Phylogeography of the blacklegged tick Ixodes scapularis throughout the United States identifies candidate loci for differences in vectorial capacity

Julia Frederick[1], Alec Thompson[1], Prisha Sharma[1], Guha Dharmarajan[1], Isobel Ronai[2], Risa Pesapane[3], Ryan Smith[4], Kellee Sundstrom[5], Jean Tsao[6], Holly Tuten[7], Michael J. Yabsley[1], and Travis Glenn[1]

[1]University of Georgia
[2]Harvard University
[3]The Ohio State University
[4]Iowa State University
[5]Oklahoma State University
[6]Michigan State University
[7]University of Illinois Urbana-Champaign

August 23, 2022

Abstract

The blacklegged tick, Ixodes scapularis, is a vector of Borrelia burgdorferi sensu stricto (s.s.), the causative agent of Lyme disease, part of a slow-moving epidemic of Lyme borreliosis spreading across the northern hemisphere. There are well-known geographic differences in the vectorial capacity of these ticks associated with genetic variation. Despite the need for detailed genetic information in this disease system, previous phylogeographic studies of these ticks have been restricted to relatively few populations or genetic loci. Here we present the most comprehensive phylogeographic study of I. scapularis conducted by using 3RAD and surveying 353 ticks from 33 counties throughout the range of I. scapularis. We found limited genetic variation among populations from the Northeast and Upper Midwest, where Lyme disease is most common, and higher genetic variation among populations from the South. We identify four genetic clusters of I. scapularis that are consistent with four major geographic regions, plus a distinct Central Florida group. In regions where Lyme disease is increasing in frequency, the I. scapularis populations genetically group with ticks from historically highly Lyme-endemic regions. Finally, we identify ten variable DNA sites that contribute the most to population differentiation. These variable sites cluster on one of the chromosome-scale scaffolds for I. scapularis and are within identified genes. Our findings illuminate the need for additional research to identify loci causing variation in the vectorial capacity of I. scapularis and where additional tick sampling would be most valuable to further understand disease trends caused by pathogens transmitted by I. scapularis.

Introduction

Determining the appropriate delineation of populations and species is critical to understanding many biological systems. Several different species concepts are being used depending on the characteristics of the organisms of interest and the lens of researchers studying them (de Queiroz, 2005). Accurately identifying the differences across closely related species, clades, and populations is beneficial to assessing the existing diversity of the group in question (Sokal & Crovello, 1970). Vectors offer opportunities to study complex ecological interactions among biological groups with varying levels of differentiation and co-varying levels of association. In vector biology, these differences can significantly change the potential risk for vector-borne

pathogen transmission within different environments (Arsnoe et al., 2019; Dolan et al., 1998; L. Eisen et al., 2017), and the organismal classification can influence how public health officials track disease risk in a region (Lane et al., 1991; Oliver et al., 1993).

The blacklegged tick, *Ixodes scapularis* , is a vector in the eastern and midwestern U.S. for at least seven human pathogens including*Anaplasma phagocytophilum, Borrelia mayonii, Borrelia miyamotoi, Ehrlichia muris eauclarensis, Babesia microti* , Powassan virus, and most notably, *Borrelia burgdorferi sensu stricto (s.s.)* , a causative agent of Lyme disease (R. J. Eisen & Eisen, 2018). Lyme disease is a nationally notifiable condition in the U.S. (CDC, 1990) and cases have been continuously increasing for years (Bacon et al., 2008; Division of Vector-Borne Diseases, 2019; Rosenberg et al., 2018). More than 95% of human Lyme disease cases occur in the northeast and the upper Midwest regions of the U.S. (Division of Vector-Borne Diseases, 2019; Kuehn, 2013), with a low incidence of cases occurring throughout the rest of the U.S. (Figure 1). While the majority of human Lyme disease cases in the U.S. are due to transmission of *B. burgdorferi s.s.* by*I. scapularis,* other *Ixodes* species have been shown to be competent vectors (Couper et al., 2020; Fleshman et al., 2021).

The taxonomy of *I. scapularis* has been debated for many years. Currently, *I. scapularis* is subclassified to the *Ixodes ricinus* species complex, notably containing the main vectors of Lyme-causing *Borrelia* spirochetes. From 1979 (Krinsky, 1979) until 1996 (Oliver et al, 1993), there were three tick species from the*I. ricinus* -species complex located within the U.S., that were associated with transmitting *B. burgdorferi s.s.* : *Ixodes pacificus* in the West, *I. scapularis* in the Southeast, and*I. dammini* in the Northeast (Lane et al., 1991). However,*I. scapularis* were publicly believed to not be a large threat as they were thought to rarely bite humans (Oliver, 1996). Due to the geographical closeness and the intermediate characteristics found in nature of *I. dammini* and *I. scapularis,* the validity of the delination between the two was questioned (Spielman et al., 1979; Oliver et al., 1993). Laboratory mating experiments determined them to be conspecific species and *I. dammini* was synonymized with*I. scapularis* (Oliver et al., 1993). Once *I. dammini* , which was previously considered the main threat, was no longer considered a valid species, the concern for and surveillance of Lyme disease increased because the main vector, *I. scapularis* , is spread throughout the eastern U.S. (Oliver et al., 1993; Sanders, 1998). This new perception of *I. scapularis* led to questions on how populations of this species differ, especially when it came to Lyme disease prevalence in a region. Subsequent genetic studies redefined*I. scapularis* into two distinct clades, delineating Northern populations (American clade) from Southern populations (Southern clade) (Norris et al., 1996).

Researchers have been using the American and Southern clades as a general way to classify *Ixodes scapularis* populations while the geographic range of the species has been expanding to the north and west. Populations have been documented in almost every state east of the Rocky Mountains while newly established populations are filling in previously uninhabited counties throughout the known range and can now even be found as far north as Québec (R. J. Eisen et al., 2016; Fleshman et al., 2021; Khatchikian et al., 2015; Ripoche et al., 2022). This continual range expansion is increasing the potential risk for Lyme disease throughout North America (Kelly et al., 2014; Kugeler et al., 2015; Leighton et al., 2012; Ogden et al., 2009). The Lyme endemic region is also expanding outward from the historical locations in the Northeast and upper Midwest, with some counties south of Virginia reporting increases in *I. scapularis* abundance (R. J. Eisen et al., 2016; Hickling et al., 2018) concomitant with increases in Lyme disease cases in humans over the past two decades (Division of Vector-Borne Diseases, 2019). States within the Ohio River Valley region are experiencing some of the highest rates of Lyme disease case spread in the U.S. For example, Lyme disease cases in Ohio have increased 10-fold over the past decade, likely due to the range convergence of the Northeastern and Upper Midwestern *I. scapularis* populations (Bisanzio et al., 2020).

The presence of *I. scapularis* alone is not enough to predict how rapidly Lyme disease risk can increase in a region because Lyme disease ecology is complex and involves multiple intersecting factors within the tick life cycle, such as dominant tick blood meal hosts and reservoir vertebrate hosts of *B. burgdorferi* (Bisanzio et al., 2020; Burtis et al., 2016; Gardner et al., 2020; Ogden et al., 2013). These intersecting factors all contribute to the vectorial capacity of*I. scapularis* for *B. burgdorferi* , defined as all "vector-related variables affecting

stability of pathogen transmission" (Spielman et al., 1984). In addition to variation in host preference, a key determinant of pathogen transmission by *I. scapularis* is likely in the host-seeking behaviors themselves, such as time spent host-seeking, or the height to which ticks will climb for host-seeking. Population level variation of host-seeking could be due to changes in humidity, because ticks are extremely sensitive to desiccation, and host community composition (i.e., mice are the most common host of tick larvae in the north, whereas lizards are important hosts in the south (Apperson et al., 1993; Ostfeld & Keesing, 2000)). However, previous research has also shown that genetic factors are likely contributing to a tick's ability to acquire and transmit pathogens and genetics could even play a role in host-seeking behavior differences (Arsnoe et al., 2015; Ginsberg et al., 2017; Ginsberg et al., 2014). Thus, the genetic composition of the expanding *I. scapularis* populations merits consideration.

While there has yet to be a study directly linking tick host-seeking behaviors to genetic variants, underlying genetic mechanisms driving population-level differences in *I. scapularis* survivability and host-seeking behaviors have been hypothesized. The consistency of these behaviors within Northern and Southern *I. scapularis* populations when exposed to varying environments suggests that host-seeking behavior is driven in part by genetics, and not solely from environmental pressures (Arsnoe et al., 2015; Tietjen et al., 2020). Identifying the populations that are leading the range expansion fronts, like those in the Ohio River Valley and the Virginia-North Carolina border, will provide public health officials with an informed view of tick-borne disease risk in currently non-Lyme endemic areas experiencing invasion (or possibly re-invasion) by blacklegged ticks. However, due to the difficult and time-consuming efforts needed to directly measure vectorial capacity of ticks, it is far more efficient to identify genetic markers that are associated with tick populations that have high or low human Lyme disease prevalence. If the markers are sufficiently dense and spread throughout the genome, then it may also be possible to identify loci associated with the traits of interest (Anandan et al., 2022; Aubry et al., 2020; Dupuis & Siegmund, 1999; Rahman et al., 2021).

Research to increase our understanding of the population genetics of*I. scapularis* has occurred for many decades. The main limitation so far has been that the studies either do not cover a wide portion of the genome or the wider geographic range of the species. They also cannot be easily compiled to accomplish both aspects as the targets and methodologies are non-compatible. Nonetheless, they have given insight into how populations of *I. scapularis* vary genetically, concluding there is significant genetic differentiation across the landscape, especially when comparing northern and southern populations (Gulia-Nuss et al., 2016; Ludwig, 2015; Norris et al., 1996; Van Zee et al., 2015). A study covering a large portion of the genome and known range of *I. scapularis* populations is needed to characterize the extent of differentiation within this species, as well as advancing an understanding that links phenotypes to genotypes.

Here, we investigate genome-wide variation within and among populations of *I. scapularis* grouped into four geographic regions that approximate major divisions described or hypothesized in previous research: Southeastern Atlantic, Southern Gulf, Upper Midwest, and Northeastern U.S. We aim to provide a basis for future research in uncovering how genetic variation among *I. scapularis* populations is driving observed phenotypic differences, thus driving Lyme disease prevalence in the U.S. Understanding the genotypic variation within and among *I. scapularis* populations, especially those at the fronts of range expansion and phenotypic change, is essential to understanding the ecology of these ticks and pathogens, which can mitigate Lyme disease risks and serve as a model system for disease ecology.

Methods

Tick Collections

We used two collection strategies to obtain *I. scapularis* adults from across the eastern half of the U.S. First, we established a network of collaborators among research institutions in the U.S. Collaborators shipped previously identified samples from past collection seasons (2010-2020) stored in 70% ethanol to the University of Georgia (UGA) Environmental Health Science (EHS) DNA laboratory for sequencing. These tick samples were collected from either drag sampling or animal hosts, with hunter-harvested white-tailed deer (*Odocoileus virginianus*) transported to check stations or local meat processors being the most common type of animal

host collection. For the second collection strategy, we implemented a community science approach to cover areas that did not have large research collection efforts established. Specifically, New England K9 Search and Rescue team, the TNC Nashua Grasslands, and other non-associated volunteers were emailed requests to submit ticks, while the Quality Deer Management Association (now National Deer Alliance) published a request in their newsletter to assist us through this community science effort. All community scientists who expressed interest in the project were mailed a tick collection kit, which included a data sheet, tweezers, tubes with 70% ethanol, a permanent marker, and a pre-paid return envelope. Community scientists were instructed to collect all ticks encountered from game, humans, and dogs during the 2020-2021 hunting seasons. These collection kits were returned to UGA for morphological identification using standard keys (Durden & Keirans, 1996; Keirans & Litwak, 1989). Only flat *I. scapularis* adults that were not significantly damaged and had a listed source location were used in this study. No samples that had been attached to humans were accepted (unattached ticks crawling on humans were accepted).

Defining Populations

For this study, collected ticks were grouped into shared geographical regions (i.e., putative populations) and collection time periods (Table 1, Figure 1). A population was defined as at least seven ticks collected within the same year and the same collection area (Davey & Blaxter, 2010). Collection areas in this study were defined at the county level and will be referred to by county name; however, for animal and passive collections we assumed that ticks in a given population may have originated in surrounding counties due to the travel of hosts and hunters. One population included in this dataset, from Owen County, Kentucky which shares a border with the sample population from Franklin County, Kentucky, has only four individuals. Our original plan was for collection sites sharing a county border to be combined as a single population for analysis, but during preliminary analysis (see below) we determined that ticks from these two counties had different genetic signatures, so we retained them as separate populations.

For reference purposes and data visualization only (i.e., not analysis), populations were further defined into quadrants of the eastern U.S. These divisions were designed to divide the collection area of *I. scapularis* in this project into four areas of similar size, while considering the two proposed convergence zones in the Ohio River Valley and at the boarder of southern Virginia. We separated north from south at latitude 36.5407 following the northern borders of North Carolina, Tennessee, and Arkansas, and then separated east from west at longitude -84.7856 following the border of Ohio and Indiana. This created our quadrants referred to as Northeast (>36.5407 and >-84.7856), Upper Midwest (>36.5407 and <-84.7856), Southeastern Atlantic (<36.5407 and >-84.7856), and Southern Gulf (<36.5407 and <-84.7856) (Figure 1).

DNA Isolation

All *I. scapularis* samples were given a unique identifier and bisected along the medial line. We retained one half of each tick in 70% ethanol, while the other half was used for DNA isolation using the Qiagen DNeasy Blood and Tissue kit (Hilden, Germany) (Ammazzalorso et al., 2015). DNA was quantified using a Qubit Fluorometer (Life Technologies, Inc., California, USA), and all DNA samples were normalized to 5-9ng/µL prior to library preparation.

RADseq Library Preparation and Sequencing

Quadruple indexed 3RAD libraries were created for every individual sample following a protocol modified from Adapterama III (Bayona-Vasquez et al., 2019) using Xbal, EcoRI-HF, and NheI-HF enzyme combination (New England Biolabs, Inc., MA, USA). In this protocol, XbaI and EcoRI-HF provide the cut sites for adapter ligation to the target DNA, whereas NheI-HF is used to cut adapter-dimers and increase efficiency of library preparation which is particularly important when only limited amounts of target DNA are available. Modifications were as follows: target DNA concentration was at a lower starting concentration than the recommended 20ng/µL, digestion time was increased from one hour to two hours, ligation cycles were increased to seven total cycles, and libraries were not cleaned between ligation and PCR steps as this does not impact library quality while saving time and reagents.

Libraries were completed by amplifying the uncleaned ligation product with 13 cycles of PCR using iTru5 and iTru7 primers (primer sequences from Glenn et al., 2019). We then visualized the libraries on a 1.5% agarose gel to ensure an appropriate size-distribution of library molecules and to screen for irregularities, which we did not observe. Individual barcoded libraries were pooled and cleaned using Speed-Beads at a 1:1.25 ratio (Pool:Speed-Beads). The cleaned pool was further size selected on a Pippin Prep (Sage Science, MA, USA) using a 1.5% dye-free Marker K agarose gel cassette (CDF1510) set to capture fragments of 550 bp +/-10%. P5 and P7 Illumina primers were used in a final PCR followed by another cleaning with Speed-Beads (see Glenn et al. 2019 for details). These pools were sequenced on an Illumina HiSeqX to obtain PE150 reads (NovoGene Co., Sacramento, CA).

Locus Identification and Filtering

Sequences were demultiplexed by NovoGene Co. using index sequences at the standard Illumina indexing positions (from the iTru5 and iTru7 primers), then cleaned and analyzed by us using Stacks v2.5 (Catchen et al., 2013). To clean the sequences, we used process_radtags, providing the restriction site and internal tags along with the parameters (-c) removing any read with an uncalled base, (-q) trimming low-quality bases using the default setting of a sliding window and a raw phred score of 10, and (-t 140) truncating reads to 140 base pairs. Cleaned sequences were aligned to the *I. scapularis* genome (GenBank: GCA_016920785.2 ASM1692078v2) using BWA-MEM v0.7.17 (Li & Durbin, 2009).

Alignments were filtered using Samtools v1.10 (Danecek et al., 2021) by keeping only uniquely mapping reads with qualities over 25 and removing unmapped reads or reads containing 5 or more variants per read. We then assembled mapped RAD loci stacks using the ref_map.pl program in Stacks and called the single nucleotide polymorphisms (SNPs) from these groups of mapped loci. To be included in the output, loci needed to be present in at least 60% of individuals. The SNP output from Stacks was exported into a variant call format (VCF) file.

VCFtools v0.1.16 (Danecek et al., 2011) was used to further filter our data to have a minor allele frequency of [?]5%. We removed loci that were missing from more than 20% of all individuals, and loci that had coverage below 6x or above 200x. All libraries were analyzed for quality and missing data, if a library did not have more than 50% of the high-quality loci present it was removed from the analysis due to poor library quality. Only 19 samples were removed, and the Stacks analysis pipeline and filtering above was re-run on the final 353 analysis-quality libraries. After all filtering steps the final dataset consists of 7,274 polymorphic loci. All analyses presented here are based on these loci.

Mitochondrial DNA amplification and sequencing

To amplify the 16S rRNA gene of samples from Osceola County, Florida, we used the protocol described in Lv et al. 2014, using rDNA primers of 16S-F 5'-TTAAATTGCTGTRGTATT-3' and 16S-R1 5'-CCGGTCTGAACTCASAWC-3'. Amplification protocol was as followed: initial denaturation (94°C, 5 minutes); followed by five cycles of 94°C for 30 s, 49°C for 30 s, and 68°C for 30 s; five cycles of 94°C for 30 s, 47°C for 30 s, and 68°C for 30 s; five cycles of 94°C for 30 s, 45°C for 30 s, and 68°C for 30 s; 25 cycles of 94°C for 30 s, 43°C for 30 s, and 68°C for 30 s; followed by a final extension step of 68°C for 5 minutes. All positive amplicons are submitted for bi-directional sequencing at Genewiz Corporation (South Plainfield, NJ).

Genetic Analysis

We calculated pairwise F_{ST} values for all populations using HIERFSTAT v0.04-22 function pairwise.neifst (Goudet, 2005). Site-specific statistics were calculated by running the 'populations' program of Stacks on the filtered VCF file, using the –fstats argument to compare select populations from each quadrant. Genetic differentiation between states was assessed with AMOVA from R package Popper (Kamvar et al., 2014). Basic diversity statistics for each population were calculated using HIERFSTAT v0.04-22, and ADEGENET v2.1.3 (Jombart & Ahmed, 2011), including number of alleles ($N_{alleles}$), allelic richness (AR), observed heterozygosity (H_o), mean gene diversity (H_s), and inbreeding coefficient (F_{IS}). Following this, we described

clustering and membership probabilities for all samples utilizing their assigned populations by running discriminant analysis of principal components (DAPC) from ADEGENET. A STRUCTURE-like analysis was then completed using R package LEA (Frichot & François, 2015) to identify clusters and calculate membership probabilities without inputting population information (Jenkins et al., 2021).

Due to noticeable uniqueness of samples from Osceola County, Florida, the species identifications of these samples were confirmed by sequencing fragments of the 16S rRNA gene (Lv et al., 2014). The phylogenetic tree was created using the 16S rRNA sequences obtained from Osceola County, Florida samples, and sequences from Genbank that were previously assigned to the American or Southern Clade for *I. scapularis* , or positively identified as *Ixodes affinis* . Sequences were aligned in Geneious Prime 2022 (Auckland, New Zealand) using the Geneious alignment algorithm with the following parameters: Global alignment with free end gaps, 65% similarity (5.0/-4.0) cost matrix, 12 for gap open penalty, 3 for gap extension penalty, and doing 2 refinement iterations. The tree was generated using the maximum-likelihood algorithm with 500 iterations in MEGA 11 with*Rhipicephalus sanguineus* (L34302) used as the outgroup (Stecher et al., 2020). The final tree was produced using the R package ggtree (Yu et al., 2017).

Isolation by distance was also tested for all populations, using Mantel tests (Wright, 1943) following the tutorial for the ADEGENET R package (Jombart & Ahmed, 2011). The coordinates used for the geographic distances are provided in Table 1. Comparisons were then visually grouped to show Northern-Northern, Northern-Southern, and Southern-Southern comparisons, as determined by the LEA analysis where populations in clusters 1 and 5 are labelled as Northern, and those in clusters 2, 3, and 4 are Southern.

Results

Tick Collections and Community Science

Throughout the course of this study, we received 94 responses from interested volunteers, who were all subsequently mailed a tick collection kit. We had 46 collection kits returned (48.9% return rate). Returned kits covered 19 states and 43 counties within the known range of *I. scapularis* . Only three of these counties overlapped with counties represented in our collaborative researcher-based samples. Community scientists were aware that this project focused on *I. scapularis,* but to simplify the process, we requested volunteers to collect all ticks encountered. This structure meant we received other species, life stages, and engorgement levels. Through community science collections we gathered 556 *I. scapularis* , and 53 unidentifiable*Ixodes* spp. due to damage to the specimens (Supplemental Table 1).

Using our dual collection strategies, we gathered a total of 1,453 flat, undamaged *I. scapularis* adults from 25 states and 82 counties. Researcher-based collections provided 1,034 individual *I. scapularis* ticks (291 females, 743 males), and community scientists contributed 419 flat, undamaged *I. scapularis* (135 females, 284 males). By combining ticks from both collection strategies, we obtained 372 flat, adult *I. scapularis* from 33 counties in 22 states, with [?]7 individuals per county (except Owen Co., KY; n=4), from which we made 3RAD libraries. Nineteen ticks were excluded due to poor library quality, leaving a total of 353 ticks in our final dataset. The 33 counties represent the 33 putative populations in our dataset that are distributed across the four quadrants: Northeast (n=12), Upper Midwest (n=9), Southern Gulf (n=6), and Southeastern Atlantic (n=6) (Figure 1, Table 1).

Data Processing

DNA sequencing was completed over four Illumina HiSeq runs with a total of 2,258,652,346 reads, targeting at least 2 million reads per individual (average reads per individual = 6,398,449). After cleaning, mapping, and filtering the raw reads to have a quality level of at least 25 and no more than 5 SNPs per read, we had a total of 301,255,910 paired mapped reads (average per individual = 853,416) for the Stacks analysis. Stacks identified 741,532 variable sites, which were then filtered to a total of 7,274 variable sites (Supplemental Table 2). These filtered, high-quality sites form the dataset used in all further analysis unless otherwise noted.

Genetic Diversity and Differentiation

Genetic variation among *I. scapularis* populations differed depending on geographic location of the populations being compared. Our overall F_{ST} value was 0.176, with paired population F_{ST} values ranging from 0 – 0.417. Northern-to-Northern comparisons had a mean F_{ST} of 0.049 (min = 0, max = 0.302, median = 0.029), and the Southern-to-Southern comparisons had a mean F_{ST} of 0.189 (min = 0.002, max= 0.367, median = 0.175) (See Supplemental Table 3 for all pairwise F_{ST}values). These comparisons showed that Northern populations are more closely related to other Northern populations than Southern populations are to each other.

To further investigate population differentiation, focal populations from each quadrant were chosen based on their dispersed geographic locations and how well they represented their quadrant (Southeastern Atlantic quadrant represented by Aiken County, SC and Osceola County, FL; Southern Gulf quadrant by Rapides County, LA; Upper Midwest quadrant by Monroe County, WI; and Northeastern quadrant by Washington County, RI), and site-specific statistics were then calculated. Comparing populations across the Northern and Southern boundary produces a higher level of genetic differentiation across more loci, and Southern populations have more differentiation at more loci with other Southern populations, than when you compare Northern to Northern (Supplemental Figures 1A-C).

Basic diversity statistics (Table 1) confirm genetic variation within all populations. In general, Northern populations have a slightly higher level of allelic richness than those in the South, and populations close to the dividing boundaries of our quadrants have higher levels than populations farther to the north or south. The Southern populations also have a slightly higher degree of inbreeding (average=0.27) compared to northern populations (average=0.22). Additional results, including AMOVA and Hardy-Weinberg equilibrium analyses, are given in the supplemental text.

Genetic Structure

The DAPC analysis was completed retaining the first 19 principal components as determined by the optimum a-score, and 19 discriminant functions which explained 30.5% of the variance (Figure 2). The scatterplot shows a main cluster that consists of 23 of the populations including all from the Northeastern quadrant, most of the Upper Midwest quadrant, Stokes and Watauga Counties in North Carolina, and overlaps with Claiborne County, Tennessee. The Southern Gulf quadrant is the only quadrant that does not have populations in the main cluster. The populations in the Southeastern Atlantic quadrant that are closest to the coastline are the furthest from the main cluster with the southernmost population from Osceola County, Florida being the most distant.

These trends were further examined using R package LEA to assess population structure of all individuals, outside of the pre-assigned populations. The cross-entropy criterion determined K=5 is optimal (Supplemental Figure 2). The five clusters can be described as assorting into Northeast, Upper Midwest, Central, Southern, and Central Florida (Figures 3, 4A). Moderate levels of admixture were apparent for most populations. However, Southern populations had lower levels of admixture with individuals mostly being assigned cluster 2, 3, or 4 and low assignments to all other clusters (Figure 3). Every tick had membership probabilities calculated for each of the five clusters, and a main cluster was assigned to each individual by finding which cluster had the highest membership probability, then the DAPC analysis redone. This iteration showed a similar pattern to the original population based DAPC, with one group being noticeably larger and consisting mainly of individuals from the North (clusters 1 and 5). There is a small distant group that is composed of the ticks from Central Florida (cluster 3), and then a group that shares no overlap with the Northern group which is primarily composed of ticks from the Southern Gulf (Figure 4A).

In this DAPC analysis we identified the top six alleles across four scaffolds that were contributing to differentiating the clusters the most (NW_024609857.1_scaffold2 position 192156357; NW024609868.1_scaffold3 position 60000181; NW024609879.1_scaffold4 positions 31884195, 31884182, 66253319, 66253456; and NW_024610117.1 position 26212) (Supplemental Figure 4). Some of these positions fall within annotated genes in the recent *I. scapularis* genome (GenBank: GCA_016920785.2 ASM1692078v2), specifically gene loci LOC8029585, LOC120844762, and LOC121834473. When examining the allele representation of all clus-

ters at these sites, cluster 3 (Central Florida) always has the opposite allele frequency to all other clusters (Supplemental Figure 4).

While morphological identifications confirmed the Central Florida ticks as *I. scapularis* , Florida has populations of *I. affinis*ticks that can be morphologically similar to *I. scapularis* . Thus, we confirmed the species identity of the ticks collected from Central Florida using 16S rRNA, finding that they are most similar to other*I. scapularis* individuals from the Southern clade (Figure 4C). However, to determine the alleles contributing to the differentiation of the other four clusters, ticks within the Central Florida cluster were removed from the dataset and the analysis rerun. In doing so, our DAPC shows more differentiation among the remaining four clusters, especially noting that ticks from the Upper Midwest and Northeast no longer completely overlapped (Figure 5B). We examined the ten SNPs that contributed the most to the differentiation of these 4 groups and found these were all on scaffold 3 (NW024609868.1 positions 42246251,42246253, 50953023, 72034653, 72034688, 72034691, 72034789, 72034790, 72034791, and 72034815; more information can be found in Supplemental Table 6). There are two other groups of SNPs, comprised of loci that are noticeably contributing to the differentiation of the four geographic clusters, located on scaffolds 4 (NW024609879.1) and 8 (NW_024609883.1) (Supplemental Figures 5-6).

The isolation by distance analysis was completed using the original 33 populations designated for the ticks. We do not see any strong correlation between genetic and geographic distance of our populations when all comparisons are included. However, Northern-to-Northern comparisons have less genetic distance between populations even as geographic distance increases, whereas the Southern-to-Southern comparisons have a varied correlation of genetic to geographic distance. The Northern-to-Southern comparisons are distinctly higher in genetic distance across all geographical distances (Figure 5).

Discussion

The *I. scapularis* and *B. burgdorferi* disease system is a model for ecological disease research. With the slow-moving epidemic of Lyme disease spreading throughout the U.S., studying this system creates a unique opportunity to describe how vector population genetics can be linked to changes in pathogen prevalence. With *I. scapularis*having notable genetic and behavioral differences associated with Lyme disease rates, identifying candidate loci and genomic regions that correlate with disease transmission gives researchers new opportunities to understand ecological disease dynamics and public health agencies the opportunity to track Lyme disease risk by genotyping relatively small numbers of ticks. Below, we discuss our overall goals to: 1) build and use a system to efficiently collect and genotype a large number of SNPs in many *I. scapularis* individuals, 2) better understand the amount of genetic variation within and among populations of *I. scapularis* , determining if variation was structured in relation to geographic regions or human Lyme disease cases (a proxy for*Borrelia burgdorferi s.s.* prevalence) and providing prospective genetic markers for vectorial capacity by identifying sites or genomic regions that may be driving differences and 3) estimate where contact zones between major genetic groups may exist.

1.a. Tick Collections and Community Science

Large-scale sampling can pose many challenges for researchers and incur substantial personnel costs. By employing a mixed collection strategy across the known geographic range of *I. scapularis,* our study was able to obtain broad sampling coverage without excessive cost or time input. Obtaining tick samples from research scientists provided consistent collections from a variety of locations and using their collections for this study provides a model for leveraging the efforts of routine tick surveillance programs into fundamental investigative research. Working with community scientists for sample collection in areas without a strong research presence had challenges. For example, not having control over which counties were targeted in the collections; some collection kits being returned with few ticks making it difficult to have a full population represented; and the added responsibility of keeping in contact with volunteers throughout the collection period. However, community scientists allowed us to scale-up the populations sampled significantly from previous *I. scapularis* population genetics studies and provided an opportunity to increase awareness about the risk of tickborne disease in many communities.

Our dataset is not balanced regarding sex ratios or populations per region. We have more male than female samples, however, completing analyses with a single sex did not produce significantly different results, and keeping both sexes represented allowed more populations to be analyzed. Tick collection efforts are higher in the North, especially the Northeast, due to high Lyme disease incidence and public health surveillance, making samples more readily available. The unevenness in sampling location does not appear to impact results, because all trends hold even with sub-selected populations to make balanced representations of the quadrants (analyses not shown).

1.b.3RAD genotyping of Ixodes scapularis

Before determining an optimized library preparation protocol, libraries created following normal 3RAD protocol parameters had larger DNA insert sequences than desired and less library quantity than expected. This led us to increase digestion time to shorten DNA insert size and increase ligation cycles to have a better final library yield. The resulting library preparation is relatively quick, easy, and low-cost (easy to process [?]1 set of 96 DNA samples per 2 days at [?]$5/sample; Bayona-Vasquez et al., 2019). Although the data in this work were accumulated via multiple partial HiSeqX lanes, the total number of reads is roughly equivalent to what could be obtained from a single NovaSeq S4 lane (˜$5,000 U.S., in many academic core labs).

The resulting sequence quantity and quality from these libraries were typically correlated to extracted DNA quality (poorer quality DNA leading to poorer quality libraries). Only 19 samples needed to be removed from the final dataset due to poor quality libraries (missing >50% of loci). There are three *I. scapularis* genome assemblies, each an improvement on the previous version. Variant loci were more readily detected with each improving assembly because more sequences were mapping to the genome assembly. For this project, we used the newest reference genome (GenBank: GCA_016920785.2 ASM1692078v2), which is likely to have made large positive difference in the quality of SNPs used and our ability to identify loci of interest.

2.a. Variation within populations

Genetic variation is abundant within all populations, but Northern and Southern populations have noticeable differences when individual populations are analyzed. Northern populations have slightly less inbreeding and slightly higher allelic richness on average than individual Southern populations, potentially indicating that Northern populations are interbreeding across larger distances, creating more genetically cohesive units across the northern U.S. The higher levels of allelic richness are also believed to be a good indicator of a population's ability to adapt to future environmental changes (Caballero & Garcia-Dorado, 2013), further supporting the hypothesis that Northern populations are leading the geographic expansion of *I. scapularis*in North America. Meanwhile, populations at the dividing boundaries of our quadrants (e.g., North Carolina and Ohio) have the highest levels of allelic richness, likely resulting from the Northern and Southern genetic groups interbreeding in these areas. In contrast, Southern populations have higher degrees of inbreeding, indicating there is less gene flow occurring among collection localities sampled in the south, creating more distinctive genetic groups.

2.b. Variation among populations

In concurrence with previous population genetic studies of *I. scapularis* (Gulia-Nuss et al., 2016; Ludwig, 2015; Norris et al., 1996; Van Zee et al., 2015), we found an overall moderate amount of genetic differentiation throughout the species range while expanding the geographic and genomic regions surveyed. We were also able to observe varying amounts of differentiation when examining different population comparisons.

2.b.i. Genetic structure relative to geographic regions

Researchers have been studying the variation of *I. scapularis*populations by identifying phenotypes that vary geographically and correlate with human Lyme disease prevalence (Arsnoe et al., 2019, Ginsberg et al, 2014). Questing height has been demonstrated to be correlated with population origin via common garden experiments, remaining correlated in subsequent generations, thus supporting the notion that this important host-seeking behavior has one or more genetic drivers (Arsnoe et al., 2019). However, genotypes or a genomic

region of interest have not yet been linked to these phenotypes. Thus, a major goal of this study was to develop a foundational database of genetic variants in *I. scapularis* populations that are located throughout the genome and can potentially be used in future studies to link genotypes to phenotypes of interest.

Genetic structure was originally analyzed through the lens of the four geographic quadrants (Figure 1). Differentiation is modest when comparing the Northeast to Upper Midwest or comparing the Southeast Atlantic to Southern Gulf. However, outside of the population originating from Osceola County, Florida, most genetic differentiation occurred in comparisons of Northern populations to Southern populations. These two major classifications can simplify the lens for how these results are viewed and mirrors the original delineation between *I. scapularis* and *I. dammini* .

2.b.ii. Southern-to-Southern Comparisons

Southern populations have the highest level of genetic diversity among populations in the sampling region and is notably not related to the geographic distance of populations. The DAPC results show many populations in the south form their own distinct units, except for a few populations that are geographically close to each other with few environmental barriers (e.g., no mountain ranges), like the Louisiana counties. The Georgia and Alabama populations are in the geographic and analytical center of our Southern populations. The central location of these populations is consistent with the hypothesis that *I. scapularis* originated in Mississippi and Georgia before expanding out to Florida and North Carolina 500,000 years ago (Van Zee et al., 2015).

The locus-level analysis between Southern populations further demonstrated the high levels of genetic diversity throughout the southern U.S. Aiken County has the lowest level of differentiation relative to other comparisons involving Osceola County (pairwise $F_{ST} = 0.194$), but there is a moderate number of loci that are still significantly different between the two populations. This is not the case for other southern population pairs that are more closely related, such as Aiken County and Rapides County (pairwise $F_{ST} = 0.065$).

We have further identified that the Southern clade, which was originally classified as one distinct unit occurring across several geographic regions, actually consists of multiple identifiable clades. This is seen in the LEA analysis, as the individuals in the south still assorted into three distinct groups out of our five major clusters, even without any associated metadata. The top six variant sites contributing the most to differentiating the five clusters are focused on differentiating Osceola County from the rest of the individuals. Osceola County samples have allele frequencies in opposite proportion to all other populations. The variant at position NW024609868.1:60000181 is within the annotated gene LOC8029585. The studied function of this gene is as a neuropeptide SIFamide receptor which is theorized to affect salivary gland regulation and/or reproductive behavior regulation (Simo et al., 2013). Changes in the functions of salivary gland regulation could affect how *B. burgdorferi* is transmitted, as the bacteria passes through the salivary glands to reach the host (Schwan & Piesman, 2002). However, we cannot be sure if the variant detected here affects function.

For this study, we assume that the Osceola County population is pure*I. scapularis* . The morphological and 16S rRNA data supported this assumption, but these individuals still form a distinct cluster with low levels of admixture. It is important to acknowledge that despite the current results, it is possible the Osceola County population is comprised of *I. scapularis* x *I. affinis* hybrids. Other studies have shown the ability of *Ixodes* species to hybridize (Kovalev et al., 2016; Patterson et al., 2017) but have not yet investigated *I. scapularis* and *I. affinis* . Further investigation of *I. scapularis* and *I. affinis* from Florida is warranted.

2.b.iii.Northern-to-Northern Comparisons

In contrast to the Southern populations, when all Northern populations are examined, there is less genetic diversity among them even at large geographic distances. The DAPC analyses further supports that the Northern populations are more similar than Southern populations because the Northeastern and Upper Midwest populations highly overlap in each iteration, only beginning to separate when the Osceola County cluster is removed.

When Osceola County is removed, the top ten differentiating sites for the four large clusters, are mainly

differentiating the Northeastern dominated group from the rest. This cluster has the inverse allele frequency to most, if not all, other clusters. All these sites are located on NW024609868.1_scaffold3 and can be found in three annotated genes (LOC8028448, LOC8052187, and LOC8029065). Gene LOC8028448 has been described as collagen alpha chain CG42342, and LOC8029065 has been described as a sphingosine-1-phosphate lyase (Simo et al., 2013). Defining the function of the noted genes in *I. scapularis* could potentially lead to understanding proximate causation of regionally variable behaviors and inference of selection pressures. Understanding how these variants could potentially impact disease spread is important on an even larger scale, as the Northern ticks are from the historically youngest populations only expanding to the area 20,000-50,000 years ago (Van Zee et al., 2015), and are hypothesized to still be expanding.

2.b.iv. Northern-to-Southern Comparisons

Only a few decades have passed since the northern dwelling *I. dammini* was assigned as a junior synonym of southern dwelling *I. scapularis* (Oliver et al., 1993). Since then, *I. scapularis* has been divided into clades to assist in defining the genetic and phenotypic differences among populations (Goddard et al., 2015; Norris et al., 1996; Sakamoto et al., 2014). In the population-level DAPC, the genetic diversity of southern and northern populations is apparent, as northern populations form one large cluster whereas the southern populations are distinct. This is further confirmed by the LEA-based DAPC analysis, in which sample metadata was not input, and each tick was assigned membership probabilities across five clusters which assorted into five major regions (Northeast, Upper Midwest, Central, Southern, and Florida/Southeastern coast). This created three clusters of southern ticks that do not overlap, and two overlapping clusters of northern ticks. The breakdown of these clusters is like the subdivisions found through the microsatellite analysis in Ludwig (2015) and can still be differentiated as broader Northern and Southern groups. The genetic and phenotypic differences found among the two groups, and even the five clusters, warrant further research into genetically defining *I. scapularis* populations to better understand how these clades impact vectorial capacity across the geographic range of *I. scapularis* . We were able to identify genomic sites that are correlated to genetic differentiation of northern and southern populations and have found regions of the chromosome-scale scaffolds that are highly variable. Both of which can be used as targets for further phenotypic-focused studies.

2.b.v. Lyme Endemic to Non-Lyme Endemic Comparisons

Rather than using geopolitical boundaries, it is also possible to view the data relative to the 95% human Lyme disease kernel, a proxy for*B. burgdorferi s.s.* prevalence. Overall, populations close to or within the Lyme disease kernel were more closely related than populations more distant from the kernel. For example, the populations in North Carolina and Tennessee are within the Southeastern region, with Stokes County, NC falling within the Lyme endemic region and Watauga County, NC and Claiborne County, TN just outside the Lyme disease kernel. All three, however, are within or extremely close to the genetic cluster containing all populations within the Lyme Disease kernel, irrespective of their distance to them (Figs. 2, 3). Alternatively, the Missouri population is in the Northern geographic area, but is well outside the Lyme disease kernel, with the genetic cluster for Missouri also well away from all other Northern populations (which are inside the Lyme disease kernel; Figs. 2, 3). This leads to the hypothesis that the genetic differentiation occurring throughout the landscape likely is not solely geographically driven, but instead is correlated with *B. burgdorferi s.s.* prevalence. Thus, the information presented herein may be used to redraw boundaries between genetically distinct groups of*I. scapularis* , relationships that seem to be correlated with human Lyme disease. Further work in this area is obviously warranted.

3. Hybridization and Range Expansion

Over the last few decades, *I. scapularis* has continued to establish new populations in counties and states that previously had recorded limited or no activity of *I. scapularis* (R. J. Eisen et al., 2016), including major expansion north through Canada (Clow et al., 2017; Ogden et al., 2006). These Canadian populations are likely to be offshoots of other Northern U.S. populations from migrating birds (Krakowetz et al., 2011; Scott et al., 2001), contributing to an increased risk of Lyme disease where risk had previously been low. Northern

populations are also converging in the Ohio River Valley region, where researchers are seeing significant increases of Lyme disease cases at the border of the Northeastern and Upper Midwestern populations (Bisanzio et al., 2020). Ticks in this region appear to be more evenly mixed genetically between the Northeastern, Upper Midwest, and Central clusters than other populations in our dataset (Fig. 2).

Range expansion is also a major concern in U.S. regions where the area was dominantly inhabited by genetically Southern populations, but now co-exist with Northern populations. Understanding how genes flow between Northern and Southern populations is important for determining how population differences could impact the maintenance and spread of the enzootic cycle of *B. burgdorferi s.s.* , and consequently, that of Lyme disease. However, directly measuring the vectorial capacity of tick populations is challenging, and it would be more efficient to identify genetic markers in *I. scapularis* that are associated with high or low human Lyme disease incidence. These markers would then facilitate tracking gene flow within these areas using hybridization zone modeling and next generation sequencing techniques. The loci identified herein provide potential candidate markers for such work.

A potential focal area for this work is the Virginia-North Carolina border which is hypothesized to be situated at the southern spatial expansion front for northern *I. scapularis* populations (Kelly et al., 2014). The North Carolina populations in our study are at this border and the outer edge of 95% of Lyme disease kernel. However, this region has many other environmental and ecological complexities to consider, such as, the varying Ecological Units (Carpenter et al., 1995) from east to west that could be affecting gene flow across the state (Van Zee et al., 2015). Using the tick mitochondrial 16S rRNA genes, Qiu (2002) found that inland populations in North Carolina have more genetic diversity than more coastal populations potentially due to the elevation changes that bisect the state (Qiu et al., 2002). The Southern coastal populations in our study tended to be less admixed than the two inland North Carolina populations. The effect hybridization of Northern and Southern populations has on tick phenotypes and Lyme disease prevalence in a region could be realized by sampling populations of *I. scapularis* east to west and north to south in North Carolina.

Conclusions

While Lyme disease is a major public health concern and prevalence has been increasing in the recent past, so has the incidence of many other tickborne diseases (R. J. Eisen & Eisen, 2018; Rosenberg et al., 2018). Understanding tick vectors, especially those like *I. scapularis* which transmits multiple human pathogens, is important not just in mitigating disease risk in humans, but to understanding these complex ecological systems in a OneHealth context. This model disease system has a complex taxonomic history and is intertwined with vector-host-pathogen-environmental relationships that still are not fully understood. The genetics of *I. scapularis* populations can be correlated with disease prevalence but a direct phenotype-genotype link had not yet been made. This study has built a foundation for tracking the genetic background, and the shifts in them, that are associated with higher disease transmission. Although the RAD loci investigated here illuminate overall patterns and potential candidate loci for further work, future whole genome resequencing studies of*I. scapularis* individuals from populations across the geographic range are needed to obtain a more complete picture of genomic variability and characterization of candidate loci.

Identifying and monitoring genetic backgrounds associated with increased occurrence of tickborne disease can provide a rapid snapshot tool for assessing potential community risk. With this 3RAD approach, we identified genomic regions that should be further studied to determine how genetic diversity is related to behavioral differences among Northern and Southern populations. By using our 3RAD approach, we can examine consistent loci across the genome in many individual ticks, which can be replicated in new *I. scapularis* populations to continually add to this dataset. Having a database of trackable loci associated with increased Lyme disease risk or prevalence could allow public health officials to implement targeted, efficient, and thus cost-effective, control strategies for distinct tick populations or undertake public health awareness campaigns in at-risk communities before they become an emerging hotspot for tickborne disease.

Acknowledgements

We thank all of the community scientists who participated in this project including Albert Krause, Alex

Brown, Bill Collins, Bill Kleiman, Bob Kidder, Craig Mott, Donna Glenn, Erwin & Gwen Thomas, Everett Fuller, Daryl Guest, Sean Burdick, Harry Bartges, Francis Culbertson, Gary Kraszewski, Greg Helfrich, James Meehan, James Roberts, Jeanne Menard, Jennifer Williston, John Sineath, Kaitlyn Sarlo, Karen Blakeley, Kirk Barkel, David Allen, Larry Strack, Matt Dettlaff, Matt Ross, Nate Hibshman, New England K9 Search and Rescue, Nikki Gammans, Norman Pollock, Paul Hicks, Quality Deer Management Association (National Deer Alliance), Randy Behm, Rich Herd, Richard Poeppel, Rick Hill, Robert Barnes, Russ Hartwell, Shane Hesting, Suffolk County Department of Health Arthropod-Borne Disease Laboratory, Ted Nixon, Tim Lewis, Todd Woida, and Vincent P. Sharkey. JCF and ATT were partially supported by Grant No. DGE-1545433 from the National Science Foundation and the IDEAS program at the University of Georgia. MJY and ATT are partially funded through Cooperative Agreements AP19VSCEAH00C004 and AP20VSCEAH00C041, Veterinary Services, Animal and Plant Health Inspection Service, USDA. Additional support was provided by SCWDS member state wildlife agencies provided by the Federal Aid to Wildlife Restoration Act (50 Stat. 917) and through federal agency partners, including the United States Geological Survey Ecosystems Mission Area and the United States Fish and Wildlife Service National Wildlife Refuge System. ATT was partially supported by and through the United States Department of Agriculture Animal Plant Health Inspection Service's National Bio- and Agro-defense Facility Scientist Training Program. Tick samples collected by HCT from Illinois drag collections resulted from work supported through an agreement with the Illinois Department of Public Health. The content is solely the responsibility of the authors and does not necessarily represent the official views of NSF, USDA, the Illinois Department of Public Health, or the Centers for Disease Control. GD was supported by funding from the US Department of Energy through Cooperative Agreement number DE-FC09-07SR22506 with the University of Georgia Research Foundation.

References

Ammazzalorso, A. D., Zolnik, C. P., Daniels, T. J., & Kolokotronis, S. O. (2015). To beat or not to beat a tick: comparison of DNA extraction methods for ticks (Ixodes scapularis). *PeerJ, 3* , e1147. Doi:10.7717/peerj.1147

Anandan, A., Nagireddy, R., Sabarinathan, S., Bhatta, B. B., Mahender, A., Vinothkumar, M., . . . Ali, J. (2022). Multi-trait association study identifies loci associated with tolerance of low phosphorus in Oryza sativa and its wild relatives. *Sci Rep, 12* (1), 4089. Doi:10.1038/s41598-022-07781-5

Apperson, C. S., Levine, J. F., Evans, T. L., Braswell, A., & Heller, J. (1993). Relative utilization of reptiles and rodents as hosts by immature Ixodes scapularis (Acari: Ixodidae) in the coastal plain of North Carolina, USA. *Exp Appl Acarol, 17* (10), 719-731. Doi:10.1007/BF00051830

Arsnoe, I., Hickling, G. J., Ginsberg, H. S., McElreath, R., & Tsao, J. I. (2015). Different populations of blacklegged tick nymphs exhibit differences in questing behavior that have implications for human lyme disease risk. *PloS One, 10* (5), e0127450. Doi:10.1371/journal.pone.0127450

Arsnoe, I., Tsao, J. I., & Hickling, G. J. (2019). Nymphal Ixodes scapularis questing behavior explains geographic variation in Lyme borreliosis risk in the eastern United States. *Ticks Tick Borne Dis, 10* (3), 553-563. Doi:10.1016/j.ttbdis.2019.01.001

Aubry, F., Dabo, S., Manet, C., Filipovic, I., Rose, N. H., Miot, E. F., . . . Lambrechts, L. (2020). Enhanced Zika virus susceptibility of globally invasive Aedes aegypti populations. *Science, 370* (6519), 991-996. Doi:10.1126/science.abd3663

Bacon, R. M., Kugeler, K. J., Mead, P. S., Centers for Disease, C., & Prevention. (2008). Surveillance for Lyme disease—United States, 1992-2006. *MMWR Surveill Summ, 57* (10), 1-9.

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

Bisanzio, D., Fernandez, M. P., Martello, E., Reithinger, R., & Diuk-Wasser, M. A. (2020). Current and

Future Spatiotemporal Patterns of Lyme Disease Reporting in the Northeastern United States. *JAMA Netw Open, 3* (3), e200319. Doi:10.1001/jamanetworkopen.2020.0319

Burtis, J. C., Sullivan, P., Levi, T., Oggenfuss, K., Fahey, T. J., & Ostfeld, R. S. (2016). The impact of temperature and precipitation on blacklegged tick activity and Lyme disease incidence in endemic and emerging regions. *Parasit Vectors, 9* (1), 606. Doi:10.1186/s13071-016-1894-6

Caballero, A., & Garcia-Dorado, A. (2013). Allelic diversity and its implications for the rate of adaptation. *Genetics, 195* (4), 1373-1384. Doi:10.1534/genetics.113.158410

Carpenter, C. A., Hooks, S. L., Koenig, F., McNab, W. H., Russell, W., & Smith, M. L. (1995). Ecological units of the eastern United States: first approximation. *US Department of Agriculture, Forest Service* , 81.

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: an analysis tool set for population genomics. *Mol Ecol, 22* (11), 3124-3140. Doi:10.1111/mec.12354

CDC. (1990). Case definitions for public health surveillance. *MMWR CDC Surveill Summ, 30* (RR-13).

Clow, K. M., Leighton, P. A., Ogden, N. H., Lindsay, L. R., Michel, P., Pearl, D. L., & Jardine, C. M. (2017). Northward range expansion of Ixodes scapularis evident over a short timescale in Ontario, Canada. *PloS One, 12* (12), e0189393. Doi:10.1371/journal.pone.0189393

Couper, L. I., Yang, Y., Yang, X. F., & Swei, A. (2020). Comparative vector competence of North American Lyme disease vectors. *Parasit Vectors, 13* (1), 29. Doi:10.1186/s13071-020-3893-x

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., . . . Genomes Project Analysis, G. (2011). The variant call format and VCFtools. *Bioinformatics, 27* (15), 2156-2158. Doi:10.1093/bioinformatics/btr330

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., . . . Li, H. (2021). Twelve years of SAMtools and BCFtools. *Gigascience, 10* (2). Doi:10.1093/gigascience/giab008

Davey, J. W., & Blaxter, M. L. (2010). RADSeq: next-generation population genetics. *Brief Funct Genomics, 9* (5-6), 416-423. Doi:10.1093/bfgp/elq031

de Queiroz, K. (2005). Ernst Mayr and the modern concept of species. *Proc Natl Acad Sci U S A, 102 Suppl 1* , 6600-6607. Doi:10.1073/pnas.0502030102

Division of Vector-Borne Diseases. (2019, February 5, 2019). Lyme Disease: Data and Surveillance. Retrieved from https://www.cdc.gov/lyme/datasurveillance/index.html

Dolan, M. C., Piesman, J., Mbow, M. L., Maupin, G. O., Peter, O., Brossard, M., & Golde, W. T. (1998). Vector competence of Ixodes scapularis and Ixodes ricinus (Acari: Ixodidae) for three genospecies of Borrelia burgdorferi. *J Med Entomol, 35* (4), 465-470. Doi:10.1093/jmedent/35.4.465

Dupuis, J., & Siegmund, D. (1999). Statistical methods for mapping quantitative trait loci from a dense set of markers. *Genetics, 151* (1), 373-386. Doi:10.1093/genetics/151.1.373

Durden, L. A., & Keirans, J. E. (1996). *Nymphs of the genus Ixodes (Acari: Ixodidae) of the United States: taxonomy, identification key, distribution, hosts, and medical/veterinary importance* : Entomological Society of America.

Eisen, L., Breuner, N. E., Hojgaard, A., Hoxmeier, J. C., Pilgard, M. A., Replogle, A. J., . . . Dolan, M. C. (2017). Comparison of Vector Efficiency of Ixodes scapularis (Acari: Ixodidae) From the Northeast and Upper Midwest of the United States for the Lyme Disease Spirochete Borrelia mayonii. *J Med Entomol, 54* (1), 239-242. Doi:10.1093/jme/tjw160

Eisen, R. J., & Eisen, L. (2018). The Blacklegged Tick, Ixodes scapularis: An Increasing Public Health Concern. *Trends Parasitol, 34* (4), 295-309. Doi:10.1016/j.pt.2017.12.006

Eisen, R. J., Eisen, L., & Beard, C. B. (2016). County-Scale Distribution ofIxodes scapularisandIxodes pacificus(Acari: Ixodidae) in the Continental United States. *Journal of Medical Entomology, 53* (2), 349-386. Doi:10.1093/jme/tjv237

Fleshman, A. C., Graham, C. B., Maes, S. E., Foster, E., & Eisen, R. J. (2021). Reported County-Level Distribution of Lyme Disease Spirochetes, Borrelia burgdorferi sensu stricto and Borrelia mayonii (Spirochaetales: Spirochaetaceae), in Host-Seeking Ixodes scapularis and Ixodes pacificus Ticks (Acari: Ixodidae) in the Contiguous United States. *J Med Entomol, 58* (3), 1219-1233. Doi:10.1093/jme/tjaa283

Frichot, E., & Francois, O. (2015). LEA: An R package for landscape and ecological association studies. *Methods in Ecology and Evolution, 6* (8), 925-929.

Ginsberg, H. S., Albert, M., Acevedo, L., Dyer, M. C., Arsnoe, I. M., Tsao, J. I., . . . LeBrun, R. A. (2017). Environmental Factors Affecting Survival of Immature Ixodes scapularis and Implications for Geographical Distribution of Lyme Disease: The Climate/Behavior Hypothesis.*PloS One, 12* (1), e0168723. Doi:10.1371/journal.pone.0168723

Ginsberg, H. S., Rulison, E. L., Azevedo, A., Pang, G. C., Kuczaj, I. M., Tsao, J. I., & LeBrun, R. A. (2014). Comparison of survival patterns of northern and southern genotypes of the North American tick Ixodes scapularis (Acari: Ixodidae) under northern and southern conditions. *Parasit Vectors, 7* , 394. Doi:10.1186/1756-3305-7-394

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

Goddard, J., Embers, M., Hojgaard, A., & Piesman, J. (2015). Comparison of Tick Feeding Success and Vector Competence for Borrelia burgdorferi Among Immature Ixodes scapularis (Ixodida: Ixodidae) of Both Southern and Northern Clades. *J Med Entomol, 52* (1), 81-85. Doi:10.1093/jme/tju005

Goudet, J. (2005). Hierfstat, a package for R to compute and test hierarchical F-statistics. *Molecular Ecology Notes, 5* (1), 184-186.

Guerra, M., Walker, E., Jones, C., Paskewitz, S., Cortinas, M. R., Stancil, A., . . . Kitron, U. (2002). Predicting the risk of Lyme disease: habitat suitability for Ixodes scapularis in the north central United States. *Emerg Infect Dis, 8* (3), 289-297. Doi:10.3201/eid0803.010166

Gulia-Nuss, M., Nuss, A. B., Meyer, J. M., Sonenshine, D. E., Roe, R. M., Waterhouse, R. M., . . . Hill, C. A. (2016). Genomic insights into the Ixodes scapularis tick vector of Lyme disease. *Nature Communications, 7* (1), 10507. Doi:10.1038/ncomms10507

Hickling, G. J., Kelly, J. R., Auckland, L. D., & Hamer, S. A. (2018). Increasing Prevalence of Borrelia burgdorferi sensu stricto-Infected Blacklegged Ticks in Tennessee Valley, Tennessee, USA. *Emerg Infect Dis, 24* (9), 1713-1716. Doi:10.3201/eid2409.180343

Jenkins, T. L., Guillemin, M. L., Simon-Nutbrown, C., Burdett, H. L., Stevens, J. R., & Pena, V. (2021). Whole genome genotyping reveals discrete genetic diversity in north-east Atlantic maerl beds. *Evol Appl, 14* (6), 1558-1571. Doi:10.1111/eva.13219

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

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

Keirans, J. E., & Litwak, T. R. (1989). Pictorial key to the adults of hard ticks, family Ixodidae (Ixodida: Ixodoidea), east of the Mississippi River. *J Med Entomol, 26* (5), 435-448. Doi:10.1093/jmedent/26.5.435

Kelly, R. R., Gaines, D., Gilliam, W. F., & Brinkerhoff, R. J. (2014). Population genetic structure of the Lyme disease vector Ixodes scapularis at an apparent spatial expansion front. *Infect Genet Evol, 27* , 543-550. Doi:10.1016/j.meegid.2014.05.022

Khatchikian, C. E., Prusinski, M. A., Stone, M., Backenson, P. B., Wang, I.-N., Foley, E., . . . Brisson, D. (2015). Recent and rapid population growth and range expansion of the Lyme disease tick vector,Ixodes scapularis, in North America. *Evolution, 69* (7), 1678-1689. Doi:10.1111/evo.12690

Kovalev, S. Y., Golovljova, I. V., & Mukhacheva, T. A. (2016). Natural hybridization between Ixodes ricinus and Ixodes persulcatus ticks evidenced by molecular genetics methods. *Ticks and tick-borne diseases* , 7(1), 113-118. doi:10.1016/j.ttbdis.2015.09.005

Krakowetz, C. N., Lindsay, L. R., & Chilton, N. B. (2011). Genetic diversity in Ixodes scapularis (Acari: Ixodidae) from six established populations in Canada. *Ticks Tick Borne Dis, 2* (3), 143-150. Doi:10.1016/j.ttbdis.2011.05.003

Krinsky, W. L. (1979). Development of the tick Ixodes dammini (Acarina: Ixodidae) in the laboratory. *Journal of medical entomology* , 16(4), 354-355. doi: 10.1093/jmedent/16.4.354

Kuehn, B. M. (2013). CDC estimates 300,000 US cases of Lyme disease annually. *JAMA, 310* (11), 1110. Doi:10.1001/jama.2013.278331

Kugeler, K. J., Farley, G. M., Forrester, J. D., & Mead, P. S. (2015). Geographic Distribution and Expansion of Human Lyme Disease, United States. *Emerging Infectious Diseases, 21* (8), 1455-1457. Doi:10.3201/eid2108.141878

Lane, R. S., Piesman, J., & Burgdorfer, W. (1991). Lyme borreliosis: relation of its causative agent to its vectors and hosts in North America and Europe. *Annu Rev Entomol, 36* , 587-609. Doi:10.1146/annurev.en.36.010191.003103

Leighton, P. A., Koffi, J. K., Pelcat, Y., Lindsay, L. R., & Ogden, N. H. (2012). Predicting the speed of tick invasion: an empirical model of range expansion for the Lyme disease vector Ixodes scapularis in Canada.*Journal of Applied Ecology, 49* (2), 457-464.

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics, 25* (14), 1754-1760. Doi:10.1093/bioinformatics/btp324

Ludwig, J. (2015). Genetic Structure of Ixodes Scapularis Say 1821 (Acari: Ixodidae), The Blacklegged Tick, By Microsatellite Analysis.*Georgia Southern University Electronic Theses and Dissertations, 1356* .

Lv, J., Wu, S., Zhang, Y., Chen, Y., Feng, C., Yuan, X., . . . Lin, X. (2014). Assessment of four DNA fragments (COI, 16S rDNA, ITS2, 12S rDNA) for species identification of the Ixodida (Acari: Ixodida).*Parasit Vectors, 7* , 93. Doi:10.1186/1756-3305-7-93

Norris, D. E., Klompen, J. S., Keirans, J. E., & Black, W. C. t. (1996). Population genetics of Ixodes scapularis (Acari: Ixodidae) based on mitochondrial 16S and 12S genes. *J Med Entomol, 33* (1), 78-89.

Ogden, N. H., Lindsay, L. R., & Leighton, P. A. (2013). Predicting the rate of invasion of the agent of Lyme disease Borrelia burgdorferi.*Journal of Applied Ecology, 50* (2), 510-518.

Ogden, N. H., Lindsay, L. R., Morshed, M., Sockett, P. N., & Artsob, H. (2009). The emergence of Lyme disease in Canada. *CMAJ, 180* (12), 1221-1224. Doi:10.1503/cmaj.080148

Ogden, N. H., Trudel, L., Artsob, H., Barker, I. K., Beauchamp, G., Charron, D. F., . . . Lindsay, L. R. (2006). Ixodes scapularis ticks collected by passive surveillance in Canada: analysis of geographic distribution and infection with Lyme borreliosis agent Borrelia burgdorferi. *J Med Entomol, 43* (3), 600-609. Doi:10.1603/0022-2585(2006)43[600:ISTCBP]2.0.CO;2

Oliver, J. H., Jr. (1996). Lyme borreliosis in the southern United States: a review. *J Parasitol, 82* (6), 926-935.

Oliver, J. H., Jr., Owsley, M. R., Hutcheson, H. J., James, A. M., Chen, C., Irby, W. S., . . . McLain, D. K. (1993). Conspecificity of the ticks Ixodes scapularis and I. dammini (Acari: Ixodidae). *J Med Entomol, 30* (1), 54-63. Doi:10.1093/jmedent/30.1.54

Ostfeld, R. S., & Keesing, F. (2000). Biodiversity and disease risk: the case of Lyme disease. *Conservation biology, 14* (3), 722-728. Doi:10.1046/j.1523-1739.2000.99014.x

Patterson, J. W., Duncan, A. M., McIntyre, K. C., & Lloyd, V. K. (2017). Evidence for genetic hybridization between Ixodes scapularis and Ixodes cookei. *Canadian Journal of Zoology* , 95(8), 527-537. doi: 10.1139/cjz-2016-0134

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* (3), 833-849. doi:10.1093/genetics/160.3.833

Rahman, S. R., Cnaani, J., Kinch, L. N., Grishin, N. V., & Hines, H. M. (2021). A combined RAD-Seq and WGS approach reveals the genomic basis of yellow color variation in bumble bee Bombus terrestris. *Sci Rep, 11* (1), 7996. doi:10.1038/s41598-021-87194-y

Ripoche, M., Bouchard, C., Irace-Cima, A., Leighton, P., & Thivierge, K. (2022). Current and future distribution of Ixodes scapularis ticks in Quebec: Field validation of a predictive model. *PLoS One, 17* (2), e0263243. doi:10.1371/journal.pone.0263243

Rosenberg, R., Lindsey, N. P., Fischer, M., Gregory, C. J., Hinckley, A. F., Mead, P. S., . . . Petersen, L. R. (2018). Vital Signs: Trends in Reported Vectorborne Disease Cases - United States and Territories, 2004-2016. *MMWR Morb Mortal Wkly Rep, 67* (17), 496-501. doi:10.15585/mmwr.mm6717e1

Sakamoto, J. M., Goddard, J., & Rasgon, J. L. (2014). Population and demographic structure of Ixodes scapularis Say in the eastern United States. *PLoS One, 9* (7), e101389. doi:10.1371/journal.pone.0101389

Sanders, M. (1998). Ixodes dammini: A Junior Synonym for Ixodes scapularis. *Emerging Infectious Diseases* , 4(1), 132. doi:10.3201/eid0401.980125.

Schwan, T. G., & Piesman, J. (2002). Vector interactions and molecular adaptations of lyme disease and relapsing fever spirochetes associated with transmission by ticks. *Emerg Infect Dis, 8* (2), 115-121. doi:10.3201/eid0802.010198

Scott, J. D., Fernando, K., Banerjee, S. N., Durden, L. A., Byrne, S. K., Banerjee, M., . . . Morshed, M. G. (2001). Birds disperse ixodid (Acari: Ixodidae) and Borrelia burgdorferi-infected ticks in Canada.*J Med Entomol, 38* (4), 493-500. doi:10.1603/0022-2585-38.4.493

Simo, L., Koci, J., & Park, Y. (2013). Receptors for the neuropeptides, myoinhibitory peptide and SIFamide, in control of the salivary glands of the blacklegged tick Ixodes scapularis. *Insect Biochem Mol Biol, 43* (4), 376-387. doi:10.1016/j.ibmb.2013.01.002

Sokal, R. R., & Crovello, T. J. (1970). The Biological Species Concept: A Critical Evaluation. *The American Naturalist, 104* (936), 127-153.

Spielman, A., Clifford, C. M., Piesman, J., & Corwin, M. D. (1979). Human babesiosis on Nantucket Island, USA: description of the vector, Ixodes (Ixodes) dammini, n. sp. (Acarina: Ixodidae). *J Med Entomol, 15* (3), 218-234. doi:10.1093/jmedent/15.3.218

Spielman, A., Levine, J. F., & Wilson, M. L. (1984). Vectorial capacity of North American Ixodes ticks. *Yale J Biol Med, 57* (4), 507-513.

Stecher, G., Tamura, K., & Kumar, S. (2020). Molecular Evolutionary Genetics Analysis (MEGA) for macOS. *Mol Biol Evol, 37* (4), 1237-1239. doi:10.1093/molbev/msz312

Tietjen, M., Esteve-Gasent, M. D., Li, A. Y., & Medina, R. F. (2020). A comparative evaluation of northern and southern Ixodes scapularis questing height and hiding behaviour in the USA. *Parasitology, 147* (13), 1569-1576. doi:10.1017/S003118202000147X

Van Zee, J., Piesman, J. F., Hojgaard, A., & Black Iv, W. C. (2015). Nuclear Markers Reveal Predominantly North to South Gene Flow in Ixodes scapularis, the Tick Vector of the Lyme Disease Spirochete. *PLoS One, 10* (11), e0139630. doi:10.1371/journal.pone.0139630

Wormser, G. P., Nadelman, R. B., Dattwyler, R. J., Dennis, D. T., Shapiro, E. D., Steere, A. C., . . . Luft, B. J. (2000). Practice Guidelines for the Treatment of Lyme Disease. *Clinical Infectious Diseases, 31* (Supplement_1), S1-S14. doi:10.1086/314053

Wright, S. (1943). Isolation by Distance. *Genetics, 28* (2), 114-138. doi:10.1093/genetics/28.2.114

Yu, G., Smith, D. K., Zhu, H., Guan, Y., & Lam, T. T. Y. (2017). ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution, 8* (1), 28-36. doi:10.1111/2041-210X.12628

Data Accessibility Statement

Genetic data:

Raw sequence reads are deposited in the SRA (BioProject PRJNA852262)

Sample metadata:

Metadata are also stored in the SRA (BioProject PRJNA852262) using the invertebrate v1.0 package

Benefit Sharing Statement

Benefits Generated: The work herein represents a collaboration among university and public health laboratories along with community scientists. The data are available for use in understanding these ticks and the diseases they transmit. The work illustrates methods that can be used for further studies on similar species, especially understudied ticks of concern to the global community.

Author contributions

JCF, TCG, and IR co-conceived the project. JCF coordinated all sample collection, completed library preparation, statistical analyses, and manuscript writing. JCF, ATT, and PS completed morphological IDs of volunteer submitted samples. ATT molecularly identified samples and ran associated analyses. JCF, ATT, GD, RP, RCS, KDS, JIT, HCT, and MJY provided samples that were previously identified. All authors completed revisions of the manuscript.

Competing Interests

The authors declare no competing interests.

Disclaimer

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government, nor any agency thereof, nor any of their employees makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions

of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

Tables

Table 1. *Ixodes scapularis* population metadata and statistics. N_ticks = total ticks in population; Nalleles = alleles in population; Ar = allelic richness; Ho = Observed heterozygosity; Hs = Mean gene diversity; Fis = Inbreeding coefficient

Location	Latitude	Longitude	N_Males	N_Females	N_Ticks	Nalleles	Ar	Ho	Hs
Lauderdale County, AL	34.8841	-87.6581	6	1	7	10581	1.19	0.14	0.19
Kent County, DE	39.1441	-75.4345	10	1	11	12615	1.22	0.15	0.2
Jefferson County, FL	30.4312	-83.8897	6	1	7	9843	1.16	0.10	0.1
Osceola County, FL	28.1020	-81.0755	7	5	12	9086	1.14	0.09	0.1
Walker County, GA	34.6858	-85.3550	4	2	6	11157	1.22	0.16	0.2
Poweshiek County, IA	41.6326	-92.6732	4	5	9	11819	1.21	0.17	0.2
Van Buren County, IA	40.8825	-92.1168	4	4	8	11970	1.22	0.17	0.2
Logan County, IL	40.0613	-89.3227	12	0	12	12720	1.22	0.16	0.2
Winnebago County, IL	42.3121	-89.1706	13	0	13	12525	1.21	0.16	0.2
Franklin County, KY	38.2481	-84.8985	6	1	7	11880	1.22	0.17	0.2
Owen County, KY	38.4965	-84.8151	4	0	4	10934	1.21	0.16	0.2
Bossier County, LA	32.7551	-93.6623	5	8	13	11258	1.17	0.12	0.1
Iberville County, LA	30.2899	-91.4048	5	3	8	10546	1.16	0.11	0.1
Rapides County, LA	31.1461	-92.5396	9	3	12	11034	1.16	0.11	0.1
Allegany County, MD	39.6255	-78.6115	9	2	11	12518	1.21	0.16	0.2
Ingham County, MI	42.7117	-84.5231	6	6	12	12497	1.21	0.16	0.2
Pettis County, MO	38.6893	-93.3389	9	0	9	11690	1.20	0.14	0.2
Stokes County, NC	36.4244	-80.2321	11	0	11	12437	1.22	0.14	0.2
Watauga Count, NC	36.1942	-81.7349	11	0	11	12691	1.22	0.16	0.2
Carroll County, NH	43.8128	-71.2506	7	6	13	12880	1.22	0.17	0.2
Oswego County, NY	43.4825	-76.1784	11	1	12	12747	1.22	0.16	0.2
Schenectady County, NY	42.8493	-73.9830	11	0	11	12264	1.21	0.16	0.2
Knox County, OH	40.4596	-82.3018	9	2	11	12507	1.22	0.16	0.2
Scioto County, OH	38.8611	-82.9932	10	2	12	12892	1.22	0.15	0.2
Payne County, OK	36.1450	-96.7500	12	0	12	11351	1.18	0.11	0.1
Centre County, PA	40.8766	-77.8367	7	3	10	12118	1.21	0.17	0.2
Westmoreland County, PA	40.2354	-79.4704	11	2	13	12566	1.21	0.15	0.2
Washington County, RI	41.4568	-71.6673	9	4	13	12646	1.21	0.16	0.2
Aiken County, SC	33.2464	-81.6679	9	4	13	11265	1.17	0.11	0.1
Claiborne County, TN	36.4830	-83.6348	7	6	13	12526	1.21	0.16	0.2
Windsor County, VT	43.4369	-72.6151	10	3	13	12739	1.22	0.16	0.2
Marinette County, WI	45.3582	-88.0650	6	6	12	12095	1.20	0.15	0.2
Monroe County, WI	44.0127	-90.6884	7	5	12	12275	1.20	0.15	0.2

Figures

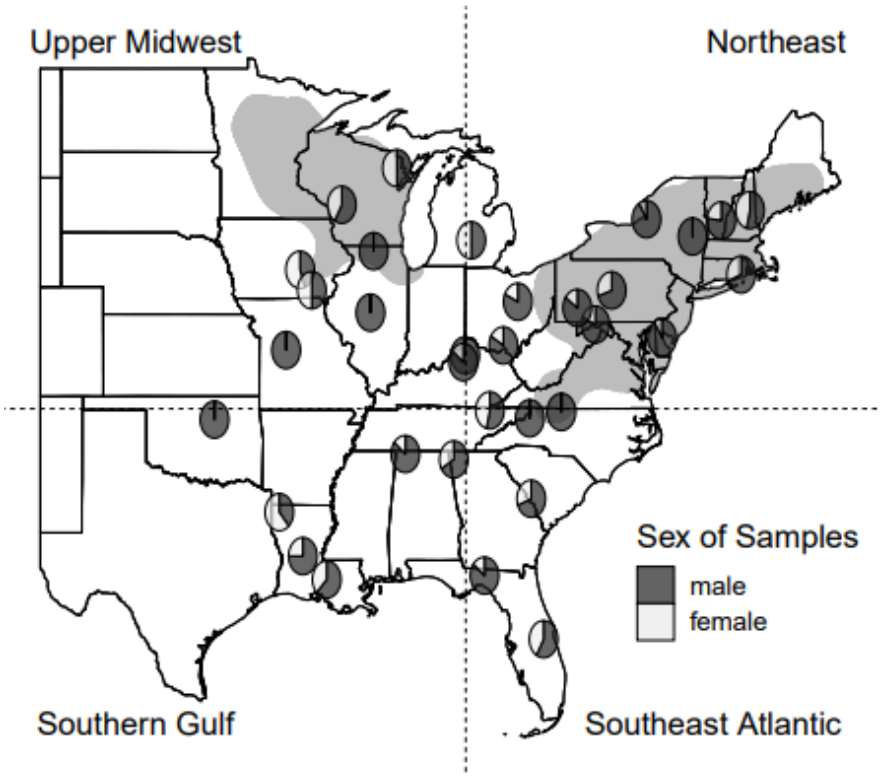

Figure 1. Collection sites of *Ixodes scapularis* adults.Map of the Eastern US. The shaded region is where 95% of Lyme disease cases are reported in the US. Each pie chart is showing the male (dark grey) to female (light grey) breakdown of the population and is centered at the longitude and latitude of the collection county.

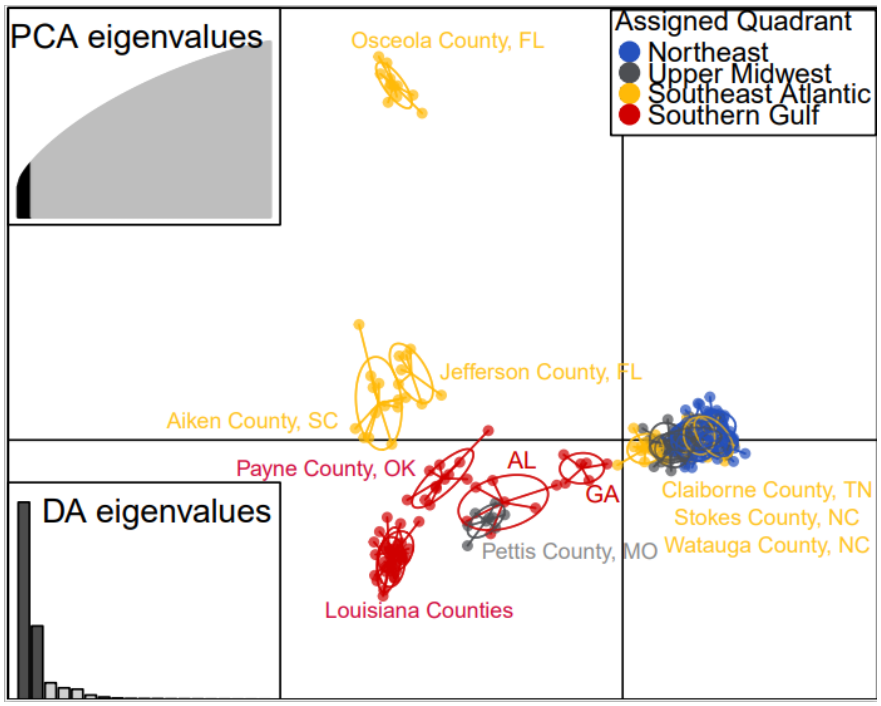

Figure 2. Discriminant Analysis of Principal Components of*I. scapularis* populations. All populations analyzed separately and assigned a color family based on the quadrant the populations originate.

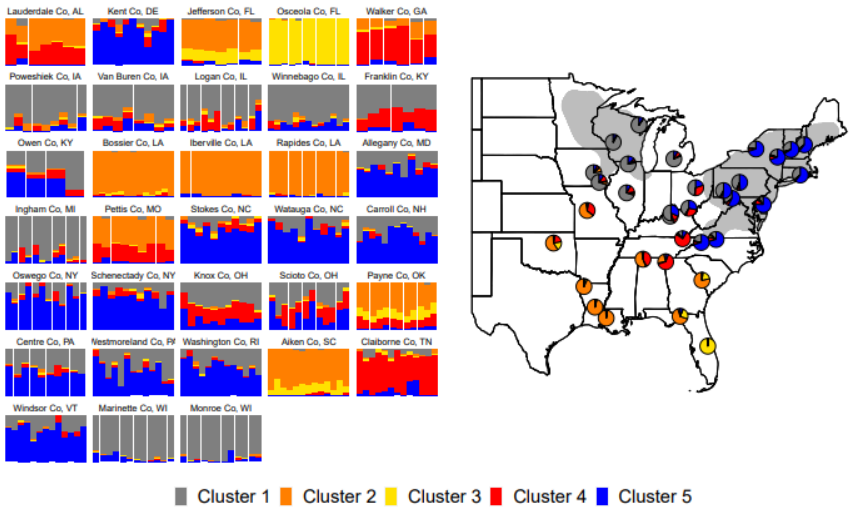

Figure 3. Genetic structure of *I. scapularis* individuals, and the relative admixture per individual and population. Grey is cluster 1 mainly representing the Northwest, orange is cluster 2 made comprised of the South, yellow is cluster 3 comprised of Osceola County Florida, red is cluster 4 representing the central region of the collection area, and blue is cluster 5 representing the Northeastern region (A) Depicts the admixture of each individual divided by population. (B) Population-level admixture proportions where the pie chart is at the longitude and latitude of the collection counties.

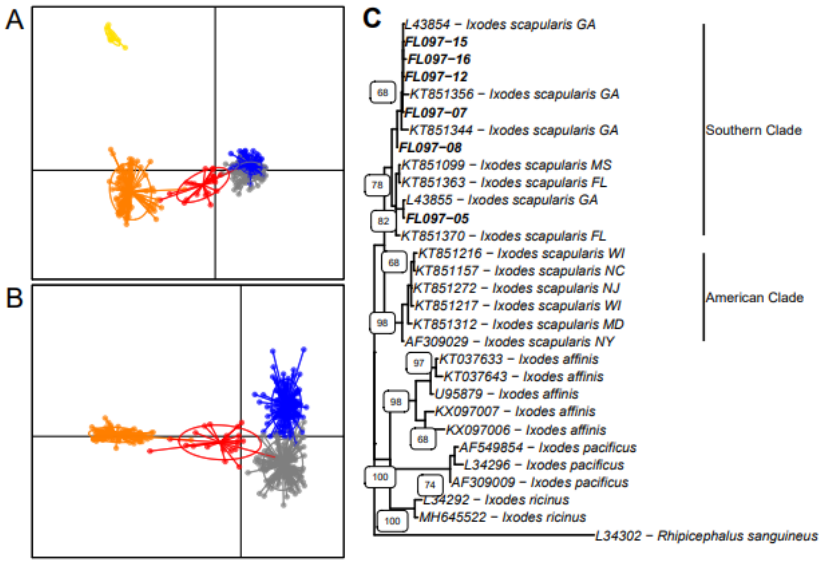

Figure 4. LEA Cluster Analyses. Grey points are cluster 1, orange points are cluster 2, yellow points are cluster 3, red points are cluster 4, and blue points are cluster 5. (A) DAPC scatter plot of all individuals assigned to the cluster which represented the most of their admixture proportions. (B) DAPC scatter plot after removing samples from the Osceola County, FL population and rerunning the analysis. (C) Phylogenetic tree of Osceola County, FL samples

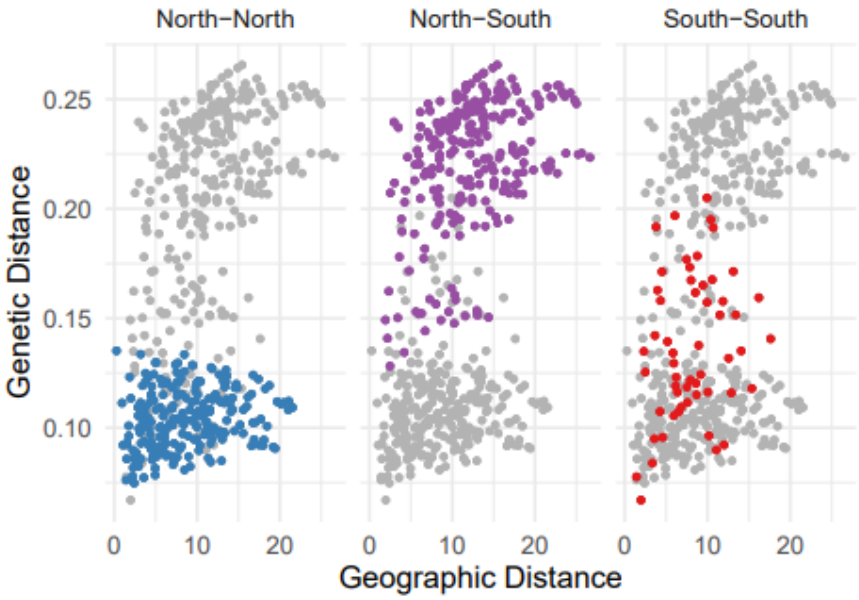

Figure 5. Isolation by Distance. Isolation by distance comparing all 33 populations. Northern to Northern comparisons in blue, Northern to Southern comparisons in purple, Southern to Southern comparisons in red.

Supplemental Figures

Supp. Figure 1A-C. Manhattan plots comparing select northern and southern populations.

Supp. Figure 2. Cross entropy from LEA determining the optimal K for the dataset.

Supp. Figure 3. Sites contributing to the differentiation of the 5 clusters.

Supp. Figure 4. Allele representation of the top contributing loci to DAPC differentiation.

Supp. Figure 5. Sites contributing to the differentiation of clusters 1, 2, 4, and 5.

Supp. Figure 6. Allele representation of the top contributing loci to DAPC differentiation for clusters after removing Osceola County, FL.

Supplemental Tables

Supp. Table 1. Community Science tick collection data for *Ixodes scapularis* .

Supp. Table 2. VCFtools filtering of Stacks output VCF. The input for each parameter in VCFtools and the retained sites after filtering.

Supp. Table 3. Pairwise Fst values calculated using hierfstat pairwise.neifst and pairwise.wcfst. Nei calculation is on top of the diagonal, and Weir-Cockerham (WC) calculation is below the diagonal.

Supp. Table 4. AMOVA Results

Supp. Table 5. Hardy-Weinberg Equilibrium Results

Supp. Table 6. Site information for the top 10 sites contributing to differentiation of Cluster 1, 2, 4, and 5.

Supplemental Text

Details of *I. scapularis* taxonomy

Details of *I. scapularis* population genetics

Supplemental Files

Supp. File 1. HWresults_per_site.csv. This file contains the Hardy-Weinberg Equilibrium results per high-quality site used in this study.