Increased diversity of zoonotic pathogens and \textit{Borrelia burgdorferi} strains in established versus incipient \textit{Ixodes scapularis} populations across the Midwestern United States

Sarah A. Hamer\textsuperscript{a,}\textsuperscript{⇑}, Graham J. Hickling\textsuperscript{b}, Edward D. Walker\textsuperscript{c}, Jean I. Tsao\textsuperscript{c}

\textsuperscript{a} Texas A&M University, 4458 TAMU, College Station, TX 77843, USA
\textsuperscript{b} University of Tennessee, 274 Ellington Plant Science Building, Knoxville, TN 37996, USA
\textsuperscript{c} Michigan State University, 13 Natural Resources Building, East Lansing, MI 48823, USA

\textbf{A R T I C L E   I N F O}

\textbf{Article info}
Received 16 January 2014
Received in revised form 2 June 2014
Accepted 6 June 2014
Available online 18 June 2014

\textbf{Keywords:}
\textit{Ixodes scapularis}
\textit{Borrelia burgdorferi}
\textit{Anaplasma}
\textit{Babesia}
\textit{American Midwest}

\textbf{A B S T R A C T}

The center of origin theory predicts that genetic diversity will be greatest near a species' geographic origin because of the length of time for evolution. By corollary, diversity will decrease with distance from the origin; furthermore, invasion and colonization are frequently associated with founder effects that reduce genetic variation in incipient populations. The blacklegged tick, \textit{Ixodes scapularis}, harbors a suite of zoonotic pathogens, and the geographic range of the tick is expanding in the upper Midwestern United States. Therefore, we posited that diversity of \textit{I. scapularis}-borne pathogens across its Midwestern range should correlate with the rate of the range expansion of this tick as well as subsequent disease emergence. Analysis of 1565 adult \textit{I. scapularis} ticks from 13 sites across five Midwestern states revealed that tick infection prevalence with multiple microbial agents (\textit{Borrelia burgdorferi}, \textit{Borrelia miyamotoi}, \textit{Babesia odocoilei}, \textit{Babesia microti}, and \textit{Anaplasma phagocytophilum}), coinfections, and molecular genetic diversity of \textit{B. burgdorferi} all were positively correlated with the duration of establishment of tick populations, and therefore generally support the center of origin – pathogen diversity hypothesis. The observed differences across the gradient of establishment, however, were not strong and were nuanced by the high frequency of coinfections in tick populations at both established and recently-invaded tick populations. These results suggest that the invasion of ticks and their associated pathogens likely involve multiple means of pathogen introduction, rather than the conventionally presented scenario whereby infected, invading ticks are solely responsible for introducing pathogens to naïve host populations.

© 2014 Elsevier B.V. All rights reserved.

\textbf{1. Introduction}

In biogeographical studies, a center of origin is a geographical area where a group of organisms first developed its distinctive properties (Cain, 1943). This concept is typically applied to studies of crop domestication and is a foundational principle of dispersal biogeography (Crisci et al., 2003). One may predict that the center of genetic diversity of a pathogen is also at its center of origin due to the length of time available for evolution (Wang et al., 1999a). Diversity is predicted to decrease with distance from the origin, because invasion and colonization events are associated with founder effects that reduce genetic variation in incipient populations (Carlson and Templeton, 1984), and less time has been available for local drift and adaptation. Because the distribution of vector-borne pathogens and the diseases they cause are inextricably linked to that of their invertebrate vectors (Gubler, 1998; Otranto et al., 2009), the patterns of invasion and establishment of vectors can often be used to predict disease risk. Similarly, pathogen diversity and prevalence may provide an indirect understanding of the history of the disease system, as well as of the underlying ecological and evolutionary processes.

\textit{Ixodes scapularis}, the blacklegged tick, is an epidemiologically important vector of multiple zoonotic pathogens in the upper Midwestern and Northeastern United States. Spielman (1988) hypothesized that historically, blacklegged ticks likely were widespread throughout eastern North America prior to the Pleistocene glaciation, but then were restricted to refugia in the northeastern United States and northwestern Wisconsin. After the glaciers receded 10,000 years ago, relict populations putatively remained (Spielman, 1988). In the mid-twentieth century, the reversion of agricultural lands to forest and implementation of hunting regulations that allowed deer populations to increase in number...
supported expansion of I. scapularis and its vertebrate hosts from these refugia (Matuschka and Spielman, 1986). I. scapularis continues to spread from endemic areas both in the Northeastern and Midwestern United States, but the mechanisms for this spread are not established. Across the Midwest, a gradient of establishment of I. scapularis is apparent (Table 1), with tick populations first detected in Wisconsin (WI), Minnesota (MN), Michigan’s (MI) Upper Peninsula, and northwestern Illinois (IL) (Bouseman et al., 1990; Drew et al., 1988; Jackson and Defoliart, 1970; Strand et al., 1992), later detected in northern Indiana (IN) (Bouseman et al., 1990; Cortinas and Kitron, 2006; Pinger and Glancy, 1989; Pinger et al., 1991, 1996), and most recently detected in the Chicago-area of Illinois (Jobe et al., 2006, 2007; Rydzewski et al., 2012) and Michigan’s Lower Peninsula (Foster, 2004; Hamer et al., 2007, 2010).

Currently, the three most epidemiologically significant I. scapularis-borne zoonotic pathogens are Borrelia burgdorferi, Anaplasma phagocytophilum, and Babesia microti. All three pathogens have similar life histories, in which wild rodents serve as important reservoirs (Donahue et al., 1987; Levin et al., 2002; Stancil, 1999; Telford and Spielman, 1993; Walls et al., 1997) and disease results when I. scapularis infected by wildlife serves as a bridge vector to humans. Several other pathogens have been recently detected in I. scapularis, including the relapsing fever spirochete B. miyamotoi. This pathogen was originally described in Xodes persulcatus ticks in Japan (Fukunaga et al., 1995), and was recently determined to cause human disease in Russia and the US (Gugliotta et al., 2013; Krause et al., 2013; Platonov et al., 2011).

We tested the center of origin theory using I. scapularis and associated pathogens not only to apply this theory to a vector-borne disease system for the first time, but also to learn more about the invasion biology of these organisms. Given the close association of I. scapularis and the pathogens it transmits, we posited that patterns of prevalence and diversity of I. scapularis-borne pathogens across its Midwestern range may reflect the order of invasion and may be useful in better understanding the mechanisms underlying the broad-scale invasion of this tick and subsequent disease emergence. Herein we hypothesize that infection prevalence, interspecific, and intraspecific diversity of pathogens within ticks will positively correlate with the duration of tick populations, due to the longer period of time for establishment of tick populations, we assigned them to three groups based on history of establishment of tick populations, as follows: Minnesota, Wisconsin, western Illinois, and Michigan Upper Peninsula sites are ‘long-established;’ northern Indiana sites are ‘intermediate-established;’ and Chicago-area and lower Michigan sites are ‘recently-invaded’. Importantly, these establishment designations were assigned not only based on the date of first documented I. scapularis detection, but also based on the results of surveys which found few or no I. scapularis in the intermediate-established and recently-invaded regions of study prior to the published establishment dates (Callister et al., 1991; Pinger et al., 1991; Walker et al., 1998).

Each site was sampled for a quota of 100 questing adult ticks by dragging a 1-m² white corduroy cloth (Falco and Fish, 1992); ticks were stored in 70% ethanol. An index of tick abundance was calculated as the number of adult I. scapularis collected per 1000 m² of drag sampling.

### 2. Materials and methods

#### 2.1. Field sites and tick samples

Ticks were sampled by drag cloth during the spring adult I. scapularis questing season (15 April–15 May) of 2006–2007 at 13 sites across the Midwest (Fig. 1, Table 1): Castle Rock State Park, Ogle Co., IL; Indiana Dunes National Lakeshore, Porter Co., IN; Tippecanoe River State Park, Pulaski Co., IN; Van Buren State Park, Van Buren Co., MI; Duck Lake State Park, Muskegon Co., MI; Fort McCoy Military Installation, Monroe Co., WI; Governor Dodge State Park, Iowa Co., WI; Saugatuck Dunes State Park, Allegan Co., MI; Menominie North and South, Menominie Co., MI; Churchill Woods Forest Preserve, DuPage Co., IL; St. corridor St. Croix State Forest, Pine Co., MN. We targeted the spring peak of activity of the adult life stage because adults have had two prior blood meals and therefore have had a greater chance of being infected in comparison to nymphs. Within each park, drag sampling was restricted to the forest floor of closed-canopy, deciduous forest habitat, as this habitat type was positively correlated with I. scapularis presence in a Midwestern habitat suitability model (Guerra et al., 2002), and a previous national tick survey also applied this sampling criterion (Diuk-Wasser et al., 2006). Due to close geographic proximity (between 30 and 70 km between sites), similar tick establishment timing, similar drag densities, and low individual sample sizes, we grouped the two Chicago area sites (Chicago North and South), the two Michigan Upper Peninsula sites (Menominie North and South), and two Michigan Lower Peninsula sites (Saugatuck and Duck Lake) for our analyses. Based on the approximate times when I. scapularis populations were first documented at these sites, we assigned them to three groups based on the current status of each site. We grouped sites based on the date of first documented I. scapularis establishment in the Midwest: (1) the population abundance of I. scapularis; (2) prevalence of Borrelia spp., A. phagocytophilum, and Babesia spp., and mixed infections in I. scapularis; and (3) strain-level diversity of B. burgdorferi.

<table>
<thead>
<tr>
<th>Establishment status</th>
<th>State</th>
<th>Year of reported I. scapularis establishment</th>
<th>Citations</th>
<th>Field sites of the current study</th>
</tr>
</thead>
<tbody>
<tr>
<td>Long-established</td>
<td>Wisconsin, northwest and central</td>
<td>1967</td>
<td>Jackson and Defoliart (1970)</td>
<td>Governor Dodge, Fort McCoy</td>
</tr>
<tr>
<td></td>
<td>Minnesota, east and central</td>
<td>1983</td>
<td>Loken et al. (1985), Drew et al. (1988)</td>
<td>St. Croix</td>
</tr>
<tr>
<td></td>
<td>Michigan, southern Upper Peninsula</td>
<td>1986</td>
<td>Strand et al. (1992)</td>
<td>Menominie North and South</td>
</tr>
<tr>
<td></td>
<td>Illinois, northwestern</td>
<td>1987</td>
<td>Bouseman et al. (1990), Kitron et al. (1991)</td>
<td>Castle Rock</td>
</tr>
<tr>
<td></td>
<td>Illinois, northeastern</td>
<td>2005</td>
<td>Jobe et al. (2006), Jobe et al. (2007)</td>
<td>Chicago North and South</td>
</tr>
</tbody>
</table>
2.2. Pathogen detection

Ticks were identified to species and stage using standard keys (Durden and Keirans, 1996; Keirans and Clifford, 1978; Sonenshine, 1979). Total DNA was extracted using the DNeasy Blood and Tissue Kit (Qiagen, Valencia, CA) following the manufacturer's animal tissue protocol, with modifications as described in Hamer et al. (2010).

All ticks were tested for the presence of three classes of pathogens in three separate PCR reactions for Borrelia species, Babesia species, and A. phagocytophilum. Subsequently, we sequenced Borrelia-positive and Babesia-positive samples to identify the species and/or strain. A PCR enzyme kit was used in all assays (PCR Supermix, Invitrogen, Carlsbad, CA), and water was used as a negative control in all assays. All assays were run in a 50 μl reaction volume. Borrelia species were detected using a nested PCR for the 16S–23S rRNA intergenic spacer region (IGS) as described by Bunikis et al. (2004a), resulting in a product size of approximately 980 bp for B. burgdorferi and 500 bp for B. miyamotoi. DNA from B. burgdorferi strain B31-infected ticks kindly provided by the CDC served as a positive PCR control. Additionally, we screened our 2007 samples for B. burgdorferi using a quantitative PCR (qPCR) targeting the 16S rRNA gene following the protocol of Tsao et al. (2004), and the reported infection prevalence includes samples positive on either assay. Babesia genus-specific PCR was performed using primers for the 18S rRNA gene to produce a fragment of variable size, including a 408-bp fragment for Ba. microti or a 437-bp fragment for Ba. odocoieli (Armstrong et al., 1998). Commercially-available Ba. microti organism (ATCC, Manassas, VA) was extracted as above and used as a positive control. The p44 gene of A. phagocytophilum (Zeidner et al., 2000) was amplified using a touchdown PCR program described in Steiner et al. (2006) to produce a 334-bp fragment. DNA from infected laboratory colony I. scapularis nymphs provided by D. Fish at Yale University was used as positive controls.

2.3. Nucleotide sequencing

Species identification and strain typing of Borrelia-positive and Babesia-positive ticks were attained through DNA sequencing using methods previously described (Hamer et al., 2010). Additionally, a subset of randomly selected samples that were PCR-positive for A. phagocytophilum was sequenced to confirm the species identity. All products were purified (Qiagen PCR Purification Kit; Qiagen, Valencia, CA) and used as a negative control in all assays. The p44 gene of A. phagocytophilum (Zeidner et al., 2000) was amplified using a touchdown PCR program described in Steiner et al. (2006) to produce a 334-bp fragment. DNA from infected laboratory colony I. scapularis nymphs provided by D. Fish at Yale University was used as positive controls.

Fig. 1. Location of field sites for collection of I. scapularis across the Midwestern United States, spring 2006–2007. Sites are categorized by establishment status of documented I. scapularis populations: highly-established (red circle); intermediate-established (blue square); recently-invaded (green triangle). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
2.4. Statistics

Kruskal–Wallis one-way analysis of variance was used to assess differences in the index of population abundance of ticks and infection prevalence among establishment groups with field sites serving as replicates. Comparisons in infection between male and female ticks were made by calculating the z-ratio and associated probability for the difference between two independent proportions. Chi-squared goodness of fit was used to assess whether infections were observed more or less frequently than expected. Expected coinfection prevalences were calculated as the product of the observed individual infection prevalence of each of two separate microbes. The effect of sample size on strain richness was assessed using a web-based rarefaction calculator (University of Alberta, Edmonton, Canada; http://wwwbiology.ualberta.ca/jbrzusto/rarefact.php). Strain richness was estimated using the nonparametric model of Chao, which uses the number of operational taxonomic units observed and the frequency with which each was observed to estimate the total population strain richness including unsampled strains (Chao and Tsung-Jen, 2003).

A series of standard ecological diversity indices were computed to assess the B. burgdorferi strain-level diversity within and among the three establishment groups. Strain richness (alpha diversity) was tabulated as a simple count of the number of strain types in each group. Evenness is a measure of the relative abundance of the different species making up the richness of an area (Krebs, 1978). This measure is constrained between zero and one in which the most even (least variable) community has a value close to one. Shannon’s Diversity Index considers both richness and evenness (Shannon and Weaver, 1949). Sørensen’s similarity index (a measure of beta diversity) was computed for each pairwise combination of establishment groups (Sørensen, 1948); this measure is constrained between zero and one in which a value of zero indicates no overlap between communities, whereas a value of one indicates that exactly the same strains are found in both communities.

Genetic differentiation within a pathogen among geographic populations was tested using an analysis of molecular variance (AMOVA) (Excoffier et al., 1992) using Arlequin 3.1 (Excoffier et al., 2005) to partition genetic variation into within- versus among-population components. Population pairwise FST values were computed to test the null hypothesis of no difference between populations by permutating haplotypes between populations (Perkins et al., 2006). The molecular diversity index ThetaK and the mismatch distribution (Rogers and Harpending, 1992) were computed for each establishment group. The mismatch distribution assesses the distribution of the number of differences (mismatches) between pairs of DNA sequences in a sample. The shape of this distribution is affected by the past demography of a population, such that the distribution is usually multimodal in samples from populations at demographic equilibrium, reflecting the highly stochastic shape of gene trees. Conversely, a unimodal mismatch distribution is characteristic of a population that has passed through a recent demographic expansion. Goodness of fit of the observed mismatch distribution to the sudden expansion model (unimodal distribution) allowed for assessment of significance through the calculation of Harpending’s raggedness index (Harpending, 1994).

To evaluate phylogenetic relationships among B. burgdorferi haplotypes, we constructed an unrooted neighbor-joining phylogeny and minimum spanning network (MSN) using Mega4 and TCS 1.2.1, (Clement et al., 2000; Tamura et al., 2007) respectively. Evolutionary distances were computed using the Kimura 2-parameter method and are in units of number of base substitutions per site. Percentage support values for clades within the neighbor-joining tree were obtained from 1000 bootstrap iterations. The MSN method determines the gene network in which the total length of the branches that connect haplotypes is minimized; discrimination among equal-length MSNs was achieved by assuming older alleles are more common than recently derived alleles, and that new mutations are likely to be found in the same population as their ancestor. Evidence for gene conversion was examined using Sawyer’s Test in GENECONV version 1.81 (http://www.math.wustl.edu/~sawyer/geneconv/), which tests the null hypothesis that nucleotide substitutions observed in a set of aligned sequences are randomly distributed (Sawyer, 1989).

3. Results

3.1. Index of abundance of I. scapularis

Across all sites, a total of 1655 questing ticks were collected, the majority (94.6%) of which were I. scapularis adults (816 females and 749 males). Also collected were 29 nymphal and one larval I. scapularis, 59 adult Dermacentor variabilis, and one adult Amblyomma americanum. The indices of abundance of I. scapularis adults were highly variable within each establishment group, and were not significantly different among groups (P = 0.95), although the trend was for abundance to correlate positively with tick establishment status (26.6, 23.7, and 17.6 adults/1000 m2 at long-established, intermediate-established, and recently-invaded sites, respectively; Table 2).

3.2. Infection of ticks with multiple microorganisms

Five different microorganisms – B. burgdorferi, B. miyamotoi, A. phagocytophilum, B. microti, and B. odoncoci – were detected among 1565 adult I. scapularis from across the Midwest, with overall infection prevalences of 51.3%, 2.2%, 8.9%, 0.3%, and 4.5%, respectively (Fig. 2; Table 2). Long-established ticks were significantly more infected with any microbe than were intermediate-established ticks (59.7% versus 53.9%; P = 0.04), whereas recently-invaded ticks (55.7%) harbored a similar overall infection prevalence to both other groups (P = 0.33 and 0.11). A difference in infection prevalence between females and males was seen only in B. miyamotoi (females, 1.3%; males, 2.9%; P = 0.03) and A. phagocytophilum (females, 7.1%; males, 10.8%; P = 0.01). All microbes except B. microti were found in ticks of all three establishment groups; B. microti was found only in two long-established populations. Infection prevalence for each microbe, except B. odoncoci, and coinfection prevalence trended highest in the long-established ticks, though only significantly higher in long-established ticks for A. phagocytophilum (11.4% versus 6.4% and 3.7% for intermediate-established and recently-invaded ticks, respectively; P = 0.004; Table 2). B. odoncoci infection prevalences trended highest in the recently-invaded ticks (7.0% versus 2.9% and 3.7% for long-established and intermediate-established ticks; P = 0.35). The IGS of all 34 B. miyamotoi-positive samples was sequenced, and all samples were 100% homologous to ‘Type 4’ North American B. miyamotoi (Bunikis et al., 2004b).

Overall, 8.9% of all ticks collected in the Midwest were coinfected with two or more of the target microbes (Table 2; Fig. 2). Male ticks had a higher prevalence of coinfections (11.2%) than females (6.9%; P = 0.003). Coinfections were found in all three establishment groups and were most prevalent among long-established ticks (10.4%; P = 0.03) as compared to intermediate-established (6.7%) and recently-invaded ticks (6.4%; Table 3). Across all samples, the most common coinfection resulted from B. burgdorferi and A. phagocytophilum, which occurred in 90 ticks (5.8% of all ticks); this rate of co-infection is 1.3 times higher than the expectation given the individual prevalence of each microbe.
within the population ($\chi^2 = 11.03; P = 0.0009$). Other coinfections that were observed in a frequency that is significantly different than expected included $B. burgdorferi$ and $B. miyamotoi$ (0.05% of all ticks; 2.1 times lower than expected; $\chi^2 = 9.89; P = 0.001$) and $B. miyamotoi$ and $B. odocoilei$ (0.3% of all ticks; 2.7 times higher than expected; $\chi^2 = 4.62; P = 0.03$). All other pairwise coinfections occurred at frequencies that did not differ from expectation. Additionally, infection of ticks with three microbes was observed 3 times, and a single tick was infected with four microbes. All of these coinfections of three or more microbes occurred in ticks collected at long-established sites.

3.3. $B. burgdorferi$ genotypes

The IGS of 458 $B. burgdorferi$-positive samples was sequenced, comprising 62.0%, 52.7%, and 44.7% of the PCR-positive samples from the long-established, intermediate-established, and recently-invaded ticks, respectively. Across all sites, 21 (4.6%) of these samples had evidence of mixed-strain infections, based on the presence of double-nucleotide peaks in the chromatogram at polymorphic sites (as defined by Bunikis et al. (2004a)). The highest proportion of mixed strain infections was found among long-established populations (5.1%), followed by intermediate-established (3.5%) and recently-invaded (2.9%; $P = 0.05$); mixed strain infections were present at all sites harboring long-established ticks and intermediate-established ticks, but were only present in one of the three sites harboring recently-invaded ticks (Table 4).

3.4. $B. burgdorferi$ RST groups

Among the 437 single strain infections, all three RST groups were represented, with RST 3 in highest abundance (58.3%) followed by RST 2 (35.5%) and RST 1 (6.2%). Strains of RST 1 were least abundant at all sites (Fig. 3), and were found at 3 of 5 long-established sites, both intermediate-established sites, and two of three recently-invaded sites.

3.5. $B. burgdorferi$ strain richness and diversity

We detected a total of 22 IGS strains within 437 $B. burgdorferi$ sequences from across the Midwest. These 22 strain types included 8 prototypical IGS strains previously described by Bunikis et al. (2004a) from the Northeastern United States (IGS 5, 1A, 2A/C, 2D, 4A, 6A, 7A/B, and 8A/C), 4 strains that we have recently detected in bird-associated ticks (Midwest A, B, F, and K (Hamer et al., 2011); and 10 novel IGS variants that differed from published strains by at least one nucleotide (Midwest C, D, E, G, H, I, J, L, M, and N; accession numbers HM015238-HM015247). Observed $B. burgdorferi$ strain richness was highest within ticks of the long-established group (19 strains), followed by intermediate-established (12 strains) and recently-invaded (10 strains), though this relationship is likely influenced by sample size (297, 74, and 66 sequences determined in the three groups, respectively; Table 4). Using a rarefaction analysis, we determined the rate at which new strains are found per sample size (Fig. 4). At a common sample size of at least 50, the long-established group (13.13 strains ± standard deviation (SD) of 1.32) is significantly richer than both the intermediate-established group (10.78 ± 0.91) and recently-invaded group (9.39 ± 0.68). In a separate analysis, the Chao-1 non-parametric estimator of true species richness using data across all sites in the Midwest is 46.5 ± 19.1. Richness is predicted to be greatest at long-established sites (27 ± 8.3) followed by intermediate-established sites (14.3 ± 2.8), and lowest at recently-invaded sites (11 ± 1.6). The percent of samples that comprised local strains (found in a single establishment group only) was greatest among long-established ticks (11.8%) as compared
to intermediate-established (2.7%) and recently-invaded (1.5%; \( \chi^2 = 11.9, df = 2, P = 0.003; \) Fig. 5).

We found that strain evenness among all three establishment groups was comparable (0.83–0.85). Shannon diversity was greatest among the long-established ticks (2.45), followed by intermediate-established (2.7%) and recently-invaded (1.5%; \( F_{ST} = 0.0089; P = 0.06 \)), indicating that the majority of haplotypes are present in all three groups and that high genetic diversity is maintained across the Midwest and continuum of establishment. Population pairwise \( F_{ST} \) analysis indicates that IGS haplotype frequencies of long-established and intermediate-established populations were not different (\( P = 0.32 \)), whereas differences were nearly significant between long-established and recently-invaded populations (\( P = 0.07 \)) and intermediate-established and recently-invaded populations (\( P = 0.03 \)). These differences in population structure are driven by the differences in relative abundance of certain strains: most notably, IGS 8A/C comprised 14.1% and 29.7% of long-established and intermediate-established strains, respectively, yet was not present in the recently-invaded population. IGS 6A comprised 18.2% of long-established strains, respectively, yet was not present in the recently-invaded population. IGS 6A comprised 18.2% of long-established and intermediate-established populations (Fig. 5).

### 3.6. *B. burgdorferi* population structure

A total of 99.11% of IGS molecular variation occurred within establishment groups, and 0.89% occurred among establishment groups. Marginally significant differences were found in the identity and frequency of IGS haplotypes among the three establishment groups (\( F_{ST} = 0.0089; P = 0.06 \)), indicating that the majority of haplotypes are present in all three groups and that high genetic diversity is maintained across the Midwest and continuum of establishment. Population pairwise \( F_{ST} \) analysis indicates that IGS haplotype frequencies of long-established and intermediate-established populations were not different (\( P = 0.32 \)), whereas differences were nearly significant between long-established and recently-invaded populations (\( P = 0.07 \)) and intermediate-established and recently-invaded populations (\( P = 0.03 \)). These differences in population structure are driven by the differences in relative abundance of certain strains: most notably, IGS 8A/C comprised 14.1% and 29.7% of long-established and intermediate-established strains, respectively, yet was not present in the recently-invaded population. IGS 6A comprised 18.2% of the recently-invaded population, but only 8.1% and 2.7% of long-established and intermediate-established populations (Fig. 5).

### 3.7. *B. burgdorferi* phylogeny and network analysis

The broad topology of our IGS phylogenetic tree (Fig. 5) – with three RST groups of which RST 3 is paraphyletic, and further delineation of *B. burgdorferi* into about a dozen intraspecific lineages – corresponds to previous descriptions (Attie et al., 2007; Bunikis et al., 2004a). Some of the tree topology is star-shaped with polytomies that occurred only at sequences that differ by one or two nucleotides. Midwest A was the most common strain across the Midwest and was the only strain to be found at all sites; it constituted 19.5%, 13.5%, and 21.2% of samples from long-established, intermediate-established, and recently-invaded ticks, respectively. Seven strains were singletons within our sampled population (IGS
2A/C, Midwest E, I, J, L, M, and N), and the percent of all strains that were singletons in each establishment group did not differ significantly (21.1%, 25.0%, and 20% in long-established, intermediate-established, and recently-invaded ticks; \( P = 0.95 \)). Strains shared by more than one establishment group were the most abundant within the sample, with eight ubiquitous strains found in all groups and three strains found in two groups. Of the 11 strains that were only found within one establishment group, the majority (72.7%) occurred in long-established ticks, whereas two such strains occurred in intermediate-established ticks, and a single such strain occurred in recently-invaded ticks. The unique IGS strains from each establishment group are interdigitated with those of the other establishment groups as well as with ubiquitous strains, suggesting a recent shared history. Within our sampled population, the maximal number of mutational steps separating a strain from the next most homologous strain was three (Fig. 6).

4. Discussion

There have been many criteria proposed for establishing the center of origin of a taxon that collectively may provide a framework for assessing diversity within and among species in relation to establishment and invasion (Cain, 1943). While these criteria were originally developed in the study of plant geography, we pose that some of the more frequently used criteria (Crisci et al., 2003) have direct application to better understanding the establishment of the \( I.\ scapularis \)-borne disease systems. Below we recapitulate our tick and pathogen data in the context of four of these criteria, and conclude that the data support the hypothesis that the long-established populations of \( I.\ scapularis \) serve as a center of origin for the Midwestern \( I.\ scapularis \)-borne pathogen systems.

4.1. The location of the area of greatest dominance and density of distribution

At the tick level, a gradient was apparent in the indices of abundance of \( I.\ scapularis \) in which the highest tick abundances occurred within the most long-established populations (in Wisconsin and Minnesota), similar to the findings of Walk et al. (2009) in New Hampshire. Similarly, in a standardized survey for \( I.\ scapularis \) nymphs across their distributional range, Diuk-Wasser et al. (2006) found that within the Midwestern sampling sites, highest densities were in Minnesota and Wisconsin, at densities comparable to those in Northeastern endemic areas. While the overall trend in our dataset was for highest tick abundance in longer-established populations, variation was apparent; most notably, the highest

---

**Table 4**

Mixed strain infections and strain richness of single strain \( B.\ burgdorferi \) infection in 458 infected adult \( I.\ scapularis \) adults across a continuum of \( I.\ scapularis \) establishment in the Midwestern United States, 2006–2007. UP = Upper Peninsula; LP = Lower Peninsula. The bold rows are a summary of each establishment group.

<table>
<thead>
<tr>
<th>Status</th>
<th>Site</th>
<th>State</th>
<th>No. sequences</th>
<th>Proportion mixed-strain</th>
<th>No. single-strain sequences</th>
<th>Strain richness</th>
<th>Theta_k</th>
</tr>
</thead>
<tbody>
<tr>
<td>Long-established</td>
<td>St. Croix State Forest</td>
<td>MN</td>
<td>47</td>
<td>0.02</td>
<td>46</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Castle Rock State Park</td>
<td>IL</td>
<td>48</td>
<td>0.15</td>
<td>41</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Governor Dodge State Park</td>
<td>WI</td>
<td>73</td>
<td>0.03</td>
<td>71</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Fort McCoy</td>
<td>WI</td>
<td>94</td>
<td>0.05</td>
<td>89</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Menominee North and South</td>
<td>MI UP</td>
<td>51</td>
<td>0.02</td>
<td>50</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td></td>
<td>All</td>
<td></td>
<td>313</td>
<td>0.05</td>
<td>297</td>
<td>19</td>
<td>4.37 (2.64-6.97)</td>
</tr>
<tr>
<td>Intermediate-established</td>
<td>Indiana Dunes</td>
<td>IN</td>
<td>19</td>
<td>0.05</td>
<td>18</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tippecanoe River State Park</td>
<td>IN</td>
<td>58</td>
<td>0.03</td>
<td>56</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td></td>
<td>All</td>
<td></td>
<td>77</td>
<td>0.04</td>
<td>74</td>
<td>12</td>
<td>3.81 (1.98-7.03)</td>
</tr>
<tr>
<td>Recently-invaded</td>
<td>Saugatuck and Duck Lake</td>
<td>MI LP</td>
<td>5</td>
<td>0.00</td>
<td>5</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Van Buren State Park</td>
<td>MI LP</td>
<td>52</td>
<td>0.04</td>
<td>50</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Chicago North and South</td>
<td>IL</td>
<td>11</td>
<td>0.00</td>
<td>11</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>All</td>
<td></td>
<td>68</td>
<td>0.03</td>
<td>66</td>
<td>10</td>
<td>3.04 (1.49-5.89)</td>
</tr>
</tbody>
</table>

---

**Fig. 3.** Variation in proportion of ribosomal spacer (RST) types 1, 2, and 3 of \( B.\ burgdorferi \) across the Midwestern United States, 2006–2007.

**Fig. 4.** Rarefaction analysis of the influence of sample size of detected strain richness of \( B.\ burgdorferi \) within adult \( I.\ scapularis \) across the Midwestern United States, 2006–2007. Predicted mean strain richness (intervals are standard deviation) is modeled at incremental sampling for each tick establishment group until the observed datapoint is reached. The observed datapoint for each group is plotted as the final datapoint on each curve.
Adelson et al. (2004) report 33.6%, Varde et al. (1998) report 43%, 17%, and 5% infection in two different Lyme disease-endemic areas in New Jersey in the late 1990s. We observe here.

The efficiencies of the vector, the degree to which other reservoirs and vectors are involved in maintenance, and the efficiency of the I. scapularis-vector relationship are key to understanding the infection prevalence we observe here.

Using a larger dataset over a broader spatial and temporal scale, we have found that newly-invaded populations had nearly equivalent B. burgdorferi prevalences with that of long-established populations, which suggests that there is not a long time lag between tick invasion and build-up of high density pathogen prevalence within these populations. This may be a consequence of ecological conditions facilitating rapid subsequent invasion of the pathogen (Ogden et al., 2013); simultaneous dual-invasion process of both the tick and the pathogen (Hamer et al., 2010); or from the existence of cryptic pathogen maintenance cycles, in which certain host species (Herwaldt et al., 2003). This parasite is an intraerythrocytic protozoan parasite associated with white-tailed deer and other cervids, and is genetically similar to Babesia divergens, the causative agent of babesiosis in cattle, and a zoonotic Babesia species (Herwaldt et al., 2003). This parasite has a north-central and northeastern distribution in the United States (Steiner et al., 2006) as well as in Texas, Oklahoma, and Virginia (Waldrup et al., 1990). Its widespread infection in human-biting ticks, combined with its genetic relatedness to the human-infectious Babesia species, implies that it may cause

Fig. 5. Unrooted neighbor-joining phylogram and frequency distribution of B. burgdorferi IGS haplotypes collected from across the Midwestern United States, 2006–2007. The percentages of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches when 60 or above. Haplotypes previously reported in Bunikis et al. (2004a) begin with ‘IGS’; all strains beginning with ‘Midwest’ were not previously reported. Strains conforming to the 16S–23S rRNA RST group 1, 2, and 3 designations by the criteria of Liveris et al. (1995) are demarcated by the labels at the top left of shaded or unshaded groups. The representation of each strain is expressed as a percent within each establishment group.

Others and we have previously found that B. burgdorferi infection prevalence is notably higher in Lyme disease endemic areas (Diuk-Wasser et al., 2012; Hamer et al., 2007; Walk et al., 2009).
disease in immunocompromised people (Armstrong et al., 1998). The ubiquity with which we found this microbe across the Midwestern *I. scapularis* range (8 of 10 sites) underscores the importance for diagnostic assays that can differentiate it from the genetically and morphologically similar agent that poses known public health risk at this time (i.e., *Ba. microti*).

4.2. The location of the greatest variety of forms of the taxon

The number of microbial taxa was greatest in the highly-established tick populations, where all five microbes were present (*B. burgdorferi*, *B. miyamotoi*, *A. phagocytophilum*, *Ba. microti*, and *Ba. odocoiyer*). *Ba. microti* was not present in intermediate-established or recently-invaded ticks. Similarly, as demonstrated by the population-specific rarefaction curves (Fig. 4), *B. burgdorferi* strain richness was greatest in highly-established tick populations. Given a common sample size of 50 infected ticks in all groups, we found statistically greater strain richness at highly-established sites versus intermediate-established and recently-invaded sites. Nevertheless, the diversity of IGS genotypes that we found within recently-invaded ticks was higher than we had anticipated given the decades less time available for colonization of these diverse strains. Similarly, Wang et al. (1999b) found that the ospC variation within a local population was almost as great as the variation of a similar-sized sample of the entire species. It may be that infected ticks are invading from areas of high pathogen prevalence, or that uninfected, invading ticks readily encounter infectious hosts in the areas into which the ticks are invaded. Hosts could be previously infected due to pathogen transmission from other tick species (Hamer et al., 2011) or could be highly vagile hosts, such as birds, that were infected originally in another endemic site and then also dispersed into the recently-tick-invaded site (and may also be a vehicle for tick dispersal) (Humphrey et al., 2010).

Diversifying selection may influence patterns of *B. burgdorferi* strain diversity through at least two mechanisms. First, host associations, facilitated by selective killing of strains by host-specific innate immune factors (Kurtenbach et al., 2002), may maintain diverse populations of Lyme disease spirochetes, at both the species level (e.g. in Europe, *B. burgdorferi* and *Borrelia afzelii* are mammal-associated, whereas *Borrelia garinii* and *Borrelia valaisiana* are bird-associated (Dubbska et al., 2009)) and at the intraspecies level of *B. burgdorferi* (Brisson and Dykhuizen, 2004). The magnitude of intraspecies host association, however, is debated (Hanincova et al., 2006). Secondly, the antigenic types that have already infected a host cannot establish a subsequent infection in the same host because of the immune response. Therefore, rarer antigenic types are able to infect previously-infected hosts, which leads to hosts acquiring multiple strains and maintenance of high levels of variation at the population level (Wang et al., 1999b).

4.3. The greatest number of overlapping distributions

Coinfections – acquired from sequential feeding events on infected hosts or from a single feeding event of a coinfected host – are generally expected to occur in highest frequency in sites with well-established transmission as opposed to recently-colonized foci with many naïve hosts and a reduced diversity of strains due to invasion by one or a few founding ticks. In addition to potentially more complex symptoms or severe disease, coinfections are epidemiologically important because human infection with multiple pathogens may lead to incomplete diagnosis and insufficient treatment (Horowitz et al., 2013; Swanson et al., 2006). At the interspecies level, we found that ticks coinfected with multiple organisms were indeed most common at long-established sites, where 10.4% of ticks harbored more than one microbe (compared with 6.7% and 6.4% at intermediate-established and recently-invaded sites, respectively). Steiner et al. (2008) also found significant coinfection of these pathogens in adult *I. scapularis* from Ft. McCoy, one of our long-established sites. In a review of coinfections of *B. burgdorferi* and *A. phagocytophilum* within *Ixodes* ticks from California, Wisconsin, and the Northeast, the prevalence of coinfections was highest among ticks collected from regions of Lyme disease endemicity in the Northeast (1–28% coinfection), where infection with each individual pathogen is also highest, relative to coinfection in Wisconsin (2%) and California (1% (Swanson et al., 2006)).

At the intraspecies level, mixed strain *B. burgdorferi* infections were negligibly most common in highly-established ticks, where 5% of *B. burgdorferi*-infected ticks harbored more than one IGS strain (compared with 4% and 3% in intermediate-established and recently-invaded ticks, respectively). There was significant variability in the proportion of mixed-strain infection within each establishment group, however, including a maximum of 15% at Castle Rock, a highly-established site in Northwestern Illinois. In comparison, Gatewood et al. (2009) assessed a longer region of the IGS, and report an overall prevalence of mixed strain infections of approximately 20% in nymphs from across the range of *I. scapularis* establishment, including sites in both the Northeast
and Midwest, with no apparent geographic bias. Using the ospA and ospC gene targets, Wang et al. (1999b) found 45% and 50% mixed-strain infections in adult I. scapularis from the endemic area of Shelter Island, NY (the sites from which B. burgdorferi was first isolated), and Guttmann et al. (1996) found 60% of adults from the same areas harbored at least two strains of B. burgdorferi. Our method of sequence analysis of the IGS appears not to be as sensitive as previous analyses of ospC in detecting multiple strains. Furthermore, sequences we obtained that appear to contain only one IGS variant (and were reported as such) may in fact be a biased result if the most abundant strain within a coinfected tick was preferentially amplified and therefore became the sole strain comprising the sequence. For these reasons, our report of mixed strain prevalence should be considered a minimum.

4.4. The identification of continuity and direction of individual variations or modifications radiating from a ‘center of origin’ along the highways of dispersal

Assessment of the structure of the genetic data could allow for speculation as to the mode or directionality of tick and pathogen dispersal, assuming a center of origin. For example, in one scenario, if the long-established I. scapularis populations were the most dense, contained the highest prevalence of infection with multiple pathogens and coinfections, and the highest diversity of B. burgdorferi strains, and if the intermediate-established and recently-invaded tick populations were characterized as nested subsets of these parameters, then a linear mode of invasion may be suggested, with length of establishment a useful predictor of the parameters. Conversely, if recently-invaded ticks were equally or more abundant and infected with diverse pathogens, then a directionality of dispersal from established foci could not be concluded. Across the Midwest, we found marginally significant global population structure (P = 0.06). The majority of strain types found in any establishment group were shared among all three groups, suggesting high rates of gene flow. However, the presence of local strains that we observed within each establishment group suggests some isolation by distance and an emerging evolutionary divergence.

In our dataset, the IGS haplotype frequencies of the recently-established population were significantly different from both the highly-established and intermediate-established populations. The network analysis exhibited some areas with star-like topology with low levels of sequence divergence and a high frequency of unique mutations, and is indicative of a rapid population expansion in the absence of selection. Similarly, analysis of mismatch distributions suggests that a recent demographic expansion has occurred in the recently-invaded population (P = 0.02) and intermediate-established population (P = 0.05), but with no evidence for recent expansion in the long-established population (P = 0.41). The trend in the molecular diversity index Theta_k was that long-established sites were the most diverse, and recently-invaded sites were least diverse, but this difference was not significant.

The diversity gradient that we hypothesized was in fact quite shallow, with regard to measures of richness, as much of our data show that recently-invaded sites harbor equally or only slightly less diverse assemblages of B. burgdorferi than long-established sites. Speculation as to the mechanisms that may account for the observed pattern may include the following: (i) There is a high propagule pressure associated with invasion, in which both the tick and the pathogen invade new sites via diverse wildlife reservoirs and/or alternative tick species. With many independent introduction events, of which each may constitute introduction of at least one different strain, founder effects are not apparent. (ii) Even if a limited number of strains are introduced to a new area, the many different reservoir species of I. scapularis-borne pathogens each present different selective pressures, which leads to increased diversity (Brisson and Dykhuizen, 2004). (iii) The time scale for build up of diverse populations may be rapid (given our evidence for rapid demographic expansion) such that even if initial invading diversity is low, our sampling occurred too late to appreciate this. (iv) A pre-existing cryptic cycle of the pathogen may be maintained in advance of the invasion front of I. scapularis, such that I. scapularis may encounter wildlife infected with diverse strains upon its own invasion (Hamer et al., 2011, 2010).

4.5. Epidemiological significance

Two main single locus genetic typing schemes have been most widely utilized to assess pathogenicity of different B. burgdorferi strains in humans: RST (a tripartite classification of the 25 or more IGS strains) and outer surface protein C (ospC; an antigenic plasmid-borne gene). Newer analyses are based on multilocus sequence typing (Hanincova et al., 2013). Some RST 1 strains are associated with a higher frequency of disseminated infection in humans and more invasive disease in experimental animals (Derdakova et al., 2004; Hanincova et al., 2008; Seinost et al., 1999; Wang et al., 2002; Wormser et al., 2008), and a bias toward the relatively less invasive RST 2 and 3 strains has been found among infected I. scapularis in the Midwest (Gatewood et al., 2009; Hamer et al., 2011, 2010; Humphrey et al., 2010). Our data similarly demonstrate a strong bias toward RST 2 and 3 infections (RST 1 was only represented in 6.2% of infections across our Midwestern sites). While multiple different RST 1 strains were found only within highly-established ticks, RST 1 strains were present within ticks of all establishment groups, including recently-invaded ticks. These data underscore the rapidity with which newly detected tick populations may pose a public health risk.

The pathogenicity in humans of the large number of novel IGS mutant strains is unknown, though we have no a priori reason to believe that single or double nucleotide polymorphisms within the non-coding intergenic spacer region would necessarily correspond to a change in virulence. To begin to assess the epidemiological risk associated with novel strains found within this cryptic cycle, we sent total DNA from 88 B. burgdorferi samples comprising five IGS strains that were not previously reported in the literature or in Genbank (Midwest A, E, J, K, and M) to University of California-Irvine for direct ospC typing. Of these, ospC was successfully amplified and sequenced from 87 samples, resulting in detection of 48 single strain infections and 39 mixed-strain infections (Barbour and Travinsky, 2010). Present in single strain infections included 18 ospC types, including representatives of the four ospC major groups that have been associated with disseminated human Lyme disease (Seinost et al., 1999). The degree to which mixed strain infections were present at the ospC locus within these samples (which had no evidence of mixing at the shorter IGS locus we analyzed) further underscores the complexity of the maintenance of pathogen diversity across the Midwest. In light of findings that recurrent Lyme disease is due to reinfection of patients with different strains of B. burgdorferi (Nadelman et al., 2012), local data on the pathogen strain diversity are helpful for assessing risk.

Across the Midwest, geographic variation in the incidence of confirmed Lyme disease (Hali-Baker et al., 2009) reflects the trends we observed in adult I. scapularis abundance and infection prevalence, although the differences we observed across tick endemicity regions were subtle. Wisconsin and Minnesota reported higher incidences (32.4 and 23.8 cases/100,000 people, respectively) in 2007, the year our tick sampling occurred. Similarly, Minnesota and Wisconsin reported higher incidences of human anaplasmosis (6.2 and 1.2 cases/100,000 people, respectively) in 2007, the year our tick sampling occurred. Similarly, Minnesota and Wisconsin reported higher incidences of human anaplasmosis (6.2 and 1.2 cases/100,000 people,
respectively) than did other states we investigated, for which only Illinois reported data (0.05 cases/100,000 people). We expect the future incidence of I. scapularis-borne disease in the Midwest to increase as I. scapularis populations continue to spread and grow, and as robust, diverse populations of *B. burgdorferi* and other pathogens circulate in areas where the tick establishes. Climate change will likely contribute to the changing dynamics of the zoonotic cycles, allowing for expansions of tick and pathogen populations. Given the linkages of climate to zoonotic cycles, allowing for expansions of tick and pathogen populations, the changing dynamics of the *Ixodes scapularis* (Acari: Ixodidae) and *Borrelia burgdorferi* genotypes (Gatewood et al., 2009), climate change may also promote the emergence of previously cryptic pathogen strains with uncertain human health consequences.

**Acknowledgements**

We thank M. Roberto Cortinas, Michelle Rosen, and Jennifer Sidge for assistance with tick collections. Funding provided by NSF Doctoral Dissertation Improvement Grant No. 0910025.

**References**


