

**GEOGRAPHIC DISTRIBUTIONS AND VERTEBRATE HOST SPECIFICITY
OF GENETIC LINEAGES OF *BORRELIA BURGDORFERI***

by

Scarlet Shifflett

A thesis submitted to the Faculty of the University of Delaware in partial fulfillment of the requirements for the degree of Master of Science in Wildlife Ecology

Spring 2023

© 2023 Scarlet Shifflett
All Rights Reserved

**GEOGRAPHIC DISTRIBUTIONS AND VERTEBRATE HOST SPECIFICITY
OF GENETIC LINEAGES OF *BORRELIA BURGENDORFERI***

by

Scarlet Shifflett

Approved: _____
Vincenzo Ellis, Ph.D.
Professor in charge of thesis on behalf of the Advisory Committee

Approved: _____
Jacob Bowman, Ph.D.
Chair of the Department of Entomology and Wildlife Ecology

Approved: _____
Calvin L. Keeler Jr., Ph.D.
Dean of the College of Agriculture and Natural Resources

Approved: _____
Louis F. Rossi, Ph.D.
Vice Provost for Graduate and Professional Education and
Dean of the Graduate College

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to my advisor, Dr. Vincenzo Ellis, for his invaluable insights and guidance throughout every stage of the project. I am also grateful to Dr. Jacob Bowman and Dr. Dustin Brisson for serving on my thesis committee and sharing their extensive knowledge to help better my research. Many thanks to all of the co-authors and collaborators, without their generosity I would not have been able to complete this research. I thank the Delaware Department of Natural Resources and Environmental Control for access to public lands for tick collection and Delaware Technical Community College and their Testing For Tickborne Pathogens grant program. Dr. Dina Fonseca's lab at Rutgers University graciously shared small mammal samples while also providing valuable suggestions to further the project.

I thank Brewster Kingham and Mark Shaw in the University of Delaware's DNA Sequencing and Genotyping Center for library preparation and Illumina sequencing. I also thank the Center for Bioinformatics and Computational Biology at the University of Delaware, and specifically Jaysheel Bhavsar, for help with data analysis. I performed bioinformatic analyses on the University of Delaware's BIOMIX cluster, which is supported by Delaware INBRE (NIH/NIGMS P20 GM103446). I thank the National Science Foundation for their financial support through their Graduate Research Fellowship Program. Many thanks to Tyler Casey, Lindsay Mann and Nicole Wagner for their help with lab work along with Ishika Sharma, Nya Wynn, Adam Rose, and Geovanny Peguero.

I am truly appreciative of my family's unwavering love and support that has allowed me to take every opportunity to further my own education.

TABLE OF CONTENTS

LIST OF TABLES	vi
LIST OF FIGURES	ix
ABSTRACT	xi

Chapter

1	PREVALENCE OF <i>BORRELIA BURGDORFERI</i> AND DIVERSITY OF ITS OUTER SURFACE PROTEIN C (<i>OSPC</i>) ALLELES IN BLACKLEGGED TICKS (<i>IXODES SCAPULARIS</i>) IN DELAWARE.....	1
2	A SPECIALIZED <i>BORRELIA BURGDORFERI</i> OUTER SURFACE PROTEIN C (<i>OSPC</i>) ALLELE MAY MITIGATE LYME DISEASE RISK IN HUMANS.....	24
	REFERENCES	47

Appendix

A	CHAPTER 1 SUPPLEMENTARY MATERIALS.....	57
B	CHAPTER 2 SUPPLEMENTARY MATERIALS.....	76
C	IACUC APPROVAL.....	82
D	PERMISSIONS	83

LIST OF TABLES

Table 1.1:	The number of nymphal blacklegged ticks by site (“Sample size”) and the number of those ticks infected with <i>B. burgdorferi</i> as determined by PCR (“PCR infections”) and qPCR (“qPCR infections”); prevalence of <i>B. burgdorferi</i> as a percentage is presented in parentheses following the number of infections.....	12
Table 1.2	<i>OspC</i> allele frequencies by site with the number of infected ticks sequenced (“sample size”) in parenthesis following the site name. Alleles associated with human disseminated infections are italicized (A, B, I, and K). The final two rows of the table show the total number of infections of each <i>ospC</i> allele (“Total alleles”) and their relative frequencies (i.e., percentage of the total number of alleles found). The final two columns show the total number of alleles (“Total alleles”) and the number of unique alleles (“Unique alleles”) found per site. Alleles are arranged in order of their relative frequencies.....	14
Table 2.1	Observed and expected host specificity of <i>ospC</i> alleles. Host specificity (“Observed”) was quantified as the number of host species an <i>ospC</i> allele infected. The number of host species an allele was expected to infect is the mean host specificity across 9,999 randomizations (“Mean expected”); the standard deviation of the mean is also presented (“SD expected”) and the <i>P</i> value indicating whether the observed host specificity differed from the random expectation. Human infectious alleles (A, B, I, and K) are italicized.	38
Table 2.2	Observed and expected <i>ospC</i> allele diversity in infected vertebrate host species (<i>M. mephitis</i> , striped skunk; <i>M. pennsylvanicus</i> , eastern meadow vole; <i>M. musculus</i> , house mouse; <i>Peromyscus</i> , white-footed/deer mice; <i>S. carolinensis</i> , eastern gray squirrel; <i>T. striatus</i> , eastern chipmunk). Allele diversity (“Observed”) was quantified as the number of unique <i>ospC</i> alleles in each host species. The expected number of alleles was found by averaging the expected values across all 9,999 randomization trials (“Mean expected”); the standard deviation of the mean is also presented (“SD expected”) and the <i>P</i> value indicating whether the <i>ospC</i> allele diversity in a host species differed from a random expectation.	39
Table A.1	Details on the seven sites across Delaware where blacklegged ticks (<i>Ixodes scapularis</i>) were collected. All ticks were collected using both drag and flag methods. Sites NC1 and K1 were sampled at two separate locations within the site on the same day. Each site shows the nearest city in Delaware, the habitat type, vegetation, and composition of the	

	sampling location along with the temperature, relative humidity, and weather conditions on the sampling day.	58
Table A.2	An example of mapping statistics used to determine the presence of an <i>ospC</i> (or <i>vsp</i>) allele. These are the results from the sample DT-3, an infected tick from site NC1. Using the metrics 1) any allele with all nucleotide positions sequenced to a depth of coverage of at least 5X and 2) any allele where mean depth of coverage was greater than the standard deviation of depth of coverage between nucleotide positions 200 and 550 (the more variable positions), we determined that DT-3 had five <i>B. burgdorferi ospC</i> alleles (N, D, M, C, and B; see column “Alleles scored as present”). The reference file we used is from Di et al. (2018) and includes 19 <i>ospC</i> alleles from <i>B. burgdorferi</i> and one <i>vsp</i> allele from <i>B. myiamotoi</i>	59
Table A.3	Quantitative metrics used to determine whether an allele was present. Each sequenced sample is presented and shows the alleles determined to be present along with the allele’s percentage of coverage (i.e., the percentage of all mapping reads that mapped to a particular allele), the mean and standard deviation of depth of coverage between nucleotide positions 200 and 550 of the reference, and the proportion of nucleotide positions in the reference with a depth of coverage of at least 5X. We considered an allele to be present in an infection if all nucleotide positions in the reference of the allele were sequenced to a depth of coverage of at least 5X and where mean depth of coverage was greater than the standard deviation of depth of coverage between nucleotide positions 200 to 550 (the more variable positions). Sample DT-4 and Is-49 partially mapped to <i>vsp</i> (a gene from <i>B. myiamotoi</i> that is present in the reference file) and are shown in the table below.	60
Table A.4	Mapping statistics for each sample. Each sequenced sample is presented (ID) with total number of reads, number of reads that mapped to the reference file, number of reads that did not map (unmapped), number of reads that mapped and paired, and the percentage of unmapped reads. Two samples (DT-4 and Is-49) had more than 50% unmapped reads and those represent the two punitive <i>B. myiamotoi</i> samples.	65
Table A.5	Bray-Curtis dissimilarities between sites based on the relative frequencies of <i>ospC</i> alleles. Only sites with more than three alleles present (Table 2) were considered.....	67
Table A.6	Samples with mixed <i>ospC</i> allele infections. A “1” indicates infection of a sample (ID) with a particular allele. Allele multiplicity refers to the total number of alleles found in an individual sample. The majority of mixed	

	infections contained two <i>ospC</i> alleles, with the greatest number of alleles in one infection being seven.	68
Table B.1	Species sample size and number of <i>B. burgdorferi</i> infections as determined by qPCR across sites. Frequency of <i>ospC</i> alleles are shown on a species level across sites along with the total number of alleles sequenced also shown.....	77
Table B.2	Diversity of the mammalian host species an <i>ospC</i> allele infects as determined by the Gini-Simpson index. Observed Gini-Simpson index values were calculated from the original <i>ospC</i> frequency data. Expected Gini-Simpson index values represent the mean Gini-Simpson index after randomizing <i>ospC</i> allele frequencies across infected hosts 9,999 times. The two-tailed P values for each of the alleles represents if the Gini-Simpson index is more extreme than expected by chance.	79
Table B.3	Diversity of <i>ospC</i> alleles a mammalian host species is infected by determined by the Gini-Simpson index. Observed Gini-Simpson index values were calculated from the original <i>ospC</i> frequency data. Expected Gini-Simpson index values represent the mean Gini-Simpson index after randomizing <i>ospC</i> allele frequencies across infected hosts 9,999 times. The two-tailed P values for each of the alleles represents if the Gini-Simpson index is more extreme than expected by chance.	80

LIST OF FIGURES

- Figure 1.1 Positive relationship between the geographic distances between sites and the Bray-Curtis dissimilarities between sites based on *ospC* allele relative frequencies suggesting that the more geographically distant sites are, the more they differ in their compositions of *ospC* alleles. Site relationships are labeled and only include sites with greater than three *ospC* alleles. The Mantel's correlation coefficient and associated P value are presented at the top of the figure. 15
- Figure 1.2 Frequency distribution of number of *ospC* alleles in mixed infections (i.e., infections with more than one *ospC* allele present). 17
- Figure 1.3 Positive correlation between relative *ospC* allele frequencies in single infections only (i.e., one allele present in an infected tick) and the relative frequencies of those alleles in mixed infections only (i.e., multiple alleles present in an infected tick). Allele identities are plotted alongside the points on the graph for reference; several alleles had identical relative frequencies and the allele names separated by commas reveal overplotting. The Spearman's correlation coefficient and associated P value are presented at the top of the figure. 18
- Figure 2.1 *B. burgdorferi* prevalence (as determine by qPCR) across sites (Goat Farm, Rutgers' Gardens, University Inn) and host species (*B. brevicauda*, northern short-tailed shrew; *D. virginiana*, Virginia opossum; *M. mephitis*, striped skunk; *M. monax*, groundhog; *M. musculus*, house mouse; *M. pennsylvanicus*, eastern meadow vole; *P. lotor*, raccoon; *Peromyscus*, white-footed/deer mice; *S. carolinensis*, eastern gray squirrel; *T. striatus*, eastern chipmunk). The number of captured individuals of each species at each site is shown above its corresponding bar. 35
- Figure 2.2 Heatmap showing *ospC* allele frequencies across host species (*T. striatus*, eastern chipmunk; *S. carolinensis*, eastern gray squirrel; *Peromyscus*, white-footed/deer mice; *M. pennsylvanicus*, eastern meadow vole; *M. musculus*, house mouse; *M. mephitis*, stripped skunk). Higher allele frequencies are redder and lower frequencies are bluer. Human infectious alleles (A, B, I, and K) are bolded. 36

Figure 2.3	Frequency of mixed allele infections by host species (<i>M. mephitis</i> , striped skunk; <i>M. pennsylvanicus</i> , eastern meadow vole; <i>M. musculus</i> , house mouse; <i>Peromyscus</i> spp., white-footed/deer mice; <i>T. striatus</i> , eastern chipmunk). The number of <i>ospC</i> alleles within a mixed infection is represented by the x-axis (“ <i>ospC</i> allele multiplicity”) and the frequencies are shown on the y-axis.	41
Figure 2.4	<i>OspC</i> allele compositions in <i>B. burgdorferi</i> infected small mammals captured multiple times (June-September 2021). Only recaptured individuals infected with <i>B. burgdorferi</i> at least once are included. All but one individual (RGCVB112) are <i>Peromyscus</i> spp. (white-footed/deer mice); RGCVB112 is an eastern chipmunk. Relative allele frequencies are shown on the y-axis (allele identity is labeled and has a unique color) and capture date on the x-axis; individuals are represented by separate panels in the figure. A point is shown on the x axis for each of the capture dates, a point without a bar indicates the individual was not infected upon capture.	42
Figure A.1	Sequence depth of coverage plotted over the reference sequences for each infected tick sample (sample name is presented at the top of each panel). The y-axis shows depth of coverage, and the x-axis refers to the sequence coordinates of the <i>ospC</i> (or <i>vsp</i>) gene. <i>OspC</i> alleles from the reference sequences are distinguished by color (see legend at the bottom of the figure) and alleles that were considered to have sufficient coverage to be considered true infections have a label next to their depth of coverage line to further distinguish them. Several samples did not map well to any of the reference sequences. For example, sample DT-1 did not fully map to any of the 19 reference alleles and may represent a novel allele. Mixed infections can be identified by multiple alleles mapping to the same sample. For example, DT-3 has five <i>ospC</i> alleles present (N, C, M, D, and B) with allele N having the highest relative coverage in the sample. Two samples (DT-4 and Is-49) partially mapped to the <i>vsp</i> gene of <i>Borrelia myiamotoi</i>	75
Figure B.1	Positive correlation between <i>ospC</i> allele relative frequencies in single infections (i.e., one allele in an infection) and mixed infections (more than one allele present in an infection). Allele identities are presented beside their points on the figure. Despite being relatively common in single infections, allele U was not found in mixed infections.	81

ABSTRACT

Characterizing the diversity of genes associated with virulence and transmission of a pathogen across the pathogen's distribution can inform our understanding of host infection risk. *Borrelia burgdorferi* is a vector-borne bacterium that causes Lyme disease in humans and is common in the United States. The outer surface protein C (*ospC*) gene of *B. burgdorferi* exhibits substantial genetic variation across the pathogen's distribution and plays a critical role in virulence and transmission in vertebrate hosts. In fact, *B. burgdorferi* infections that disseminate across host tissues in humans are associated with only a subset of *ospC* alleles. Delaware has a high incidence of Lyme disease, but the diversity of *ospC* in *B. burgdorferi* in the state has not been evaluated. In chapter 1, we used PCR to amplify *ospC* in *B. burgdorferi*-infected blacklegged ticks (*Ixodes scapularis*) in sites statewide and used short-read sequencing to identify *ospC* alleles. *B. burgdorferi* prevalence in blacklegged ticks varied across sites, but not significantly so. We identified 15 previously characterized *ospC* alleles accounting for nearly all of the expected diversity of alleles across the sites as estimated using the Chao1 index. Nearly 40% of sequenced infections (23/58) had more than one *ospC* allele present suggesting mixed strain infections and the relative frequencies of alleles in single infections were positively correlated with their relative frequencies in mixed infections. Turnover of *ospC* alleles was positively related to distance between sites with closer sites having more similar allele compositions than more distant sites. This suggests a degree of *B. burgdorferi* dispersal limitation or habitat specialization. *OspC* alleles known to cause disseminated infections in humans were found at the highest frequencies across sites, corresponding to Delaware's high incidence of Lyme disease.

Negative interactions among pathogen genotypes during infection may affect overall transmission dynamics in multi-host systems. For example, specialist pathogens may limit the impact of generalists, which may become emerging infectious diseases. *Borrelia burgdorferi*, a bacterium that causes Lyme disease, is transmitted primarily among wildlife hosts and *Ixodes* ticks. Despite significant bacterial genetic diversity, most human infections are caused by bacteria with specific alleles of the outer surface protein C (*ospC*) gene (“human infectious alleles”; HIAs). In chapter 2, we tested 272 individuals of 11 mammalian species for *B. burgdorferi* and sequenced *ospC* alleles. Multiple allele (“mixed”) infections were common in individuals of all species. HIAs were most common in mice (*Peromyscus* spp.) with only one HIA in a site with almost no mice. Chipmunks (*Tamias striatus*) were hosts of allele U, the only specialist. Surprisingly, while an unexpectedly large number of alleles were recovered from chipmunks, including HIAs, allele U was not found in mixed infections. These results are consistent with allele U excluding other alleles, perhaps through indirect host-immune mediated mechanisms, thereby reducing the capacity of chipmunks to act as reservoirs for HIAs. This suggests that specialized pathogens may be able to mitigate the infection risk of hosts they do not infect.

Chapter 1

PREVALENCE OF *BORRELIA BURGdorFERI* AND DIVERSITY OF ITS OUTER SURFACE PROTEIN C (OSPC) ALLELES IN BLACKLEGGED TICKS (*IXODES SCAPULARIS*) IN DELAWARE

Introduction

Population genetics can reveal pathogen dispersal and distribution patterns and the evolutionary and ecological processes shaping those patterns (Huyse et al., 2005; McDonald and Linde, 2002). Since the virulence and transmissibility of pathogens often depends on specific pathogen genes (Musser, 1996; Tonetti et al., 2015), characterizing the genetic diversity and distributions of those genes may be especially relevant with regards to host infection risk (e.g., Aguilera-Arreola et al., 2005). Importantly, population genetic patterns can be scale-dependent (Huyse et al., 2005) and therefore sampling throughout the geographic distribution of a pathogen is critical to fully understanding its population genetic structure and distribution.

Borrelia burgdorferi is a zoonotic pathogen that causes Lyme disease in humans and is the most common vector-borne infectious disease in the United States (Kugeler et al., 2021). The bacterium is primarily transmitted to vertebrate hosts, including humans, by the bite of blacklegged ticks (*Ixodes scapularis*) in the upper Midwest and eastern United States and western blacklegged ticks (*Ixodes pacificus*) in the western United States (Movila et al., 2012). In humans, *B. burgdorferi* can cause generalized symptoms and without treatment can develop into mono- or oligo-articular arthritis and a variety of neurologic complications (Cardenas-de la Garza et al., 2019; Steere, 1987). Interestingly, not all *B. burgdorferi* (*sensu stricto*) infections in humans

disseminate from the primary site of infection (skin) into secondary sites (blood and cerebrospinal fluid) and cause disease (Seinost et al., 1999; Wormser et al., 2008). It is hypothesized that infectiousness of *B. burgdorferi* in humans is associated with particular alleles of the outer surface protein C (*ospC*) gene (Seinost et al., 1999; Wormser et al., 2008).

OspC is one of the most polymorphic genes in the *B. burgdorferi* genome (Baranton et al., 2001; Jauris-Heipke et al., 1993; Wang et al., 1999) and its expression is required to establish vertebrate infections (Earnhart et al., 2010; Grimm et al., 2004a; Tilly et al., 2006). *OspC* alleles have been classified into major allele groups that typically differ by more than 8% in their genetic sequences, but by no more than 2% within major groups (Wang et al., 1999). There are at least 28 distinct *ospC* major allele groups (Rudenko et al., 2013). Of those major alleles, four (A, B, I, and K) have been found to cause the majority of disseminated infections in humans (Seinost et al., 1999; Wormser et al., 2008). Therefore, the distributions of *ospC* alleles are an important part of understanding Lyme disease risk in humans.

OspC allele diversity varies across the United States (Barbour and Travinsky, 2010; Tyler et al., 2018; Walter et al., 2017). Barbour and Travinsky (2010) examined infected ticks and found 20 *ospC* alleles (18 major allele groups) in the Northeast, 26 in the Midwest (23 major groups), and 12 in northern California (12 major groups); 18 *ospC* major allele groups occurred in both the Northeast and Midwest but differed in frequency. Furthermore, individual locations often have high *ospC* allele diversity (Alghaferi et al., 2005; Bunikis et al., 2004; Di et al., 2018; Wang et al., 1999). The *ospC* gene shows high rates of apparent horizontal gene transfer and recombination

(Walter et al., 2017; Wang et al., 1999), although this may occur more frequently in the midwestern than the northeastern U.S. (Brisson et al., 2010).

Several methods have been used to detect *ospC* major alleles. PCR and Sanger sequencing of the *ospC* gene can detect and sequence unique alleles (Jauris-Heipke et al., 1993), but can often not separate the genetic sequences of alleles when they occur in mixed infections (i.e., more than one allele present in a single infection) without cloning and sequencing of clones (Di et al., 2018). Since multiple infections can be common in both ticks and vertebrate hosts (Brisson and Dykhuizen, 2004), alternative and less costly methods have also been employed. For example, reverse line blot hybridization and more recently LUMINEX genotyping have been used to detect *ospC* major alleles (Brisson and Dykhuizen, 2004; Pearson et al., 2022; Qiu et al., 2002). These approaches are cost-effective and useful when targeting known *ospC* alleles, but cannot detect novel or recombinant alleles (e.g., allele C is likely a result of genetic exchange between alleles I, B, and E; Wang et al. 1999, Qiu et al. 2002). Furthermore, these methods have typically relied on nested PCR reactions which can increase contamination risk relative to single round PCR (Bensch et al., 2021). Recently, Di et al. (2018) designed primers in regions that flank the *ospC* gene for a single round PCR and a protocol for sequencing the PCR product with short-read next-generation sequencing (Di et al., 2018). This protocol has the potential to resolve mixed infections and allow for sequencing of full *ospC* alleles.

While the distributions of *ospC* alleles have been characterized in many locations (Alghaferi et al., 2005; Anderson and Norris, 2006; Brisson and Dykhuizen, 2004; Bunikis et al., 2004; Di et al., 2018; Tyler et al., 2018; Wang et al., 1999), there are areas of the *B. burgdorferi* distribution which have not been investigated. For

example, *ospC* distribution data have not been reported from the state of Delaware. Delaware is a notable state in the distribution of *B. burgdorferi* because it has a high incidence of Lyme disease (<https://www.cdc.gov/lyme/datasurveillance/maps-recent.html>) and it was likely part of a historical refugium of blacklegged ticks during the last glacial maximum (Xu et al., 2020). Furthermore, the genetic clade of blacklegged ticks that expanded from the Delmarva Peninsula (Delaware and parts of Maryland and Virginia) as the ice sheet receded is largely responsible for transmitting *B. burgdorferi* (Xu et al., 2020). Here we characterize *ospC* allele diversity and *B. burgdorferi* prevalence in blacklegged ticks (i.e., the proportion of ticks infected by *B. burgdorferi*) in several sites spanning all three counties in Delaware. We also compare the sensitivity of the PCR designed by Di et al. (2018) to target the *ospC* gene with a more commonly employed quantitative PCR assay (qPCR) used to detect *B. burgdorferi* (Hojgaard et al., 2014).

Methods

Tick collection

We collected nymphal blacklegged ticks from seven locations on state-owned land across Delaware's three counties in the spring and summer of 2019 (Table A.1) using standard drag and flag methodology. Briefly, we used a 1 m² white corduroy flag and checked for ticks every 10 m; total survey distance per site varied. We stored all ticks in 90% ethanol in the field. Two sampling locations (NC1, NC2) are in New Castle County, Delaware's most northern and urban county. Site NC1 is closest to the city of Newark, Delaware (DE) with the majority of the vegetative community consisting of tulip poplar (*Liriodendron tulipifera*) and a variety of oak species (*Quercus* spp.). In Rockland, DE NC2 is composed of freshwater marsh and tulip poplar. Two sampling locations (K1, K2) are in Kent County, with K1 near Leipsic,

DE and K2 near Felton, DE. Tidal salt marsh and a variety of oak and pine (*Pinus* spp.) species are found in K1 with a millpond and hardwood forest found at K2. Three sites are found in Sussex County (S1: near Dagsboro, DE; S2: near Milton, DE; S3: near Laurel, DE). Site S1 is composed of hardwood forests and resides on the shores of the inland bay. Black-barrier salt marsh, grassland dunes, and hardwood forests are all found at S2. The majority of the vegetative community at site S3 consists of oak and beech (*Fagus* spp.) species with successional tulip poplar. Kent and Sussex Counties represent the central and southern part of Delaware, respectively, and are not as developed and urbanized as the northern New Castle County. DNA was extracted from all ticks across the seven sites and used to calculate prevalence of *B. burgdorferi* infection and to quantify *ospC* allele diversity in infected ticks.

As part of a separate study, nymphal and adult blacklegged ticks were collected in a similar fashion from four sites; three of the original seven (NC1, K1, K2) and a third site near Wilmington, DE in New Castle County (NC3). Site NC3 is dominated by oak and beech trees. Genomic DNA was extracted from those ticks, and they were screened for *B. burgdorferi* (results not shown). Prior to DNA extraction ticks were washed in distilled water, bisected with a sterile scalpel, with half of the tick once again bisected and used for DNA extraction. The specific extraction protocol followed the Insect Supplementary Protocol for the Qiagen DNeasy Blood and Tissue Kit (Qiagen cat# 69506). We sequenced the *ospC* alleles from 31 of those infected ticks (nymphs and adults were not differentiated prior to DNA extraction) and used those data as part of our analysis of *ospC* allele diversity. The 31 infected ticks from this separate study were not included in our *B. burgdorferi* prevalence analyses because nymphs and adults had not been differentiated.

Molecular screening of ticks for *B. burgdorferi* and *ospC* sequencing

For the nymphal ticks collected from the seven sites listed above (NC1, NC2, K1, K2, S1, S2, S3) we washed individual ticks in a 10% bleach solution for 30 seconds followed by two separate one-minute washes in ultrapure water prior to DNA extraction. Ticks were then bisected with a sterile scalpel or razor blade inside a biosafety cabinet and both parts were used for the DNA extraction. We extracted DNA from each tick individually using a DNeasy Blood and Tissue kit (Qiagen cat# 69506) following the manufacturer's protocol. DNA concentrations were measured with a Qubit 4 Fluorometer (Thermo Fisher Scientific, Waltham, MA, USA). The average DNA concentration from black-legged tick extractions was $0.7 \text{ ng}/\mu\text{l} \pm 0.24 \text{ s.d.}$ in a total volume of $400 \mu\text{l}$.

We tested each sample for *B. burgdorferi* using PCR and qPCR. The PCR (carried out in a Bio-Rad T100 thermocycler) protocol amplified the *ospC* gene using a single round of PCR with the primers of Di et al. (2018). Our $25 \mu\text{l}$ PCR reaction included $12.5 \mu\text{l}$ of hot start master mix (DreamTaq Hot Start Master Mix 2X, Thermo Fisher Scientific, Waltham, MA, USA cat# K9011), $1.25 \mu\text{l}$ of a $10 \mu\text{M}$ solution of each primer (Oc-Fwd and Oc-Rev), $8 \mu\text{l}$ of ultrapure water and $2 \mu\text{l}$ of DNA template. The PCR thermal profile protocol used one cycle of initial denaturation (2 min at 95°C), followed by 35 cycles of denaturation (30 s at 95°C), annealing (30 s at 52°C), and extension (1 min at 72°C), and a final extension at 72°C for 10 min. All PCRs included at least one negative control (ultrapure water) and one positive control (DNA from a *B. burgdorferi*-infected tick).

We loaded the PCR product onto a 1.5% agarose gel stained with GelRed dye (Biotium, Inc., Hayward, CA, USA cat# NC9594719) to visualize amplification. We

combined PCR amplicon with 6X Trirack DNA loading dye (Thermo Fisher Scientific, Waltham, MA, USA cat# SM0241) and performed gel electrophoresis in a 1X TAE solution at 90 V for 40 min. Gels were visualized on the Bio-Rad GelDoc Go imaging system.

The University of Delaware's Genotyping and Sequencing Center at the Delaware Biotechnology Institute (DBI) prepared Illumina libraries using 1 ng of input DNA (PCR product from infected ticks) and the QuantaBio Library prep kit and sparQ unique dual index adapters (QuantaBio, Beverly, MA, USA). Library preparation included a 20 min fragmentation time and 15 PCR cycles. Final library concentrations were quantified using a Qubit 3.0 fluorometer and Advanced Analytical Technologies, Inc Fragment Analyzer (Thermo Fisher Scientific, Waltham, MA, USA). Libraries were pooled in equimolar ratios and sequenced on an Illumina Miseq using a 500-cycle nano sequencing kit (Illumina, San Diego, CA, USA). The sequencing was paired-end with 251 cycles per read. Raw sequence files can be accessed from an online Dryad repository (S. Shifflett et al., 2023).

We also tested the same samples for *B. burgdorferi* using a commonly used qPCR protocol (Hojgaard et al., 2014) for comparison of the effectiveness of the single round PCR to detect infections. While samples were run once in the PCR protocol (described above), they were run in duplicate in the qPCR (all duplicates gave the same result). One qPCR reaction included four primers (two primer sets) and two probes to target *B. burgdorferi* and *I. scapularis*. We combined 10 μ l of TaqMan Fast Advanced Master Mix (Thermo Fisher Scientific, Waltham, MA, USA cat# 4444557), 0.18 μ l of a 10 μ M solution of each primer (fliD-Fwd and fliD-Rev, targeting *B. burgdorferi*, and actin-Fwd and actin-Rev, targeting the tick *I. scapularis*),

0.05 μ l of a 10 μ M solution of each probe (*fliD* and actin probes), and 7.18 μ l of ultrapure water in each reaction. Primer and probe sequences were taken from Hojgaard et al. (2014). Each well in the qPCR plate was brought up to 20 μ l with the addition of 2 μ l of DNA extract. The qPCR thermal profile protocol was 50°C for two minutes followed by two minutes at 95°C for the initial stage, followed by 40 cycles of 95°C for one second and 60°C for 20 seconds. The qPCR reactions were performed on an Applied Biosystems QuantStudio 3 Real-Time PCR system (Thermo Fisher Scientific, Waltham, MA, USA).

Bioinformatic analysis

OspC alleles were identified from the short-read sequences following the protocol outlined in Di et al. (2018) with minor modifications (bioinformatic code can be accessed in the online Dryad repository; Shifflett et al., 2023). For each sample, we trimmed adaptors using trimgalore v.0.6.6 (Krueger, 2007) followed by a post-trimming quality assessment using FastQC v.0.11.9 (Andrews, 2010). Reads were then mapped to the *ospC* fasta reference file from Di et al. (2018) using bwa v.0.7.17 (Li and Durbin 2009) and saved as a sam file. Sam files were converted to bam files and reads were sorted and indexed using SAMtools v.1.10 (Danecek et al., 2021). We generated depth of coverage at each base pair position of the reference sequences using the SAMtools stats function. Depth of coverage was then plotted for each reference sequence in R v.4.0.1 (R Core Team, 2020) using the ggplot2 v.3.3.5 R package (Wickham, 2016). We identified the presence of particular *ospC* alleles visually based on depth of coverage of mapping to the reference file (Supplementary Fig. 1). We also investigated quantitative metrics to determine whether an allele was present. We used the following metrics to identify sequenced alleles: 1) any allele with

all nucleotide positions sequenced to a depth of coverage of at least 5X and 2) any allele where mean depth of coverage was greater than the standard deviation of depth of coverage between nucleotide positions 200 and 550 in the reference (the more variable positions of the gene; Table A.2). We generated the allele identification metrics using the R package dplyr v.1.0.8 (Wickham et al., 2021). We also quantified the proportion of unmapped reads in each sample using SAMtools stats. Quantifying the proportion of unmapped reads also allowed us to identify potentially novel, unmapped alleles.

Statistical analyses

We used the Chao1 estimator (Chao et al., 2009) to determine the expected number of *ospC* major alleles across our sample sites using the estimateR function in the R package vegan v.2.6-2 (Oksanen et al., 2022) based on the frequencies that each allele appeared in the entire dataset.

We calculated the relative frequency of each *ospC* allele at each site by dividing the number of each allele recovered by the total number of all alleles recovered at the site. We then calculated Bray-Curtis dissimilarities between sites using the relative allele frequencies as an index of allele turnover. Bray-Curtis dissimilarities were calculated using the vegdist function in vegan. We also calculated the relative frequency of each allele across all sites.

We calculated the Great circle distances between sites using the rdist.earth function in the R package fields v.14.1 (Nychka et al., 2021). Geographic distances and Bray-Curtis dissimilarities were correlated using a Mantel correlation test with the mantel function in vegan. A positive correlation would suggest that the more geographically distant sites are, the more different their *ospC* allele compositions are.

We defined mixed infections as individual ticks with more than one sequenced *ospC* allele. For alleles found in mixed infections we calculated the frequency of the alleles in mixed infections only and correlated this (Pearson's correlation) with the frequency of alleles in single infections only using the `cor.test` function in R. Allele frequency was determined by the number of times the allele appeared in a particular infection type (i.e., mixed or single) divided by the total number of all alleles in all infections of the same type). A positive correlation would suggest that alleles common in mixed infections are common overall and may exist in mixed infections largely by chance.

Finally, we quantified the correlation between the PCR and qPCR results of the infection status of each tick (infected, 1; uninfected, 0) from the seven sites used to estimate prevalence. We did this with the `cor` function in R. We also tested whether prevalence of infection (as determined by PCR and qPCR separately) varied among sites using Fisher's exact test. This was calculated with the `fisher.test` function in R. We used R v.4.0.1 (R Core Team, 2020) for all statistical analyses and R code is presented in the online Dryad repository (S. Shifflett et al., 2023).

Results

We tested 111 nymphal *I. scapularis* ticks collected across seven sites in Delaware for *B. burgdorferi* (this does not include the 31 positive ticks collected for another study; Table 1.1). The qPCR protocol found four more *B. burgdorferi* infections than the PCR protocol, however, one infection identified by the PCR and not by the qPCR partially mapped to the *vsp* gene of *B. miyamotoi* (sample Is-49; Figure A.1, Table A.3). In total, the PCR identified 29 infected ticks (one putatively infected by *B. miyamotoi*, the rest by *B. burgdorferi*) and the qPCR identified 32 *B.*

burgdorferi infected ticks (Table 1.1). The two methods were therefore strongly correlated ($r = 0.913$; the putative *B. miyamotoi* infection was not counted as an infection when calculating *B. burgdorferi* prevalence). *B. burgdorferi* prevalence varied across sites, but not significantly (PCR results, Fisher's exact test $P = 0.07$; qPCR results, $P = 0.144$; Table 1.1), including when we dropped the two sites with the lowest sample size (S2 and S3 each had only five ticks sampled; PCR results, Fisher's exact test $P = 0.06$; qPCR results, $P = 0.136$).

Table 1.1: The number of nymphal blacklegged ticks by site (“Sample size”) and the number of those ticks infected with *B. burgdorferi* as determined by PCR (“PCR infections”) and qPCR (“qPCR infections”); prevalence of *B. burgdorferi* as a percentage is presented in parentheses following the number of infections.

Site	County	Nearest city	Sample size	PCR infections (prevalence)	qPCR infections (prevalence)
NC1	New Castle	Newark	23	4 (17%)	4 (17%)
NC2	New Castle	Rockland	14	2 (14%)	3 (21%)
K1	Kent	Leipsic	40	17 (43%)	18 (45%)
K2	Kent	Felton	13	2 (15%)	3 (23%)
S1	Sussex	Dagsboro	11	1 (9%)	2 (18%)
S2	Sussex	Milton	5	2 (40%)	2 (40%)
S3	Sussex	Laurel	5	0 (0%)	0 (0%)

We sequenced *ospC* from 29 infected ticks; 28 ticks had *ospC* alleles from *B. burgdorferi* and another infected tick mapped partially to the *vsp* gene of *B. miyamotoi*. We additionally amplified and sequenced the *ospC* gene from the extracted DNA of 31 infected ticks from sites NC1 (n = 11), NC3 (n = 8), K1 (n = 5), and K2 (n = 7). One sample from these additional 31 infected ticks also partially mapped to the *vsp* gene of *B. miyamotoi* (sample DT-4, Figure A.1 and Table A.3). After removing the two samples that mapped to the *vsp* reference, we identified 15 *ospC* major alleles among all 58 sequenced *B. burgdorferi* infections (Table 1.2). At least one allele known to cause disseminated infections in humans (“human infectious alleles”) was found at each site (Table 1.2). Allele A, a human infectious allele, occurred at the highest frequency, infecting a total of 21 ticks at six sites (Table 1.2).

Allele U, previously found to specialize on eastern chipmunks (*Tamias striatus*; Brisson and Dykhuizen 2004), occurred in one tick at site NC1. On average, each allele was found at $2.93 \text{ sites} \pm 0.384 \text{ s.e.}$

Two samples had reads mapping incompletely to several alleles (DT-1 and Is-89; Figure A.1 and Table A.3) suggesting a possible novel allele formed through recombination, however we did not confirm this finding with additional sequencing. One sample (Is-89) from site NC1 fully mapped to two *ospC* alleles (as defined by our criteria above) with multiple alleles also showing relatively high mapping but not meeting the defined inclusion criteria; this sample may represent a mixed infection with two previously known *ospC* alleles (C and K) and one potentially novel allele. In addition, two samples had > 50% of reads unmapped and belonged to the putative *B. miyamotoi* infections (samples DT-4 and Is-49, Table A.4; these two samples were not resequenced). Unmapped reads generally BLASTed to *I. scapularis* or *Borrelia* (results not shown).

We calculated Bray-Curtis dissimilarities between sites based on relative *ospC* allele frequencies. We removed sites with three or fewer alleles detected leaving four sites to compare (NC1, NC3, K1, and K2). K1 was most similar to K2 and K2 was least similar to NC3; the mean dissimilarity between sites was $0.531 \pm 0.048 \text{ s.e.}$ (Table A.5). We correlated Bray-Curtis dissimilarities with geographic distances between sites (Figure 1.1). We found a positive correlation between *ospC* allele turnover and distance between sites (Mantel $r = 0.75$, $P = 0.042$) with closer sites showing more similar *ospC* allele compositions. Sites NC1 and NC3 had a higher-than-expected Bray-Curtis dissimilarity based on their close geographic distance (Figure 1.1).

Table 1.2 *OspC* allele frequencies by site with the number of infected ticks sequenced (“sample size”) in parenthesis following the site name. Alleles associated with human disseminated infections are italicized (*A*, *B*, *I*, and *K*). The final two rows of the table show the total number of infections of each *ospC* allele (“Total alleles”) and their relative frequencies (i.e., percentage of the total number of alleles found). The final two columns show the total number of alleles (“Total alleles”) and the number of unique alleles (“Unique alleles”) found per site. Alleles are arranged in order of their relative frequencies.

Site (sample size)	<i>A</i>	<i>K</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>T</i>	<i>E</i>	<i>N</i>	<i>G</i>	<i>M</i>	<i>F</i>	<i>H</i>	<i>I</i>	<i>O</i>	<i>U</i>	Total alleles	Unique alleles
NC1 (14)	4	3	3	8	2	2	0	3	2	1	1	0	0	1	1	31	12
NC2 (2)	1	0	1	0	0	0	1	0	0	0	0	0	0	0	0	3	3
NC3 (8)	2	1	1	0	1	3	4	0	1	1	1	1	1	0	0	17	11
K1 (22)	8	9	6	1	2	0	0	1	1	2	2	2	2	0	0	36	11
K2 (9)	5	5	1	0	2	0	0	0	0	0	0	0	0	0	0	13	4
S1 (1)	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1
S2 (2)	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	2	2
S3 (0)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Total alleles	21	19	12	9	7	5	5	5	4	4	4	3	3	1	1	103	15
Relative frequency (%)	20.4	18.4	11.7	8.7	6.8	4.9	4.9	4.9	3.9	3.9	3.9	2.9	2.9	1	1		

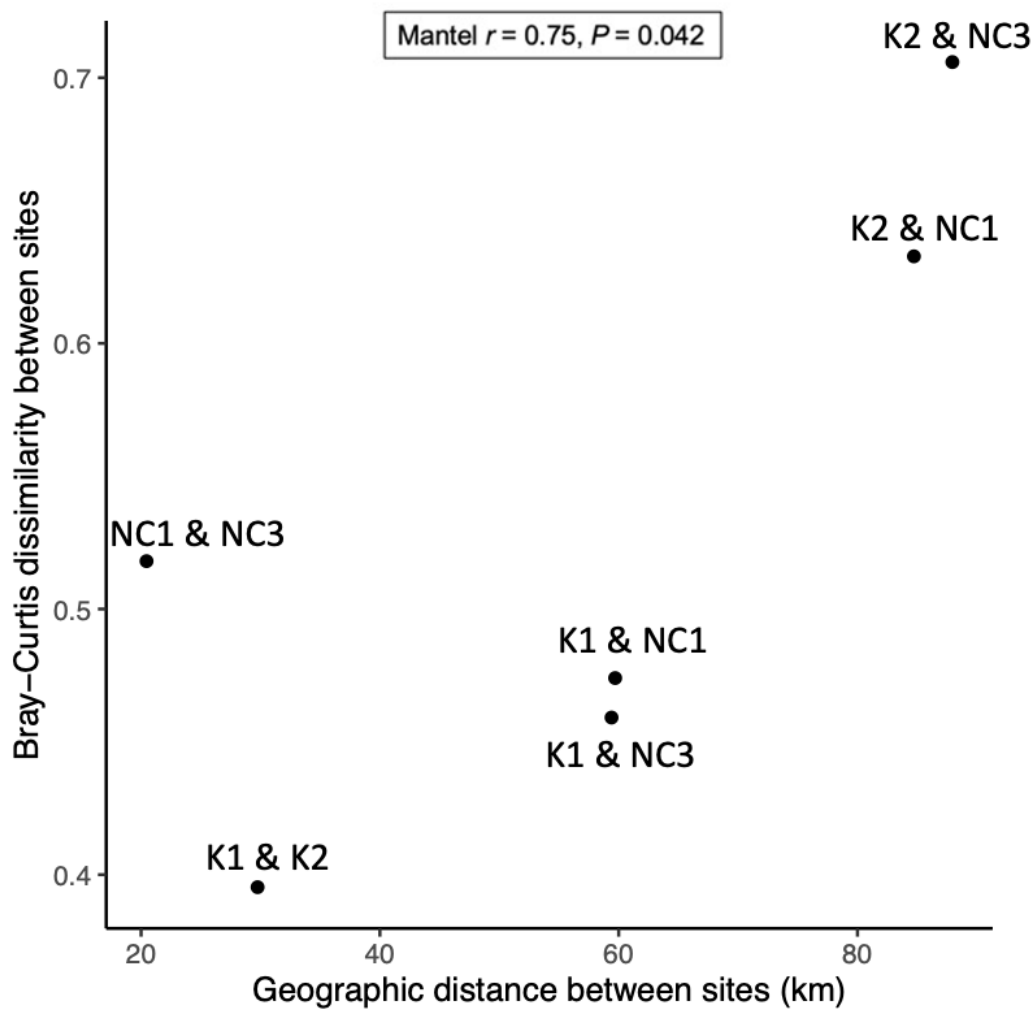


Figure 1.1 Positive relationship between the geographic distances between sites and the Bray-Curtis dissimilarities between sites based on *ospC* allele relative frequencies suggesting that the more geographically distant sites are, the more they differ in their compositions of *ospC* alleles. Site relationships are labeled and only include sites with greater than three *ospC* alleles. The Mantel's correlation coefficient and associated P value are presented at the top of the figure.

Of 58 infected ticks, 23 had mixed infections as determined by the presence of more than one *ospC* allele (Table 1.1, Table A.6). Of these mixed infections, the majority occurred at site K1 (10 mixed infections) followed by site NC1 (6 mixed infections), NC3 (4 mixed infections), K2 (2 mixed infections), and NC2 (1 mixed infection). The mean number of alleles present in mixed infections was 3.0 ± 0.32 s.e., with mixed infections of two alleles occurring most frequently ($n = 14$) and one mixed infection having seven alleles (the maximum found; Figure 1.2). The relative frequency of *ospC* alleles in single infections was positively correlated with the relative frequency of alleles in mixed infections ($r = 0.537$, $P = 0.039$; Figure 1.3), a result consistent with the assemblage of mixed infections being a random process. However, allele B was somewhat of an outlier appearing more often in mixed infections (15.9% of alleles in mixed infections) than might be expected based on its frequency in single infections (2.9% of alleles in single infections; Figure 1.3).

We estimated the total number of *ospC* alleles across our study sites using the Chao1 estimator based on the frequencies of the alleles we recovered that could be assigned unambiguously to alleles in the reference file (Table 1.2). We estimated that there are 16.0 ± 1.64 s.e. major alleles, suggesting that we recovered nearly all (15/16) of the allele diversity across the study sites.

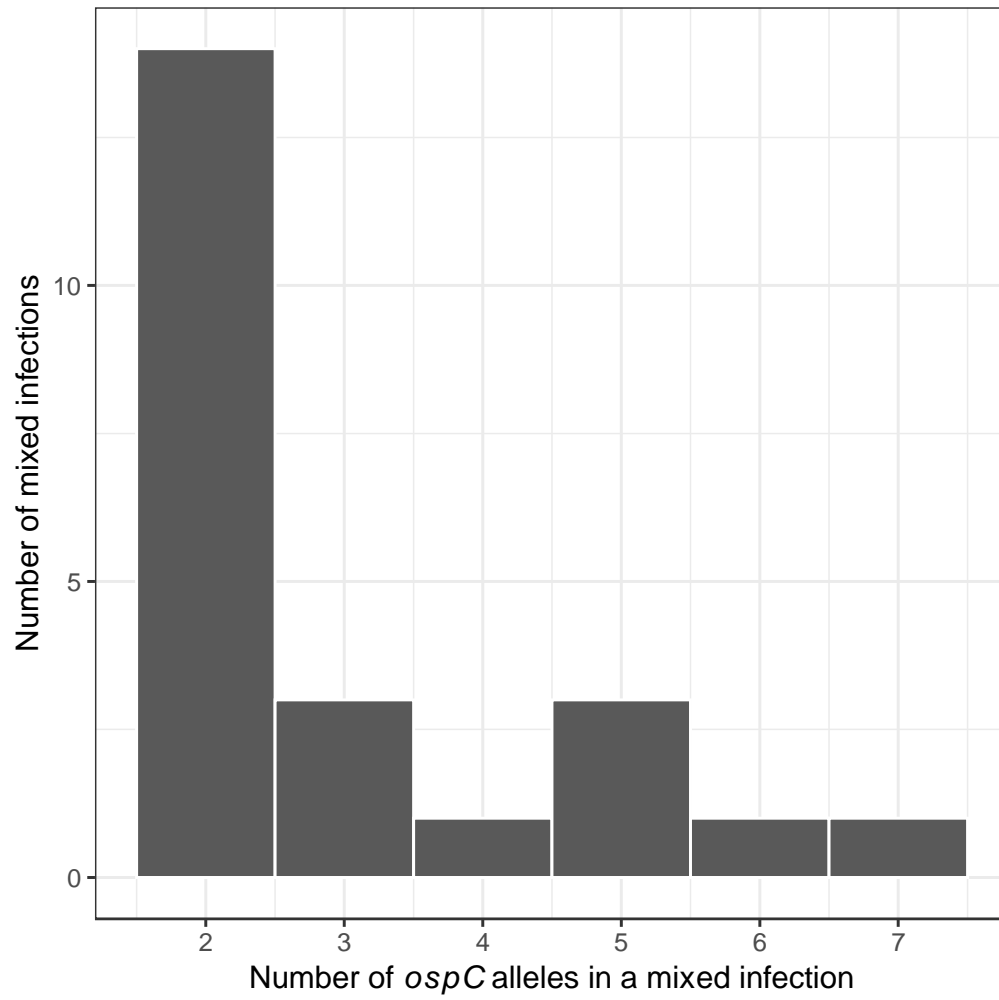


Figure 1.2 Frequency distribution of number of *ospC* alleles in mixed infections (i.e., infections with more than one *ospC* allele present).

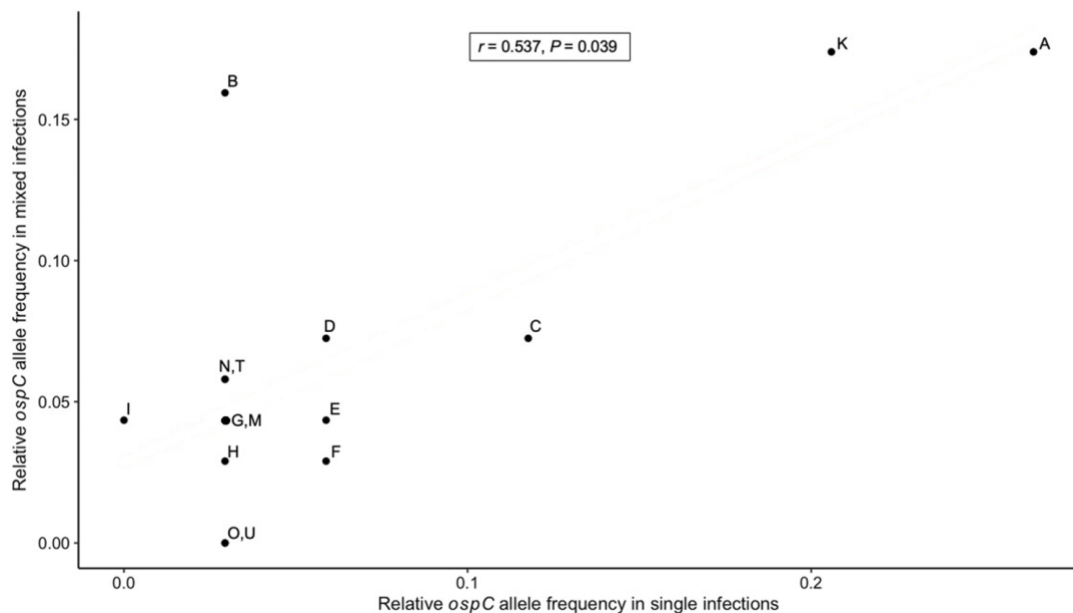


Figure 1.3 Positive correlation between relative *ospC* allele frequencies in single infections only (i.e., one allele present in an infected tick) and the relative frequencies of those alleles in mixed infections only (i.e., multiple alleles present in an infected tick). Allele identities are plotted alongside the points on the graph for reference; several alleles had identical relative frequencies and the allele names separated by commas reveal overplotting. The Spearman's correlation coefficient and associated P value are presented at the top of the figure.

Discussion

The prevalence of *B. burgdorferi* in nymphal blacklegged ticks varied, but not significantly so across seven sites spanning all three counties of Delaware (Table 1.1). Furthermore, we found human infectious *ospC* major alleles (A, K, B, and I) in 65.5% of the infected ticks and at all sites with infected ticks (Table 1.2), consistent with Delaware being classified as a high incidence state for Lyme disease. The turnover of *ospC* alleles was greater between sites that were more geographically distant from one another (Table A.5, Figure 1.1), suggesting a degree of limitation in *ospC* allele

dispersal or habitat specialization of the *ospC* alleles. Among the four sites with greater than three alleles found (NC1, NC3, K1, and K2), allele richness was high (mean = 9.5 ± 1.85 s.e.) as has been found elsewhere (Bunikis et al., 2004; Di et al., 2018; Wang et al., 1999). In fact, we estimate that we identified nearly all of the *ospC* alleles at the sites we surveyed. Mixed infections (i.e., infections with more than one *ospC* major allele present) were common and may partly reflect the random association of common alleles (Figure 1.3), even if there may be negative interactions between strains in mixed infections (Di et al., 2018).

The prevalence of *B. burgdorferi* in nymphal blacklegged ticks across all seven sites was approximately 27% (qPCR = 28.8%, PCR = 25.2%), similar to the prevalence reported by Di et al. (2018) for nymphal blacklegged ticks in New York. However, prevalence varied from 0 to 45% among the sites (Table 1.1). Some variation in prevalence may be due to error associated with low sample sizes and some may be due to the abundance of suitable habitat and wildlife reservoir hosts. *B. burgdorferi* prevalence has been shown to vary over relatively small spatial scales within states in the US. For example, Adelson et al. (2004) found a 33.6% prevalence of *B. burgdorferi* in nymphal and adult blacklegged ticks in Union County, New Jersey while Courtney et al. (2003) found a 43% prevalence in Hunterdon County New Jersey, just 48 miles (77 km) away. While those differences could reflect temporal variation in prevalence, Courtney et al. (2003) also documented a 61.6% prevalence in northwestern Pennsylvania and a 13.1% prevalence in southeastern Pennsylvania within the same collection years. Many of these studies sampled for both nymphal and adult *I. scapularis* which will factor into prevalence rates. Adults should have a higher prevalence of *B. burgdorferi* than nymphs because they have had an

additional blood meal and therefore more chances to become infected. Within New Castle County, Delaware, Adalsteinsson et al. (2018) found an average *B. burgdorferi* prevalence of 11.2% in nymphal blacklegged ticks, however they found a 15.9% prevalence in sample sites invaded by non-native multiflora rose (*Rosa multiflora*). This prevalence more closely resembled our findings for sites in New Castle County, highlighting that prevalence can vary over small spatial scales and/or over time. Prevalence was similar between the PCR and qPCR methods, however, the qPCR identified four more infections, suggesting it is more sensitive. Nevertheless, the *ospC* single round PCR described by Di et al. (2018) was effective at estimating prevalence in our sample and also identified putative *B. miyamotoi* infections.

Human infection risk depends on both the prevalence of *B. burgdorferi* in ticks and the density of infected ticks (especially nymphal ticks; Barbour and Fish, 1993; Mather et al., 1996; Piesman et al., 1987), and on the presence of human infectious *ospC* alleles. We found such human infectious alleles at all sites containing *B. burgdorferi* infected ticks, a result consistent with previous studies in other locations (Di et al., 2018; Tyler et al., 2018). Furthermore, we found a high diversity of *ospC* alleles (Table 1.2). We identified 15 known *ospC* alleles (Table 1.2) and two potentially novel alleles (Table A.3) through sequencing. Despite limited sampling, we likely identified nearly all of the alleles in the sites according to the Chao1 estimator. Similar *ospC* allele diversity was found in New York and Canada. Di et al. (2018) found 19 *ospC* alleles across four sites in New York, with alleles L, J, and C14 not found in our samples. In southern Canada (southern Manitoba, northwest Ontario, and Nova Scotia), Tyler et al. (2018) found 18 *ospC* alleles with alleles Y and J not found in our samples. All the alleles we identified were found in both the Di et al. (2018) and

the Tyler et al. (2018) studies. In the bordering state of Maryland, a lower number of *ospC* alleles (eight) were found in white-footed mice (Anderson and Norris, 2006; Swanson and Norris, 2008), however both our study and the Maryland study found alleles A and K to occur at the greatest frequencies (Anderson and Norris, 2006; Swanson and Norris, 2008). We found allele K, a human infectious allele, at a relatively high frequency (19% of all alleles found), similar to a study by Qiu et al. (2002) across the Northeast. However, Wang et al. (1999) found allele K to occur at the lowest frequency in Shelter Island, New York. In Di et al.'s (2018) study in New York, alleles not known to cause disseminated infections in humans (T and G) occurred at the highest frequencies (19.26% of all alleles found). The habitat composition of the sites we surveyed differed from wetland forest dominated with softwood pines (*Pinus spp.*) to hardwood forest comprised of oak (*Quercus spp.*) and tulip poplar (*Liriodendron tulipifera*; Table A.1) supporting previous results of *ospC* alleles being found in multiple habitat types (Barbour, 1998; Humphrey et al., 2010).

OspC allele composition was more similar among geographically close sites than among geographically distant sites, suggesting a degree of dispersal limitation (Figure 1.1). Barbour and Travinsky (2010) also identified *ospC* dispersal limitation with allele frequencies varying between geographic regions and geographically close regions (Northeast and Midwest) having more similar allele compositions. Birds and deer play an important role in transporting ticks and their pathogens across large geographic distances (Comstedt et al., 2006; Madhav et al., 2004; Vuong et al., 2014). However, bird species have been shown to vary in their ability to competently pass on *B. burgdorferi* infections to ticks (Ginsberg et al., 2005; Mather et al., 1989; Richter, 2000) while white-tailed deer (*Odocoileus virginianus*) are incompetent hosts (Telford

et al., 1988). Vertebrate hosts that are competent for the majority of *ospC* alleles (e.g., white footed mice and shrews; Brisson and Dykhuizen (2004)) may have more limited dispersal (Stickel, 1968). Thus, variation in the dispersal abilities, propensity to disperse with ticks, and reservoir competence of the many vertebrate hosts of *B. burgdorferi*, may contribute to the dispersal limitation of *ospC* alleles.

Sites K1 and K2 had the most similar *ospC* allele compositions (Table A.5). These sites were both located in the same county and composed of wetland forest and large agricultural fields. Sites K2 and NC3 showed the greatest difference in *ospC* allele composition (Table A.5) and are the most geographically distant from each other. Site K2 is in the coastal plains of Delaware with agricultural lands as the largest land cover type, while site NC3 represents Delaware's northern Piedmont habitat with impervious surfaces taking up the largest proportion of land cover (Coxe, 2012a, 2012b). We found *ospC* allele compositions at sites NC1 and NC2 to differ more than expected by chance given their close geographic distance. This difference is likely a result of the habitat composition of the sites. As previously mentioned, NC3 is mainly composed of impervious surfaces and occurs in a relatively urbanized area. NC1 consists of a significantly larger natural area.

We found mixed infections in 39.7% of the infected blacklegged ticks. Mixed infections should be more common in adult ticks relative to nymphal ticks as adults have one more feeding opportunity in which to become infected (but see Di et al., 2018). We could not test that prediction with our data because adults and nymphs were not distinguished in the sample of 31 infected ticks that included adults. While there may be negative interactions between *B. burgdorferi* strains with different *ospC* alleles in a mixed infection (Di et al. 2018), the relative frequency of *ospC* alleles in mixed

infections was positively correlated with the relative frequency of *ospC* alleles in single infections, suggesting that many associations of alleles in mixed infections may occur at random (Figure 1.2; Table A.6). We found mixed infections to be common as shown in other studies (Brisson and Dykhuizen, 2004; Di et al., 2018) and for the distribution of multiplicity of infection (i.e., number of alleles in a mixed infection) to be right-skewed with a mode of two alleles per mixed infection (Figure 1.2). Qiu et al. (2001) similarly found an average of 2.6 *ospC* alleles per infected tick and Di et al. (2018) also found *ospC* allele multiplicity to be right skewed. Some alleles have been shown to co-occur in mixed infections more frequently than expected by chance, suggesting a deterministic mechanism involved in allele associations (Di et al., 2022), perhaps mediated by the host immune system. While our dataset is not large enough to conduct a similar analysis, such patterns of allele co-occurrence deserve further investigation.

Here we quantified the prevalence of *B. burgdorferi* and the diversity of its *ospC* alleles in ticks at several sites across the three counties of Delaware. The prevalence of *B. burgdorferi* (Table 1.1) and the composition of *ospC* alleles varied across sites (Table 1.2), with human infectious alleles particularly prevalent. These patterns likely contribute to the high incidence of Lyme disease in Delaware and add to our understanding of the distribution of *B. burgdorferi ospC* allele diversity in the northeastern U.S. Future *ospC* studies should investigate the degree to which *ospC* allele frequencies change over time in particular locations (Qiu et al., 2002). Incorporating *ospC* allele identities into estimates of *B. burgdorferi* prevalence may increase the relevance of those estimates to human risk of Lyme disease and inform public health interventions to prevent Lyme disease cases.

Chapter 2

A SPECIALIZED *BORRELIA BURGDORFERI* OUTER SURFACE PROTEIN C (OSPC) ALLELE MAY MITIGATE LYME DISEASE RISK IN HUMANS

Introduction

Host specificity (i.e., the diversity of host species a pathogen infects) is an important pathogen phenotype linked to infectious disease emergence. Specialist pathogens, which infect a single host species or a few closely related host species, are less likely to become emerging infectious diseases (EIDs) than generalist pathogens, which are capable of infecting multiple host species (Woolhouse and Gowtage-Sequeria, 2005). Host specificity may vary even among closely related, cryptic pathogen species or lineages (e.g., Ellis and Bensch, 2018). Therefore, identifying the genotypes responsible for EIDs (Daszak et al., 2000) or potential future EIDs will require characterizing the genetic diversity of pathogens and the markers associated with host specificity (Llewellyn et al., 2009). Furthermore, specialists are hypothesized to be more effective at infecting their hosts than generalists – the “jack-of-all trades is a master of none hypothesis” (Medeiros et al., 2014), although evidence for the trade-off is mixed (Hellgren et al., 2009). Pathogen genotypes may exclude each other from mixed-genotypes or “coinfections” (Bremermann and Thieme, 1989; Wille et al., 2002), thereby affecting the overall pathogen presence in an area. Genotype exclusion may be direct (e.g., through resource competition) or indirect if antibodies to one pathogen genotype are effective at blocking infection by other pathogen genotypes, but not vice-versa (“asymmetric cross reactivity”; Nouri et al., 2021). Specialized pathogens that indirectly exclude generalists (Adams et al., 2021) may reduce the transmission rate of the generalists to other host species. In the case of human infections, specialists could lower human infection risk by reducing the overall

population sizes of generalists. Such a phenomenon, while theoretically possible, requires empirical support to prove that it exists in nature.

Lyme disease is an EID caused by the bacterium *Borrelia burgdorferi* (*sensu stricto*; hereafter *B. burgdorferi*) and is the most common vector-borne disease in the United States (Kugeler et al., 2021). *B. burgdorferi* is transmitted to vertebrate hosts, including humans, by blacklegged ticks (*Ixodes scapularis*) in the upper Midwest and northeastern United States and western blacklegged ticks (*Ixodes pacificus*) in the western United States (Movila et al., 2012). *B. burgdorferi* is maintained in a zoonotic life cycle between *Ixodes* ticks and vertebrate hosts. Common mammalian hosts for *B. burgdorferi* in the upper Midwest and northeastern U.S. include deer mice (*Peromyscus maniculatus*), white-footed mice (*P. leucopus*), eastern chipmunks (*Tamias striatus*), masked shrews (*Sorex cinereus*), Northern short-tailed shrews (*Blarina brevicauda*), and eastern meadow voles (*Microtus pennsylvanicus*; Donahue et al., 1986; Levine et al., 1984; Markowski et al., 1998; Ostfeld et al., 2014; Telford et al., 1990). The white-footed mouse is considered the main reservoir host (i.e., species infected by the pathogen that serves as a source of infection to other species) because they are natural hosts for the immature stages of *I. scapularis* and infected mice have a high probability of transmitting the bacterium to feeding blacklegged ticks (Brisson et al., 2008; Donahue et al., 1986; LoGiudice et al., 2003).

The most polymorphic gene in the *B. burgdorferi* genome is the outer surface protein C gene (*ospC*) (Baranton et al., 2001; Jauris-Heipke et al., 1993; Wang et al., 1999), and it is expressed when *B. burgdorferi* migrates from the tick's midgut to its salivary glands while feeding on a vertebrate host (Schwan et al., 1995; Schwan and Piesman, 2000). It is one of the first and most targeted antigens by the vertebrate

host's immune system (Grimm et al., 2004; Wilske et al., 1993) and is required to establish infection in vertebrate hosts (Earnhart et al., 2010; Grimm et al., 2004; Tilly et al., 2006). Infected hosts develop antibodies to *ospC* and therefore cannot be reinfected by the same *ospC* allele (Gilmore et al., 1996; Probert et al., 1997). Moreover, *ospC*-mediated asymmetric cross-reactivity has been shown in *B. afzelii*, a species in the Lyme disease group, and this cross reactivity is dependent on host species genotype (Nouri et al., 2021). The *ospC* gene shows high rates of apparent horizontal gene transfer and recombination (Walter et al., 2017), although this may occur more frequently in the upper midwestern than in the northeastern U.S. (Barbour and Travinsky, 2010; Brisson et al., 2010). Isolates of *B. burgdorferi* belonging to the same *ospC* “major allele” group (hereafter, “alleles”) can differ by up to 2% in their *ospC* gene sequences, but genetic distances between alleles are greater than 8% (Brisson et al., 2010; Wang et al., 1999). There are at least 28 distinct *ospC* alleles (Rudenko et al., 2013) which are generally named with single letters (i.e., alleles A, B, C, etc.).

While *B. burgdorferi* infects many vertebrate species, alleles vary in their host specificity, with some alleles infecting multiple host species and others restricted to one or a few host species (Brisson and Dykhuizen, 2004). Moreover, only a subset of *B. burgdorferi* *ospC* alleles (alleles A, B, K, and I) commonly cause disseminated infections in humans (i.e., infections that spread from the skin to the blood and cerebrospinal fluid; Dykhuizen et al., 2008; Seinost et al., 1999); these are known as human infectious alleles (hereafter, “HIAs”). Brisson and Dykhuizen (2004) looked at the distribution of *ospC* alleles in four small mammal species in Millbrook, NY and showed that different hosts share some alleles while other alleles are species-specific.

Among the hosts analyzed, eastern chipmunks harbored the most alleles (14), followed by the white-footed mouse (12), northern short-tailed shrew (11), and the eastern gray squirrel (8). Half of the 14 alleles identified infected all four hosts, but two alleles (alleles L and M) infected only two hosts each, and one allele (allele U) only infected eastern chipmunks. Three of the four host species were infected with all four HIAs, while the eastern grey squirrel carried three. Based on surveys of ticks in the environment, all four host species in the study would likely have been exposed to all 14 *ospC* alleles, suggesting that differences in allele host specificity were not an artefact of incomplete sampling and instead represented real differences in host specificity (Brisson and Dykhuizen, 2004).

Host specificity of *B. burgdorferi ospC* alleles has received limited attention since Brisson and Dykhuizen (2004), and it remains to be determined whether the results of their study hold for other locations in the range of *B. burgdorferi*. Furthermore, *ospC* alleles exhibit non-random patterns of association in ticks (Di et al., 2022), but little is known about their interactions in vertebrate hosts. Specifically, it is unclear whether interactions between *ospC* alleles of differing host specificities can have impacts on human infection risk, possibly through the exclusion of HIAs by specialists like allele U. To address these issues, we quantified *ospC* diversity and prevalence in small mammals and mesocarnivores from three sites in New Jersey. We tested 1) if any *ospC* alleles were more specialized or generalized than expected by chance, 2) if mammalian species hosted more or less diversity of *ospC* alleles than expected by chance, and 3) if specialized alleles might exclude other alleles from mixed allele infections. We also quantified the prevalence of *B. burgdorferi* among mammalian host species. These investigations allowed us to better understand the

relationship between host specificity and human infection risk in this widespread and common EID.

Methods

Sample collection

We trapped 272 individuals of 11 species of small mammals and mesocarnivores between May and September 2021 at three locations on the Rutgers University campus in New Brunswick, New Jersey using Sherman and Tomahawk live traps. The study sites were Goat Farm (40.48 latitude, -74.44 longitude), Rutgers' Garden (40.47, -74.42), and University Inn (40.48, -74.43). We collected a 2mm ear biopsy from individuals of 10 of the 11 species in the field using a sterile ear punch (Fisherbrand™ Animal Ear Punch). For Northern short-tailed shrews, we removed the tip of their tails using sterilized surgical scissors. We stored each biopsy in 200 μ L of RNAlater (Thermo Fisher Scientific, Waltham, MA, USA cat# AM7021) at 4°C for one to three days, followed by long-term storage in -20°C until DNA extraction. Trapping effort was similar across all sites, however trapping effort increased at University Inn for select species. All sampling was conducted under a scientific collection permit for game species from the New Jersey Division of Fish and Wildlife (Permit NO: SC 2021-01) and approval from the Rutgers University Institutional Animal Care and Use Committee (IACUC ID# AMEND202100256).

All three sites are located in an urbanized part of New Jersey. Natural vegetation at the sites was primarily deciduous forest dominated by oak trees (*Quercus* spp.) and huckleberry shrubs (*Gaylussacia* spp.). The University Inn site was unique in that it had almost no white-footed mice or eastern meadow voles despite more intense sampling efforts to trap small mammals at this site (Ferreira et al., in prep), providing

a natural experiment to test the dependency of allele diversity and composition on host species diversity and composition.

Molecular screening of mammal tissue for *B. burgdorferi* and *ospC* sequencing

To extract DNA, RNAlater was first removed from each sample, and the ear tissue was rinsed twice with 400 μL of 1X PBS. Next, five 2.8 mm stainless steel beads and 200 μL of 1X PBS were added to the sample tube. Tubes were transferred to a TissueLyser (Qiagen, Germantown, MD, USA), and ear samples were disrupted for two minutes at the maximum oscillation frequency (30 Hz). This step was repeated until the mixture was homogeneous. Half of the homogenate (100 μL) was stored in a -80°C freezer for future use, and DNA was extracted from the other half using a DNeasy Blood and Tissue kit (Qiagen, Germantown, MD, USA cat #69506) following the manufacturer's protocol.

We tested each sample in duplicate for *B. burgdorferi* using a qPCR protocol (Hojgaard et al., 2014). Individuals were considered positive for *B. burgdorferi* if at least one of the replicates tested positive. The qPCR protocol included one primer set and one probe to target *B. burgdorferi*. We combined 10 μL of TaqMan Fast Advanced Master Mix (Thermo Fisher Scientific, Waltham, MA, USA), 0.18 μl of a 10 μM solution of each primer (fliD-Fwd and fliD-Rev), 0.05 μl of a 10 μM solution of the probe (fliD), and 7.18 μl of ultrapure water in each reaction. Primer and probe sequences were taken from Hojgaard et al. (Hojgaard et al., 2014). Each well in the qPCR plate was brought up to 20 μl with the addition of 2 μl of DNA extract. The qPCR thermal profile protocol was 50°C for two minutes followed by two minutes at 95°C for the initial stage, followed by 40 cycles of 95°C for one second and 60°C for 20 seconds. The qPCR reactions were performed on an Applied Biosystems

QuantStudio 3 Real-Time PCR system (Thermo Fisher Scientific, Waltham, MA, USA). *B. burgdorferi* prevalence was determined by qPCR infection status.

The *ospC* gene of *B. burgdorferi* was amplified for all positive qPCR samples using a PCR protocol developed by Di et al. (2018). The 25 μ l PCR reaction included 12.5 μ l of master mix (DreamTaq Hot Start Master Mix 2X, Thermo Fisher Scientific, Waltham, MA, USA), 1.25 μ l of a 10 μ M solution of each primer (Oc-Fwd and Oc-Rev), 7 μ l of ultrapure water, and 3 μ l of DNA template. The PCR thermal profile protocol used one cycle of initial denaturation (2 min at 95°C), followed by 35 cycles of denaturation (30 s at 95°C), annealing (30 s at 52°C), and extension (1 min at 72°C), and a final extension at 72°C for 10 min. All PCRs and qPCRs included at least one negative control (ultrapure water) and two positive controls (DNA from a *B. burgdorferi*-infected tick and white-footed mouse). We visualized PCR amplification through gel electrophoresis with 1.5% agarose gels.

We used the nested PCR protocol of Bunikis et al. (2004) to amplify *ospC* in samples that did not amplify with the Di et al. (2018) protocol, but were infected according to qPCR. For the Bunikis et al. (2004) PCR, our first (i.e., outer) PCR was a 25 μ l reaction containing 12.5 μ l of master mix (DreamTaq Hot Start Master Mix 2X, Thermo Fisher Scientific, Waltham, MA, USA), 1.25 μ l of a 10 μ M solution of each primer (*ospC*_F and *ospC*_R), 7 μ l of ultrapure water, and 3 μ l of DNA extract. The thermal profile for the outer PCR reaction was the same as the single-round PCR of Di et al. (2018) described above. The nested (i.e., inner) 25 μ l reaction contained 12.5 μ l of master mix, 1.25 μ l of a 10 μ M solution of each primer (*ospC*_Fn and *ospC*_Rn), 8 μ l of ultrapure water, and 2 μ l of PCR amplicon from the first reaction. The PCR thermal profile for the nested reaction was one cycle of initial denaturation (2 min at

95°C), followed by 30 cycles of denaturation (30 s at 95°C), annealing (30 s at 52°C), and extension (1 min at 72°C), and a final extension at 72°C for 10 min. All PCRs included at least one negative control (ultrapure water; also, negative controls from the first reaction were run in the nested reaction and confirmed negative in the gel) and two positive controls and visualized as before.

We cleaned the PCR product of amplified samples using AMPure XP beads (Beckman Coulter, Horsham, PA, USA), and diluted to 0.2 ng/μl to build Illumina libraries with the Illumina Nextera XT DNA Library Prep kit (Illumina, San Diego, CA, USA) using the i7 and i5 unique dual index adapters from the Nextera XT Index kit (Illumina, San Diego, CA, USA). Libraries were cleaned with AMPure XP beads and were pooled in equimolar ratios and sent to the University of Delaware's Genotyping and Sequencing Center at the Delaware Biotechnology Institute (DBI) for sequencing. Libraries were sequenced on an Illumina Miseq using a 500-cycle nano sequencing kit (Illumina, San Diego, CA, USA). The sequencing was paired-end with 251 cycles per read.

Bioinformatic analysis

OspC alleles were identified from the short-read sequences following the protocol described by Di et al. (2018) with modifications as presented in Shifflett et al. (2023). Briefly, reads were trimmed, checked for quality control and mapped to the FASTA *ospC* reference file in Di et al. (2018). Depth of coverage was then calculated at each base pair position across the reference sequences. Alleles were considered present if all nucleotide positions were sequenced to a depth of coverage of at least 5X between nucleotide positions 200 and 550 of the reference (the more variable region of the gene). We also quantified the proportion of unmapped reads in each sample using

SAMtools stats (Danecek et al., 2021). One sample (R3) had reads incompletely mapping to several alleles suggesting a possible novel allele formed through recombination however, we were unable to recover a sequence using Sanger sequencing. An additional five samples showed unrecovered alleles due to sequence reads not mapping to the reference file. Samples with unresolved alleles were not included in the *ospC* analyses.

Statistical analyses

We removed recaptures from all *ospC* and prevalence analyses. We also removed the single eastern cottontail (*Sylvilagus floridanus*) sample from all *ospC* analyses. While the majority of *Peromyscus* mice were likely white-footed mice (*P. leucopus*), we called them *Peromyscus* spp. to account for potential misidentification between the white-footed mouse and the deer mouse (*P. maniculatus*). We tested whether *B. burgdorferi* prevalence varied across sites using a Fisher's exact test and across species using a Pearson's chi-squared test in R v. 4.2.1 (R Core Team, 2020).

We calculated the host specificity of *ospC* alleles as 1) the number and 2) the Gini-Simpson diversity index of host species each allele infected. We also quantified the diversity of *ospC* alleles found in hosts by calculating 1) the number and 2) the Gini-Simpson diversity index of *ospC* alleles that infected each host species. We calculated the Gini-Simpson index using the diversity function (with argument "index = simpson") in the R package *vegan* v.2.6-2 (Oksanen et al., 2022). To test if allele host specificity and allele diversity per host differed from a random expectation, we randomly shuffled *ospC* alleles among infected individuals of all host species 9,999 times. For each randomization, we calculated a randomized allele host specificity and allele diversity per host species and used those values to create a random distribution

to compare with each measured value. We then calculated two-tailed P values for each of the measured values to determine if they were more extreme than expected by chance.

We classified individuals with mixed infections as those having more than one sequenced *ospC* allele. We used the `cor.test` function in R v. 4.2.1 (R Core Team, 2020) to correlate the frequency of *ospC* alleles found in mixed infections with allele frequencies found in single infections. A positive correlation would suggest that alleles found in mixed infections are due to chance (Shifflett et al., 2023). We also calculated the frequency of mixed infections across host species.

We used the Chao1 estimator (Chao et al., 2009) to determine the expected number of *ospC* alleles across all sampling sites and the number of alleles expected to infect each host species. We used the `estimateR` function in the R package `vegan` v.2.6-2 (Oksanen et al., 2022) based on allele frequencies to calculate the Chao1 estimator.

Results

We tested 272 individual hosts across 11 species for *B. burgdorferi* and found 81 infections by qPCR; after removing recaptures (23 individuals) there were 68 qPCR infections (27.3%). Three host species were not infected despite substantial sampling (Virginia opossums, *Didelphis virginiana*, n = 11; groundhogs, *Marmota monax*, n = 8; racoons, *Procyon lotor*, n = 32; Figure 2.1; the lone eastern cottontail was also not infected) and removal of nymphal blacklegged ticks potentially infected with *B. burgdorferi* (Ferreira et al., *in prep*). *B. burgdorferi* prevalence varied across host species with sample sizes greater than five ($\chi^2 = 46.99$, df = 8, $P < 0.005$) and when removing the three uninfected species ($\chi^2 = 20.1$, df = 5, $P < 0.005$) and reached

the highest levels on *Peromyscus* spp. (52%) and on eastern chipmunk (41%). Among recaptured individuals, those infected with *B. burgdorferi* were also infected at subsequent captures. Furthermore, three uninfected individuals were infected upon recapture a month later; 10 individuals remained uninfected between captures.

B. burgdorferi prevalence among all individuals, irrespective of host species, varied among the three sites (Fisher's exact test $P = 0.047$). Prevalence was highest at the Goat Farm site (31%), which also had the highest sample size ($n = 136$; Table B.1). University Inn, the site with almost no *Peromyscus* spp. or eastern meadow voles (Figure 2.1), had the lowest sample size and prevalence ($n = 41$, 12%). Two host species (eastern chipmunks and *Peromyscus* spp.) had sufficient sample sizes at more than one site to compare *B. burgdorferi* prevalence. However, prevalence did not vary among sites for either species (eastern chipmunks, $n = 3$ sites in the comparison, Fisher's exact test $P = 0.901$; *Peromyscus* spp., $n = 2$, Fisher's exact test $P = 0.752$).

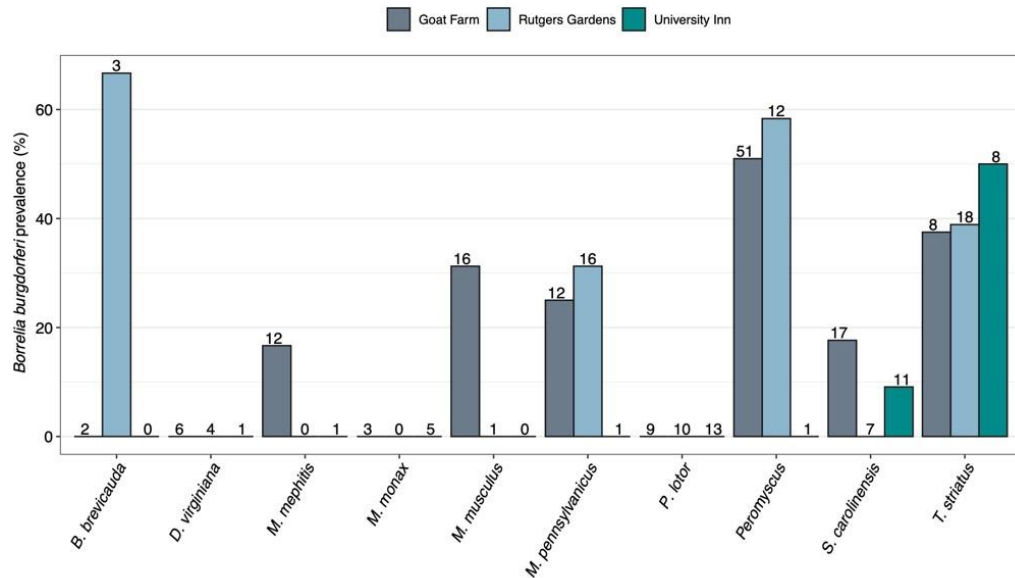


Figure 2.1 *B. burgdorferi* prevalence (as determined by qPCR) across sites (Goat Farm, Rutgers' Gardens, University Inn) and host species (*B. brevicauda*, northern short-tailed shrew; *D. virginiana*, Virginia opossum; *M. mephitis*, striped skunk; *M. monax*, groundhog; *M. musculus*, house mouse; *M. pennsylvanicus*, eastern meadow vole; *P. lotor*, raccoon; *Peromyscus*, white-footed/deer mice; *S. carolinensis*, eastern gray squirrel; *T. striatus*, eastern chipmunk). The number of captured individuals of each species at each site is shown above its corresponding bar.

ospC allele diversity

We sequenced *ospC* in 70 infected individuals (13 recaptures) and identified 14 alleles across six species and three sites (Figure 2.2). HIAs were found at the greatest relative frequencies (Figure 2.2, Table B.1), with allele B occurring in 25 individuals (27.8% of all infected individuals), K in 17 (18.9%), and A in 15 (16.7%). Alleles B, K, and A (all HIAs) made up the majority of alleles at Rutgers Garden (62.5% of all alleles) and Goat Farm (68.5%; Table B.1). University Inn had the

lowest allele diversity of the sites (only alleles U, T, and I were found), with one HIA found. University Inn is also the site that had only one individual of *Peromyscus* spp. sampled. We estimated the total number of *ospC* alleles across the study sites using the Chao1 estimator as 14.6 ± 1.18 s.e., suggesting we recovered almost all (14/14.6) of the allele diversity.

While eastern chipmunks had half the sample size (14 infected individuals) of *Peromyscus* spp. (28 infected individuals), they were infected with more unique *ospC* alleles (12) than *Peromyscus* spp. (8). The eastern meadow vole had six infected individuals with five unique *ospC* alleles. Eastern gray squirrels showed the lowest number of unique *ospC* alleles (1), but this was likely due to only finding two infected individuals. The striped skunk had three *ospC* alleles (B, A, and C) in two infected individuals.

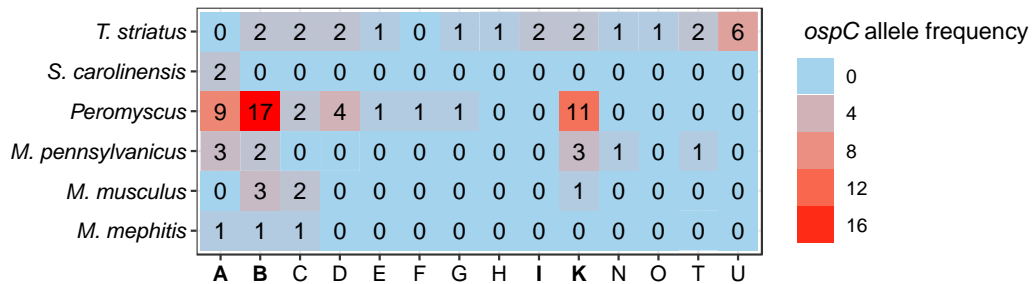


Figure 2.2 Heatmap showing *ospC* allele frequencies across host species (*T. striatus*, eastern chipmunk; *S. carolinensis*, eastern gray squirrel; *Peromyscus*, white-footed/deer mice; *M. pennsylvanicus*, eastern meadow vole; *M. musculus*, house mouse; *M. mephitis*, striped skunk). Higher allele frequencies are redder and lower frequencies are bluer. Human infectious alleles (A, B, I, and K) are bolded.

Host specificity of *ospC* alleles and diversity of *ospC* alleles in host species

Host specificity did not differ from a random expectation for any allele except U, which was more specialized than expected by chance. Allele U only infected eastern chipmunks (Figure 2.2) but was expected to infect about three host species by chance (Table 2.1). The frequencies of alleles F, H, and O were too low (one of each) for those alleles to be included in the analysis. We repeated this analysis using the Gini-Simpson index instead of number of host species per allele and found the same results (Table B.2).

While eastern chipmunks appeared to be the only host species infected by a specialized *ospC* allele (U), they also were infected with more *ospC* alleles than expected by chance (12 alleles observed, *ca.* 9 expected; Table 2.2). Furthermore, the species with the highest prevalence (*Peromyscus* spp.) was infected by fewer alleles than expected by chance (8 observed, *ca.* 11 expected). The number of alleles infecting the other host species did not differ from a random expectation (Table 2.2). We repeated this analysis using the Gini-Simpson index instead of number of *ospC* alleles per host species with similar results (Table B.3).

Table 2.1 Observed and expected host specificity of *ospC* alleles. Host specificity (“Observed”) was quantified as the number of host species an *ospC* allele infected. The number of host species an allele was expected to infect is the mean host specificity across 9,999 randomizations (“Mean expected”); the standard deviation of the mean is also presented (“SD expected”) and the *P* value indicating whether the observed host specificity differed from the random expectation. Human infectious alleles (A, B, I, and K) are italicized.

	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>	<i>G</i>	<i>I</i>	<i>K</i>	<i>N</i>	<i>T</i>	<i>U</i>
Observed	4	5	4	2	2	2	1	4	2	2	1
Mean expected	4.26	4.94	3.22	3.02	1.67	1.67	1.66	4.42	1.66	2.13	3.02
SD expected	0.85	0.77	0.83	0.81	0.47	0.47	0.47	0.84	0.47	0.64	0.81
P value	1	1	0.70	0.52	1	1	0.67	1	1	1	0.03

Table 2.2 Observed and expected *ospC* allele diversity in infected vertebrate host species (*M. mephitis*, striped skunk; *M. pennsylvanicus*, eastern meadow vole; *M. musculus*, house mouse; *Peromyscus*, white-footed/deer mice; *S. carolinensis*, eastern gray squirrel; *T. striatus*, eastern chipmunk). Allele diversity (“Observed”) was quantified as the number of unique *ospC* alleles in each host species. The expected number of alleles was found by averaging the expected values across all 9,999 randomization trials (“Mean expected”); the standard deviation of the mean is also presented (“SD expected”) and the P value indicating whether the *ospC* allele diversity in a host species differed from a random expectation.

	<i>M. mephitis</i>	<i>M. pennsylvanicus</i>	<i>M. musculus</i>	<i>Peromyscus</i> spp.	<i>S. carolinensis</i>	<i>T. striatus</i>
Observed	3	5	3	8	1	12
Mean expected	2.59	5.82	4.26	11.44	1.85	8.68
SD expected	0.54	1.14	0.91	1.19	0.36	1.32
P value	1	0.78	0.39	0.01	0.30	0.03

Mixed infections

After removing recaptures, leaving 57 infected hosts with recovered *ospC* alleles, 18 (32%) had mixed infections (i.e., more than one allele present). The mean number of alleles present in mixed infections (also referred to as “allelic multiplicity”) was 2.83 ± 1.29 s.d. with two individuals (a *Peromyscus* spp. and an eastern chipmunk) each having six alleles (Figure 2.3). Most mixed infections occurred at Goat Farm where 12 individuals had more than one allele and infections with two alleles occurred most frequently. Similarly, Rutgers Garden had six individuals with mixed infections ranging from two to six alleles per infection. University Inn did not have mixed infections. Overall, *Peromyscus* spp. had the greatest number of mixed infections (11) followed by eastern chipmunks (3) and eastern meadow voles (2), but there was not a significant difference in mixed infection frequency among species with a sample size of at least five infections (Fisher’s exact test $P = 0.659$). We did not find any mixed infections with more than two alleles in striped skunks or in house mice (Figure 2.3). The mean number of *ospC* alleles within mixed infections was greatest in eastern chipmunks (4 ± 1 s.e.), followed by eastern meadow voles (3 ± 1 s.e.), and *Peromyscus* spp. (2.64 ± 0.36 s.e.). The relative frequency of *ospC* alleles in single infections was positively correlated with the relative frequency of alleles in mixed infections ($r = 0.708$, $P = 0.005$; Figure B.1). However, allele U occurred in 15% of single infections but did not occur in any mixed infections and appeared as an outlier in the analysis (Figure B.1). Thus, while eastern chipmunks had the highest average number of alleles per mixed infection compared to the other hosts, they had far fewer mixed infections than *Peromyscus* spp., because allele U, the allele that is specialized on eastern chipmunks and is relatively common, did not occur in mixed infections (Figure B.1).

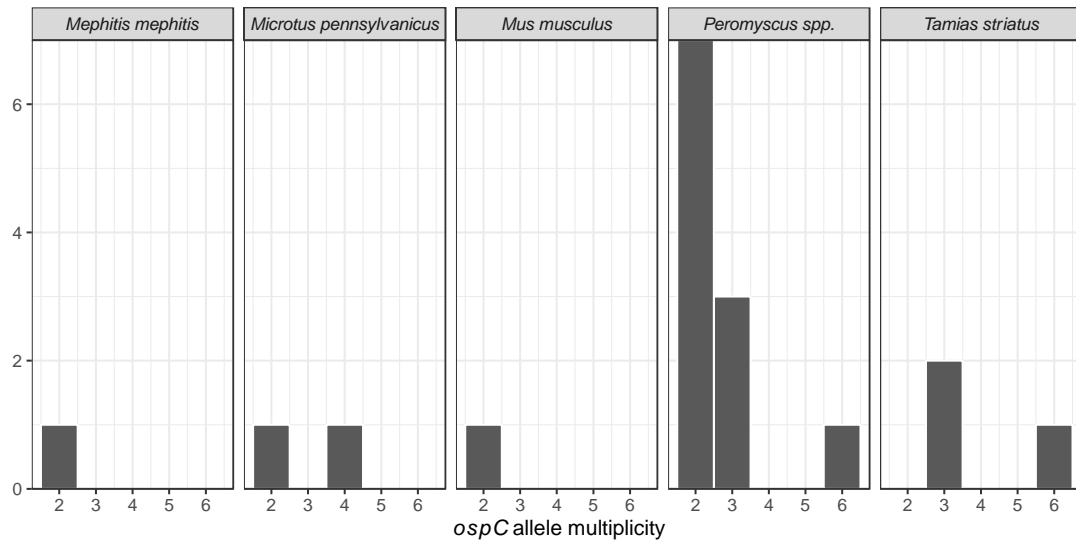


Figure 2.3 Frequency of mixed allele infections by host species (*M. mephitis*, striped skunk; *M. pennsylvanicus*, eastern meadow vole; *M. musculus*, house mouse; *Peromyscus* spp., white-footed/deer mice; *T. striatus*, eastern chipmunk). The number of *ospC* alleles within a mixed infection is represented by the x-axis (“*ospC* allele multiplicity”) and the frequencies are shown on the y-axis.

Recapture infection dynamics

Two recaptures maintained the same *ospC* allele between captures (Figure 2.4). For example, one *Peromyscus* spp. was infected with allele A both in August and September. The only recaptured and infected eastern chipmunk was infected with allele U in July and September. However, several recaptured individuals had dynamic *ospC* allele compositions (Figure 2.4). Some *ospC* alleles may also have been undetected at some points. For example, one *Peromyscus* spp. was captured four times and was infected by allele A at the first, third, and fourth captures, but not at the second capture

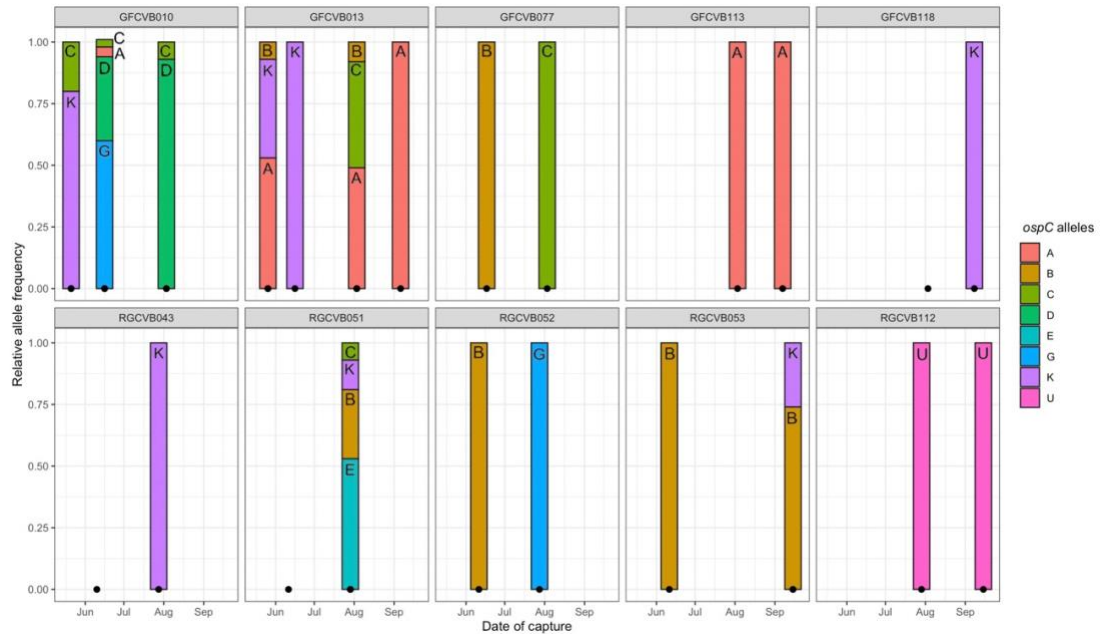


Figure 2.4 *OspC* allele compositions in *B. burgdorferi* infected small mammals captured multiple times (June-September 2021). Only recaptured individuals infected with *B. burgdorferi* at least once are included. All but one individual (RGCVB112) are *Peromyscus* spp. (white-footed/deer mice); RGCVB112 is an eastern chipmunk. Relative allele frequencies are shown on the y-axis (allele identity is labeled and has a unique color) and capture date on the x-axis; individuals are represented by separate panels in the figure. A point is shown on the x axis for each of the capture dates, a point without a bar indicates the individual was not infected upon capture.

Discussion

The prevalence of *B. burgdorferi* varied among host species (Figure 2.1), and we identified 14 *ospC* alleles (Figure 2.2), nearly the total number estimated to occur in the sample sites. While the host specificity of most alleles did not differ from a random expectation, allele U (a non-HIA) appeared specialized on eastern chipmunks, as shown previously (Brisson and Dykhuizen, 2004). Eastern chipmunks were infected with more alleles than expected by chance, while *Peromyscus* spp., which were most

frequently infected by HIAs, were infected by fewer alleles than expected by chance (Table 2.2). Of all infections with recovered *ospC* alleles, *ca.* 32% were mixed allele infections, and the frequency of mixed infections was positively correlated with the frequency of single infections among host species (Figure B.1), consistent with the random assembly of mixed infections. However, despite the high diversity of *ospC* alleles infecting eastern chipmunks, the specialized allele U only occurred in single infections, suggesting it excludes other alleles from mixed infections, perhaps through indirect mechanisms involving the host immune system. If this is true, allele U may reduce the overall frequency of HIAs in areas that have eastern chipmunks.

Competitive exclusion between coinfecting parasites has been seen in multiple disease systems (Chappell, 1969; Devevey et al., 2015; Massey et al., 2004; Mideo, 2009; Putaporntip et al., 2010; Read and Taylor, 2001). Within the Lyme disease system, Devevey et al. (2015) investigated experimental coinfections between *ospC* alleles A, K, and N in mice and found the *ospC* allele causing the initial infection to have the greatest fitness, regardless of identity, suggesting a “priority effect”. Furthermore, they found no evidence of cross-reactivity of *ospC* specific antibodies (Devevey et al., 2015) and hypothesized that resource competition may underly the negative interactions between *ospC* alleles. However, Nouri et al. (2021) found different results in a similar experimental system with *ospC* from *B. afzelii*. Those authors found that the degree of *ospC* antibody cross reactivity was stronger in one mouse strain than in others. We hypothesize that allele U is able to exclude other *ospC* alleles in eastern chipmunks, although it is unclear exactly how this might happen. Devevey et al. (2015) hypothesized that competitive exclusion of *ospC* alleles by the priority effect mechanism is most likely to occur in areas with low prevalence of *B.*

burgdorferi, for example in areas where the pathogen has only recently begun to occur. The sites we studied occur in an established *B. burgdorferi* area and prevalence and *ospC* allele diversity was generally high. We therefore suspect that indirect host immune system effects underlie the distribution of allele U, perhaps through antibody cross reactivity in eastern chipmunks. Experimental investigations are needed to better understand how allele U is able to prevent mixed infections and specialize on eastern chipmunks.

Both our work and the work of Brisson and Dykhuizen (2004) show that *ospC* allele U specializes on eastern chipmunks. The fact that allele U does not occur in mixed infections and reaches a higher prevalence on its host species than any other allele (Figure 2.2), is consistent with the hypothesis that specialists are more effective than generalists. Such a result has been shown elsewhere (Leggett et al., 2013; Malpica et al., 2006; Medeiros et al., 2014; Regoes et al., 2000; Straub et al., 2011), but see Hellgren et al. (2009) and Bedhomme et al. (2012) for counter examples. Importantly, *ospC* allele diversity may not correlate perfectly with *B. burgdorferi* strain diversity since *ospC* alleles can be horizontally transferred (Walter et al., 2017). Therefore, whole genome sequencing of infections with allele U will be required to decipher the mechanisms underlying the specialization of allele U.

The exclusion of HIAs in eastern chipmunks by allele U may be an important aspect of the transmission dynamics of the Lyme disease system. HIAs commonly occur in mixed infections, with eastern chipmunks having on average more alleles in mixed infections than any other vertebrate host (also seen in Brisson and Dykhuizen (2004)), supporting the importance of this species in maintaining *ospC* allele diversity. However, without the presence of allele U, eastern chipmunks would likely host more

HIA infections and consequently lead to more HIA infected ticks in the environment, thereby increasing human infection risk. While the degree to which allele U could mitigate human infection risk is unclear, primarily because HIAs are maintained at high levels by *Peromyscus* spp., the reduction may not be negligible and will depend on relative mammalian host species population sizes. The possibility that a specialized *ospC* allele could reduce human infection risk by excluding HIAs should be further investigated to inform future novel biocontrol strategies to minimize Lyme disease risk and understand how host diversity may mitigate human infection risk (Ostfeld and Keesing, 2000).

Peromyscus spp. had the highest prevalence (Figure 2.1) and were most frequently infected by HIAs (Figure 2.2), consistent with other studies (Brisson and Dykhuizen, 2004; Hanincová et al., 2006; Oliver et al., 2006) and with the hypothesis that they are major drivers of Lyme disease risk in humans (Donahue et al., 1986; Levine et al., 1984). Between their high prevalence of infection (52% in this study) and high population sizes (Myers et al., 2009), especially in human-modified landscapes (Cummings and Vessey, 1994; Myers et al., 2009; Nupp and Swihart, 1998), *Peromyscus* spp. appear to be important for maintaining *B. burgdorferi* infections that can be transmitted to feeding ticks. The importance of *Peromyscus* spp. to the maintenance of HIAs in the environment is highlighted by an unplanned natural experiment in our study design. The University Inn site had almost no *Peromyscus* spp. Perhaps as a consequence, University Inn also only had a single infection with a HIA in an eastern chipmunk.

We found *ospC* allele composition to be dynamic within *Peromyscus* spp. across time. Our results are similar to Swanson and Norris (2008) who reported

variation in *ospC* allele composition over time in *P. leucopus* in Maryland. Our results suggest that *Peromyscus* spp. may clear *B. burgdorferi* infections of certain alleles over short time periods (*ca.* a month).

Eastern gray squirrels were infrequently infected by *B. burgdorferi* (4 infections in 35 individuals sampled). This is consistent with other studies that have shown eastern gray squirrels to have lower prevalence than other small mammals (Brisson and Dykhuizen, 2004; Hanincová et al., 2006). Three species with substantial sampling (raccoons, Virginia opossums, and groundhogs) were not infected with *B. burgdorferi* (Figure 2.1). Studies that examined engorged larval ticks removed from the animals have proposed that raccoons and Virginia opossums are infected with *B. burgdorferi* (Hanincová et al., 2006). However, previous work screening raccoon ear tissue for *B. burgdorferi* also failed to find infected individuals (Slajchert et al., 1997). *B. burgdorferi* infection intensity may be too low for accurate PCR amplification in these hosts and may only be detectable by screening engorged ticks; future work is needed to determine the infection intensity levels that correspond to transmission by ticks and PCR detection. Alternatively, differences in prevalence in these hosts may result from true spatial variation in prevalence.

Our study demonstrates that a specialized pathogen genotype can reduce the frequency of more generalized, co-occurring pathogen genotypes in nature through some means of exclusion, plausibly indirect host immune responses. In the case of Lyme disease, the presence of allele U and its eastern chipmunk host may reduce the overall frequency of HIAs and thereby reduce human infection risk. Understanding the ecology of pathogen interactions in the wild may lead to novel avenues of research for mitigating the effects of Lyme disease and other EIDs.

REFERENCES

- Adalsteinsson, S.A., Shriver, W.G., Hojgaard, A., Bowman, J.L., Brisson, D., D'Amico, V., Buler, J.J., 2018. Multiflora rose invasion amplifies prevalence of Lyme disease pathogen, but not necessarily Lyme disease risk. *Parasites Vectors* 11, 54. <https://doi.org/10.1186/s13071-018-2623-0>
- Adams, B., Walter, K.S., Diuk-Wasser, M.A., 2021. Host specialization, immune cross-reaction and the composition of communities of co-circulating *Borrelia* strains. *Bull Math Biol* 83, 66. <https://doi.org/10.1007/s11538-021-00896-2>
- Adelson, M.E., Rao, R.-V.S., Tilton, R.C., Cabets, K., Eskow, E., Fein, L., Occi, J.L., Mordechai, E., 2004. Prevalence of *Borrelia burgdorferi*, *Bartonella* spp., *Babesia microti*, and *Anaplasma phagocytophila* in *Ixodes scapularis* ticks collected in northern New Jersey. *J Clin Microbiol* 42, 2799–2801. <https://doi.org/10.1128/JCM.42.6.2799-2801.2004>
- Aguilera-Arreola, Ma.G., Hernandez-Rodriguez, C., Zuniga, G., Figueras, M.J., Castro-Escarpulli, G., 2005. *Aeromonas hydrophila* clinical and environmental ecotypes as revealed by genetic diversity and virulence genes. *FEMS Microbiology Letters* 242, 231–240. <https://doi.org/10.1016/j.femsle.2004.11.011>
- Alghaferi, M.Y., Anderson, J.M., Park, J., Auwaerter, P.G., Aucott, J.N., Norris, D.E., Dumler, J.S., 2005. *Borrelia burgdorferi ospC* heterogeneity among human and murine isolates from a defined region of Northern Maryland and Southern Pennsylvania: Lack of correlation with invasive and noninvasive genotypes. *J Clin Microbiol* 43, 1879–1884. <https://doi.org/10.1128/JCM.43.4.1879-1884.2005>
- Anderson, J.M., Norris, D.E., 2006. Genetic diversity of *Borrelia burgdorferi* sensu stricto in *Peromyscus leucopus*, the primary reservoir of Lyme disease in a region of endemicity in Southern Maryland. *Appl Environ Microbiol* 72, 5331–5341. <https://doi.org/10.1128/AEM.00014-06>
- Andrews, S., 2010. FastQC: A quality control tool for high throughput sequence data [Online]. Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
- 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, 149–156. [https://doi.org/10.1016/S0923-2508\(01\)01186-X](https://doi.org/10.1016/S0923-2508(01)01186-X)
- Barbour, A.G., 1998. Fall and rise of Lyme disease and other *Ixodes* tick-borne infections in North America and Europe. *British Medical Bulletin* 54, 647–658. <https://doi.org/10.1093/oxfordjournals.bmb.a011717>
- Barbour, A.G., Fish, D., 1993. The biological and social phenomenon of Lyme disease. *Science* 260, 1610–1616. <https://doi.org/10.1126/science.8503006>

- Barbour, A.G., Travinsky, B., 2010. Evolution and distribution of the *ospC* gene, a transferable serotype determinant of *Borrelia burgdorferi*. *mBio* 1. <https://doi.org/10.1128/mBio.00153-10>
- Bedhomme, S., Lafforgue, G., Elena, S.F., 2012. Multihost experimental evolution of a plant RNA virus reveals local adaptation and host-specific mutations. *Molecular Biology and Evolution* 29, 1481–1492. <https://doi.org/10.1093/molbev/msr314>
- Bensch, S., Inumaru, M., Sato, Y., Lee Cruz, L., Cunningham, A.A., Goodman, S.J., Levin, I.I., Parker, P.G., Casanueva, P., Hernández, M., Moreno-Rueda, G., Rojo, M., 2021. Contaminations contaminate common databases. *Mol Ecol Resour* 21, 355–362. <https://doi.org/10.1111/1755-0998.13272>
- Bremermann, H.J., Thieme, H.R., 1989. A competitive exclusion principle for pathogen virulence. *J. Math. Biology* 27, 179–190. <https://doi.org/10.1007/BF00276102>
- Brisson, D., Dykhuizen, D.E., 2004. *OspC* diversity in *Borrelia burgdorferi*: Different hosts are different niches. *Genetics* 168, 713–722. <https://doi.org/10.1534/genetics.104.028738>
- Brisson, D., Dykhuizen, D.E., Ostfeld, R.S., 2008. Conspicuous impacts of inconspicuous hosts on the Lyme disease epidemic. *Proceedings of the Royal Society B: Biological Sciences* 275, 227–235.
- Brisson, D., Vandermause, M.F., Meece, J.K., Reed, K.D., Dykhuizen, D.E., 2010. Evolution of Northeastern and Midwestern *Borrelia burgdorferi*, United States. *Emerg. Infect. Dis.* 16, 911–917. <https://doi.org/10.3201/eid1606.090329>
- 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, 1741–1755. <https://doi.org/10.1099/mic.0.26944-0>
- Cardenas-de la Garza, J.A., De la Cruz-Valadez, E., Ocampo-Candiani, J., Welsh, O., 2019. Clinical spectrum of Lyme disease. *Eur J Clin Microbiol Infect Dis* 38, 201–208. <https://doi.org/10.1007/s10096-018-3417-1>
- Chao, A., Colwell, R.K., Lin, C.-W., Gotelli, N.J., 2009. Sufficient sampling for asymptotic minimum species richness estimators. *Ecology* 90, 1125–1133. <https://doi.org/10.1890/07-2147.1>
- Chappell, L.H., 1969. Competitive exclusion between two intestinal parasites of the three-spined stickleback, *Gasterosteus aculeatus* L. . *The Journal of Parasitology* 55, 775. <https://doi.org/10.2307/3277217>
- Comstedt, P., Bergström, S., Olsen, B., Garpmo, U., Marjavaara, L., Mejlom, H., Barbour, A.G., Bunikis, J., 2006. Migratory Passerine Birds as Reservoirs of Lyme Borreliosis in Europe. *Emerg. Infect. Dis.* 12, 1087–1102. <https://doi.org/10.3201/eid1207.060127>
- 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. *J Clin Microbiol* 41, 1569–1573. <https://doi.org/10.1128/JCM.41.4.1569-1573.2003>
- Coxe, R., 2012a. Historical Analysis and Map of Vegetation Communities, Land Covers, and Habitats of Alapocas Run State Park New Castle County, Delaware. Department of Natural Resources and Environmental Control, Delaware Division of Fish and Wildlife.
- Coxe, R., 2012b. Historical Analysis and Map of Vegetation Communities, Land Covers, and Habitats of Killens Pond State Park Kent County, Delaware. Department of Natural Resources and Environmental Control, Delaware Division of Fish and Wildlife.
- Cummings, J.R., Vessey, S.H., 1994. Agricultural influences on movement patterns of white-footed mice (*Peromyscus leucopus*). *American Midland Naturalist* 132, 209. <https://doi.org/10.2307/2426575>
- Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., Li, H., 2021. Twelve years of SAMtools and BCFtools. *GigaScience* 10, giab008. <https://doi.org/10.1093/gigascience/giab008>
- Daszak, P., Cunningham, A.A., Hyatt, A.D., 2000. Emerging Infectious Diseases of Wildlife-- Threats to Biodiversity and Human Health. *Science* 287, 443–449. <https://doi.org/10.1126/science.287.5452.443>
- Devevey, G., Dang, T., Graves, C.J., Murray, S., Brisson, D., 2015. First arrived takes all: inhibitory priority effects dominate competition between co-infecting *Borrelia burgdorferi* strains. *BMC Microbiol* 15, 61. <https://doi.org/10.1186/s12866-015-0381-0>
- Di, L., Akther, S., Bezrucenkovas, E., Ivanova, L., Sulkow, B., Wu, B., Mneimneh, S., Gomes-Solecki, M., Qiu, W.-G., 2022. Maximum antigen diversification in a Lyme bacterial population and evolutionary strategies to overcome pathogen diversity. *ISME J* 16, 447–464. <https://doi.org/10.1038/s41396-021-01089-4>
- Di, L., Wan, Z., Akther, S., Ying, C., Larracuente, A., Li, L., Di, C., Nunez, R., Cucura, D.M., Goddard, N.L., Krampis, K., Qiu, W.-G., 2018. Genotyping and quantifying Lyme pathogen strains by deep sequencing of the outer surface protein C (*ospC*) locus. *Journal of Clinical Microbiology* 56, e00940-18, [/jcm/56/11/e00940-18.atom](https://doi.org/10.1128/JCM.00940-18). <https://doi.org/10.1128/JCM.00940-18>
- Donahue, J., Piesman, J., Spielman, A., 1986. Reservoir competence of White-footed mice for Lyme disease spirochetes. *The American Journal of Tropical Medicine and Hygiene* 36, 92–96. <https://doi.org/10.4269/ajtmh.1987.36.92>
- Dykhuisen, D.E., Brisson, D., Sandigursky, S., Wormser, G.P., Nowakowski, J., Nadelman, R.B., Schwartz, I., 2008. Short report: The propensity of different *Borrelia burgdorferi sensu stricto* genotypes to cause disseminated infections in humans. *American Journal of Tropical Medicine and Hygiene* 78, 806–810. <https://doi.org/10.4269/ajtmh.2008.78.806>
- Earnhart, C.G., LeBlanc, D.V., Alix, K.E., Desrosiers, D.C., Radolf, J.D., Marconi, R.T., 2010. Identification of residues within ligand-binding domain 1 (LBD1)

- of the *Borrelia burgdorferi* OspC protein required for function in the mammalian environment. *Molecular Microbiology* 76, 393–408.
<https://doi.org/10.1111/j.1365-2958.2010.07103.x>
- Ellis, V.A., Bensch, S., 2018. Host specificity of avian haemosporidian parasites is unrelated among sister lineages but shows phylogenetic signal across larger clades. *International Journal for Parasitology* 48, 897–902.
<https://doi.org/10.1016/j.ijpara.2018.05.005>
- Ferreira, F.C., González, J., Milholland, M., Tung, G., Fonseca, D.M., 2023. Ticks (Acari: Ixodida) on synanthropic small and medium-sized mammals in areas of the northeastern United States infested with the Asian longhorned tick, *Haemaphysalis longicornis*. in preparation.
- Gilmore, R.D., Kappel, K.J., Dolan, M.C., Burkot, T.R., Johnson, B.J., 1996. Outer surface protein C (OspC), but not P39, is a protective immunogen against a tick-transmitted *Borrelia burgdorferi* challenge: evidence for a conformational protective epitope in OspC. *Infect Immun* 64, 2234–2239.
<https://doi.org/10.1128/iai.64.6.2234-2239.1996>
- Ginsberg, H.S., Buckley, P.A., Balmforth, M.G., Zhioua, E., Mitra, S., Buckley, F.G., 2005. Reservoir Competence of Native North American Birds for the Lyme Disease Spirochete, *Borrelia burgdorferi*. *JOURNAL OF MEDICAL ENTOMOLOGY* 42, 5.
- Grimm, D., Tilly, K., Byram, R., Stewart, P.E., Krum, J.G., Bueschel, D.M., Schwan, T.G., Policastro, P.F., Elias, A.F., Rosa, P.A., 2004. Outer-surface protein C of the Lyme disease spirochete: A protein induced in ticks for infection of mammals. *Proceedings of the National Academy of Sciences* 101, 3142–3147.
<https://doi.org/10.1073/pnas.0306845101>
- Hanincová, K., Kurtenbach, K., Diuk-Wasser, M., Brei, B., Fish, D., 2006. Epidemic spread of Lyme Borreliosis, Northeastern United States. *Emerg. Infect. Dis.* 12, 604–611. <https://doi.org/10.3201/eid1204.051016>
- Hellgren, O., Pérez-Tris, J., Bensch, S., 2009. A jack-of-all-trades and still a master of some: prevalence and host range in avian malaria and related blood parasites. *Ecology* 90, 2840–2849. <https://doi.org/10.1890/08-1059.1>
- Hojgaard, A., Lukacik, G., Piesman, J., 2014. Detection of *Borrelia burgdorferi*, *Anaplasma phagocytophilum* and *Babesia microti*, with two different multiplex PCR assays. *Ticks and Tick-borne Diseases* 5, 349–351.
<https://doi.org/10.1016/j.ttbdis.2013.12.001>
- Humphrey, P.T., Caporale, D.A., Brisson, D., 2010. Uncoordinated phylogeography of *Borrelia burgdorferi* and its tick vector, *Ixodes scapularis*: vector-pathogen postglacial phylogeography. *Evolution* 64, 2653–2663.
<https://doi.org/10.1111/j.1558-5646.2010.01001.x>
- Huyse, T., Poulin, R., Théron, A., 2005. Speciation in parasites: a population genetics approach. *Trends in Parasitology* 21, 469–475.
<https://doi.org/10.1016/j.pt.2005.08.009>

- Jauris-Heipke, S., Fuchs, R., Preac-Mursic, V., Schwab, E., Soutschek, E., Will, G., Wilske, B., 1993. Genetic heterogeneity of the genes coding for the outer surface protein C (*ospC*) and the flagellin of *Borrelia burgdorferi*. *Medical Microbiology and Immunology* 182, 37–50.
- Krueger, F., 2007. Trim Galore [online]. available at: https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/.
- Kugeler, K.J., Schwartz, A.M., Delorey, M.J., Mead, P.S., Hinckley, A.F., 2021. Estimating the frequency of Lyme disease diagnoses, United States, 2010–2018. *Emerg. Infect. Dis.* 27, 616–619. <https://doi.org/10.3201/eid2702.202731>
- Leggett, H.C., Buckling, A., Long, G.H., Boots, M., 2013. Generalism and the evolution of parasite virulence. *Trends in Ecology & Evolution* 28, 592–596. <https://doi.org/10.1016/j.tree.2013.07.002>
- Levine, J., Wilson, M., Spielman, A., 1984. Mice as reservoirs of the Lyme disease spirochete. *The American Journal of Tropical Medicine and Hygiene* 34, 335–360. <https://doi.org/10.4269/ajtmh.1985.34.355>
- Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25, 1754–1760.
- Llewellyn, M.S., Miles, M.A., Carrasco, H.J., Lewis, M.D., Yeo, M., Vargas, J., Torrico, F., Diosque, P., Valente, V., Valente, S.A., Gaunt, M.W., 2009. Genome-scale multilocus microsatellite typing of *Trypanosoma cruzi* discrete typing unit I reveals phylogeographic structure and specific genotypes linked to human infection. *PLoS Pathog* 5, e1000410. <https://doi.org/10.1371/journal.ppat.1000410>
- LoGiudice, K., Ostfeld, R.S., Schmidt, K.A., Keesing, F., 2003. The ecology of infectious disease: Effects of host diversity and community composition on Lyme disease risk. *Proc. Natl. Acad. Sci. U.S.A.* 100, 567–571. <https://doi.org/10.1073/pnas.0233733100>
- Madhav, N.K., Brownstein, J.S., Tsao, J.I., Fish, D., 2004. A dispersal model for the range expansion of blacklegged tick (Acari: Ixodidae). *J Med Entomol* 41, 842–852. <https://doi.org/10.1603/0022-2585-41.5.842>
- Malpica, J.M., Sacristán, S., Fraile, A., García-Arenal, F., 2006. Association and host selectivity in multi-host pathogens. *PLoS ONE* 1, e41. <https://doi.org/10.1371/journal.pone.0000041>
- Markowski, D., Howard, G.S., Hyland, K.E., Hu, R., 1998. Reservoir competence of the meadow vole (Rodentia: *Cricetidae*) for the Lyme disease spirochete *Borrelia burgdorferi*. *Journal of Medical Entomology* 35, 804–808. <https://doi.org/10.1093/jmedent/35.5.804>
- Massey, R.C., Buckling, A., French–Constant, R., 2004. Interference competition and parasite virulence. *Proc. R. Soc. Lond. B* 271, 785–788. <https://doi.org/10.1098/rspb.2004.2676>
- Mather, T.N., Iii, S.R.T., MacLachlan, A.B., Spielman, A., 1989. Incompetence of catbirds as reservoirs for the Lyme disease spirochete (*Borrelia*

- burgdorferi
- . The Journal of Parasitology 75, 66.
-
- <https://doi.org/10.2307/3282938>
- Mather, T.N., Nicholson, M.C., Donnelly, E.F., Matyas, B.T., 1996. Entomologic index for human risk of Lyme disease. American Journal of Epidemiology 144, 1066–1069. <https://doi.org/10.1093/oxfordjournals.aje.a008879>
- McDonald, B.A., Linde, C., 2002. Pathogen population genetics, evolutionary potential, and durable resistance. Annu. Rev. Phytopathol. 40, 349–379. <https://doi.org/10.1146/annurev.phyto.40.120501.101443>
- Medeiros, M.C.I., Ellis, V.A., Ricklefs, R.E., 2014. Specialized avian Haemosporida trade reduced host breadth for increased prevalence. J. Evol. Biol. 27, 2520–2528. <https://doi.org/10.1111/jeb.12514>
- Mideo, N., 2009. Parasite adaptations to within-host competition. Trends in Parasitology 25, 261–268. <https://doi.org/10.1016/j.pt.2009.03.001>
- Movila, A., Toderas, I., Dubinina, H., Uspenskaia, I., Alekseev, A., 2012. Zoonotic peculiarities of *Borrelia burgdorferi s.l.*: vectors competence and vertebrate host specificity, in: Lyme Disease. InTech, pp. 28–54.
- Musser, J., 1996. Molecular population genetic analysis of emerged bacterial pathogens: selected insights. Emerg. Infect. Dis. 2, 1–17. <https://doi.org/10.3201/eid0201.960101>
- Myers, P., Lundrigan, B.L., Hoffman, S.M.G., Haraminac, A.P., Seto, S.H., 2009. Climate-induced changes in the small mammal communities of the Northern Great Lakes Region. Global Change Biology 15, 1434–1454. <https://doi.org/10.1111/j.1365-2486.2009.01846.x>
- Nouri, M., Latorre-Margalef, N., Czopek, A., Råberg, L., 2021. Cross-reactivity of antibody responses to *Borrelia afzelii ospC* : asymmetry and host heterogeneity. Infection, Genetics and Evolution 91, 104793. <https://doi.org/10.1016/j.meegid.2021.104793>
- Nupp, T.E., Swihart, R.K., 1998. Effects of forest fragmentation on population attributes of white-footed mice and eastern chipmunks. Journal of Mammalogy 79, 1234–1243. <https://doi.org/10.2307/1383014>
- Nychka, D., Furrer, R., Paige, J., Sain, S., 2021. fields: Tools for spatial data R package version 14.1. [online]. Available at: <https://github.com/dnychka/fieldsRPackage>.
- Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O’Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., Wagner, H., Barbour, M., Bedward, M., Bolker, B., Borcard, D., Carvalho, G., Chirico, M., De Caceres, M., Durand, S., Antoniazzi Evangelista, H.B., FitzJohn, R., Friendly, M., Furneaux, B., Hannigan, G., Hill, M.O., Lahti, L., McGlenn, D., Ouellette, M.-H., Ribeiro Cunha, E., Smith, T., Stier, A., Ter Braak, C.J.F., Weedon, J., 2022. vegan: Community Ecology Package [online]. Available at: <https://github.com/vegandevs/vegan>.
- Oliver, J., Means, R.G., Kogut, S., Prusinski, M., Howard, J.J., Layne, L.J., Chu, F.K., Reddy, A., Lee, L., White, D.J., 2006. Prevalence of *Borrelia burgdorferi* in

- small mammals in New York state. *Journal of Medical Entomology* 43, 12. [https://doi.org/10.1603/0022-2585\(2006\)43\[924:pobbis\]2.0.co;2](https://doi.org/10.1603/0022-2585(2006)43[924:pobbis]2.0.co;2)
- Ostfeld, R.S., Keesing, F., 2000. Biodiversity and disease risk: The case of Lyme disease. *Conservation Biology* 14, 722–728. <https://doi.org/10.1046/j.1523-1739.2000.99014.x>
- Ostfeld, R.S., Levi, T., Jolles, A.E., Martin, L.B., Hosseini, P.R., Keesing, F., 2014. Life history and demographic drivers of reservoir competence for three tick-borne zoonotic pathogens. *PLoS ONE* 9, e107387. <https://doi.org/10.1371/journal.pone.0107387>
- Pearson, P., Skaltsis, O., Luo, C.-Y., Xu, G., Oppler, Z., Brisson, D., Rich, S.M., 2022. A *Borrelia burgdorferi* outer surface protein C (*ospC*) genotyping method using Luminex technology. *PLoS ONE* 17, e0269266. <https://doi.org/10.1371/journal.pone.0269266>
- Piesman, J., Mather, T.N., Dammin, G.J., Telford, S.R., Lastavica, C.C., Spielman, A., 1987. Seasonal variation of transmission risk of Lyme disease and human Babesiosis. *American Journal of Epidemiology* 126, 1187–1189. <https://doi.org/10.1093/oxfordjournals.aje.a114757>
- Probert, W.S., Crawford, M., Cadiz, R.B., LeFebvre, R.B., 1997. Immunization with outer surface protein (*osp*) A, but not *ospC*, provides cross-protection of mice challenged with North American isolates of *Borrelia burgdorferi*. *Journal of Infectious Diseases* 175, 400–405. <https://doi.org/10.1093/infdis/175.2.400>
- Putaporntip, C., Jongwutiwes, S., Thongaree, S., Seethamchai, S., Grynberg, P., Hughes, A.L., 2010. Ecology of malaria parasites infecting Southeast Asian macaques: evidence from cytochrome b sequences: Ecology of malaria parasites. *Molecular Ecology* 19, 3466–3476. <https://doi.org/10.1111/j.1365-294X.2010.04756.x>
- Qiu, W.-G., Dykhuizen, D.E., Acosta, M.S., Luft, B.J., 2002. Geographic uniformity of the Lyme disease spirochete (*Borrelia burgdorferi*) and its shared history with tick vector (*Ixodes scapularis*) in the northeastern United States. *Genetics* 160, 833–849. <https://doi.org/10.1093/genetics/160.3.833>
- R Core Team, 2020. R: A language and environment for statistical computing [online]. Available at: <https://www.R-project.org/>.
- Read, A.F., Taylor, L.H., 2001. The ecology of genetically diverse infections. *Science* 292, 1099–1102. <https://doi.org/10.1126/science.1059410>
- Regoes, R.R., Nowak, M.A., Bonhoeffer, S., 2000. Evolution of virulence in a heterogeneous host population. *Evolution* 54, 64–71. <https://doi.org/10.1111/j.0014-3820.2000.tb00008.x>
- Richter, D., 2000. Competence of American robins as reservoir hosts for Lyme disease spirochetes. *Emerg. Infect. Dis.* 6, 133–138. <https://doi.org/10.3201/eid0602.000205>
- Rudenko, N., Golovchenko, M., Hönig, V., Mallátová, N., Krbková, L., Mikulášek, P., Fedorova, N., Belfiore, N.M., Grubhoffer, L., Lane, R.S., Oliver, J.H., 2013. Detection of *Borrelia burgdorferi* Sensu Stricto *ospC* alleles associated with

- human Lyme Borreliosis worldwide in non-human-biting tick *Ixodes affinis* and rodent hosts in Southeastern United States. *Appl Environ Microbiol* 79, 1444–1453. <https://doi.org/10.1128/AEM.02749-12>
- 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. *J Clin Microbiol* 38, 382–388. <https://doi.org/10.1128/JCM.38.1.382-388.2000>
- 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. *Proc. Natl. Acad. Sci. U.S.A.* 92, 2909–2913. <https://doi.org/10.1073/pnas.92.7.2909>
- Seinost, G., Dykhuizen, D.E., Dattwyler, R.J., Golde, W.T., Dunn, J.J., Wang, I.-N., Wormser, G.P., Schriefer, M.E., Luft, B.J., 1999. Four clones of *Borrelia burgdorferi sensu stricto* cause invasive infection in humans. *Infect Immun* 67, 3518–3524. <https://doi.org/10.1128/IAI.67.7.3518-3524.1999>
- Shifflett, S.A., Wiedmeyer, T., Kennedy, A., Maestas, L., Buoni, M., Ciloglu, A., Ellis, V., 2023. Data from Prevalence of *Borrelia burgdorferi* and diversity of its outer surface protein C (*ospC*) alleles in blacklegged ticks (*Ixodes scapularis*) in Delaware. <https://doi.org/10.5061/dryad.bcc2fqzh8>.
- Shifflett, S.A., Wiedmeyer, T., Kennedy, A., Maestas, L., Buoni, M., Ciloglu, A., Ellis, V.A., 2023. Prevalence of *Borrelia burgdorferi* and diversity of its outer surface protein C (*ospC*) alleles in blacklegged ticks (*Ixodes scapularis*) in Delaware. *Ticks and Tick-borne Diseases* 14, 102139. <https://doi.org/10.1016/j.ttbdis.2023.102139>
- Shifflett, S.A., Ferreira, F.C., González, J., Toledo, A., Fonseca, D.M., Ellis, V.A., *in prep.* A specialized *Borrelia burgdorferi* outer surface protein C (*ospC*) may mitigate Lyme disease risk in humans.
- Slajchert, T., Kitron, U.D., Jones, C.J., Mannelli, A., 1997. Role of the eastern chipmunk (*Tamias striatus*) in the epizootiology of Lyme Borreliosis in Northwestern Illinois, USA. *Journal of Wildlife Diseases* 33, 40–46. <https://doi.org/10.7589/0090-3558-33.1.40>
- Steere, A.C., 1987. The clinical evolution of Lyme arthritis. *Ann Intern Med* 107, 725. <https://doi.org/10.7326/0003-4819-107-5-725>
- Stickel, L.F., 1968. Home range and travels, in: King, J.A. (Ed.), *Biology of Peromyscus* (Rodentia). American Society of Mammalogists, pp. 373–411.
- Straub, C.S., Ives, A.R., Gratton, C., 2011. Evidence for a trade-off between host-range breadth and host-use efficiency in aphid parasitoids. *The American Naturalist* 177, 389–395. <https://doi.org/10.1086/658177>
- Swanson, K.I., Norris, D.E., 2008. Presence of multiple variants of *Borrelia burgdorferi* in the natural reservoir *Peromyscus leucopus* throughout a transmission season. *Vector-Borne and Zoonotic Diseases* 8, 397–406. <https://doi.org/10.1089/vbz.2007.0222>
- Telford, Sam.R., Mather, T.N., Moore, S.I., Wilson, M.L., Spielman, A., 1988. Incompetence of deer as reservoirs of the Lyme disease spirochete. *The*

- American Journal of Tropical Medicine and Hygiene 39, 105–109.
<https://doi.org/10.4269/ajtmh.1988.39.105>
- Telford, S.R., Mather, T.N., Adler, G.H., Spielman, A., 1990. Short-tailed shrews as reservoirs of the agents of Lyme disease and human Babesiosis. *The Journal of Parasitology* 76, 681. <https://doi.org/10.2307/3282982>
- Tilly, K., Krum, J.G., Bestor, A., Jewett, M.W., Grimm, D., Bueschel, D., Byram, R., Dorward, D., VanRaden, M.J., Stewart, P., Rosa, P., 2006. *Borrelia burgdorferi* ospC protein required exclusively in a crucial early stage of mammalian infection. *Infect Immun* 74, 3554–3564.
<https://doi.org/10.1128/IAI.01950-05>
- Tonetti, N., Voordouw, M.J., Durand, J., Monnier, S., Gern, L., 2015. Genetic variation in transmission success of the Lyme Borreliosis pathogen *Borrelia afzelii*. *Ticks and Tick-borne Diseases* 6, 334–343.
<https://doi.org/10.1016/j.ttbdis.2015.02.007>
- Tyler, S., Tyson, S., Dibernardo, A., Drebot, M., Feil, E.J., Graham, M., Knox, N.C., Lindsay, L.R., Margos, G., Mechai, S., Van Domselaar, G., Thorpe, H.A., Ogden, N.H., 2018. Whole genome sequencing and phylogenetic analysis of strains of the agent of Lyme disease *Borrelia burgdorferi* from Canadian emergence zones. *Sci Rep* 8, 10552. <https://doi.org/10.1038/s41598-018-28908-7>
- 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. <https://doi.org/10.1016/j.meegid.2013.12.011>
- Walter, K.S., Carpi, G., Caccone, A., Diuk-Wasser, M.A., 2017. Genomic insights into the ancient spread of Lyme disease across North America. *Nat Ecol Evol* 1, 1569–1576. <https://doi.org/10.1038/s41559-017-0282-8>
- Wang, I.-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, 15–30. <https://doi.org/10.1093/genetics/151.1.15>
- Wickham, H., 2016. *ggplot2: elegant graphics for data analysis* dp. Springer-Verlag New York.
- Wickham, H., Francois, R., Henry, L., Muller, K., 2021. *dplyr: a grammar of data manipulation* [online]. Available at: <https://CRAN.R-project.org/package=dplyr>. R package 1.0.7.
- Wille, P., Boller, T., Kaltz, O., 2002. Mixed inoculation alters infection success of strains of the endophyte *Epichloë bromicola* on its grass host *Bromus erectus*. *Proc. R. Soc. Lond. B* 269, 397–402. <https://doi.org/10.1098/rspb.2001.1889>
- Wilske, B., Preac-Mursic, V., Jauris, S., Hofmann, A., Pradel, I., Soutschek, E., Schwab, E., Will, G., Wanner, G., 1993. Immunological and molecular polymorphisms of ospC, an immunodominant major outer surface protein of *Borrelia burgdorferi*. *Infect Immun* 61, 2182–2191.
<https://doi.org/10.1128/iai.61.5.2182-2191.1993>

- Woolhouse, M.E.J., Gowtage-Sequeria, S., 2005. Host range and emerging and reemerging pathogens. *Emerg. Infect. Dis.* 11, 1842–1847.
<https://doi.org/10.3201/eid1112.050997>
- Wormser, G.P., Brisson, D., Liveris, D., Hanincová, K., Sandigursky, S., Nowakowski, J., Nadelman, R.B., Ludin, S., Schwartz, I., 2008. *Borrelia burgdorferi* genotype predicts the capacity for hematogenous dissemination during early Lyme disease. *J INFECT DIS* 198, 1358–1364.
<https://doi.org/10.1086/592279>
- Xu, G., Wielstra, B., Rich, S.M., 2020. Northern and southern blacklegged (deer) ticks are genetically distinct with different histories and Lyme spirochete infection rates. *Sci Rep* 10, 10289. <https://doi.org/10.1038/s41598-020-67259-0>

Appendix A

CHAPTER 1 SUPPLEMENTARY MATERIALS

Table A.1 Details on the seven sites across Delaware where blacklegged ticks (*Ixodes scapularis*) were collected. All ticks were collected using both drag and flag methods. Sites NC1 and K1 were sampled at two separate locations within the site on the same day. Each site shows the nearest city in Delaware, the habitat type, vegetation, and composition of the sampling location along with the temperature, relative humidity, and weather conditions on the sampling day.

Site	County	Nearest city	Collection date	Habitat	Vegetation and habitat composition	Weather	Temp(°C)	Wetness	Relative humidity	Wind
NC1	New Castle	Newark	June 1, 2019	Field edge	Tulip popular (<i>Liriodendron tulipifera</i>), variety of oaks (<i>Quercus</i> spp.)	Partly cloudy	28.8	Damp	84.7	None
NC2	New Castle	Rockland	June 22, 2019	Trail edge	Tulip popular (<i>L. tulipifera</i>), freshwater marsh	Sunny	24.2	Damp	44.2	light
K1	Kent	Leipsic	June 5, 2019	Field edge	Variety of oak and pines (<i>Pinus</i> spp.) tidal salt marsh	Partly cloudy	31.9	Dry	69.1	Light
K2	Kent	Felton	June 5, 2019	Forest edge	Hardwood forest, millpond	Partly cloudy	21.0	Dry	78.2	Light
S1	Sussex	Dagsboro	April 24, 2019	Field edge	Hardwood forests, shores of the inland bays	Partly cloudy	21.8	Dry	65.5	Strong
S2	Sussex	Milton	June 26, 2019	Trail edge	hardwood forests, back-barrier salt marsh, and grassland dunes	Sunny	24.2	Dry	88.0	None
S3	Sussex	Laurel	June 28, 2019	Trail edge	Mostly oak and beech (<i>Fagus</i> spp.) with successional tulip popular	Overcast	27.9	Wet	61.0	Strong

Table A.2 An example of mapping statistics used to determine the presence of an ospC (or vsp) allele. These are the results from the sample DT-3, an infected tick from site NC1. Using the metrics 1) any allele with all nucleotide positions sequenced to a depth of coverage of at least 5X and 2) any allele where mean depth of coverage was greater than the standard deviation of depth of coverage between nucleotide positions 200 and 550 (the more variable positions), we determined that DT-3 had five *B. burgdorferi* ospC alleles (N, D, M, C, and B; see column “Alleles scored as present”). The reference file we used is from Di et al. (2018) and includes 19 ospC alleles from *B. burgdorferi* and one vsp allele from *B. myiamotoi*.

<i>ospC</i> allele	Allele scored as present	Proportion of nucleotide positions in the reference with at least 5X depth of coverage	Mean depth of coverage	Standard deviation of depth of coverage	Relative coverage (%)
N	Yes	1	1969.36	172.98	0.42
D	Yes	1	1385.83	40.56	0.29
M	Yes	1	676.71	39.67	0.14
C	Yes	1	641.9	23.65	0.14
B	Yes	1	32.3	6.83	0.01
I	No	0.37	1.44	1.81	0
O	No	0.34	0.52	1.02	0
A	No	0.34	1.77	1.58	0
E	No	0.3	0.11	0.45	0
H	No	0.69	5.56	4.96	0
K	No	0.31	1.39	1.35	0
F	No	0.32	0	0	0
G	No	0.41	1.01	5.67	0
J	No	0.29	0	0	0
T	No	0.27	0.13	0.81	0
U	No	0.25	0	0	0
B3	No	0.24	0.63	2.45	0
L	No	0.06	0	0	0
C14	No	0	0	0	0
vsp	No	0	0	0	0

Table A.3 Quantitative metrics used to determine whether an allele was present. Each sequenced sample is presented and shows the alleles determined to be present along with the allele's percentage of coverage (i.e., the percentage of all mapping reads that mapped to a particular allele), the mean and standard deviation of depth of coverage between nucleotide positions 200 and 550 of the reference, and the proportion of nucleotide positions in the reference with a depth of coverage of at least 5X. We considered an allele to be present in an infection if all nucleotide positions in the reference of the allele were sequenced to a depth of coverage of at least 5X and where mean depth of coverage was greater than the standard deviation of depth of coverage between nucleotide positions 200 to 550 (the more variable positions). Sample DT-4 and Is-49 partially mapped to *vsp* (a gene from *B. myiamotoi* that is present in the reference file) and are shown in the table below.

ID	ospC allele	Proportion of nucleotide positions in the reference with at least 5X depth of coverage	Mean depth of coverage	Standard deviation of depth of coverage	Relative coverage (%)
DT-1	Potentially novel allele	-	-	-	-
DT-3	N	1	1969.36	172.98	42
	D	1	1385.83	40.56	29
	M	1	676.71	39.67	14
	C	1	641.9	23.65	14
	B	1	32.3	6.83	1
DT-4	<i>vsp</i>	0.995	183.05	447.34	68.7
DT-5	C	1	6212.27	188.09	100
DT-7	O	1	5318.35	428.42	99
DT-8	A	1	2566.45	71.49	46

	B	1	1763.19	72.74	31
	C	1	876.65	65.46	16
	K	1	135.18	18.45	2
	G	1	120.31	9.61	2
	D	1	67.49	11.08	1
	N	1	48.69	4.73	1
DT-11	F	1	7531.95	373.25	100
DT-13	C	1	5232.33	195.04	100
DT-15	T	1	5054.02	261.37	62
	C	1	2603.67	95.35	32
	B	1	294.94	19.05	4
	A	1	187.57	12.92	2
	K	1	60.18	7.75	1
DT-21	U	1	7305.66	299.43	100
DT-23	T	1	6161.01	205.3	87
	A	1	920.63	63.39	13
DT-26	D	1	5716.13	303.21	100
DT-27	K	1	7110.62	255.4	99
DT-32	K	1	6011.79	206.35	99
DT-34	K	1	2313.99	97.88	47
	E	1	1468.46	175.9	30
	B	1	747.89	41.85	15
	M	1	190.13	11.97	4
	G	1	100.13	11.41	2
	H	1	64.27	6.78	1

DT-37	E	1	7650.88	215.68	92
	D	1	673.54	36.32	8
DT-38	A	1	7589.02	193.37	100
DT-39	K	1	3513.15	153.08	53
	B	1	1636.96	77.34	25
	D	1	1382.49	57.01	21
	A	1	10.41	8.97	0
DT-40	A	1	5223.12	168.77	85
	K	1	904.06	74.79	15
DT-42	A	1	6427.58	185.51	100
DT-43	K	1	4446.13	178.97	74
	N	1	1491.83	150.57	25
DT-44	A	1	6694.56	222.58	100
DT-45	F	1	7776.79	215.05	98
	K	1	115.22	7.56	1
DT-47	A	1	7649.81	243.43	99
	K	1	79.82	5.91	1
DT-50	I	1	5256.19	137.15	62
	A	1	3263.45	87.5	38
DT-52	T	1	7455.04	211.36	100
DT-55	A	1	6021.57	160.81	94
	T	1	352.74	19.93	5
	I	1	49.19	11.97	1
DT-57	A	1	4124.79	190.53	100
DT-60	E	1	5835.38	368.94	100
DT-59	E	1	7111.82	430	100
DT-56	T	1	3149.3	147.11	74

	F	1	1112.75	97.51	26
ls-29	A	1	4591.26	168.18	100
ls-33	K	1	7493.49	253.25	99
ls-5	A	1	6424.66	238.86	100
ls-9	N	1	5986.68	527.12	100
ls-13	A	1	2544.64	93.67	100
ls-16	B	1	7188.76	397.38	94
ls-16	E	1	391.24	32.97	5
ls-49	vsp	1	288.01	683.49	75
ls-50	F	1	7027.86	413.77	100
ls-56	K	1	5064.28	174.63	99
ls-76	G	1	4957.17	231.3	100
ls-79	C	1	6254.67	161.91	100
ls-81	M	1	5994.42	237.07	100
ls-85	H	1	5217.33	315.01	99
ls-86	K	1	7020.42	415.74	99
ls-89	C	1	2174.02	954.07	33
	K	1	1813.19	514.62	27
	Potentially novel allele	-	-	-	-
ls-92	D	1	6470.69	155.49	100
ls-96	K	1	7594.56	290.75	99
ls-97	A	1	5328.07	156.1	74
	B	1	1821.25	87.04	25
ls-98	B	1	4069.4	243.96	99
ls-99	B	1	6328.34	277.54	98
ls-99	K	1	115.35	9.39	2

Is-103	K	1	7604.57	303.34	99
Is-107	K	1	6079.73	284.44	95
	B	1	283.65	22.51	4
Is-108	M	1	4091.76	108.25	58
	I	1	1163.87	139.4	16
	B	1	1115.73	57.37	16
	A	1	342.75	14.8	5
	D	1	311.77	16.33	4
Is-110	A	1	7874.38	98.88	100
Is-111	K	1	7902.31	142.25	96
	G	1	135.73	13.43	2
	H	1	68.13	5.82	1
Is-38	N	1	7879.1	219.38	69
	A	1	2220.54	94.77	20
	C	1	601.2	255.72	5
Is-45	C	1	7877.87	267.21	99
Is-53	A	1	7815.65	425.82	99
Is-63	B	1	7860.49	228.66	71
	A	1	3212.49	113.02	29

Table A.4 Mapping statistics for each sample. Each sequenced sample is presented (ID) with total number of reads, number of reads that mapped to the reference file, number of reads that did not map (unmapped), number of reads that mapped and paired, and the percentage of unmapped reads. Two samples (DT-4 and Is-49) had more than 50% unmapped reads and those represent the two punitive B. myiamotoi samples.

ID	Total number of reads	Reads mapped	Reads unmapped	Reads mapped and paired	% Unmapped reads
DT-1	35574	34808	766	34732	2.15
DT-3	18896	21511	1449	21430	7.67
DT-4	18896	5133	13763	5056	72.84
DT-5	28450	27592	858	27530	3.02
DT-7	24848	23629	1219	23566	4.91
DT-8	27070	25785	1285	25732	4.75
DT-11	32196	30451	1745	30396	5.42
DT-13	24572	23116	1456	23074	5.93
DT-15	35574	34808	766	34732	2.15
DT-21	32310	31336	974	31262	3.01
DT-23	32294	31298	996	31224	3.08
DT-26	25778	24762	1016	24706	3.94
DT-27	31146	30738	408	30676	1.31
DT-32	26804	26428	376	26368	1.4
DT-34	23088	22467	621	22428	2.69
DT-37	37780	37033	747	36946	1.98
DT-38	36078	34846	1232	34756	3.41

DT-39	30150	29606	544	29544	1.8
DT-40	29920	28445	1475	28382	4.93
DT-42	29208	27134	2074	27062	7.1
DT-43	27850	27328	522	27274	1.87
DT-44	31172	29417	1755	29364	5.63
DT-45	38872	37649	1223	37566	3.15
DT-47	31042	30565	477	30498	1.54
DT-50	36448	35703	745	35634	2.04
DT-52	34094	33302	792	33198	2.32
DT-55	32432	31228	1204	31138	3.71
DT-57	27226	19255	7971	19186	29.28
DT-60	31934	30651	1283	30568	4.02
DT-59	36326	34375	1951	34286	5.37
DT-56	22172	20744	1428	20660	6.44
Is-29	24386	18938	5448	18898	22.34
Is-33	31540	31018	522	30954	1.66
Is-5	28532	28197	335	28136	1.17
Is-9	27650	26914	736	26840	2.66
Is-13	12828	12436	392	12380	3.06
Is-16	33114	32473	641	32414	1.94
Is-49	28550	6499	22051	6374	77.24
Is-50	32534	29510	3024	29442	9.29
Is-56	28024	22048	5976	22000	21.32
Is-76	24378	21420	2958	21390	12.13
Is-79	29588	28103	1485	28022	5.02
Is-81	28846	28278	568	28208	1.97
Is-85	24312	23692	620	23642	2.55
Is-86	35056	34526	530	34422	1.51

ls-89	36708	35812	896	35738	2.44
ls-92	29760	28810	950	28746	3.19
ls-96	35888	34914	974	34848	2.71
ls-97	31656	30974	682	30898	2.15
ls-98	20786	18934	1852	18880	8.91
ls-99	32344	30190	2154	30114	6.66
ls-103	41392	37769	3623	37670	8.75
ls-107	28956	28354	602	28266	2.08
ls-108	33460	32669	791	32592	2.36
ls-110	38456	37902	554	37832	1.44
ls-111	54810	54379	431	54268	0.79
ls-38	71026	70005	1021	69804	1.44
ls-45	75874	74785	1089	74550	1.44
ls-53	75390	74175	1215	73970	1.61
ls-63	70714	69333	1381	69188	1.95

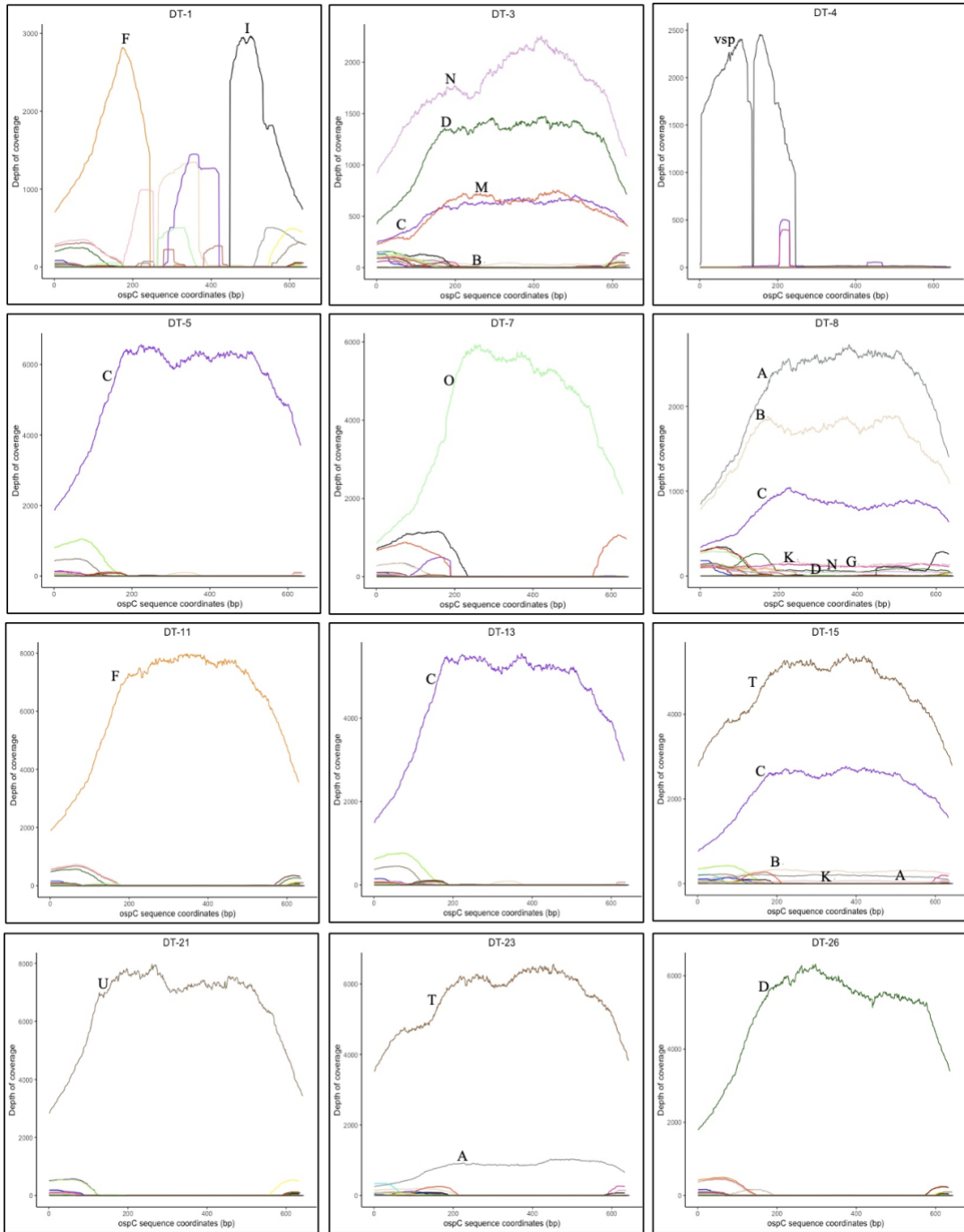
Table A.5 Bray-Curtis dissimilarities between sites based on the relative frequencies of *ospC* alleles. Only sites with more than three alleles present (Table 2) were considered.

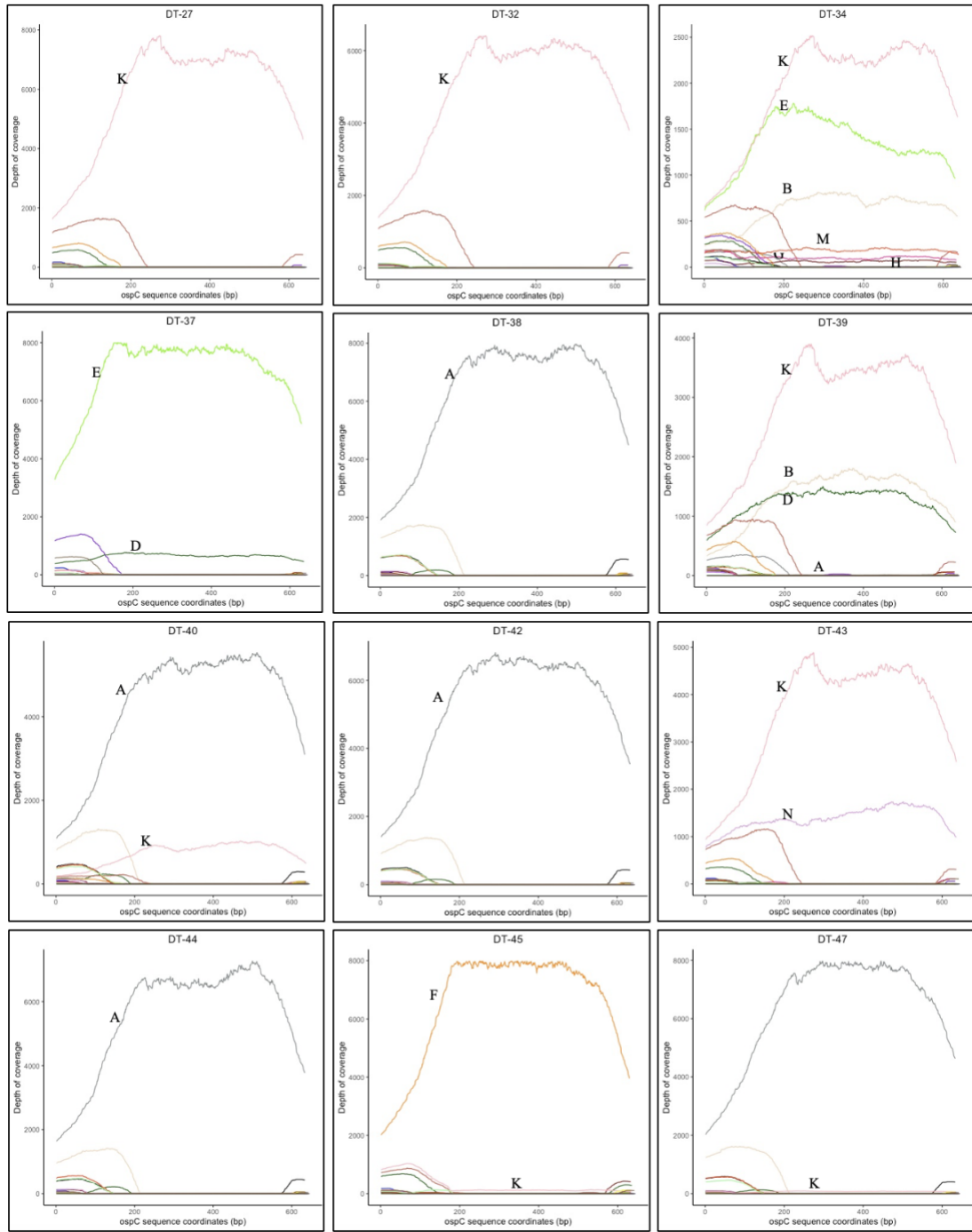
	NC1	NC3	K1
NC3	0.518		
K1	0.474	0.459	
K2	0.633	0.706	0.395

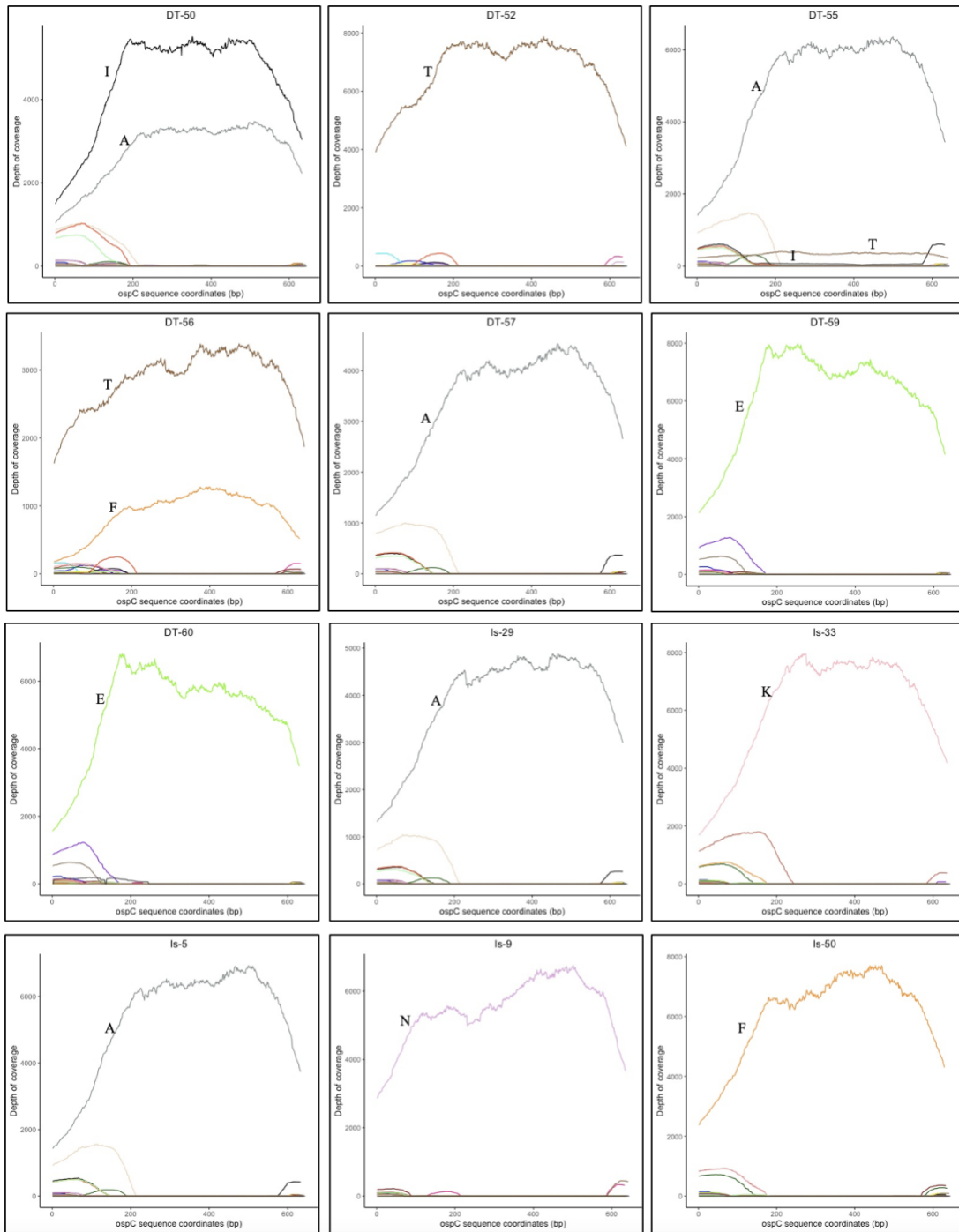
Table A.6 Samples with mixed *ospC* allele infections. A “1” indicates infection of a sample (ID) with a particular allele. Allele multiplicity refers to the total number of alleles found in an individual sample. The majority of mixed infections contained two *ospC* alleles, with the greatest number of alleles in one infection being seven.

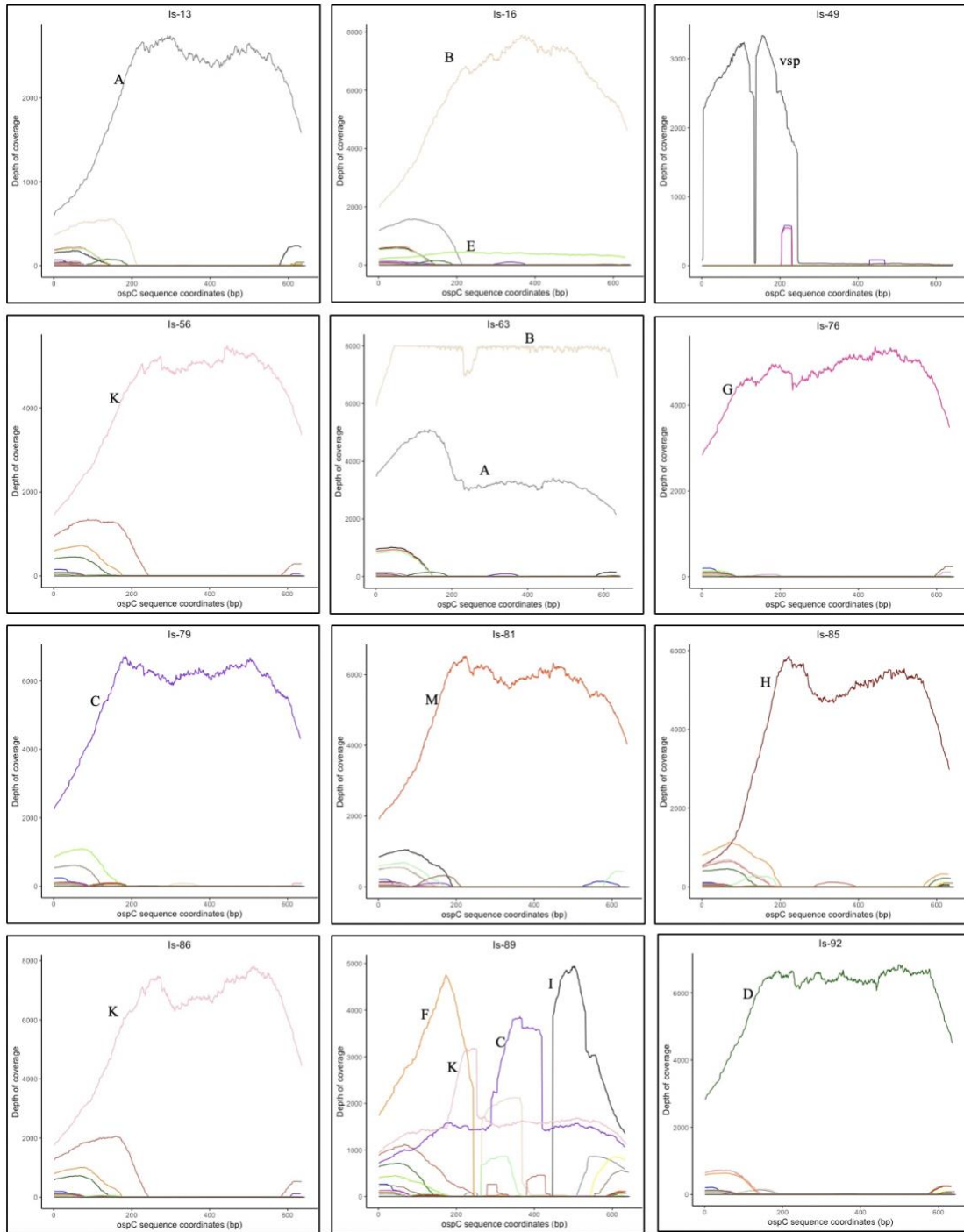
ID	Site	U	T	A	D	K	E	B	M	G	H	N	C	F	I	O	Allele multiplicity
DT-3	NC1	0	0	0	1	0	0	1	1	0	0	1	1	0	0	0	5
DT-8	NC1	0	0	1	1	1	0	1	0	1	0	1	1	0	0	0	7
DT-15	NC1	0	1	1	0	1	0	1	0	0	0	0	1	0	0	0	5
DT-23	NC1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	2
Is-38	NC1	0	0	1	0	0	0	0	0	0	0	1	1	0	0	0	3
Is-89	NC1	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	2
Is-16	NC2	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	2
DT-34	NC3	0	0	0	0	1	1	1	1	1	1	0	0	0	0	0	6
DT-37	NC3	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	2
DT-55	NC3	0	1	1	0	0	0	0	0	0	0	0	0	0	1	0	3
DT-56	NC3	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	2
DT-43	K1	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	2
DT-45	K1	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	2
DT-47	K1	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	2
DT-50	K1	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	2
Is-97	K1	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	2
Is-99	K1	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	2
Is-107	K1	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	2
Is-108	K1	0	0	1	1	0	0	1	1	0	0	0	0	0	1	0	5

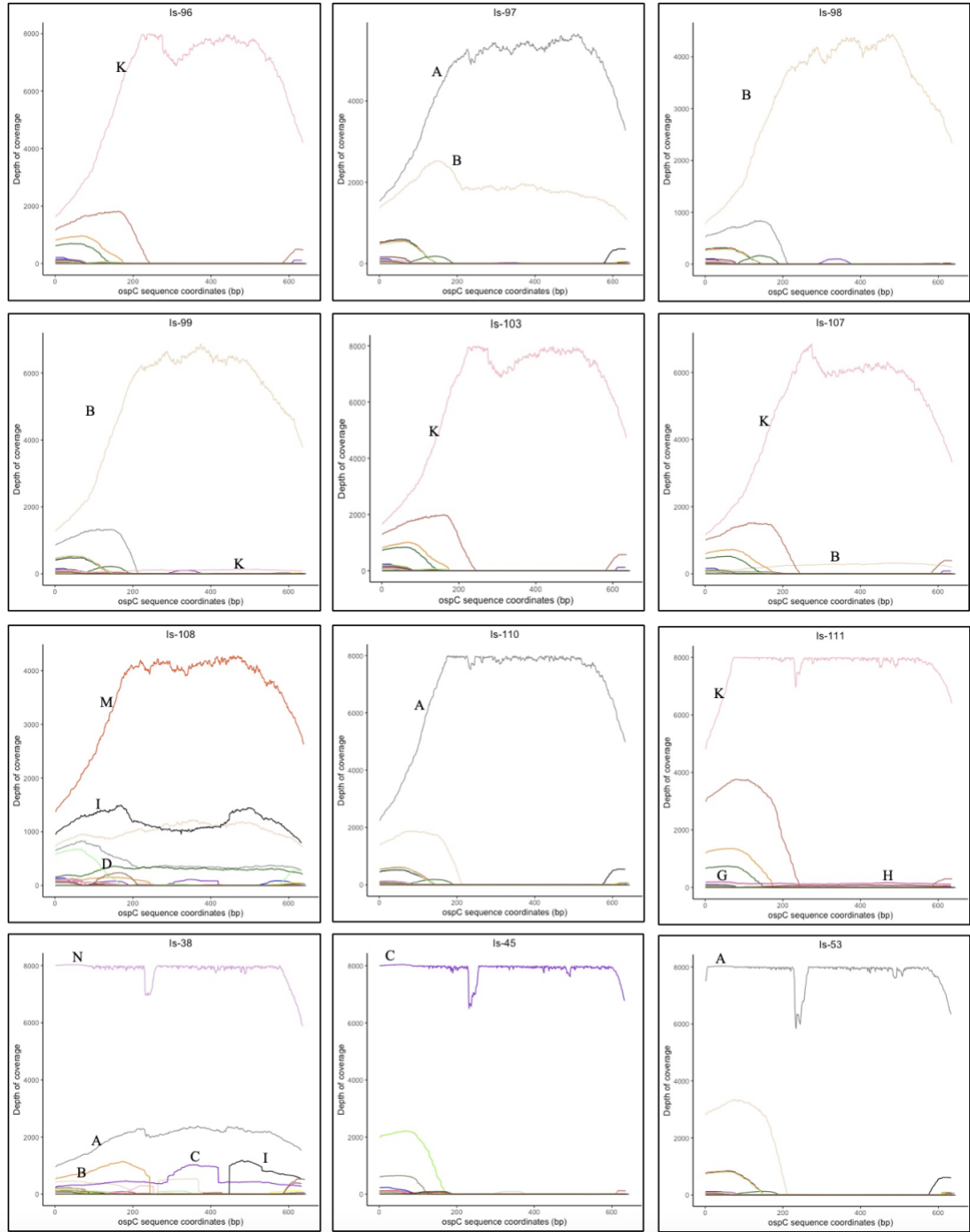
Is-111	K1	0	0	0	0	1	0	0	0	1	1	0	0	0	0	0	3
Is-63	K1	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	2
DT-39	K2	0	0	1	1	1	0	1	0	0	0	0	0	0	0	0	4
DT-40	K2	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	2
Total allele frequency		0	4	12	5	12	3	11	3	3	2	4	5	2	3	0	69











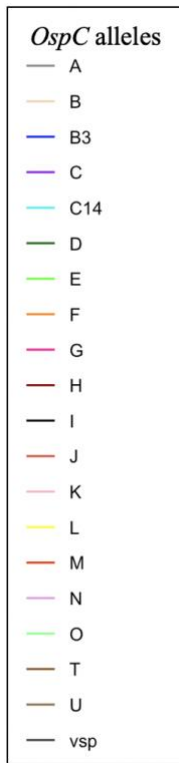


Figure A.1 Sequence depth of coverage plotted over the reference sequences for each infected tick sample (sample name is presented at the top of each panel). The y-axis shows depth of coverage, and the x-axis refers to the sequence coordinates of the *ospC* (or *vsp*) gene. *OspC* alleles from the reference sequences are distinguished by color (see legend at the bottom of the figure) and alleles that were considered to have sufficient coverage to be considered true infections have a label next to their depth of coverage line to further distinguish them. Several samples did not map well to any of the reference sequences. For example, sample DT-1 did not fully map to any of the 19 reference alleles and may represent a novel allele. Mixed infections can be identified by multiple alleles mapping to the same sample. For example, DT-3 has five *ospC* alleles present (N, C, M, D, and B) with allele N having the highest relative coverage in the sample. Two samples (DT-4 and Is-49) partially mapped to the *vsp* gene of *Borrelia myiamotoi*.

Appendix B

CHAPTER 2 SUPPLEMENTARY MATERIALS

	<i>P. lotor</i>	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	<i>S. carolinensis</i>	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	<i>T. striatus</i>	18	7	2	2	0	1	1	3	1	1	1	1	1	0	0	0
	Total	71	21	9	7	4	1	1	3	2	1	1	1	2	0	0	0
University Inn	<i>D. virginiana</i>	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	<i>M. monax</i>	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	<i>M. mephitis</i>	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	<i>M. pennsylvanicus</i>	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	<i>Peromyscus spp.</i>	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	<i>P. lotor</i>	13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	<i>S. carolinensis</i>	11	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	<i>T. striatus</i>	8	4	0	0	0	0	0	2	1	0	0	1	0	0	0	0
	Total	41	5	0	0	0	0	0	2	1	0	0	1	0	0	0	0
	Total alleles across all sites	248	68	2	5	17	15	7	6	6	3	2	2	2	2	1	1
Relative frequency (%)			.8	.9	.7	.8	.7	.6	.6	.3	.2	.2	.2	.2	1.	1.	1.

Table B.2 Diversity of the mammalian host species an *ospC* allele infects as determined by the Gini-Simpson index. Observed Gini-Simpson index values were calculated from the original *ospC* frequency data. Expected Gini-Simpson index values represent the mean Gini-Simpson index after randomizing *ospC* allele frequencies across infected hosts 9,999 times. The two-tailed P values for each of the alleles represents if the Gini-Simpson index is more extreme than expected by chance.

	A	B	C	D	E	F	G	H	I	K	N	O	T	U
observed	0.578	0.509	0.735	0.444	0.5	0	0.5	0	0	0.533	0.5	0	0.444	0
mean	0.62	0.635	0.569	0.554	0.329	0	0.332	0	0.329	0.624	0.328	0	0.441	0.553
s.d.	0.087	0.063	0.135	0.146	0.237	0	0.236	0	0.237	0.08	0.237	0	0.208	0.145
P	0.514	0.089	0.146	0.426	1	0	1	0	0.685	0.293	1	0	1	0.029

Table B.3 Diversity of *ospC* alleles a mammalian host species is infected by determined by the Gini-Simpson index. Observed Gini-Simpson index values were calculated from the original *ospC* frequency data. Expected Gini-Simpson index values represent the mean Gini-Simpson index after randomizing *ospC* allele frequencies across infected hosts 9,999 times. The two-tailed P values for each of the alleles represents if the Gini-Simpson index is more extreme than expected by chance.

	<i>Mephitis</i>	<i>Microtus pennsylvanicus</i>	<i>Mus musculus</i>	<i>Peromyscus spp.</i>	<i>Sciurus carolinensis</i>	<i>Tamias striatus</i>
observed	0.667	0.76	0.611	0.757	0	0.877
Expected	0.568	0.765	0.711	0.832	0.423	0.813
sd.expected	0.145	0.066	0.093	0.02	0.181	0.036
P value	1	0.974	0.352	0.002	0.309	0.034

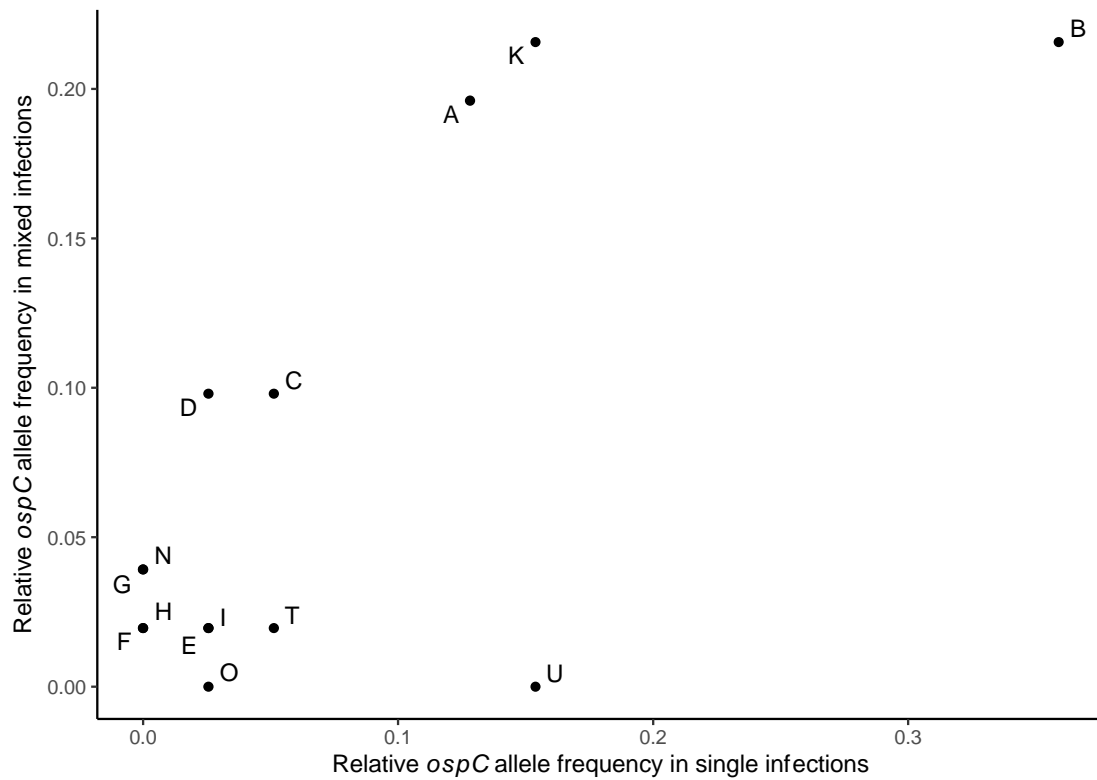


Figure B.1 Positive correlation between *ospC* allele relative frequencies in single infections (i.e., one allele in an infection) and mixed infections (more than one allele present in an infection). Allele identities are presented beside their points on the figure. Despite being relatively common in single infections, allele U was not found in mixed infections.

Appendix C

IACUC APPROVAL



Institutional Animal Care and Use Program
Rutgers, The State University of New Jersey
<https://orra.rutgers.edu>

RWJMS Research Tower
675 Hoes Lane West, Room 115
Piscataway, New Jersey 08854

ADMC 14, Room 1421
30 Bergen Street
Newark, New Jersey 07107

APPROVAL OF SUBMISSION

May 6, 2021

[Dina Fonseca](#)
Headlee Research Lab
New Brunswick, NJ 08901-8536

848-932-3146
dina.fonseca@rutgers.edu

Dear Dina Fonseca:

On 5/3/2021, the IACUC reviewed the following submission:

Type of Review:	Amendment
Title of Protocol:	Amendment for PROTO201900163
Investigator:	Dina Fonseca
IACUC ID:	AMEND202100256

The protocol is approved from 5/3/2021 to 11/18/2022. Your next annual review is due 11/18/2021. Your next triennial review is due 11/18/2022.

Sincerely,

The Institutional Animal Care and Use Committee (IACUC)

Appendix D

PERMISSIONS

Chapter 1, “Prevalence of *Borrelia burgdorferi* and diversity of its outer surface protein C (*ospC*) alleles in blacklegged ticks (*Ixodes scapularis*) in Delaware” has been published in the open access journal Ticks and Tick-borne Diseases. This work is under a creative commons license

(<https://www.sciencedirect.com/science/article/pii/S1877959X23000213#:~:text=a%20Creative%20Commons-,license,-open%20access>) with the rights and content link available here:

<https://www.sciencedirect.com/science/article/pii/S1877959X23000213#:~:text=Get%20rights%20and%20content>.

