Four Clones of *Borrelia burgdorferi* Sensu Stricto Cause Invasive Infection in Humans

GERALD SEINOST,1 DANIEL E. DYGHIUSSEN,2 RAYMOND J. DATTWYLER,1* WILLIAM T. GOLDE,1 JOHN J. DUNN,3 ING-NANG WANG,2 GARY P. WORMSER,4 MARTIN E. SCHRIEFER,5 AND BENJAMIN J. LUFT1

Department of Medicine and Department of Ecology and Evolution,1 State University of New York at Stony Brook, Stony Brook, New York 11794; Biology Department, Brookhaven National Laboratory, Upton, New York 11973;2 Department of Medicine, New York Medical College, Valhalla, New York 10595;3 and Centers for Disease Control, Fort Collins, Colorado 80522

Received 20 October 1998/Returned for modification 6 December 1998/Accepted 14 April 1999

Lyme disease begins at the site of a tick bite, producing a primary infection with spread of the organism to secondary sites occurring early in the course of infection. A major outer surface protein expressed by the spirochete early in infection is outer surface protein C (OspC). In *Borrelia burgdorferi* sensu stricto, OspC is highly variable. Based on sequence divergence, alleles of ospC can be divided into 21 major groups. To assess whether strain differences defined by ospC group are linked to invasiveness and pathogenicity, we compared the frequency distributions of major ospC groups from ticks, from the primary erythema migrans skin lesion, and from secondary sites, principally from blood and spinal fluid. The frequency distribution of ospC groups from ticks is significantly different from that from primary sites, which in turn is significantly different from that from secondary sites. The major groups A, B, I, and K had higher frequencies in the primary sites than in ticks and were the only groups found in secondary sites. We define three categories of major ospC groups: one that is common in ticks but very rarely if ever causes human disease, a second that causes only local infection at the tick bite site, and a third that causes systemic disease. The finding that all systemic *B. burgdorferi* sensu stricto infections are associated with four ospC groups has importance in the diagnosis, treatment, and prevention of Lyme disease.

OspC is protective against tick-transmitted *Borrelia* infection (12). However, the protection is limited to the immunizing OspC allele, as challenge with heterologous isolates results in infection (27).

OspC is very diverse (16). Wang et al. (40) found 13 ospC alleles of *B. burgdorferi* sensu stricto within a small tick population, while Livey et al. (17) found 34 alleles in 76 *B. burgdorferi* sensu lato isolates. Wang et al. (40) defined major ospC groups by using the observation that ospC alleles are either very similar (less than 2% sequence divergence) or very different (greater than 8% sequence divergence), with most having greater than 14% sequence divergence. Using these parameters and including all available sequences from GenBank plus the alleles from the tick population, 19 major ospC groups, groups A through S (Table 1), were defined and strains were associated with each group listed in Wang et al. (40). Strain B31 belongs to group A, strain HB19 belongs to group I, strain N40 belongs to group E, and strain 297 is a mixture of groups I and K. In the present study, we describe two new groups (groups T and U).

There is evidence that ospC has been transferred between strains (7a, 17) and even between genospecies (40). This is not true of *Borrelia* chromosomal genes (8, 20). We find that ospA and ospC alleles in *B. burgdorferi* sensu stricto are almost completely linked (40). This suggests that once an ospC allele has been transferred into a particular background, there is little or no selection for another similar recombinant. Thus each major ospC group represents a clonal population descended from a single recombination. We call each major ospC group a clone and use the terms group and clone interchangeably.

In this study, we compared the frequency distributions of these major groups from ticks with those from human skin and those from skin with those isolated from secondary sites of...
human infection. While many groups found ticks are found in primary skin lesions, the frequency distributions are significantly different. All groups are found more or less commonly in ticks but in the primary skin lesions, only four groups are commonly found, and others found are rarely or not at all. More importantly, only four ospC groups are found in secondary sites.

**MATERIALS AND METHODS**

*B. burgdorferi* sensu stricto groups P through S are found only in Europe. Groups R and S are excluded from the analysis because nearly identical ospC alleles are found in *B. afzelii* and *B. garinii*, showing that these groups were recently created by cross-species transfer.

### RESULTS

*ospC mobility classes in human B. burgdorferi isolates.

We initiated our analysis of *ospC* sequence diversity by accumulating a large collection of isolates of *B. burgdorferi* sensu stricto (*n* = 132) from patient samples of skin, blood, and CSF (Table 2). Each was propagated in vitro and used as a source of DNA for analysis. The *ospC* genotypes of each strain was determined by cold SSCP analysis of the 5′ end (340 bp) of the gene and was later confirmed by SSCP analysis of the 3′ end (314 bp) of *ospC*. In all *B. burgdorferi* isolates, the genetic variation at the 5′ end of the gene corresponded to the variation at the 3′ end. At least two representatives of each SSCP mobility class were subsequently sequenced. The sequences of the same mobility classes were identical in all samples and each mobility class had a unique sequence. Therefore, the sensitivity and specificity of SSCP analysis were 100%. This result confirms the findings of previous studies, which also utilized SSCP analysis to deter-
mine genetic variability in *B. burgdorferi* (13, 40). Each SSCP mobility class was designated an allele. Wang et al. (40) recently described 13 *ospC* alleles. In this study, we present an additional five *ospC* mobility classes, OC14 to -18. OC14 has the same *ospC* sequence as the *ospC* in strain 2591 (GenBank accession no. U01892) and OC15 has the same sequence as the *ospC* in strain 26815 (accession no. L42897).

**Nucleotide sequence accession numbers.** OC16, OC17, and OC18 are alleles that have not been previously described in the literature. The GenBank accession numbers for these *ospC* alleles are AF065143, AF065144, and AF097915, respectively.

**Multiple infections.** Of the 132 primary isolates from patients with Lyme disease in this study, most contained only a single strain. Seven skin isolates and one CSF isolate contained two different strains as determined by SSCP analysis, thus giving a total of 140 different strains. The *ospC* allele pairs found in multiply infected erythema migrans biopsy specimens were OC1 and OC12, OC1 and OC14, OC2 and OC3 (two sets), OC2 and OC12 (two sets), and OC8 and OC18. CSF isolate NY940657 contained *ospC* alleles OC1 and OC12. For CSF isolate 297, which was isolated in Connecticut, there are two *ospC* sequences published in GenBank, L42893 (analogous to OC10) and U08284 (analogous to OC12). The pairwise difference of *ospC* sequences of both strains is 16.4%, suggesting central nervous system infection with two different strains in this isolate. Overall, 5.5% of the isolates in our collection contained two strains. Wang et al. (40) found that 50% of ticks collected on the eastern end of Long Island were infected with multiple strains. This suggests that exposure to multiple strains in a single tick bite is common and raises the possibility that different strains are differentially pathogenic.

To these 140 strains for which we determined the *ospC* allele, we added 22 strains of known *ospC* sequence from GenBank to give a total of 162. The majority of these strains were obtained from either eastern Long Island (n = 51) or Westchester County, N.Y. (n = 77), with the remainder from other areas in the United States (n = 22) and Europe (n = 12) in which the strains are endemic. The isolates were divided into those from sites of the primary infection, erythema migrans skin lesions (n = 118), and secondary sites, where the infection had disseminated (n = 44). This latter group included 20 isolates from CSF, 23 isolates from blood, and 1 isolate from synovial fluid.

**TABLE 2.** *B. burgdorferi* sensu stricto isolates classified according to geographic region, source, and major *ospC* group

<table>
<thead>
<tr>
<th>Geographical source of isolates</th>
<th>Biological source (no. of isolates)</th>
<th>Major <em>ospC</em> group* (no. observed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>New York (Westchester County)</td>
<td>Skin (62)</td>
<td></td>
</tr>
<tr>
<td>Blood (14)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CSF (1)</td>
<td>A (8), B (10), D (1), E (1), G (6), H (4), I (4), J (6), K (18), M (2), U (2)</td>
<td></td>
</tr>
<tr>
<td>New York (eastern Long Island)</td>
<td>Skin (40)</td>
<td></td>
</tr>
<tr>
<td>Blood (1)</td>
<td>A (10), B (6), C (2), G (1), I (4), K (12), N (3), O (1), T (1)</td>
<td></td>
</tr>
<tr>
<td>CSF (10)</td>
<td>A (1)</td>
<td></td>
</tr>
<tr>
<td>Connecticut</td>
<td>Skin (2)</td>
<td></td>
</tr>
<tr>
<td>Blood (3)</td>
<td>C (1), K (1)</td>
<td></td>
</tr>
<tr>
<td>CSF (3)</td>
<td>A (1), I (1), K (1)</td>
<td></td>
</tr>
<tr>
<td>California</td>
<td>Skin (4)</td>
<td></td>
</tr>
<tr>
<td>Blood (2)</td>
<td>A (1), H (1), I (1), M (1)</td>
<td></td>
</tr>
<tr>
<td>CSF (1)</td>
<td>A (1)</td>
<td></td>
</tr>
<tr>
<td>Arkansas</td>
<td>Blood (1)</td>
<td></td>
</tr>
<tr>
<td>Wisconsin</td>
<td>Skin (1)</td>
<td></td>
</tr>
<tr>
<td>Pennsylvania</td>
<td>Skin (1)</td>
<td></td>
</tr>
<tr>
<td>CSF (1)</td>
<td>H (1)</td>
<td></td>
</tr>
<tr>
<td>Texas</td>
<td>Skin (1)</td>
<td></td>
</tr>
<tr>
<td>Michigan</td>
<td>Skin (2)</td>
<td></td>
</tr>
<tr>
<td>Europe (Austria, Denmark, France, Germany, Italy)</td>
<td>Skin (5)</td>
<td></td>
</tr>
<tr>
<td>Blood (2)</td>
<td>A (3), B (2)</td>
<td></td>
</tr>
<tr>
<td>CSF (4)</td>
<td>A (2)</td>
<td></td>
</tr>
<tr>
<td>Synovial fluid (1)</td>
<td>A (4)</td>
<td></td>
</tr>
</tbody>
</table>

* A major *ospC* group is defined as a set of sequences with more than 98% similarity within each group and less than 92% sequence similarity between groups.

**TABLE 3.** Contingency table for major *ospC* group data for lesions and ticks as published in GenBank* and the literature

<table>
<thead>
<tr>
<th>Isolate source (n)</th>
<th>No. of observations in isolates of major <em>ospC</em> group*;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Erythema migrans lesions (22)</td>
<td>A</td>
</tr>
<tr>
<td>11</td>
<td>4</td>
</tr>
<tr>
<td>Isodes scapularis ticks (25)</td>
<td>5</td>
</tr>
</tbody>
</table>

* GenBank data include all *ospC* submissions made before February 1998.
* The numbers of observations for lesions and for ticks were significantly different (χ² = 11.321 with 3 degrees of freedom; P < 0.025).
* Combined major groups (Comb) are defined by individual frequencies of 0.025 or less and include groups D, E, F, I, J, L, M, O, Q, and R.
The present analysis revealed two sensu stricto fall into two classes. Either a pair is different in low-frequency groups. When combined, these groups make up 12.7% of the total number of strains. When all groups with three or fewer isolates are combined to give the lfg (Fig. 1B), a 2-by-8 contingency test comparing the frequencies of the tick samples from eastern Long Island is not significantly different from the GenBank and literature sample ($\chi^2 = 10.84$ with 8 degrees of freedom; $P > 0.25$). Our analysis shows that the frequency distribution of alleles in human isolates is significantly different from the distribution in ticks. This difference does not depend upon geography. Thus, we combined the data from diverse locations for the tests below. There is one caveat to the postulate that there is no geographical substructure. When the local and GenBank tick samples are compared without combining any groups, the test is significant ($\chi^2 = 31.94$ with 15 degrees of freedom; $P < 0.01$). The importance of this result is that it shows that not all groups are found at every sampling site. Consequently, there could be a group not present in southeastern New York state which is infectious.

Collective analysis of all ospC groups presented in this study shows that most groups are found in both ticks and humans (Table 1). However, while the different groups are in near-equal frequency in ticks (40), major groups A and K predominate in humans (Fig. 1).

Comparison of the frequencies of the various groups in primary and secondary sites. We further investigated the pattern of pathogenicity of the various clones by comparing their frequencies in the primary site of infection, the skin, with their frequencies in secondary sites. Only four major groups (A, B, I, and K) were found in both the skin and secondary sites (Fig. 1B and C). All other major groups were found only in the skin. When all groups with three or fewer isolates are combined to give the lfg (Fig. 1B), a 2-by-8 contingency test comparing the frequency distributions of skin isolates versus those of secondary site isolates indicates significance at $P < 0.005$ (Table 5). When no groups are combined, a 2-by-15 contingency test is still significant ($\chi^2 = 24.07$ with 14 degrees of freedom; $P < 0.05$). The distribution of strains from primary and secondary infections.

**TABLE 4. Contingency table for major ospC groups retrieved from lesions and ticks collected from eastern Long Island**

<table>
<thead>
<tr>
<th>Isolate source (n)</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
<th>K</th>
<th>Comb</th>
<th>b</th>
</tr>
</thead>
<tbody>
<tr>
<td>Erythema migrans lesions (46)</td>
<td>13</td>
<td>6</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>4</td>
<td>16</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Ixodes scapularis</em> ticks (74)</td>
<td>12</td>
<td>12</td>
<td>11</td>
<td>9</td>
<td>6</td>
<td>5</td>
<td>7</td>
<td>1</td>
<td>5</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Combined major groups (Comb) are defined by individual frequencies of 0.025 or less and include groups E, J, N, and O.

**Major ospC groups in human *B. burgdorferi* isolates.** The pairwise differences of ospC sequences within *B. burgdorferi* sensu stricto fall into two classes. Either a pair is different in less than 2% of its nucleotides or it is different in more than 8% of its nucleotides. Wang et al. (40) defined 19 major ospC groups, designated A to S. The present analysis revealed two additional ospC groups, designated T and U. OC16 represents major group T and OC17 represents major group U (Table 1). The lowest pairwise differences of group T and U from other major ospC groups are 16.1% and 20.5%, respectively.

OC18 is an allele in group J along with OC11. They have 99.6% sequence identity and were both found in skin isolates. OC11 was isolated five times and OC18 was isolated twice from erythema migrans biopsies. Otherwise, all ospC alleles found in this sample were from different major groups.

**B. burgdorferi clones are differentially pathogenic.** We wished to show that clones are differentially pathogenic by comparing the frequencies of the various major groups in ticks, in the initial infection in the skin, and in disseminated infections. Our samples are not ideally suited for the comparisons that we wish to make since most of the samples are from southeastern New York state. However, we do not think this is a serious problem because the frequency distributions seem to be similar across the species range.

The strains in GenBank and the literature for which theospC sequence has been determined are widely sampled from the entire geographic range of the species. They give a small but random sample of the frequencies of the major ospC groups in ticks and humans. The frequency of the major ospC groups from human isolates is significantly different from the frequency found in ticks (Table 3) (17, 18). This is also true locally. Table 4 shows that the frequency distribution of strains from skin from eastern Long Island is significantly different from tick strains collected in the same area.

To support the contention that geography is an insignificant component in the differences in the frequencies of the various groups in ticks and humans, we performed the two following tests. The frequency distributions of human isolates from eastern Long Island and the human isolates from GenBank and the literature are not significantly different ($\chi^2 = 7.07$ with 7 degrees of freedom; $P > 0.25$). Further, the frequency distribution of the tick samples from eastern Long Island is not significantly different from the GenBank and literature sample ($\chi^2 = 10.84$ with 8 degrees of freedom; $P > 0.25$).
Some groups cause a local infection at the tick bite site but not systemic disease, e.g., ospC groups G, H, J, and T. Finally, there are some groups which are responsible for systemic disease; these are ospC groups A, B, I, and K. Our findings indicate that all systemic B. burgdorferi sensu stricto infections in humans are caused by strains in these four ospC groups.

The data of Steere et al. (33) are consistent with the results of this study. They found that 20% of untreated erythema migrans lesions cleared spontaneously without causing any systemic complications (33). This is not significantly different (P = 0.25 for a 2-by-2 contingency test with double dichotomy) from the percentage of noninvasive strains found in the skin, suggesting that the erythema migrans lesions that clear spontaneously are caused by noninvasive clones.

There is extensive genetic and antigenic diversity of OspC in all three pathogenic genospecies of B. burgdorferi sensu lato (17, 18, 25, 36, 40). In this study, we show that ospC alleles are linked to both infectivity and invasiveness and that invasiveness is confined to a small number of ospC clones. One might expect that systemic disease in the other pathogenic genospecies is also caused by a small number of clones. Thus, ospC is a good marker for human pathogenicity and perhaps its determinator. These findings have important implications not only for our understanding of the pathogenesis of this disease but for its diagnosis and prevention.

Possibility of other invasive clones. Although we feel that our findings are broadly applicable, two factors, the limited geographic distribution of most of the isolates studied and the source of the invasive isolates, primarily blood and CSF, leave open the possibility that additional invasive groups may be defined. This possibility in no way diminishes the overall importance of the finding that only a limited subset of the ospC groups are associated with invasive disease.

Spirochetemia is a transient phenomenon but is presumably key in seeding secondary skin sites, the heart, joints, and nervous system, where these Borrelia organisms cause the secondary and tertiary clinical manifestations of Lyme disease. All four invasive groups were found in isolates from blood and CSF. The one joint isolate belonged to group A. However, it can be inferred that groups not found in the blood will not be found in the joints, since most if not all dissemination of Borrelia to secondary sites is via blood.

Population genetics analysis. This paper relies on population genetic analysis, which is an unusual approach to the study of infectious disease. Thus, we would like to point out the strengths and weaknesses of this analysis. Normally, model organisms are used as substitutes for experiments on humans. However, this substitution works only as long as the properties of the model organism and of humans are the same for the studied phenomena. In this study, the human immune system plays a critical role which is expected to be different from the immune response in model organisms, particularly the mouse. Humans are accidental and usually dead-end hosts, while the mouse is a critical host reservoir. Consequently, one cannot do direct experiments but must reach an inferential conclusion from survey data. The field of population genetics has developed sound procedures for reaching conclusions from survey data. However, this analysis still relies on correlation and does not explain cause.

Another major assumption in the analysis is that we are dealing with clones and using ospC as a marker for the clones. If ospC is the determinator of the observed differences in pathogenesis, this assumption is unimportant. However, if other genes are involved, then the assumption that there are clones with the various alleles of different genes linked in clone complexes is important. Ryan et al. (29) have challenged the...
clone concept. They discovered an interesting phenomenon. In the mixed culture of a major ospC group L strain and a group G strain, the group L strain is selected in the mammal host and the group G strain is selected in the tick. These strains are different not only at the ospC locus but also at ospA and in their plasmid profile. In their discussion, Ryan et al. reject this explanation, instead suggesting that the genome of one strain contains this diversity—two ospC alleles, two ospC alleles, and plasmid content rearranged depending upon the environment. Their reasons for rejecting the mixed culture explanation are as follows. (i) The culture was colony purified before the experiment started. However, the plating efficiency was so low that the colonies that grew up could be mixed colonies where both strains were required for growth. Thus, one need not assume that the culture is a single clone arising from a single cell. (ii) Passage of the rarer strain in the mammal to the tick is unlikely. However, cultures were started using multiple ticks and 10 ticks were used to passage the Borrelia to a new mammalian host. Thus, even if all infected ticks do not pick up the rarer strain, which is then selected in the infected ticks, both the culture and the mammal are likely to contain this strain. Also, we have shown that different clones are not picked up randomly but are likely to be picked up together (28). It is clear that the ospA and ospC alleles are tightly linked even though they are on different plasmids (40). Thus, if the invasiveness is caused by allelic variation at another locus, this variation is likely to be tightly linked to the ospC variation.

**OspC as a virulence factor.** OspC could be the important virulence factor that determines the differences in pathogenicity between clones. Using immunoglobulin M antibodies from patients with neuroborreliosis, Mathiesen et al. (19) recently showed that the C-terminal region of OspC is exposed on the surface of the spirochete. The C-terminal polyproline II-like helix, which is conserved in all *B. burgdorferi* strains, and its adjacent residues are likely to be important for host-parasite interactions. OspC is not required for growth in culture and maintenance of the cp26 plasmid (37).

**Practical implications.** Our results have important implications for the development of vaccines and serum-based diagnostic tests for Lyme disease. Two independent trials of first-generation vaccines for the prevention of Lyme disease recently studied the efficacy and safety of a vaccine that is based on recombinant OspA (32, 35). However, a vaccine that consists of recombinant OspA may require frequent booster immunizations, since high antibody titers are necessary to eliminate *B. burgdorferi* in the tick vector (9). Natural infection with *B. burgdorferi* does not elicit antibody response to this outer membrane protein, as it does to OspC. Our findings suggest that all systemic *B. burgdorferi* sensu stricto infections in humans are associated with four ospC groups. Because vaccines against OspC are known to be protective, but have been limited by the diversity of OspC (27), our finding could lead to a highly protective vaccine. We propose that a vaccine that includes these four forms of OspC would be a second level of protection against disseminated infection of the *B. burgdorferi* spirochete. SSCP analysis may provide a rapid and powerful tool to monitor vaccine efficacy by detecting rare or new invas-ive ospC groups.

New diagnostic assays based on major ospC groups A, B, I, and K could identify those at risk for progressive illness. A number of investigators have used OspC as a serodiagnostic antigen for early Lyme disease (10, 11, 24). In these tests, the use of OspC as a diagnostic antigen gave highly specific but not sensitive results. However, these studies included only one *B. burgdorferi* strain and therefore only one type of OspC. Routine tests for the diagnosis of Lyme disease also use a single-strain protocol and therefore a single ospC allele for detection of antibody to the spirochete. Given that OspC proteins are antigenically variable (27), individuals infected with one strain may produce an antibody response that is not reactive with an OspC protein from a different major group. We therefore predict that antibody detection using antigen preparations incorporating a proper mix of invasive clones of *B. burgdorferi* will be much more sensitive than the present single-strain protocols.

**ACKNOWLEDGMENTS**

This work was supported by grants from the State of New York (10505-4-3R-97) and the National Institutes of Health (AIAR 37256). G.S. was supported by an Erwin Schroedinger research fellowship of the Austrian Academy of Sciences. M.J. was supported by an Erwin Schroedinger research fellowship of the Austrian Academy of Sciences. J.P. was supported by the National Institutes of Health (AIAR 37256).

We gratefully thank D. Qui, D. S. Dunkin, M. Randesi, J. Medalle, B. Lade, L. Butler, M. L. Hayes, and P. Munoz for technical assistance; J. Kieleczawa for assistance with sequencing PCR samples; K. Gintner-Davis (Centers for Disease Control), I. Schwartz (New York Medical College), B. W. Berger, P. K. Coyle, R. Kahlish, and A. Natarro (all from NY at Stony Brook) for providing strains used in this study; B. Rannala for discussions on the statistical tests; and R. Gasser for thoughtful comments on the manuscript.

**REFERENCES**


