
Evolution of Northeastern and Midwestern *Borrelia burgdorferi*, United States

Dustin Brisson, Mary F. Vandermause, Jennifer K. Meece, Kurt D. Reed, and Daniel E. Dykhuizen

The per capita incidence of human Lyme disease in the northeastern United States is more than twice that in the Midwest. However, the prevalence of *Borrelia burgdorferi*, the bacterium that causes Lyme disease, in the tick vector is nearly identical in the 2 regions. The disparity in human Lyme disease incidence may result from a disparity in the human invasiveness of the bacteria in the Northeast and Midwest caused by fundamentally different evolutionary histories. *B. burgdorferi* populations in the Northeast and Midwest are geographically isolated, enabling evolutionary divergence in human invasiveness. However, we found that *B. burgdorferi* populations in the Northeast and Midwest shared a recent common ancestor, which suggests that substantial evolutionary divergence in human invasiveness has not occurred. We propose that differences in either animal ecology or human behavior are the root cause of the differences in human incidence between the 2 regions.

Lyme disease, caused by the bacterium *Borrelia burgdorferi*, is the most common vector-borne disease in the United States (1). *B. burgdorferi* is transmitted to humans during the blood meal of an infected *Ixodes* tick, predominantly *Ixodes scapularis* in North America (2). The prevalence and density of *B. burgdorferi*-infected *I. scapularis* ticks are nearly identical in the northeastern and midwestern United States, the regions with the highest incidence of Lyme disease in humans (3–6); however, the number of human Lyme disease cases reported in the Northeast and

Midwest is not (1). The overwhelming majority of Lyme disease cases in the United States are reported from the Northeast (82%), followed distantly by the Midwest (9%) (1). Similarly, per capita Lyme disease incidence is >2× greater in the Northeast than the Midwest. We address the hypothesis that *B. burgdorferi* populations in the Northeast and Midwest have fundamentally different evolutionary histories, which may result in differing degrees of human invasiveness.

The evolutionary and demographic histories of *B. burgdorferi* in the Northeast have been intensively studied. The effective population size of northeastern *B. burgdorferi* is small because of its recent colonization of the northern United States and its life-history strategy (7,8). The strikingly impoverished neutral genetic diversity and high linkage disequilibrium within *B. burgdorferi* populations likely result from small effective population sizes (9–11). Genetic loci are found in perfect or near-perfect association in *B. burgdorferi* in the Northeast (9–12). Strong linkage disequilibrium among genetic loci can result from several evolutionary and ecologic forces in addition to small population size (drift), such as lack of recombination machinery or limited opportunity for gene exchange (13). Genetically diverse strains of *B. burgdorferi* often are found within the same tick or same vertebrate host, suggesting ample opportunity for genetic exchange (4,14). Evidence is strong that recombination occurs within a genomic lineage of *B. burgdorferi* (15–17). Thus, *B. burgdorferi* has the opportunity and the recombination system needed for genetic exchange. A historically small effective population size is a parsimonious explanation for the low neutral genetic diversity and strong linkage disequilibrium.

The evolutionary and demographic histories of *B. burgdorferi* in the Midwest are comparatively understudied. On the coarse scale, the evolutionary history and ecol-

Author affiliations: University of Pennsylvania, Philadelphia, Pennsylvania, USA (D. Brisson); Marshfield Clinic Research Foundation, Marshfield, Wisconsin, USA (M.F. Vandermause, J.K. Meece); Northwestern University/Feinberg School of Medicine, Chicago, Illinois, USA (K.D. Reed); and Stony Brook University, Stony Brook, New York, USA (D.E. Dykhuizen)

DOI: 10.3201/eid1606.090329

ogy of *B. burgdorferi* in the Northeast and Midwest appear similar. Both regions have oak-maple-dominated forests ideal for deer and the small mammals that maintain the *I. scapularis* and *B. burgdorferi* populations, were under the Pleistocene ice sheet, and were recently colonized by *B. burgdorferi*. The differences accounting for a lower Lyme disease incidence in the Midwest than the Northeast are not clear but most likely can be found on a finer scale in the evolutionary history of *B. burgdorferi* (this study), the tick vector (8), or human exposure.

The differences accounting for a lower Lyme disease incidence in the Midwest than the Northeast are not clear but are likely to be found on a finer evolutionary or ecologic scale. A recent report suggested that a lower proportion of the 16S–23S rRNA intergenic spacer (IGS) type 1 (RST-1) allele in midwestern *B. burgdorferi* populations, an allele associated with human invasiveness in the northeastern United States, may account for the differences in human Lyme disease incidence (8). However, the 16S–23S rRNA IGS does not directly influence *B. burgdorferi* invasiveness; it is in linkage disequilibrium with a gene of major effect in the Northeast. The linkage patterns in midwestern *B. burgdorferi* are as yet unstudied. A fundamentally different evolutionary history would result in divergent linkage disequilibrium patterns between the Northeast and the Midwest and potentially result in differing degrees of human invasiveness associated with alleles at the 16S–23S rRNA spacer.

For this study, we used a phylogenetic framework to analyze the evolutionary and demographic histories of *B. burgdorferi* in the midwestern and northeastern United States (18). A previous multilocus sequencing typing study of 4 *B. burgdorferi* loci (outer surface protein C [*ospC*], 16S–23S rRNA (*rrs-rrlA*), *ospA*, and outer membrane protein [p66]) from strains isolated in the Northeast identified 9 distinct lineages with complete linkage among alleles at the 4 loci (10). That study analyzed statistical associations of haplotypes at each locus without regard to the underlying evolutionary relationships of sequences. In this study, we analyzed *ospC*, *rrs-rrlA*, *ospB*, and *ospA* (p66 contains little evolutionary information [10]) from midwestern and northeastern populations in a phylogenetic framework to investigate the shared and vicariant evolutionary and demographic histories of *B. burgdorferi* from geographically isolated regions.

Methods

B. burgdorferi Isolates

All isolates were derived from skin biopsy specimens of 47 adult patients who had erythema migrans, at the Marshfield Clinic in central Wisconsin during 1995–2001. Specimens were collected and cultured as described elsewhere (19).

DNA Extraction and Amplification

The DNA sequences of *ospA*, *ospB*, *rrs-rrlA*, and *ospC* were determined for use in our phylogenetic analysis. Cultivated isolates were harvested by centrifugation at $5,000 \times g$ for 15 min, resuspended in sterile water, and lysed by boiling for 5 min. PCR conditions are described by Bunikis et al. (10) for *ospA* and *rrs-rrlA*, by Caporale and Kocher (20) for *ospB*, and by Brisson and Dykhuizen (4) for *ospC*. Negative controls were included for DNA extraction and PCR procedures to monitor for contamination. Amplified PCR products were sequenced in both directions. *ospC* products were subject to the reverse-line blot procedure as described previously (4,7). *ospC* amplicons that could not be classified to a major allelic group were sequenced.

Analyses

All analyses included sequence data collected in this study and reported in Bunikis et al. (10). DNA sequences were aligned using the Clustal X algorithm (21) with default settings, and alignments were refined manually where necessary. Descriptive statistics (π , synonymous and nonsynonymous polymorphisms) were determined using DNAsp (ver. 4.50.3) (22). Two tests for recombination within genes (Sawyer test and maximum χ^2 test) and a test for allelic association (index of association [I_A]) were performed in START2 (23). Sawyer runs test compares pairs of alleles to determine whether regions of the sequence space have more consecutive identical polymorphisms (runs) than expected by chance (24). The maximum χ^2 test uses the distribution of polymorphic sites to identify potential recombination events between pairs of alleles (25). The I_A assesses the extent of association between loci using allele frequencies without regard to the underlying sequences (26).

Phylogenetic reconstruction was performed using Bayesian (MrBayes; <http://mrbayes.csit.fsu.edu>) and maximum likelihood (PAUP*, version 4.0; <http://paup.csit.fsu.edu>) approaches. Trees were constructed from the sequence data from each gene individually and from combinations of genes. IGS, *ospA*, and *ospB* provide reliable data for phylogenetic inference and can be easily compared with previous analyses (10). *ospC* DNA sequences provide information about disease invasiveness (27,28) but were not used in phylogenetic reconstruction because the gene tree topology is star-shaped with short and unsupported internal branches, long terminal branches, and many polytomies. This topology provides no information about the evolutionary history of this gene or the relationships among alleles. When information about *ospC* was included in the data set, it was included as a heavily weighted morphologic character in MrBayes, not as DNA sequence data. The morphologic data constrain the isolates with the same *ospC* allele cluster but do not affect the topology of the internal branches of the tree, which depend only on the DNA sequences of

the other genes in the data set. For each data set, we used Modeltest 3.4 (29) to select the appropriate model of molecular evolution. We used this model to find the Bayesian tree and the maximum likelihood tree.

We used the Shimodaira-Hasegawa (SH) test as implemented in PAUP to test for differences in evolutionary histories among genes using the resampling of estimated log-likelihoods approximation with 1,000 bootstrap replicates (18). This method tests for significant differences in the likelihood of several given tree topologies given a data set. The likely tree topologies produced by each locus analyzed in this study were used as the set of topologies. If no true difference exists, each data set is equally likely to produce any of the given tree topologies, and all genes in this study share similar evolutionary histories. Alternatively, a significantly lower likelihood suggests that recombination has occurred between the genes in the data set and the gene used to create the lower-likelihood tree topology.

Results

Sequence Diversity

B. burgdorferi isolates were cultured from primary or secondary skin lesions of 47 adults visiting the Marshfield Clinic during 1995–2001. The sequences at the IGS, *ospA*, *ospB*, and *ospC* loci were determined from 39, 31, 44, and 44 strains of these *B. burgdorferi* isolates, respectively. Erythema migrans represented the bulk of the diversity found in tick samples (27) and provided a representative sample of *B. burgdorferi* bacteria in the Midwest. Furthermore, we found all of the IGS types found by Bunikis et al. (10), suggesting the sample from the Midwest was not biased in terms of the types found. *ospA* showed little evolutionarily information and was not determined from 14 of the isolates. One sequence was found at each locus in each culture.

Nucleotide diversity varied considerably among loci (Table 1). We found 47 polymorphic sites in the IGS sequence sample, resulting in 16 unique haplotypes (online Technical Appendix, www.cdc.gov/EID/content/16/6/911-Techapp.pdf). Six haplotypes were identical to haplotypes previously found in the Northeast, and 38 of the 47 polymorphic sites found in the Northeast were present in this sample from the Midwest, supporting a recent shared ancestry of *B. burgdorferi* from the 2 populations (online Technical Appendix). The 10 unique haplotypes in the Midwest

and 9 unique polymorphic sites suggest some isolation by distance and an emerging evolutionary divergence.

ospA and *ospB* are considerably less diverse than IGS (Table 1). Nevertheless, there are 14 unique *ospA* haplotypes and 14 unique *ospB* haplotypes in this sample (online Technical Appendix). Nine of the *ospA* haplotypes in the samples from the Midwest also are found in *B. burgdorferi* in the Northeast. The diversity of *ospB* has not been examined in the Northeast. Amino acid sequence evolution in *ospA* and *ospB* appears to be constrained because synonymous changes greatly exceed nonsynonymous changes ($dN_{ospA}/dS_{ospA} = 0.334$; $dN_{ospB}/dS_{ospB} = 0.303$).

The mean pairwise diversity at *ospC* was much greater than at the other loci analyzed ($\pi = 0.207$). Sixty-three changes resulted in amino acid substitutions. However, we found no evidence of directional selection ($dN/dS = 0.591$). Eighteen major groups of alleles, which differ by >8% in sequence, are represented in the isolates from the Midwest (online Appendix Table, www.cdc.gov/EID/content/16/6/911-appT.htm). Fourteen of the 18 major group alleles are identical to those found in the Northeast, although types L and O are rare in the Northeast. We found 4 novel *ospC* major group alleles (W, V, Y, Z) in these samples (online Appendix Table). *ospC* major groups W, V, and X also were detected in tick populations in the Midwest (D. Brisson, unpub. data), although we did not see type X in this set of human isolates. Strains with *ospC* major groups I, J, and T, which are present in tick and human samples in the Northeast, were absent in this sample from the Midwest.

Phylogenies

We used Bayesian and likelihood algorithms to reconstruct phylogenies for each gene. The 2 approaches resulted in identical topologies for each gene except *ospC*. The IGS gene tree forms a strongly supported phylogeny with most nodes bifurcating (Figure). Polychotomies occur only at sequences that are identical or differed by 1 nucleotide. The phylogeny generally supports the previously described divisions of IGS sequences into the 3 restriction fragment length subgroups, arbitrarily called RST 1, RST2, and RST3 (31). However, the diverse RST3 group is not monophyletic. The IGS sequences from strains in the Northeast and Midwest are interdigitated, suggesting a recent shared history.

The gene trees for *ospA* and *ospB* individually have many weakly supported nodes and polychotomies because

Table 1. Genetic markers and population diversity of *Borrelia burgdorferi*, midwestern and northeastern United States, 1999–2001*

Marker	No. samples	π	dN/dS	Maximum χ^2 , p value	Sawyer test, significant fragments
IGS	39	0.014	NA	107, p<0.01	0
<i>ospA</i>	31	0.0033	0.334	0	0
<i>ospB</i>	44	0.0042	0.303	739, p<0.05	0
<i>ospC</i>	44	0.207	0.591	74, 83, 90, 137, 338; p<0.001	4

*dN/dS, directional selection; IGS, intergenic spacer; NA, not applicable; osp, outer surface protein.

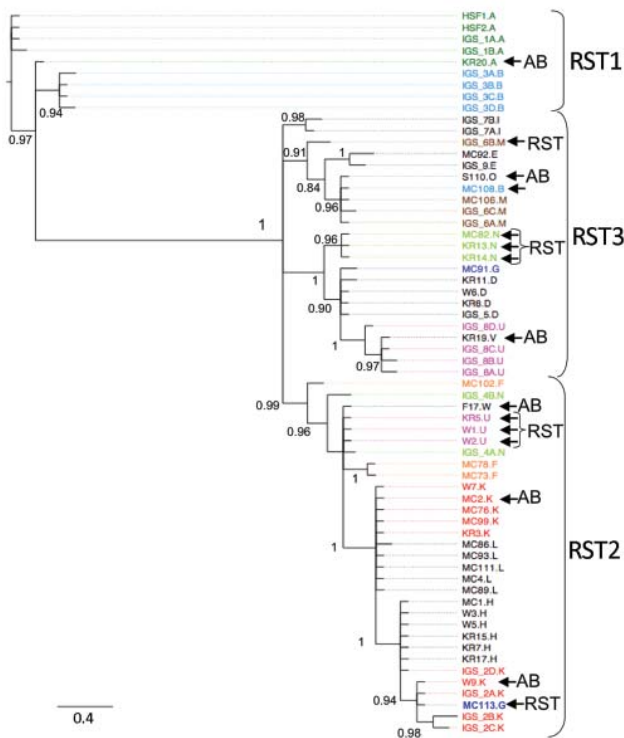


Figure. Phylogeny of *Borrelia burgdorferi* isolates in the northeastern and midwestern United States based on intergenic spacer (IGS) sequence. operational taxonomic unit names beginning with IGS were isolated in the northeastern United States (10); all other isolates are from patients in the Midwest. The letter after period designates the outer surface protein C (*ospC*) major allele of the isolate. Colored isolate names highlight isolates with the same *ospC* major group that cluster in different clades, which suggests horizontal gene transfer. The *ospC* of several strains is not linked to the IGS ribosomal spacer type (RST) to which it is commonly linked in the Northeast (10,34). AB indicates differences between the *ospAB* tree and the IGS tree. This tree is midpoint rooted. Scale bar indicates number of substitutions per site.

of the limited number of phylogenetically informative sites. The phylogeny reconstructed from the *ospA/ospB* operon (concatenated data set) result in much stronger support for internal nodes. Nevertheless, the *ospAB* phylogeny contains 1 polychotomy. Concatenating the data sets for *ospA* and *ospB* sequences is appropriate because these loci lie contiguously in a single operon, and little evidence exists for recombination either within or between these genes (see below). The northeastern and midwestern strains were interdigitated on the *ospA*, *ospB*, and *ospAB* trees, which suggests a recent, shared population history.

Analysis of the *ospC* data set yielded a phylogeny with short internal branches, long terminal branches, and no supported internal nodes, commonly referred to as a star phylogeny (10,13). *ospC* thus appeared to have evolved by recombination (9,32). Phylogenies of highly recombining

loci are poorly supported because the evolutionary history differed in different segments of the allele (33).

Intragenic Recombination

We found little evidence of recombination within IGS, *ospA*, or *ospB* in the Midwest data set or the combined northeastern/midwestern data set (Table 1). Sawyer runs test found no runs of polymorphisms that are significantly incongruous within the sequence. Maynard Smith's maximum χ^2 test found 1 recombination event within the IGS sequence at nt 107 (position compared with B31 type strain) in both data sets. We found no evidence for recombination within *ospA* and only 1 potential site of recombination at the extreme 3' end of *ospB*. Neither maximum χ^2 nor Sawyer test found evidence for recombination between *ospA* and *ospB* in the *ospAB* operonic data set. However, *ospA* haplotypes are often associated with multiple *ospB* haplotypes suggesting recombination or independent evolution (online Technical Appendix). Intragenic recombination was evident within *ospC* sequences. Sawyer test identified 4 runs of polymorphisms that appear to be transferred from another *ospC* major allelic group. Maynard Smith maximum χ^2 test found evidence of recombination at 5 locations along the sequence (Table 1).

Intergenic Recombination

We assessed intergenic recombination using the I_A and the SH test. The index of association detects nonrandom associations of haplotypes among genes but does not account for sequence variation or relatedness. The I_A indicated that all loci examined in this study were significantly nonrandomly associated, i.e., in linkage disequilibrium ($p < 0.0001$) (Table 1).

The SH test uses phylogenetic information to identify horizontal transfer events by comparing the evolutionary histories of genes as represented by the gene-tree topologies. The likelihood of the IGS sequence data resulting in the *ospAB* tree topology is significantly less likely than these data resulting in the IGS topology (Table 2). That is, the evolutionary history of the IGS locus is not congruent with the *ospAB* evolutionary history, providing evidence of lateral gene transfers. Strains found in different strongly supported clades on the *ospAB* and the IGS trees are identified in the Figure ($p < 0.0001$).

The phylogenies reconstructed from midwestern *ospAB* sequences were compared with *ospAB* phylogenies that constrain strains with the same *ospC* allele to cluster (*ospABC* tree). We found no statistical support for recombination between *ospC* and *ospAB* ($p = 0.313$). However, 4 *ospC* type K strains (W9, W7, MC2, KR3) are separated from the other K strains by supported nodes in the *ospAB* tree, and 1 *ospC* type F strain (MC102) is separated from the other *ospC* type F strains. We also found no evidence of

Table 2. Summary of Shimodaira-Hasegawa test results of potential horizontal gene transfer events in *Borrelia burgdorferi*, midwestern and northeastern United States*

Test comparison	Data set	Δ lnL	p value
<i>ospAB</i> vs.			
IGS	<i>ospAB</i>	8.50709	0.390
<i>ospAB/ospC</i>	<i>ospAB</i>	21.26004	0.313
IGS vs.			
<i>ospAB</i>	IGS	642.826545	<0.0001
IGS/ <i>ospC</i>	IGS	67.53033	0.006
IGS/ <i>ospC</i> †	IGS	388.21228	<0.0001
<i>ospB</i> vs. <i>ospB/ospC</i>	<i>ospB</i>	18.91326	0.338
<i>ospA</i> vs. <i>ospA/ospC</i>	<i>ospA</i>	45.30135	0.146
<i>ospA</i> vs. <i>ospA/ospC</i> †	<i>ospA</i>	79.62589	0.080

* Δ lnL, difference in log-likelihood; *osp*, outer surface protein; IGS, intergenic spacer.

†Included sequence data from Bunikis et al. (10).

recombination between *ospA* and *ospC* or *ospB* and *ospC* ($p > 0.1$). However, the *ospA* tree is nearly significantly more likely than the *ospAC* tree when strains from the Northeast are included in the analysis ($p = 0.08$).

Several instances of recombination are evident between IGS and *ospC*. The differences in evolutionary history were observable by using the midwestern IGS data set alone ($p = 0.006$) and the combined midwestern/northeastern IGS data set ($p < 0.001$). These analyses provide evidence that linkage patterns in *B. burgdorferi* from the Northeast differ from that of the linkage patterns in *B. burgdorferi* from the Midwest.

Discussion

The divergence in human Lyme disease incidence between the Northeast and Midwest does not result from independent evolution of human invasiveness because of geographic isolation. Although pathogen populations in the Midwest appear geographically isolated from those in the Northeast (6), evolutionary and demographic analyses indicate that they share a recent common ancestor. Both populations have little standing genetic variation, as indicated by the limited number of polymorphic sites, suggesting small effective population sizes and similar life-history strategies. The combination of linked alleles also is similar in both regions, supporting the recent shared ancestor hypothesis. *B. burgdorferi* strains isolated in the Midwest are interleaved with northeastern strains on phylogenetic trees evincing their close evolutionary relationship. However, there is some genetic divergence and differing linkage groups between the regions, intimating that gene flow is limited between these populations, allowing them to differentiate. The recent common ancestor in the northeastern and midwestern *B. burgdorferi* and limited genetic divergence suggests that human Lyme disease incidence cannot be explained by fundamentally different evolutionary histories resulting in differing degrees of human infectiousness.

None of the loci investigated show substantial genetic divergence between regions, suggesting a recent common ancestor and similar phenotypes. Northeastern haplotypes are interleaved with midwestern haplotypes such that the time to coalescence of alleles within a region is equivalent to the time to coalescence for alleles from both regions (Figure). These data suggest that northeastern and midwestern strains have a recent common ancestor. The limited genetic diversity in *B. burgdorferi* in the Midwest and Northeast (online Technical Appendix) suggests that the populations have retained the life-history strategy of their common ancestor (10,12). However, isolation by distance and subsequent divergence resulted in unique alleles in each region.

The IGS gene tree reconstructed from midwestern and northeastern data broadly supports the RST system described using northeastern populations (31). RST types 1 and 2 form strongly supported monophyletic groups. RST3 is polyphyletic and should be split into 3 groups as defined by the strongly supported clades (Figure). Supporting this suggestion, RST3 is diverse genetically and phenotypically (34). Interestingly, this division would separate *ospC* major group I bearing strains, a particularly invasive group in humans, from the other RST3 strains that rarely cause disseminated infections in humans (27,28,30).

The *ospC* data support the hypothesis that the strains from the Northeast and Midwest have a common ancestor but are currently isolated and have begun to diverge. Most *ospC* major groups are found in both regions. Given the genetic distance between major group alleles, the exact set of alleles is unlikely to have occurred twice independently. Additionally, the linkage relationships between *ospC* alleles and IGS alleles are similar in both regions. Both lines of evidence suggest that most of the diversity at *ospC* originated before the northeastern and midwestern populations diverged. Differences in invasiveness between *B. burgdorferi* in the Northeast and Midwest do not result from fundamentally different evolutionary histories.

Four novel *ospC* major group alleles appear to be unique to the Midwest (Table 2). In addition to these novel *ospC* major groups, a type C–like allele appears to have been generated independently in the Midwest and the Northeast. The group C allele in the Midwest shares 96.8% similarity with the group C allele in the Northeast. Whether the unique *ospC* major group alleles were generated recently in only 1 region or whether they were shared in the ancestral population and subsequently lost in only 1 region is not clear.

B. burgdorferi lineages exchange DNA, contrary to previous reports (10,11,13). We found at least some evidence of recombination between all genetic loci examined; even the *ospAB* operon has several homoplasious mutations, suggesting potential recombination (online Technical Appendix). However, recombination between *ospA*

and *ospB* is not statistically supported and may have arisen from recurrent mutation (online Appendix Table). Despite evidence for recombination, the linkage relationships are similar in the Northeast and the Midwest, supporting the recent common ancestry of these populations.

Recombination is more apparent in Lyme disease foci in the Midwest than in the Northeast (Figure; online Appendix Table). This is likely to be caused by neutral divergence of linkage patterns resulting from small effective population sizes in both regions coupled with gene flow from the Northeast to the Midwest but not in the other direction. Small effective population sizes eliminate most of the linkage combinations in each region such that they are in perfect linkage disequilibrium. Gene flow from the northeastern population then introduces linkage pattern variation into the midwestern population. Linkage patterns unique to the Midwest are absent from the Northeast, suggesting that gene flow from the Midwest to the Northeast is rare.

B. burgdorferi in the Northeast and Midwest share a remarkably similar evolutionary history. Independent evolution of human invasiveness in the 2 regions does not explain the lower human Lyme disease incidence in the Midwest. Other potential causes for the differences in human Lyme disease incidence include differences in human exposure to *B. burgdorferi*-infected ticks and ecologic differences in the reservoir host community. Lyme disease typically is contracted peridomestically in the Northeast (35), but similar studies reporting the peridomestic acquisition of *B. burgdorferi* have not been reported from the Midwest. Human risk for exposure to Lyme disease also may be exaggerated in the Northeast because of the immense suburban populations around the major metropolitan areas. Current ecologic conditions yielding differences in the composition of reservoir host species could alter the prevalence of *B. burgdorferi* lineages that are particularly invasive in humans (27,28,30,36,37). For example, midwestern ticks are rarely infected with *ospC* genotypes A or B (RST I) (3), 2 of the 4 genotypes that are common in the Northeast and regularly cause human Lyme disease (27,28,30).

This work was supported by National Institute of Health grants AI076342 (D.B.) and GM31912 (D.E.D) and Centers for Diseases Control and Prevention grant U01CK000170 (D.B.).

Dr Brisson is an assistant professor in the biology department at the University of Pennsylvania. His primary research interests focus on the ecology and evolution of infectious diseases.

References

- Bacon RM, Kugeler KJ, Mead PS. Surveillance for Lyme disease—United States, 1992–2006. *MMWR Surveill Summ.* 2008;57:1–9.
- Burgdorfer W, Barbour AG, Hayes SF, Benach JL, Grunwaldt E, Davis JP. Lyme disease—a tick-borne spirochetosis. *Science.* 1982;216:1317–9. DOI: 10.1126/science.7043737
- Gatewood AG, Liebman KA, Vourc’h G, Bunikis J, Hamer SA, Cortinas R, et al. Climate and tick seasonality are predictors of *Borrelia burgdorferi* genotype distribution. *Appl Environ Microbiol.* 2009;75:2476–83. DOI: 10.1128/AEM.02633-08
- Brisson D, Dykhuizen DE. *ospC* diversity in *Borrelia burgdorferi*: different hosts are different niches. *Genetics.* 2004;168:713–22. DOI: 10.1534/genetics.104.028738
- Caporale DA, Johnson CM, Millard BJ. Presence of *Borrelia burgdorferi* (Spirochaetales: Spirochaetaceae) in southern Kettle Moraine State Forest, Wisconsin, and characterization of strain W97F51. *J Med Entomol.* 2005;42:457–72. DOI: 10.1603/0022-2585(2005)042[0457:POBSS]2.0.CO;2
- Diuk-Wasser MA, Gatewood AG, Cortinas MR, Yaremych-Hamer S, Tsao J, Kitron U, et al. Spatiotemporal patterns of host-seeking *Ixodes scapularis* nymphs (Acari: Ixodidae) in the United States. *J Med Entomol.* 2006;43:166–76. DOI: 10.1603/0022-2585(2006)043[0166:SPOHIS]2.0.CO;2
- Qiu WG, Dykhuizen DE, Acosta MS, Luft BJ. Geographic uniformity of the Lyme disease spirochete (*Borrelia burgdorferi*) and its shared history with tick vector (*Ixodes scapularis*) in the northeastern United States. *Genetics.* 2002;160:833–49.
- Humphrey PT, Caporale DA, Brisson D. Uncoordinated biogeography of the Lyme disease pathogen, *Borrelia burgdorferi*, and its tick vector, *Ixodes scapularis*. *Evolution.* 2010. In press.
- Qiu WG, Schutzer SE, Bruno JF, Attie O, Xu Y, Dunn JJ, et al. Genetic exchange and plasmid transfers in *Borrelia burgdorferi* sensu stricto revealed by three-way genome comparisons and multilocus sequence typing. *Proc Natl Acad Sci U S A.* 2004;101:14150–5. DOI: 10.1073/pnas.0402745101
- Bunikis J, Tsao J, Berglund J, Fish D, Barbour AG. Sequence typing reveals extensive strain diversity of the Lyme borreliosis agents *Borrelia burgdorferi* in North America and *Borrelia afzelii* in Europe. *Microbiology.* 2004;150:1741–55. DOI: 10.1099/mic.0.26944-0
- Dykhuizen DE, Polin DS, Dunn JJ, Wilske B, Preac-Mursic V, Dattwyler RJ, et al. *Borrelia burgdorferi* is clonal: implications for taxonomy and vaccine development. *Proc Natl Acad Sci U S A.* 1993;90:10163–7. DOI: 10.1073/pnas.90.21.10163
- Attie O, Bruno JF, Xu Y, Qiu D, Luft BJ, Qiu WG. Co-evolution of the outer surface protein C gene (*ospC*) and intraspecific lineages of *Borrelia burgdorferi* sensu stricto in the northeastern United States. *Infect Genet Evol.* 2007;7:1–12. DOI: 10.1016/j.meegid.2006.02.008
- Dykhuizen DE, Baranton G. The implications of a low rate of horizontal transfer in *Borrelia*. *Trends Microbiol.* 2001;9:344–50. DOI: 10.1016/S0966-842X(01)02066-2
- Guttman DS, Wang PW, Wang IN, Bosler EM, Luft BJ, Dykhuizen DE. Multiple infections of *Ixodes scapularis* ticks by *Borrelia burgdorferi* as revealed by single-strand conformation polymorphism analysis. *J Clin Microbiol.* 1996;34:652–6.
- Casjens S, Palmer N, van Vugt R, Huang WM, Stevenson B, Rosa P, et al. A bacterial genome in flux: the twelve linear and nine circular extrachromosomal DNAs in an infectious isolate of the Lyme disease spirochete *Borrelia burgdorferi*. *Mol Microbiol.* 2000;35:490–516. DOI: 10.1046/j.1365-2958.2000.01698.x
- Stevenson B, Miller JC. Intra- and interbacterial genetic exchange of Lyme disease spirochete *erp* genes generates sequence identity amidst diversity. *J Mol Evol.* 2003;57:309–24. DOI: 10.1007/s00239-003-2482-x
- Zhang J-R, Norris SJ. Genetic variation of the *Borrelia burgdorferi* gene *vlsE* involves cassette-specific, segmental gene conversion. *Infect Immun.* 1998;66:3698–704.

18. Shimodaira H, Hasegawa M. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol Biol Evol.* 1999;16:1114–6.
19. Wormser GP, Liveris D, Nowakowski J, Nadelman RB, Cavaliere LF, McKenna D, et al. Association of specific subtypes of *Borrelia burgdorferi* with hematogenous dissemination in early Lyme disease. *J Infect Dis.* 1999;180:720–5. DOI: 10.1086/314922
20. Caporale DA, Kocher TD. Sequence variation in the outer-surface-protein genes of *Borrelia burgdorferi*. *Mol Biol Evol.* 1994;11:51–64.
21. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 1994;22:4673–80. DOI: 10.1093/nar/22.22.4673
22. Rozas J, Rozas R. DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics.* 1999;15:174–5. DOI: 10.1093/bioinformatics/15.2.174
23. Jolley KA, Feil EJ, Chan MS, Maiden MC. Sequence type analysis and recombinational tests (START). *Bioinformatics.* 2001;17:1230–1. DOI: 10.1093/bioinformatics/17.12.1230
24. Sawyer S. Statistical tests for detecting gene conversion. *Mol Biol Evol.* 1989;6:526–38.
25. Smith JM. Analyzing the mosaic structure of genes. *J Mol Evol.* 1992;34:126–9.
26. Smith JM, Smith NH, O'Rourke M, Spratt BG. How clonal are bacteria? *Proc Natl Acad Sci U S A.* 1993;90:4384–8. DOI: 10.1073/pnas.90.10.4384
27. Seinst G, Dykhuizen DE, Dattwyler RJ, Golde WT, Dunn JJ, Wang IN, et al. Four clones of *Borrelia burgdorferi* sensu stricto cause invasive infection in humans. *Infect Immun.* 1999;67:3518–24.
28. Dykhuizen DE, Brisson D, Sandigursky S, Wormser GP, Nowakowski J, Nadelman RB, et al. The propensity of different *Borrelia burgdorferi* sensu stricto genotypes to cause disseminated infections in humans. *Am J Trop Med Hyg.* 2008;78:806–10.
29. Posada D, Crandall KA. MODELTEST: testing the model of DNA substitution. *Bioinformatics.* 1998;14:817–8. DOI: 10.1093/bioinformatics/14.9.817
30. Wormser GP, Brisson D, Liveris D, Hanincová K, Sandigursky S, Nowakowski J, et al. *Borrelia burgdorferi* genotype predicts the capacity for hematogenous dissemination during early Lyme disease. *J Infect Dis.* 2008;198:1358–64. DOI: 10.1086/592279
31. Liveris D, Wormser GP, Nowakowski J, Nadelman R, Bittker S, Cooper D, et al. Molecular typing of *Borrelia burgdorferi* from Lyme disease patients by PCR-restriction fragment length polymorphism analysis. *J Clin Microbiol.* 1996;34:1306–9.
32. Livey I, Gibbs CP, Schuster R, Dorner F. Evidence for lateral transfer and recombination in *OspC* variation in Lyme disease *Borrelia*. *Mol Microbiol.* 1995;18:257–69. DOI: 10.1111/j.1365-2958.1995.mmi_18020257.x
33. Ragan MA. Detection of lateral gene transfer among microbial genomes. *Curr Opin Genet Dev.* 2001;11:620–6. DOI: 10.1016/S0959-437X(00)00244-6
34. Wang G, Ojaimi C, Wu H, Saksenberg V, Iyer R, Liveris D, et al. Disease severity in a murine model of Lyme borreliosis is associated with the genotype of the infecting *Borrelia burgdorferi* sensu stricto strain. *J Infect Dis.* 2002;186:782–91. DOI: 10.1086/343043
35. Cromley EK, Cartter ML, Mrozinski RD, Ertel SH. Residential setting as a risk factor for Lyme disease in a hyperendemic region. *Am J Epidemiol.* 1998;147:472–7.
36. Brisson D, Dykhuizen DE. A modest model explains the distribution and abundance of *Borrelia burgdorferi* strains. *Am J Trop Med Hyg.* 2006;74:615–22.
37. Brisson D, Dykhuizen DE, Ostfeld RS. Conspicuous impacts of inconspicuous hosts on the Lyme disease epidemic. *Proc Biol Sci.* 2008;275:227–35. DOI: 10.1098/rspb.2007.1208

Address for correspondence: Dustin Brisson, Department of Biology, University of Pennsylvania, Leidy Laboratories 209, 433 S University Ave, Philadelphia, PA 19104-6018, USA; email: dbrisson@sas.upenn.edu

All material published in *Emerging Infectious Diseases* is in the public domain and may be used and reprinted without special permission; proper citation, however, is required.

Get the content you want delivered to your inbox.



Table of Contents
Podcasts
Ahead of Print Articles
Medscape CME™
Specialized Content

Online subscription: www.cdc.gov/ncidod/eid/subscrib.htm

Appendix Table. Distance matrix of DNA sequences from the outer surface protein Cgene of *Borrelia burgdorferi**

Sequence	A	B	C	C _{KR10}	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V _{MM25} †	W _{F17}	X _{WI30}	Y _{BC2}	Z _{KR18}	
A																												
B	79.8																											
C	83	84.3																										
C _{KR10}	81.9	84.7	96.8†																									
D	84.5	79.9	79.1	78.6																								
E	78.5	77.8	79.8	78.5	76.3																							
F	78.3	76.1	80.4	78.7	79.8	78.1																						
G	79.8	76.6	76.8	75.7	78.5	80.1	78.5																					
H	78.5	74.4	81.1	80.2	77.6	77	82.8	75.9																				
I	81.7	80.2	87.7	86.5	81.6	79	79.6	74.8	80.2																			
J	81.3	76.8	83.2	82.2	81.9	78.5	84.5	78.5	89.7	82.4																		
K	82.4	78.7	83	81.9	81.5	81.5	83	78.3	81.5	82.6	83.7																	
L	74.2	78.5	77.2	77.4	76.1	75.9	76.6	73.8	80.9	79.8	80.6	76.3																
M	80.4	74.6	79.1	77.6	78.1	81.3	81.1	77.4	76.3	78.7	78.7	81.5	75.1															
N	76.1	79.8	79.1	78.7	77	76.8	76.6	82	75.3	77	77.2	79.6	74.6	76.4														
O	75.9	83.3	80.4	80	78.7	81.3	80.9	78.1	75.3	77.4	77	77.2	78.5	80.4	78.5													
P	77.8	76.3	76.3	75.9	75.7	79.4	79.6	83.7	75.7	75.5	79.1	76.6	74.2	78.5	78.5	76.1												
Q	81.1	76.6	84.3	83.7	77.8	77.4	80.2	76.8	83	84.5	82.2	80.2	77.2	78.1	77.6	77	76.6											
R	77.4	74.2	77.6	75.9	76.1	78	78.3	82.4	75.1	77.4	77.6	78.5	73.5	79.8	77.6	77.4	83.7	75.3										
S	78.7	82.6	80.4	79.4	79.4	79.1	77.6	81.3	75.6	78.1	79.2	78.3	77	76.6	79.4	81.8	79.8	75.9	82									
T	78.9	78.9	77.6	76.8	78.3	76.4	76.1	80.4	72.7	77.2	75.1	78.5	74.9	77	78.3	77.5	80.2	74.8	76.1	80								
U	79.8	76.8	78.3	77	74.6	74.2	77.1	81.7	74.8	75.3	76.8	74.8	76.3	75.9	76.7	78.6	81.1	75.2	80.9	84.1	83.2							
V _{MM25} †	83.1	78.3	85.5	84.1	83.2	85.1	86.5	82.1	87.9	86	89.4	90.8	81.6	87	82.2	82.1	80.7	85.5	83.6	80.2	78.8	78.7						
W _{F17}	79.4	79.4	82.4	82.2	78.2	79.6	78.9	77.6	75.5	81.7	79.6	79.6	80	84.1	77	81.1	77.2	78.3	77.6	79.8	76.6	78.3	84.1					
X _{WI30}	82.6	76.8	82.6	80.9	80.9	76.8	79.6	79.1	78.9	80.2	82.2	81.3	75.7	78.5	77.8	79.4	77.2	79.1	80.4	79.8	76.6	78.7	86.5	78.1				
Y _{BC2}	75.9	77.4	78.3	78.5	75.7	79.1	78.9	75.9	79.4	76.1	79.6	78.1	81.9	78.7	75.5	80	77.8	79.1	75.7	79.3	76.8	78.7	86	78.9	75.5			
Z _{KR18}	77.4	77.4	80	79.4	78.9	77.6	79.3	77.6	76.5	77	78.9	82.6	74.4	78.7	78.8	77	77.2	75.9	78	77.6	75.7	75.3	86	76.6	76.3	75.5		

*Used only first 300 bp, compared with NC_001903. GenBank accession nos HM047873, HM047874, HM047875, HM047876, HM047877, and HM047878.

†Strain designations..

‡KR10 is >2% and <8% different from C; <2% means the same allelic group, >8% means a different group.

Evolution of Northeastern and Midwestern *Borrelia burgdorferi*, United States

Technical Appendix

Statistical Support for Horizontal Gene Transfer

Several instances of recombination are evident between intergenic spacer (IGS) and the outer surface protein C (*ospC*) gene. Differences in evolutionary history are observable by using the midwestern IGS data set alone ($p = 0.006$) and the combined midwestern/northeastern IGS data set ($p < 0.001$). The different p values in the 2 analyses could result simply from different sample sizes used in the Shimodaira-Hasegawa test (63 strains in the midwestern/northeastern data set; 39 strains in the midwestern data set). We constructed 10 data sets by randomly sampling 39 strains from the midwestern/northeastern data set. The constrained and unconstrained phylogenies were reconstructed and compared for each randomly sampled data set. The topologies of the unconstrained trees from all the random data sets were significantly more likely than constrained trees, supporting evidence of lateral gene transfer. Moreover, the p values were < 0.006 in 9 of the 10 cases, suggesting that including northeastern linkage relationships increases differences between constrained and unconstrained trees ($p \leq 0.001$, < 0.001 , < 0.001 , 0.001 , 0.002 , 0.002 , 0.004 , 0.004 , 0.006 , 0.008), despite the perfect linkage disequilibrium between IGS and *ospC* in the northeastern United States.

Technical Appendix Table 1. Midwestern and northeastern IGS haplotypes

Strain	Alignment	Midwest frequency	Bunikis type*
	1111122222222222233333444444444455555555555666666666666777 1346777777778899000011224557888833346113368889012347788801225567889111 6134134567893658267998698050235626959594568235065551602903463622245136		
Hap_1	ACAAA-----AGTTAGGTTCTATAGGACAGAGTCTCACGGACTGGTACTTCGGGTCCAATCACTCCC	2	1A
Hap_2T.T....	1	
Hap_3G.....	0	1B
Hap_4G.....G.....T.TC...	0	3C
Hap_5G.....G.....T.T....	0	3A
Hap_6G.....G.....T.T....	0	3B
Hap_7G.....G.....C.....T.T....	0	3D
Hap_8	..TGGATTTTTTTAGCGAA.GT..C..A.GGAGC...C...A.G..ATCG..C.T....C.CTGT...	3	6A
Hap_9	..TGGATTTTTTTAGCGAACGT..CG.A.GG.GC...C...A.GA..ATCG.CC.T....C.CTGT...	1	
Hap_10	G.TGGATTTTTTTAGCGAA.GTCG....GG.GC...CT..A.G.CA.CG..C.T....C.CTGT...	1	
Hap_11	G.TGGATTTTTTTAGCGAA.GTC....GG.GC...CT..A.G.CA.CG.T.C.T....C.CTGT...	1	
Hap_12	GTTGGATTTTTTTAGCGAA.GTC....GG.GC..TCT..A.G.CA.CG..C.T...T.C.CTGT...	1	8C
Hap_13	G.TGGATTTTTTTAGCGAA.GTC....GG.GC...CT..A.G.CA.CG..C.T....C.CTGT...	2	5
Hap_14	G.TGGATTTTTTTAGCGAA.GTC.CG....GG.GC...CT..A.G.CA.CG..C.T....C.CTGT...	3	
Hap_15	..TGGATTTTTTTAGCGAA.GT..C.A..GGAGC...C...A.G..A.CG..CTT....C.CTGT...	1	
Hap_16	..TGGATTTTTTTAGCGAA.GT..C.A..GG.GCC..C..T..G..A.CG..CTT..C..C.CTGT...	2	
Hap_17	..TGGATTTTTTTAGCGAA.GT..C.A..GG.GC...C...G..A.CG..CTT..C..C.CTGT...	3	
Hap_18	G.TGG-----TAGCGAA.GT..C.A..GG.GC...C...TG..A.CG..CTT..C..CCCTGT.T...	1	
Hap_19	G.TGG-----TAGCGAA.GT..C.A..GG.GC...C...TG..A.CG..CTT..C..C.CTGT.T...	9	
Hap_20	G.TGG-----TAGCGAA.GT..C.A..GG.GC.T.CT...TG..A.CG..CTT..C..C.CTGT.TT...	2	2A
Hap_21	G.TGG-----TAGCGAA.GT..C.A..GG.GC...CT...TG..A.CG..CTT..C..C.CTGT.TT...	6	2D
Hap_22	G.TGG-----TAGCGAA.GT..C.A..GG.GC.T.CT...TG..A.CG..C.T..C...CTGT.TT...	0	2B
Hap_23	G.TGG-----TAGCGAA.GT..C.A..GG.GC.T.CT...TG..A.CG..CTT..C...TGT.TT...	0	2C
Hap_24	..TGGATTTTTTTAGCGAA.GT..C.A..GG.GC...C...G..A.CG..CTT..C..C.CTGT...T	0	4A
Hap_25	..TGGATTTTTTTAGCGAA.GT..C.A..GG.GC...C.C.A.G..A.CG..CTT..C..C.CTGT...T	0	4B
Hap_26	GTTGGATTTTTTTAGCGAA.GTC....GG.GC..TC...A.G.CA.CG..C.T...T.C.CTGT...	0	8B
Hap_27	GTTGGATTTTTTTAGCGAA.GTC....GG.GC..TCT..A.G.CA.CG..C.T..T.T.C.CTGT...	0	8A
Hap_28	GTTGGATTTTTTTAGCGAA.GTC....GG.GC...CT..A.G.CA.CG..C.T...T.C.CTGT...	0	8D
Hap_30	..TGGATTTTTTTAGCGAA.GT..C...GG.GC...C...A.G..A.CG..C.TT...TC.CTGT...	0	7B
Hap_31	..TGGATTTTTTTAGCGAA.GT..C...GG.GC...C...A.G..A.CG..C.TT...TC.CTGT...	0	7A
Hap_32	..TGGATTTTTTTAGCGAA.GT..C..A.GG.GC...C...A.G..ATCG..C.TT...TC.CTGT...	0	6B
Hap_33	..TGGATTTTTTTAGCGAA.GT..C..A.GGAGC...CT..A.G..ATCG..C.T....C.CTGT...	0	6C
Hap_34	..TGGATTTTTTTAGCGAA.GT..CG.A.GG.GC...C...A.G.CATCG.CC.T....C.CTGT...	0	9

*As defined in Bunikis et al. (1).

Technical Appendix Table 2. *ospAB* haplotypes*

Strain	<i>ospA</i> <i>ospB</i> 2344771222333455777 258805551467666917134 989762891232069407392	A	B	C	D
Hap_1	GACGTCAACTTTGGGCGTTGA	2	1	1	1†
Hap_2	...A.....G.TC	1	2	2	2
Hap_3	...A....C.CAA...G..C	1	2	2	3
Hap_4	...AC....C.CAA...G..C	1	3	3	3
Hap_5	...AC....C..A...G..C	1	3	3	7
Hap_6	...ACTC...C.AA...GC.C	1	4		5
Hap_7	.C.AC....C.CAA...G..C	7	5		3
Hap_8	.C.AC....C..A...G..C	1	5		7
Hap_9	...A...G..C.AA...GC.C	1	6		5†
Hap_10	...A...G.C..A...G..C	1	6		7
Hap_11	...A...G.CC.AA.T.G...	1	6		11
Hap_12	.C.ACTC...C..A...G..C	2	7		7
Hap_13	.C.A.TC..C.CAA...G..C	1	8	8	3†
Hap_14	.C.A.TC..C..A...GC.C	3	8	8	8†
Hap_15	.C.A.T...C...AAA.AG..C	2	9	9	6†
Hap_16	.CTA...G...A...G..C	1	10		4
Hap_17	.C.AC.C.AC..AA...G..C	1	11		10
Hap_18	...A.TC..C..A...G...	2	12		9
Hap_19	CC.A.TC..C..A...G...	1	13		9†

*A, frequency in midwestern patients; B, *ospA* haplotypes; C, *ospA* haplotypes as defined in Bunikis et al. (1); D, *ospB* haplotypes.

†In the 13 isolates from which *ospA* was not sequenced, 1 *ospB* haplotype 1 and 3 were found, 2 *ospB* haplotype 6 were found, and 3 *ospB* haplotypes 5, 8, and 9 were found.

References

1. Bunikis J, Tsao J, Berglund J, Fish D, Barbour AG. Sequence typing reveals extensive strain diversity of the Lyme borreliosis agents *Borrelia burgdorferi* in North America and *Borrelia afzelii* in Europe. *Microbiology*. 2004;150:1741–55. [Medline DOI: 10.1099/mic.0.26944-0](https://doi.org/10.1099/mic.0.26944-0)