2017

Population Genetics Of Borrelia Burgdorferi In The Eastern And Midwestern United States

Stephanie Nicole Seifert
University of Pennsylvania, seifst@sas.upenn.edu

Population Genetics Of Borrelia Burgdorferi In The Eastern And Midwestern United States

Abstract
Changes in climate and land-use are influencing the spatial distributions and population structure of species throughout the world. Among those species predicted to be affected by changes in climate and land-use are the hosts and vectors of many zoonotic pathogens, including the hosts and vectors of the Lyme borreliosis pathogen, Borrelia burgdorferi. In this study, we investigate the population structure of B. burgdorferi at the interface of two expanding vector populations using a population genetic and phylogeographic approach. We then examine the influence of recombination between major B. burgdorferi lineages on the population structure of B. burgdorferi within and between geographic regions. We found that B. burgdorferi lineages were heterogeneously distributed between the Northeast and Southeast United States, suggesting that gene flow between geographic regions is limited for some lineages while other lineages showed evidence of long distance dispersal. We identified at least two clonal B. burgdorferi lineages present in both competent vector species found in southern Virginia, suggesting that vector species is not a barrier to dispersal between regions. Population genetic and phylogeographic analyses of B. burgdorferi reveal a complicated pattern of population structure within and between geographic regions. Population genomic analyses show recombination between lineages despite evidence of strong linkage disequilibrium and a lack of genetic variation within B. burgdorferi lineages, suggesting a small effective population size.

Degree Type
Dissertation

Degree Name
Doctor of Philosophy (PhD)

Graduate Group
Biology

First Advisor
Dustin Brisson

Second Advisor
Paul S. Schmidt

Keywords
Borrelia, Disease Ecology, Microbiology, Phylogenetics, Phylogeography, Recombination

Subject Categories
Biology | Ecology and Evolutionary Biology | Genetics

POPULATION GENETICS OF BORRELIA BURGDORFERI IN THE EASTERN AND MIDWESTERN UNITED STATES

Stephanie N. Seifert

A DISSERTATION

in

Biology

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2017

Supervisor of Dissertation

_____

Dr. Dustin Brisson, Associate Professor of Biology, University of Pennsylvania

Graduate Group Chairperson

_____

Dr. Michael Lampson, Associate Professor of Biology, University of Pennsylvania

Dissertation Committee:

Dr. Paul Schmidt, Associate Professor of Biology, University of Pennsylvania

Dr. Dustin Brisson, Associate Professor of Biology, University of Pennsylvania

Dr. Michael Levy, Associate Professor of Epidemiology, University of Pennsylvania

Dr. Timothy Linksvayer, Assistant Professor of Biology, University of Pennsylvania

Dr. Paul Sniegowski, Professor of Biology, University of Pennsylvania

DEDICATION

For my parents, who nurtured my independence and curiosity. Without their unconditional love and support, I would not have taken the leap from working at coffee shops and record stores to pursuing higher education. I thank them for encouraging me in all my endeavors, especially this one.

ACKNOWLEDGMENT

I would like to thank my thesis advisor, Dustin Brisson, for giving me the opportunity to work in the Evolution and Ecology of Infectious Disease Systems Laboratory (EEDS Lab) at Penn. Dustin has provided me with incredible laboratory resources as well as the autonomy to develop my own research topic and interests. I am grateful for the patience, intellectual input, and support of my committee members: Paul Schmidt, Tim Linksvayer, Mike Levy, and Paul Sniegowski. I thank David Roos for both discussions about pathogen systems and for career guidance.

I have greatly benefited from working with the past and present members of the Brisson Lab: Godefroy Devevey, Ozlem Onder, Tam Tran, Camilo Khatchikian, Wilbert Long, Erica Foley, Emily Avis, and Chloe Kaczvinsky. Special thanks to Wei Zhou, Alex Berry, Damien Lekkas, and Jill Devine for their many constructive suggestions over the years and for creating a great environment to discuss everything science and science-fiction.

I am thankful to my graduate student cohort, especially Mitra Eghbal, Emily Behrman, and Lee Dietterich for our time spent discussing science and grad student life, reading each other's research and grant proposals, playing board games, and exploring Philadelphia. I have also valued lively academic discussions with Mitchell Newberry, Vincent Luczak, Ruby Cortes, Joe LaChance, Aaron Leichty, Sesh Sundararaman, Erik Clarke, Ozan Kiratli, Aurora MacRae-Crerar, Vinayak Mathur, and Dan Song. I want to extend my gratitude to the Bio-Computing group, especially Adam Linder, Paul Dougherty, and Shaheen Beg for their good nature and dedication to keeping our computers functioning and programs running.

I would like to acknowledge the kindness and generosity of Deborah Brandenburg who gave me a place to stay near my partner, a friendly ear, and many great conversations.

Finally, I am grateful for my partner, Matthew Brandenburg, for his unwavering support and encouragement, for helping me crush or collect ticks on weekends, for caring for our dogs when I worked late or had to travel, for believing in me when my confidence faltered, and for giving me a reason to smile every single day. SoHvaD tlqwlj vl'ang.

ABSTRACT

POPULATION GENETICS OF BORRELIA BURGDORFERI IN THE EASTERN AND MIDWESTERN UNITED STATES

Stephanie N Seifert

Dustin Brisson

Changes in climate and land-use are influencing the spatial distributions and population structure of species throughout the world. Among those species predicted to be affected by changes in climate and land-use are the hosts and vectors of many zoonotic pathogens, including the hosts and vectors of the Lyme borreliosis pathogen, *Borrelia burgdorferi.* In this study, we investigate the population structure of *B. burgdorferi* at the interface of two expanding vector populations using a population genetic and phylogeographic approach. We then examine the influence of recombination between major *B. burgdorferi* lineages on the population structure of *B. burgdorferi* within and between geographic regions. We found that *B. burgdorferi* lineages were heterogeneously distributed between the Northeast and Southeast United States, suggesting that gene flow between geographic regions is limited for some lineages while other lineages showed evidence of long distance dispersal. We identified at least two clonal *B. burgdorferi* lineages present in both competent vector species found in southern Virginia, suggesting that vector species is not a barrier to dispersal between regions. Population genetic and phylogeographic analyses of *B. burgdorferi* reveal a complicated pattern of population structure within and between geographic regions. Population genomic analyses show recombination between lineages despite evidence of strong linkage disequilibrium and a lack of genetic variation within *B. burgdorferi* lineages, suggesting a small effective population size.

TABLE OF CONTENTS

ACKNOWLEDGMENT III

ABSTRACT IV

LIST OF TABLES VII

LIST OF FIGURES VIII

CHAPTER 1: A GENERAL INTRODUCTION 1

1.1 The importance of population structure in microbial communities 1

1.2 Population genetics and phylogeography 3

1.3 Introduction to the Borrelia burgdorferi system 4

1.4 Scientific questions and dissertation structure 7

CHAPTER 2: POPULATION GENETICS AND PHYLOGEOGRAPHY OF BORRELIA BURGDORFERI BETWEEN THE NORTHEASTERN AND SOUTHEASTERN UNITED STATES 10

2.1 Introduction 10

2.2 Methods 14

2.3 Results and Discussion 18

2.4 Conclusions 25

2.5 Figures and tables 27

CHAPTER 3: POPULATION GENOMICS OF BORRELIA BURGDORFERI REVEAL LOW LEVELS OF RECOMBINATION AT THE OSPC LOCUS, DISRUPTING CLONAL ASSOCIATION WITH OSPC MAJOR ALLELE GROUPS 35

3.1 Introduction 35

3.2 Methods 37

3.3 Results and Discussion 40

3.4 Conclusions 43

3.5 Figures and tables 47

CHAPTER 4: DISCUSSION 52

SUPPLEMENTARY MATERIAL 54

REFERENCES 57

LIST OF TABLES

Table 2.1. Tick collection data and results of diagnostic PCR targeting *B. burgdorferi* IGS and *ospC*……..…27

Table 2.2. Origin, sample sizes, and sequence diversity of *Borrelia burgdorferi* plasmid sequences from the Northeast, Virginia, and North Carolina..28

Table 2.3. Analysis of molecular variance for plasmid sequence data of *B. burgdorferi* among the Northeast, Virginia, and North Carolina and sampling sites within regions……29

Table 3.1. Origin, sample sizes, sequence diversity, and linkage across the *B. burgdorferi* chromosome and plasmid sequences……..47

Table 3.2. Recombination statistics for each *B. burgdorferi* genome segment as estimated in ClonalFrameML……48

Supplementary Table S1A. Long-Range PCR primers and positions on genome segment……..……54

Supplementary Table S1B. Thermal cycling conditions for LR-PCR……...…..……55

LIST OF FIGURES

Figure 1.1 Map showing approximate sampling locations…….…….7

Figure 2.1. Distribution of oMGs from questing ticks, human erythema migrans lesions, and human disseminated infections (Seinost et al. 1999)…...30

Figure 2.2. Tick sampling locations and infection prevalence……....31

Figure 2.3. The distribution of *ospC* major allele groups among ticks…….32

Figure 2.4. Bayesian phylogenetic tree of (A) cp26 sequences and (B) lp54 sequences from *B. burgdorferi* collected in the NE, VA, and NC……..33

Figure 2.5. Constrained analysis of principal coordinates based on tick species and geographic origin of *B. burgdorferi* oMGs……..34

Figure 3.1A Maximum likelihood tree and recombination graph of *B. burgdorferi* chromosome sequences……49

Figure 3.1B Maximum likelihood tree and recombination graph of *B. burgdorferi* lp54 plasmid sequences……..…50

Figure 3.1C Maximum likelihood tree and recombination graph of *B. burgdorferi* cp26 plasmid sequences……..…51

Supplementary Figure S2 Sequences used as the *ospC* reference database for local BLAST……..……..…56

CHAPTER 1: A GENERAL INTRODUCTION

1.1 The importance of population structure in microbial communities

Evolutionary forces including mutation, natural selection, genetic drift, gene flow, and recombination shape the diversity of species and leave a detectable molecular signature on populations. Evolutionary forces play out on populations which are distributed in geographic space and time. Thus, population structure is often described in the context of biogeography.

The study of microbial biogeography has largely centered around the Baas-Becking hypothesis which suggests that microbial species experience boundless dispersal and environmental or ecological conditions shape the distribution of microbial species so that "everything is everywhere, but the environment selects" (Finlay and Clark 1999). Nevertheless, dispersal in host-associated microbes is limited by the biogeography of their hosts (Smith et al. 2007, Martiny et al. 2006, O'Malley 2008) leading to the emergence of population structure and geographic endemism (O'Malley 2008).

Human migration and globalization has led to the spread and subsequent establishment of many human-associated pathogens with previously limited geographic distributions including West Nile virus (Spielman et al. 2004), *Mycobacterium tuberculosis* (Comas et al. 2013), and *Helicobacter pylori* (Linz et al. 2007). Consequently, efforts to elucidate the factors contributing to the population structure of pathogenic microbes have been increasing.

With rapid generation times and large populations, microbes are thought to have high evolutionary potential and high genetic diversity. In contrast, many highly pathogenic microbes have very low genetic diversity and relatively small effective population sizes (Achtman 2008). There are several models of microbial evolution which

could give rise to small effective population sizes including A) differentiation of populations by ecotype (e.g. host or vector) followed by periodic selection, B) metapopulation dynamics in which isolated patches (e.g. host or vector) are randomly colonized by a subpopulation with occasional patch turnover, C) genetic drift alone, or D) predator prey dynamics in which the bacterial population experiences regular oscillations in population size (Fraser et al. 2009).

Pathogen systems are typically better sampled than other microbial systems, which makes them an excellent choice for testing ecological and evolutionary hypotheses. For example, influenza viruses are frequently discussed in terms of antigenic shifts in which novel viral subtypes are generated through recombination of gene segments from two or more viral strains; however, individual influenza lineages undergo antigenic drift, in which a single strain accumulates mutations through time without recombination (Earn et al. 2002). Thorough sampling of individual influenza strains revealed a low effective population size (Rambaut et al. 2008). The authors found that seasonal influenza outbreaks in the temperate zone follow a cycle of increasing diversity through the “flu season” followed by a genome wide selective sweep each spring, resulting in low genetic diversity relative to similar viruses which infect fewer people (Rambaut et al. 2008). Recombination (or reassortment) of genetic segments would then rapidly add genetic variation to the genome.

In bacteria, recombination involves the non-vertical transfer and incorporation of genetic material between a donor and recipient and is considered an important driver of bacterial evolution (Polz et al. 2013, Gyles & Boerlin 2014). Recombination can disrupt linkage groups by moving gene fragments between different genetic backgrounds, nullifying the effects of clonal interference (Vos 2009). Recombination facilitates both the accumulation of beneficial mutations in a single genetic background and the purging

of deleterious mutations (Vos 2009). Thus, recombination increases the effective strength of natural selection and allows for rapid adaptation in bacterial populations (Wiedenbeck & Cohan 2011).

The rate of recombination can vary between closely related bacteria (Didelot & Maiden 2010) and in a single bacterial lineage through time (Holt et al. 2008). It is unclear whether the observed rate of recombination in bacterial populations is reflective of the true frequency of recombination events or the strength of selection for advantageous alleles (Didelot & Maiden 2010). The low observed rate of recombination in many bacterial pathogens may be consequence of stabilizing selection that is purging deleterious alleles acquired through recombination or a true lack of recombination events (Achtman 2008, Achtman 2012, Haven et al. 2011, Casjens et al. 2012).

1.2 Population genetics and phylogeography

The ecological and evolutionary history of a species leaves an imprint on the genetic structure of populations. In the presence of strong selection and local adaptation, genetic exchange should be highest between similar environments resulting in a pattern of genetic isolation by environment (Cooke et al. 2012). In terms of pathogen populations, the environment can be thought of as different host or vector species. If gene flow is reduced between different host types, then host specialization may be favored, resulting in genetic structure between host types. Alternatively, persisting gene flow between host environments and recombination could constrain host adaptation resulting in a panmixing population (Cooke et al. 2012). In addition, a similar process can be found when geographical regions are considered, as restricted gene flow among geographic areas allows genetic drift to differentiate populations with increasing geographic distance, in a process known as isolation by distance or IBD (Wright 1943).

Phylogeography is the study of the evolutionary and ecological processes that are responsible for current species distributions (Avise et al. 1987). The integration of the principles of population genetics, biogeography, and ecology gives the context in which to interpret patterns of present day species distributions. Phylogeographic analyses provide a framework for testing ecological and evolutionary hypotheses for drivers of population structure and geographic distributions.

Using phylogenetic trees coupled with geographic, we can visualize evolutionary relationships between individuals in a population and make inferences regarding population structure across geographic space. With improvements in computational power and the development of new statistical tools are developed which allow for the simultaneous estimation of the phylogeny with associated traits researchers have been able to make predictions about the spread of pathogens through space and time (Lemey et al. 2014). However, resolution and accuracy of phylogeographic studies are generally improved by increasing the number of samples and informative loci (Benavides et al. 2014).

1.3 Introduction to the Borrelia burgdorferi system

Borrelia burgdorferi is a globally distributed obligate vector-borne bacterial pathogen that is transmitted between vertebrate hosts by ticks in the genus *Ixodes* (Eisen & Lane 2002). *B. burgdorferi* has been identified in dozens of vertebrate species ranging from squirrels, shrews, birds, and mice (Barbour and Hayes 1986). As *B. burgdorferi* has no free-living stage, dispersal is dependent on movement by vertebrate hosts and vectors, and the ability for B. burgdorferi to persist in an environment is inexorably linked to presence of both suitable hosts and suitable vectors.

The Ixodid tick vectors are blood-feeding ectoparasites that undergo four primary stages of development: the egg, the larval stage, the nymphal stage, and the adult stage. An ixodid tick requires a host bloodmeal at each life stage, detaching from the host between feedings to molt. *B. burgdorferi* spirochetes are ingested by an Ixodid larva while the tick feeds on an infected vertebrate host (Eisen & Lane 2002). The pathogen must then persist in the midgut of the tick through both molting between life stages and through the winter diapause (Eisen & Lane 2002). The pathogen multiplies within the midgut of the replete larval tick until the bloodmeal has been digested and then experiences a precipitous decline in abundance when the tick molts (Piesman et al. 1990). Upon ingestion of the second bloodmeal, the spirochetes begin to multiply in the tick midgut and migrate through the hemolymph and enter the salivary glands where the spirochetes exit the tick and infect the vertebrate host (de Silva and Fikrig 1995, Anguita et al. 2003).

The *B. burgdorferi* genome is comprised of a single linear chromosome, containing most of the housekeeping genes, and a variable number of circular and linear plasmids (Chaconas & Kobryn 2010). There is little diversity in the size and content of the linear chromosome between *B. burgdorferi* isolates (Schutzer et al. 2011); however, the complement and content of the plasmids varies widely between lineages (Casjens et al. 2012). The plasmids contain many decaying pseudogenes as well as many paralogs, suggesting a history of duplicative rearrangement (Casjens et al. 2000). Plasmid cp26 appears to be the only plasmid required for growth in culture, as it encodes a telomere resolvase necessary for replication of linear molecules and the host-associated *ospC* locus (Byram et al. 2004). Additional plasmids are necessary for maintenance *in vivo* and contain genes encoding for host-regulated surface expressed lipoproteins (Samuels 2011).

The outer surface proteins OspC and OspA have been thoroughly studied in *B. burgdorferi*. The highly polymorphic lipoprotein, OspC, is expressed by *B. burgdorferi* in the tick migut upon encountering a bloodmeal and is then subsequently expressed while in the vertebrate host (Gilmore & Piesman 2000). OspC is one of the first *B. burgdorferi* antigens to be recognized and attacked by the immune system of the vertebrate host (Fung et al. 1994), and is one of few loci to undergo horizontal gene transfer in whole or in part between *B. burgdorferi* lineages (Barbour & Travinsky 2010). The *ospC* locus in *B. burgdorferi* has a longer coalescence time than expected relative to neutrally evolving genes (Wang et al. 1999) and is under balancing selection (Brisson et al. 2004, Haven et al. 2011). Importantly, major *ospC* allele groups vary in their pathology and propensity to cause disseminated infections in humans (Dykhuizen et al. 2008).

The surface expressed protein OspA mediates the interaction between *B. burgdorferi* and the tick midgut (Pal et al. 2000) and *ospA* is upregulated while in the tick host but downregulated in the vertebrate host (Schwan et al. 1995, Schwan & Piesman 2000). Multiple studies have reported strong linkage disequilibrium between loci in *B burgdorferi* (Balmelli & Piffaretti 1996, Dykhuizen & Baranton 2001). However, patterns of linkage disequilibrium between loci are inconsistent between geographic regions (Travinsky et al. 2010), which may suggest adaptation to geographically distinct hosts and vectors.

There is evidence that *B burgdorferi* experiences differential survival between competent vectors (Dolan et al. 1997) and that the *Borrelia* genospecies exhibit varying degrees of vector specificity (Piesman & Gern 2004, Wodecka 2011). A study of *B burgdorferi* across the US found that populations of the pathogen cluster geographically, loosely corresponding to the primary vector ranges which could suggest adaptive evolution in *B. burgdorferi* to the vector species or neutral evolution as a result of vicariance and genetic drift (Margos et al. 2008, Margos et al. 2012).

For this dissertation, we collected samples from the Northeast, Midwest, Virginia, and North Carolina in the United States. We gathered additional samples from the NCBI dataset which were derived from the Northeast and California in the US and from France and Germany in Europe (Fig 1.1).

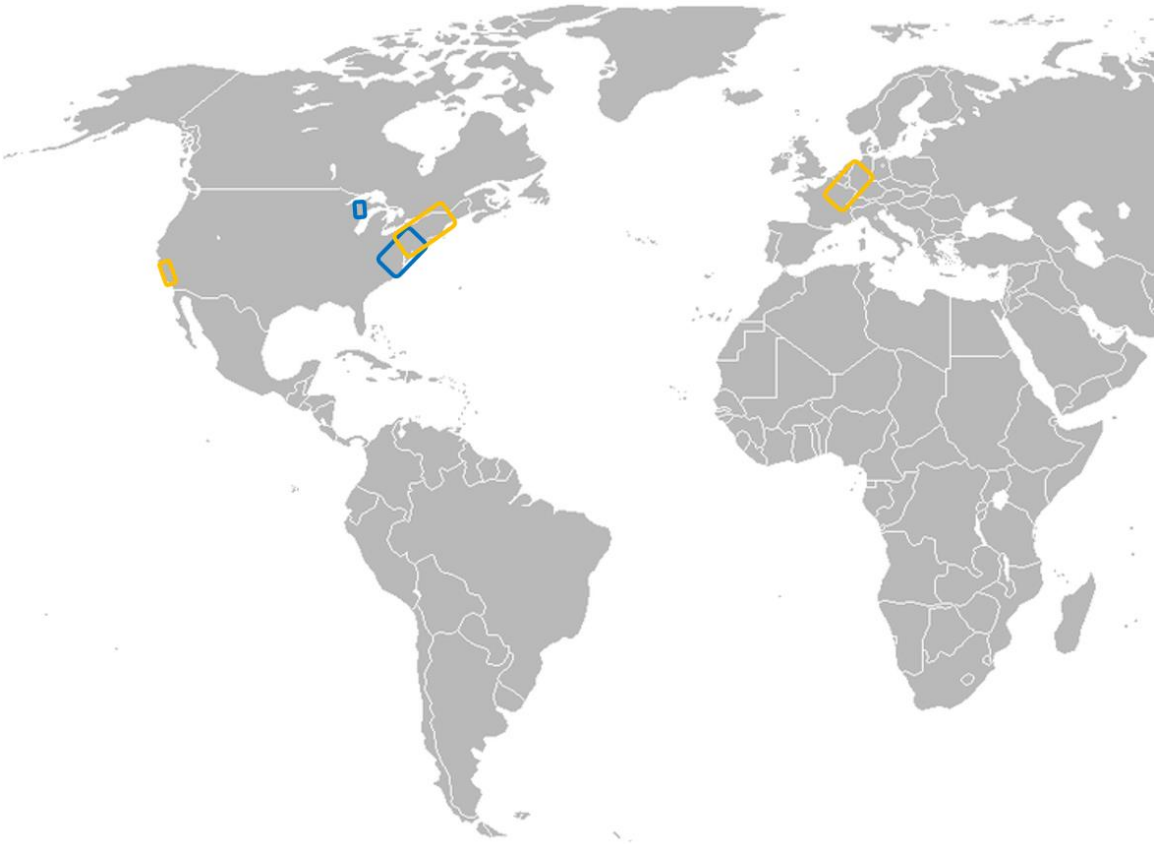


Figure 1.1 Map showing approximate sampling locations for samples sequenced for this thesis (blue) and samples acquired from the NCBI database (gold).

1.4 Scientific questions and dissertation structure

With changing host and vector distributions and a general increase in reported cases of Lyme borreliosis in the United States (Brinkerhoff et al. 2014), elucidating the factors contributing to *B. burgdorferi* population structure is vital for predicting future human risk.

Though *B. burgdorferi* is known to be vectored by multiple tick species, there are few studies investigating local adaptation to vectors and no studies in which population genetic data was analyzed from a region where *B. burgdorferi* is maintained by two competent vector species overlapping in geographic space. If *B. burgdorferi* is locally adapted to a tick vector at the cost of reduced fitness in other vectors, this could have profound implications for limitations to dispersal of *B. burgdorferi* populations. In this thesis, we investigate the population structure of *B. burgdorferi* at the interface of two expanding vector populations using a population genetic and phylogeographic approach. We then examine the influence of recombination between major *B. burgdorferi* lineages on the population structure of *B. burgdorferi* within and between geographic regions.

Chapter 2: We investigated the distribution of host-associated *ospC* major allele groups (oMGs) between geographic regions and competent vector species. We then describe the phylogeographic relationships and population structure between *B. burgdorferi* isolates in the Northeast, Virginia, and North Carolina. We found that the distribution of oMGs is associated with vector species and geographic regions. The phylogenetic tree showed that *B. burgdorferi* isolates tended to cluster by oMG rather than geographic origin, with some evidence of phylogenetic incongruity which may indicate recombination. Our population genetic analyses showed statistically significant population structure between and within geographic regions, suggesting that *B. burgdorferi* populations are structured both ecologically (by host-associated oMG) and geographically.

Chapter 3: We investigated the extent of linkage disequilibrium and recombination in *B. burgdorferi* populations within and between geographic regions. We found that loci are

nonrandomly associated (in linkage disequilibrium) throughout *B. burgdorferi* populations in the upper Midwest, Northeast, Virginia, and North Carolina. We also found evidence of limited recombination throughout the genome, though mutation is still more common, with relatively higher rates of recombination detected at the *ospC* locus and the *dbpA/dpbB* loci. Our study revealed low levels of genetic diversity with lineages, suggesting a small effective populations size. We also report that lineages are not always defined by oMG, and we concur with reports that the southern oMG-L strain is likely the result of trans-Atlantic dispersal.

CHAPTER 2: POPULATION GENETICS AND PHYLOGEOGRAPHY OF *BORRELIA BURGDORFERI* BETWEEN THE NORTHEASTERN AND SOUTHEASTERN UNITED STATES

2.1 Introduction

Changes in climate and land-use have altered species distributions throughout the world. Climate change is predicted to shift species distributions toward increasing latitudes and elevations with increasing average temperatures (Chen et al. 2011). At smaller geographic scales, climate change may lead to expansion or contraction of species ranges, changes in phenology, and changes in community interactions (Walther et al. 2002). Changes in land-use patterns can lead to habitat degradation and fragmentation, influencing species population structure and community composition (Haddad et al. 2015). Among those species predicted to be affected by climate change and changes in land-use are the hosts and vectors of many zoonotic pathogens, including the hosts and vectors of the Lyme borreliosis pathogen, *Borrelia burgdorferi* (Ogden et al. 2008, Mills et al. 2010).

An estimated 300,000 people are diagnosed with Lyme borreliosis in the US every year, manifesting in a wide range of symptoms or combinations of symptoms including erythema migrans rashes, Bells palsy, and arthritis (CDC 2015). *B. burgdorferi* is an obligate vector-borne pathogen, transmitted between vertebrate hosts by ticks in the genus *Ixodes*. *B. burgdorferi* has been found in a broad range of host species comprised of vertebrates from highly divergent taxonomic groups including songbirds, field mice, squirrels, and shrews (Barbour and Hayes 1986).

In the US, human Lyme borreliosis incidence is highest in the Northeast and upper Midwest where transmission is maintained by the tick, *Ixodes scapularis* (CDC 2015). The geographic distribution of *Ix. scapularis* extends from southern Canada to Florida (CDC 2016). A cryptic zoonotic *B. burgdorferi* transmission cycle is maintained in the southeastern US by the tick, *Ixodes affinis*, which has an infection prevalence in the southeast that is comparable to the infection prevalence of *Ix. scapularis* in the Northeast and Midwest (Maggi et al. 2010, Rudenko et al. 2013). While *Ix. affinis* is not believed to frequently parasitize humans, these ticks can amplify and maintain a zoonotic cycle of *B. burgdorferi* transmission which may then enter the *Ix. scapularis* population (Oliver 1996). Reports of a recent northern expansion of *Ix. affinis* into southern Virginia have generated renewed interest in studying the southeastern population of *B. burgdorferi* (Maggi et al. 2010, Rudenko et al. 2013). Very little is known about the southeastern *B. burgdorferi* population and its relationship to its southern vector, *Ix. affinis*.

Despite overlapping with *Ix. affinis* in the southeast, the *B. burgdorferi* infection prevalence in *Ix. scapularis* is low (0-4%) in the southeastern US relative to the infection prevalence in northern populations of *Ix. scapularis* (20-80%) (Courtney et al. 2003, Maggi et al. 2010, Clark 2004, Rudenko et al. 2013, Brisson personal communication). Population genetic studies suggest that *Ix. scapularis* is expanding southward in Virginia from the northern population (Kelly et al. 2014), corresponding to an increase in *B. burgdorferi* prevalence in *Ix. scapularis* and in humans in Virginia (Brinkerhoff et al 2014).

A study of *B. burgdorferi* across the northern and western US found that populations of the pathogen cluster geographically, with barriers to gene flow between the pacific coast and Midwest, and the Midwest and Northeast (Margos et al. 2012).

However, Humphrey et al. (2010) found that only a subset of *B. burgdorferi* haplotypes experienced a barrier to gene flow between the Midwest and Northeast. A recent multiple locus sequence typing (MLST) study from several locations in Canada identified multiple clonal complexes of *B. burgdorferi* distributed across a broad geographic region, suggesting population structure is defined ecologically rather than geographically (Mechai et al. 2015). Qiu et al. (2008) found population structure within geographic regions associated with an outer surface protein, OspC.

The *B. burgdorferi* lipoprotein OspC is required for early dissemination in the host (Tilly et al. 2006) and is among of the first antigens targeted by the host immune system during infection (Fung et al. 1994). OspC is expressed on the surface of *B. burgdorferi* in the tick midgut when the tick begins to engorge on host blood and is subsequently expressed during the early stages of infection in the vertebrate host (Gilmore & Piesman 2000, Schwan and Piesman 2000). The *ospC* locus in *B. burgdorferi* has a long coalescence time relative to neutrally evolving genes (Wang et al. 1999) suggesting that the locus is under balancing selection (Brisson and Dykhuizen 2004).

Genetic diversity at the *ospC* locus is partitioned into major allele groups (oMGs) which are classified by differences in the nucleotide sequence of >8% between oMGs and less than 2% within an oMG (Wang et al 1999). There are at least 28 unique oMGs identified with varied geographic distributions throughout North America and Europe. *B. burgdorferi* oMGs are not randomly distributed among different vertebrate host species in the environment, and each host species is susceptible to only a subset of oMGs (Brisson and Dykhuizen 2004). The pattern of oMG-associated host specialization is also reflected in human *B. burgdorferi* infections. Seinost et al. found that only eight of the eleven oMGs isolated from questing ticks in New York state were regularly recovered from human skin lesions, and only four of those oMGs were regularly

recovered in disseminated infections (1999, Figure 2.1). Subsequent studies have found an additional five oMG types frequently identified in human disseminated infections (Earnhart et al. 2005, Dykhuizen et al. 2008). Characterizing the distribution of oMGs in an environment is not only important in understanding the general ecology of *B. burgdorferi*, but also in interpreting human risk.

Previous studies have employed MLST to investigate population genetic structure in *B. burgdorferi*, however, a lack of genetic variation in the pathogen population has hindered the resolution of population genetic studies based on MLST (Margos et al. 2012, Qiu and Martin 2014, Seifert et al. 2015). Population genomics studies would be an ideal solution; however, culturing *B. burgdorferi* from ticks is challenging and may lead to culture bias (Liveris et al. 1999). We designed a novel long-range polymerase chain reaction (PCR) technique to amplify the conserved *B. burgdorferi* plasmids cp26 and lp54. The plasmid cp26 is the only plasmid required for growth in culture, and it encodes both *ospC* and a telomere resolvase necessary for replication of linear DNA molecules including the chromosome (Byram et al. 2004). Several studies have reported linkage disequilibrium between the *ospC* locus and chromosomal loci (Qiu et al. 2002, Bunikis et al. 2004, Barbour and Travinsky 2010). The plasmid lp54 is required to maintain the zoonotic transmission cycle and it encodes for 16 host regulated open reading frames including those coding for the surface expressed proteins OspA and OspB which are required for colonization of the tick midgut (Pal et al. 2000, Brooks et al. 2003, Yang et al. 2004). The expression of *ospA* and *ospB* is upregulated while in the tick host and downregulated while in the vertebrate host (Schwan et al. 1995, Schwan & Piesman 2000). Both cp26 and lp54 were found to be informative in a population genomic study of *B. burgdorferi* in France (Jacquot et al. 2014).

In this study, we investigated the distribution of *B. burgdorferi* oMGs in the southeastern US where *Ix. affinis* is the predominant vector, relative to the Northeast US where *Ix. scapularis* is the dominant vector, as well as southern Virginia where both tick species were found to be infected with *B. burgdorferi*. We then examined the effect of tick species and geographic region on the population structure of *B. burgdorferi* using a phylogeographic approach.

2.2 Methods

Tick collection

Questing ticks were collected in Pennsylvania (2014 & 2016), Delaware (2014), southern Virginia (2014), and North Carolina (2014)(Table 2.1). Ticks were collected using the drag-flag method, in which a $1m^2$ white corduroy flag is dragged along the forest floor and undergrowth. Ticks were then morphologically identified as *Ixodes* sp., removed from the flag with forceps, and stored live in 2mL microcentrifuge tubes until they could be deposited in a -80℃ freezer. To prevent desiccation before freezing, one ~4mm cube cut from an autoclaved sponge and dampened with DI water was placed in each microcentrifuge tube.

DNA extraction, diagnostic PCR, tick species identification

Individual ticks were crushed into a homogenate using a barrier pipette tip in 20µL (nymphs) or 30µL (adults) of filter sterilized PBS (pH 7.4). DNA was then extracted from 15µL of the tick homogenate with the Qiagen PureGene Tissue Kit for gDNA. The protocol was modified for high molecular weight DNA extraction from nymphal and adult ticks (protocol upon request). Individual ticks were then screened for *B. burgdorferi* using two semi-nested PCRs, one targeting the rrs-rrlA intergenic spacer (IGS) and one targeting the *ospC* locus (IGS - Bunikis et al. 2004, OspC - Vuong et al.

2014). *Ixodes* ticks from VA and NC were identified as *Ix. affinis* or *Ix. scapularis* using a multiplexed real-time PCR assay (Wright et al. 2014).

LR-PCR, next-generation sequencing

Six semi-nested primer sets were designed to amplify 12-15kb fragments of the *ospC*-encoding *B. burgdorferi* plasmids, cp26 and lp54, for amplification via long range PCR (LR-PCR) using the PrimeSTAR GXL Polymerase (Takara Bio). Primers and thermal cycling conditions are listed in Supplementary Table S1A-B. Amplified LR-PCR products were run on a 0.5% agarose gel for purification using the QIAEX II Gel Extraction kit (Qiagen). DNA concentration of each purified LR-PCR product was measured on a qubit 3.0 fluorometer (Thermo Fisher Scientific) and the six LR-PCR products were pooled at equal concentrations for each tick. We used 15ng of the pooled LR-PCR product to create short-insert libraries (Nextera-XT DNA Library Prep Kit) and sequenced the LR-PCR products on an Illumina MiSeq or Miniseq, producing 150bp paired-end reads. In order to minimize the potential for batch effects due to multiple sequencing runs, samples from multiple regions and sites were multiplexed in each lane. The reads were trimmed for Illumina adapters and low quality bases (q<25) using Trim Galore! (Krueger 2015) and mapped to the cp26 and lp54 from the *Borrelia burgdorferi* B31 reference genome (BioProject PRJNA3) using smalt version 0.7.4 (sanger.ac.uk/resources/software/smalt/).

Variant calling and sequence alignment

Variant calling for singly infected ticks was completed using FreeBayes with the following parameters: ploidy = 1, min coverage = 4, no population priors, minimum mapping quality = 30, and minimum alternate fraction = 0.75. Consensus sequences for each *B. burgdorferi* plasmid from singly infected ticks were generated with an in-house script against the B31 reference, with ‘N’ inserted when coverage was too low for variant

calling (<4X). Multiple sequence alignment for cp26 and lp54 was completed using MAFFT v. 7.222 (Katoh et al. 2002) implemented in Geneious vs. 10.0.2 (www.geneious.com, Kearse et al. 2012). Individual sequences were not included in further analysis if more than 25% of the sequence contained 'N' not found in other sequences. Remaining sequences were then trimmed for ambiguities and gaps.

Identification of oMGs

For samples with >50X average coverage at the *ospC* locus, mapped reads were extracted and each read was BLASTed against a local database of 42 *ospC* sequences belonging to 29 distinct oMGs, including *B. bissettii* (supplementary figure S2). Reads matching >92% to an oMG were assigned to that group while reads with no match >92% in the local database were blasted against the full NCBI database. In 14 instances when reads had no match in the local database, the NCBI BLAST resulted in matches to the closely related genospecies, *Borrelia bissettii*. oMG assignments comprising >5% of total reads were counted as present in the tick. Corrections were made to account for sequences in the local blast database in which multiple sequences belong to a single oMG. For example, oMG-Fa and oMG-Fb share >98% nucleotide sequence similarity. Matches to either of those references in the local database were counted as a match to "oMG-F". When a single read matched to multiple oMGs in the library, it only contributed to the final count if additional reads provided further discriminatory power by mapping to a different section of the *ospC* locus. We were unable to identify oMGs from reads mapped to samples in which the tick was infected with more than 4 oMGs.

In order to test our LR-PCR and oMG identification method, we mixed gDNA extracted from cultured *B. burgdorferi* isolates representing oMG-A, oMG-E, and oMG-J at equal concentrations with a combination of two oMGs (A & J) and a combination of three oMGs (A & J &E), we then amplified cp26 and lp54 using our LR-PCR protocol and

computationally identified which oMGs were present using the local BLAST and bin method described above. We were able to identify both combinations accurately.

Statistical analysis of ospC major allele group distributions

We investigated the effects of geographic region and tick species on the distribution of oMGs using Constrained Analysis of Principal Coordinates (CAP) implemented in R as part of the vegan package (R Core Team 2016, Oksanen et al. 2016). CAP ordinates a dissimilarity matrix, then tests for differences among treatment groups using a redundancy analysis (Anderson and Willis 2003). We used the Raup-Crick distance metric which is a probabilistic index that calculates the dissimilarity between communities under a null model from presence/absence data (Pos et al. 2014, Oksanen et al. 2016). We split the data into four grouping variables which include a tick species factor and a geographic origin factor in order to best disentangle the effects of tick species and geographic origin on the distribution of oMGs given the geographic distribution of ticks sampled: *Ix. scapularis* in the Northeast, *Ix. scapularis* in Virginia, *Ix. affinis* in Virginia, *Ix. affinis* in North Carolina. Statistical significance for the CAP analysis was determined by an ANOVA-like permutation test (n = 999) to assess significance of tick species and geographic region on the distribution of oMGs (Oksanen et al. 2016).

Population genetics

Population genetic parameters including the number of segregating sites, haplotype diversity, and nucleotide diversity (Table 2.2) were calculated using DnaSP v5.10.01 (Librado and Rozas 2009). Population genetic structure between regions (Northeast, Virginia, and North Carolina) and sites (sampling locations within each region) was assessed using the Analysis of MOlecular Variance (AMOVA)(Table 2.3) calculated using the POPPR package in R (R Core Team 2016, Kamvar et al. 2014).

Phylogenetic analyses

Phylogenetic relationships for each plasmid were reconstructed using both maximum likelihood and Bayesian methods in order to confirm tree topology. We used the BIC, implemented in jModelTest, to determine the best-fit model of nucleotide substitution for cp26 and lp54 sequence alignments; the general time reversible model of sequence evolution and inverse gamma-distributed rate variation (GTR+I+G) was then used in the Bayesian phylogenetic reconstruction and GTR + G was used in the PhyML phylogenetic reconstruction (Darriba et al. 2012). Bayesian phylogenetic trees were inferred with MrBayes v3.2.5 and the Monte Carlo Markov Chain was run for 1,000,000 generations with four heated chains and sampled every 1000 generations (Ronquist et al. 2012). Posterior probability at nodes show support for Bayesian tree topology and trees were rooted with an outgroup (*Borrelia afzelii*, BioProject PRJNA224116). Final trees were visualized with Figtree v1.4.3 and branches were labeled by geographic origin of samples (tree.bio.ed.ac.uk/software/figtree). Maximum likelihood phylogenetic trees were inferred with PhyML v 3.0 with parameters set to estimate equilibrium frequencies, transition:transversion ratios, and gamma shape (Guindon et al. 2010). We calculated the relative rates of observed recombination and mutation using ClonalFrameML with transition:transversion ratios and starting trees estimated using PhyML (Didelot and Wilson 2015).

2.3 Results and Discussion

Overall prevalence

A total of 79 of 133 (59%) *Ix. affinis* ticks collected from North Carolina and Virginia were PCR positive for *B. burgdorferi* (Table 2.1). We were able to amplify, sequence, and identify *B. burgdorferi* oMGs for 65 *Ix. affinis* ticks. Among *Ix. affinis*

ticks, we identified a total of 12 oMGs in NC and 8 oMGs in VA (Figure 2.2). A total of 45 of 220 (20%) *Ix. scapularis* ticks collected from the Northeast and Virginia were PCR positive for *B. burgdorferi* and none of 14 *Ix. scapularis* ticks in North Carolina were PCR positive for *B. burgdorferi* (Table 2.1). We were able to amplify, sequence, and identify 11 oMGs for 22 *Ix. scapularis* ticks (Figure 2.3). A total of 17 different oMGs were identified in this study, and six oMGs were found in both tick species (A, B, C, D, G, H).

The coastal region of southern Virginia contains the upper geographic boundaries for several *B. burgdorferi* host species including the cotton mouse, *P. gossypinus* (Wolfe and Linzey 1977), and the hispid rat, *Sigmodon hispidus* (Hall 1959). *B. burgdorferi* cultures isolated from hispid rats and cotton mice collected in South Carolina and Georgia were found to have a high incidence of *B. burgdorferi* oMG-L, which is rare in the Northeast (Rudenko et al. 2013). We found oMG-L in *Ix. affinis* in North Carolina and in Virginia, but we did not recover oMG-L from *Ix. scapularis*.

We found that *Ix. affinis* ticks in NC had the most oMGs with an average of 2.04 oMGs per infected tick, while *Ix. affinis* in VA had on average 1.33 oMGs per infected tick. *Ix. scapularis* in the Northeast and VA had an average of 1.60 and 1.24 oMGs per tick respectively. Four individual ticks in the NC population of *Ix. affinis* appeared to contain more than four strains and were excluded from the analysis. We found that among collections in April 2014, all *Ix. scapularis* in VA and nearly all (1 adult, 193 nymphs) *Ix. scapularis* in the Northeast were nymphs while all *Ix. scapularis* in North Carolina were adults (Table 2.1). All *Ix. affinis* ticks collected in the spring were adults, with only a few (7 nymphs,11 adults) nymphs recovered in the late summer.

The observed differences in life history structure between *Ix. scapularis* and *Ix. affinis* likely influenced both the number of *B. burgdorferi* oMGs per tick and the yield of the LR-PCR method. Because adult ticks have fed twice, once as a larva and once as a

nymph, they are more likely to carry multiple *B. burgdorferi* oMGs than nymphal ticks, which have only fed once. In addition, a nymph will take a larger bloodmeal than a larva, resulting in more *B. burgdorferi* on average in the midgut of an infected adult tick than an infected nymph. Given that the LR-PCR is sensitive to both DNA quality and concentration, it is likely that the low yield of amplified LR-PCR fragments from *Ix. scapularis* samples relative to *Ix. affinis* samples is a result of the reduced number of *B. burgdorferi* in each nymph relative to each adult. The results of the diagnostic PCRs for the IGS and *ospC* locus were within expectations given prevalence reports in previous publications (Maggi et al. 2010, Rudenko et al. 2013, Clark 2004, Courtney et al. 2003, Brisson personal communication).

High *B. burgdorferi* prevalence in *Ix. scapularis* is reinforced by seasonally structured synchronous phenology (Gatewood et al. 2009). When there is a large difference between summer and winter temperatures, as in the Northeast where winter temperatures are frequently near or below 0°C, unfed nymphs will enter diapause during the winter months and emerge in large numbers in the spring. The emergence of nymphs in the spring is followed by the emergence of larvae in the summer. The newly hatched larvae then feed on the host community, some proportion of those hosts will have recently been infected with *B. burgdorferi* by one of the nymphs. One hypothesis for low *B. burgdorferi* prevalence in the southeastern population of *Ix. scapularis* is that larvae and nymphs emerge asynchronously throughout the year and transmission of *B. burgdorferi* is not reinforced by the nymph-then-larvae pattern of synchronous emergence. In 2014, the winter in southern VA was unusually harsh and protracted, with an unseasonable snowfall occurring as late as March (usclimatedata.com), which may explain the unusual synchronous emergence of nymphs in April. Our collections were not stratified in time and space in a way that allow for a rigorous study of the

influence of climate on the phenology of ticks and the distribution of oMGs. However, the lack of life stage diversity in our tick collections within regions does suggest a synchronous phenology which is linked to climate in *Ix. scapularis*.

Though the sample size was considerably smaller in VA than in NC for *Ix. affinis* (n = 16, n = 125 respectively), we still recovered seven of the twelve oMGs found in NC. It was surprising to find such diversity in oMGs given the recent expansion of *Ix. affinis* into southern Virginia (Nadolny et al. 2011) and the low expected infection prevalence for *Ix. scapularis* in the southeast (Clark et al. 2004, Maggi et al. 2010, Gaff personal communication). None of the oMGs in the *Ix. affinis* from VA were unique relative to the oMGs found in *Ix. affinis* from NC; however, five of the seven oMGs found in *Ix. affinis* from VA were not found in *Ix. scapularis* from the NE or VA. We would expect this pattern if *Ix. affinis* and *Ix. scapularis* fed on different host communities, though both vector species are generalists and there is overlap in the known host range (Harrison et al. 2010).

oMG community composition

Our analysis shows a statistically significant difference between the composition of oMGs between tick species and geographic regions ($F_{3,80} = 4.69$, $P = 0.001$). The first and second axes from the constrained analysis of principal coordinates (CAP1 & CAP2) account for 89.19% of the constrained variation (Figure 2.5). The CAP plot shows a separation in ordination space driven by both tick species and geographic origin, suggesting that the geography and tick species account for a substantial amount of variation in oMG. This analysis is likely confounded by the small sample size for oMGs from *Ix. scapularis* ticks which may not be representative of the oMG diversity in the region (5 ticks, 8 *B. burgdorferi*, 3 oMGs); however, the skewed proportion of oMG-A in

both tick species in VA may reflect the early establishment of this oMG at the leading edge of both expanding vector ranges.

Human strains

Several oMGs reported to cause EM lesions and/or disseminated infections in humans (Seinost et al. 1999, Earnhart et al. 2005, Dykhuizen et al. 2008) were identified in both tick species, including oMG-A, oMG-B, oMG-C, oMG-D, oMG-G, and oMG-H (Figure 2.3). We found oMG-B and oMG-H in both *Ix. scapularis* the Northeast and *Ix. affinis* in NC, but neither oMG was identified in VA, possibly due to small sample sizes. In the NE and NC, oMG-A comprised ~19% of the total distribution of oMGs, while in VA oMG-A comprised ~37% the total distribution of oMGs in *Ix. scapularis* and 50% of the total distribution of oMGs in *Ix. affinis*.

Population genetics

We found 627 polymorphic sites among 26 individuals across 18,283bp of the plasmid cp26 and 669 polymorphic sites among 22 individuals across 38,682bp of the plasmid lp54 (Table 2.2). For both cp26 and lp54, haplotype diversity was lower in Virginia than in North Carolina or the Northeast (Table 2.2). This pattern was driven by the abundance of oMG-A isolates in VA which comprised 8 of the 10 samples from VA for each plasmid and lacked any variation across both cp26 and lp54. The lack of diversity found in oMG-A is consistent with a previous study which suggested that oMG-A is a recent migrant to North America from Europe (Qiu et al. 2008). We found no variation among oMG-A haplotypes in VA between *Ix. scapularis* (n=1) and *Ix. affinis* ticks (n=7) across cp26 and lp54, which lends support to the potential for *B. burgdorferi* to move between tick species. Our motivation for developing a novel method for amplification of partial plasmids was to provide better resolution for population genetics

studies. The low haplotype diversity suggests that we were unable to capture variation within clonal lineages even when 56,965bp are included in the analysis.

Across all geographic regions, nucleotide diversity was higher in cp26 than lp54 (Table 2.2). Much of the diversity in cp26 is localized at the host-associated *ospC* locus, though we found SNPs between oMGs across the plasmid. Similarly, we found SNPs distributed across plasmid lp54 between oMGs. Nucleotide diversity at lp54 across all regions in this study, 0.005 +/- 0.001, was similar to the genome-wide nucleotide diversity reported for 22 *B. burgdorferi* isolates, 0.004 +/- 0.003, recently sequenced in Europe (Becker et al. 2016), whereas cp26 was double that value at 0.010 +/- 0.001.

Results from the AMOVA revealed significant variation within and between regions but not between sites within regions for both cp26 and lp54 (Table 2.3). We found that differentiation within sites accounted for more of the variation than between regions (Table 2.3). Natural selection can enhance geographic differentiation when populations adapt to local environments (Slatkin 1987), but in obligate pathogen systems the host is the local environment (Brooks et al. 2006). For *B. burgdorferi*, there may be many potential host species present in a single geographic region. The identification of population structure within sites in *B. burgdorferi* suggests that the effects of ecological isolation are greater than the effects of geographic isolation in structuring *B. burgdorferi* populations.

Phylogenetic analyses

A phylogenetic tree was inferred from plasmid cp26 with 26 *B. burgdorferi* sequences from an alignment of 18,283bp (Figure 2.4). A phylogenetic tree for lp54 was inferred with 23 *B. burgdorferi* sequences from an alignment of 38,682bp (Figure 2.4). Both trees show similar patterns of phylogenetic clusters associated with oMG rather than geographic origin or vector species. Within oMG lineages, there is some evidence

of clustering by geographic region. For example, in the oMG-A clade for both cp26 and lp54 we find within-region clustering for samples from VA and samples from the NE.

There were some instances of phylogenetic incongruity among the trees for cp26 and lp54. Specifically, a Northeast oMG-U is clustered within the oMG-G group in the lp54 tree and the relationships between individuals within the oMG-A clade are inconsistent between cp26 and lp54, though the oMG-A clade remains monophyletic in both trees (Figure 2.4). These results indicate that possible recombination or incomplete lineage sorting has influenced the tree topology in *B. burgdorferi* and disrupted linkage disequilibrium between *ospC* and other loci.

We estimated the rate of recombination to mutation (R/Θ) for cp26 and lp54 at 1.462 with a posterior variance of 0.0075 and 0.166 with a posterior variance of 0.0003 respectively. The average length of the recombinant DNA fragment (*δ)* for cp26 and lp54 was 98bp and 50bp respectively. Most of the detected recombination in cp26 centered around the *ospC* locus which is expected given that new oMGs often arise via recombination (Barbour and Travinsky 2010). A study of the core genomes across multiple Borrelia genospecies found a R/Θ rate of 0.182 (Becker et al. 2016), which is similar to our reported value for lp54. The clonal frame model does not allow for multiple recombination events at any position, thus when the parameter *δ*R is greater than one, the clonal frame model is prone to underestimating the value of R (Didelot and Wilson 2015). The *δ*R for cp26 was approximately 1.39, indicating that recombination is likely to have occurred more than once at some positions, and 14 of 39 recombination events detected across cp26 occurred at the *ospC* locus. These analyses show that recombination is an important force in generating diversity at the *ospC* locus.

2.4 Conclusions

In this study, we investigated the population genetic structure of *B. burgdorferi* in the Northeast, where *Ix. scapularis* is the dominant vector, in North Carolina, where *Ix. affinis* is the predominant vector, and in Virginia at the interface of these two competent vector species. We found that the distribution of *ospC* major allele groups is associated with geographic region and vector species, suggesting that the differences in the host communities between the northeastern and the southeastern US could be contributing to the differences in oMG distribution between the regions. Despite differences in the distribution of oMGs, we found no evidence of differentiation in *B. burgdorferi* between vectors within an oMG, which refutes the hypothesis that local adaptation to the vector *Ix. affinis* suppresses infection prevalence in the sympatric competent vector, *Ix. scapularis*.

Our population genetic and phyogeographic analyses further showed population structuring within geographic regions associated with oMGs, which supports the hypothesis of ecological isolation rather than geographic isolation as the primary driver of population structure in *B. burgdorferi*. The independent reconstruction of phylogenies showing similar topologies from different plasmid sequences suggests that *B. burgdorferi* is largely clonal, though we found evidence of some recombination and phylogenetic incongruence, suggesting a disruption of clonal lineages associated with the *ospC* locus.

Acknowledgements

The authors would like to thank Marcee Toliver for sharing expertise regarding the ecology and distribution of *Ix. affinis* and *Ix. scapularis* in North Carolina. Collections in VA at the Hoffler Creek Wildlife Preserve were completed under the guidance of Dr. Holly Gaff. Collecting permits for the North Carolina State Park System were granted by the North Carolina Department of Environment and Natural Resources (R14-23). Permissions for tick collection in Pennsylvania at Crow's Nest Preserve and Wawa Preserve were granted by the Natural Lands Trust, and we thank Dan Barringer for helpful tips. We thank Dr. Solny Adalsteinsson and the Forest Fragments in Managed Ecosystems program for collection assistance in Delaware. Sequencing and DNA library preparations were completed in the Hahn Lab Core facility at the University of Pennsylvania. Emily Avis contributed to computational sequence processing. The authors would like to thank Dr. Lee Dietterich for guidance and assistance on statistical data analysis as well as Alex Berry, Dr. Sandy Olkowski, and Jill Devine for helpful discussions. Stephanie Seifert was supported by NSF GRFP DGE-1321851.

2.5 Figures and tables

Table 2.1. Tick collection data and results of diagnostic PCR targeting *B. burgdorferi* IGS and *ospC*

State	Location	Tick Species	Number of Ticks	% Nymphs	Positive for Bb	% Bb	Latitude	Longitude
Delaware (NE)	FRAME site	*Ix affinis*	0	NA	0	NA	39.64534	-75.75786
		Ix scapularis	46	100%	5	11%		
Pennsylvania (NE)	Crow's Nest Preserve	*Ix affinis*	0	NA	0	NA	40.18846	-75.75512
		Ix scapularis	138	99%	29	21%		
Pennsylvania (NE)	Wawa Preserve	*Ix affinis*	0	NA	0	NA	39.90791	-75.45748
		Ix scapularis	9	100%	2	22%		
Virginia	Hoffler Creek Wildlife Foundation	*Ix affinis*	17	0%	13	76%	36.89452	-76.39748
		Ix scapularis	27	100%	9	33%		
North Carolina	Cliffs of the Neuse State Park	*Ix affinis*	14	0%	7	50%	35.24124	-77.88387
		Ix scapularis	0	NA	0	NA		
North Carolina	Medoc Mountain State Park	*Ix affinis*	31	0%	13	42%	36.26194	-77.88714
		Ix scapularis	1	0%	0	0%		
North Carolina	Merchant's Millpond State Park	*Ix affinis*	71	10%	46	65%	36.43446	-76.69985
		Ix scapularis	13	8%	0	0%		
Total	All locations	*Ix affinis*	133	5%	79	59%	NA	NA
		Ix scapularis	234	94%	45	19%		

Table 2.2. Origin, sample sizes, and sequence diversity of *Borrelia burgdorferi* plasmid sequences from the Northeast, Virginia, and North Carolina.

Region	Plasmid	n	S	h	Hd	π
Northeast	cp26	10	519	8	0.956	0.011 (0.001)
	lp54	10	501	10	1.000	0.005 (0.001)
Virginia	cp26	10	225	2	0.356	0.004 (0.002)
	lp54	10	260	3	0.378	0.002 (0.001)
North Carolina	cp26	6	334	3	0.600	0.008 (0.003)
	lp54	2	280	2	1.000	0.007 (0.003)
All Locations	cp26	26	625	13	0.886	0.010 (0.001)
	lp54	22	669	15	0.879	0.005 (0.001)

n = sample size, S = segregating sites, h = number of haplotypes, Hd = haplotype diversity, π = nucleotide diversity, standard deviation in parentheses

Table 2.3. Analysis of molecular variance for plasmid sequence data of *B. burgdorferi* among the Northeast, Virginia, and North Carolina (regions) and sampling sites within regions (sites)[a]

	Variation (%)	Φ	*P*
cp26			
Φ_{CT} (among regions)	29.3	0.2927	0.016*[b]
Φ_{SC} (among sites within regions)	6.5	0.0919	0.630[c]
Φ_{ST} (within sites)	64.2	0.3577	<0.001*[d]
lp54			
Φ_{CT} (among regions)	35.6	0.357	<0.001*[b]
Φ_{SC} (among sites within regions)	-5.10	-0.079	0.883[c]
Φ_{ST} (within sites)	69.4	0.306	0.002*[d]

[a] Variance was partitioned among and within three geographic regions and within sampling sites

[b] Probability of obtaining value for Φ was determined by 1000 randomizations by permuting samples between regions

[c] Probability of obtaining value for Φ was determined by 1000 randomizations by permuting samples within regions

[d] Probability of obtaining value for Φ was determined by 1000 randomizations by permuting samples within sites

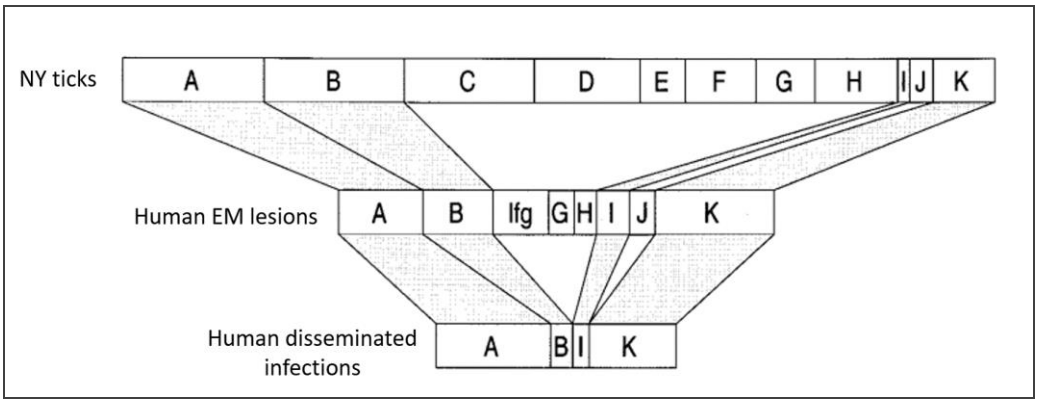


Figure 2.1. Distribution of oMGs from questing ticks, human erythema migrans lesions, and human disseminated infections in New York State.
(adapted from Seinost et al. 1999)

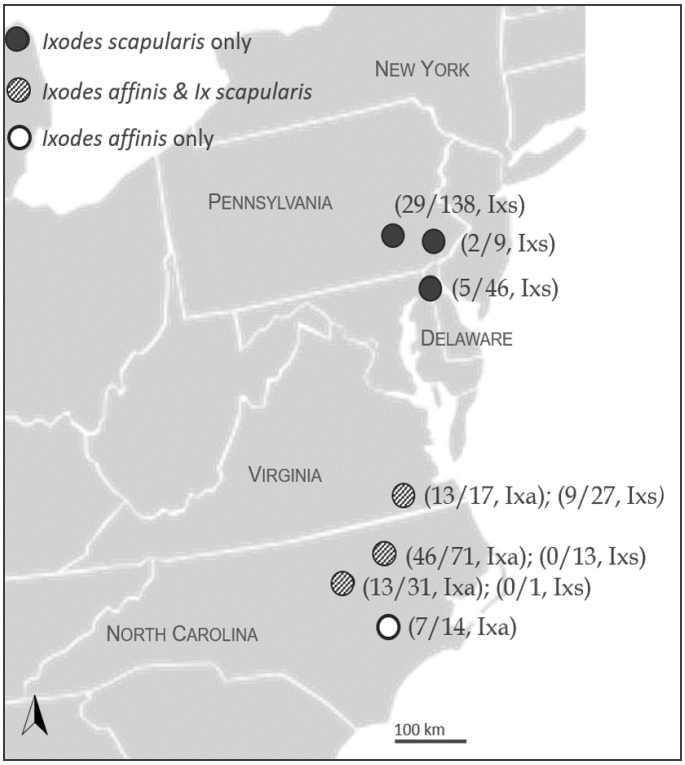


Figure 2.2. Tick sampling sites in Pennsylvania and Delaware (Northeast), Virginia, and North Carolina. Values in parentheses next to each sampling location represent the number of *B. burgdorferi* positive ticks, the total number of ticks, and the tick species (Ixa = *Ix. affinis*, Ixs = *Ix. scapularis*).

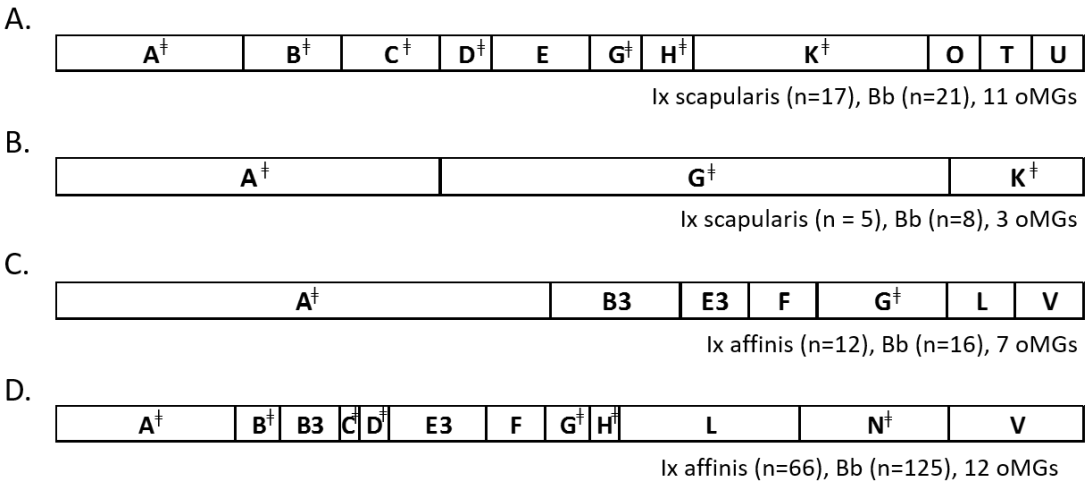


Figure 2.3. The distribution of *ospC* major allele groups among *Ix. scapularis* ticks collected in the (A.) Northeast, (B.) *Ix. scapularis* ticks collected in Virginia, (C.) *Ix. affinis* ticks collected in Virginia, and (D.) *Ix. affinis* ticks collected in North Carolina.
‡ indicates an oMG indicated in human skin or disseminated infections.

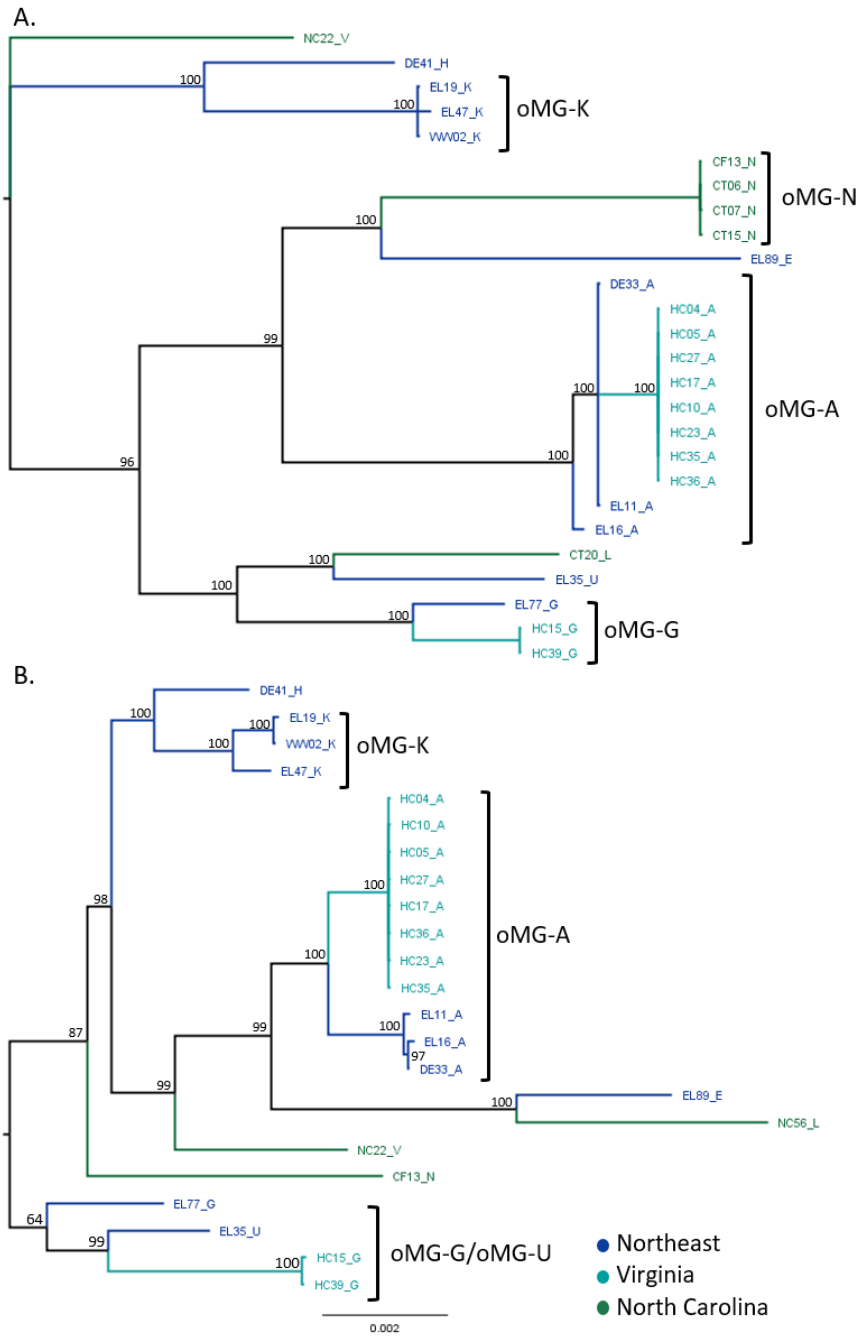


Figure 2.4. Bayesian phylogenetic tree of (A) cp26 sequences and (B) lp54 sequences from *Borrelia burgdorferi* collected in the NE (blue), VA (aquamarine), and NC (dark green). The tree for cp26 is based on 26 sequences of 18,283bp. The tree for lp54 is based on 23 sequences of 38,682bp. Support at nodes are given by the posterior probability. The scale bar indicates substitutions per site.

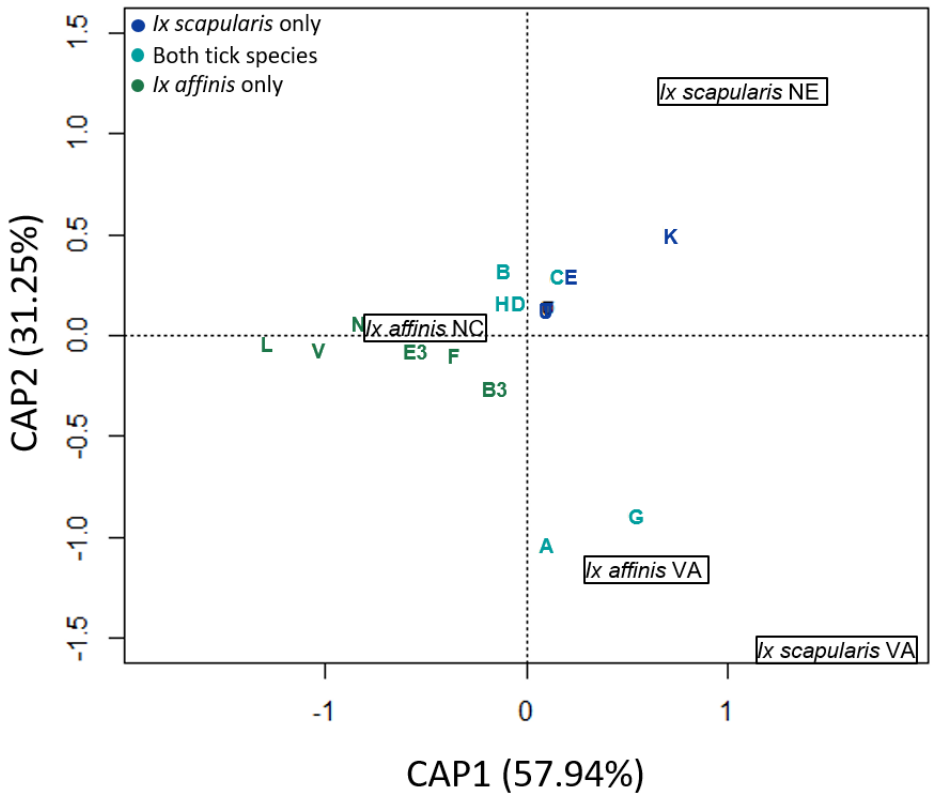


Figure 2.5. Constrained analysis of principal coordinates of Raup-Crick distance based on tick species and geographic origin of *B. burgdorferi* oMGs. Labels represent the centroids of the average composition of oMGs among tick species by geographic region; NE = Northeast, VA = Virginia, NC = North Carolina. The letters represent each oMG and the colors indicate whether that oMG is found in *Ix. affinis* only (green), *Ix. affinis* and *Ix. scapularis* (aquamarine), or *Ix. scapularis* only (blue).

CHAPTER 3: POPULATION GENOMICS OF *BORRELIA BURGDORFERI* REVEAL LOW LEVELS OF RECOMBINATION AT THE OSPC LOCUS, DISRUPTING CLONAL ASSOCIATION WITH *OSPC* MAJOR ALLELE GROUPS

3.1 Introduction

Microbial populations are shaped by population genetic processes including mutation, natural selection, genetic drift, and recombination. Though bacteria are generally described as clonal organisms, vertically transmitting genetic material from parent to offspring, the exchange of genetic material can also occur via three alternative mechanisms: 1) transduction, in which a genetic material is moved between bacterial cells by a viral intermediary, 2) transformation, in which donor DNA is taken up from the environment by the recipient, and 3) conjugation, in which DNA is exchanged through cell-to-cell contact. Recombination has been shown to contribute to diversification in structured bacterial populations if genetic material is horizontally transferred between independently evolving lineages (Guttman & Dykhuizen 1994) and can serve to accelerate adaptation by reducing competition among beneficial mutations by beneficial mutations together (Vos et al. 2009) or allowing deleterious mutations to be purged (Kim and Orr 2005, Vos et al. 2009).

Recombination breaks up the clonal pattern of inheritance, thus the strength of linkage disequilibrium is negatively correlated with recombination (Maynard Smith 1993). When reconstructing the phylogenetic relationships between clonal individuals, regions where linkage disequilibrium is disrupted by recombination would be incongruous with the clonal tree (Smith and Smith 1998, Vos 2009, Didelot and Wilson 2015). As a single recombination event can introduce multiple mutations, recombination can inflate estimates of genomic divergence between closely related taxa (Didelot and Wilson

2015). Therefore, accounting for recombination is important when making phylogenetic and population genetic inferences.

Interest in the study of recombination in bacteria has been fueled by the decrease in DNA sequencing costs and an interest in tracing the spread of antimicrobial resistance and pathogenicity islands between lineages of bacterial pathogens (Dobrindt et al. 2004). Bacteria range from highly clonal (e.g. *Salmonella enterica*), with nonrandom association between loci, to highly recombinogenic (e.g. *Helicobacter pylori*), with nearly random association between loci (Spratt 2004). The causative agent of Lyme borreliosis, *Borrelia burgdorferi*, has been described in contradictory terms as completely clonal (Dykhuizen et al. 1993) and subject to pervasive recombination (Haven et al. 2011). Early studies found little evidence of recombination, but only a few chromosomal loci were evaluated (Dykhuizen et al. 1993, Smith and Smith 1998). When the first whole *B. burgdorferi* genomes were sequenced, researchers selected very divergent *B. burgdorferi* isolates for sequencing in order to capture the most diversity. Population genetic inferences based on these early genomes likely overestimated recombination as they were based on a sample set selected to represent global diversity rather than a random selection from a population (Seifert et al. 2015). Subsequent studies which included sequences from chromosomal and plasmid-borne loci or whole genomes have suggested that *B. burgdorferi* is mostly clonal with limited recombination (Qiu et al. 2008, Jacquot et al. 2014).

B. burgdorferi is an obligate vector-borne pathogen with a broad range of vertebrate hosts. Using long-range PCR to amplify the *B. burgdorferi* plasmids cp26 and lp54 from ticks in the Northeast, Virginia, and North Carolina we identified population structure associated with a the *ospC* locus (Chapter 2). We found that most of the reported recombination across the cp26 plasmid was localized at the *ospC* locus. Intra-

population recombination can be difficult to detect relative to the importation of DNA from more distantly related pathogens as fewer, if any, mutations may exist in the recombined fragment, thus intra-population recombination rates are likely to be under-estimated (Didelot and Wilson 2015). Recombination at the *ospC* locus may be more readily detected than recombination at other loci because the locus is under balancing selection (Livey et al. 1995).

The *ospC* locus is partitioned into major allele groups (oMGs) which differ by less than 2% in nucleotide identity within groups and more than 8% between groups (Wang et al 1999). There is evidence of linkage disequilibrium between ospC and chromosomal loci (Dykhuizen et al. 1993, Qiu et al. 2008), however there is also evidence of recombination in part or in whole at the *ospC* locus (Qiu et al. 2008, Barbour and Travinsky 2010, Jacquot et al. 2014). There has not been a consensus on the extent of linkage associated with the *ospC* locus in *B. burgdorferi*. In this study, we investigate patterns of linkage disequilibrium in *B. burgdorferi* associated with the *ospC* locus.

3.2 Methods

Bacterial cultures

We resurrected twenty-four *B. burgdorferi* isolates previously cultured for a 2010 study on evolution of *B. burgdorferi* in the Midwest which had been identified to contain a single oMG (Brisson et al. 2010). We also cultured a standard laboratory strain, B31 (passage = 5), for which the whole genome sequence has been resolved (BioProject PRJNA3) in order to test our sequencing and variant calling pipeline. The Midwest isolates were originally cultured from skin biopsy specimens, with the exception of BC3_N which was isolated from cerebrospinal fluid, from adult patients at the Marshfield

Clinic in central Wisconsin from 1995-2001. Isolates were grown in 6mL of BSK-II culture medium in 8mL polystyrene round-bottom screw cap tubes at 34℃ under BioSafety Level 2 conditions as described by Zuckert (2007). High molecular weight gDNA was extracted from *B. burgdorferi* cultures in late-exponential-phase using the Qiagen PureGene DNA extraction kit following the protocol for cultured cells.

LR-PCR, next generation sequencing

DNA libraries were prepared from the whole genome DNA from the *B. burgdorferi* cultures using the Nextera-XT DNA Library Prep Kit and multiplexed on a single Miniseq lane yielding 150bp paired-end reads. Reads were then trimmed for Illumina indexing primers, adapters, and quality <25 using Trim Galore! (Krueger 2015). Trimmed reads were then mapped against reference sequences for the chromosome, cp26, and lp54 using the *B. burgdorferi* B31 reference genome (BioProject PRJNA3) using smalt version 0.7.4 (sanger.ac.uk/resources/software/smalt/). Variant calling and oMG identification for the *B. burgdorferi* cultures was completed as described in Chapter 2 for singly infected ticks. Twenty-Four *B. burgdorferi* cultures from WI were included in this study in addition to B31 and the cp26 and lp54 plasmid sequences from singly infected ticks as described in Chapter 2. In addition to the cultured *B. burgdorferi* isolates and LR-PCR sequences from singly infected ticks in the eastern US (Chapter 2), we included *B. burgdorferi* sequences acquired through the NCBI database in the analyses which included samples from the northeast, Europe, and California (supplementary table S3). Sequences for each genome segment (chromosome, lp54, and cp26) were aligned using MAFFT v. 7.222 (Katoh et al. 2002) implemented in Geneious vs. 10.0.2 (www.geneious.com, Kearse et al. 2012).

Population genetics

Population genetic parameters for the chromosome, cp26, and lp54 were calculated in DnaSP v5.10.01 (Librado and Rozas 2009) (Table 3.1). We determined the extent of linkage disequilibrium within *B. burgdorferi* populations by calculating the Index of association (I_A) as described by Maynard-Smith et al. (1993). The I_A is defined as:

$$I_A = (V_o/V_e) - 1$$

where V_o is the observed variance of differences between loci (pairwise distance) and V_e is the expected variance of differences between loci given free recombination (Maynard Smith et al. 1993, Agapow et al. 2001). Because I_A is sensitive to the number of loci which makes the index difficult to interpret when comparing between populations and genetic segments, we also calculated $\overline{r}_d$ which converts I_A into a correlation coefficient that varies from 0 (no linkage disequilibrium) to 1 (complete linkage disequilibrium (Agapow et al. 2001). We calculated $\overline{r}_d$ with and without clone correction (reducing the dataset to include one of each haplotype per population) and the significance of $\overline{r}_d$ was evaluated by resampling the data to produce a null distribution. Both I_A and $\overline{r}_d$ were calculated in the R package, POPPR (R Core Team 2016, Kamvar et al. 2014).

Phylogenetic analyses

The general time reversible model of sequence evolution with gamma distributed rates (GTR+G) was determined to be the best-fit model of nucleotide substitution for all three of our multiple sequence alignments (chromosome, lp54, and cp26) using the BIC implemented in jModelTest (Darriba et al. 2012). Maximum likelihood trees were inferred using PhyML v 3.0 (Guindon et al. 2010). We confirmed the topology of the tree through Bayesian phylogenetic reconstruction for each genome segment rooted with an outgroup (*Borrelia afzelii*, BioProject PRJNA224116) and implemented in MrBayes

v3.2.5; we ran the Monte Carlo Markov Chain for 10,000,000 generations with four heated chains, sampling every 10,000 generations with 2,500,000 generations burn-in (Ronquist et al. 2012). PhyML trees and the corresponding sequence alignments were used to infer the ratio of the rates of recombination/mutation (R/Θ), average length of recombined fragment (δ), and the location of recombination along the sequence alignment in ClonalFrameML and generate a clonal phylogeny with recombination corrected branch lengths (Didelot and Wilson 2015).

3.3 Results and Discussion

Population genetics

Among all 29 samples used in the chromosome analysis we found a total of 14,692 segregating sites across 798,363bp (Table 3.1). We found remarkably little genetic diversity within oMGs across the chromosome. If we only consider isolates from oMGs that are represented by more than one isolate (oMG-A, oMG-E, oMG-F, oMG-G, oMG-H, oMG-O, oMG-N, oMG-U), then the average number of pairwise differences within and between an oMG is 1,256bp +/- 1557bp and 4149bp +/- 78bp respectively. If we exclude oMGs which do not form a monophyletic clade (oMG-A, oMG-F, and oMG-U), then the average number of pairwise differences within an oMG falls to 277bp +/- 242bp while the average number of pairwise differences between oMGs remains relatively high at 4151bp +/- 86bp. We sequenced two oMG-O samples that were isolated from different patients and found that they contained only 8 pairwise differences across the entire chromosome. These data suggest that fine scale inferences of migratory history and demography within an oMG will require a full genome dataset.

Interestingly, several plasmid sequences recovered from field-collected ticks in PA shared haplotypes with sequences in the NCBI database despite comparisons of

large plasmid fragments. For instance, two of the northeastern tick-derived oMG-K cp26 sequences of 18,252bp in length shared a haplotype with New York isolate 297 (NCBI: CP002268). However, these same three isolates varied by 11bp (297_K : EL19_K), 38bp (EL47_K : EL19_K), and 45bp (297_K : EL47_K) in pairwise comparisons across plasmid lp54.

Both the I_A and $\overline{rd}$ indicate that loci are nonrandomly associated (Table 3.1). As haplotype diversity was low across lp54 and cp26 in North Carolina and Virginia, we assessed the association among loci with clone-corrected data which revealed a similar pattern of significant linkage disequilibrium among loci with the exception of the cp26 plasmid in Virginia which had only two haplotypes among 10 samples and was therefore inadequate for calculating variance of pairwise differences. Low neutral genetic diversity and high linkage disequilibrium among loci suggests a low effective population size in *B. burgdorferi* (Dykhuizen et al. 1993).

Phylogenetic trees

We reconstructed three phylogenetic trees using ClonalframeML recombination corrections: a phylogenetic tree based on 798,363bp of the chromosomal sequence from 29 *B. burgdorferi* isolates, a phylogenetic tree based on 38,581bp of the lp54 plasmid sequence from 56 *B. burgdorferi* isolates, and a phylogenetic tree based on 18,252bp of the cp26 plasmid sequence from 62 *B. burgdorferi* isolates (Figure 1A-C). Our phylogenetic trees revealed a pattern of clustering by oMG rather than geographic origin with only a few exceptions indicating recombination. Notably, in Wisconsin we found that an *ospC* locus from an oMG-U appears to have recombined onto the genetic background of an oMG-F which is reflected in cp26, and the chromosomal phylogenetic trees, but not the lp54 tree. The limited genetic diversity across the chromosome between the oMG-U and oMG-F isolates suggests a recent recombination event at the

ospC locus in which an oMG-U recombined onto an oMG-F genetic background. The most internal (ancestral) nodes across all trees had low support in the starting maximum likelihood tree and at the base of the rooted Bayesian tree was a polychotomy.

Clustering at the external nodes within an oMG type was generally supported across trees reconstructed from each genetic segment, though there were some exceptions. In the chromosomal phylogeny, we found that an oMG-B isolate and a European-derived oMG-B2 isolate disrupt the oMG-A clade which is otherwise comprised of samples from the Midwest and Northeast. This pattern was also resolved with high posterior probability in the Bayesian tree. In the cp26 and lp54 trees, oMG-A isolates form a monophyletic clade with internal clustering by geographic region. Interestingly, the oMG-L isolate from North Carolina consistently clusters with an oMG-L isolate from France in both the cp26 and lp54 trees while the oMG-L isolate from Wisconsin shows no consistent relationships among trees, lending support to the hypothesis of trans-Atlantic dispersal of the southern oMG-L (Rudenko et al. 2013). We found that an oMG-V isolate from the Midwest did not cluster with an oMG-V isolate from North Carolina in either the lp54 or the cp26 trees. This pattern was repeated in the oMG-N isolates from the Midwest and North Carolina in the cp26 tree (no data for oMG-N in NC for lp54), where oMG-N from Wisconsin consistently clusters with the oMG-E group (across all trees) and does not cluster with oMG-N from North Carolina. These data show differences in linkage patterns between geographic regions.

We estimated the posterior mean and posterior variance for the ratio of rates of recombination to mutation (R/Θ) and the mean length of imported DNA (δ) for the chromosome, lp54, and cp26 (Table 3.2). We found a relatively higher rate of recombination detected at the *ospC* locus (~11kb in Figure 3.1) in cp26 which likely accounts for relatively high rate of recombination to mutation when compared to the

chromosome and lp54. Across lp54 we found a higher rate of recombination corresponding to the loci *decorin binding protein A* and *decorin binding protein B* which are host-associated virulence factors (Shi et al. 2008).

We found that mutation was identified more frequently than recombination (R/Θ= 0.5526, Table 3.2) using the ClonalFrameML method. Interestingly, the recombination graph shows many mutations throughout the chromosome for which the ClonalFrameML model assigned a moderate to high chance of homoplasy. This may reflect shortcomings of the model when applied to intra-population datasets. The underlying principle is that recombination will introduce more mutations than observed in the clonal genealogy (where mutation occurs but recombination does not). Detection of recombination from an inter-population donor is generally more robust than detection of recombination from an intra-population donor because the imported sequence should be more divergent resulting a region with more than the expected number of mutations given a constant mutation rate. Though the ClonalFrameML model performed reasonably well when tested on a simulated intra-population dataset (Didelot and Wilson 2015), the *B. burgdorferi* chromosome is highly conserved and is likely under purifying selection (Chaconas & Kobryn 2010, Haven et al. 2011), therefore, recombination in this case could lead to purging of mutations rather than the accumulation of mutations which would confound the accurate estimation of recombination.

3.4 Conclusions

In this study, we used whole chromosomes as well as partial plasmid sequences to investigate the extent of recombination in the Lyme borreliosis pathogen, *B. burgdorferi*, between the Midwest, Northeast, and Virginia. We concur with previous findings that *B. burgdorferi* is neither completely clonal, though there is significant

linkage disequilibrium among loci, nor highly recombinogenic given that the ratio of recombination to mutation is relatively low across the chromosome and lp54. We identified host-associated loci with evidence of elevated levels of recombination on lp54 at the *dbpA* and *dbpB* loci and on cp26 at *ospC*. Several groups appear to be in linked by oMG depending on which populations are sampled and which genomic sequences are examined which likely explains some of the discordance between studies. For example, if we had only sampled *B. burgdorferi* from the Northeast and Midwest, then we would have concluded complete linkage among oMG-G isolates.

Most notably, we identified strikingly little variation between individuals within a clade, though the clade identity cannot always be predicted by oMG. For example, between two oMG-G isolates from Wisconsin we found only one mutation per 100kb across the chromosome. Some clades were in perfect alignment with the oMG, and within an oMG there appeared to be sub-structure correlating to geographic region. For example, oMG-H and oMG-D form a monophyletic clade with subtrees correlating to geographic regions across all three phylogenetic trees and a similar pattern is revealed among oMG-A isolates in the lp54 and cp26 trees. These data suggest some isolation between geographic regions. We identified stronger association between loci in North Carolina and Virginia than in the Northeast or Midwest. The apparent clonality of the North Carolina and Virginia populations may reflect the recent expansion and lower effective population size in the region (Chapter 2) whereas the Northeast and Midwest populations of *B. burgdorferi* are relatively well-established populations.

Population structure in *B. burgdorferi* appears to be driven by a multitude of factors, including geographic and ecological isolation. It is clear that *B. burgdorferi* has the necessary machinery to undergo recombination and it is likely that different oMGs have the opportunity to exchange genetic material since both ticks and vertebrate hosts

are frequently found to be infected by multiple oMGs (Brisson and Dykhuizen 2004). Therefore, the observation of strong linkage disequilibrium and limited neutral genetic diversity likely reflect a small effective population size. We found the lowest genetic diversity in *B. burgdorferi* isolated from ticks collected in southern Virginia. Taken together with the recent studies suggesting that both vector species are experiencing a range expansion toward this region, the low genetic diversity found in the *B. burgdorferi* populations suggests a founder effect.

Acknowledgements

We thank Marcee Toliver and Dr. Holly Gaff for assistance in collecting infected ticks in North Carolina and the Hoffler Creek Wildlife Preserve in VA, respectively. Collecting permits for the North Carolina State Park System were granted by the North Carolina Department of Environment and Natural Resources (R14-23). Permissions for tick collection in Pennsylvania at Crow's Nest Preserve and Wawa Preserve were granted by the Natural Lands Trust, and we thank Dan Barringer for helpful tips. We thank Dr. Solny Adalsteinsson and the Forest Fragments in Managed Ecosystems program for collection assistance in Delaware. We thank Dr. Kurt Reed for access to the Wisconsin *B. burgdorferi* cultures. Sequencing and DNA library preparations were completed in the Hahn Lab Core facility at the University of Pennsylvania. The authors would like to thank Mitra Eghbal, Alex Berry and Jill Devine for helpful comments on an early draft of this manuscript. Stephanie Seifert was supported by NSF GRFP DGE-1321851.

3.5 Figures and tables

Table 3.1. Origin, sample sizes, sequence diversity, and linkage across the *Borrelia burgdorferi* chromosome and plasmid sequences

Region	Genome Segment	n	S	h	Hd	π	I_A	$\overline{r}d$	$\overline{r}dcc$
Northeast	chromosome	4	7739	4	1	0.0052 (0.0010)	131.5*	0.017*	0.017*
	lp54	18	601	16	0.987	0.0053 (0.0004)	52.6*	0.090*	0.088*
	cp26	19	596	12	0.947	0.0116 (0.0005)	48.5*	0.083*	0.052*
Wisconsin	chromosome	24	13262	24	1	0.0048 (0.0002)	436.9*	0.017*	0.017*
	lp54	24	751	24	1	0.0058 (0.0003)	38.9*	0.054*	0.054*
	cp26	24	664	22	0.993	0.0118 (0.0003)	22.4*	0.035*	0.034*
Virginia	chromosome	NA	NA	NA	NA	NA	NA	NA	NA
	lp54	10	255	3	0.378	0.0024 (0.0011)	253.2*	0.997*	0.988*
	cp26	10	226	2	0.356	0.0044 (0.0020)	225*	1.000*	NA
North Carolina	chromosome	NA	NA	NA	NA	NA	NA	NA	NA
	lp54	2	274	2	1	0.0071 (0.0036)	NA	NA	NA
	cp26	6	346	3	0.6	0.0082 (0.0028)	184.5*	0.535*	0.034*
ALL	chromosome	29	14692	29	1	0.0049 (0.0001)	408.4*	0.029*	0.029*
	lp54	56	979	47	0.981	0.0058 (0.0002)	38.7*	0.044*	0.042*
	cp26	62	751	42	0.976	0.0115 (0.0003)	33.1*	0.047*	0.032*

n = sample size, S = segregating sites, h = number of haplotypes, Hd = haplotype diversity, π = nucleotide diversity, standard deviation in parentheses, IA = Index of association, $\overline{r}d$ = standardized I_A, $\overline{r}dcc$ = standardized I_A on clone-corrected data
* indicates P < .001

Table 3.2. Recombination statistics for each *B. burgdorferi* genome segment as estimated in ClonalFrameML

Genome Segment	*R/Θ*	*δ*
chromosome	0.5526 (4.6e-05)	319.8
lp54	0.3619 (2.3e-4)	26.0
cp26	1.5124 (4.4e-3)	123.6

R/Θ = Ratio of recombination to mutation with posterior variance in parentheses, *δ* = the mean length of imported DNA fragments

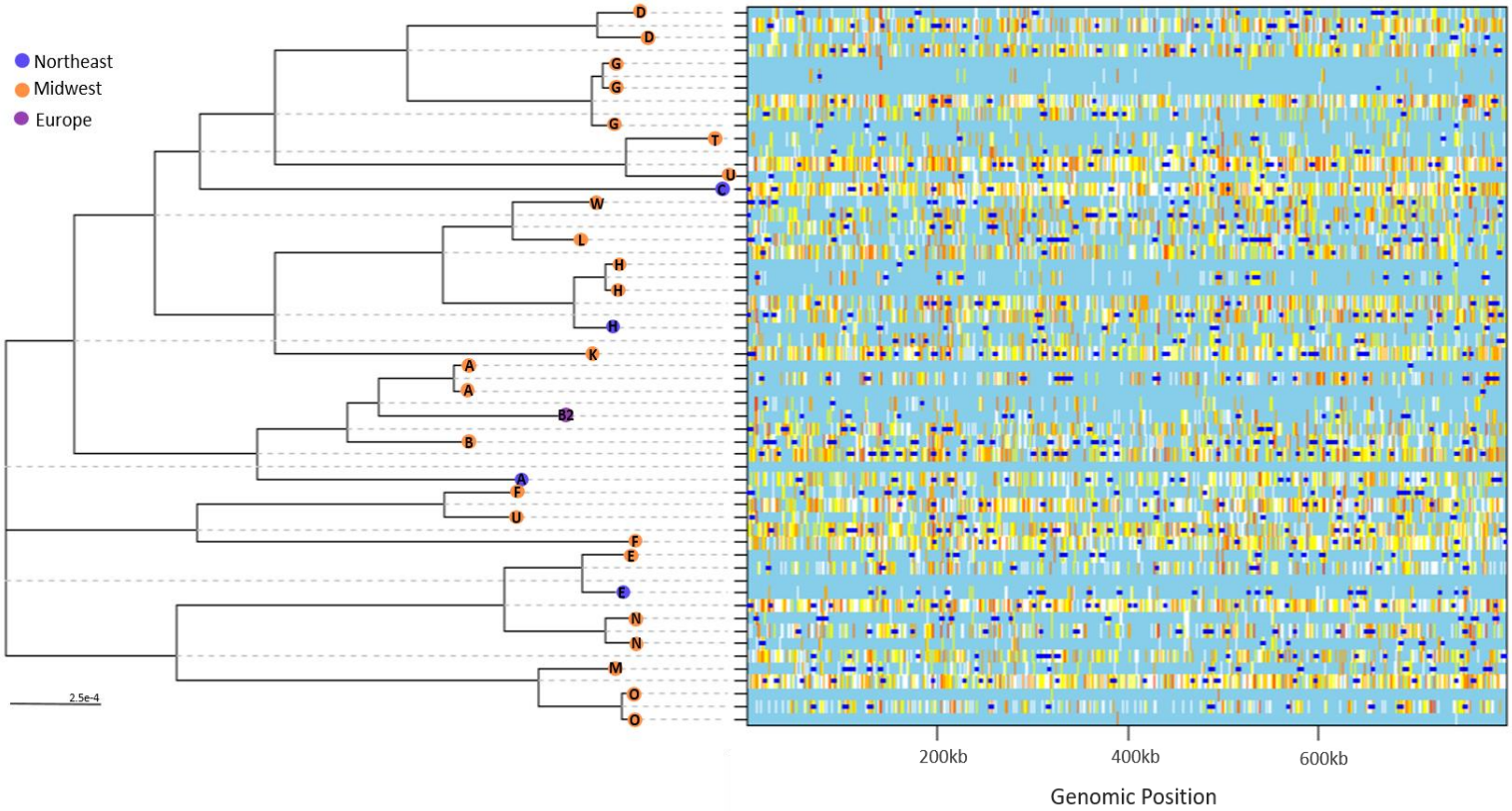


Figure 3.1-A

Maximum likelihood tree for *B. burgdorferi* chromosome sequences with graphical illustration of inferred recombination. The tips of the phylogeny are color coded by geographic origin for the isolate, the letter represents the oMG. Recombination events are highlight by dark blue horizontal lines in the graph, vertical bars show substitutions with yellow to red indicating increase in homoplasy and white indicating a non-homoplasic substitution. Light blue indicates no substitution.

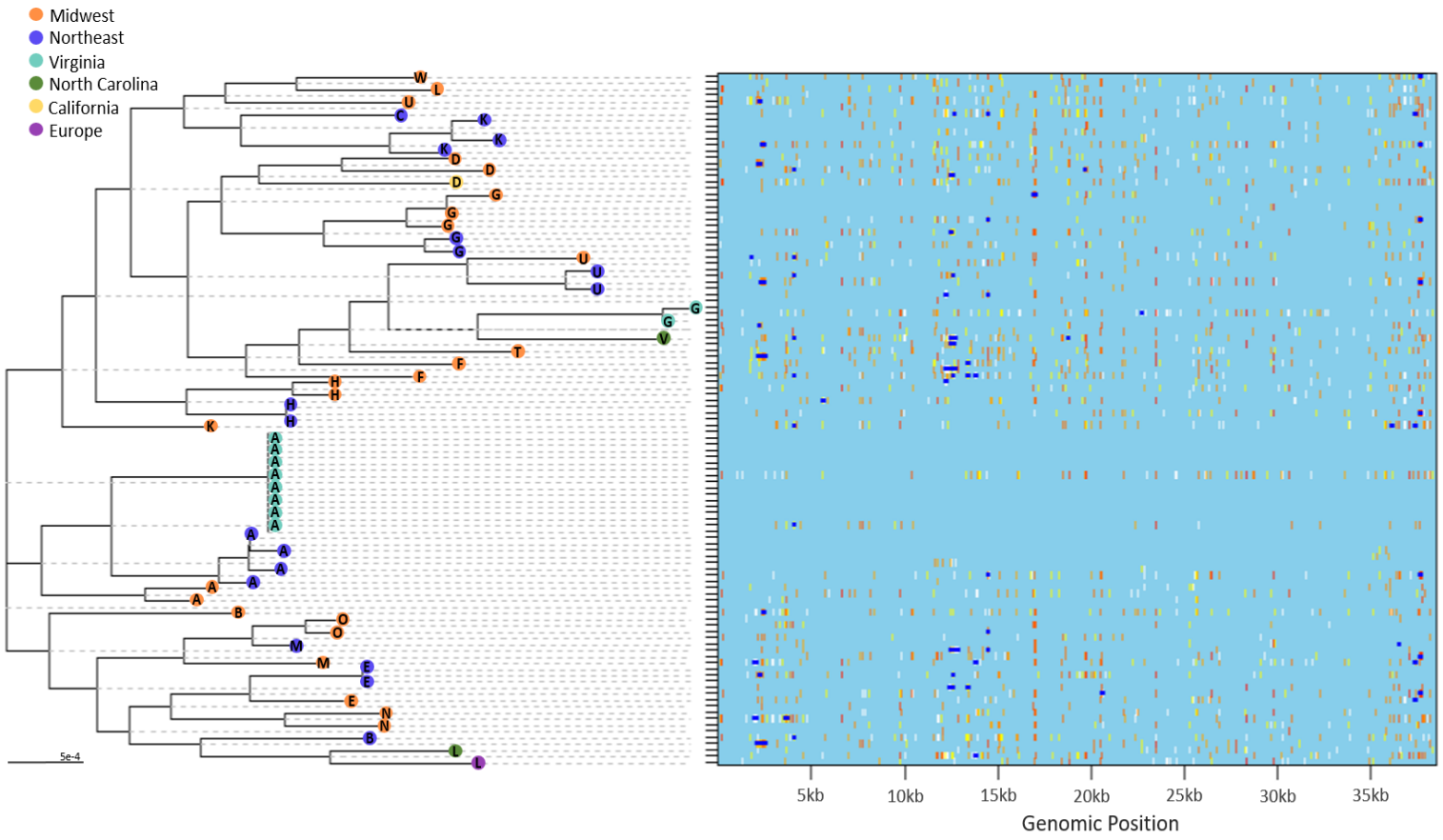


Figure 3.1-B

Maximum likelihood tree for *B. burgdorferi* lp54 sequences with graphical illustration of inferred recombination. The tips of the phylogeny are color coded by geographic origin for the isolate, the letter represents the oMG. Recombination events are highlight by dark blue horizontal lines in the graph, vertical bars show substitutions with yellow to red indicating increase in homoplasy and white indicating a non-homoplasic substitution. Light blue indicates no substitution.

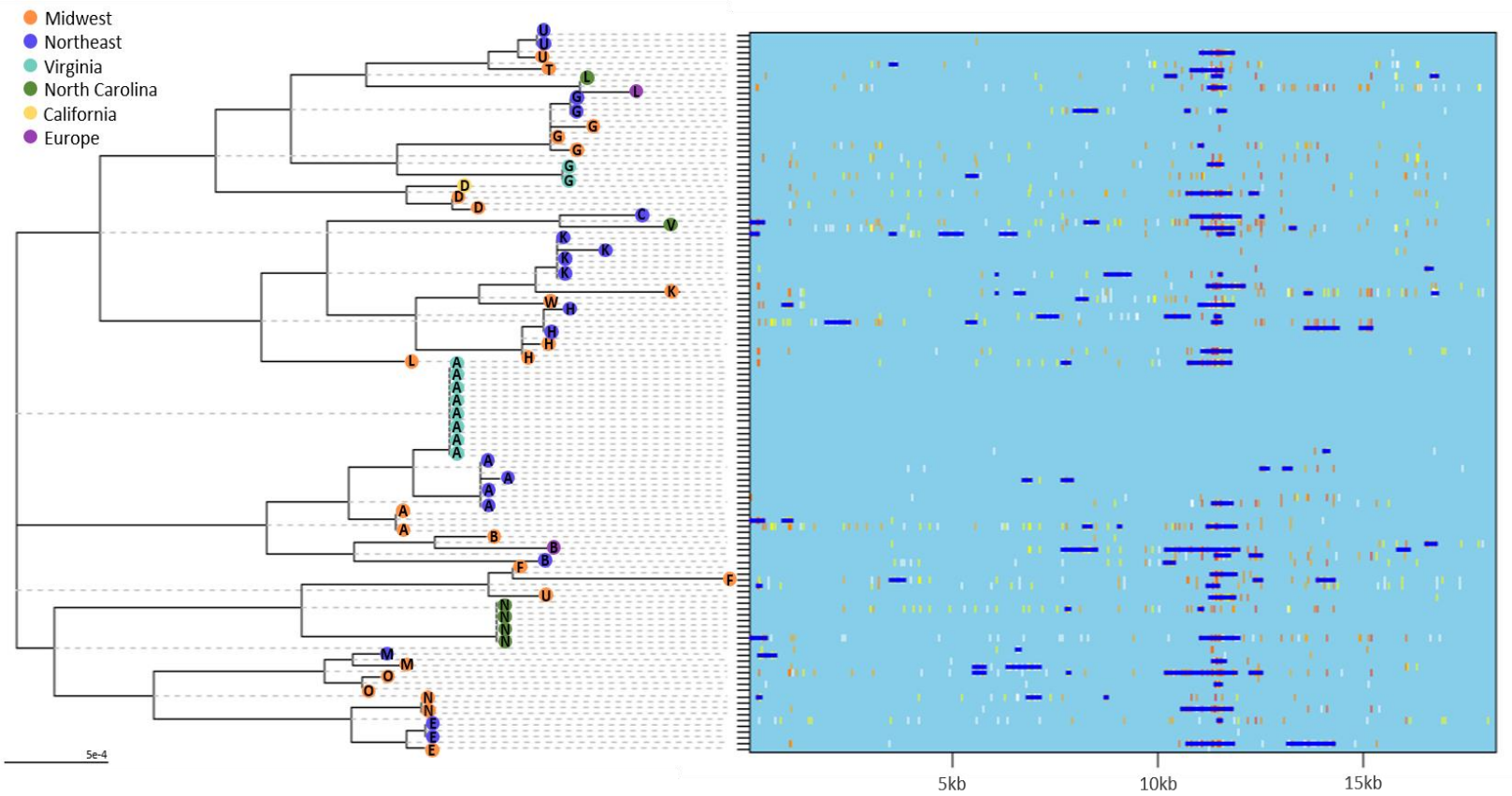


Figure 3.1-C

Maximum likelihood tree for *B. burgdorferi* cp26 sequences with graphical illustration of inferred recombination. The tips of the phylogeny are color coded by geographic origin for the isolate, the letter represents the oMG. Recombination events are highlight by dark blue horizontal lines in the graph, vertical bars show substitutions with yellow to red indicating increase in homoplasy and white indicating a non-homoplasic substitution. Light blue indicates no substitution.

CHAPTER 4: DISCUSSION

Though miniscule, microbes have a disproportionately large impact on the earth's ecosystems; they contribute to nutrient cycling in soils and in the ocean, aid in digestion, and some are pathogenic. The extent microbial dispersal abilities has been debated for nearly a century (Finlay and Clark 1999), however, the advent of DNA sequencing technologies and an increasing number of population genetic studies has revealed that population structure is evident in many bacterial communities (O'Malley 2008). Host-associated microbes, in particular, are constrained by the same biogeographic barriers as their host species (Martiny et al. 2006).

In pathogenic microbes, population structure arising as a result of limited gene flow has implications for the spread of genes and beneficial alleles by recombination between populations. Recombination in bacteria can increase genetic diversity by introducing multiple mutations in a single recombination event (Didelot and Maiden 2010), accelerate natural selection by efficiently combining beneficial mutations from different genetic backgrounds (Wiedenbeck & Cohan 2011), and disrupting Muller's ratchet by allowing deleterious mutations to be purged from the genome (Vos 2009).

In this thesis, we studied the population structure of the Lyme borreliosis pathogen, *B. burgdorferi*, and found that there is population structure between geographic regions and within geographic regions associated with major allele groups of the host associated *ospC* locus (oMGs). We report that at least two, apparently clonal, oMGs are associated with both competent vector species in southern Virginia, suggesting that tick species is not a barrier to dispersal for these oMGs. We identified many oMGs in Virginia and North Carolina which were not identified in the northeast. This may indicate that some oMGs are associated with hosts that do not travel long

distances (e.g. hispid rats rather than birds) or that the oMGs arrive in the northeast either exist in low frequencies so that we did not detect them or that they fail to establish.

Our population genetic and phylogeographic data support that *B. burgdorferi* is structured by oMGs and within oMGs by geographic region, though this is primarily supported by data from oMG-A and oMG-G for which we have data from three separate regions. We found evidence of recombination disrupting linkage with the ospC locus which indicates that clade identity is not always predicted by oMG. Finally, we report finding a low effective population size, as evidenced by strong linkage disequilibrium and low neutral genetic diversity, in all populations studied.

There are several models of microbial evolution which could give rise to small effective population sizes including differentiation of populations by ecotype (e.g. host or vector) followed by periodic selection, metapopulation dynamics in which isolated patches (e.g. host or vector) are randomly colonized by a subpopulation with occasional patch turnover, or genetic drift alone (Fraser et al. 2009). Our dataset is too coarse to facilitate differentiation between these models, but we can postulate that different host species can serve as an ecotype or as an isolated patch. Brisson and Dykhuizen (2004) found that *B. burgdorferi* oMGs were nonrandomly distributed across host species which lends some support to the model of period selection rather than simple genetic drift or metapopulations. However, these models are not mutually exclusive. Even with periodic selection driven by host-associated multiple niche polymorphism, several oMGs are capable of infecting each host species (Brisson and Dykhuizen 2004, Vuong et al. 2014) which would allow for both metapopulation dynamics (among oMGs capable of infecting a particular host species) and patch turnover (driven by strain specific immunity). Future studies using computational models could potentially tease apart the forces driving small effective population sizes in *B. burgdorferi*.

Supplementary Material

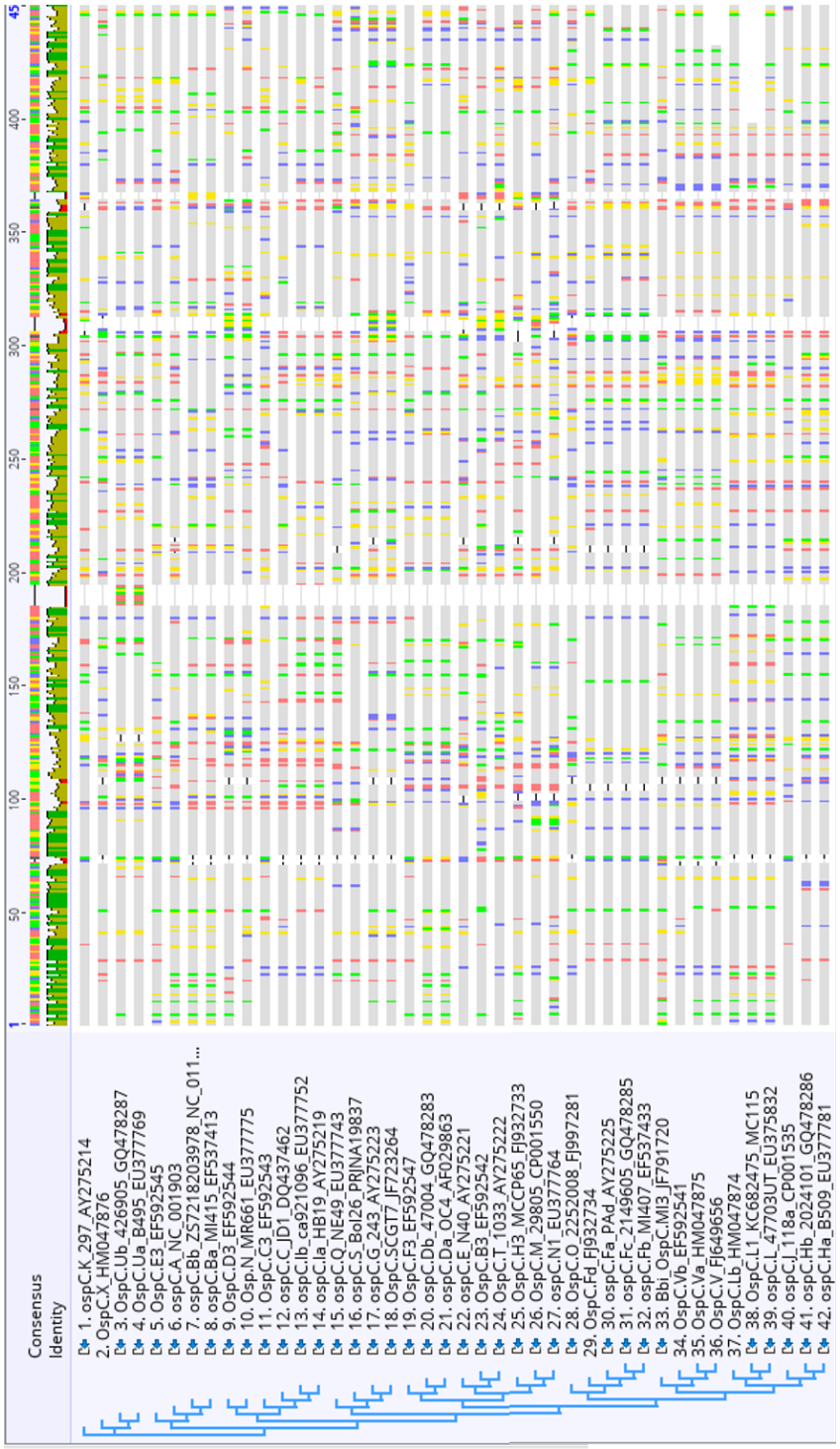
Supplementary Table S1A: Long-Range PCR primers and positions on genome segment.

Genome segment (region)	name	sequence	position	round
plasmid lp54 (2543-17263)	lp54-1-1F	5-TCGGCAATCATATCACTCCAAGTTCCA-3	2202-2228	1
	lp54-1-2F	5-ACTCCCCAAACAAATCTCTGGCTTTAA-3	2543-2571	2
	lp54-1-BR	5-CAAATGCAGCTCTTGAATCGTCCTCTA-3	17237-17263	1+2
plasmid lp54 (17095-27169)	lp54-2-BF	5-TGGCTTGTACTTTTGCTTCTCTTAAGGCT-3	17095-17124	1+2
	lp54-2-1R	5-GCCGGTTCTTGAAAAATCGTCAAAGTCT-3	27835-27864	1
	lp54-2-2R	5-CGGGTCAATCTCTTCAAGGTTGTAAGTT-3	27141-27169	2
plasmid lp54 (23309-34274)	lp54-3-1F	5-ACGAGATTACAGTAGCAATGCCCAGTATTA-3	23309-23338	1
		5-ACGAGATTACAACAGCAATGCCCAGTATTA-3		
	lp54-3-2F	5-GTTAAAGCAAAAATTGGCGCTGATACCAA-3	26450-26478	2
	lp54-3-BR	5-GTCTAATCCAGGCTCAAAGGCAGTGTT-3	34248-34274	1+2
plasmid lp54 (32425-50450)	lp54-4-2F	5-CCATGTACGATTGGGACAGCGTT-3	32425-32447	1
	lp54-4-1F	5-GTATGACTACATTAATGAGCGCTGGAGTTA-3	32972-33001	2
	lp54-4-BR	5-GGCGAATATGACAAAAATACGGGAGGAT-3	50423-50450	1+2
plasmid cp26 (158-15868)	cp26-1-1F	5-CGATGGAAGGGTAGAAATTGTAGCTTTCTT-3	158-190	1
		5-TTCATTTCCATGTTTTTRGGCGCAAATGCAAGAT-3	1251-1284	2
	cp26-1-BR	5-CAAGACTTCATGACTGCAGAATGGACT-3	15842-15868	1+2
plasmid cp26 (16509-26755)	cp26-2-1F	5-CAGAGAGCGTTCTTAATGATAGTTCTATTCCTACAA-3	13509-13544	1
	cp26-2-2F	5-GCTATTGCCATTTGGCTTTCTGTAACA-3	15516-15540	2
	cp26-2-BR	5-GTTAGTCCAACTCCTACGTTTAAAAGGTGTGT-3	26724-26755	1+2

* primer lp54-3-1F contains a two-nucleotide inversion, we mixed these two primers for each reaction

Supplementary Table S1B: Thermal Cycling conditions for LR-PCR

Touch Down PCR thermal cycling conditions for LR-PCR
Step 01: 98°C 2min
Step 02: 98°C 10sec
Step 03: 72°C* 15sec -1°C/cycle
Step 04: 72°C 15min
Step 05: go to step 2, repeat 6X
Step 06: 98°C 10sec
Step 07: 65°C 15sec
Step 08: 72°C 15min
Step 09: go to step 6, repeat 24X
Step 10: 4°C hold



Supplementary Figure S2: Multiple sequence alignment of ospC locus sequences used as a reference for local BLAST to identify ospC sequences. Note oMG designations for some sequences have been altered to reflect calculated oMG identity. oMG group is followed by NCBI accession number in the sequence titles.

References

Achtman, M., & Wagner, M. (2008). Microbial diversity and the genetic nature of microbial species. *Nature Reviews Microbiology*, 6(6), 431-440.

Agapow, P. M., & Burt, A. (2001). Indices of multilocus linkage disequilibrium. *Molecular Ecology Notes*, 1(1-2), 101-102.

Anderson, M. J., & Willis, T. J. (2003). Canonical analysis of principal coordinates: a useful method of constrained ordination for ecology. *Ecology*, 84(2), 511-525.

Anguita, J., Hedrick, M. N., & Fikrig, E. (2003). Adaptation of *Borrelia burgdorferi* in the tick and the mammalian host. *FEMS Microbiology Reviews*, 27(4), 493-504.

Avise, J. C., Arnold, J., Ball, R. M., Bermingham, E., Lamb, T., Neigel, J. E., . . . Saunders, N. C. (1987). Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annual Review of Ecology and Systematics*, 18(1), 489-522.

Balmelli, T., & Piffaretti, J.-C. (1996). Analysis of the genetic polymorphism of *Borrelia burgdorferi* sensu lato by multilocus enzyme electrophoresis. International *Journal of Systematic and Evolutionary Microbiology*, 46(1), 167-172.

Baranton, G., Seinost, G., Theodore, G., Postic, D., & Dykhuizen, D. (2001). Distinct levels of genetic diversity of *Borrelia burgdorferi* are associated with different aspects of pathogenicity. *Research in Microbiology*, 152(2), 149-156.

Barbour, A. G., & Hayes, S. F. (1986). Biology of Borrelia species. *Microbiological Reviews*, 50(4), 381.

Barbour, A. G., & Travinsky, B. (2010). Evolution and distribution of the ospC gene, a transferable serotype determinant of *Borrelia burgdorferi*. *MBio*, 1(4), e00153-00110.

Becker, N. S., Margos, G., Blum, H., Krebs, S., Graf, A., Lane, R. S., . . . Fingerle, V. (2016). Recurrent evolution of host and vector association in bacteria of the *Borrelia burgdorferi* sensu lato species complex. *BMC Genomics*, 17(1), 734.

Benavides, J. A., Cross, P. C., Luikart, G., & Creel, S. (2014). Limitations to estimating bacterial cross-species transmission using genetic and genomic markers: inferences from simulation modeling. *Evolutionary Applications*, 7(7), 774-787.

Brinkerhoff, R. J., Gilliam, W. F., & Gaines, D. (2014). Lyme disease, Virginia, USA, 2000–2011. *Emerging Infectious Diseases*, 20(10), 1661.

Brisson, D. (2010). Evolution of Northeastern and Midwestern *Borrelia burgdorferi*, United States-Volume 16, Number 6—June 2010-*Emerging Infectious Disease Journal*-CDC.

Brisson, D., & Dykhuizen, D. E. (2004). ospC diversity in *Borrelia burgdorferi*. Genetics, 168(2), 713-722.

Brooks, C. S., Hefty, P. S., Jolliff, S. E., & Akins, D. R. (2003). Global analysis of *Borrelia burgdorferi* genes regulated by mammalian host-specific signals. *Infection and Immunity*, 71(6), 3371-3383.

Brooks, D. R., León-Règagnon, V., McLennan, D. A., & Zelmer, D. (2006). Ecological fitting as a determinant of the community structure of platyhelminth parasites of anurans. *Ecology*, 87(sp7).

Bunikis, J., Garpmo, U., Tsao, J., Berglund, J., Fish, D., & Barbour, A. G. (2004). Sequence typing reveals extensive strain diversity of the Lyme borreliosis agents *Borrelia burgdorferi* in North America and Borrelia afzelii in Europe. *Microbiology*, 150(6), 1741-1755.

Byram, R., Stewart, P. E., & Rosa, P. (2004). The essential nature of the ubiquitous 26-kilobase circular replicon of *Borrelia burgdorferi*. *Journal of Bacteriology*, 186(11), 3561-3569.

Casjens, S. (2000). Borrelia genomes in the year 2000. *Journal of Molecular Microbiology and Biotechnology*, 2(4), 401-410.

Casjens, S. R., Mongodin, E. F., Qiu, W.-G., Luft, B. J., Schutzer, S. E., Gilcrease, E. B., . . . Vargas, L. C. (2012). Genome stability of Lyme disease spirochetes: comparative genomics of *Borrelia burgdorferi* plasmids. *PloS One*, 7(3), e33280.

Chaconas, G., & Kobryn, K. (2010). Structure, function, and evolution of linear replicons in Borrelia. *Annual Review of Microbiology*, 64, 185-202.

Chen, I.-C., Hill, J. K., Ohlemüller, R., Roy, D. B., & Thomas, C. D. (2011). Rapid range shifts of species associated with high levels of climate warming. *Science*, 333(6045), 1024-1026.

Clark, K. (2004). Borrelia species in host-seeking ticks and small mammals in northern Florida. *Journal of Clinical Microbiology*, 42(11), 5076-5086.

Comas, I., Coscolla, M., Luo, T., Borrell, S., Holt, K. E., Kato-Maeda, M., . . . Thwaites, G. (2013). Out-of-Africa migration and Neolithic coexpansion of Mycobacterium tuberculosis with modern humans. *Nature Genetics*, 45(10), 1176-1182.

Cooke, G. M., Chao, N. L., & Beheregaray, L. B. (2012). Divergent natural selection with gene flow along major environmental gradients in Amazonia: insights from genome scans, population genetics and phylogeography of the characin fish Triportheus albus. *Molecular Ecology*, 21(10), 2410-2427.

Courtney, J. W., Dryden, R. L., Montgomery, J., Schneider, B. S., Smith, G., & Massung, R. F. (2003). Molecular characterization of Anaplasma phagocytophilum and

Borrelia burgdorferi in *Ixodes scapularis* ticks from Pennsylvania. *Journal of Clinical Microbiology*, 41(4), 1569-1573.

Darriba, D., Taboada, G. L., Doallo, R., & Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods*, 9(8), 772-772.

De Silva, A. M., & Fikrig, E. (1995). Growth and migration of *Borrelia burgdorferi* in Ixodes ticks during blood feeding. *The American Journal of Tropical Medicine and Hygiene*, 53(4), 397-404.

Didelot, X., & Maiden, M. C. (2010). Impact of recombination on bacterial evolution. *Trends in Microbiology*, 18(7), 315-322.

Didelot, X., & Wilson, D. J. (2015). ClonalFrameML: efficient inference of recombination in whole bacterial genomes. *PLoS Computational Biology*, 11(2), e1004041.

Dobrindt, U., Hochhut, B., Hentschel, U., & Hacker, J. (2004). Genomic islands in pathogenic and environmental microorganisms. *Nature Reviews Microbiology*, 2(5), 414-424.

Dolan, M. C., Maupin, G. O., Panella, N. A., Golde, W. T., & Piesman, J. (1997). Vector competence of *Ixodes scapularis, I. spinipalpis*, and *Dermacentor andersoni* (Acari: Ixodidae) in transmitting *Borrelia burgdorferi*, the etiologic agent of Lyme disease. *Journal of Medical Entomology*, 34(2), 128-135.

Dykhuizen, D. E., & Baranton, G. (2001). The implications of a low rate of horizontal transfer in Borrelia. *Trends in Microbiology*, 9(7), 344-350.

Dykhuizen, D. E., Brisson, D., Sandigursky, S., Wormser, G. P., Nowakowski, J., Nadelman, R. B., & Schwartz, I. (2008). The propensity of different *Borrelia burgdorferi* sensu stricto genotypes to cause disseminated infections in humans. *The American Journal of Tropical Medicine and Hygiene*, 78(5), 806-810.

Dykhuizen, D. E., Polin, D. S., Dunn, J. J., Wilske, B., Preac-Mursic, V., Dattwyler, R. J., & Luft, B. J. (1993). *Borrelia burgdorferi* is clonal: implications for taxonomy and vaccine development. *PNAS*, 90(21), 10163-10167.

Earn, D. J., Dushoff, J., & Levin, S. A. (2002). *Ecology and Evolution* of the flu. *Trends in Ecology & Evolution*, 17(7), 334-340.

Earnhart, C. G., Buckles, E. L., Dumler, J. S., & Marconi, R. T. (2005). Demonstration of OspC type diversity in invasive human Lyme disease isolates and identification of previously uncharacterized epitopes that define the specificity of the OspC murine antibody response. *Infection and Immunity*, 73(12), 7869-7877.

Eisen, L., & Lane, R. S. (2002). Vectors of *Borrelia burgdorferi* sensu lato. *Lyme borreliosis: Biology, Epidemiology and Control*, 91-115.

Finlay, B. J., & Clarke, K. J. (1999). Ubiquitous dispersal of microbial species. *Nature*, 400(6747), 828-828.

Fraser, C., Alm, E. J., Polz, M. F., Spratt, B. G., & Hanage, W. P. (2009). The Bacterial Species Challenge: Making Sense of Genetic and Ecological Diversity. *Science*, 323(5915), 741-746. doi:10.1126/*Science*.1159388

Fung, B. P., McHUGH, G. L., Leong, J. M., & Steere, A. C. (1994). Humoral immune response to outer surface protein C of *Borrelia burgdorferi* in Lyme disease: role of the immunoglobulin M response in the serodiagnosis of early infection. *Infection and Immunity*, 62(8), 3213-3221.

Gatewood, A. G., Liebman, K. A., Vourc'h, G., Bunikis, J., Hamer, S. A., Cortinas, R., . . . Tsao, J. (2009). Climate and tick seasonality are predictors of *Borrelia burgdorferi* genotype distribution. *Applied and Environmental Microbiology*, 75(8), 2476-2483.

Gilmore, R. D., & Piesman, J. (2000). Inhibition of *Borrelia burgdorferi* migration from the midgut to the salivary glands following feeding by ticks on OspC-immunized mice. *Infection and Immunity*, 68(1), 411-414.

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology*, 59(3), 307-321.

Guttman, D. S., & Dykhuizen, D. E. (1994). Detecting selective sweeps in naturally occurring Escherichia coli. *Genetics*, 138(4), 993-1003.

Gyles, C., & Boerlin, P. (2014). Horizontally transferred genetic elements and their role in pathogenesis of bacterial disease. *Veterinary Pathology*, 51(2), 328-340.

Haddad, N. M., Brudvig, L. A., Clobert, J., Davies, K. F., Gonzalez, A., Holt, R. D., . . . Collins, C. D. (2015). Habitat fragmentation and its lasting impact on Earth's ecosystems. *Science Advances*, 1(2), e1500052.

Hall, E. R., & Kelson, K. R. (1959). The mammals of North America. v. 1-2.

Harrison, B. A., Rayburn Jr, W. H., Toliver, M., Powell, E. E., Engber, B. R., Durden, L. A., . . . Whitt, P. B. (2010). Recent discovery of widespread *Ixodes affinis* (Acari: Ixodidae) distribution in North Carolina with implications for Lyme disease studies. *Journal of Vector Ecology*, 35(1), 174-179.

Haven, J., Vargas, L. C., Mongodin, E. F., Xue, V., Hernandez, Y., Pagan, P., . . . Casjens, S. R. (2011). Pervasive recombination and sympatric genome diversification driven by frequency-dependent selection in *Borrelia burgdorferi*, the Lyme disease bacterium. *Genetics*, 189(3), 951-966.

Holt, K. E., Parkhill, J., Mazzoni, C. J., Roumagnac, P., Weill, F.-X., Goodhead, I., . . . Wain, J. (2008). High-throughput sequencing provides insights into genome variation and evolution in Salmonella Typhi. *Nature Genetics*, 40(8), 987-993.

Humphrey, P. T., Caporale, D. A., & Brisson, D. (2010). Uncoordinated phylogeography of *Borrelia burgdorferi* and its tick vector, *Ixodes scapularis*. *Evolution*, 64(9), 2653-2663.

Ing-Nang Wang, D. E. D., Weigang Qiu, John J. Dunn, Edward M. Bosler, Benjamin Luft (1999). Genetic diversity of ospC in a local population of *Borrelia burgdorferi* sensu stricto. *Genetics*, 151(1), 15-30.

J Bunikis, U. G., J Tsao, J Berglund, D Fish, AG Barbour. (2004). Sequence typing reveals extensive strain diversity of the Lyme borreliosis agents *Borrelia burgdorferi* in North America and *Borrelia afzelii* in Europe. *Microbiology*, 150, 1741-1755. doi:10.1099/mic.0.26944-0

Jacquot, M., Gonnet, M., Ferquel, E., Abrial, D., Claude, A., Gasqui, P., . . . Faure, B. (2014). Comparative population genomics of the *Borrelia burgdorferi* species complex reveals high degree of genetic isolation among species and underscores benefits and constraints to studying intra-specific epidemiological processes. *PloS One*, 9(4), e94384.

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

Katoh, K., Misawa, K., Kuma, K. i., & Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, 30(14), 3059-3066.

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

Kelly, R. R., Gaines, D., Gilliam, W. F., & Brinkerhoff, R. J. (2014). Population genetic structure of the Lyme disease vector *Ixodes scapularis* at an apparent spatial expansion front. Infection, *Genetics and Evolution*, 27, 543-550.

Kim, Y., & Orr, H. A. (2005). Adaptation in sexuals vs. asexuals: clonal interference and the Fisher-Muller model. *Genetics*, 171(3), 1377-1386.

Krueger, F. (2015). Trim Galore. A wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files. (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/)

Lemey, P., Rambaut, A., Bedford, T., Faria, N., Bielejec, F., Baele, G., . . . Brockmann, D. (2014). Unifying viral genetics and human transportation data to predict the global transmission dynamics of human influenza H3N2. *PloS Pathogens*, 10(2), e1003932.

Librado, P., & Rozas, J. (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, 25(11), 1451-1452.

Linz, B., Balloux, F., Moodley, Y., Manica, A., Liu, H., Roumagnac, P., . . . van der Merwe, S. W. (2007). An African origin for the intimate association between humans and Helicobacter pylori. *Nature*, 445(7130), 915-918.

Linzey, J. L. W. a. A. V. (1977). *Peromyscus gossypinus*. *Mammalian Species,* 70, 1-5. doi:DOI: 10.2307/3503898

Liveris, D., Varde, S., Iyer, R., Koenig, S., Bittker, S., Cooper, D., . . . Wormser, G. P. (1999). Genetic diversity of *Borrelia burgdorferi* in Lyme disease patients as

determined by culture versus direct PCR with clinical specimens. *Journal of Clinical Microbiology*, 37(3), 565-569.

Livey, I., Gibbs, C., Schuster, R., & Dorner, F. (1995). Evidence for lateral transfer and recombination in OspC variation in Lyme disease Borrelia*. Molecular Microbiology*, 18(2), 257-269.

Maggi, R. G., Reichelt, S., Toliver, M., & Engber, B. (2010). Borrelia species in *Ixodes affinis* and *Ixodes scapularis* ticks collected from the coastal plain of North Carolina. *Ticks and Tick-borne Diseases*, 1(4), 168-171.

Margos, G., Gatewood, A. G., Aanensen, D. M., Hanincová, K., Terekhova, D., Vollmer, S. A., . . . Kurtenbach, K. (2008). MLST of housekeeping genes captures geographic population structure and suggests a European origin of *Borrelia burgdorferi*. *PNAS*, 105(25), 8730-8735. doi:10.1073/pnas.0800323105

Margos, G., Tsao, J. I., Castillo-Ramírez, S., Girard, Y. A., Hamer, S. A., Hoen, A. G., . . . Ogden, N. H. (2012). Two boundaries separate *Borrelia burgdorferi* populations in North America. *Applied and Environmental Microbiology*, 78(17), 6059-6067.

Martiny, J. B. H., Bohannan, B. J., Brown, J. H., Colwell, R. K., Fuhrman, J. A., Green, J. L., . . . Kuske, C. R. (2006). Microbial biogeography: putting microorganisms on the map. *Nature Reviews Microbiology*, 4(2), 102-112.

Mechai, S., Margos, G., Feil, E., Lindsay, L., & Ogden, N. (2015). Complex population structure of *Borrelia burgdorferi* in southeastern and south central Canada as revealed by phylogeographic analysis. *Applied and Environmental Microbiology*, 81(4), 1309-1318.

Mills, J. N., Gage, K. L., & Khan, A. S. (2010). Potential influence of climate change on vector-borne and zoonotic diseases: a review and proposed research plan. *Environmental Health Perspectives*, 118(11), 1507.

Nadolny, R., Gaff, H., Carlsson, J., & Gauthier, D. (2015). Comparative population genetics of two invading ticks: Evidence of the ecological mechanisms underlying tick range expansions. *Infection*, *Genetics and Evolution*, 35, 153-162.

Nadolny, R. M., Wright, C. L., Hynes, W. L., Sonenshine, D. E., & Gaff, H. D. (2011). *Ixodes affinis* (Acari: Ixodidae) in southeastern Virginia and implications for the spread of *Borrelia burgdorferi*, the agent of Lyme disease. *Journal of Vector Ecology*, 36(2), 464-467.

O'Malley, M. A. (2008). 'Everything is everywhere: but the environment selects': ubiquitous distribution and ecological determinism in microbial biogeography. *Studies in History and Philosophy of Biological and Biomedical Sciences*, 39(3), 314-325.

Ogden, N., Lindsay, L., Hanincova, K., Barker, I., Bigras-Poulin, M., Charron, D., . . . Schwartz, I. (2008). Role of migratory birds in introduction and range expansion of *Ixodes scapularis* ticks and of *Borrelia burgdorferi* and *Anaplasma phagocytophilum* in Canada. *Applied and Environmental Microbiology*, 74(6), 1780-1790.

Oksanen, J. (2016). Vegan: an introduction to ordination.

Oliver Jr, J. H. (1996). Lyme borreliosis in the southern United States: a review. *The Journal of Parasitology*, 926-935.

Page, A. J., Cummins, C. A., Hunt, M., Wong, V. K., Reuter, S., Holden, M. T., . . . Parkhill, J. (2015). Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics*, 31(22), 3691-3693.

Pal, U., De Silva, A. M., Montgomery, R. R., Fish, D., Anguita, J., Anderson, J. F., . . . Fikrig, E. (2000). Attachment of *Borrelia burgdorferi* within *Ixodes scapularis*

mediated by outer surface protein A. *The Journal of Clinical Investigation*, 106(4), 561-569.

Pal, U., Yang, X., Chen, M., Bockenstedt, L. K., Anderson, J. F., Flavell, R. A., . . . Fikrig, E. (2004). OspC facilitates *Borrelia burgdorferi* invasion of *Ixodes scapularis* salivary glands. *The Journal of Clinical Investigation*, 113(2), 220-230.

Piesman, J., & Gern, L. (2004). Lyme borreliosis in Europe and North America. *Parasitology*, 129(S1), S191-S220.

Piesman, J., Oliver, J. R., & Sinsky, R. J. (1990). Growth kinetics of the Lyme disease spirochete (*Borrelia burgdorferi*) in vector ticks (*Ixodes dammini*). *The American Journal of Tropical Medicine and Hygiene*, 42(4), 352-357.

Polz, M. F., Alm, E. J., & Hanage, W. P. (2013). Horizontal gene transfer and the evolution of bacterial and archaeal population structure. *Trends in Genetics*, 29(3), 170-175.

Pos, E., Guevara Andino, J. E., Sabatier, D., Molino, J. F., Pitman, N., Mogollón, H., . . . Di Fiore, A. (2014). Are all species necessary to reveal ecologically important patterns? *Ecology and Evolution*, 4(24), 4626-4636.

Posada, D. (2008). jModelTest: phylogenetic model averaging. *Molecular Biology and Evolution*, 25(7), 1253-1256.

Qiu, W.-G., Bruno, J. F., McCaig, W. D., Xu, Y., Livey, I., Schriefer, M. E., & Luft, B. J. (2008). Wide distribution of a high-virulence *Borrelia burgdorferi* clone in Europe and North America.

Qiu, W.-G., & Martin, C. L. (2014). Evolutionary genomics of *Borrelia burgdorferi* sensu lato: findings, hypotheses, and the rise of hybrids. Infection, *Genetics and Evolution*, 27, 576-593.

Qiu, W.-G., Schutzer, S. E., Bruno, J. F., Attie, O., Xu, Y., Dunn, J. J., . . . Luft, B. J. (2004). Genetic exchange and plasmid transfers in *Borrelia burgdorferi* sensu stricto revealed by three-way genome comparisons and multilocus sequence typing. *PNAS*, 101(39), 14150-14155.

Rambaut, A., Pybus, O. G., Nelson, M. I., Viboud, C., Taubenberger, J. K., & Holmes, E. C. (2008). The genomic and epidemiological dynamics of human influenza A virus. *Nature*, 453(7195), 615-619.

Rebecca R. Kelly, D. G., Will Gilliam, R. Jory Brinkerhoff. (2014). Population genetic structure of the Lyme disease vector *Ixodes scapularis* at an apparent expansion front. *Infection, Genetics, and Evolution*, 27, 543-550. doi:http://dx.doi.org/10.1016/j.meegid.2014.05.022

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S., . . . Huelsenbeck, J. P. (2012). MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. *Systematic Biology*, 61(3), 539-542. doi:10.1093/sysbio/sys029

Ronquist, F., Teslenko, M., Van Der Mark, P., Ayres, D. L., Darling, A., Höhna, S., . . . Huelsenbeck, J. P. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*, 61(3), 539-542.

Rozas, P. L. a. J. (2009). DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, 25, 1451-1452.

Rudenko, N., Golovchenko, M., Grubhoffer, L., & Oliver, J. H. (2013). The rare ospC allele L of *Borrelia burgdorferi* sensu stricto, commonly found among samples collected in a coastal plain area of the southeastern United States, is associated

with *Ixodes affinis* ticks and local rodent hosts *Peromyscus gossypinus* and *Sigmodon hispidus*. *Applied and Environmental Microbiology*, 79(4), 1403-1406.

Samuels, D. S. (2011). Gene regulation in *Borrelia burgdorferi*. *Annual Review of Microbiology*, 65, 479-499.

Schutzer, S. E., Fraser-Liggett, C. M., Casjens, S. R., Qiu, W.-G., Dunn, J. J., Mongodin, E. F., & Luft, B. J. (2011). Whole-genome sequences of thirteen isolates of *Borrelia burgdorferi*. *Journal of Bacteriology*, 193(4), 1018-1020.

Schwan, T. G., & Piesman, J. (2000). Temporal changes in outer surface proteins A and C of the Lyme disease-associated spirochete, *Borrelia burgdorferi*, during the chain of infection in ticks and mice. *Journal of Clinical Microbiology*, 38(1), 382-388.

Schwan, T. G., Piesman, J., Golde, W. T., Dolan, M. C., & Rosa, P. A. (1995). Induction of an outer surface protein on *Borrelia burgdorferi* during tick feeding. *PNAS*, 92(7), 2909-2913.

Seifert, S. N., Khatchikian, C. E., Zhou, W., & Brisson, D. (2015). Evolution and population genomics of the Lyme borreliosis pathogen, *Borrelia burgdorferi*. *Trends in Genetics*, 31(4), 201-207.

Seinost, G., Dykhuizen, D. E., Dattwyler, R. J., Golde, W. T., Dunn, J. J., Wang, N., . . . Luft, B. J. (1999). Four clones of *Borrelia burgdorferi*sensu stricto cause invasive infection in humans. *Infection and Immunity*, 67(7), 3518-3524.

Shapiro, B. J., Friedman, J., Cordero, O. X., Preheim, S. P., Timberlake, S. C., Szabó, G., . . . Alm, E. J. (2012). Population genomics of early events in the ecological differentiation of bacteria. *Science*, 336(6077), 48-51.

Shi, Y., Xu, Q., McShan, K., & Liang, F. T. (2008). Both decorin-binding proteins A and B are critical for the overall virulence of *Borrelia burgdorferi*. *Infection and Immunity*, 76(3), 1239-1246.

Slatkin, M. (1987). Gene flow and the geographic structure of natural populations. *Science*, 236, 787-793.

Smith, J. M. (1968). Evolution in Sexual and Asexual Populations. *The American Naturalist*, 102(927), 469-473.

Smith, J. M., & Smith, N. H. (1998). Detecting recombination from gene trees. *Molecular Biology and Evolution*, 15(5), 590-599.

Smith, J. M., Smith, N. H., O'Rourke, M., & Spratt, B. G. (1993). How clonal are bacteria? *PNAS*, 90(10), 4384-4388.

Smith, K. F., Sax, D. F., Gaines, S. D., Guernier, V., & Guégan, J.-F. (2007). Globalization of human infectious disease. *Ecology*, 88(8), 1903-1910.

Spielman, A., Andreadis, T., Apperson, C., Cornel, A., Day, J., Edman, J., . . . Lampman, R. (2004). Outbreak of West Nile virus in North America. *Science*, 306(5701), 1473-1475.

Spratt, B. G. (2004). Exploring the Concept of Clonality in Bacteria. In N. Woodford & A. P. Johnson (Eds.), *Genomics, Proteomics, and Clinical Bacteriology: Methods and Reviews* (pp. 323-352). Totowa, NJ: Humana Press.

Tilly, K., Krum, J. G., Bestor, A., Jewett, M. W., Grimm, D., Bueschel, D., . . . Stewart, P. (2006). *Borrelia burgdorferi* OspC protein required exclusively in a crucial early stage of mammalian infection. *Infection and Immunity*, 74(6), 3554-3564.

Travinsky, B., Bunikis, J., & Barbour, A. G. (2010). Geographic differences in genetic locus linkages for *Borrelia burgdorferi*. *Emerging Infectious Diseases*, 16(7).

Vos, M. (2009). Why do bacteria engage in homologous recombination? *Trends in Microbiology*, 17(6), 226-232.

Vuong, H. B., Canham, C. D., Fonseca, D. M., Brisson, D., Morin, P. J., Smouse, P. E., & Ostfeld, R. S. (2014). Occurrence and transmission efficiencies of *Borrelia burgdorferi* ospC types in avian and mammalian wildlife. Infection, *Genetics and Evolution*, 27, 594-600.

Walther, G.-R., Post, E., Convey, P., Menzel, A., Parmesan, C., Beebee, T. J., . . . Bairlein, F. (2002). Ecological responses to recent climate change. *Nature*, 416(6879), 389-395.

Wang, N., Dykhuizen, D. E., Qiu, W., Dunn, J. J., Bosler, E. M., & Luft, B. J. (1999). Genetic diversity of ospC in a local population of *Borrelia burgdorferi* sensu stricto. *Genetics*, 151(1), 15-30.

WG Qiu, D. D., MS Acosta, BJ Luft. (2002). Geographic uniformity of the Lyme disease spirochete (*Borrelia burgdorferi*) and its shared history with the tick vector (*Ixodes scapularis*) in the northeastern United States. *Genetics*, 160, 833-849.

Wiedenbeck, J., & Cohan, F. M. (2011). Origins of bacterial diversity through horizontal genetic transfer and adaptation to new ecological niches. *FEMS Microbiology Reviews*, 35(5), 957-976.

Wodecka, B. (2011). flaB gene as a molecular marker for distinct identification of Borrelia species in environmental samples by the PCR-restriction fragment length polymorphism method. *Applied and Environmental Microbiology*, 77(19), 7088-7092.

Wolfe, J. L., & Linzey, A. V. (1977). *Peromyscus gossypinus*. *Mammalian Species*. (70), 1-5.

Wright, C. L., Hynes, W. L., White, B. T., Marshall, M. N., Gaff, H. D., & Gauthier, D. T. (2014). Single-tube real-time PCR assay for differentiation of *Ixodes affinis* and *Ixodes scapularis*. *Ticks and Tick-borne Diseases*, 5(1), 48-52.

Wright, S. (1943). Isolation by distance. *Genetics*, 28(2), 114.

Yang, X. F., Pal, U., Alani, S. M., Fikrig, E., & Norgard, M. V. (2004). Essential role for OspA/B in the life cycle of the Lyme disease spirochete. *Journal of Experimental Medicine*, 199(5), 641-648.

Zückert, W. R. (2007). Laboratory maintenance of *Borrelia burgdorferi*. *Current Protocols in Microbiology*, 12C. 11.11-12C. 11.10.