View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Dartmouth Digital Commons (Dartmouth College)

Dartmouth College

Dartmouth Digital Commons

Dartmouth Scholarship

Faculty Work

6-22-2012

Two Boundaries Separate Borrelia burgdorferi Populations in North America

Gabriele Margos
University of Bath

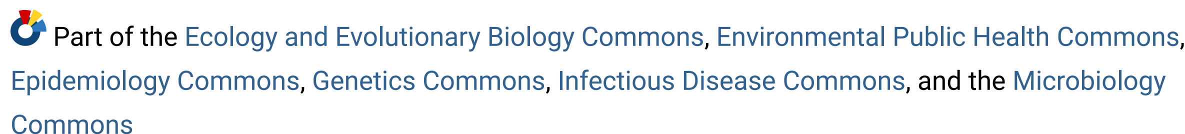
Jean I. Tsao
Michigan State University

Santiago Castillo-Ramirez
University of Bath

Yvette A. Girard
University of California, Davis

Anne G. Hoen
Dartmouth College

Follow this and additional works at: https://digitalcommons.dartmouth.edu/facoa



Part of the Ecology and Evolutionary Biology Commons, Environmental Public Health Commons, Epidemiology Commons, Genetics Commons, Infectious Disease Commons, and the Microbiology Commons

Dartmouth Digital Commons Citation

Margos, Gabriele; Tsao, Jean I.; Castillo-Ramirez, Santiago; Girard, Yvette A.; and Hoen, Anne G., "Two Boundaries Separate Borrelia burgdorferi Populations in North America" (2012). *Dartmouth Scholarship*. 478.
https://digitalcommons.dartmouth.edu/facoa/478

This Article is brought to you for free and open access by the Faculty Work at Dartmouth Digital Commons. It has been accepted for inclusion in Dartmouth Scholarship by an authorized administrator of Dartmouth Digital Commons. For more information, please contact dartmouthdigitalcommons@groups.dartmouth.edu.

AEM Journals.ASM.org

Two Boundaries Separate *Borrelia burgdorferi* Populations in North America

Gabriele Margos,[a]* Jean I. Tsao,[b,f] Santiago Castillo-Ramírez,[a] Yvette A. Girard,[c] Sarah A. Hamer,[b] Anne Gatewood Hoen,[d] Robert S. Lane,[e] Steve L. Raper,[a] and Nicholas H. Ogden[g]

Department of Biology and Biochemistry, University of Bath, Claverton Down, Bath, United Kingdom[a]; Department of Fisheries and Wildlife, Michigan State University, East Lansing, Michigan, USA[b]; Wildlife Health Center, University of California, Davis, California, USA[c]; Department of Community and Family Medicine, Dartmouth Medical School, Hanover, New Hampshire, USA[d]; Department of Environmental Science, Policy and Management, University of California, Berkeley, California, USA[e]; Department of Large Animal Clinical Sciences, Michigan State University, East Lansing, Michigan, USA[f]; and Zoonoses Division, Centre for Food-borne, Environmental and Zoonotic Infectious Diseases, Public Health Agency of Canada, Ottawa, Canada[g]

Understanding the spread of infectious diseases is crucial for implementing effective control measures. For this, it is important to obtain information on the contemporary population structure of a disease agent and to infer the evolutionary processes that may have shaped it. Here, we investigate on a continental scale the population structure of *Borrelia burgdorferi*, the causative agent of Lyme borreliosis (LB), a tick-borne disease, in North America. We test the hypothesis that the observed population structure is congruent with recent population expansions and that these were preceded by bottlenecks mostly likely caused by the near extirpation in the 1900s of hosts required for sustaining tick populations. Multilocus sequence typing and complementary population analytical tools were used to evaluate *B. burgdorferi* samples collected in the Northeastern, Upper Midwestern, and Far-Western United States and Canada. The spatial distribution of sequence types (STs) and inferred population boundaries suggest that the current populations are geographically separated. One major population boundary separated western *B. burgdorferi* populations transmitted by *Ixodes pacificus* in California from Eastern populations transmitted by *I. scapularis*; the other divided Midwestern and Northeastern populations. However, populations from all three regions were genetically closely related. Together, our findings suggest that although the contemporary populations of North American *B. burgdorferi* now comprise three geographically separated subpopulations with no or limited gene flow among them, they arose from a common ancestral population. A comparative analysis of the *B. burgdorferi* outer surface protein C (*ospC*) gene revealed novel linkages and provides additional insights into the genetic characteristics of strains.

Populations of infectious agents are shaped by evolutionary and demographic processes, as well as by the population dynamics of their hosts and (in the case of vector-borne pathogens) arthropod vectors. These processes can leave signatures in the pathogens' genomes. Their frequency and distribution in space and time can be exploited via strain typing for important utilitarian purposes, such as associating specific genotypes with specific ecological niches and determining the geographic distribution of genotypes or of phenotypic characteristics, such as pathogenicity. By investigating the evolutionary history of pathogens, we may be able to elucidate how environmental drivers have in the past shaped patterns of their spread or dispersal and infer how they may do so in the future. Here, by studying *Borrelia burgdorferi*, the causative agent of Lyme borreliosis (LB) in North America, we investigate the population structure of a vector-borne agent on a continental scale and test the hypothesis that it reflects recent environmentally driven demographic events (population bottlenecks) experienced by vector populations and their reproductive hosts.

LB is a tick-borne bacterial infection which, if not diagnosed and treated early, may develop into a debilitating multisystemic disorder (48). The LB group of spirochetes currently comprises 18 named species, but *B. burgdorferi sensu stricto* (hereinafter called *B. burgdorferi*) is the species regularly associated with human disease in North America (34). In North America, *B. burgdorferi* is transmitted primarily by the blacklegged tick, *Ixodes scapularis*, east of the Rocky Mountains and by the western blacklegged tick, *Ixodes pacificus*, in the Far-Western United States, and a large variety of vertebrates serve as reservoir hosts (e.g., see reference 50). In the United States, LB risk currently occurs primarily in three regions: the Northeast, where >80% of reported cases in the United States occur, the Upper Midwest, where 11% of cases occur, and the Pacific coastal region (California and Washington) (2). In Canada, LB is also an emerging infectious disease that became a national reportable disease in 2010. Risk of exposure to vector ticks has been identified in the Maritimes and southeastern Quebec, southern central Canada (Ontario and Manitoba), and southern British Columbia (38–41).

Geographic variation in LB incidence arises in part from environmental factors (climate, natural habitats, and associated anthropogenic changes) that affect the occurrence, abundance, and activity of tick vectors and hosts, the efficiency of transmission cycles, regional variation in the frequencies of more- or less-pathogenic genotypes, and/or the behavior of humans (40, 50). Thus, the spatial distribution of human cases is driven by a com-

Received 22 February 2012 Accepted 5 June 2012
Published ahead of print 22 June 2012
Address correspondence to Gabriele Margos, gmargos1@gmail.com.
* Present address: Gabriele Margos, Bavarian Public Health and Food Safety Authority, Branch Oberschleißheim, Veterinärstrasse, Oberschleißheim, Germany.
Supplemental material for this article may be found at http://aem.asm.org/.
Copyright © 2012, American Society for Microbiology. All Rights Reserved.
doi:10.1128/AEM.00231-12

bination of factors that vary regionally (17, 21, 22, 31) and that may have been influenced by recent historical events impacting the distribution of ticks and their reproductive hosts (reviewed in references 3 and 46).

The historical distribution of *B. burgdorferi* in North America can perhaps be inferred from early records of deer and vector distributions (3, 18, 36). It has been suggested that white-tailed deer had been driven almost to extinction by 1900 through both unmanaged exploitation and habitat loss, which would also have affected the availability of other hosts for ticks (7, 12, 24, 36). Deer are key hosts for adult *I. scapularis*, and their decline would be expected to have driven *I. scapularis* populations to a low level, which in turn probably produced severe bottlenecks in *B. burgdorferi* populations (reviewed in references 8, 28, 43, and 46). Reforestation caused reexpansion of deer populations (as well as that of other woodland hosts), followed by expansions of both vector-tick and *B. burgdorferi* populations (3, 46). Several recent studies reporting on contemporary *I. scapularis* and *B. burgdorferi* populations found that their current distribution in the Northeast and Midwest is discontinuous (4, 13, 14, 26). We hypothesize that these recent and past demographic events have left genetic signatures in the genome of *B. burgdorferi*, which we sought to investigate in the present study.

Multilocus sequence typing (MLST) allows the characterization of diverse bacterial populations (1, 32, 47) by analyzing the genetic variation of multiple loci encoding proteins essential for cell maintenance, or "housekeeping." Previous studies using MLST suggested that North American *B. burgdorferi* populations were subdivided between the Northeast and Upper Midwest (26, 41). Other authors, using different genetic markers, such as the plasmid-encoded outer surface protein C (*ospC*) or the 16S-23S rRNA intergenic spacer region (IGS), suggested that the population genetic structure of *B. burgdorferi* isolates from the Northeast and Upper Midwest overlaps that from California (4, 5, 22). It has, however, been proposed that the genetic variation in *ospC* does not reflect the organisms' evolutionary history but, as a result of recombination and/or horizontal plasmid transfer, it instead reflects the evolution of the locus (4).

Here, we seek to elucidate the evolutionary history of *B. burgdorferi* in North America and its impact on the contemporary population structure using MLST (26, 33, 38, 41). Expecting that multiple chromosomal genetic markers will better reflect the organism's history, we determine whether the genetic structure of *B. burgdorferi* in North America fits a pattern consistent with long-term and/or more recent environmental changes. For this, we analyzed strains from the Northeast (New England and southeastern Canada), the Upper Midwest (including south-central Canada), and the Far West (California) in North America to investigate potential population subdivisions in an explicitly spatial context. Parallel to the MLST analyses, we typed the *ospC* locus to understand how variation at this locus may relate to the population divisions as determined by MLST.

MATERIALS AND METHODS

Collection of ticks and screening for *Borrelia*. Overall, of the 295 samples analyzed in this study, 162 from Canada were described previously by Ogden and coauthors (41), 78 from the Northeast and Upper Midwest by Hoen and coauthors (26), and 7 from California by Margos and coauthors (33). Two GenBank submissions, i.e., strains WI91-23 (genome project ID 28627) and CA11.2A (genome project ID 28629) (45), were included in MLST only, not the spatial analysis. An additional 25 samples from the Upper Midwest and 21 from California were analyzed by MLST for the purpose of this study. Although the number of samples from the Western coastal region is relatively small, the sample set includes representatives of the most common strains found in this region. A summary table of samples (Table 1) is given above, and a complete list of the samples evaluated and their geographic origins is given in Table S1 in the supplemental material.

TABLE 1 *B. burgdorferi* samples included in this study, number of STs, and their geographic origin

Region	No. of samples from: Canada (mainly adult ticks)	USA (nymphs and adult ticks)	Total no. of samples	Total no. of STs	Samples/ST
Northeast	152[a]	41	193	37	5.2
Upper Midwest	10[b]	62	72	30	2.4
California		28	28	18	1.5
Total			293[c]	85	

[a] New Brunswick, Nova Scotia, Prince Edward Islands, Newfoundland, Quebec, and Eastern Ontario.
[b] Western Ontario, Manitoba, and Alberta.
[c] Two additional samples from GenBank were included for MLST analysis only (not geographically referenced).

I. scapularis ticks (mainly adults) from Canada were collected from 2005 to 2007 by means of a passive surveillance program from companion animals and humans at veterinary clinics or medical clinics as described previously (41). The collections comprised ticks from resident populations and others that probably had dispersed on migratory birds from locations in the United States where the *I. scapularis* and *B. burgdorferi* populations are established (10, 42). Questing adult *I. scapularis* ticks were sampled in Upper Midwestern sites between 2004 and 2008 by cloth dragging (25). Tick collections from California consisted of questing *I. pacificus* nymphs that were collected in 2004 from 78 dense woodlands in Mendocino County (16, 22). We are aware that adult ticks may not pose a risk for spirochete transmission, but the pathogens they harbor in many of our samples are still representative of the local *Borrelia* populations and constitute an important component of the overall population genetic pool of *B. burgdorferi* in North America. Although some ticks collected in Canada and their associated *Borrelia* organisms may not be from locally established transmission cycles, they potentially represent the propagule pool of *Borrelia* that may eventually become established locally. These samples were therefore included in analyses that explored bacterial relationships. However, *Borrelia* samples collected in Canada from regions that are not known to have resident tick populations were excluded from the spatial population analysis ($n = 191$; see Table S1 in the supplemental material). Samples were stored either in 70% or 95% ethanol until processed for DNA purification using Qiagen DNeasy blood and tissue purification kits (Qiagen) (16, 25, 41).

MLST and *ospC* amplification. Nested PCRs for the eight housekeeping genes (*clpA*, *clpX*, *nifS*, *pepX*, *pyrG*, *recG*, *rplB*, and *uvrA*) were carried out using HotStarTaq (Qiagen, Germany) and a touchdown PCR protocol for the primary reaction (35, 38). For the secondary reaction, an annealing temperature of 50°C was chosen. Denaturation and elongation were run at temperatures of 95°C and 72°C for 30 s and 1 min, respectively.

PCR fragments were sequenced in forward and reverse directions (Qiagen) and manually compared using DNASTAR (Lasergene). Sequences that contained two base peaks at the same position in forward and reverse sequences were considered mixed infections and, therefore, were not included in our analyses. Besides the 293 samples we processed, data for two samples were obtained from GenBank (45).

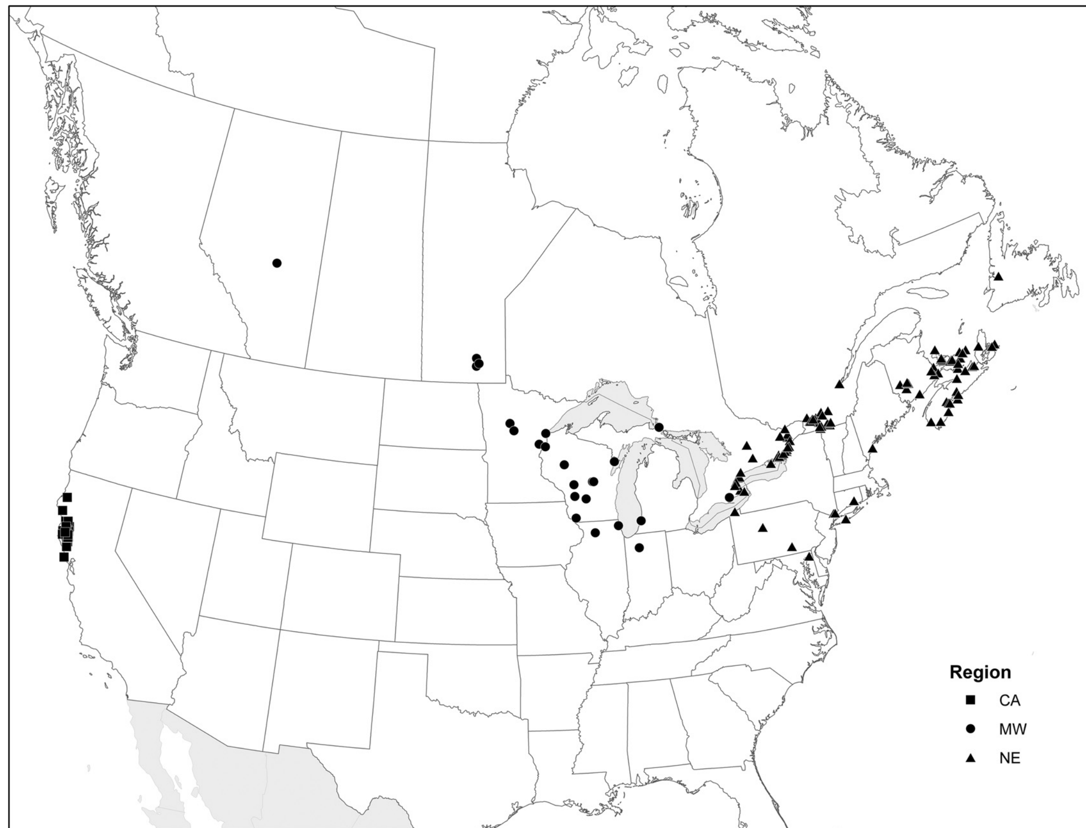


FIG 1 Geographic distribution of *B. burgdorferi* sequence types (STs) determined using ArcGIS. STs that were first described in the Northeast, Upper Midwest, and California are labeled with triangles, circles, and squares, respectively.

Sequences for the individual genes were compared to sequences previously deposited in the borrelia.mlst.net database, and unique sequences were given new, consecutive allele numbers. Allele numbers for all eight loci make up the allelic profile that determines the sequence type (ST) for each strain. Novel allelic profiles were given consecutive ST numbers.

The plasmid-encoded *ospC* gene was amplified in a majority of samples using primers and conditions following Bunikis and coauthors (6).

Population analyses. To investigate the genetic population structure of *B. burgdorferi*, we used six analytical tools available for MLST data. The first three assess contemporary population structure, whereas the last three assess the population structure with respect to the relatedness of strains; both permit inferences of past events. (i) The geographical distribution of STs was visualized using ArcGIS version 10 and the website http://www.spatialepidemiology.net; (ii) a permutation test was performed (Margos et al. [33]) to assess to what extent the spatial distribution of STs corresponds to geographic location; (iii) the geographic occurrence of genetic boundaries was identified by "wombling" (11) to establish whether the genetic data fit the notion of genetically separated populations in North America. The relatedness of strains from the populations in different geographic locations was explored using (iv) eBurst and goeBurst (19, 20), (v) Bayesian analysis of population structure (BAPS) (9), and (vi) phylogenetic analysis (23). These analyses provided information on several hierarchical levels, as follows: (i) the spatial distribution of STs and population boundaries provide information about the contemporary population structure; (ii) the information contained in the allelic profiles corresponds to the degree of relatedness between STs, and the eBURST and goeBurst analyses can be used to infer the pattern of descent; and (iii) the BAPS analysis and gene genealogies use the information available in the sequences and provide information about the historical past of the populations. Detailed information on the software and settings used can be found in the supplemental material.

Nucleotide sequence accession numbers. All housekeeping gene sequences have been deposited to the *B. burgdorferi* MLST database hosted at Imperial College London, United Kingdom, and are available at http://borrelia.mlst.net. *ospC* sequences have been submitted to GenBank under accession numbers JQ951094 to JQ951225.

RESULTS

Contemporary *B. burgdorferi* population structure and subpopulation boundaries. The 295 *B. burgdorferi* samples analyzed were resolved into 85 sequence types (STs) (Table 1; also see Table S1 in the supplemental material). In general, the distribution of STs corresponded to geographic location. None of the STs determined for the 28 samples from California were present in the samples from the other geographic regions (Upper Midwest or Northeast) (Fig. 1; also see Fig. S1 in the supplemental material). Only 5 of the 65 STs found in the Northeast and Upper Midwest occurred in both regions (STs 12, 29, 19, 221, and 222). This differential spatial distribution of ST frequencies in the subpopulations was corroborated by a permutation test using the allelic profiles of *B. burgdorferi* (see Table S1). This test provided significant evidence (dissimilarity value, 5.5; $P = 0.0001$) (also see Fig. S2 in the supplemental material) that the geographic distribution of *B. burgdorferi* STs found in the Northeast, the Upper Midwest, and California differed. The test also provided significant support for spatial differences when Canadian samples from regions without established *I. scapularis* populations ($n = 191$; dissimilarity value, 5.5; $P = 0.001$; data not shown) were omitted.

Wombling of 191 samples from regions with known resident *I. scapularis* populations revealed two population boundaries that

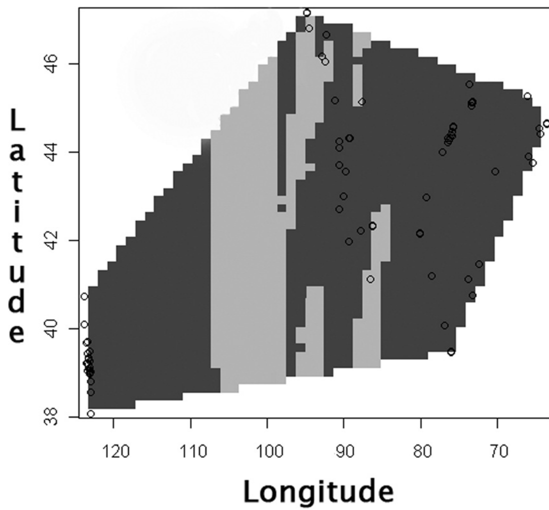


FIG 2 Locations of *B. burgdorferi* population boundaries obtained with wombsoft software. The dark gray pixels indicate the study area, and the circles correspond to samples. The light gray areas represent regions of abrupt genetic change and determine population boundaries. Two main regions for population boundaries were found. One separates *I. scapularis* from *I. pacificus*-dominated habitats, and the second, less-pronounced boundary is located in Indiana and Ohio. Latitudes are °N and longitudes are °W.

divided North American *B. burgdorferi* populations into three subpopulations. A pronounced boundary was located between longitudes 93°W and 110°W, which coincided with the Great Plains and the Rocky Mountains and separated the Californian and Upper Midwestern populations (Fig. 2). The boundary was also identified with more stringent and more relaxed settings, resulting in slightly differing geographic ranges (data not shown). A more accurate identification of the division of *B. burgdorferi* subpopulations between the Upper Midwest and California will require the investigation of more samples from both geographic regions. A second potential boundary was located between longitude 83°W (western Ohio) and 88°W (western Indiana), thereby separating the Northeastern and Upper Midwestern populations. Figure 2 shows the population boundaries determined using wombsoft (h-value of 1.4, and $p_B = 0.4$). The boundary between the Northeastern and Upper Midwestern sites became more pronounced when only samples from these two regions were considered (data not shown).

Ancestral *B. burgdorferi* population structure. Despite the contemporary geographic structure of North American *B. burgdorferi* populations, the eBURST/goeBURST, BAPS, and phylogenetic analyses revealed that these populations of *B. burgdorferi* are genetically closely related (Fig. 3 and 4). Application of the eBURST algorithm divided the strains into 17 clonal complexes.

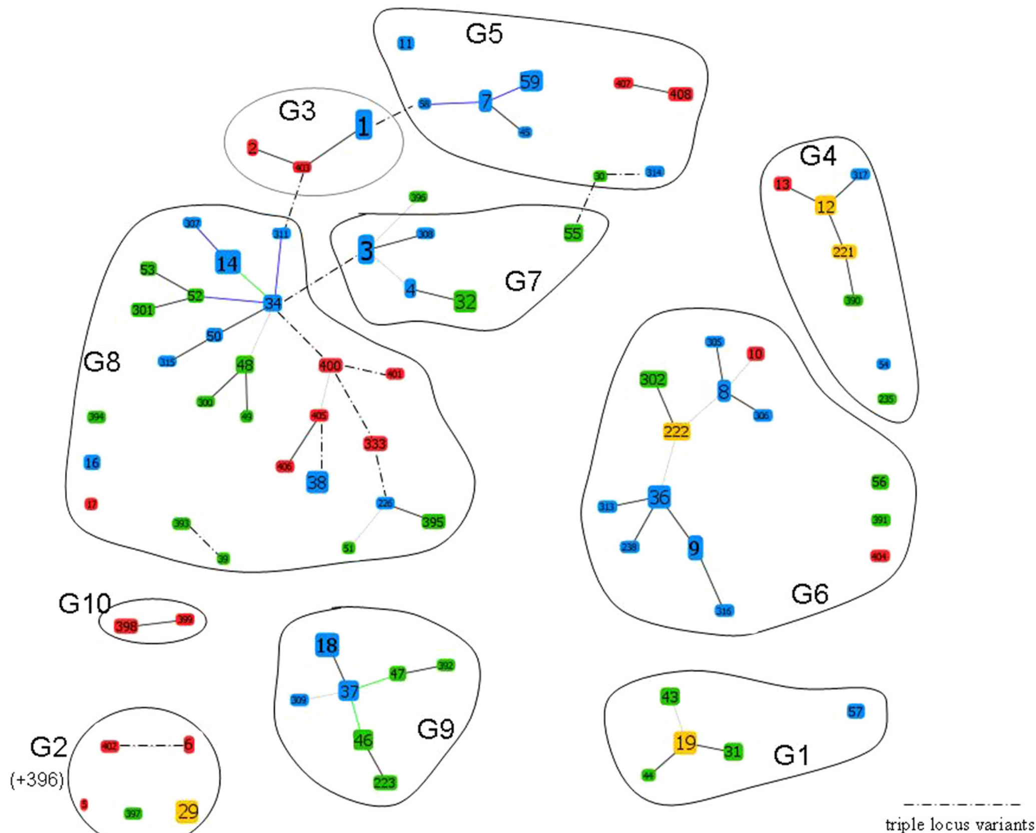


FIG 3 A population snapshot of *B. burgdorferi* isolates from the Northeast, Upper Midwest, and California. STs are color coded according to their geographic location: blue, Northeastern United States; green, Upper Midwest; yellow, Northeast and Upper Midwest; red, California. Colored lines connecting STs indicate descending order of certainty: black lines are inferred without tiebreak rules, blue lines are inferred using tiebreak rule 1 (number of SLV), and green lines are inferred using tiebreak rule 2 (number of DLV). Light-gray lines correspond to DLV. TLV are indicated by dashed lines. BAPS group assignments (G1 to G10) are indicated as circles around clonal complexes and singletons determined in goeBURST.

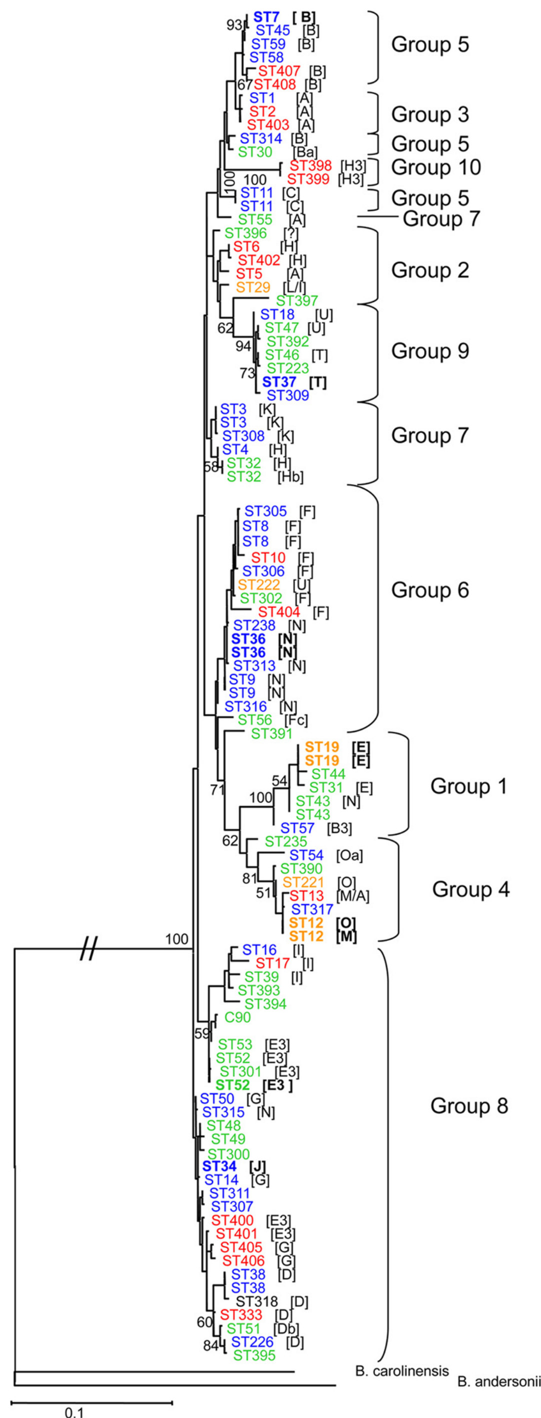


FIG 4 Phylogenetic inference of STs generated using maximum likelihood. Major *ospC* groups are indicated in brackets next to the ST. BAPS groups are also indicated. STs are color coded according to geographic location: blue, Northeast; green, Upper Midwest; yellow, Northeast and Upper Midwest; red, California. *Borrelia carolinensis* and *Borrelia andersonii* sequences were used as outgroups to root the phylogeny. Scale bar = 10% divergence. The branch length of the outgroup is not according to scale, as indicated by slashes.

Founder strains were identified with reasonable bootstrap support (approximately 60%) in four clonal complexes, whereas founders were less certain in the other complexes (Fig. 3 and Table 2). In the goeBURST diagram, the groupings of strains were more "bushy" than in previous analyses and, consistent with previous data, the clonal complexes consisted of strains from the different geographic regions, especially the Northeast and Upper Midwest (Fig. 3) (26). STs from California were mostly found to be triple-locus variants (TLV) of strains from the Northeast and Upper Midwest, except for ST403 and ST2. ST403 was the first single-locus variant (SLV) for ST1, an abundant ST in the Northeast, and it connects ST1 with ST2. These findings further support the genetic relatedness of the subpopulations. Similarly, the Californian ST13 was found to be an SLV of ST12, which occurs in the Northeast and Upper Midwest.

To corroborate the findings made by eBURST, we used BAPS because it infers population structuring. The BAPS analysis divided the strains into 10 groups (Table 3 and Fig. 3), which mirrored the relationship assigned by the eBURST analysis. In Fig. 3, the BAPS groups are projected onto the goeBURST diagram. Small differences were discerned by the two analyses. Specifically, ST55 (a TLV of ST30) was assigned to BAPS group 7, whereas ST30 and ST314 were assigned to group 5. ST396 was assigned to group 2, while in goeBURST, it formed a double-locus variant (DLV) with ST3, a member of BAPS group 7. These differences, however, were also reflected in the phylogeny, wherein samples assigned to groups 5 and 7 were split into different phylogenetic clusters (Fig. 4).

STs that demonstrated significant admixture between groups in the BAPS analysis are given in Table 4. These STs shared genetic information from several groups, suggesting they had undergone recombination. This included STs that manifested disagreement between goeBURST and BAPS: ST30 and ST314 both yielded significant admixture.

The phylogenetic analyses revealed that strains from all three regions were represented in the major lineages (Fig. 4). Although some of the strains from California were closely related to strains from other regions, no single ST from California was found in either of the other regions. In fact, two STs from California, 398 and 399, were distant from neighboring clades. These two STs had unique alleles for most of the genes, except *rplB* (see Table S1 in the supplemental material).

The major *ospC* groups that had been determined for many of the STs are shown adjacent to the tree in Fig. 4. Many of the major *ospC* groups matched the ST clusters (Fig. 4; also see Table S1). Some STs, however, were matched with more than one major *ospC* group, for example, ST12 was found to carry group M *ospC* and group O *ospC*. Other major *ospC* groups, such as group A, were matched locally with different STs, e.g., ST1 in the Northeast, ST2 and ST5 in California, or ST55 in the Upper Midwest. Several major *ospC* groups fell into different phylogenetic clusters (e.g., N or H), presumably as a result of horizontal transfer.

DISCUSSION

In this study, we analyzed North American *B. burgdorferi* populations by MLST in an explicitly spatial context. Samples from the geographic regions most affected by LB, that is, the Northeast and Upper Midwest, and from a region in northern coastal California with a high prevalence of tick infection (29), were included in our analyses. All samples were derived from ixodid ticks (nymphs and adults), collected either by dragging vegetation (Northeast, Upper Midwest, and California) or by means of passive surveillance (southern Canada) between 2004 and 2008.

TABLE 2 The eBURST analysis grouped 295 *B. burgdorferi* samples comprising 85 STs into 17 clonal complexes and a number of singletons[a]

Clonal complex	No. of samples/ no. of STs	ST	FREQ	SLV	DLV	TLV	% bootstrap support for clonal complex founder
1	27/9	**34**	**3**	**4**	**4**	**0**	**65**
		52	2	4	3	1	48
		14	13	3	4	1	12
		50	2	2	4	2	2
		311	1	2	3	3	0
		307	1	2	2	3	6
		53	2	1	4	3	0
		301	2	1	3	3	0
		315	1	1	1	4	0
2	26/6	46	5	3	2	0	35
		37	5	3	2	0	34
		47	2	3	2	0	39
		18	11	1	2	2	0
		223	2	1	2	2	0
		392	1	1	2	2	0
3	11/5	**12**	**5**	**3**	**1**	**0**	**62**
		221	2	2	2	0	18
		13	2	1	2	1	0
		317	1	1	2	1	0
		390	1	1	1	2	0
4	23/5	**36**	**8**	**3**	**1**	**0**	**59**
		9	12	2	2	0	19
		313	1	1	2	1	0
		238	1	1	2	1	0
		316	1	1	1	2	0
5	17/4	**7**	**8**	**3**	**0**	**0**	**63**
		59	7	2	1	0	5
		58	1	2	1	0	9
		45	1	1	2	0	0
6	6/3	**48**	**4**	**2**	**0**	**0**	**31**
		49	1	1	1	0	0
		300	1	1	1	0	0
7	12/3	**19**	**8**	**2**	**0**	**0**	**32**
		31	3	1	1	0	0
		44	1	1	1	0	0
8	41/3	**403**	**1**	**2**	**0**	**0**	**31**
		1	37	1	1	0	0
		2	3	1	1	0	0
9	10/3	**8**	**8**	**2**	**0**	**0**	**33**
		306	1	1	1	0	0
		305	1	1	1	0	0
10	2/2	228	1	1	0	0	
		393	1	1	0	0	
11	3/2	395	2	1	0	0	
		226	1	1	0	0	
12	6/2	222	3	1	0	0	
		302	3	1	0	0	
13	3/2	408	2	1	0	0	
		407	1	1	0	0	
14	2/2	406	1	1	0	0	
		405	1	1	0	0	
15	33/2	3	32	1	0	0	
		308	1	1	0	0	
16	13/2	32	8	1	0	0	
		4	5	1	0	0	
17	3/2	398	2	1	0	0	
		399	1	1	0	0	

[a] For each clonal complex, the number of samples and the number of STs constituting the clonal complex is indicated. The predicted founder of the clonal complexes is given in boldface, but for some clonal complexes, no single founder was predicted. The frequency (FREQ) of each ST and the numbers of single (SLV), double (DLV), and triple (TLV) locus variants for each ST are given. Statistical support of founder assignment was obtained by a bootstrap procedure (resampling for bootstrapping = 1,000). Singletons (n = 28): 57, 56, 55, 54, 51, 314, 43, 404, 309, 402, 401, 400, 39, 38, 30, 29, 6, 5, 17, 16, 11, 10, 333, 397, 396, 394, 235, and 391.

TABLE 3 Relationship of BAPS groups and STs[a]

BAPS group	STs
1	19, 31, 43, 44, 57
2	29, 5, 6, 396, 397, 402
3	1, 2, 403
4	12, 13, 54, 221, 235, 317, 390
5	7, 11, 30, 45, 58, 59, 314, 407, 408
6	8, 9, 10, 36, 56, 222, 238, 302, 305, 306, 313, 316, 391, 404
7	3, 4, 32, 55, 308
8	14, 16, 17, 34, 38, 39, 48, 49, 50, 51, 52, 53, 226, 300, 301, 307, 311, 315, 333, 393, 394, 395, 400, 401, 405, 406
9	18, 37, 46, 47, 223, 309, 392
10	398, 399

[a] The genetic relationship of *B. burgdorferi* samples from North America was analyzed using a Bayesian approach (BAPS) which divided the STs into 10 groups. A comparison of BAPS groups and goeBurst clonal complexes is shown in Fig. 4.

Contemporary population structure: evidence for a recent bottleneck. The geographic distribution of STs suggests that present day populations are separated along geographical lines: the Northeast, Upper Midwest, and California. The permutation test provided evidence that the distribution of STs was significantly different among the three regions, and we hypothesize that this probably reflects the local expansion of regional refuge populations in the Upper Midwest and Northeast during a time much more recent than events reflected in the phylogenetic tree.

Furthermore, the inference of genetic boundaries by wombling analysis between North American *B. burgdorferi* populations provided additional support for the present-day population structure. A western population boundary separated *B. burgdorferi* populations east and west of the Great Plains and the Rocky Mountains, landscape features that are credible natural barriers to *B. burgdorferi* gene flow. The finding of closely related strains, i.e., SLV east and west of the Rocky Mountains, suggests some population connectivity despite these landscape features. Although mutations in housekeeping genes accumulate slowly and the genes evolve over long periods, the uplifting of the Rocky Mountains dates back to about 50 to 100 million years ago (http://www.geo.arizona.edu/geo5xx/geo527/Rockies/uplift.html), which is likely to be further in the past than the mutation(s) found in an SLV. It is possible that the following hypotheses may alone or collectively explain gene flow among these geographic regions: (i) occasional *Borrelia* transport occurred (and may be still occurring) across these barriers (for example, by birds); (ii) the now divided sub-

TABLE 4 STs that showed significant evidence for admixture among the BAPS groups

Sample	ST	DNA sequence represented by BAPS group[a]: 1	2	3	4	5	6	7	8	9	10	P value[b]
ON06-194	226	0	0	0	0	0.06	0	0.31	0.63	0	0	0.035
ON07-441	314	0	0.01	0.18	0	0.62	0	0.18	0	0	0.01	0.03
PEI05-15	316	0	0	0	0	0	0.81	0.19	0	0	0	0.045
51405	30	0	0.02	0.19	0	0.48	0	0.31	0	0	0	0.005
1472505	39	0	0	0.11	0.2	0.01	0	0.05	0.63	0	0	0.035
1468904	49	0	0	0	0	0	0	0.42	0.58	0	0	0.02
501427	54	0	0.13	0	0.67	0	0	0	0	0.2	0	0.005
C462	397	0	0.48	0	0	0	0.15	0	0.1	0.24	0.03	0.005

[a] The numbers in columns 1 to 10 are the Bayesian posterior mean estimates of the proportion of the DNA sequence represented by each BAPS group.
[b] A P value of <0.05 was considered significant, and only significant cases are shown in the table.

populations of *B. burgdorferi* constituted an admixed population when deer (36) and, possibly, ticks and *B. burgdorferi* had a much wider distribution range than at present; and (iii) gene flow occurred in an ancestral admixed population that might have been located further south during Pleistocene glacial periods from 2.5 million years ago to about 18,000 years ago (see below) (27).

The location of the boundary between the Northeastern and Upper Midwestern *B. burgdorferi* subpopulations is roughly consistent with a subdivision described for *B. burgdorferi ospC* at a longitude of 83°W (4). This boundary coincides with a region where a previous study described an absence of *I. scapularis* (13). However, a model for risk of Lyme borreliosis in North America based on the density of host-seeking nymphs predicted higher nymphal densities in northern Indiana and Ohio than were observed during the project (14), which suggests that the division of *B. burgdorferi* populations found between the Northeast and Upper Midwest may be the result of a population contraction as a consequence of vector population decline. The collapses of white-tailed deer, vector tick, and presumably, *Borrelia* populations that occurred from 500 to 110 years ago (46) are consistent with our findings and those of others (6, 26, 44) that local *B. burgdorferi* populations expanded out of refuge populations in the Northeast and Upper Midwest. If so, this scenario may explain the clonality and linkage disequilibrium between plasmid-encoded and chromosomal genes described in the initial population studies on *B. burgdorferi* in the Northeast (6, 44). Our study confirms and extends results from previous studies showing that on a wider geographic scale, the initially reported linkage disequilibrium between *ospC* and chromosomal markers was not as pronounced (5, 49). However, *B. burgdorferi* populations, especially at the edges of their distributional ranges, are likely to show an element of clonality that is related to their parasitic life style, i.e., the bacterium resides only in the tick or its vertebrate host. For recombination or horizontal gene transfer to occur, close physical contact among bacteria is required and depends upon the frequency of infection in vertebrate hosts or tick vectors with multiple strains.

Evidence for an admixed ancestral *B. burgdorferi* population structure. Although the geographic distribution of STs provided information about the contemporary population structure of *B. burgdorferi* in North America, BAPS, eBURST, goeBURST, and phylogeny yielded information about its putative ancestral population structure. We hypothesize that the signals observed using these analytical tools reflect an ancestral population structure that existed before the recent population bottleneck caused by the decline of deer and the consequent effects on tick vector and *Borrelia* populations (3, 13, 14, 18, 36). Our findings suggest that before these events, *B. burgdorferi* strains formed overlapping populations and/or that gene flow between populations was more pronounced than today. For example, the denser "forest" shown in the goeBURST diagram and the additional clonal complexes obtained compared to the ones described before (26) underpin the genetic relatedness of strains from the different regions. In addition, the BAPS analysis largely agreed with the relationships inferred from goeBURST and the phylogenetic tree. Together, these data suggest that to understand the evolutionary history of *B. burgdorferi* as a species, samples from all three regions must be considered.

Bootstrap support for some clonal complex founder assignments was slightly higher than 60%. This was probably due to a lack of SLV, which may have been caused by loss of genetic diversity as a result of the severe bottleneck. However, SLV of STs from the Upper Midwestern and Northeastern sites were found in California (e.g., ST12/ST13 and ST403/ST1), suggesting that additional strains like this may be found during more exhaustive analyses of samples from these regions. An increased number of SLV or DLV will improve the bootstrap support for clonal complex founders and will increase our understanding of the direction of gene flow. Previous studies suggested that gene flow occurred in an east-to-west direction in North America (5, 26). This conclusion is consistent with our current and previous eBURST results that most founders of clonal complexes are found in the Northeast (26), with the fact that *B. burgdorferi* is absent in Asia and therefore would not have populated North America via the Bering Strait, and with the hypothesis that *B. burgdorferi* originated in Europe (33). SLV and DLV were found more frequently in Northeastern and Upper Midwestern sites than in either the Northeast and California or the Upper Midwest and California, intimating a closer relationship between subpopulations of Northeastern and Upper Midwestern spirochetes. Moreover, the analysis using BAPS found significant admixtures between strains from the Northeast and Upper Midwest, providing support for an overlap of these *B. burgdorferi* subpopulations in the past. We speculate that the current population structure reflects the strong bottleneck that *B. burgdorferi* has undergone due to loss of forest habitats (at least in the Northeast and Upper Midwest) and that this bottleneck caused a severe loss of diversity in the Northeast.

The finding of identical STs in the Upper Midwest and Northeast suggests that either these populations once overlapped, reflecting a random distribution of these STs in both regions at the onset of the bottleneck, or alternatively, there is currently a limited gene flow between the two regions. Differentiating between these possibilities requires the use of more sensitive methods, such as the analysis of genome-wide single nucleotide polymorphisms or ecological studies to determine the host associations of these STs. In this respect, it is interesting to note that although migratory birds are considered important to transport ticks and, possibly, tick-borne pathogens over long distances (39), their movements, which respond (directly or indirectly) to seasonal changes in temperature-driven resource availability, are mostly north-south in direction rather than east-west, and this behavior probably evolved or emerged due to increasing availability of northern habitat during interglacials (37). An alternative potential explanation for east-to-west spread of LB spirochetes might be spread via rodent hosts or their ancestors. This would require the concomitant spread of ticks and/or the presence of established vector tick populations. It is conceivable that *Ixodes* ticks may have had a much wider distribution range when its main reproductive host, the deer, was more widespread (36). For some rodent hosts, phylogeographic patterns similar to that seen here for *B. burgdorferi* have been shown, i.e., Northeastern, Midwestern, and Western populations, with the Midwest and Californian clades being linked geographically (and via interbreeding at hybridization zones) by Rocky Mountain-Great Plains clades (15, 51). Such a pattern is consistent with the phylogeographies of Hanta virus and its deer mouse reservoirs (15). We recognize, however, that gene flow might have occurred in ancestral populations located further south in North America.

In agreement with previous studies (4, 49), the phylogenetic analysis revealed that strains that carried the same *ospC* genotype were matched regionally with different MLST sequence types. Al-

though the samples from California represented the most common major *ospC* groups present in questing *I. pacificus* nymphs in this region, *ospC* type H3 was described as the most prevalent in questing *I. pacificus* nymphs in northwestern California (22). Curiously, the three strains carrying *ospC* type H3 (ST398 and ST399) were genetically the most divergent in our analysis, suggesting that ST398, ST399, and perhaps, related strains probably represent a long-term separated population in California. Ecologically, *ospC* type H3 strains have been associated with hardwood-dominated forests where the western gray squirrel (*Sciurus griseus*) is the primary reservoir of *B. burgdorferi* (30). Epidemiologically, these strains have not been associated with human patients, and it had been speculated that their high prevalence might be one reason for the low human LB incidence in northern California (22).

In conclusion, this study has revealed that *B. burgdorferi* in North America comprises three separate but genetically related populations. These are separated by geographic barriers, with California populations being the most isolated, while Northeastern and Upper Midwestern populations overlap and could have been isolated more recently by land use changes. Nevertheless, both the geographic occurrence and the evidence for close relatedness of the three populations also correspond with known phylogeographic patterns among rodent hosts for the vector tick and bacterium, perhaps influencing current *B. burgdorferi* phylogeographic structure. Further studies of *B. burgdorferi* phylogeography, coupled with similar studies on vertebrate hosts and tick vectors, are required to enhance our understanding of how environmental drivers in the past have shaped the population and geographic structure of *B. burgdorferi*. Our study provides optimism that mining signatures in the *B. burgdorferi* genome may assist in prediction of the effects of current and future environmental changes on the risk of vector-borne diseases.

ACKNOWLEDGMENTS

We are grateful to M. Hurn for support with statistical analyses, S. Manel for help interpreting genetic boundaries, D. Aanensen for maintaining the MLST database, and Natalia Fedorova for critical comments on the manuscript. We are indebted to Klaus Kurtenbach (died March 2009) for his inspirational ideas.

REFERENCES

1. **Aanensen DM, Spratt BG.** 2005. The multilocus sequence typing network: mlst.net. Nucleic Acids Res. **33**:W728–W733.
2. **Bacon RM, Kugeler KJ, Mead PS.** 2008. Surveillance for Lyme disease—United States, 1992-2006. MMWR Surveill. Summ. **57**:1–9.
3. **Barbour AG, Fish D.** 1993. The biological and social phenomenon of Lyme disease. Science **260**:1610–1616.
4. **Barbour AG, Travinsky B.** 2010. Evolution and Distribution of the *ospC* gene, a transferable serotype determinant of *Borrelia burgdorferi*. mBio **1**(4):e00153-10. doi:10.1128/mBio.00153-10.
5. **Brisson D, Vandermause MF, Meece JK, Reed KD, Dykhuizen DE.** 2010. Evolution of northeastern and midwestern *Borrelia burgdorferi*, United States. Emerg. Infect. Dis. **16**:911–917.
6. **Bunikis J, et al.** 2004. Sequence typing reveals extensive strain diversity of the Lyme borreliosis agents *Borrelia burgdorferi* in North America and *Borrelia afzelii* in Europe. Microbiology **150**:1741–1755.
7. **Christensen EM.** 1959. A historical view of the ranges of the white-tailed deer in northern Wisconsin forests. Am. Midl. Nat. **61**:230–238.
8. **Collins DL, Nardy RV, Glasgow RD.** 1949. Further notes on the host relationships of ticks on Long Island. J. Econ. Entomol. **42**:159.
9. **Corander J, Marttinen P.** 2006. Bayesian identification of admixture events using multilocus molecular markers. Mol. Ecol. **15**:2833–2843.
10. **Courtney JW, Kostelnik LM, Zeidner NS, Massung RF.** 2004. Multiplex real-time PCR for detection of anaplasma phagocytophilum and Borrelia burgdorferi. J. Clin. Microbiol. **42**:3164–3168.
11. **Crida A, Manel S.** 2007. Wombsoft: an R package that implements the Wombling method to identify genetic boundary. Mol. Ecol. Notes **7**:588–591.
12. **Cronon W.** 1983. Changes in the land: Indians, colonists, and the ecology of New England, 1st ed. Hill and Wang, New York, NY.
13. **Diuk-Wasser MA, et al.** 2006. Spatiotemporal patterns of host-seeking *Ixodes scapularis* nymphs (Acari: Ixodidae) in the United States. J. Med. Entomol. **43**:166–176.
14. **Diuk-Wasser MA, et al.** 2010. Field and climate-based model for predicting the density of host-seeking nymphal *Ixodes scapularis*, an important vector of tick-borne disease agents in the eastern United States. Glob. Ecol. Biogeogr. **19**:504–514.
15. **Dragoo JW, et al.** 2006. Phylogeography of the deer mouse (*Peromyscus maniculatus*) provides a predictive framework for research on hantaviruses. J. Gen. Virol. **87**:1997–2003.
16. **Eisen L, Eisen RJ, Lane RS.** 2006. Geographical distribution patterns and habitat suitability models for presence of host-seeking ixodid ticks in dense woodlands of Mendocino County, California. J. Med. Entomol. **43**:415–427.
17. **Eisen RJ, Lane RS, Fritz CL, Eisen L.** 2006. Spatial patterns of Lyme disease risk in California based on disease incidence data and modeling of vector-tick exposure. Am. J. Trop. Med. Hyg. **75**:669–676.
18. **Falco RC, Daniels TJ, Fish D.** 1995. Increase in abundance of immature *Ixodes scapularis* (Acari: Ixodidae) in an emergent Lyme disease endemic area. J. Med. Entomol. **32**:522–526.
19. **Feil EJ, Li BC, Aanensen DM, Hanage WP, Spratt BG.** 2004. eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. J. Bacteriol. **186**:1518–1530.
20. **Francisco AP, Bugalho M, Ramirez M, Carrico JA.** 2009. Global optimal eBURST analysis of multilocus typing data using a graphic matroid approach. BMC Bioinformatics **10**:152. doi:10.1186/1471-2105-10-152.
21. **Girard YA, Fedorova N, Lane RS.** 2011. Genetic diversity of *Borrelia burgdorferi* and detection of *B. bissettii*-like DNA in serum of north-coastal California residents. J. Clin. Microbiol. **49**:945–954.
22. **Girard YA, et al.** 2009. Population structure of the lyme borreliosis spirochete *Borrelia burgdorferi* in the western black-legged tick (*Ixodes pacificus*) in Northern California. Appl. Environ. Microbiol. **75**:7243–7252.
23. **Guindon S, Lethiec F, Duroux P, Gascuel O.** 2005. PHYML Online—a web server for fast maximum likelihood-based phylogenetic inference. Nucleic Acids Res. **33**:W557–W559.
24. **Halls LK.** 1984. White-tailed deer: ecology and management. Stackpole Books, Harrisburg, PA.
25. **Hamer SA, Tsao JI, Walker ED, Hickling GJ.** 2010. Invasion of the Lyme disease vector *Ixodes scapularis*: implications for *Borrelia burgdorferi* endemicity. Ecohealth **7**:47–63. doi:10.1007/s10393-010-0287-0.
26. **Hoen AG, Margos G, Bent SJ, Kurtenbach K, Fish D.** 2009. Phylogeography of *Borrelia burgdorferi* in the eastern United States reflects multiple independent Lyme disease emergence events. Proc. Natl. Acad. Sci. U. S. A. **106**:15013–15018.
27. **Humphrey PT, Caporale DA, Brisson D.** 2010. Uncoordinated phylogeography of *Borrelia burgdorferi* and its tick vector, *Ixodes scapularis.* Evolution **64**:2653–2663.
28. **Jackson JO, DeFoliart GR.** 1970. *Ixodes scapularis* Say in northern Wisconsin. J. Med. Entomol. **7**:124–125.
29. **Lane RS, et al.** 1992. Risk factors for Lyme disease in a small rural community in northern California. Am. J. Epidemiol. **136**:1358–1368.
30. **Lane RS, Mun J, Eisen JERL.** 2005. Western gray squirrels (Rodentia: Sciuridae): a primary reservoir host of *Borrelia burgdorferi* in Californian oak woodlands? J. Med. Entomol. **42**:388–396.
31. **Lane RS, Piesman J, Burgdorfer W.** 1991. Lyme borreliosis—relation of its causative agent to its vectors and hosts in North America and Europe. Annu. Rev. Entomol. **36**:587–609.
32. **Maiden MC.** 2006. Multilocus sequence typing of bacteria. Annu. Rev. Microbiol. **60**:561–588.
33. **Margos G, et al.** 2008. MLST of housekeeping genes captures geographic population structure and suggests a European origin of *Borrelia burgdorferi.* Proc. Natl. Acad. Sci. U. S. A. **105**:8730–8735.
34. **Margos G, et al.** 2010. Multilocus sequence analysis of *Borrelia bissettii* strains from North America reveals a new *Borrelia* species, *Borrelia kurtenbachii.* Ticks Tick Borne Dis. **1**:151–158.

35. **Margos G, et al.** 2009. A new *Borrelia* species defined by multilocus sequence analysis of housekeeping genes. Appl. Environ. Microbiol. **75**: 5410–5416.
36. **McCabe TR, McCabe RE.** 1997. Recounting whitetails past, p 11–26. *In* McShea WJ, Underwood HB, Rappole JH (ed), The science of overabundance: deer ecology and population management. Smithsonian Institution Press, Washington, DC.
37. **Milá B, Smith TB, Wayne RK.** 2006. Postglacial population expansion drives the evolution of long-distance migration in a songbird. Evolution **60**:2403–2409.
38. **Ogden NH, et al.** 2010. Active and passive surveillance and phylogenetic analysis of *Borrelia burgdorferi* elucidate the process of Lyme disease risk emergence in Canada. Environ. Health Perspect. **118**:909–914.
39. **Ogden NH, et al.** 2008. Role of migratory birds in introduction and range expansion of *Ixodes scapularis* ticks and of *Borrelia burgdorferi* and *Anaplasma phagocytophilum* in Canada. Appl. Environ. Microbiol. **74**:1780–1790.
40. **Ogden NH, et al.** 2006. Climate change and the potential for range expansion of the Lyme disease vector *Ixodes scapularis* in Canada. Int. J. Parasitol. **36**:63–70.
41. **Ogden NH, et al.** 2011. Investigation of genotypes of *Borrelia burgdorferi* in *Ixodes scapularis* ticks collected in surveillance in Canada. Appl. Environ. Microbiol. **77**:3244–3254.
42. **Ogden NH, et al.** 2006. *Ixodes scapularis* ticks collected by passive surveillance in Canada: analysis of geographic distribution and infection with Lyme borreliosis agent *Borrelia burgdorferi*. J. Med. Entomol. **43**:600–609.
43. **Piesman J.** 2002. Ecology of *Borrelia burgdorferi* sensu lato in North America, p 223–249. *In* Gray JS, Kahl O, Lane RS, Stanek G (ed), Lyme borreliosis: biology of the infectious agents and epidemiology of disease. CABI Publishing, Wallingford, United Kingdom.
44. **Qiu WG, Dykhuizen DE, Acosta MS, Luft BJ.** 2002. Geographic uniformity of the Lyme disease spirochete (*Borrelia burgdorferi*) and its shared history with tick vector (*Ixodes scapularis*) in the Northeastern United States. Genetics **160**:833–849.
45. **Schutzer SE, et al.** 2011. Whole-genome sequences of thirteen isolates of *Borrelia burgdorferi*. J. Bacteriol. **193**:1018–1020.
46. **Spielman A.** 1994. The emergence of Lyme disease and human babesiosis in a changing environment. Ann. N. Y. Acad. Sci. **740**:146–156.
47. **Spratt BG.** 1999. Multilocus sequence typing: molecular typing of bacterial pathogens in an era of rapid DNA sequencing and the Internet. Curr. Opin. Microbiol. **2**:312–316.
48. **Steere AC, Coburn J, Glickstein L.** 2004. The emergence of Lyme disease. J. Clin. Invest. **113**:1093–1101.
49. **Travinsky B, Bunikis J, Barbour AG.** 2010. Geographic differences in genetic locus linkages for *Borrelia burgdorferi*. Emerg. Infect. Dis. **16**:1147–1150.
50. **Tsao JI.** 2009. Reviewing molecular adaptations of Lyme borreliosis spirochetes in the context of reproductive fitness in natural transmission cycles. Vet. Res. **40**:36.
51. **Waters JH.** 1963. Biochemical relationships of the mouse *Peromyscus* in New England. Syst. Zool. **12**:122–133.