travinsky, A.G. Hoen, M.A. Diuk-Wasser, D. Fish, and J.I. Tsao, *Niche partitioning of Borrelia burgdorferi and Borrelia miyamotoi in the*

- same tick vector and mammalian reservoir species*. Am J Trop Med Hyg, 2009. **81**(6): p. 1120-31.
35. Scoles, G.A., M. Papero, L. Beati, and D. Fish, *A relapsing fever group spirochete transmitted by Ixodes scapularis ticks*. Vector Borne Zoonotic Dis, 2001. **1**(1): p. 21-34.
 36. Brownstein, J.S., T.R. Holford, and D. Fish, *Effect of Climate Change on Lyme Disease Risk in North America*. Ecohealth, 2005. **2**(1): p. 38-46.
 37. Gatewood, A.G., K.A. Liebman, G. Vourc'h, J. Bunikis, S.A. Hamer, R. Cortinas, F. Melton, P. Cislo, U. Kitron, J. Tsao, A.G. Barbour, D. Fish, and M.A. Diuk-Wasser, *Climate and tick seasonality are predictors of Borrelia burgdorferi genotype distribution*. Appl Environ Microbiol, 2009. **75**(8): p. 2476-83.
 38. Lindsay, L.R., I.K. Barker, G.A. Surgeoner, S.A. McEwen, and G.D. Campbell, *Duration of Borrelia burgdorferi infectivity in white-footed mice for the tick vector Ixodes scapularis under laboratory and field conditions in Ontario*. J Wildl Dis, 1997. **33**(4): p. 766-75.
 39. Taft, S.C., M.K. Miller, and S.M. Wright, *Distribution of borreliae among ticks collected from eastern states*. Vector Borne Zoonotic Dis, 2005. **5**(4): p. 383-9.
 40. Bacon, R.M., R.D. Gilmore, Jr., M. Quintana, J. Piesman, and B.J. Johnson, *DNA evidence of Borrelia lonestari in Amblyomma americanum (Acari: Ixodidae) in southeast Missouri*. J Med Entomol, 2003. **40**(4): p. 590-2.
 41. Rawlings, J.A. and G.J. Teltow, *Prevalence of Borrelia (Spirochaetaceae) spirochetes in Texas ticks*. J Med Entomol, 1994. **31**(2): p. 297-301.

42. Benson, M.J., J.D. Gawronski, D.E. Eveleigh, and D.R. Benson, *Intracellular symbionts and other bacteria associated with deer ticks (Ixodes scapularis) from Nantucket and Wellfleet, Cape Cod, Massachusetts*. Appl Environ Microbiol, 2004. **70**(1): p. 616-20.
43. Moreno, C.X., F. Moy, T.J. Daniels, H.P. Godfrey, and F.C. Cabello, *Molecular analysis of microbial communities identified in different developmental stages of Ixodes scapularis ticks from Westchester and Dutchess Counties, New York*. Environ Microbiol, 2006. **8**(5): p. 761-72.
44. Heise, S.R., M.S. Elshahed, and S.E. Little, *Bacterial diversity in Amblyomma americanum (Acari: Ixodidae) with a focus on members of the genus Rickettsia*. J Med Entomol. **47**(2): p. 258-68.
45. Schmitt, S.M., S.D. Fitzgerald, T.M. Cooley, C.S. Bruning-Fann, L. Sullivan, D. Berry, T. Carlson, R.B. Minnis, J.B. Payeur, and J. Sikarskie, *Bovine tuberculosis in free-ranging white-tailed deer from Michigan*. J Wildl Dis, 1997. **33**(4): p. 749-58.
46. Corn, J.L. and V.F. Nettles, *Health protocol for translocation of free-ranging elk*. J Wildl Dis, 2001. **37**(3): p. 413-26.
47. Courtney, J.W., R.L. Dryden, J. Montgomery, B.S. Schneider, G. Smith, and R.F. Massung, *Molecular characterization of Anaplasma phagocytophilum and Borrelia burgdorferi in Ixodes scapularis ticks from Pennsylvania*. J Clin Microbiol, 2003. **41**(4): p. 1569-73.

VITA

Daniel Tau Yuan was born in Houston, Texas on November 3rd 1982 to Hsien-Hwa "Steven" Yuan and Cindy Yuan. He graduated from Clear Lake High School in 2001 and attended the University of Texas at Austin in the Fall 2001. In August of 2005 he graduated from the University of Texas at Austin with a bachelors of science in Microbiology. He went on to work at Baylor College of Medicine, first at the Human Genome Sequencing Center and then in Dr. Aleksandar Rajkovic's lab in the OB/GYN department. In the Fall of 2008, he enrolled in The University of Texas Health Science Center Graduate School of Biomedical Sciences at Houston.