Genetic Diversity of *ospC* in a Local Population of
Borrelia burgdorferi sensu stricto

**Ing-Nang Wang,* Daniel E. Dykhuizen,* Weigang Qiu,* John J. Dunn,†
Edward M. Bosler‡ and Benjamin J. Luft‡**

**Department of Ecology and Evolution, State University of New York, Stony Brook, New York 11794-5245, †Biology Department, Brookhaven
National Laboratory, Upton, New York 11973 and ‡Department of Medicine, State University of New York, Stony Brook,
New York 11794-8153*

ABSTRACT

The outer surface protein, OspC, is highly variable in *Borrelia burgdorferi sensu stricto*, the agent of Lyme disease. We have shown that even within a single population OspC is highly variable. The variation of *ospA* and *ospC* in the 40 infected deer ticks collected from a single site on Shelter Island, New York, was determined using PCR-SSCP. There is very strong apparent linkage disequilibrium between *ospA* and *ospC* alleles, even though they are located on separate plasmids. Thirteen discernible SSCP mobility classes for *ospC* were identified and the DNA sequence for each was determined. These sequences, combined with 40 GenBank sequences, allow us to define 19 major *ospC* groups. Sequences within a major *ospC* group are, on average, <1% different from each other, while sequences between major *ospC* groups are, on average, ∼20% different. The tick sample contains 11 major *ospC* groups, GenBank contains 16 groups, with 8 groups found in both samples. Thus, the *ospC* variation within a local population is almost as great as the variation of a similar-sized sample of the entire species. The Ewens-Watterson-Slatkin test of allele frequency showed significant deviation from the neutral expectation, indicating balancing selection for these major *ospC* groups. The variation represented by major *ospC* groups needs to be considered if the OspC protein is to be used as a serodiagnostic antigen or a vaccine.

LYME disease is the most important tick-borne disease in the United States (Barbour and Fish 1993). The causative agent is *Borrelia burgdorferi*, a spirochete, first isolated from Shelter Island, New York (Burgdorfer *et al.* 1982). Subsequent studies have subdivided this species into a species complex which includes *B. burgdorferi sensu stricto*, *B. afzelii*, *B. garinii*, *B. japonica* (Baranton *et al.* 1992; Kawabata *et al.* 1993), and several other groups (Postic *et al.* 1994; Balmelli and Piffaretti 1996; Valsangiacomo *et al.* 1997). The first three species are pathogenic and have been shown to be associated with different chronic symptoms (Assous *et al.* 1993; van Dam *et al.* 1993; Anthonissen *et al.* 1994; Balmelli and Piffaretti 1995). All three pathogenic species are found in Europe. In contrast, only *B. burgdorferi sensu stricto* is found in the United States.

In the eastern United States, the deer tick (*Ixodes scapularis*) is the major vector responsible for the transmission of the spirochetes from one animal to another. The tick has four main stages in its 2-yr life cycle: egg, larva, nymph, and adult (Wilson and Spielman 1985; Yuval and Spielman 1990). Typically, eggs are laid during June and July and hatch into larvae in August. The larvae feed on small mammals such as mice, voles, and chipmunks and develop into nymphs during August and September. The nymphs overwinter and feed on these small mammals and some larger mammals during the summer, developing into adults in August. The adults prefer larger mammals such as deer, horses, dogs, and raccoons. In either October or after overwintering, adults mate on a host and only the females feed. For a detailed model of tick population dynamics see Sandberg *et al.* (1992).

The spirochete population is maintained by an infection cycle from ticks to mammals and then back into younger ticks. Transovarial passage of the spirochete is considered insignificant in the life cycle of the bacteria because <1% of the unfed *I. scapularis* larvae caught in nature are infected (Piesman *et al.* 1986; Magnarelli *et al.* 1987). All ticks feed at least twice: once as larvae and again as nymphs, while females feed a third time as adults (Bosler 1993). A tick is infected either as a larva or as a nymph and can pass the infection on to another mammal only after molting to the next stage. The infection of young mice and other small mammals in late spring by infected nymphs and the infection of larvae in late summer by these small mammals seems to be the major transmission route of the spirochete (Spielman 1988; Spielman *et al.* 1993).

B. burgdorferi infection induces a strong humoral immune response to as many as 11 proteins (Craft *et*

Corresponding author: Daniel E. Dykhuizen, Department of Ecology and Evolution, State University of New York, Stony Brook, NY 11794-5245. E-mail: dandyk@life.bio.sunysb.edu

al. 1986). Some of these proteins are within the outer membrane (Luft *et al.* 1992) and are not exposed to immune selection even though they induce a strong immune response. Other proteins are part of the membrane (Cunningham *et al.* 1988) and might be exposed to immune selection. Of these outer surface proteins (Osp's), OspA and OspC, which are encoded by genes located on two separate plasmids (Barbour and Garon 1988; Marconi *et al.* 1993; Sadziene *et al.* 1993), have been the most studied. Experimental vaccines made against either of these Osp's are successful in warding off challenges from the same strain (Fikrig *et al.* 1990; Simon *et al.* 1991; Gilmore *et al.* 1996; Probert *et al.* 1997), but not necessarily from different strains (Fikrig *et al.* 1992; Golde *et al.* 1995; Probert *et al.* 1997).

The expression pattern of *ospA* and *ospC* is suggestive of their roles in the infection cycle of the spirochetes (Schwan *et al.* 1995; Stevenson *et al.* 1995). In unfed ticks, spirochetes express OspA but not OspC. However, when the tick starts feeding on mammals, OspC synthesis is induced and OspA repressed. The switch is in part due to the change in temperature; OspC is induced at 32–37°, but not at 24°. Evidence suggests coregulation of these two genes (Jonsson and Bergstrom 1995).

ospC is located on a 27-kb circular plasmid (Marconi *et al.* 1993; Sadziene *et al.* 1993), and encodes a lipoprotein of ∼22 to 23 kD in size (Padula *et al.* 1993). The expression level of OspC varies greatly in different strains (Theisen *et al.* 1993; Hofmeister and Childs 1995; Picken *et al.* 1995). It is well known that *ospC* is highly polymorphic (Theisen *et al.* 1993, 1995; Wilske *et al.* 1993; Livey *et al.* 1995). However, most of the information about variation in *ospC* is from strains arbitrarily collected in different places at different times. Little is known about the diversity in a small area, although there is some evidence suggesting that local populations are variable (Livey *et al.* 1995; Masuzawa *et al.* 1997).

There are practical implications to how diverse *ospC* is locally. If local diversity is low, serodiagnostic antigens and anti-OspC vaccines would be effective in the local area where they were developed. On the other hand, if local *ospC* diversity is high, similar to worldwide diversity, serodiagnostic antigens or vaccines might be more generally effective, perhaps worldwide.

To assess the genetic diversity of *ospC* in a local population, we used PCR-SSCP, a technique that allowed us to survey *ospA* variation from single ticks (Guttman *et al.* 1996). In this study, we show that the genetic diversity of *ospC* in a single population is nearly as high as that found worldwide. Furthermore, population genetic evidence suggested that there is very little plasmid transfer between spirochetes, again demonstrating the clonality of *B. burgdorferi* (Dykhuizen *et al.* 1993). A model of frequency-dependent selection on *ospC* by the host immune system is proposed to explain the maintenance of *ospC* variation.


```
ext(+):  5'-AAAGAATACATTAAGTGCGATATT-3'

ext(-):  5'-GGGCTTGTAAGCTCTTTAACTG-3'      (a)
         5'-TTAAGGTTTTTTTGGACTTTCTGC-3'    (b)

int(+):  5'-TTGTTAGCAGGAGCTTATGCAATATC-3'

int(-):  5'-CAATCCACTTAATTTTTGTGTTATTAG-3'  (a)
         5'-TAATGTGCTCAATTTTTCTATTATTAC-3'  (b)

1st PCR:  ext(+)/ext(-)(a)

2nd PCR:  front: ext(+)/int(-)(a or b)
          back:  int(+)/ext(-)

cloning:  ext(+)/ext(-)(a or b)
```
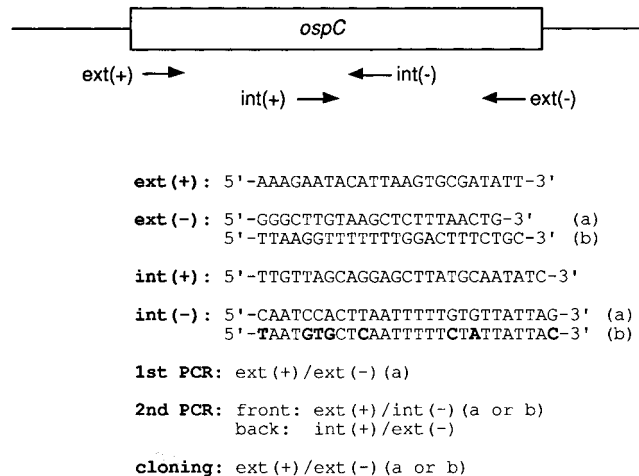
Figure 1.—Schematic drawing of the *ospC* regions that are subjected to PCR-SSCP analysis. The primer sequences and the primer pairs used for either the first round or the second round of PCR are also shown. Note that the boldface in primer int(−)(b) shows the eight sites where this primer is different from int(−)(a). For cloning of *ospC* genes, primer ext(+) and either primer ext(−)(a) or ext(−)(b) were used. However, only sequences between primers ext(+) and ext(−)(a) were used for sequence analyses.

MATERIALS AND METHODS

Extraction of DNA: Adult deer ticks (*I. scapularis*) collected in 1994 on Shelter Island, Long Island, New York, were bisected and DNA was extracted as described before (Guttman *et al.* 1996).

PCR amplifications: Two genes, *ospA* and *ospC*, both encoding outer surface lipoproteins in *B. burgdorferi*, were studied.

ospA amplification: Procedure for *ospA* PCR amplification was the same as described previously (Guttman *et al.* 1996). This amplification combined both nested and "touch-down" PCR. The first round of PCR was performed with low stringency and low cycling number using a pair of external primers to generate a heterogeneous population of amplified DNAs. The second round of PCR, using primer pairs internal to the previously amplified DNAs, was started out at very high stringency; as amplification proceeded the stringency was reduced gradually to increase yield.

ospC amplification: On the basis of our previous experience with *ospA* and preliminary experiments with *ospC*, the optimal size of the DNA fragment for SSCP analysis is ∼300 bp. A two-step seminested PCR amplification strategy was used for *ospC*. The first round of PCR used extracted DNA as template and two external primers for extension. The second round of PCR used the PCR products from the first round as template and either of two internal primers with the paired external primer. This divided the *ospC* into two roughly equal-sized fragments. Figure 1 shows the regions of *ospC* being amplified and the primer sequences. The two external primers are as follows: 5'-AAA GAA TAC ATT AAG TGC GAT ATT-3' (+) beginning at base 6; and 5'-GGG CTT GTA AGC TCT TTA ACT G-3' (−) ending at base 602. The front half of *ospC* was amplified using the external primer to the (+) strand and the primer 5'-CAA TCC ACT TAA TTT TTG TGT TAT TAG-3' (−) ending at base 345. The back half of *ospC* was amplified using the primer 5'-TTG TTA GCA GGA GCT TAT GCA ATA TC-3' (+) beginning at base 289 and the external primer to the (−) strand. The external primers amplified a 597-bp fragment. Amplifica-

tion of the front half produced a 340-bp fragment, while amplification of the back half produced a 314-bp piece. All the base numbers and amplified fragment sizes are based on *ospC* sequence of strain B31 (GenBank accession number U01894), with start codon as base 1.

The first round of PCR amplification used 10 μl of extracted DNA, dNTP at 0.2 mm per nucleotide, 1× PCR buffer [50 mm KCl, 10 mm Tris-HCl (pH 8.3)], 2.5 mm MgCl$_2$, 1 unit of native or ampli-*Taq* polymerase (Perkin Elmer-Cetus, Norwalk, CT), 0.5 μm of each external primer, and sterile HPLC water to a final volume of 50 μl. The reaction mixture was overlaid with two drops of mineral oil or without mineral oil when the top lid was heated (MJ Research, Watertown, MA). The temperature cycling profile of the first round of PCR was 1 min at 96° for 1 cycle linked to 40 sec at 95°, 35 sec at 54°, and 2 min at 72° for 20 cycles. For the second round of PCR, 0.5 μl of the first round PCR product was used as template and a pair of primers was added to amplify half of the gene. All the other components of reaction mixture were kept the same. The cycling profile was 1 min at 96° for 1 cycle linked to 40 sec at 95°, 35 sec at 54°, and 1 min at 72° for 35 cycles. To ensure that this two-step PCR strategy amplifies only *ospC* from *B. burgdorferi*, negative controls at various stages of amplification were performed for each PCR reaction. No nonspecific amplification was found.

Cold SSCP analysis: To survey genetic variation at both *ospA* and *ospC* genes, PCR-amplified fragments were subjected to cold SSCP analysis as described by Hongyo *et al.* (1993) and modified to suit our purpose (Guttman *et al.* 1996). The solution containing the PCR product, the amount depending on the DNA yield as estimated on an agarose gel, was reduced to 5 μl in a speed vac. The samples were then mixed with 0.4 μl of 1 m mercury hydroxide (Alfa Aesar, Ward Hill, MA) and 9.6 μl of loading buffer (Hongyo *et al.* 1993), heated to 95° for 4 min to denature the double-stranded DNA, and plunged into ice to prevent reannealing. The chilled samples were loaded onto a precast 20% TBE polyacrylamide gel (Novex, San Diego) with 1.5× TBE buffer. For both *ospA* and *ospC*, running condition and sample preparation were the same as described previously (Guttman *et al.* 1996), except that the *ospC* samples were run at a constant 240 V for 16.5 hr.

A potential major problem with PCR-SSCP for population genetics studies is that certain variants may not be amplified and included in the sample. This is not a problem here for the following reasons: (1) We previously have shown that the expected frequency of ticks with positive amplification of *ospA* is not significantly different from the frequency of infected ticks by visual observation (Guttman *et al.* 1996). (2) The *ospA* classes as distinguished by SSCP are all different by a single nucleotide from another class, showing there are no missing classes (Guttman *et al.* 1996). Thus we are sampling all the major variants of *ospA*. Any variant that we have not included would have to be rare. In this study, we amplified both *ospA* and *ospC*. (3) There were no ticks in which one gene amplified and not the other, giving us confidence that we have amplified all or nearly all the variants. (4) Both the front and back halves of *ospC* were amplified and analyzed by SSCP and there were no cases of extra or missing classes in this analysis. The analysis of the association of *ospA* alleles and *ospC* alleles allows the possibility of an occasional missing class, but these data can also be explained by recombination.

Cloning of *ospC* PCR fragments: We identified 13 different *ospC* mobility classes. Where possible, the cloning was done from ticks that were infected with a single clone of *B. burgdorferi* (see Table 1). The primer pair for cloning was either the same as that used for the first round of PCR amplification for *ospC* or the same (+) strand primer with a new (−) strand primer 5′-TTA AGG TTT TTT TGG ACT TTC TGC-3′ (−), which

ends at bp 633. This later primer pair gives a 627-bp fragment. The PCR products were separated on a 2% agarose gel and eluted from an agarose block using the Prep-A-Gene kit (Bio-Rad, Hercules, CA). Eluted fragments were cloned into pGEM-T cloning vector (Promega, Madison, WI). Each *ospC* mobility class was cloned from two different infected ticks. Before sequencing, the identity of each cloned gene was confirmed using PCR-SSCP analysis.

DNA sequencing: Plasmids containing cloned PCR fragments were purified using the high pure plasmid isolation kit (Boehringer Mannheim, Indianapolis) following manufacturer's instructions and cycle-sequenced using FS-TaqDyeDeoxy terminator chemistry and commercially available primers flanking the T-vector cloning sites. Samples were precipitated with ethanol and analyzed on an automated sequencer (Perkin Elmer ABI 373S). Data assembly and editing were done using Sequencher software (Gene Codes, Ann Arbor, MI).

Sequence analyses: Clustal W (1.4), running on a Sun Sparc workstation, was used for primary sequence alignment. Minor adjustment of the output alignment was performed by hand. MEGA (Kumar *et al.* 1993) was then used for further sequence analyses, including estimation of the ratio of synonymous *vs.* nonsynonymous substitutions per site (d_S / d_N) and the significance testing of this ratio. When sequences were compared, the sequence of the external primers was removed. Thus all sequence comparisons, including sequences from GenBank, start at base 31 and end at 579 (numbers based on the *ospC* sequence of strain B31 with the start codon as base 1).

An Ewens-Watterson-Slatkin test on allele frequency distribution was performed by Arlequin (Schneider *et al.* 1996), a software package for population genetics. There are two proposed tests for neutrality based on Ewens sampling theory (Ewens 1972): one is the homozygosity test by Watterson (1977), and the other is an exact test by Slatkin (1994, 1996). For an excellent description of the Ewens sampling theory and the mechanics of generating sample configurations see Chapter 3 of Hartl and Clark (1989). For large sample sizes, both tests are required to generate a specified number of random configurations. These configurations are contingent upon the total number of sampled genes (*n*) and the number of alleles (*k*) in the sample. Each random configuration was then calculated for its homozygosity *F* (for Watterson test) and for its exact probability (for Slatkin test). Each calculated value was compared with the value from the observed sample configuration. The test statistic is defined as the proportion of the random configurations with values equal to or smaller than the observed values. The null hypothesis (that the observed sample configuration was drawn from a neutral allele frequency distribution) is rejected if the test probability is $< s/2$, or $> 1 − s/2$, where *s* is the prespecified significance level (a two-tailed test). When the sample size is small, both tests lead to similar probabilities. However, the differences could be substantial when the sample size is relatively large (Slatkin 1996). We performed both tests.

RESULTS

ospC **mobility classes:** In a previous study from the same tick population we showed that there were only four *ospA* mobility classes, which we named MC1 to MC4 (Guttman *et al.* 1996). In this study we found only these same four *ospA* mobility classes (here we rename them OA1 to OA4). Cold SSCP analysis of *ospC* revealed more genetic variation than in *ospA* (Figure 2). Various *ospC* mobility classes were first identified by studying the variation in the front half of the gene. Of the 40 ticks

TABLE 1

Four *ospA* and *ospC* SSCP mobility classes found in each tick

Tick no.	OA mobility class	OC mobility class	Tick no.	OA mobility class	OC mobility class
1	1	1[a]	21	1/3/4	4/9[b]/12
2	4	2	22	1/2	1/3
3	4	2[a]	23	4	6
4	1	1, 4	24	3	13[a]
5	1/4/[c]	1, 4/6/5	25	1/4	1/2
6	3/4	9[b]/2	26	1/2	4/3
7	4	2	27	3	9[a,b]
8	2/4	3/2	28	1/2	1/3
9	1/2/4	1/3/2 (11)[a,b,d]	29	2	5[a]
10	3/4	13[a]/2	30	1/2/[c]	1, 4/[c]/[a]2
11	1	1[a]	31	[c]/4	9[b]/8[a], 10[a]
12	4	6[a]	32	1/4	1/6 (11)[a,b,d]
13	2/4	5[a]/8[a] (7)[a,d]	33	2	3[a]
14	1	4	34	1/4	4/ 2, 6, 8
15	2	3[a]	35	1	1
16	2/3	3, 8/ 9[b], 12	36	4	2
17	1	4[a]	37	1/2	1/3
18	1/2/3	[c]/3, 8/ 12	38	4	6[a]
19	3	9[b]	39	3	9[a,b]
20	1/3	4[a]/ 12	40	2	3

Slashes separate different OAs and corresponding OCs.
[a] Alleles being cloned and sequenced in this study. Both OC12 alleles were cloned from another population.
[b] PCR amplification using redesigned internal primer, int(−)(b) (see Figure 1 for sequence).
[c] Possible missing mobility classes for either *ospA* or *ospC*, given the linkage association determined from this data (see text).
[d] Numbers in parentheses denote association status not certain (see text and Table 3).

positive for *ospA*, only 38 were positive for *ospC*, while all 22 ticks that were negative for *ospA* were negative for *ospC*. We were able to amplify the *ospC* genes from the other two *ospA*-positive ticks using other primers (see below).

Ticks containing similar mobility classes were run side by side to confirm their identity. Among the ticks containing identical mobility classes, two independent *ospC* genes were PCR-amplified and cloned to serve as mobility class standards. This process was repeated several times until we obtained all of the *ospC* mobility classes present in our tick sample. For the two ticks in which the front half could not be amplified, the entire *ospC* gene was amplified. These amplified genes were cloned. Sequencing showed that the internal primer sequences contained eight mismatches (see Figure 1). Using a redesigned internal primer, two more mobility classes (9 and 11) were uncovered by screening all 40 *ospA*-positive ticks. In the end, we obtained 13 different *ospC* mobility classes (OCs) as shown in Figure 2A.

The mobility classes for the back half were also established by the same methodology as the classes for the front half of the gene. There were 13 classes and each class for the back half corresponded to a class for the front half (Figure 2B). The back half of the gene corroborated the results from the front half of the gene. Table 1 shows the results of all the SSCP mobility classes of both *ospA* and *ospC* found in each of the 40 positive ticks. The implications of this table are discussed below.

Multiple infections: Out of 40 *ospA*-positive ticks, 18 (45%) of them were multiply infected with two or more different *ospA* alleles, and the average number of strains per infected tick is 1.53 (61/40). For *ospC*, 50% were multiply infected, and the average is 1.88 (76/40) strains per infected tick. Because the number of distinguishable *ospC* alleles is over three times the number of *ospA* alleles, the results from the *ospC* survey provide a more accurate means to determine multiple infection rates and the average number of strains per infected tick. Because both of the average number of strains infecting a tick are relatively low (<2), we can infer the allele frequencies from the frequencies of the SSCP bands (Qiu *et al.* 1997; W. Qiu, unpublished results).

Allele frequencies: The number of ticks infected with each *ospC* mobility class and the estimated allele frequencies are given in Table 2. The distribution of allele frequencies is more even than expected for a neutral model, suggesting the action of balancing selection (see *Ewens-Watterson-Slatkin test*, below).

Linkage: A close inspection of Table 1 shows that the *ospA* and *ospC* mobility classes are strongly associated even though they are on different plasmids. This associa-
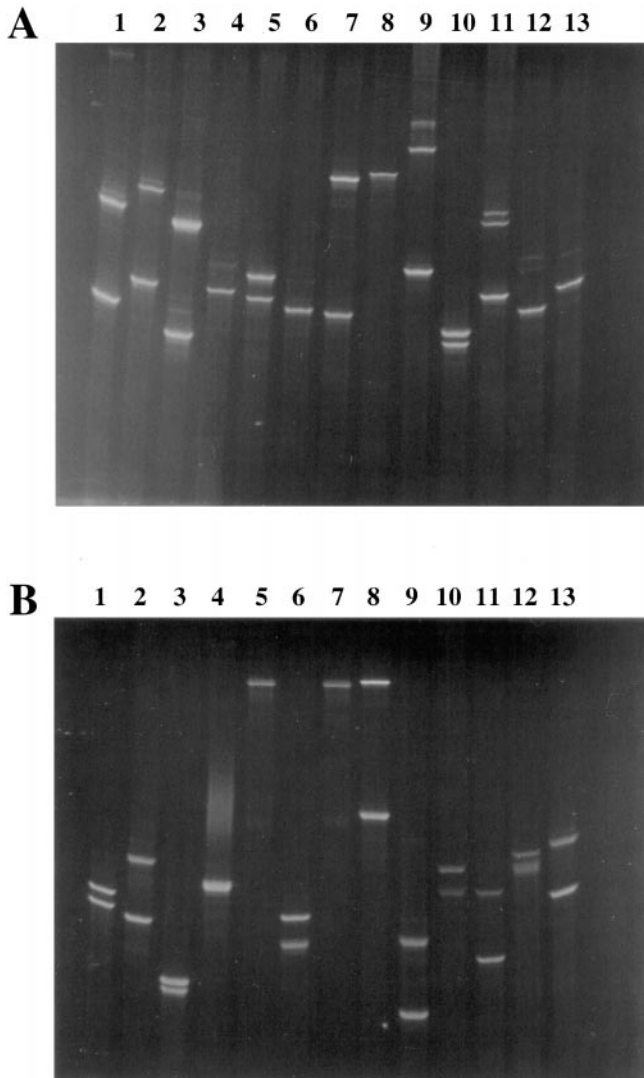
Figure 2.—PCR-SSCP of 13 cloned *ospC* mobility classes. (A) The front half of the gene. (B) The back half of the gene. Lanes 1–13 correspond to OC1–OC13 in both A and B. The exact banding pattern on a SSCP gel is unpredictable. The most frequent pattern is two bands, with each band corresponding to a single-strand DNA of the double-stranded PCR product. However, sometimes both bands comigrate to give a single band.

tion is functionally like chromosomal linkage, so we refer to this association as linkage. The linkage relationships are summarized in Table 3. For the great majority of infected ticks in our sample, OA1 is always associated with OC1, OC4, and possibly OC11; OA2 with OC3, OC5, OC8, and possibly OC7; OA3 with OC9, OC12, and OC13; and OA4 with OC2, OC6, OC8, OC10, and possibly OC7 and OC11. Since both OC7 and OC10 appeared only once in the sample, their association is uncertain. However, based on surveys from other populations, we are able to assign OC10 to OA4 (I.-N. Wang, unpublished data). OC11 cannot be assigned to a specific OA with certainty because in both instances (tick numbers 9 and 32 in Table 1) it co-occurred with OA1

TABLE 2

Number of ticks found carrying a specific *ospC* mobility class

OCs	No. of ticks	Estimated frequency
1	12	0.162
2	12	0.162
3	11	0.149
4	9	0.122
5	3	0.041
6	6	0.081
7	1	0.014
8	5	0.068
9	7	0.095
10	1	0.014
11	2	0.027
12	3	0.041
13	2	0.027

and OA4. An expanded survey should be able to resolve its linkage status. The appearance of OC8 with both OA2 and OA4 is not an artifact, because the same association pattern was also observed in tick populations from other localities (I.-N. Wang, unpublished data). In addition, OC12 can be assigned to OA3 because strains isolated from other ticks from this population show this linkage relationship.

The linkage relationships listed in Table 3 suggest there are a few instances of missing mobility class or violation of these linkage relationships. These are noted by *c* in Table 1. The presence of missing mobility classes can be explained three ways: (1) Our PCR-SSCP analysis did not uncover all the variation within a single tick because of chance variation in the amplification; (2) the missing mobility classes could not be amplified because of mutational changes, most likely deletions, in the gene; (3) the linkage relationship has changed. The first explanation can be eliminated. Another aliquot was taken from the DNA sample of all infected ticks and the PCR-SSCP analysis was redone. We obtained the same results; no classes appeared or disappeared. We cannot distinguish between the other two explanations from these data. However, tick 30 (Table 1) with the

TABLE 3

Linkage relationship between *ospA* and *ospC* mobility classes

OA mobility class	OC mobility class
1	1, 4, (11)[a]
2	3, 5, 8, (7)[a]
3	9, 12, 13
4	2, 6, 8, 10, (7)[a], (11)[a]

[a] Numbers in parentheses denote linkage status not certain yet.

pair of missing classes is likely to be a case in which the linkage relationship changed: OA2 was transferred into a background carrying OC2. Deletions joining *ospA* and *ospB* together have been found in nature (Rosa *et al.* 1992). If this has happened in one or more of the strains in our sample, the *ospA* allele like this would not be amplified and consequently be counted as missing. In any case, the overall association or linkage between *ospA* and *ospC* is surprisingly high, given that these two genes are on different plasmids.

ospC sequences: To confirm that there are 13 different *ospC* alleles and only 13 in this sample, 22 genes, as well as 2 genes from other cultures with the same *ospC* mobility class, were cloned and sequenced. For the 11 different *ospC* mobility classes where 2 different genes of the same class were cloned and sequenced, the sequences were identical (data not shown). Each mobility class has a different sequence. Figure 3 shows the aligned DNA sequences and the deduced amino acid sequences.

Population variation compared to species variation in *ospC*: While it was known that *ospC* is very polymorphic (Theisen *et al.* 1993; Jauris-Heipke *et al.* 1995; Livey *et al.* 1995), it was not known whether this *ospC* variation is found within a single population or only across the species. To create a sample of the variability of the entire species to compare to our sample from a local population, we used 9 different sequences of *B. burgdorferi sensu stricto* that were found using PCR-RFLP analysis (Livey *et al.* 1995) and 4 different sequences from GenBank, MIL (accession number U91802), CA-11.2A (accession number L25413), CA4 (accession number L81131), and 272 (accession number X84785) to give a sample of 13 different sequences. Of these 13 strains, 3 are from Connecticut, 3 from California, 2 from elsewhere in the United States (probably Minnesota), and 1 strain each from France, Germany, Slovakia, Pennsylvania, and Maryland. Out of 567 nucleotide sites compared, 244 (43%) of them are polymorphic (gaps in the aligned sequences are omitted from all the analyses) for the local sample compared to 259 (46%) for the sample from the entire species. For the local population, the number of amino acid polymorphic sites is 97 out of 189 residues, the same for the species sample. Clearly, most of the diversity found within the species is found within the local population. Masuzawa *et al.* (1997) reached a similar conclusion with a much smaller sample size.

Distribution of variation: A sliding-window plot of our 13 sequences, as shown in Figure 4, indicates that the polymorphic sites are not evenly distributed along the *ospC* sequence. The general trend is that the proportion of polymorphic sites increases as we move from the 5′ end to the 3′ end, with most of the variation clustered in the last two-thirds of the protein sequence. We used five relatively low points to divide the gene into six

regions. There are two prominent peaks, one each in regions 3 and 5, which might be antigenically important.

Major *ospC* groups: The distribution of pairwise differences of DNA sequence among our 13 sequences is bimodal. Most of the pairs are quite different, ranging from 9.7% (OC9 *vs.* OC11) to 22.3% (OC8 *vs.* OC10), with an average of 17.5%. Several pairs are quite similar with only 0.7% (OC5 *vs.* OC7) and 1.8% (OC12 *vs.* OC13) of their nucleotides different.

A similar pattern also emerges from worldwide surveys. As of September 1997, there were 40 *ospC* sequences from *B. burgdorferi sensu stricto* in the GenBank database. We have excluded the *ospC* sequence from strain KIPP (accession number X84782) from this analysis. When KIPP is aligned with the other sequences, its differences are scattered throughout the sequence and frequently at positions that are monomorphic in all other sequences, indicating numerous sequencing errors. The distribution of pairwise percentage differences for the remaining 39 sequences is also bimodal.

The same pattern again emerges when the two sets of sequences are combined. This bimodal pattern is clearly illustrated in Figure 5. From this we can define major *ospC* groups (Table 4). Members of the same group will have sequence differences <2%, while members of different groups will differ by >8%. The major *ospC* groups should not be confused with SSCP mobility classes. For example, within our 13 mobility classes, OC5 and OC7 are from the same major *ospC* group, as are OC12 and OC13. While the RFLP groups as defined by Livey *et al.* (1995) are very similar to our major *ospC* groups, they are not identical. The *ospC* sequence from strain BUR (GenBank accession number X84765) is missing a *Dra*I site found in RFLP group 4, so BUR would not be placed in this RFLP group, even though sequence similarity reveals that it belongs to the same major *ospC* group as the other RFLP group 4 strains.

The 53 *ospC* sequences (from GenBank and our current study) fall into 19 major *ospC* groups (Table 4). Ten of our 13 OCs belong to groups that include strains previously sequenced. The remaining 3 (OC3, OC8, and OC9) are new. Also, there are 8 major *ospC* groups that we did not find in our sample. The data allow three possibilities: (1) Each local population contains the entire variation found within the species—the remaining 8 major *ospC* groups would be found with more extensive sampling of the population; (2) each population contains endemics—some groups have a wide range and some groups are found only in one geographic area; (3) each population contains a subset of all groups found in the species—each group would be found across a wide geographical range, but because of the process of extinction and recolonization, only some of the groups are found in any population at any one time. More extensive sampling within the population and across populations could determine which possibility dominates.

A

```
                                                                                             111 111 111 111 111
             333 333 333 444 444 444 455 555 555 556 666 666 666 777 777 777 788 888 888 889 999 999 999 000 000 000 011 111
             123 456 789 012 345 678 901 234 567 890 123 456 789 012 345 678 901 234 567 890 123 456 789 012 345 678 901 234
OC1   ATG ACT TTA TTT TTA TTT ATA TCT TGT AAT AAT TCA GGG AAA GAT GGG AAT ACA TCT GCA AAT TCT GCT GAT GAG TCT GTT AAA
OC2   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
OC3   ... ... ... ... ... ... ... ... ... ... ... ..A ... ... ... G.. ... ... ... ... ... ... ... ... ... ... ... ...
OC4   ... ... ... ... ... ... ... ... ... ... ... ..A ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
OC5   ... ... ... ... ... ... ... ... ... ... ... ..A ... ... ... G.. ... ... ... ... ... ... ... ... ... ... ... ...
OC7   ... ... ... ... ... ... ... ... ... ... ... A.A ... ... ... G.. ... A.. ... ... ..C ... ... ... ... ... ... ...
OC6   ... ... ... ... ... ... ... ... ... ... ... ..A ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
OC8   ... ... ... ... ... ... ... ... ... ... ... ..A ... ... ... G.. ... A.. ... ... ..C ... ... ... ... ... ... ...
OC9   ... ... ... ... ... ... ... ... ... ... ... ..A ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
OC10  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
OC11  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
OC12  ... ... ... ... ... ... ... ... ... ... ... ..A ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
OC13  ... ... ... ... ... ... ... ... ... ... ... ..A ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
```

```
            111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
            111 112 222 222 222 333 333 333 344 444 555 555 555 666 666 667 777 777 778 888 888 888 999 999 999 999 999
            567 890 123 456 789 012 345 678 901 234 567 890 123 456 789 012 345 678 901 234 567 890 123 456 789 012 345 678
OC1   GGG CCT AAT CTT ACA GAA ATA AGT AAA AAA ATT ACG GAT TCT AAT GCG GTT TTA CTT GCT GTG AAA GAG GTT GAA GCG TTG CTG
OC2   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
OC3   ... ... ... ... ... ... ... ... ..A ..A ... ..C ..A ... G.T ..C ..C ... ... ..A ... ..G A.C ..A ..T
OC4   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... .T. ... ...
OC5   ... ... ... ... ... ... ... ... ..A ..A ... ..C ..A ... G.T ..G ..C ... ... ..A ... ..G A.C ..A ..T
OC7   ... ... ... ... ... ... ... ... ..A ..A ... ..C ..A ... G.T ..G ..C ... ... ..A ... ..G A.C ..A ..T
OC6   ... ... ... ... ... ... ... ... ..A ..A ... ..C ..A ... G.T ..G ... ..A ... ..A A.. ... A.T ... ..T
OC8   ... ... ... ... G.. ... ... ... ..A ..A ... ... ..A ... G.. ..G ... ..A ... ..A ... .CG ... ... ..T
OC9   ... ... ... ... ... ... ... ... ..A ..A ... ..C ..A ... G.T ..G ... ..A ... ..A ... A.C ... ..T
OC10  ... ... ... ... ... ... ... ... ..A ..A ... ..C ..A ... G.T ..C ..C ... ... ..A ... ... A.T ... ..T
OC11  ... ... ... ... ... ... ... ... ..A ..A ... ..C ..A ... G.T ..G ... ... ..A A.. ... A.T ... ..T
OC12  ... ... ... ... ... ... ... ... ..A ..A ... ..C ..A ... G.T ..G ... ... ... ..A A.. ... A.T ... ..T
OC13  ... ... ... ... ... ... ... ... ..A ..A ... ..C ..C ... G.T ..G ... ... ... ..A A.. ... A.T ... ..T
```

```
            122 222 222 222 222 222 222 222 222 222 222 222 222 222 222 222 222 222 222 222 222 222 222 222 222 222 222 222
            900 000 000 001 111 111 111 222 222 222 233 333 333 334 444 444 444 555 555 555 566 666 666 667 777 777 777 888
            901 234 567 890 123 456 789 012 345 678 901 234 567 890 123 456 789 012 345 678 901 234 567 890 123 456 789 012
OC1   TCA TCT ATA GAT GAA ATT GCT GCT AAA GCT ATT GGT AAA AAA ATA CAC CAA AAT AAT GGT TTG GAT ACC GAA AAT AAT CAC AAT
OC2   ... ... ... ... ..G C.. ... --- ... ... ... ... ... ... ... A.A A.C G.. GG. A.. ..A ... .AT ... GCA ... .G. ..C
OC3   G.. ... ... ... ... C.. ... --- ... ... ... ... ... ... ... A.A A.C G.. GT. A.. ..A ... .AT ..G GCA G.. A.. ..C
OC4   ... ... ... ... ... C.. ... AAG ... ... ... ... ... ... ... G.T ... ..C ... .C. ..A .GC ..T CT. ... ... ..T ..C
OC5   G.. ... ... ... ... C.. ... A.C ... ... ... ... ... ... ... GG. --- ... ... ... ..A ..G G.. A.T C.G .G. A.A ..C
OC7   G.. ... ... ... ... C.. ... A.C ... ... ... ..G ... ... ... GG. --- ... ... ... ..A ..G G.. A.T C.G .G. A.A ..C
OC6   ... ... ... ... ... C.. ... A.. ... ... ... ... ... C.. ... G.T --- GCA ... ... ..A .GC GTT C.G GCG ... ..A ..C
OC8   ... ... ... ... ... C.. ... --- ... ... ... ... ... ... ... G.G --- C.A ... ... .G. G.T ..T GCG ... ... ..C
OC9   G.. ... ... A.C C.. C.. ... --- ... ... ... ... ... ... ... G.T ... ... GGC AC. ... .G. GA. ..T GGG GGC ..A ...
OC10  A.. ... ... ... ..G C.. ... --- ... ... ... ... ... ... ... A.A A.C G.. GT. A.. ..A ... .AT ..G GCA G.. ... ..C
OC11  G.. ... ... ... ... C.. ... A.. ... ... ... ... ... ... ... G.T A.C ... GC. ... ... .G. G.T ... GTG GG. ..A ..C
OC12  G.. ... ... ... ... C.. ... A.. ... ... ... ... ... ... ... ..A ... ... GG. ... ..A .C. GT. ... GCG GGG ..T ...
OC13  G.. ... G.. ... ... C.. ... A.. ... ... ... ... ... ... ... ..A ... ... GG. ... ..A .C. GT. ... GCG GGG ..T ...
```

```
            222 222 222 222 222 223 333 333 333 333 333 333 333 333 333 333 333 333 333 333 333 333        333 333 333 333
            888 888 899 999 999 990 000 000 000 111 111 111 122 222 222 223 333 333 333 444 444 444 455    555 555 556 666
            345 678 901 234 567 890 123 456 789 012 345 678 901 234 567 890 123 456 789 012 345 678 901    234 567 890 123
OC1   GGA TCA TTG TTA GCG GGA GCT TAT GCA ATA TCA ACC CTA ATA AAA CAA AAA TTA GAT GGA TTG AAA AAT --- GAA GGA TTA AAG
OC2   .AG ... ... ... ..A ... ... A.. ... ... ... T.. ... .C. ... ... ... AG. AA. ..A ..C GGA TCA ... .T ... ...
OC3   ... ... ..A A.. ..A ... ... A.. ... ... ... T.. ... .C. ... ... ... AG. AA. ..A ..C GGA TCA ... ..T ... ...
OC4   ... ... ... ..G ... ... ... ... ..T ... ... G.T ... ... .C. G.. ... ... AG. TC. A.A ... G.. TCA .G. .A. ..G ...
OC5   AC. ... ... ... T.A ... ... ... ... ..T GA. ... ... GC. G.. ... ... A.. .T. ... ... ... ... .A. ... ...
OC7   AC. ... ... ... T.A ... ... ... ... ..T GA. ... ... GC. G.. ... ... A.. .T. ... ... ... ... .A. ... ...
OC6   ... ... ... ..A ..T ..C ... ... ... ... ... .C. ... ... ... AG. .C. ..A --- ... TCA ... .AC ... ..A
OC8   ACC ... ... ..A ... ... C.. .A. ... ... ... ... ... ... ... T.. ... ... ... ... ... ... ... ..T
OC9   ... ... ... C.. ..A ... ... ... ... ..C ... ..T G.. ... .T. G.. ... ..G AGC AC. ..A ... ... GTA ... .A. ... ..A
OC10  ... ... ..A A.. T.A ... ..A ... TT. ..T ... .A. T.. ... .C. A.. ... A.. AG. .C. A.A ... G.. TCA .G. .A. ..G ...
OC11  ... ... ... C.. ..A ... ... ... ... ..C ... ..T G.. ... .T. G.. ... ..G AGC AC. ..A ... ... GTA ... .A. ... ..A
OC12  ... A.. ... ... ..A ..T ... A.. ... ... .AA ... .C. ... ... ... ... ... ... ... ... ... TCA ... AA. ... ...
OC13  ... A.. ... ... ..A ..T ... ..C A.. ... ... .AA ... .C. ... G.. ... ... ... ... ... ... TCA ... AA. ... ...
```

Figure 3.—DNA and amino acid sequences of the 13 *ospC* mobility classes. (A) Nucleotide sequences (including gaps). (B) Deduced amino acid sequences (including gaps) from A. The nucleotide and amino acid numbers are the numbers for strain B31, which is identical to OC1. The GenBank accession numbers for these sequences are given in Table 4.
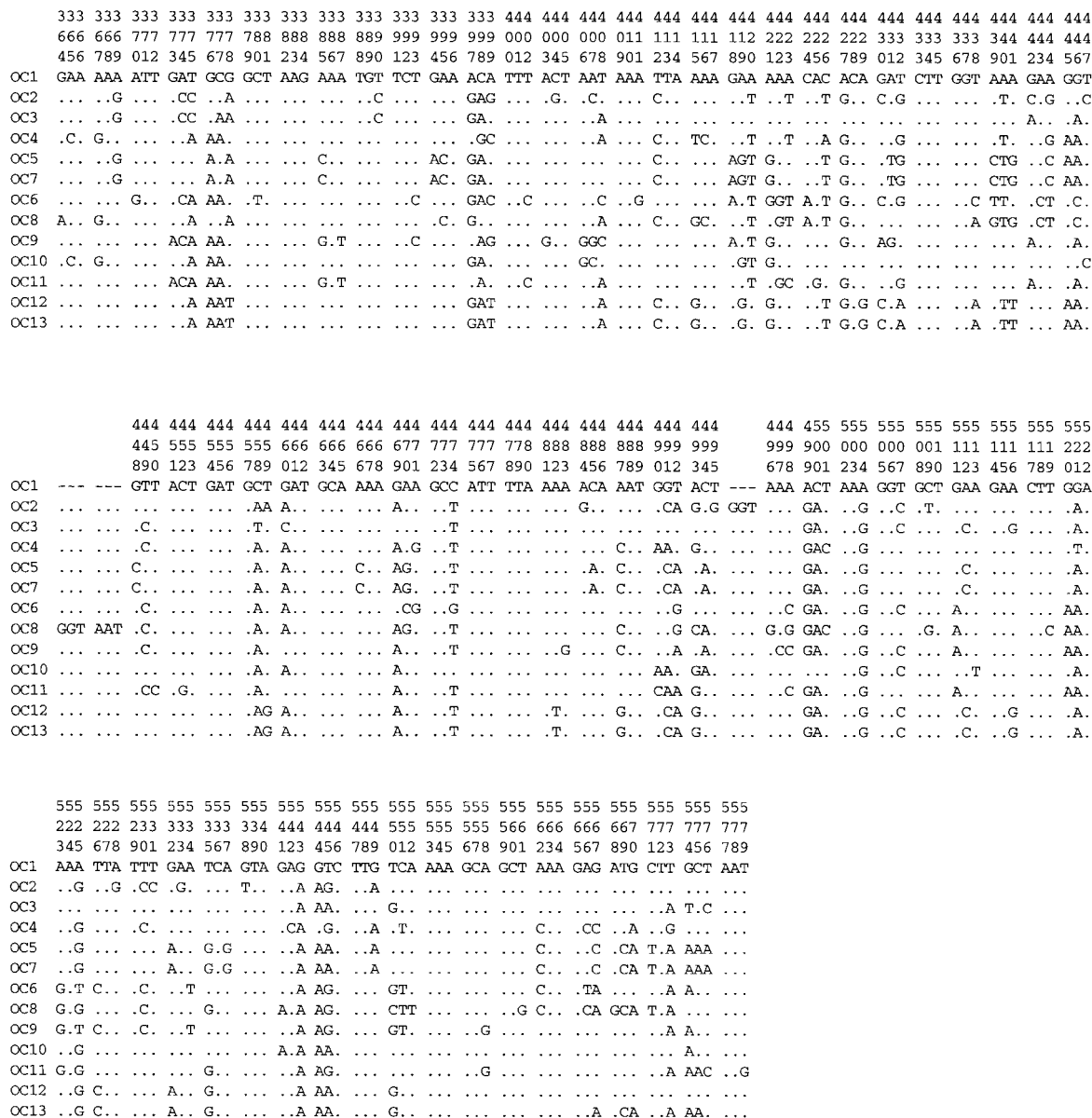
```
        333 333 333 333 333 333 333 333 333 333 333 333 444 444 444 444 444 444 444 444 444 444 444 444 444 444 444 444
        666 666 777 777 777 788 888 888 889 999 999 999 000 000 000 011 111 111 112 222 222 222 333 333 333 344 444 444
        456 789 012 345 678 901 234 567 890 123 456 789 012 345 678 901 234 567 890 123 456 789 012 345 678 901 234 567
OC1     GAA AAA ATT GAT GCG GCT AAG AAA TGT TCT GAA ACA TTT ACT AAT AAA TTA AAA GAA CAC ACA GAT CTT GGT AAA GAA GGT
OC2     ... ..G ... .CC ..A ... ... ... ..C ... ... GAG ... .G. .C. ... C.. ... ..T ..T ..T G.. C.G ... ... .T. C.G ..C
OC3     ... ..G ... .CC .AA ... ... ... ..C ... ... GA. ... ... ..A ... ... ..A ... ... ... ... ... ... ... A.. .A.
OC4     .C. G.. ... ..A AA. ... ... ... ... ... ... .GC ... ... ..A ... C.. TC. ..T ..T ..A G.. ..G ... ... .T. ..G AA.
OC5     ... ..G ... ... A.A ... ... C.. ... ... AC. GA. ... ... ... ... ... C.. ... AGT G.. ..T G.. .TG ... ... CTG ..C AA.
OC7     ... ..G ... ... A.A ... ... C.. ... ... AC. GA. ... ... ... ... ... C.. ... AGT G.. ..T G.. .TG ... ... CTG ..C AA.
OC6     ... ... G.. .CA AA. .T. ... ... ..C ... GAC ..C ... ..C ..G ... ... A.T GGT A.T G.. C.G ... ... ..C TT. .CT .C.
OC8     A.. G.. ... ..A ..A ... ... ... ..C G... ... ... ... ..A ... C.. GC. ..T .GT A.T G.. ... ... ..A GTG .CT .C.
OC9     ... ... ... ACA AA. ... ... G.T ... ..C ... .AG ... G.. GGC ... ... ... A.T G.. ... G.. AG. ... ... ... A.. .A.
OC10    .C. G.. ... ..A AA. ... ... ... ... ... ... GA. ... ... GC. ... ... ... .GT G.. ... ... ... ... ... ... ... ..C
OC11    ... ... ... ACA AA. ... ... G.T ... ... .A. ..C ... ..A ... ... ..T .GC .G. G.. ..G ... ... ... A.. .A.
OC12    ... ... ... ..A AAT ... ... ... ... ... ... GAT ... ... ..A ... C.. G.. .G. G.. ..T G.G C.A ... ..A .TT ... AA.
OC13    ... ... ... ..A AAT ... ... ... ... ... ... GAT ... ... ..A ... C.. G.. .G. G.. ..T G.G C.A ... ..A .TT ... AA.


        444 444 444 444 444 444 444 444 444 444 444 444 444 444 444 444      444 455 555 555 555 555 555 555 555
        445 555 555 555 666 666 666 677 777 777 778 888 888 888 999 999      999 900 000 000 001 111 111 111 222
        890 123 456 789 012 345 678 901 234 567 890 123 456 789 012 345      678 901 234 567 890 123 456 789 012
OC1     GTT ACT GAT GCT GAT GCA AAA GAA GCC ATT TTA AAA ACA AAT GGT ACT ---  AAA ACT AAA GGT GCT GAA GAA CTT GGA
OC2     ... ... ... ... .AA A.. ... ... A.. ..T ... ... G.. ... .CA G.G GGT  ... GA. ..G ..C .T. ... ... ... .A.
OC3     ... .C. ... ... .T. C.. ... ... ..T ... ... ... ... ... ... ... ...  GA. ..G ..C ... .C. ..G ... .A.
OC4     ... .C. ... ... ..A. A.. ... ..A.G ..T ... ... ... ... C.. AA. G..   ... GAC ..G ... ... ... ... ... ..T.
OC5     ... C.. ... ... .A. A.. ... C.. AG. ..T ... ... .A. C.. .CA .A. ...   GA. ..G ... ... ..C. ... ... .A.
OC7     ... C.. ... ... .A. A.. ... C.. AG. ..T ... ... .A. C.. .CA .A. ...   GA. ..G ... ... ..C. ... ... .A.
OC6     ... .C. ... ... .A. A.. ... ..CG ..G ... ... ... ... ..G ... ...     .C GA. ..G ..C ... A.. ... ... AA.
OC8     GGT AAT .C. ... ..A. A.. ... AG. ..T ... ... C.. ..G CA. ... G.G GAC  ..G ... .G. A.. ... ..C AA.
OC9     ... .C. ... ... .A. ... ... A.. ..T ... ...G ... C.. .A. .CC GA. ..G  ..C ... A.. ... ... AA.
OC10    ... ... ... ... .A. A.. ... A.. ... ... ... ... ... AA. GA. ... ...   ..G ..C ... ..T ... ... .A.
OC11    ... ... .CC .G. ... .A. ... ... A.. ..T ... ... ... ... CAA G.. ...   ..C GA. ..G ... A.. ... ... AA.
OC12    ... ... ... ... .AG A.. ... ... A.. ..T ... ... .T. ... G.. .CA G..   ... GA. ..G ..C ... .C. ..G ... .A.
OC13    ... ... ... ... .AG A.. ... ... A.. ..T ... ... .T. ... G.. .CA G..   ... GA. ..G ..C ... .C. ..G ... .A.


        555 555 555 555 555 555 555 555 555 555 555 555 555 555 555 555 555 555 555
        222 222 233 333 333 334 444 444 444 555 555 555 566 666 666 667 777 777 777
        345 678 901 234 567 890 123 456 789 012 345 678 901 234 567 890 123 456 789
OC1     AAA TTA TTT GAA TCA GTA GAG GTC TTG TCA AAA GCA GCT AAA GAG ATG CTT GCT AAT
OC2     ..G ..G .CC .G. ... T.. ..A AG. ..A ... ... ... ... ... ... ... ... ... ...
OC3     ... ... ... ... ..A AA. ... G.. ... ... ... ... ... ... ... ..A T.C ...
OC4     ..G ... .C. ... ... ... .CA .G. ..A .T. ... ... ... C.. .CC ..A ..G ... ...
OC5     ..G ... ... A.. G.G ... ..A AA. ..A ... ... ... C.. ..C .CA T.A AAA ...
OC7     ..G ... ... A.. G.G ... ..A AA. ..A ... ... ... C.. ..C .CA T.A AAA ...
OC6     G.T C.. .C. ..T ... ... ..A AG. ... GT. ... ... ... C.. .TA ... A A.. ...
OC8     G.G ... .C. ... G.. ... A.A AG. ... CTT ... ... ..G C.. .CA GCA T.A ... ...
OC9     G.T C.. .C. ..T ... ... ..A AG. ... GT. ... ..G ... ... ... ..A A.. ...
OC10    ..G ... ... ... ... ... A.A AA. ... ... ... ... ... ... ... ... A.. ...
OC11    G.G ... ... ... G.. ... ..A AG. ... ... ... ..G ... ... ... ..A AAC ..G
OC12    ..G C.. ... A.. G.. ... ..A AA. ... G.. ... ... ... ... ... ... ... ...
OC13    ..G C.. ... A.. G.. ... ..A AA. ... G.. ... ... ... ... ..A .CA ..A AA. ...
```

Figure 3A.—(*continued*)

Within-group pairwise differences: Table 4 also shows the average within-group pairwise percentage differences. Out of the 10 groups with two or more members, 9 have all their pairwise differences at 1% or less. However, for group K, the average difference is 1.35% (ranging from 0 to 2.98%). The strain that is most different from the others is OC13. Omission of this sequence from analysis drops the average within-group difference to 0.83%. Most of the differences between OC13 and the rest are found clustered at the very end of the sequence. The last 18 nucleotides of OC13, which include 6 of the 10 changes where OC13 is different from the other sequences within group K, are identical to the homologous sequence found in group M, suggesting that OC13 was generated from a recombination between group K and group M.

Between-group pairwise differences: There is a tail of low pairwise differences between major *ospC* groups (Figure 5). The two lowest points (9.0 and 9.8%) are the pairwise comparisons between major *ospC* groups H and J. For 116 bp between bases 253 and 368, which is one of the most polymorphic regions of the *ospC*, the sequences are identical, as if a short piece of DNA has been transferred from one to the other. Most of the remaining pairwise differences between groups that are <14.5% are comparisons between major *ospC* group C and groups I and B. Groups C and I are very similar over the first 255 bases that include the first major peak. Starting at base 263 and going to base 374, Groups C and B are identical except for one base. This covers the variable region between the two peaks (see Figure 4). Thus it looks as if major *ospC* group C is composed of

B

```
                                                                                      1 1111111111
           1111111112 2222222223 3333333334 4444444445 5555555556 6666666667 7777777778 8888888889 9999999990 0000000001
           1234567890 1234567890 1234567890 1234567890 1234567890 1234567890 1234567890 1234567890 1234567890 1234567890
OC1        MTLFLFISCN NSGKDGNTSA NSADESVKGP NLTEISKKIT DSNAVLLAVK EVEALLSSID EIAAKAIGKK IHQNNGLDTE NNHNGSLLAG AYAISTLIKQ
OC2        .......... .......... .......... .......... .......... .L.-...... .KNDGS..N. A.R.E..... ..T.....T.
OC3        .......... .......A.. .......... .......... E....V.... ...T..A... .L.-...... .KNDVS..N. ADN....I.. ..T.....T.
OC4        .......... .......... .......... .......... ...V..... .L.K...... .D...A.G.L .......... .....A..TE
OC5        .......... .......A.. .......... .......... E....V.... ...T..A... .L.T...... .G-....EAN QSK.T...S. .....D..AE
OC6        .......... .......... .......... .......... E....V.... .I.T...... .L.T....Q. .D-A...GVQ A.Q....... ........T.
OC7        .......... ..R....A.T .......... ..A....... E....V.... ...T..A... .L.T...... .G-....EAN QSK.T...S. .....D..AE
OC8        .......... .......A.T .......... ..A....... E....V.... ..A....... .L.-...... .E-Q...GAD A...T..... .HE.......
OC9        .......... .......... .......... .......... E....V.... ...T..A..N QL.-...... .D..GT.GDD GGQ....... .....V.IE
OC10       .......... .......... .......... .......... E....V.... ...T..T... .L.-...... .KNDVS..N. AD.....IS. ..L..N..TK
OC11       .......... .......... .......... .......... E....V.... .I.T..A... .L.T...... .DN.A..GA. VGQ....... .....V.IE
OC12       .......... .......... .......... .......... E....V.... .I.T..A... .L.T...... .Q..G..AV. AG...T.... ..T..K..T.
OC13       .......... .......... .......... .......... E....V.... .I.T..A.V. .L.T...... .Q..G..AV. AG...T.... ..T..K..T.
```

```
           1111111 11 1111111111 1111111111 1111111111   11111111 11111111 1 1111111111 1111111111 1111111
           1111111 11 2222222222 3333333333 4444444444   55555555 55666666 6 6667777777 7778888888 8889999
           1234567 89 0123456789 0123456789 0123456789    01234567 89012345 6 7890123456 7890123456 7890123
OC1        KLDGLKN-EG LKEKIDAAKK CSETFTNKLK EKHTDLGKEG --VTDADAKE AILKTNGT-K TKGAEELGKL FESVEVLSKA AKEMLAN
OC2        ..SK.NGS.. .....A.... ...E.ST... DN.AQ..IQ. .....EN..K ....A.AAG. D..V...E.. SG.L.S.... .......
OC3        ..SK.NGS.. .....AE... ...E..K... ........KD .A..VH... .......... D...A..E.. .....N.A.. .....S.
OC4        ..SSI.DSGE ..AE.EK... ...S..K..S DNQAE..I.N ..A..DN..K .....HNA.. D......V.. S...AG.L.. .QAI...
OC5        ..NV.....E ......T..Q ..TE...... SE.AV..LDN ..L..DN.QR ....KHAN.. D...A..E.. .KA..N.... .QDT.K.
OC6        ..SA.-.S.D ....VAKV.. ...D...... NGNAQ..LAA ..A..DN..A .........N D...K..KD. SD...S.V.. .QV..T.
OC7        ..NV.....E ......T..Q ..TE...... SE.AV..LDN ..L..DN.QR ....KHAN.. D...A..E.. .KA..N.... .QDT.K.
OC8        ..Y....... .NKE.E.... ..AA..K..A DSNA...VAA GNA..DN..R .....H.H.E D..GK..KE. S.A.KS.L.. .QAA...
OC9        ..ST...V.E .....TK..D ...K.AG... NE.AS...KD ..A..D...K .....H.N.T D...K..KD. SD...S.V.. .....T.
OC10       .ISAI.DSGE ..AE.EK... ...E..A... GE........ .....DN..K ......ND.. ...D..E... ....KN.... .....T.
OC11       ..ST...V.E .....TK..D ...K..K... DSRAE...KD ..AS.D...K ......QA.N D...K..KE. ..A..S.... .....NK
OC12       .......S.K .....EN... ...D..K..E GE.AQ..I.N .....EN..K ...I.DAA.. D...A..E.. .KA..N.A.. .......
OC13       E......S.K .....EN... ...D..K..E GE.AQ..I.N .....EN..K ...I.DAA.. D...A..E.. .KA..N.A.. ...T.N.
```

Figure 3.—(*continued*).

three pieces, one from group I, another from group B, and a third from a group yet to be described. Major *ospC* group C, while common in our sampled population, has been found only on Shelter Island. Consequently, this may be a local endemic, created by recombination and likely to become extinct because of immunological overlap with other groups. None of the pairwise comparisons with differences >14.5% show any extended similarities.

Ewens-Watterson-Slatkin test: On the basis of Ewens sampling theory (Ewens 1972), Watterson (1977) proposed a test to see if an observed allele frequency distribution conforms to neutral expectation. The null hypothesis is that the distribution of allele frequencies is compatible with the neutral hypothesis; *i.e.*, mutation and genetic drift alone explain the distribution of allele frequencies when a Wright-Fisher population equilibrium model is used to determine population structure. The test statistic is F, the probability that two genes chosen at random will be the same allele. For the test, F is conditioned on the number of alleles found in the sample. If F is significantly different from the expected value (the null hypothesis), then selection is important or the population structure differs significantly from the Wright-Fisher model. A significant value of F larger than the expected value implies purifying selection. On the other hand, a significantly smaller F implies balancing selection (Hartl and Clark 1989, Chap. 3). Table 5 shows that both for the alleles as the 13 mobility classes or as the 11 major *ospC* groups, the value of F is
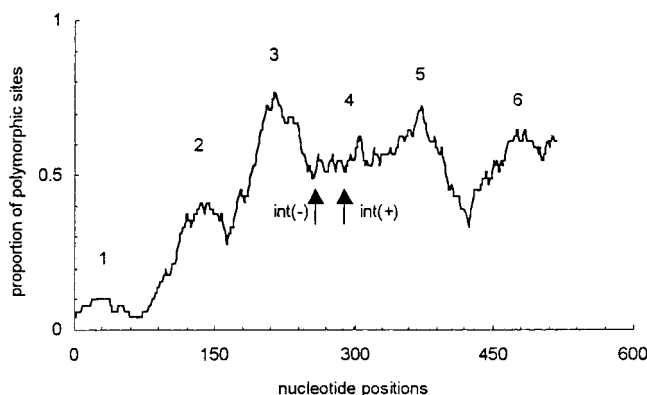
Figure 4.—Sliding-window representation of nucleotide polymorphism of the 13 *ospC* mobility classes. The DNA sequences in Figure 3 are used. A window size is 51 nucleotides (17 amino acids), which is the optimal size as determined by Tajima's method (Tajima 1991). The six regions (1–6) are defined by localizing regions of relatively low variability. The last nucleotide of this 51-nucleotide window is then specified as the border nucleotide. The five border nucleotides that define these regions are 149, 243, 336, 417, and 519, where the numbers are according to the B31 sequence numbering. The locations of two internal primer sequences used in PCR-SSCP analysis, int(−) and int(+), are indicated by broad-headed arrows.
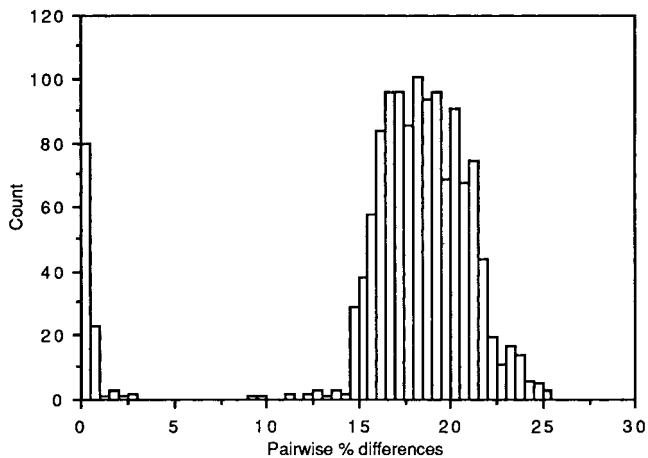
Figure 5.—Distribution of percentage pairwise differences among 52 *ospC* sequences. Thirty-nine *ospC* sequences from GenBank together with 13 sequences from this study were used. The class interval is 0.5%.

significantly lower than expected given neutrality, indicating balancing selection.

Slatkin (1994, 1996) proposed an exact test also on the basis of the same sampling theory. The test provides the probability of obtaining a configuration of alleles with the same or smaller probability as the observed configuration using the Ewens sampling distribution of neutral alleles. This is different from the Watterson test because the Slatkin test does not incorporate a model of selection to test deviation from the distribution of alleles expected under the null hypothesis of neutrality. On the basis of Slatkin's test, the frequency of the 11 major *ospC* groups is significantly different from that of random expectation, but the frequency of the 13 mobility classes is not significantly different from neutrality (Table 5). The 10-fold lower significance levels for the major *ospC* groups compared to the mobility classes arise because 2 of the rarer groups are divided to generate the mobility classes. We conclude from these tests that the frequencies of the major *ospC* groups are not compatible with the neutral model and that the deviation indicates balancing selection.

Balancing selection: The results presented so far strongly suggest that the major *ospC* groups are under strong balancing selection. The ratio of synonymous *vs.* nonsynonymous substitutions per site (d_S/d_N) can be used to investigate various modes of selection (Hughes and Nei 1988; Hughes 1991, 1992; Hughes *et al.* 1993; Riley 1993; Caporale and Kocher 1994; Hughes and Hughes 1995; Seibert *et al.* 1995; Wagner and Riley 1996). It is constructed such that if there is no selection on the gene, as in a pseudogene, the value of this ratio is expected to be 1. For most functional proteins, the ratio is higher, usually ~5, because amino acid sequences are constrained by selection (Li and Graur 1991). A ratio <1 implies positive Darwinian selection for diversity at the amino-acid level. As the Ewens-Watt-

erson-Slatkin test strongly indicates that *ospC* is selected to be diverse by balancing selection, we might expect the d_S/d_N ratio to be <1. To estimate the ratio, we used the same 13 sequences as we used for the sliding window. As shown in Table 6, the synonymous substitutions per site ($d_S = 0.268$) are slightly more than the nonsynonymous substitutions per site ($d_N = 0.155$), and the ratio (1.73) is significantly >1 ($P < 0.001$), rather than <1 as expected.

To understand why this result does not contradict the evidence of balancing selection from the Ewens-Watterson-Slatkin test, one has to understand the d_S/d_N ratio. Synonymous substitutions are assumed to be neutral, and the diversity increases at a rate proportional to time. The nonsynonymous substitutions can be neutral, detrimental, or advantageous. The neutral nonsynonymous substitutions will accumulate at the same rate as the synonymous substitutions, giving a ratio of one. Because of selection, detrimental nonsynonymous substitutions will accumulate at a much slower rate, giving a smaller distance than expected from neutral mutations and thus giving a ratio of much greater than one. And, because of selection, advantageous nonsynonymous substitutions will accumulate at a much faster rate than the synonymous substitutions, giving a larger distance than expected and thus giving a ratio much less than one. The actual ratio will be made up of the combination of these three ratios.

We can use this understanding of the dynamics of the d_S/d_N ratio to conclude something about the function of the different regions of the *ospC* gene. The first region (bp 1–123) shows a ratio of 2.66 with a low number of synonymous substitutions, much lower than any of the other regions. These low d_S and d_N values are consistent with the first region having a recent common ancestor, and the high value for the ratio suggests that functional constraint plays a dominant role. Were functional constraint also responsible for the low values of d_S and d_N, then these values from between-species comparisons should also be low. Yet the d_S and d_N values between the species (*B. burgdorferi sensu stricto*, *B. afzelii*, and *B. garinii*) are almost 10-fold higher than the within-species values.

For the second region (bp 124–213), the ratio is 5.02, which is significantly ($P < 0.001$) >1, indicating that most nonsynonymous mutations are detrimental and that functional constraint prevents nonsynonymous divergence in this region. Indeed a similar pattern of a high d_S/d_N ratio in the between-species comparisons confirms this interpretation.

Regions 3–5 (bp 214–480) of *ospC* are the regions likely to be subject to balancing selection. The d_S/d_N ratio of these regions is larger than one, although not significantly ($P > 0.1$). The reduction in the d_S/d_N ratio is because there are more nonsynonymous changes within this region than in region 2, while the rate of synonymous substitutions is the same. The large synonymous distance suggests a distant common ancestor.

TABLE 4

**The 19 major *ospC* allele groups for *Borrelia burgdorferi sensu stricto*
and the within-group percentage differences**

Major group (%)[a]	*ospC* allele	Accession no.	Origin
A (0.02)	B31	U01894	New York
	CA4	L81131	California
	HII	U91792	Italy
	IP1	U91799	France
	IP2	L42887	France
	IP3	U91797	France
	L5	U91798	Austria
	P1F	U91801	Austria
	PKa	X69589	Germany
	Txgw	X84783	Texas
	OC1[b]	AF029860	New York
B (0.42)	35B808	U91794	Germany
	61BV3	U91795	Germany
	BUR	X84765	New York
	DK7	X73625	Denmark
	PBre	X81522	Germany
	ZS7	L42868	Germany
	OC2[b]	AF029861	New York
C	OC3[b]	AF029862	New York
D (0.18)	CA-11.2A	L25413	California
	OC4[b]	AF029863	New York
E (0.37)	Z8691	L42894	Pennsylvania
	N40	U04240	Connecticut
	OC5[b]	AF029864	New York
	OC7[b]	AF029866	New York
F (0.39)	27579	L42896	Connecticut
	B. pacificus	X83555	California
	Son188	L81130	California
	OC6[b]	AF029865	New York
G	OC8[b]	AF029867	New York
H	OC9[b]	AF029868	New York
I (0.25)	297[c]	L42893	Connecticut
	HB19	U04281	Connecticut
	OC10[b]	AF029869	New York
J (0.37)	MIL	U91802	Slovakia
	OC11[b]	AF029870	New York
K (1.35)	272	X84785	United States
	297[c]	U08284	Connecticut
	28354	L42895	Maryland
	KIPP[d]	X84782	New York
	MUL	X84779	New York
	OC12[b]	AF029871	New York
	OC13[b]	AF029872	New York
L (0.57)	21347	L42899	United States
	T255	X81524	Germany
M	2591	U01892	Connecticut
N	26815	L42897	Connecticut
O	DUNKIRK	X84778	New York
P	20006	U91796	France
Q	212	U91790	France
R (0.00)	ESP-1	U91791	Spain
	NE56	U91800	Switzerland
S	Z136	U91793	Germany

[a] The percentage difference within each major *ospC* group.
[b] Sequence from present study.
[c] Both *ospC* sequences from strain 297 were found in GenBank entries. The one in group I was submitted by Livey *et al.* (1995), and the one in group K was submitted by Fung *et al.* (1994).
[d] KIPP was excluded from all the analyses containing GenBank sequences (see text for explanation).

TABLE 5

Ewens-Watterson-Slatkin test of allele frequency distribution

Category	Mobility classes	Major *ospC* groups
No. of classes or groups	13	11
Sample size	74	74
Observed F	0.115	0.118
Expected F	0.189	0.227
No. of simulations	10,000	10,000
Probability	0.0238[a]	0.0017[a]
Slatkin's exact probability	0.0398[a]	0.0045[a]

[a] Two-tailed test.

Thus, the combined effects of selection for advantageous mutations and selection against detrimental ones give a ratio larger than one, but much smaller than the ratio in regions 1 and 2.

Region 6 (bp 481–567) is like regions 3–5, but with a ratio significantly ($P < 0.05$) greater than one. Region 6 shows more constraint than regions 3–5. The d_S/d_N ratios of regions 3–6 are consistent with a pattern of balancing selection when the entire region is being maintained for diversity.

This result suggests that the immunological differences are multifactorial, and different clones are selected. This is in contrast to immune escape, where a single amino acid change is selected within a clone. We propose that these two types of selection for diversity be given different names, balancing selection for the first and diversifying selection for the second. In balancing selection, alleles will be maintained in the population for a very long time either by frequency or by niche-dependent selection. Balancing selection will be characterized by a very long time to the common ancestor and a d_S/d_N ratio greater than one and will usually show significant deviation from the Ewens distribution. In diversifying selection, new alleles are constantly being selected and so each of these new alleles is expected to remain in the population a relatively short time. Diversifying selection will be characterized by a d_S/d_N ratio significantly less than one, a very short time to the common ancestor, and sometimes can show significant deviation from the Ewens distribution.

DISCUSSION

Population genetics survey using PCR-SSCP: We expect PCR-SSCP analysis to be commonly used in population genetics studies where large sample sizes and intensive data collection are necessary. This was noted by Aguade *et al.* (1994), "SSCP analysis followed by stratified DNA sequencing is clearly efficient for surveying regions of low polymorphism" (p. 4662). But then they go on to say, "But there are questions about the nature

and proportion of polymorphisms that can be reliably surveyed" (p. 4662). While they estimate that as much as 50% of the single nucleotide variation might be missed, they conclude that in general the fraction missed will be of the order of 14%. The large gel format (Hydrolink-MDE gel) used by Aguade *et al.* (1994) has been shown to be less sensitive (79 *vs.* 95%) in detecting nucleotide substitutions than the small gel format (PhastSystem gel; Vidal-Puig and Moller 1994) we use. We found the small gel format was easy to use and gave clear, reliable results.

Major *ospC* groups: We identified 11 major *ospC* groups within a single population. When combined with sequences obtained worldwide, 19 major *ospC* groups can be distinguished. So far, 4 groups (groups A, B, J, and L) have been found both in Europe and the United States, 4 groups (groups P, Q, R, and S) have been found only in Europe, while the remaining 11 groups have been found only in the United States.

Foretz *et al.* (1997) hypothesized that *B. burgdorferi sensu stricto* originated in North America and only recently spread to Europe. Our data support this idea. Sequences within the same *ospC* group are identical if they are of European origin (seven sequences for group A, five sequences for group B, and two sequences for group R), while in the United States, there is considerable variation within groups. For the major *ospC* allele group A, the four United States sequences (from California, New York, and Texas), only OC1 and B31, both isolated from Shelter Island, New York, are identical. In group B, OC2 is different from BUR (from New York); in group D, OC4 is different from CA-11.2A (from California); in group E, OC7 is different from the rest (from Pennsylvania and Connecticut); in group F, Son188 (from California) is different from the rest (from California, Connecticut, and New York); in group I, OC10 is different from the rest (from Connecticut); and in group K, only OC12 and 28354 (from Maryland) are alike. In group A, the sequence found on the east coast of the United States, *i.e.*, the sequence in B31 and OC1, is identical to the sequence found in Europe; however, in group B, the sequence in Europe is about equally distant from all the North American sequences. These data suggest that a strain like B31 was transferred recently from the east coast of the United States and has spread throughout Europe. However, it does not preclude continuous transfer of strains from the United States to Europe over a long period of time.

How can one explain the presence of four major groups found only in Europe given the hypothesis of the United States origin of this genospecies? There are three possible explanations: (1) These groups are common in the United States but simply have not yet been discovered here; (2) these groups could represent much older migrations of strains that are now rare or extinct in the United States; (3) these groups could have recombined with the European genospecies and now have

TABLE 6

d_S/d_N ratios of different *ospC* gene regions

Gene region[a]	$d_S \pm SE^b$	$d_N \pm SE^b$	d_S/d_N
Total	0.268 ± 0.019	0.155 ± 0.008	1.73
1 (1–123)	0.040 ± 0.022	0.015 ± 0.007	2.66
2 (124–213)	0.261 ± 0.057	0.052 ± 0.016	5.02
3 (214–312)	0.334 ± 0.056	0.269 ± 0.029	1.24
4 (313–387)	0.321 ± 0.078	0.190 ± 0.028	1.69
5 (388–480)	0.273 ± 0.063	0.249 ± 0.029	1.10
6 (481–567)	0.334 ± 0.062	0.185 ± 0.027	1.81

[a] The division of *ospC* was according to the sliding-window plot of Figure 4.

[b] Both d_S and d_N are estimated using Nei and Gojobori's method (Nei and Gojobori 1986) and their respective standard errors (SEs) are estimated using Nei and Jin's method (Nei and Jin 1989). Both of these methods are implemented in MEGA (Kumar *et al.* 1993).

sequences similar to those found in another species. To test this last possibility we did a BLAST search for gene regions 3–5 (bp 166–429, see Table 6) of these four groups. We found that the central region of major *ospC* group P is almost identical (one and three differences over 257 bp) with the central region of *B. afzelii* strains DK1 (GenBank accession number X73627) and PLud (GenBank accession number X83552), and the central region of major *ospC* group S is almost identical (246 out of 249 bp) with *B. garinii* strains TIs1 (GenBank accession number X81525) and H13 (GenBank accession number L42889). The central regions of groups Q and R did not match any known sequences. The N terminus of all four groups showed the correct species-specific signature (Livey *et al.* 1995), excluding the possibility that these strains were misidentified. This shows genetic transfer across genospecies is possible and, if the migration of *B. burgdorferi sensu stricto* is from United States to Europe, that *B. burgdorferi sensu stricto* has picked up the central region of *ospC* from the European species.

Species-wide distribution of major *ospC* groups: One interesting and important finding from this study is that most of the major *ospC* groups found worldwide (as listed in GenBank) are found within a single population. This pattern indicates that the geographic distribution of major *ospC* groups is relatively homogeneous. Different parts of the species range have very similar sets of *ospC* alleles and local diversity is high. Evidence based on very limited microsequencing of OspC proteins purified from *B. burgdorferi* strains isolated in Illinois suggests high local diversity of *ospC* (Picken *et al.* 1995). Similarly, as reported by Livey *et. al.* (1995), five out eight *ospC* RFLP groups for *B. afzelii* were found near the city of Vienna. Strains isolated from infected humans from a local area also showed high *ospC* diversity (Masuzawa *et al.* 1997). Obviously, a more expanded and detailed

survey is needed to see if the pattern also holds in other parts of the world and for other genospecies of Borrelia.

Frequency-dependent selection in maintaining major *ospC* groups: The evidence for balancing selection at the *ospC* locus is: (1) There exist a large number of alleles within a local population; (2) the allele frequencies are more even than expected by neutrality (Ewens-Watterson-Slatkin test); (3) the alleles differ at many nucleotides, suggesting a common ancestor a long time ago, even before the genospecies split (D. E. Dykhuizen, unpublished data); and (4) almost all the variation in *ospC* is between major *ospC* groups.

One of the striking results of this study is the lack of variation within major *ospC* groups within a population. Nine of the 11 major groups are isosequential. This lack of polymorphism within major groups indicates that the effective population size of each group is very small, which is typical for parasites. Because a successful transmission of spirochetes from one host to another involves many random events at every stage of the infection cycle, the most likely fate of a newly derived mutant is extinction within a short time. Nevertheless, new mutations will be fixed, by chance, within local populations, causing geographical variation. Variation is found within groups across the United States. Together these facts, lack of variation within a population and variation across populations, suggest that Borrelia is broken into many semi-isolated populations.

Dobzhansky (1970) suggested that balancing selection can arise when different environments favor different genotypes. Because adult and nymphal ticks each prefer different kinds of animals and because the frequencies of *ospA* alleles are different in nymphal and adult ticks, we have concluded that this type of balancing selection is important in *B. burgdorferi* (Qiu *et al.* 1997). However, the difference in the *ospA* allele frequencies between the nymphal and adult ticks is not large, suggesting this type of balancing selection is not sufficient to maintain the number of alleles and the frequency distribution of the major *ospC* groups as seen here in a single population of adult ticks.

Thus, rather than adaptation to different animal hosts, we propose that frequency-dependent selection, another form of balancing selection, is the major force maintaining the variation in *ospC*. We propose that this frequency-dependent selection is created by the host immune system. OspC is clearly a protective antigen. It has been demonstrated that OspC is expressed on the surface of spirochetes during tick feeding on its host (Schwan *et al.* 1995), and the presence of OspC protein elicits the host to mount a strong IgM response at an early phase of infection (Dressler *et al.* 1993; Fung *et al.* 1994; Engstrom *et al.* 1995). Furthermore, animals immunized with anti-OspC vaccine are protected against subsequent infection by the same strain of *B. burgdorferi* (Gilmore *et al.* 1996) but not by heterologous strains (Probert *et al.* 1997). Passive immunization with

polyclonal and monospecific mouse immune sera to recombinant OspC can cure chronically infected mice (Zhong *et al.* 1997). Thus anti-OspC antibodies can prevent reinfection with the same strain and clear present infections.

In an endemic area like Long Island and Shelter Island, many ticks are infected with *B. burgdorferi*, often with multiple clones (Qiu *et al.* 1997). Moreover, the mice and chipmunks are simultaneously infested with many ticks (Hofmeister and Childs 1995; Qiu *et al.* 1997). Thus, every host will be infected with a large and heterogeneous population of *B. burgdorferi*, expressing many different OspC proteins. The most abundant *ospC* group will, on average, be the group that a host responds to immunologically both first and most strongly. First because it is the OspC protein most often seen with an initial infection and most strongly because there will be the greatest number of repeated infections with this type. Thus a rare type is more likely to establish a persistent infection than a common one and be passed onto the next generation of ticks.

There is no evidence for immune escape and consequently diversifying selection in these Borrelia. When mice are infected in the laboratory, they become chronically infected, but the genotype of the strain does not change. They are stable in their plasmid profile (Persing *et al.* 1994), and the *ospC* sequence remains constant during the infection period (Stevenson *et al.* 1994). As pointed out by Hofmeister and Childs (1995), the above-mentioned studies used a monoculture containing only a single clone. However, the plasmid profiles usually change over time if a large number of clones are used (Barthold *et al.* 1993). This change is due to clonal selection among different strains of *B. burgdorferi* present in the inocula. A similar phenomenon would be important in the field—selection for the rare clones and against the common ones.

Population dynamics of Borrelia: From the data presented, we propose that Borrelia is differentiated into local populations; *i.e.*, the migration rate is not so high as to make the species a panmictic population. Within each local population, the effective population size is quite small, making genetic drift an important force in molding the population structure of Borrelia. If this drift were not countered by strong balancing selection and some migration, local populations would lack diversity. Frequency-dependent selection mediated by the immune system of the mammalian host acts to preserve the diversity between major *ospC* groups.

Recombination causes divergence of strains within the same major group and homogenizes strains in different major groups. As reported in the results, those within-group pairwise differences that showed diversity >1% showed evidence of this diversity being created by recombination. Likewise, those between-group pairwise differences that showed differences of <14.5% showed

evidence of similarities created by recombination. Balancing selection must select against these recombinants, otherwise the tight bunching of the major groups as seen in Figure 5 would be lost. From this, we conclude that the major *ospC* groups will be stable over time.

It is interesting to note that the five groups (E, G, I, J, and K) that show the lowest frequency in this data show the evidence for migration. Only groups E and K contain two mobility classes. Also, group G is the only group where the *ospC* allele is linked to two *ospA* alleles (OC8, Table 3). We do not expect the same *ospC* allele to be linked to two different *ospA* alleles within a single population, if the effective population size is small and the balancing selection on major *ospC* groups is strong. This diversity should have been lost by drift. However, in different populations, there is no reason to expect the linkage relationships to be the same. Thus, we conclude that this diversity in linkage patterns is created by migration. This result suggests that there is a low level of migration, but so low that migrant alleles are detected only when their frequency is increased by selection.

Clinical implications: Since OspC is one of the few spirochetal proteins that elicit an early strong IgM response specific to *B. burgdorferi* (Dressler *et al.* 1993; Fung *et al.* 1994; Engstrom *et al.* 1995), it is only natural to use OspC as an early serodiagnostic antigen for Lyme borreliosis. However, many studies on the efficacy of using OspC protein as a serodiagnostic antigen either on human patients (Fung *et al.* 1994; Padula *et al.* 1994; Gerber *et al.* 1995; Mathiesen *et al.* 1996) or infected dogs and horses (Magnarelli *et al.* 1997) showed that it gave specific, but not sensitive, results. That is, there were few false-positive identifications, but many false-negative ones. This result should not come as a total surprise since OspC proteins are extremely variable even in a local area. Because all of these studies used only one type of OspC protein (strain 297, 2591, or MUL) as a diagnostic antigen to test against a bank of collected sera, specific but not sensitive results can be expected. Clearly, to increase sensitivity and the chance of early detection, it is necessary that the array of immunologically distinct OspC proteins be used as serodiagnostic antigens.

OspC and other surface proteins have been the subjects of intensive efforts to develop a protective vaccine against *B. burgdorferi* infection. Many animal studies indicate the effectiveness of an anti-OspC vaccine (Preac-Mursic *et al.* 1992; Probert and LeFebvre 1994; Gilmore *et al.* 1996). However, all the studies were conducted with only one type of OspC protein. A proper mix of various types of OspCs will be necessary for an anti-OspC vaccine to be effective. The tremendous amount of genetic variation at *ospC* decreases the suitability of using OspC as a protective antigen unless a way of handling this variation can be found.

LITERATURE CITED

Aguade, M., W. Meyers, A. D. Long and C. H. Langley, 1994 Single-strand conformation polymorphism analysis coupled with stratified DNA sequencing reveals reduced sequence variation in the *su(s)* and *su(w^a)* regions of the *Drosophila melanogaster* X chromosome. Proc. Natl. Acad. Sci. USA **91:** 4658–4662.

Anthonissen, F. M., M. De Kesel, P. P. Hoet and G. H. Bigaignon, 1994 Evidence for the involvement of different genospecies of *Borrelia* in the clinical outcome of Lyme disease in Belgium. Res. Microbiol. **145:** 327–331.

Assous, M. V., D. Postic, G. Paul, P. Nevot and G. Baranton, 1993 Western blot analysis of sera from Lyme borreliosis patients according to the genomic species of the *Borrelia* strains used as antigens. Eur. J. Clin. Microbiol. Infect. Dis. **12:** 261–268.

Balmelli, T., and J.-C. Piffaretti, 1995 Association between different clinical manifestations of Lyme disease and different species of *Borrelia burgdorferi* sensu lato. Res. Microbiol. **146:** 329–340.

Balmelli, T., and J.-C. Piffaretti, 1996 Analysis of the genetic polymorphism of *Borrelia burgdorferi* sensu lato by multilocus enzyme electrophoresis. Int. J. Syst. Bacteriol. **46:** 167–172.

Baranton, G., D. Postic, I. Saint Girons, P. Boerlin, J.-C. Piffaretti *et al.*, 1992 Delineation of *Borrelia burgdorferi* sensu stricto, *Borrelia garinii* sp. nov., and group VS 461 associated with Lyme borreliosis. Int. J. Syst. Bacteriol. **42:** 378–383.

Barbour, A. G., and D. Fish, 1993 The biological and social phenomenon of Lyme disease. Science **260:** 1610–1616.

Barbour, A. G., and C. F. Garon, 1988 The genes encoding major surface proteins of *Borrelia burgdorferi* are located on a plasmid. Ann. NY Acad. Sci. **539:** 144–153.

Barthold, S. W., M. S. de Souza, J. L. Janotka, A. L. Smith and D. H. Persing, 1993 Chronic Lyme borreliosis in the laboratory mouse. Am. J. Pathol. **143:** 959–971.

Bosler, E. M., 1993 Tick vectors and hosts, pp. 18–26 in *Lyme Disease*, edited by P. K. Coyle. Mosby-Year Book, St. Louis.

Burgdorfer, W. A., A. G. Barbour, S. F. Hayes, J. L. Nenach, E. Gruwldt *et al.*, 1982 Lyme disease a tick borne spirochetosis? Science **216:** 1317–1319.

Caporale, D. A., and T. D. Kocher, 1994 Sequence variation in the outer-surface-protein genes of *Borrelia burgdorferi*. Mol. Biol. Evol. **11:** 51–64.

Craft, J. E., D. K. Fischer, G. T. Shiamamoto and A. R. Steere, 1986 Antigens of *Borrelia burgdorferi* recognized during Lyme disease. J. Clin. Invest. **78:** 934–939.

Cunningham, T. M., D. D. Thomas, S. D. Thompson, J. N. Miller and M. A. Lovett, 1988 Identification of *Borrelia burgdorferi* surface components by Triton X-114 phase partitioning. Ann. NY Acad. Sci. **539:** 376–378.

Dobzhansky, T., 1970 *Genetics of the Evolutionary Process.* Columbia University Press, New York.

Dressler, F., J. A. Whalen, B. N. Reinhardt and A. C. Steere, 1993 Western blotting in the serodiagnosis of Lyme disease. J. Infect. Dis. **167:** 392–400.

Dykhuizen, D. E., D. S. Polin, J. J. Dunn, B. Wilske, V. Preac-Mursic *et al.*, 1993 *Borrelia burgdorferi* is clonal: implications for taxonomy and vaccine development. Proc. Natl. Acad. Sci. USA **90:** 10163–10167.

Engstrom, S. M., E. Shoop and R. C. Johnson, 1995 Immunoblot interpretation criteria for serodiagnosis of early Lyme disease. J. Clin. Microbiol. **33:** 419–427.

Ewens, W. J., 1972 The sampling theory of selectively neutral alleles. Theor. Popul. Biol. **3:** 87–112.

Fikrig, E., S. W. Barthold, F. S. Kantor and R. A. Flavell, 1990 Protection of mice against the Lyme disease agent by immunizing with recombinant ospA. Science **250:** 552–556.

Fikrig, E., S. W. Barthold, D. H. Persing, X. Sun, F. S. Kantor *et al.*, 1992 *Borrelia burgdorferi* strain 25015: characterization of outer surface protein A and vaccination against infection. J. Immunol. **7:** 2256–2260.

Foretz, M., D. Postic and G. Baranton, 1997 Phylogenetic analysis of *Borrelia burgdorferi* sensu stricto by arbitrarily primed PCR and pulsed-field gel electrophoresis. Int. J. Syst. Bacteriol. **47:** 11–18.

Fung, B. P., G. L. McHugh, J. M. Leong and A. C. Steere, 1994 Humoral immune response to outer surface protein C of *Borrelia burgdorferi* in Lyme disease: role of the immunoglobulin M response in the serodiagnosis of early infection. Infect. Immun. **62:** 3213–3221.

Gerber, M. A., E. D. Shapiro, G. L. Bell, A. Sampieri and S. Padula, 1995 Recombinant outer surface protein C ELISA for the diagnosis of early Lyme disease. J. Infect. Dis. **171:** 724–727.

Gilmore, R. D., K. J. Kappel, M. C. Dolan, T. R. Burkot and B. J. B. 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.

Golde, W. T., T. R. Burkot, J. Piesman, M. C. Dolan, C. Capiau *et al.*, 1995 The Lyme vaccine candidate outer surface protein A (OspA) in a formulation compatible with human use protects mice against natural tick transmission of *B. burgdorferi*. Vaccine **13:** 435–441.

Guttman, D. S., P. W. Wang, I.-N. Wang, E. M. Bosler, B. J. Luft *et al.*, 1996 Multiple infections of *Ixodes scapularis* ticks by *Borrelia burgdorferi* as revealed by single-strand conformation polymorphism analysis. J. Clin. Microbiol. **34:** 652–656.

Hartl, D. L., and A. G. Clark, 1989 *Principles of Population Genetics.* Sinauer Associates, Sunderland, MA.

Hofmeister, E. K., and J. E. Childs, 1995 Analysis of *Borrelia burgdorferi* sequentially isolated from *Peromyscus leucopus* captured at a Lyme disease enzootic site. J. Infect. Dis. **172:** 462–469.

Hongyo, T., G. S. Buzard, R. J. Calvert and C. M. Weghorst, 1993 'Cold SSCP': a simple, rapid and non-radioactive method for optimized single-stranded conformation polymorphism analyses. Nucleic Acids Res. **21:** 3637–3642.

Hughes, A. L., 1991 Circumsporozoite protein genes of malaria parasites (*Plasmodium* spp.): evidence for positive selection on immunogenic regions. Genetics **127:** 345–353.

Hughes, A. L., 1992 Positive selection and interallelic recombination at the merozoite surface antigen-1 (MSA-1) locus of *Plasmodium falciparum*. Mol. Biol. Evol. **9:** 381–393.

Hughes, A. L., and M. Nei, 1988 Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. Nature **335:** 167–170.

Hughes, C. A., S. M. Engstrom, L. A. Coleman, C. B. Kodner and R. C. Johnson, 1993 Protective immunity is induced by a *Borrelia burgdorferi* mutant that lacks OspA and OspB. Infect. Immun. **61:** 5115–5122.

Hughes, M. K., and A. L. Hughes, 1995 Natural selection on *Plasmodium* surface proteins. Mol. Biochem. Parasitol. **71:** 99–113.

Jauris-Heipke, S., G. Liegl, V. Preac-Mursic, D. Robler, E. Schwab *et al.*, 1995 Molecular analysis of genes encoding outer surface protein C (OspC) of *Borrelia burgdorferi* sensu lato: relationship to *ospA* genotype and evidence of lateral gene exchange of *ospC*. J. Clin. Microbiol. **35:** 1860–1866.

Jonsson, M., and S. Bergstrom, 1995 Transcriptional and translational regulation of the expression of the major outer surface proteins in Lyme disease *Borrelia* strains. Microbiology **141:** 1321–1329.

Kawabata, H., T. Masuzawa and Y. Yanagihara, 1993 Genomic analysis of *Borrelia japonica* sp. nov. isolated from *Ixodes ovatus* in Japan. Microbiol. Immunol. **37:** 843–848.

Kumar, S., K. Tamara and M. Nei, 1993 MEGA: molecular evolutionary genetics analysis. Version 1.01. The Pennsylvania State University, University Park, PA.

Li, W.-H., and D. Graur, 1991 *Fundamentals of Molecular Evolution.* Sinauer Associates, Sunderland, MA.

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.

Luft, B. J., S. Mudri, W. Jiang, R. J. Dattwyler, P. D. Gorevic *et*

al., 1992   The 93-kilodalton protein of *Borrelia burgdorferi*: an immunodominant protoplasmic cylinder antigen. Infect. Immun. **60:** 4309–4321.

Magnarelli, L. A., J. F. Anderson and D. Fish, 1987   Transovarial transmission of *Borrelia burgdorferi* in *Ixodes dammini* (Acari:Ixodidae). J. Infect. Dis. **156:** 234–236.

Magnarelli, L. A., R. A. Flavell, S. J. Padula, J. F. Anderson and E. Fikrig, 1997   Serological diagnosis of canine and equine borreliosis: use of recombinant antigens in enzyme-linked immunosorbent assays. J. Clin. Microbiol. **35:** 169–173.

Marconi, R. T., D. S. Samuels and C. F. Garon, 1993   Transcriptional analyses and mapping of the *ospC* gene in Lyme disease spirochetes. J. Bacteriol. **175:** 926–932.

Masuzawa, T., T. Komikado and Y. Yanagihara, 1997   PCR-restriction fragment length polymorphism analysis of the *ospC* gene for detection of mixed culture and for epidemiological typing of *Borrelia burgdorferi* sensu stricto. Clin. Diag. Lab. Immunol. **4:** 60–63.

Mathiesen, M. J., K. Hansen, N. Axelsen, L. Halkier-Sorensen and M. Theisen, 1996   Analysis of the human antibody response to outer surface protein C (OspC) of *Borrelia burgdorferi* sensu stricto, *B. garinii*, and *B. afzelii.* Med. Microbiol. Immunol. **185:** 121–129.

Nei, M., and T. Gojobori, 1986   Simple method for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. **3:** 418–426.

Nei, M., and L. Jin, 1989   Variances of the average numbers of nucleotide substitutions within and between populations. Mol. Biol. Evol. **6:** 290–300.

Padula, S. J., A. Sampieri, F. Dias, A. Szczepanski and R. W. Ryan, 1993   Molecular characterization and expression of p23 (OspC) from a North American strain of *Borrelia burgdorferi.* Infect. Immun. **61:** 5097–5105.

Padula, S. J., F. Dias, A. Sampieri, R. B. Craven and R. W. Ryan, 1994   Use of recombinant OspC from *Borrelia burgdorferi* for serodiagnosis of early Lyme disease. J. Clin. Microbiol. **32:** 1733–1738.

Persing, D. H., D. Mathiesen, D. Podzorski and S. W. Barthold, 1994   Genetic stability of *Borrelia burgdorferi* recovered from chronically infected immunocompetent mice. Infect. Immun. **62:** 3521–3527.

Picken, R. N., Y. Cheng, D. Han, J. A. Nelson, A. G. Reddy *et al.*, 1995   Genotypic and phenotypic characterization of *Borrelia burgdorferi* isolated from ticks and small animals in Illinois. J. Clin. Microbiol. **33:** 2304–2315.

Piesman, J., J. G. Donahue, T. N. Mather and A. Spielman, 1986   Transovarially acquired Lyme disease spirochetes (*Borrelia burgdorferi*) in field-collected larval *Ixodes dammini* (Acari: Ixodidae). J. Med. Entomol. **23:** 219.

Postic, D., M. V. Assous, P. A. D. Grimont and G. Baranton, 1994   Diversity of *Borrelia burgdorferi* sensu lato evidenced by restriction fragment length polymorphism of rrf(5S)-ffl(23S) intergenic spacer amplicons. Int. J. Syst. Bacteriol. **44:** 743–752.

Preac-Mursic, V., B. Wilske, E. Patsouris, S. Jauris, G. Will *et al.*, 1992   Active immunization with pC protein of *Borrelia burgdorferi* protects gerbils against *B. burgdorferi* infection. Infection **20:** 342–349.

Probert, W. S., and R. B. LeFebvre, 1994   Protection of C3H/HeN mice from challenge with *Borrelia burgdorferi* through active immunization with OspA, OspB, or OspC, but not with OspD or the 83-kilodalton antigen. Infect. Immun. **62:** 1920–1926.

Probert, W. S., M. Crawford, R. B. Cadiz and R. B. LeFebvre, 1997   Immunization with outer surface protein (Osp) A, but not OspC, provides cross-protection of mice challenged with North American isolates of *Borrelia burgdorferi.* J. Infect. Dis. **175:** 400–405.

Qiu, W., E. M. Bosler, J. R. Campbell, G. D. Ugine, I.-N. Wang *et al.*, 1997   A population genetic study of *Borrelia burgdorferi* sensu stricto from eastern Long Island, New York, suggested frequency-dependent selection, gene flow and host adaptation. Hereditas **127:** 203–216.

Riley, M. A., 1993   Positive selection for colicin diversity in bacteria. Mol. Biol. Evol. **10:** 1048–1059.

Rosa P. A., T. Schwan and D. Hogan, 1992   Recombination between genes encoding major outer surface protein A and protein B of *Borrelia burgdorferi*. Mol. Microbiol. **6:** 3031–3040.

Sadziene, A., B. Wilske, M. Ferdows and A. Barbour, 1993   The cryptic ospC gene of *Borrelia burgdorferi* B31 is located on a circular plasmid. Infect. Immun. **61:** 2192–2195.

Sandberg, S., T. E. Awerbuch and A. Spielman, 1992   A comprehensive multiple matrix model representing the life cycle of the tick that transmits the agent of Lyme disease. J. Theor. Biol. **157:** 203–220.

Schneider, S., J.-M. Kueffer, D. Roessli and L. Excoffier, 1996   Arlequin: a software package for population genetics. Version 1.0. Genetics & Biometry Lab, Department of Anthropology, University of Geneva.

Schwan, T. G., J. Piesman, W. T. Golde, M. C. Dolan and P. A. Rosa, 1995   Induction of an outer surface protein on *Borrelia burgdorferi* during tick feeding. Proc. Natl. Acad. Sci. USA **92:** 2909–2913.

Seibert, S. A., A. Y. Howell, M. K. Hughes and A. L. Hughes, 1995   Natural selection on the *gag*, *pol*, and *env* genes of human immunodeficiency virus 1 (HIV-1). Mol. Biol. Evol. **12:** 803–813.

Simon, M. M., U. E. Schaible, M. D. Kramer, C. Eckerskorn, C. Museteanu *et al.*, 1991   Recombinant OspA from *Borrelia burgdorferi* induces antibodies protective against spirochete infection in mice. J. Infect. Dis. **164:** 123–132.

Slatkin, M., 1994   An exact test for neutrality based on the Ewens sampling distribution. Genet. Res. **64:** 71–74.

Slatkin, M., 1996   A correction to the exact test based on the Ewens sampling distribution. Genet. Res. **68:** 259–260.

Spielman, A., 1988   Prospects for suppressing transmission of Lyme disease. Ann. NY Acad. Sci. **539:** 212–220.

Spielman, A., S. R. Telford, III and R. J. Pollack, 1993   The origins and course of the present outbreak of Lyme disease, pp. 83–96 in *Ecology and Environmental Management of Lyme Disease*, edited by H. S. Ginsberg. Rutgers University Press, New Brunswick, NJ.

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.

Stevenson, B., T. G. Schwan and P. A. Rosa, 1995   Temperature-related differential expression of antigens in the Lyme disease spirochete, *Borrelia burgdorferi.* Infect. Immun. **63:** 4535–4539.

Tajima, F., 1991   Determination of window size for analyzing DNA sequences. J. Mol. Evol. **33:** 470–473.

Theisen, M., B. Frederiksen, A.-M. Lebech, J. Vuust and K. Hansen, 1993   Polymorphism in *ospC* gene of *Borrelia burgdorferi* and immunoreactivity of OspC protein: implications for taxonomy and for use of OspC protein as a diagnostic antigen. J. Clin. Microbiol. **31:** 2570–2576.

Theisen, M., M. Borre, M. J. Mathiesen, B. Mikkelsen, A.-M. Lebech *et al.*, 1995   Evolution of the *Borrelia burgdorferi* outer surface protein OspC. J. Bacteriol. **177:** 3036–3044.

Valsangiacomo, C., T. Balmelli and J.-C. Piffaretti, 1997   A phylogenetic analysis of *Borrelia burgdorferi* sensu lato based on sequence information from the *hbb* gene, coding for a histone-like protein. Int. J. Syst. Bacteriol. **47:** 1–10.

van Dam, A. J., H. Kuiper, K. Vos, A. Widjojkusumo, B. M. de Jongh *et al.*, 1993   Different genospecies of *Borrelia burgdorferi* are associated with distinct clinical manifestation of Lyme disease. Clin. Infect. Dis. **17:** 708–717.

Vidal-Puig, A., and D. E. Moller, 1994   Comparative sensitivity of alternative single-strand conformation polymorphism (SSCP) methods. BioTechniques **17:** 490–496.

Wagner, R. R., and M. A. Riley, 1996   Low synonymous site variation at the *lacY* locus in *Escherichia coli* suggests the action of positive selection. J. Mol. Evol. **42:** 79–84.

Watterson, G. A., 1977   Heterosis or neutrality? Genetics **85:** 789–814.

Wilske, B., V. Preac-Mursic, S. Jauris, A. Hofmann, I. Pradel *et al.*, 1993   Immunological and molecular polymorphisms of OspC, an immunodominant major outer surface protein of *Borrelia burgdorferi.* Infect. Immun. **61:** 2182–2191.

Wilson, M. L., and A. Spielman, 1985   Seasonal activity of immature *Ixodes dammini* (Acari: Ixodidae). J. Med. Entomol. **22:** 408–414.

Yuval, B., and A. Spielman, 1990   Duration and regulation of the developmental cycle of *Ixodes dammini* (Acari: Ixodidae). J. Med. Entomol. **27:** 196–201.

Zhong, W., T. Stehle, C. Museteanu, A. Siebers, L. Gern *et al.*, 1997   Therapeutic passive vaccination against chronic Lyme disease in mice. Proc. Natl. Acad. Sci. USA **94:** 12533–12538.

Communicating editor: W. F. Eanes