Title

Evolution and Distribution of the ospC Gene, a Transferable Serotype Determinant of Borrelia burgdorferi

Journal

mBio, 1(4)

ISSN

2150-7511

Authors

Barbour, Alan G
Travinsky, Bridgit

Publication Date

2010-09-28

DOI

10.1128/mBio.00153-10

Evolution and Distribution of the *ospC* Gene, a Transferable Serotype Determinant of *Borrelia burgdorferi*

Alan G. Barbour and Bridgit Travinsky

Departments of Microbiology and Molecular Genetics, Medicine, and Ecology and Evolutionary Biology, University of California, Irvine, California, USA

ABSTRACT ***Borrelia burgdorferi*, an emerging bacterial pathogen, is maintained in nature by transmission from one vertebrate host to another by ticks. One of the few antigens against which mammals develop protective immunity is the highly polymorphic OspC protein, encoded by the *ospC* gene on the cp26 plasmid. Intragenic recombination among *ospC* genes is known, but the extent to which recombination extended beyond the *ospC* locus itself is undefined. We accessed and supplemented collections of DNA sequences of *ospC* and other loci from ticks in three U.S. regions (the Northeast, the Midwest, and northern California); a total of 839 *ospC* sequences were analyzed. Three overlapping but distinct populations of *B. burgdorferi* corresponded to the geographic regions. In addition, we sequenced 99 *ospC* flanking sequences from different lineages and compared the complete cp26 sequences of 11 strains as well as the cp26 *bbb02* loci of 56 samples. Besides recombinations with traces limited to the *ospC* gene itself, there was evidence of lateral gene transfers that involved (i) part of the *ospC* gene and one of the two flanks or (ii) the entire *ospC* gene and different lengths of both flanks. Lateral gene transfers resulted in different linkages between the *ospC* gene and loci of the chromosome or other plasmids. By acquisition of the complete part or a large part of a novel *ospC* gene, an otherwise adapted strain would assume a new serotypic identity, thereby being comparatively fitter in an area with a high prevalence of immunity to existing OspC types.**

IMPORTANCE **The tick-borne zoonosis Lyme borreliosis is increasing in incidence and spreading geospatially in North America. Further understanding of the evolution and genetics of its cause, *Borrelia burgdorferi*, in its environments fosters progress toward ecologically based control efforts. By means of DNA sequencing of a large sample collection of the pathogen from across the United States, we studied the gene for the bacterium's highly diverse OspC protein, protective immunity against which develops in animals. We found that the distributions and frequencies of types of OspC genes differed between populations of *B. burgdorferi* in the Northeast, the Midwest, and California. Over time, OspC genes were transferred between strains through recombinations involving the whole or parts of the gene and one or both flanks. Acquisitions of OspC genes that are novel for the region confer to recipients unique identities to host immune systems and, presumably, selective advantage when immunity to existing types is widespread among hosts.**

Received 31 May 2010 **Accepted** 18 August 2010 **Published** 28 September 2010

Citation Barbour, A. G., and B. Travinsky. 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.

Editor Paul Keim, Northern Arizona University

Address correspondence to Alan G. Barbour, abarbour@uci.edu.

Lyme borreliosis is a vector-borne zoonosis with multiple reservoir hosts within sylvatic cycles in temperate regions of the northern hemisphere (1). In North America, the agent is the spirochete *Borrelia burgdorferi*. A related, often sympatric species is *Borrelia bissettii*, but this species has not been associated with human disease. Among the reservoirs for *B. burgdorferi* are rodents, other small mammals, and ground-foraging birds. Its tick vectors in North America are *Ixodes scapularis* and *Ixodes pacificus*. Sequential and mixed infections of reservoirs and tick vectors are common (2, 3). In regions where *B. burgdorferi* is enzootic, sites as small as a few hectares have between 9 and 15 strains (4–6).

Lyme borreliosis is the most frequently reported arthropod-borne infection in the United States and continues to increase in incidence and geographic area of risk (7). Three regions in the United States have enzootic transmission of *B. burgdorferi* and moderate-to-high risk of infection for humans; these regions comprise several states of the Northeast (Connecticut, Massachusetts, New York, Rhode Island, New Jersey, Maryland, and Pennsylvania), the Midwest (Michigan, Illinois, Minnesota, and Wisconsin), and northern California. The absence or rarity of *B. burgdorferi* in Ohio and in the Rocky Mountain and Great Basin regions suggests that these three populations of *B. burgdorferi* are geographically if not ecologically isolated.

Our aim was not to reconstruct the history of the species *B. burgdorferi* in North America. Other studies have endeavored this (8, 9). Rather, our attention was on a systematically collected set of samples of *B. burgdorferi* from across the United States and the use of those samples to infer the evolution of the highly polymorphic *ospC* gene, which encodes OspC, a surface-exposed lipoprotein of *B. burgdorferi* and other Lyme borreliosis-related species. OspC is homologous to the Vsp protein of the sister taxon of species that cause relapsing fever (10). The OspC and Vsp fam-

ilies of proteins have marked differences in their primary sequences within each family while retaining an α-helical bundle structure (11, 12). Whereas relapsing fever *Borrelia* genomes each have several different *vsp* alleles (13, 14), *B. burgdorferi* genomes have a single *ospC* gene, which is located on cp26, a circular plasmid of 26 kb (15). There is no antigenic variation of OspC during experimental infection (16), but within a given geographic area, several strains, each expressing a different OspC protein, coexist (4, 6).

A serotype is "a serologically distinguishable strain of a microorganism" (*Oxford English Dictionary*, 2nd ed., 1989). If any single component of *B. burgdorferi* determined serotypic identity, OspC could be it, for the following reasons. (i) No other gene of *B. burgdorferi* approaches *ospC* in diversity of alleles (17). (ii) The type-specific immunity conferred by immunization with OspC corresponds to the strain-specific immunity to *B. burgdorferi* in natural infections (18, 19). An animal immunized with one OspC type is protected against infection by a strain with the same OspC type but not a strain with a different one. (iii) The duration of expression of *ospC* by *B. burgdorferi* in mice coincides with the 1- to 3-week period in which *B. burgdorferi* circulates in the blood (20–22).

The population structure of *B. burgdorferi* reflects the contributions of mutation and recombination and contains both clonal and nonclonal elements (4, 17, 23–25). Traces of lateral gene transfer and recombination were evident in the plasmids of this species (17, 23, 26). While the *ospC* gene is heavily marked by the recombination process (27–29), this locus was reported to have phylogenetic consistency with chromosomal loci for strain collections largely limited to the Northeast (4, 23, 30). When more strains from the Midwest were included in the analysis, exceptions to strict linkage disequilibrium between *ospC* alleles and chromosomal markers were observed (24).

In our study of strains from the Midwest and California as well as the Northeast, we noted incongruities between *ospC* alleles and chromosomal loci across the three geographic areas (25). Some *ospC* alleles occurred in two different lineages, and there were indications that the same lineage could have two different *ospC* alleles. Differences between the *B. burgdorferi* populations of the Northeast and the Midwest were also apparent by multilocus sequence typing (MLST) of the spirochete in *I. scapularis* ticks collected over different transmission seasons (9).

For the present study, we first completed determination of the *ospC* genotypes of *B. burgdorferi* from a survey of infected *I. scapularis* ticks from the Northeast and the Midwest (8, 31) and then included samples from *I. pacificus* ticks from northern California (5) and additional ticks from the Midwest. For a fuller accounting of the variety of recombinations involving *ospC* genes, we extended the sequence analysis to the flanking regions of *ospC* genes in a large sample of strains and to the entire cp26 plasmid for a subset of these. These results provide insight into the evolution and distribution of *ospC* of Lyme borreliosis-related *Borrelia* spp.

RESULTS

Diversity and geographical distribution of *ospC* alleles. We accessed the most extensive and systematic collection of *B. burgdorferi* samples yet assembled. Twenty-five *ospC* types were identified in populations in the Northeast, the Midwest, and California (see Table S1 in the supplemental material). There were two subtypes, a and b, each for types D, H, I, and U, and three subtypes, a, b, and c, for type F. The codon-based nucleotide alignment of the 32 types and subtypes excluded the coding sequence for the signal peptide, the conserved first 11 residues of the mature lipoprotein, and the conserved last 4 C-terminal residues (see data set W1 at http://spiro.mmg.uci.edu/data/ospC). There were 549 sites in the alignment with at least one gap at 30 positions; 294 of the sites were variable. The overall nucleotide diversity per site (π) was 0.189. The 300 pairwise distances between the 25 *ospC* type sequences of *B. burgdorferi* were normally distributed, with a mean and median of 0.29 and a standard deviation of 0.05 (see Fig. S1 in the supplemental material). Figure 1 shows the unrooted, network-based phylogram of the 32 type and subtype sequences of *B. burgdorferi* and an *ospC* sequence of *B. bissettii*. With some exceptions, which are considered below, the DNA sequences of the alleles are approximately equidistant from each other.

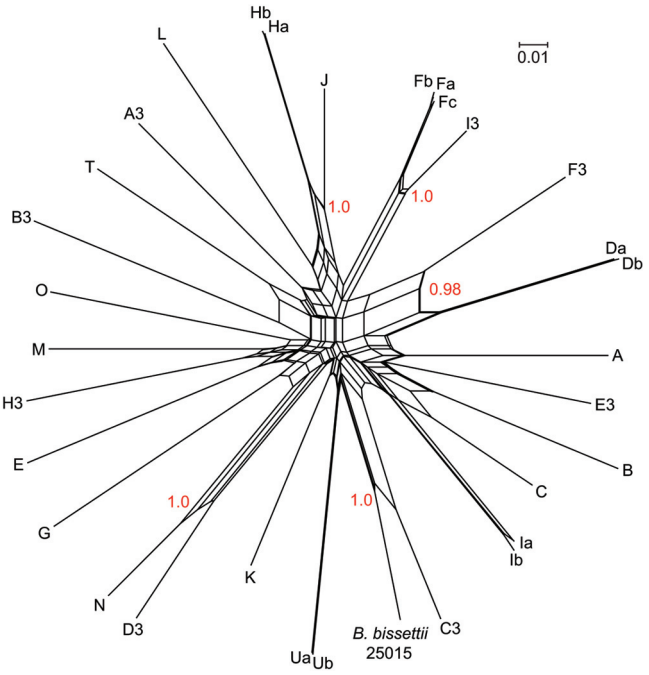

FIG 1 Unrooted distance phylogram with recombination network for codon-aligned *ospC* genes of 31 strains of *Borrelia burgdorferi* and 1 strain of *B. bissettii*, as implemented by SplitsTree. Nodes with posterior probabilities of ≥0.95 by Bayesian inference are shown. The scale bar indicates the distance.

Table 1 summarizes the frequencies of *ospC* alleles in *I. scapularis* nymphs in 3 geographic regions; Data Set S1 in the supplemental material provides the determinations by site and geographic coordinates. Without subtype distinctions, there were 18 alleles in the Northeast, 23 alleles in the Midwest, and 12 alleles at the northern California sites. The tick specimens from the Northeast ($n = 396$) and the Midwest ($n = 443$) were collected by the same protocol and over the same time periods (8). There were 18 types (exclusive of subtypes) common to both regions, but their individual prevalences among 396 sequences from the Northeast and 393 sequences from the Midwest were statistically different (likelihood ratio statistic = 211; $P = 10^{-9}$). The coefficient of determination (R^2) for rankings paired for the 18 types was only 0.18 (Spearman rank $Z = 1.52$; $P > 0.05$).

Figure 2 shows the collection sites for *I. scapularis* nymphs in the Northeast ($n = 23$) and the Midwest ($n = 28$) and the Bayesian

TABLE 1 Prevalence of *ospC* alleles of *B. burgdorferi* in *Ixodes* ticks in 3 geographic areas

ospC type or subtype	Northeast			Midwest			California		
	No. of *ospC* alleles	Prevalence	Rank[a]	No. of *ospC* alleles	Prevalence	Rank	No. of *ospC* alleles	Prevalence	Rank
A	57	0.144	2	30	0.068	3	22	0.111	4
B	48	0.121	3	14	0.032	11	10	0.051	8
C	15	0.038	11	6	0.014	20	0	0.000	
Da	15	0.038	12	0	0.000	5	0	0.000	6
Db	0	0.000		23	0.052		17	0.086	
E	16	0.040	10	6	0.014	20	1	0.005	11.5
Fa	18	0.045	7	17	0.038	2	19	0.096	5
Fb	2	0.005		33	0.074		0	0.000	
Fc	0	0.000		8	0.018		0	0.000	
G	10	0.025	13	13	0.029	14	12	0.061	7
Ha	21	0.053	6	0	0.000	1	0	0.000	3
Hb	1	0.003		101	0.228		25	0.126	
I[b]	19	0.048	8	14	0.032	11	0	0.000	
J	3	0.008	16	7	0.016	18	0	0.000	
K	77	0.194	1	17	0.038	9	0	0.000	
L	1	0.003	17	29	0.065	4	0	0.000	
M	4	0.010	15	18	0.041	8	7	0.035	9
N	37	0.093	4	20	0.045	7	0	0.000	
O	6	0.015	14	8	0.018	17	0	0.000	
T	17	0.043	9	13	0.029	14	1	0.005	11.5
Ua	28	0.071	5	10	0.023	11	0	0.000	
Ub	0	0.000		4	0.009		0	0.000	
A3	0	0.000		6	0.014	20	0	0.000	
B3	1	0.003	18	2	0.005	22	0	0.000	
C3	0	0.000		13	0.029	14	0	0.000	
D3	0	0.000		1	0.002	23	0	0.000	
E3	0	0.000		9	0.020	16	31	0.157	2
F3	0	0.000		21	0.047	6	0	0.000	
H3	0	0.000		0	0.000		48	0.242	1
I3	0	0.000		0	0.000		5	0.025	10
Total	396	1.000		443	1.000		198	1.000	

[a] If counts were equal, then the same rank was assigned; subtypes of D, F, H, and U were combined.
[b] All type I alleles were subtype Ia.

posterior probability contours for a model of two different populations of *B. burgdorferi*, as defined by *ospC* alleles (see Data Set S1 in the supplemental material). With a longitude of −83°W, which passes through Detroit, MI, and Columbus, OH, as a dividing line, the probabilities for membership in one or the other cluster were >0.99 for the first cluster for 396 of the 396 samples from the Northeast and >0.99 for the second cluster for 442 of the 443 samples from the Midwest. The probabilities were lower overall when the models specified 3, 4, or 5 populations. When the *ospC* alleles and their geographic locations for the 839 samples were randomly permuted as a negative control, only one population was defined.

The *I. pacificus* ticks were collected in California under a different protocol. While data from the two collections are not fully commensurate, the relative frequencies of the different alleles between the three regions can be compared (Table 1). Some *ospC* types, such as A and H, occurred in all 3 regions, but others were absent from one or two regions. For example, type K was the most prevalent allele in the Northeast sample but was a minor type in the Midwest and was not detected at all in California. Only 9 (50%) of the 18 *ospC* alleles from the Northeast were found in the California samples. Nevertheless, for each region that was sampled, the pairwise distances in sequence between the prevalent alleles were very similar in their distributions: the means ± standard deviations were 0.29 ± 0.05, 0.29 ± 0.05, and 0.28 ± 0.05 for the indigenous alleles from the Northeast, the Midwest, and California, respectively. This suggests that across the three regions, the balancing selections were of the same magnitude.

Recombination within *ospC* genes. These *ospC* findings confirmed the results obtained with chromosomal loci (5, 9): the three populations of *B. burgdorferi* overlap but are genetically distinct. Not only did the *ospC* alleles differ in frequency between regions, they also differed in their associations with chromosomal and other loci across regions (25). One explanation for the latter finding is lateral gene transfer of *ospC* genes or their fragments. Before considering recombination events extending beyond the *ospC* locus, we examined the evidence for recombination within *ospC* alleles.

The "star" pattern and the tangle at the root of the *ospC* tree in Fig. 1 indicate a mosaic genetic structure for *ospC* in these populations. The recombination breakpoints and diversity levels were highest in the middle of the coding sequence (see Fig. S2 in the supplemental material). Most *ospC* alleles were patchworks, the accumulated effects of multiple recombinations involving donors within the species and genus (27). A possible relic of an event is the 15-bp region (positions 532 to 546), by which subtypes Ia and Ib solely differ. In Ia, the sequence is GAA TCA GTA AAA AAC, which encodes the peptide ESVKN, while in Ib, it is AAA GCA GTA GAG GTC, which encodes KAVEV. The former nucleotide

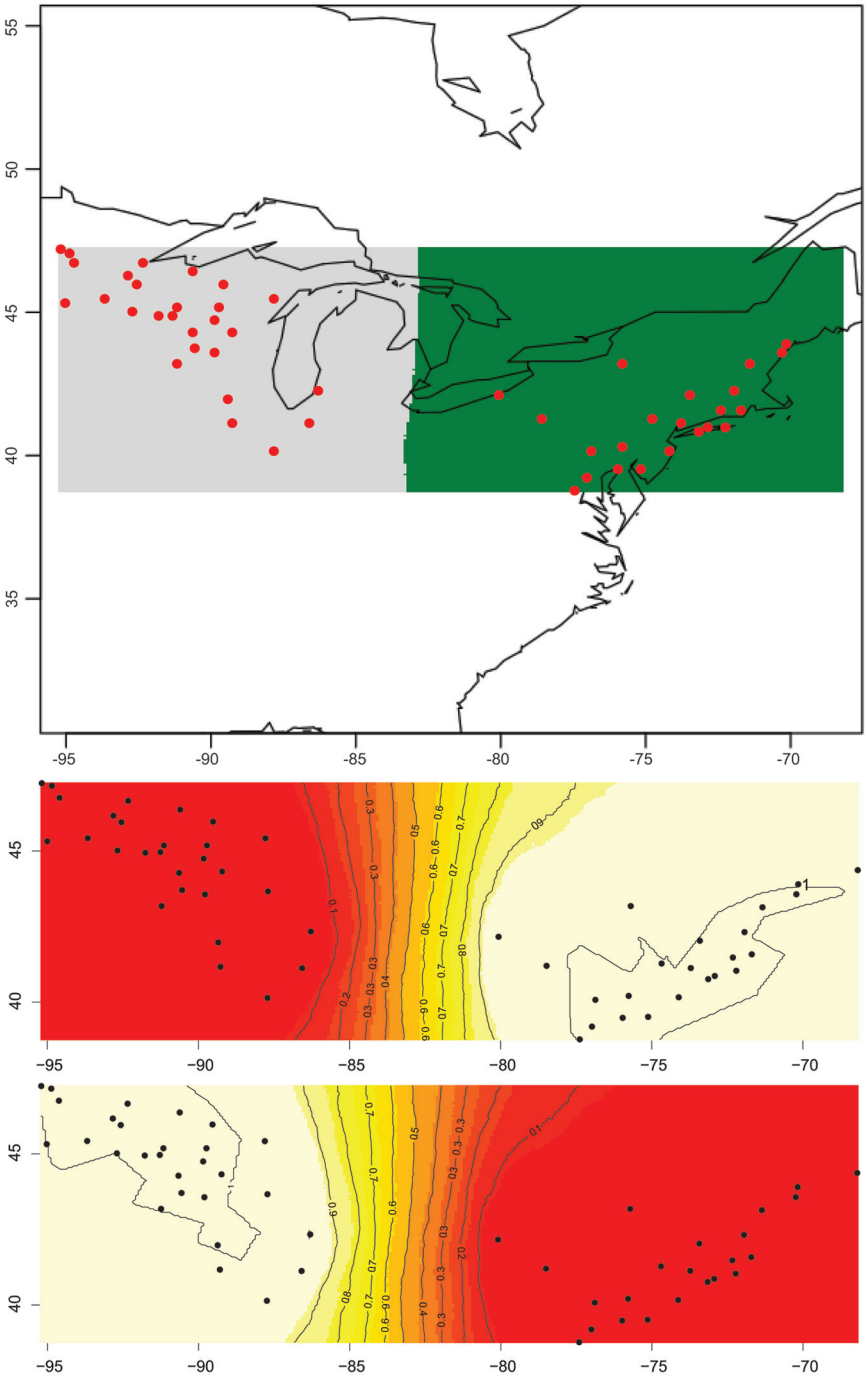

FIG 2 Population structure of the *ospC* gene of *B. burgdorferi* in *Ixodes scapularis* nymphs in the Northeast (23 sites) and the Midwest (28 sites) regions of the United States for the years 2004 to 2007. Upper panel, overview map of posterior mode of population membership. Middle and lower panels, contour of posterior probabilities (0.1 to 1.0) for two populations defined by 27 *ospC* sequence types and subtypes identified in these regions. The *y* coordinates are latitude, and the *x* coordinates are longitude. The data for the analysis and the resultant posterior probabilities for cluster 1 or 2 for each site are given in Data Set S1 in the supplemental material.

sequence is identical to the corresponding sequence of the type M allele, while the latter is identical to type E3's sequence.

Some pairs of *ospC* alleles plausibly have a closer evolutionary relationship. In the phylogram in Fig. 1, the pairs comprising H and J, F and I3, N and D3, F3 and D, and C3 and an *ospC* locus of *B. bissettii* are indicated by significant support for common internal nodes. These pairs featured longer stretches of identity or near-identity between the types than was observed for other pairs (see Fig. S3 in the supplemental material). I3 and A, comprising a type pair, do not stand out as recombinants in Fig. 1, but this is more clearly evident in the alignment of their sequences. As Girard et al. noted (5), type I3 OspC is a chimera of type F for the first two-thirds of the protein and of type A for the last third.

Recombination of cp26 plasmids. California strains with a type I3 *ospC* locus had *rrs-rrlA* and *rrfA-rrlB* intergenic spacers that were identical in sequence to those of type Fa strains but not those of type A strains (25), an indication that a type A-bearing strain was the donor for the chimeric gene. We return to this particular event after first examining the more general case of recombination of cp26 plasmids.

For this analysis, we assembled complete sequences of the 26-kb plasmids bearing *ospC* in the species. The cp26 sequences of the following 10 strains were publicly available: B31, 64b, ZS7, WI91-23, 29805, 94a, 72a, 118a, CA-11.2a, and 156a. All but 3 strains (ZS7, WI91-23, and CA-11.2a) were from the Northeast. For additional representation from outside this region, we determined the sequence of cp26 of the California isolate CA8. Since the mosaic character of the *ospC* locus itself could overwhelm the detection process, we removed its coding sequence from the alignment. The resultant 11 aligned sequences had 25,934 positions, with 749 informative sites (see data set W2 at http://spiro.mmg.uci.edu/data/ospC).

The numbers of recombination events per site (rho) and mutation events per site (theta) were 0.051 and 0.011, respectively, for a rho/theta ratio of 4.7. For the entire 11-sequence alignment with a window size of 100 characters, the PhiTest for recombination gave a mean test value of 0.563, with a variance of 10^{-4} and an observed value of 0.203 ($P < 10^{-5}$). Figure 3 shows the inferred recombination breakpoints and their significances along the lengths of the aligned sequences. The highest density of breakpoints surrounded the position which *ospC* would otherwise occupy. The only other location with a near-comparable density of breakpoints was centered on position 6400, which is in the open

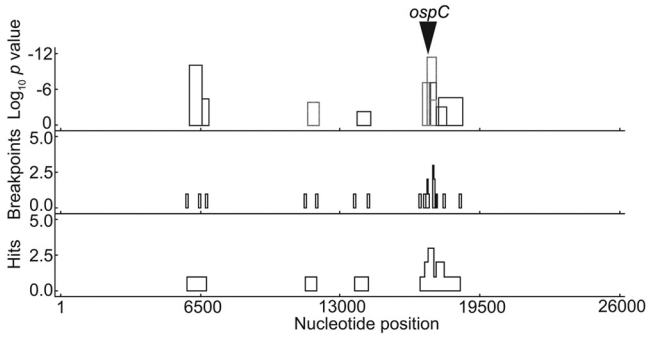

FIG 3 Recombination detection analysis of cp26 plasmids of 11 strains of *B. burgdorferi*. The *ospC* coding sequences were removed at the position indicated by the point of the arrowhead. The alignment was subjected to the SciScan algorithm with a window size of 200, a step size of 20, and 100 permutations. The *x* axis provides the nucleotide positions of the alignment. The bottom panel ("Hits" on the *y* axis) indicates the sequence region where recombination was detected. The middle panel gives the number of estimated breakpoints per window. The top panel indicates the $\log_{10}$ of the *P* value for the test statistic for recombination detection.

reading frame BBB08, encoding a hypothetical lipoprotein. There was little or no evidence of recombination at the beginnings and ends of the sequences in the alignment, which are actually contiguous in these circular plasmids.

As a comparison to the cp26 plasmids, we aligned the nucleotide sequences of the lp54 linear plasmids of the 11 strains and carried out the same analysis (see Fig. S4 in the supplemental material). The alignment included the polymorphic alleles for decorin binding protein A (*dbpA*). The rho/theta ratio for lp54 was lower, at 2.9, than that for cp26 with the *ospC* locus excluded. There was evidence of recombination in the *dbpA* gene for some strains but, in contrast to the discordance between trees for *ospC* and other loci, the topology for *dbpA* largely matched the tree topology for the full-length plasmids (Fig. S4).

The gene for hypothetical protein BBB02A was the cp26 locus that had several sites informative for phylogenetic inference, but without evidence of recombination. The infrequency of recombination may be attributable to the adjacency of this gene to *bbb03*, the gene for telomere resolvase, which is essential for replication (32). With the exception of one node, the tree topology of *bbb02* sequences was concordant with that of the entire group of plasmids (see Fig. S5 in the supplemental material). For the following strains, the presence or absence of a 3-bp insertion at position 397 of the 441- or 444-bp gene sufficed to define the cluster or genotype, respectively: (i) 118a, 72a, CA-11.2a, 94a, CA8, and 156a or (ii) WI91-23, 29805, ZS7, 64b, and B31. These characteristics qualified *bbb02* as a proxy for the plasmid, and accordingly, we determined the *bbb02* sequences of 45 additional *B. burgdorferi* strains, with an emphasis on strains with different linkages of *ospC* types to chromosomal loci (see data set W3 at http://spiro.mmg.uci.edu/data/ospC).

Recombination beyond the *ospC* genes. Examining flanking regions for the *ospC* gene, we noted that the sequences on each side could be conveniently grouped into 13 sets of oligonucleotide characters of 2 to 20 nucleotides (nt) that each included informative polymorphisms (see Fig. S6 in the supplemental material). (Here, "character" is defined in accordance with systematics usage: a variable feature with two or more different states.) For instance, among the 11 cp26 plasmids, there were only 4 variants for the sequence beginning 67 nucleotides 5′ of the *ospC* start site: CAAATA, CAAAT–, ATTTG–, and ATTTGA. There were 5 oligonucleotide characters, designated *a*, *b*, *c*, *d*, and *e*, that were upstream of the *ospC* coding region, and downstream of the stop codon were the characters *h*, *i*, *j*, *k*, *l*, and *m*. We also included in the analysis the characters *f* and *g*, from the front and end of the *ospC* gene, respectively. With the exception of the *g* character, which typified *ospC* diversity, there were no more than 6 variants per character.

We amplified and sequenced a cp26 fragment that extended from a position 229 nucleotides upstream of the *ospC* gene start codon to one 534 nucleotides downstream of the stop codon. This ~1.5-kb fragment corresponded to the plasmid region with the high density of probable breakpoints (Fig. 3). The sequencing was carried out on 99 selected isolates or tick extracts with *ospC* alleles that were linked to different *rrs-rrlA* loci

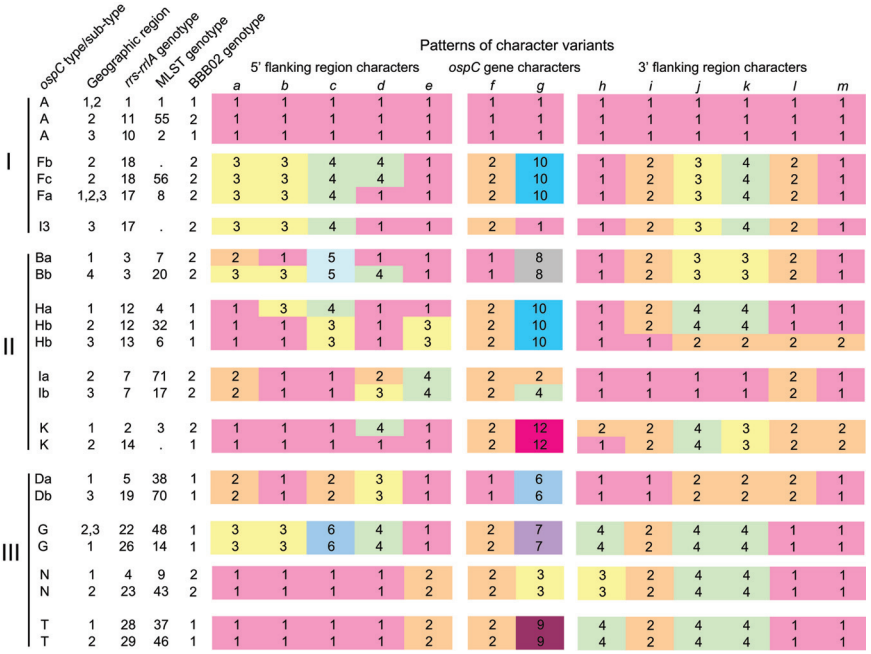

FIG 4 Patterns of variants of flanking regions for *ospC* genes of pairs and trios of *B. burgdorferi* strains. The 13 oligonucleotide characters and their variants are given in Fig. S6 in the supplemental material, and the alignment is given in Data Set S2 in the supplemental material. The characters are indicated by italicized lowercase letters. Five characters (*a* to *e*) are from the 5′ flanking region for the *ospC* gene, two characters (*f* and *g*) are from the *ospC* gene itself, and six characters (*h* to *m*) are from the 3′ flanking region. The *ospC* type or subtype is indicated in the leftmost column. The geographic regions are the Northeast (1), the Midwest (2), northern California (3), and Europe (4). The *rrs-rrlA* intergenic spacer and MLST genotypes were as defined by Travinsky et al. (25). Groups I, II, and III are described in the text.

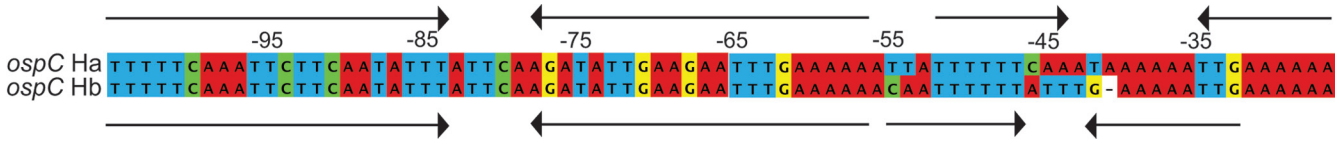

FIG 5 Alignment of sequences of the upstream and promoter regions for *ospC* genes of subtype Ha and Hb strains of *B. burgdorferi*. The locations of inverted repeats are indicated by arrows. Positions are numbered with respect to the transcriptional start site for *ospC*.

and were from different geographic regions (see Data Set S2 in the supplemental material). For the alignment, we also included the corresponding sequences from the 11 strains for which the complete cp26 sequences were available. Figure 4 schematically represents the variety of patterns for the 13 oligonucleotide characters before, within, and following *ospC* genes. Included in the figure are the geographic origins, the *rrs-rrlA* intergenic spacer and MLST genotypes, and the cp26 classification by *bbb02* genotype 1 or 2.

The type I3 *ospC* gene observed in several samples from California is attributable to a recombination between a recipient strain bearing a cp26 plasmid with a type Fa *ospC* gene and a donor strain bearing a cp26 plasmid with a type A gene (Fig. 4, group I). The I3 isolates had the same *rrs-rrlA* and *rrfA-rrlB* intergenic spacer sequences (25) and the same *bbb02* sequences (see data set W3 at http://spiro.mmg.uci.edu/data/ospC) as Fa isolates. The I3 *ospC* gene characters *f* and *g* were the same as the corresponding characters of type F and type A, as befits a chimera. But the 5′ and 3′ flanking regions for the I3 *ospC* gene were identical to those for subtype Fa *ospC* and not those for type A-bearing cp26 plasmids. The presumptive proximal crossover was within the sequence TACTGATG, which both type A and subtype Fa share at positions 450 to 457 of the 630-bp-long type A gene. The I3, Fa, and A strains all had variant 1 of oligonucleotide character *h* but differed over characters *i*, *j*, *k*, and *l*, suggesting that the distal crossover point was either among the coding sequence's last 30 nucleotides or among the following 106 nucleotides.

Strains bearing *ospC* subtype Ha or Hb were an example of another type of recombination (Fig. 5, group II). The three strains had different MLST profiles and different *ospA* sequences (25). They can also be distinguished by their 5′ and 3′ flanking regions. The Ha-bearing strain in the Northeast has the same 3′ flanking region as Hb-bearing strains of the Midwest, but over their 5′ flanks, the Midwest and California strains with Hb alleles are identical. The Ha and Hb alleles differ by a single synonymous substitution, which is near the gene's 5′ end, consistent with a recombination involving the 5′ flanking region and the *ospC* gene itself. All the type H representatives were *bbb02* genotype 1.

The other pairs of strains with the same or near-identical *ospC* genes, different MLST or rrs-*rrlA* spacer genotypes, and differences in their 5′ and/or 3′ flanking regions involved types B, I, and K. Two strains, exemplified by isolates 64b from the Northeast and ZS7 from Europe, had subtypes Ba and Bb, respectively. There are several differences between 64b and ZS7 in the 5′ flank to the *ospC* genes (Fig. 4; see also Data Set S2 in the supplemental material). Three of the four polymorphic positions distinguishing Ba and Bb occur in the first third of the sequence and cluster within 15 positions. This is consistent with a lateral transfer of a fragment that included the 5′ end of *ospC* and the adjacent upstream region. Whereas the strains with subtypes Ia and Ib differed in their 5′ flanking regions, the two polymorphic positions between the *ospC* alleles occurred in character *g* at their 3′ ends. The pairs involving the B and I strains had the same *bbb02* genotypes. The type K strains from the Northeast and the Midwest had identical *ospC* sequences but differed in their 5′ and 3′ flankings as well as in both the *rrs-rrlA* spacer and the *bbb02* genotypes, suggestive of the transfer of the entire *ospC* gene between lineages.

Another group, group III (Fig. 4), comprised the pairs involving types D, G, N, and T, which had different chromosomal genotypes but the same *ospC* genes and the same flanking regions to the extent of our sequencing. Also qualifying for this group were the three type A strains in the sampling. The two type A strains from the Northeast or California and the three samples of a Midwest type A strain had *bbb02* sequences that were in the two different clades, suggestive of transfer of either an entire plasmid or an extensive length of a plasmid.

Recombination in the promoter region. A possible consequence of replacement of all or part of *ospC* is a collateral effect on adjacent loci or on regulatory regions. We noted that the oligonucleotide character *e* was located just upstream of the "−35" box of the *ospC* promoter. Substitutions in this area could affect the inverted repeats implicated in regulation of *ospC* expression as an operator or through supercoiling (33, 34). Two type H strains, one of subtype Ha and the other of subtype Hb, differed in character *e* (Fig. 4). Figure 5 shows for these two strains the upstream sequences, numbered in reference to the transcriptional start site (35); oligonucleotide *e* corresponds to positions −55 to −42. The first inverted repeats, which spanned positions −105 to −54, were the same in sequence and location for the pair, but the second inverted repeats, which included the "−35" σ^{70} promoter element, were different by an indel and 5 substitutions. Although it was shifted upstream by 4 nucleotides and had a different sequence, the Hb strain still had a predicted stem-loop and a ΔG and a melting temperature (T_m) of −25 kcal/mol and 62.7°C, instead of −22.8 kcal/mol and 61.6°C, respectively, for the Ha strain.

Transfer of an entire *ospC* gene. To this point, we have examined pairs or trios of strains with the same or near-identical *ospC* alleles and found evidence of lateral gene transfer of all or part of the *ospC* gene and, in addition, different lengths of sequence on one or both sides of the locus. We next looked at another possible outcome of lateral gene transfer, namely, the occurrence of substantially different *ospC* genes in members of the same lineage. Notwithstanding the cumulative effects of intra- and interspecies recombination on the chromosome as well as plasmids (17, 36, 37), there was evidence that two strains, 72a and 118a, occupied an internal node of comparatively recent origin. Although strains 72a and 118a have type G and J *ospC* alleles (24) and different *ospA* alleles on their lp54 plasmids (25), strains 72a and 118a had the same *rrfA-rrlB* intergenic spacer (25) and the same *dbpA* sequences (see Fig. S4 in the supplemental material).

We extended the comparison to include sequences for 8 ribosomal protein genes, which are considered informational genes

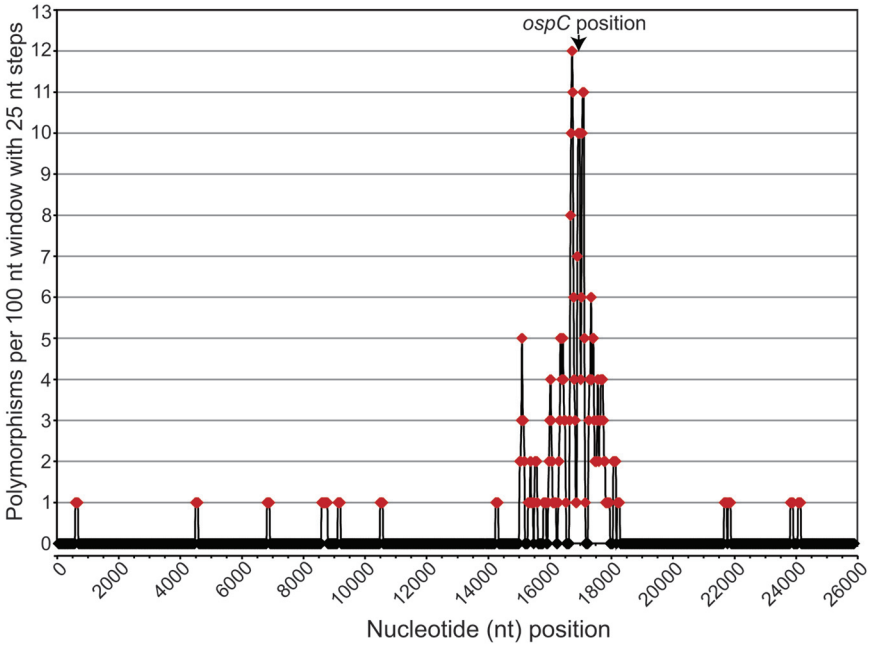

FIG 6 Nucleotide polymorphisms between the cp26 plasmid sequences, excluding the *ospC* genes, of strains 72a and 118a of *B. burgdorferi*. The position in the sequence where the *ospC* gene sequence was deleted is shown by an arrow.

and, thus, less susceptible to whole or partial replacement than are operational genes, such as that for a metabolic enzyme (38). Among the 11 strains with genome sequences, only 72a and 118a had identical sequences for each of these 8 ribosomal protein genes (see Fig. S7 in the supplemental material). With the MLST set of eight operational housekeeping genes, base substitutions between 72a and 118a were noted, but these were fewer than was observed between other pairs of strains (Fig. S7), and the two strains retained their positions with respect to strain CA-11.2a. Strains 72a and 118a also had the same *bbb02* genotypes and the most closely related cp26 and lp54 plasmids among the 11 strains examined (Fig. S4 and S5). The taxonomic relationship of 72a and 118a with CA-11.2a that was observed with the two sets of chromosomal loci held true for the cp26 sequences. Figure 6 shows the locations of nucleotide differences between the cp26 plasmids of strains 72a and 118a with exclusion of *ospC* coding sequences. The greatest difference, by far, between the two cp26 sequences was at positions on each side of *ospC*, a region extending for ~2 kb on the 5′ side and ~1 kb on the 3′ side.

DISCUSSION

Shakespeare's late plays were more collaborative in authorship than was previously thought (39), and the same can be said of the origins of existing *B. burgdorferi* strains. Intraspecies recombination had a greater role in shaping the evolution of *B. burgdorferi* than was previously appreciated. With representation from three geographic regions, there were several exceptions to linkage disequilibrium between the plasmid-borne *ospC* gene and chromosomal loci (25). The present study extended the geospatial analysis and demonstrated that the population structures for the *ospC* locus overlapped between the three regions but were distinguishable, thereby confirming the results with MLST and *rrs-rrlA* spacer loci from smaller sample sizes (5, 8, 9). By analyzing whole cp26 sequences of 11 strains and the *ospC* flanking regions for a larger set of strains, we identified a variety of recombination events that contributed to the nature of North American *B. burgdorferi*.

In our view, strain-specific immunity of reservoir hosts is sufficient to account for the strong balancing selection at the *ospC* locus that is notable in *B. burgdorferi* population structures (4, 6, 40). But OspC has also been characterized in functional terms: some strains, defined by their *ospC* alleles, are associated with higher likelihoods of dissemination beyond the skin in humans or experimental animals (41, 42). One study attributed the different OspC phenotypes to differential binding of plasminogen (43). Although there are other candidates for host range determinants, such as complement-regulator factor H binding proteins (26, 44), a role for OspC in adaptations to different niches cannot be excluded. So, *ospC* diversity could arguably reflect the outcome of niche selection processes (41, 45). Nevertheless, we doubt that the observed antigenic diversity of OspC is merely epiphenomenal to functional differences between proteins. The range of pairwise sequence distances among *ospC* alleles nearly matches that of the highly polymorphic family of surface proteins of the relapsing fever agent *Borrelia hermsii*, which employs antigenic variation to evade host immunity (14). Possibly, both immune and niche selective forces are in play, but their relative contributions remain to be determined.

Retention of a gene, like *ospC*, that is necessary for tick-to-vertebrate transmission is more ensured by its location on cp26, apparently the only indispensable plasmid (15, 46). While possession of a cp26 plasmid is required for cell replication, it need not be the original cp26 plasmid. The plasmid encodes compatibility functions, and one cp26 plasmid can be displaced by another if they are incompatible (46). This potentially allows for replacement of entire cp26 plasmids through lateral gene transfer as well as a range of products of recombination between two plasmids that transiently coexist in the same cell. Transfer between *B. burgdorferi* of segments of DNA of ≤1 kb was noted (27), but recent findings indicate that lateral gene transfer may involve longer lengths of DNA.

We classify recombinations involving *ospC* into 5 patterns. The first is intragenic, that is, effectively limited to the OspC-coding sequence itself. Recombination within *ospC* was noted in several reports, beginning with Livey et al. (29), and accounts for its mosaic genetic structure. Possible examples of intragenic recombination include the *ospC* pairs comprising H and J, C and B, and E and H3 (see Fig. S3 in the supplemental material). But our focus here is on recombination outcomes that extend beyond *ospC*'s boundaries. The other four patterns are those that involve (i) part of the *ospC* coding sequence and a sequence extending into the 5′ flanking region, (ii) part of the *ospC* coding sequence and a sequence extending into the 3′ flanking region, (iii) the *ospC* gene and both

flanks but not the entire plasmid, and (iv) replacement of the entire cp26 plasmid.

Inclusion of a sequence upstream of the *ospC* coding sequence in the recombinant fragment could affect the promoter and inverted repeats that may constitute a regulatory element (33, 34). There resides also the *guaA-guaB* (*bbb17* and *bbb18*) operon, beginning on the complementary strand 185 nt before the *ospC* start site. On the 3′ side, there is a highly conserved sequence that would form a stem-loop typical of the rho-independent terminator (12, 35). There are also two short open reading frames for hypothetical peptides of 36 (BBB20) and 31 (BBB21) amino acids (aa) before the stop codon 431 nucleotides downstream on the complementary strand for the open reading frame BBB22, which is homologous to xanthine/uracil permeases of other bacteria. While there may be greater scope for rearrangements without disruptive effects downstream of *ospC*, a transcription-regulatory element may be changed in sequence without necessarily altering its function (Fig. 5).

When both flanking regions are involved in the recombination, an entire *ospC* allele may be substituted. The evidence is strongest for the closely related strains 72a and 118a. The presence of the same *ospC* type in different lineages is exemplified by the type K strains from the Northeast and the Midwest in the collection (Fig. 4). Other possible examples of the latter phenomenon involve the D, G, N, and T strains from different geographic areas. Recombinations that involved part of the *ospC* gene and either the 5′ or the 3′ flanking region were exemplified by strains of types H, B, and I.

An outcome (outcome 5), i.e., a displacement of one cp26 plasmid by an incompatible plasmid, has been observed in the laboratory (46). We observed discordant tree topologies for the cp26 sequences and the two sets of chromosomal loci (see Fig. S5 and S7 in the supplemental material). Only the strains 118a, 72a, and CA-11.2a maintained the same taxonomic relationship in all 3 phylogenies. But definitive examples of the outcome (outcome 5) were not found in the subset of strains for which the whole-genome sequence was available. The traces of this may be more apparent as more whole-genome sequences are available for comparative analysis.

The mechanisms for lateral gene transfer in *B. burgdorferi* in nature are unknown. As mixed infections are common (2, 3, 25, 47), there are opportunities for genetic exchange in both reservoir hosts and vectors. There is no evidence of conjugation in the genus, and transformation of *B. burgdorferi* is less efficient in the laboratory than is the case for many other bacterial species (48). But membrane vesicles or blebs, which have been shown to contain plasmid DNA (49), could be the vehicles for the higher frequency of transformation events in nature. The cp26 plasmid itself is not a prophage, but transduction of cp26 fragments via another virus, such as the prophage constituting the cp32 replicons (50), is possible.

The *ospC* gene is clearly transferable, but is it a mobile genetic element? It does not appear to have accessory genes associated with it. The flanking genes, *guaA* and *guaB* on one side and *bbb22* on the other, encode enzymes for nucleotide metabolism or uptake; these enzymes are not discernibly transposases or integrases. There may be a role for the sets of inverted repeats, which are on each side of *ospC* and potentially form recombinogenic stem-loop structures. These inverted-repeat regions may be included in the transfer, as we have demonstrated. But they need not be, since both recipient and donor have them for their transcription regulation and termination functions. Although we have found evidence of transfer of entire *ospC* genes in some lineages, a single recombination with incorporation of only part of the gene may be sufficient to confer a new antigenic identity for the recipient cell.

Finally, can we accommodate both the phylogeography and the inferred mechanisms of genetic diversity into a model of the evolution of this pathogen? OspC's prominence as an abundant, immunogenic surface protein, which is expressed as the spirochete enters the host's skin and then circulates in the blood, makes this protein an important target for protective immunity. The vertebrate hosts' immune responses subject the *ospC* gene to frequency-dependent balancing selection. While recombination does not create polymorphisms at the single-nucleotide level, interstrain exchange of two or more suitably distant sequences can yield novel combinations of substitutions and indels and, as an eventual consequence, a set of antigenically distinctive proteins. One now sees the cumulative effects of intragenic recombination, involving both long and short fragments and occurring in multiple rounds, in the highly polymorphic repertoires of *ospC* genes in three different populations of *B. burgdorferi*. But we have also seen that intragenic recombination may involve either of its flanking regions. Indeed, it may depend on one of these flanks for a stable heteroduplex if homologous rather than illegitimate recombination is the more common mechanism.

If both flanks were the substrates for a recombination with a heterologous sequence stretch between them, as occurs in the relapsing fever agent *B. hermsii* during antigenic variation (14), transfer of an entire *ospC* gene into a different strain would be achieved. This fits the general category of serotype shift and is distinguished from serotype replacement, in which the population structure of a pathogen changes as newly introduced strains gain a foothold in the presence of herd immunity to existing strains. In the case of *ospC*, a serotype shift would not create a novel allele *per se* when the greater population of *B. burgdorferi* is taken into account. But within a partially isolated geographic area, such as the Northeast, with the introduction, e.g., through migratory birds, of a new strain with an *ospC* locus that is locally unique, acquisition of this one determinant would presumably suffice for enhanced fitness when the beneficiary faces the prospect of reservoir hosts, a large proportion of which are immune to existing strains. The invading bacterial strain itself may not prosper in the new environment, perhaps for lack of other adaptations, e.g., a putative tick midgut adhesin or host-specific complement resistance, suited for parasitism of local ticks and reservoir hosts, such as *I. scapularis* and the deer mouse (*Peromyscus leucopus*) at the Northeast sites or *I. pacificus* and the western gray squirrel (*Sciurus griseus*) in California (5).

We propose that the *ospC* phenomenon is analogous to a single gene that upon acquisition and expression provides for a bacterium resistance to an antibiotic in an antibiotic-rich environment, like a hospital or poultry facility. We acknowledge that there may be unrecognized epistatic relationships for *ospC* that operate to constrain the variety of genetic backgrounds of *B. burgdorferi* in which a particular OspC protein can effectively function. But as long as a newly acquired *ospC* gene is faithfully positioned next to the promoter, retains the coding sequence for the conserved signal peptide, and is not truncated at its 3′ end, we presume that the novel OspC protein will be expressed and successfully transported

to the outer membrane and the cell's surface and function in its new bacterial host (51).

MATERIALS AND METHODS

Strains and culture. The cultivated *B. burgdorferi* strains were B31 (ATCC 35210), N40, JD 1, Sh2, 2665, and HB19 from the Northeast (52) and CA8, CA11, CA12, CA15, CA16, CA17, CA172, CA337, CA533, and CA534 from northern California (53). Strains VGQ, WQR, WQR27, and QQQ from the Northeast were provided by Merial Limited, Athens, GA. The strains were cultivated in modified Barbour-Stoenner-Kelly II medium and harvested by centrifugation at 9,500 × *g* for 20 min at 22°C (52).

DNA samples from ticks. The states of the Northeast represented in the study (with the number of *ospC* sequences per state indicated in parentheses) were Connecticut (28), Massachusetts (8), Maryland (55), Maine (39), New Hampshire (3), New Jersey (16), New York (176), Pennsylvania (47), Rhode Island (11), and Virginia (13); the Midwest states were Iowa (8), Illinois (17), Indiana (4), Michigan (20), Minnesota (218), and Wisconsin (176). The procedures for (i) collection of 7,749 questing *I. scapularis* nymphs during the years 2004 to 2007 at 23 collection sites in the Northeast and 28 in the Midwest, with recording of geospatial coordinates, (ii) extraction of DNA from the ticks, (iii) quantitative PCR for identification of ticks with *B. burgdorferi*, and (iv) genotyping of the *B. burgdorferi* isolates were described previously (8, 25). *B. burgdorferi* was identified in 1,540 (20%) of the ticks. PCR amplification of *ospC* genes was carried out blindly with respect to geographic location and was attempted on the 1,522 specimens for which sufficient DNA was available. We reported previously on 741 extracts, in which only a single *ospC* sequence was detected (25). Here, we include the results from an additional 198 extracts, out of a total of 241 with evidence of mixed infections, in which one of the *ospC* types in the mixture could be determined by sequencing, thereby bringing the total number of *ospC* type determinations from this study of the Northeast and the Midwest to 839 (see Data Set S1 in the supplemental material). Girard et al. described the collection of 214 *B. burgdorferi*-infected *I. pacificus* nymphs from 78 woodland sites in Mendocino County, CA, in 2004 (5); *ospC* was amplified and sequenced from 198 (93%) of these nymphs. DNA samples were stored in single-use aliquots at −80°C until use. To confirm the results from the aforementioned collections, we also characterized *ospC* sequences and other loci for 48 infected *I. scapularis* adults from the Midwest (provided by Sarah Hamer and Jean Tsao, Michigan State University) and *B. burgdorferi* isolates VGQ, WQR, WQR27, QQQ, and 2665.

Existing sequences. Table S1 in the supplemental material gives the GenBank accession numbers for existing chromosome, cp26, and *ospC* sequences. The naming conventions for *ospC* were described by Travinsky et al. (25). The MLST loci were *clpA*, *clpX*, *nifS*, *pepX*, *pyrG*, *recG*, *rplB*, and *uvrA* (54). The 8 ribosomal proteins were L1 (*rplA*), L2 (*rplB*), L3 (*rplC*), L4 (*rplD*), L5 (*rplE*), S2 (*rpsB*), S3 (*rpsC*), and S4 (*rpsD*). For strains for which annotation was incomplete, these genes were identified by using the sequence of strain B31 for a search with the BLASTn algorithm at the GenBank website or, in the case of the CA8 chromosome (see below), on a local server. These sequences as well as the MLST sequences were codon aligned and concatenated.

PCR. Spirochetes were lysed by suspending the pellet in 1 mM EDTA and then incubating it in boiling water for 30 min. Phusion DNA polymerase (Finnzymes, Woburn, MA) was used in Phusion HF buffer with 1.5 mM $MgCl_2$ and 0.4 μg/ml of bovine serum albumin. The final concentrations of each deoxynucleoside triphosphate (dNTP) and primer were 200 μM and 0.5 μM, respectively. Amplification of the *ospC* sequence corresponding to nucleotide positions 91 to 618 for strain B31's *ospC* gene was done by the method of Bunikis et al. (4), with the exception of 5′ GACTTTATTTTTCCAGTTACTTTTT 3′ for the reverse outer primer. The 5′ and 3′ flanking regions of *ospC*, corresponding to positions 16674 to 18070 for cp26 of strain B31, were obtained with the forward and reverse primers 5′ GGGATCCAAAATCTAATACAA 3′ and 5′ CCCTTAACATACAATATCTCTTC 3′, respectively. For the reaction, the 3-min denaturation step at 98°C was followed by 40 cycles at 98°C for 30 s, 60°C for 30 s, and 72°C for 90 s and finally a 7-min extension at 72°C. For strain B31, the size of the product was 1,397 bp. The bbb02 gene was amplified by nested PCR with outer forward, outer reverse, inner forward, and inner reverse primers of 5′ TTTAATTATAAGCTATAGTTTTT GTTTTT 3′, 5′ TGAAAAATTATTAAATGGGAATAAG 3′, 5′ ATTTGG GAAATATTAGGAAATATT 3′, and 5′ TGGGAATAAGTATTCAAA CATT 3′, respectively. The PCR conditions were the same, except the annealing temperature was 55°C and the extension was for 30 s.

DNA sequencing. PCR products were purified using DNA Clean & Concentrator-5 (Zymo Research, Orange, CA) kits and were sequenced directly over both strands by the dideoxy method with a CEQ 8000 DNA sequencer (Beckman-Coulter, Fullerton, CA) or an Applied Biosystems 3730xl DNA analyzer. The sequences of 110 *ospC* genes and their 5′ and 3′ flanking regions are given in Data Set S2 in the supplemental material; sequences of 56 *bbb02* genes are given in data set W3 at http://spiro.mmg.uci.edu/data/ospC.

Genome sequencing. DNA was extracted from strain CA8 by use of a DNeasy tissue kit (Qiagen, Valencia, CA). At Ambry Genetics (Aliso Viejo, CA), the DNA was sheared to an average size of 200 bp, the ends were filled in, and adapters were added. The ligated products were size selected by gel purification and then amplified by PCR with primers for the adapters. Library size and fragment concentration were assessed with an Agilent Bioanalyzer (Santa Clara, CA). The paired-end library yielded ~150 × 10^3 clusters per tile and was sequenced using 39 cycles with an Illumina Genome Analyzer IIx instrument (Hayward, CA). Initial data processing was performed with the Illumina RTA program (SCS version 2.4). Base calling and sequence quality filtering scripts were executed with the Illumina pipeline software program (version 1.4). The assembly was *de novo*, and the depth of coverage was ≥20×.

Phylogenetic, recombination detection, and population genetic analyses. Sequences were aligned using Clustal X (55). DNA distances were determined with the DNADIST algorithm, as implemented by Mobyle (http://mobile.pasteur.fr). Nucleotide polymorphism was assessed with DnaSP version 5.10 (56). Phylogenetic inference for coding sequences was carried out by Bayesian estimation as implemented by MrBayes version 3.1.2 (http://mrbayes.csit.fsu.edu/) (57), by maximum likelihood estimation as implemented by PhyML version 3.0 (http://www.atgc-montpellier.fr/phyml/) (58), or by phylogenetic network analysis as implemented by SplitsTree version 4.10 with the NeighborNet protocol (http://www.splitstree.org/) (59). The evolutionary model for protein-encoding regions was estimated with ModelTest (60). For Bayesian analysis, there were 10^6 generations with the first 2,000 sampled trees discarded. For maximum likelihood analysis, there were 1,000 iterations. For whole-plasmid sequences, neighbor-joining and maximum likelihood phylograms were generated with Phylo-win (http://pbil.univ-lyon1.fr/software/); the observed differences were the distance setting, and the empirical transition-to-transversion ratio was the maximum likelihood setting. Recombination detection and analysis were carried out with the RDP3 suite (http://darwin.uvigo.es/rdp/rdp.html) (61). The SciScan method was used for assessing signals of recombination (62). The PhiTest was also used to assess the likelihood of recombination (63). The *R* statistical package Geneland was used for stochastic simulation and MCMC (Markov chain Monte Carlo)-based inference of population structure from genetic and geographical data (http://www2.imm.dtu.dk/~gigu/Geneland/) (64). There were 100,000 iterations and thinning by 100; the assumptions were the false-null-allele model and an uncorrelated allele frequency. The postprocess setting was 300 × 150, and the burn-in setting was 200. The goodness-of-fit tests (StatXact version 6.3; Cytel Software, Boston, MA) and hypothesis tests (Stata version 10.1; Stata Corp., College Station, TX) were 2 tailed.

Nucleotide sequence accession numbers. The cp26 sequence was assigned GenBank accession number GU569091. The sequence determined in the whole-genome shotgun project has been deposited in GenBank under accession number ADMY00000000. Detailed analysis of strain

CA8's chromosome will be presented elsewhere. Alignments not included in the supplemental material are posted as data sets at http://spiro.mmg.uci.edu/data/ospC.

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at http://mbio.asm.org/lookup/suppl/doi:10.1128/mBio.00153-10/-/DCSupplemental.

Table S1, DOC file, 0.076 MB.
Figure S1, PDF file, 0.038 MB.
Figure S2, PDF file, 0.101 MB.
Figure S3, PDF file, 0.089 MB.
Figure S4, PDF file, 0.107 MB.
Figure S5, PDF file, 0.079 MB.
Figure S6, PDF file, 0.369 MB.
Figure S7, PDF file, 0.080 MB.
Data Set S1, TXT file, 0.055 MB.
Data Set S2, TXT file, 0.255 MB.

ACKNOWLEDGMENTS

We thank Anne Gatewood Hoen, Maria Diuk-Wasser, and Durland Fish of Yale University, Yvette Girard and Robert Lane of University of California Berkeley, Sarah Hamer and Jean Tsao of Michigan State University, and Deborah Grosenbaugh of Merial Limited for providing specimens. We are grateful to Claire Fraser-Liggett, E. F. Mongodin, Sherwood Casjens, John Dunn, Ben Luft, Wei-Gang Qiu, and Steve Schutzer for providing public access to the whole-genome shotgun sequences of the *B. burgdorferi* strains.

This work was supported by Public Health Service grants AI-065359 from the National Institute of Allergy and Infectious Diseases and CI 00171-01 from the Centers for Disease Control and Prevention.

REFERENCES

1. **Steere, A. C., J. Coburn, and L. Glickstein.** 2005. Lyme borreliosis, p. 176–206. *In* J. L. Goodman, D. T. Dennis, and D. E. Sonenshine (ed.), Tick-borne diseases of humans. ASM Press, Washington, DC.
2. **Bunikis, J., J. Tsao, C. J. Luke, M. G. Luna, D. Fish, and A. G. Barbour.** 2004. *Borrelia burgdorferi* infection in a natural population of *Peromyscus leucopus* mice: a longitudinal study in an area where Lyme borreliosis is highly endemic. J. Infect. Dis. **189**:1515–1523.
3. **Seinost, G., W. T. Golde, B. W. Berger, J. J. Dunn, D. Qiu, D. S. Dunkin, D. E. Dykhuizen, B. J. Luft, and R. J. Dattwyler.** 1999. Infection with multiple strains of *Borrelia burgdorferi* sensu stricto in patients with Lyme disease. Arch. Dermatol. **135**:1329–1333.
4. **Bunikis, J., U. Garpmo, J. Tsao, J. Berglund, D. Fish, and A. G. Barbour.** 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.
5. **Girard, Y. A., B. Travinsky, A. Schotthoefer, N. Federova, R. J. Eisen, L. Eisen, A. G. Barbour, and R. S. Lane.** 2009. Population structure of the Lyme disease spirochete *Borrelia burgdorferi* in the western black-legged tick (*Ixodes pacificus*) in Northern California. Appl. Environ. Microbiol. **75**:7243–7252.
6. **Wang, I. N., D. E. Dykhuizen, W. Qiu, J. J. Dunn, E. M. Bosler, and B. J. Luft.** 1999. Genetic diversity of ospC in a Local population of *Borrelia burgdorferi* sensu stricto. Genetics **151**:15–30.
7. **Bacon, R. M., K. J. Kugeler, and P. S. Mead.** 2008. Surveillance for Lyme disease—United States, 1992-2006. MMWR Surveill. Summ. **57**:1–9.
8. **Gatewood, A. G., K. A. Liebman, G. Vourc'h, J. Bunikis S. A. Hamer, R. Cortinas, F. Melton, P. Cislo, U. Kitron, J. Tsao, A. G. Barbour, D. Fish, and M. A. Diuk-Wasser.** 2009. Climate and tick seasonality predict *Borrelia burgdorferi* genotype distribution. Appl. Environ. Microbiol. **75**:2476–2483.
9. **Gatewood Hoen, A., G. Margos, S. J. Bent, M. A. Diuk-Wasser, A. G. Barbour, K. Kurtenbach, and D. Fish.** 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.
10. **Carter, C. J., S. Bergström, S. J. Norris, and A. G. Barbour.** 1994. A family of surface-exposed proteins of 20 kilodaltons in the genus *Borrelia*. Infect. Immun. **62**:2792–2799.
11. **Kumaran, D., S. Eswaramoorthy, B. J. Luft, S. Koide, J. J. Dunn, C. L. Lawson, and S. Swaminathan.** 2001. Crystal structure of outer surface protein C (OspC) from the Lyme disease spirochete, *Borrelia* burgdorferi. EMBO J. **20**:971–978.
12. **Lawson, C. L., B. H. Yung, A. G. Barbour, and W. R. Zuckert.** 2006. Crystal structure of neurotropism-associated variable surface protein 1 (Vsp1) of *Borrelia turicatae*. J. Bacteriol. **188**:4522–4530.
13. **Barbour, A. G., Q. Dai, B. I. Restrepo, H. G. Stoenner, and S. A. Frank.** 2006. Pathogen escape from host immunity by a genome program for antigenic variation. Proc. Natl. Acad. Sci. U. S. A. **103**:18290–18295.
14. **Dai, Q., B. I. Restrepo, S. F. Porcella, S. J. Raffel, T. G. Schwan, and A. G. Barbour.** 2006. Antigenic variation by *Borrelia hermsii* occurs through recombination between extragenic repetitive elements on linear plasmids. Mol. Microbiol. **60**:1329–1343.
15. **Sadziene, A., B. Wilske, M. S. Ferdows, and A. G. Barbour.** 1993. The cryptic *ospC* gene of *Borrelia burgdorferi* B31 is located on a circular plasmid. Infect. Immun. **61**:2192–2195.
16. **Stevenson, B., L. K. Bockenstedt, and S. W. Barthold.** 1994. Expression and gene sequence of outer surface protein C of *Borrelia burgdorferi* reisolated from chronically infected mice. Infect. Immun. **62**:-3568–3571.
17. **Qiu, W. G., S. E. Schutzer, J. F. Bruno, O. Attie, Y. Xu, J. J. Dunn, C. M. Fraser, S. R. Casjens, and B. J. Luft.** 2004. 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. **101**:14150–14155.
18. **Bockenstedt, L. K., E. Hodzic, S. Feng, K. W. Bourrel, A. de Silva, R. R. Montgomery, E. Fikrig, J. D. Radolf, and S. W. Barthold.** 1997. *Borrelia burgdorferi* strain-specific Osp C-mediated immunity in mice. Infect. Immun. **65**:4661–4667.
19. **Gilmore, R. D., Jr., K. J. Kappel, M. C. Dolan, T. R. Burkot, and B. J. Johnson.** 1996. Outer surface protein C (OspC), but not P39, is a protective immunogen against a tick-transmitted *Borrelia burgdorferi* challenge: evidence for a conformational protective epitope in OspC. Infect. Immun. **64**:2234–2239.
20. **Barthold, S. W., D. H. Persing, A. L. Armstrong, and R. A. Peeples.** 1991. Kinetics of *Borrelia burgdorferi* dissemination and evolution of disease after intradermal inoculation of mice. Am. J. Pathol. **139**:263–273.
21. **Crother, T. R., C. I. Champion, J. P. Whitelegge, R. Aguilera, X. Y. Wu, D. R. Blanco, J. N. Miller, and M. A. Lovett.** 2004. Temporal analysis of the antigenic composition of *Borrelia burgdorferi* during infection in rabbit skin. Infect. Immun. **72**:5063–5072.
22. **Dolan, M. C., J. Piesman, B. S. Schneider, M. Schriefer, K. Brandt, and N. S. Zeidner.** 2004. Comparison of disseminated and nondisseminated strains of *Borrelia burgdorferi* sensu stricto in mice naturally infected by tick bite. Infect. Immun. **72**:5262–5266.
23. **Attie, O., J. F. Bruno, Y. Xu, D. Qiu, B. J. Luft, and W. G. Qiu.** 2007. 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. **7**:1–12.
24. **Qiu, W. G., J. F. Bruno, W. D. McCaig, Y. Xu, I. Livey, M. E. Schriefer, and B. J. Luft.** 2008. Wide distribution of a high-virulence *Borrelia burgdorferi* clone in Europe and North America. Emerg. Infect. Dis. **14**:1097–1104.
25. **Travinsky, B., J. Bunikis, and A. G. Barbour.** 2010. Geographic differences in genetic locus linkages for *Borrelia* burgdorferi. Emerg. Infect. Dis. **16**:1147–1150.
26. **Stevenson, B., and J. C. Miller.** 2003. Intra- and interbacterial genetic exchange of Lyme disease spirochete erp genes generates sequence identity amidst diversity. J. Mol. Evol. **57**:309–324.
27. **Dykhuizen, D. E., and G. Baranton.** 2001. The implications of a low rate of horizontal transfer in Borrelia. Trends Microbiol. **9**:344–350.
28. **Earnhart, C. G., and R. T. Marconi.** 2007. OspC phylogenetic analyses support the feasibility of a broadly protective polyvalent chimeric Lyme disease vaccine. Clin. Vaccine Immunol. **14**:628–634.
29. **Livey, I., C. P. Gibbs, R. Schuster, and F. Dorner.** 1995. Evidence for lateral transfer and recombination in OspC variation in Lyme disease Borrelia. Mol. Microbiol. **18**:257–269.
30. **Hanincova, K., D. Liveris, S. Sandigursky, G. P. Wormser, and I. Schwartz.** 2008. *Borrelia burgdorferi* sensu stricto is clonal in patients with early Lyme borreliosis. Appl. Environ. Microbiol. **74**:5008–5014.
31. **Barbour, A. G., J. Bunikis, B. Travinsky, A. G. Hoen, M. A. Diuk-Wasser, D. Fish, and J. I. Tsao.** 2009. Niche partitioning of *Borrelia burgdorferi* and *Borrelia miyamotoi* in the same tick vector and mammalian reservoir species. Am. J. Trop. Med. Hyg. **81**:1120–1131.

32. **Chaconas, G.** 2005. Hairpin telomeres and genome plasticity in Borrelia: all mixed up in the end. Mol. Microbiol. **58**:625–635.
33. **Alverson, J., S. F. Bundle, C. D. Sohaskey, M. C. Lybecker, and D. S. Samuels.** 2003. Transcriptional regulation of the ospAB and ospC promoters from *Borrelia* burgdorferi. Mol. Microbiol. **48**:1665–1677.
34. **Xu, Q., K. McShan, and F. T. Liang.** 2008. Verification and dissection of the ospC operator by using flaB promoter as a reporter in *Borrelia burgdorferi*. Microb. Pathog. **45**:70–78.
35. **Margolis, N., D. Hogan, W. Cieplak, Jr., T. G. Schwan, and P. A. Rosa.** 1994. Homology between *Borrelia burgdorferi* OspC and members of the family of *Borrelia hermsii* variable major proteins. Gene **143**:105–110.
36. **Casjens, S., N. Palmer, R. van Vugt, W. M. Huang, B. Stevenson, P. Rosa, R. Lathigra, G. Sutton, J. Peterson, R. J. Dodson, D. Haft, E. Hickey, M. Gwinn, O. White, and C. M. Fraser.** 2000. 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. **35**:490–516.
37. **Stevenson, B., S. Casjens, and P. Rosa.** 1998. Evidence of past recombination events among the genes encoding the Erp antigens of *Borrelia burgdorferi*. Microbiology **144**:1869–1879.
38. **Jain, R., M. C. Rivera, and J. A. Lake.** 1999. Horizontal gene transfer among genomes: the complexity hypothesis. Proc. Natl. Acad. Sci. U. S. A. **96**:3801–3806.
39. **Shapiro, J.** 2010. Contested will: who wrote Shakespeare? Simon & Schuster, New York, NY.
40. **Qiu, W. G., D. E. Dykhuizen, M. S. Acosta, and B. J. Luft.** 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.
41. **Dykhuizen, D. E., D. Brisson, S. Sandigursky, G. P. Wormser, J. Nowakowski, R. B. Nadelman, and I. Schwartz.** 2008. The propensity of different *Borrelia burgdorferi* sensu stricto genotypes to cause disseminated infections in humans. Am. J. Trop. Med. Hyg. **78**:806–810.
42. **Seinost, G., D. E. Dykhuizen, R. J. Dattwyler, W. T. Golde, J. J. Dunn, I. N. Wang, G. P. Wormser, M. E. Schriefer, and B. J. Luft.** 1999. Four clones of *Borrelia burgdorferi* sensu stricto cause invasive infection in humans. Infect. Immun. **67**:3518–3524.
43. **Lagal, V., D. Portnoi, G. Faure, D. Postic, and G. Baranton.** 2006. *Borrelia burgdorferi* sensu stricto invasiveness is correlated with OspC-plasminogen affinity. Microbes Infect. **8**:645–652.
44. **Kurtenbach, K., S. De Michelis, S. Etti, S. M. Schafer, H. S. Sewell, V. Brade, and P. Kraiczy.** 2002. Host association of *Borrelia burgdorferi* sensu lato—the key role of host complement. Trends Microbiol. **10**: 74–79.
45. **Brisson, D., and D. E. Dykhuizen.** 2004. ospC diversity in *Borrelia burgdorferi*: different hosts are different niches. Genetics **168**:713–722.
46. **Jewett, M. W., R. Byram, A. Bestor, K. Tilly, K. Lawrence, M. N. Burtnick, F. Gherardini, and P. A. Rosa.** 2007. Genetic basis for retention of a critical virulence plasmid of *Borrelia burgdorferi*. Mol. Microbiol. **66**: 975–990.
47. **Swanson, K. I., and D. E. Norris.** 2008. Presence of multiple variants of *Borrelia burgdorferi* in the natural reservoir *Peromyscus leucopus* throughout a transmission season. Vector Borne Zoonotic Dis. **8**:397–405.
48. **Samuels, D. S., and C. F. Garon.** 1997. Oligonucleotide-mediated genetic transformation of *Borrelia burgdorferi*. Microbiology **143**:519–522.
49. **Dorward, D. W., and C. F. Garon.** 1990. DNA is packaged within membrane-derived vesicles of gram-negative but not gram-positive bacteria. Appl. Environ. Microbiol. **56**:1960–1962.
50. **Eggers, C. H., B. J. Kimmel, J. L. Bono, A. F. Elias, P. Rosa, and D. S. Samuels.** 2001. Transduction by phiBB-1, a bacteriophage of *Borrelia burgdorferi*. J. Bacteriol. **183**:4771–4778.
51. **Schulze, R. J., S. Chen, O. S. Kumru, and W. R. Zückert.** 2010. Translocation of *Borrelia burgdorferi* surface lipoprotein OspA through the outer membrane requires an unfolded conformation and can initiate at the C-terminus. Mol. Microbiol. **76**:1266–1278.
52. **Luke, C. J., K. Carner, X. Liang, and A. G. Barbour.** 1997. An OspA-based DNA vaccine protects mice against infection with *Borrelia burgdorferi*. J. Infect. Dis. **175**:91–97.
53. **Postic, D., N. M. Ras, R. S. Lane, M. Hendson, and G. Baranton.** 1998. Expanded diversity among Californian *Borrelia* isolates and description of *Borrelia bissettii* sp. nov. (formerly *Borrelia* group DN127). J. Clin. Microbiol. **36**:3497–3504.
54. **Margos, G., A. G. Gatewood, D. M. Aanensen, K. Hanincova, D. Terekhova, S. A. Vollmer, M. Cornet, J. Piesman, M. Donaghy, A. Bormane, M. A. Hurn, E. J. Feil, D. Fish, S. Casjens, G. P. Wormser, I. Schwartz, and K. Kurtenbach.** 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.
55. **Larkin, M. A., G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez, J. D. Thompson, T. J. Gibson, and D. G. Higgins.** 2007. Clustal W and Clustal X version 2.0. Bioinformatics **23**:2947–2948.
56. **Librado, P., and J. Rozas.** 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics **25**:1451–1452.
57. **Huelsenbeck, J. P., and F. Ronquist.** 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics **17**:754–755.
58. **Guindon, S., and O. Gascuel.** 2003. A simple, fast and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. **52**: 696–704.
59. **Huson, D. H., and D. Bryant.** 2006. Application of phylogenetic networks in evolutionary studies. Mol. Biol. Evol. **23**:254–267.
60. **Posada, D., and K. A. Crandall.** 2001. Selecting the best-fit model of nucleotide substitution. Syst. Biol. **50**:580–601.
61. **Martin, D. P.** 2009. Recombination detection and analysis using RDP3. Methods Mol. Biol. **537**:185–205.
62. **Gibbs, M. J., J. S. Armstrong, and A. J. Gibbs.** 2000. Sister-scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. Bioinformatics **16**:573–582.
63. **Bruen, T. C., H. Philippe, and D. Bryant.** 2006. A simple and robust statistical test for detecting the presence of recombination. Genetics **172**: 2665–2681.
64. **Guillot, G., A. Estoup, F. Mortier, and J. F. Cosson.** 2005. A spatial statistical model for landscape genetics. Genetics **170**:1261–1280.