JOURNAL OF BACTERIOLOGY, June 1995, p. 3036–3044 Vol. 177, No. 11
0021-9193/95/$04.00+0

Evolution of the *Borrelia burgdorferi* Outer Surface Protein OspC

MICHAEL THEISEN,[1]* MARTIN BORRE,[1] MARIANNE J. MATHIESEN,[1] BO MIKKELSEN,[2] ANNE-METTE LEBECH,[1] AND KLAUS HANSEN[1]

Statens Seruminstitut, Department of Infection-Immunology, DK-2300 Copenhagen S,[1] *and University of Århus, Universitetsparken, Århus,*[2] *Denmark*

Received 21 September 1994/Accepted 23 March 1995

The genes coding for outer surface protein OspC from 22 *Borrelia burgdorferi* strains isolated from patients with Lyme borreliosis were cloned and sequenced. For reference purposes, the 16S rRNA genes from 17 of these strains were sequenced after being cloned. The deduced OspC amino acid sequences were aligned with 12 published OspC sequences and revealed the presence of 48 conserved amino acids. On the basis of the alignment, OspC could be divided into an amino-terminal relatively conserved region and a relatively variable region in the central portion. The distance tree obtained divided the *ospC* sequences into three groups. The first group contained *ospC* alleles from all ($n = 13$) sensu stricto strains, the second group contained *ospC* alleles from seven *Borrelia afzelii* strains, and the third group contained *ospC* alleles from five *B. afzelii* and all ($n = 9$) *Borrelia garinii* strains. The ratio of the mean number of synonymous (d_S) and nonsynonymous (d_N) nucleotide substitutions per site calculated for *B. burgdorferi* sensu stricto, *B. garinii*, and *B. afzelii ospC* alleles suggested that the polymorphism of OspC is due to positive selection favoring diversity at the amino acid level in the relatively variable region. On the basis of the comparison of 16S rRNA gene sequences, *Borrelia hermsii* is more closely related to *B. afzelii* than to *B. burgdorferi* sensu stricto and *B. garinii*. In contrast, the phylogenetic tree obtained for the *B. hermsii* variable major protein, Vmp33, and 18 OspC amino acid sequences suggested that Vmp33 and OspC from *B. burgdorferi* sensu stricto strains share a common evolutionary origin.

The outer membrane protein OspC is an immunodominant surface protein of the spirochete *Borrelia burgdorferi*, which causes Lyme borreliosis. The first detectable antibody response to *B. burgdorferi* consists of immunoglobulin M antibodies to OspC as well as to the 41-kDa flagellin (29). Thus, OspC, like the flagellum (9), may become an important diagnostic antigen.

The OspC protein is also considered a vaccine candidate, since active immunization with recombinant OspC protected gerbils (22) and mice (23) against subsequent challenge with the homologous *B. burgdorferi* strain. Moreover, protection against *B. burgdorferi* challenge was observed in recent studies involving hamsters immunized with a mutant of strain 297 expressing *ospC* but lacking OspA and OspB (12). These findings strongly suggest that a Lyme borreliosis vaccine may benefit from inclusion of OspC in conjunction with other antigens.

Significant heterogeneity exists in OspC at the immunological level (25, 28) and at the genetic level (13, 17, 25). On the basis of the analysis of the nucleotide sequences from seven *B. burgdorferi* strains, we found that *ospC* sequences fall into three groups, each corresponding to one of the genospecies *B. burgdorferi* sensu stricto, *Borrelia garinii*, and *Borrelia afzelii* (25).

In order to further investigate the evolution of the *ospC* locus, *ospC* genes from 22 *B. burgdorferi* strains were cloned and sequenced and compared with *ospC* nucleotide sequences retrieved from the EMBL database. Additionally, the nucleotide sequences of the 16S rRNA genes from 17 of the *B. burgdorferi* strains were determined. The sequence data were then analyzed by using phylogenetic methods. The phylogenetic relationship to Vmp33 of the related spirochete *Borrelia hermsii* was also investigated.

* Corresponding author. Mailing address: Department of Infection-Immunology, Statens Seruminstitut, Artillerivej 5, Copenhagen, DK-2300, Denmark. Phone: (45) 32683779. Fax: (45) 32683871. Electronic mail address: theisen@biobase.dk.

MATERIALS AND METHODS

Bacterial strains, plasmids, and growth conditions. *Escherichia coli* XL1-blue (*recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac* [F′ *proAB*$^+$ *lacI*^{q}ZΔ*M15* Tn*10*]) was obtained from Stratagene (La Jolla, Calif.), and the *B. burgdorferi* strains used in the present study were from the laboratory collection. The genospecies according to the criteria of Baranton et al. (1) of the 34 *B. burgdorferi* strains are given in Table 1 and were determined as described by Marconi and Garon (16). *E. coli* strains were grown in Luria-Bertani (LB) medium (18), and *B. burgdorferi* strains were grown in BSK medium (2). When required, LB medium was supplemented with ampicillin at 100 μg/ml.

DNA techniques. Restriction enzymes and T4 DNA ligase were used as recommended by the supplier (New England Biolabs). DNA sequencing was carried out by the *Taq* DyeDeoxy termination cycle sequencing method using an ABI373 automatic sequencer.

Cloning and sequencing of *ospC*. The near-complete *ospC* genes from 22 *B. burgdorferi* sensu lato strains were amplified from genomic DNA by PCR using four primer sets: primers BF43 (5′-AAT AAT TCA TAT AAA AAG GAG GCA CA) and BF2 (25) were used with strains DK29, DK32, SL14, and SL20; primers BF44 (5′-AAA AGG AGG CAC AAA TTA ATG) and BF2 were used with DK35; primers BF46 (5′-TGC GAT ATT AAT GAC TTT ATT TTT ATT T) and BF2 were used with strain SL10; and primers BF1 (25) and BF2 were used to amplify *ospC* from the remaining strains. PCR was performed as previously described (25), and the products were inserted into the *Eco*RV site of pBluescript (Stratagene) by the TA cloning procedure (8). One strand was sequenced for at least two PCR clones of each *ospC* allele for verification.

Cloning and sequencing of the 16S rRNA gene. The near-complete 16S rRNA gene was amplified from genomic DNA by using standard PCR conditions and the degenerated primers DA73 [5′-CGG CGA ATT CAG AGT TTG ATC (C/A)TG GCT CAG] and DA76 [5′-CGG CGG ATC CAA GGA GGT G(A/T)T CCA (G/A/C)CC (C/A/G)CA]. PCR products were cloned into pBluescript by using the primer-integrated restriction enzyme sites for *Eco*RI and *Bam*HI (underlined). The sequencing primers used were 21M13 forward; M13 reverse; *Borrelia*-specific sense primers DA115 (5′-GCA ACC CTT GTT ATC TGT TAC C) (nucleotides [nt] 1107 to 1128), DA123 (5′-CGG GAG GCA GCA GCT AAG AA) (nt 344 to 363), and DA124 (5′-CGA TGC ACA CTT GGT GTT AA) (nt 816 to 835); and antisense primers DA72 [5′-GAC GTC (A/G)TC C(G/A/T/C)C (A/G/T)CC TTC CTC] (nt 1174 to 1193), DA113 (5′-TAT TAC CGC GGC TGC TGG CAC G) (nt 513 to 534), DA114 (5′-GCG AGC ATA CTC CCC AGG CGG C) (nt 97 to 998), and DA122 (5′-TAG GAG TCT GGA CCG TAT CT) (nt 324 to 343). Nucleotide numbering was obtained by aligning the sequences to the *E. coli rrnB* sequence. The nucleotide sequence of one clone derived from each strain was determined. More than 95% of the sequence was determined for both strands.

Data analysis. The *ospC* nucleotide sequences obtained were aligned with

TABLE 1. *B. burgdorferi* sensu lato strains studied in this analysis

Genospecies and strain[a]	Origin[b]	Accession no.[c]		Reference or source[d]
		ospC	16S rDNA	
B. burgdorferi sensu stricto				
DK7	Skin, ACA	X73625	X85195	Our collection
BUR	Skin	X84765	ND	R. C. Johnsson
DUN	Skin	X84778	X85201	R. C. Johnsson
TxGW	Skin	X84783	ND	R. C. Johnsson
272	Skin	X84785	X85189	A. C. Steere
297	CSF	U08284	X85204	A. C. Steere
LIP	Skin	ND	X85203	R. C. Johnsson
Kipp	Skin	X84782	X85196	R. C. Johnsson
MUL	Skin	X84779	X85200	R. C. Johnsson
B31	Tick	U01894	ND	EMBL database
DN127c19-2	Tick	U04280	ND	EMBL database
HB19	Tick	U04281	ND	EMBL database
N40	Tick	U04240	ND	EMBL database
25015	Tick	U04282	ND	EMBL database
B. garinii				
DK6	CSF, LMR	X73628	ND	Our collection
DK27	Skin, EM	X73623	X85193	Our collection
DK29	Skin, EM	X84770	X85202	Our collection
DK32	Skin, EM	X84772	ND	Our collection
DK35	Skin, EM	X84773	ND	Our collection
SL10	CSF, LMR	X84780	ND	M. Karlsson
SL14	CSF, LMR	X84784	ND	M. Karlsson
SL20	CSF, LMR	X84781	X85198	M. Karlsson
PBi	CSF, LMR	X69595	X85199	V. Preac-Mursic
B. afzelii				
DK1	Skin, EM	X73627	X85190	Our collection
DK2	Skin, ACA	X84766	X85188	Our collection
DK3	Skin, ACA	X84771	X85192	Our collection
DK4	Skin, EM	X84774	X85194	Our collection
DK5	Skin, ACA	X84775	X85197	Our collection
DK8	Skin, ACA	X84776	ND	Our collection
DK9	Skin, ACA	X84777	ND	Our collection
DK15	Skin, EM	X84768	ND	Our collection
DK21	Skin, EM	X84767	X85191	Our collection
DK22	Skin, EM	X84769	ND	Our collection
DK26	Skin, EM	X73624	ND	Our collection
PKo	Skin, EM	X62162	ND	V. Preac-Mursic

[a] Genospecies were determined by the method of Marconi and Garon (16).
[b] ACA, acrodermatitis chronica atrophicans; CSF, cerebrospinal fluid; LMR, lymphocytic meningoradiculitis; EM, erythema migrans.
[c] The sequences for the 22 strains are available for electronic retrieval from EMBL/GenBank under the accession numbers listed. rDNA, gene coding for rRNA; ND, sequence not determined.
[d] R. C. Johnsson, Minneapolis, Minn.; A. C. Steere, Boston, Mass.; M. Karlsson, Stockholm, Sweden; V. Preac-Mursic, Munich, Germany.

previously published *ospC* sequences by using the Clustal V program (11). Phylogenetic analyses were based on a distance matrix output generated by the PHYLIP version 3.42 (7) DNADIST program using the method of Jukes and Cantor which corrects for multiple base changes. Trees were constructed by using FITCH with random sequence addition and global rearrangements. DRAWTREE was used to draw phylogenetic trees, and CONSENSE was used to generate 50% majority rule consensus trees. DNA-maximum likelihood analysis was performed with the fastDNAml program. The reproducibility of tree nodes was analyzed by using the bootstrapping program SEQBOOT.

Alignment of OspC and Vmp33 sequences, determination of the percent amino acid sequence identity, and construction of the subsequent phylogenetic tree were performed by using the alignment program of Hein (10).

The numbers of synonymous nucleotide substitutions per site (d_S) and nonsynonymous nucleotide substitutions per site (d_N) were calculated by the method of Nei and Gojobori (19), using the NAG version 3.0 program (21). Since the number of allelic comparisons was too large, we present only the means of d_S and d_N of *ospC* and *ospA* for each of the three genospecies.

The Genetics Computer Group (GCG) program PlotSimilarity was used to calculate the average identity among all OspC amino acid sequences at each position in the alignment, using a sliding window of comparison of 10 amino acids, and the GCG program Antigenic was used to localize putative antigenic domains (5).

RESULTS

The deduced amino acid sequences from the *ospC* genes of the 22 strains determined in this work and the 12 strains retrieved from the EMBL database are shown in Fig. 1. The signal peptide and the amino-terminal and carboxy-terminal eight amino acid residues of the mature proteins were not included in this analysis, as they were not available for all the sequences. The alignment required both insertions and deletions, since the length of the OspC alleles varied between 175 and 181 amino acids. There are 48 completely conserved sites, the majority of which are located in the amino-terminal one-quarter of the alignment (Fig. 1).

Identities in OspC sequences range from 60 to 100%, with an average of 74.4%. Figure 2 shows the percent identity calculated for each position in the amino acid alignment using a sliding window of comparison of 10 amino acid residues. It appears that the majority of the amino-terminal 58 amino acids

FIG. 1. Alignment of the predicted amino acid sequences of OspC variants from 34 *B. burgdorferi* sensu lato isolates according to the Clustal V alignment program (11). Gaps introduced to obtain optimal homology (dashes), positions which are identical in all sequences (asterisks), and positions which differ in only one sequence (dots) are indicated. The bottom line shows positions where the amino acid sequence of *B. hermsii* Vmp33 is identical to the OspC sequences.

of the alignment are relatively conserved among the OspC alleles, whereas the central region is relatively variable. The variability is particularly located to three areas containing amino acid residues 59 to 79, 89 to 109, and 119 to 139. These areas often coincide with domains of high hydrophilicity and predicted antigenicity (data not shown).

For estimation of the evolutionary relationship among the *ospC* alleles, only the nucleotide sequence of the region coding for the amino acids shown in Fig. 1 was used. The *ospC* distance tree shown in Fig. 3A divides the 34 taxa into three major phylogenetic groups: one consisting of all 13 sensu stricto strains, one containing 7 *B. afzelii* strains, and one containing all 9 *B. garinii* strains and 5 *B. afzelii* strains. The topology shown was supported by the indicated bootstrap values. A similar result was obtained by using DNA-maximum likelihood analysis (not shown).

16S rRNA gene sequences of selected *B. burgdorferi* strains. The almost-complete sequence of a continuous stretch of 1,488 nt of the 16S rRNA gene from 17 strains was also used to analyze the evolutionary relationship of the *B. burgdorferi* sensu lato strains. The similarity between these sequences ranged from 98 to 100%. The distance tree obtained (Fig. 3B) split the 17 strains into three major lineages, each corresponding to one of the genospecies *B. burgdorferi* sensu stricto, *B. garinii*, and *B. afzelii*. The topology shown was supported by bootstrap analyses with values of 100, 100, and 92 for the sensu stricto, *B. afzelii*, and *B. garinii* groups, respectively.

Phylogenetic analysis of the conserved and variable regions of *ospC*. On the basis of the amino acid alignments (Fig. 1 and 2), two regions of *ospC* are distinguished: a relatively conserved region including codons 1 to 58 and a relatively variable region consisting of codons 59 to 150 (hereafter referred to as the conserved and variable regions, respectively). A gene tree was constructed for each of the two regions, and bootstrap values for selected nodes which are supported in more than 75 of 100 trees are indicated (Fig. 4). The phylogenetic trees show that the conserved regions fall into three major groups, each corresponding to one of the genospecies, while the variable regions do not fall into such groups. When the conserved and variable regions were subjected to DNA-maximum likelihood analysis, the same general branching order was observed (not shown).

Number of nucleotide substitutions. The mean number of nucleotide substitutions in the conserved and variable regions was estimated for each of the three genospecies. The sequences were aligned by using the Clustal V program, and nucleotide insertions were removed prior to the analysis. Table 2 shows the mean d_S and d_N values for each group of allelic comparisons. d_S is higher than d_N in the conserved region of the sensu stricto and *B. afzelii* groups, while the two parameters are of equal magnitude in the *B. garinii* group. Both d_S and d_N are much higher, and of equal magnitude, in the variable region than in the conserved region, indicating that the variable region evolves more rapidly than the conserved region. The lower d_N than d_S in the conserved region may reflect functional constraints on the OspC protein. d_S values obtained from allelic comparisons of the complete *ospA* sequence retrieved from the EMBL database are also listed in Table 2. These values are almost identical to the values obtained for the conserved region but are different from those for the variable region. Recently, d_S and d_N values for *ospA* were also calculated by Caporale and Kocher (3); however, those authors did not distinguish between alleles from strains belonging to different genospecies.

Relatedness of OspC to *B. hermsii* Vmp. It was recently shown by Carter et al. (4) that Vmp33 from *B. hermsii* is related to OspC from PKo and B31. To further analyze the relationship between OspC and Vmp33, the amino acid sequences of the mature OspC proteins from 18 *B. burgdorferi* strains were aligned with Vmp33 from *B. hermsii*. We find that OspC is on the average 36% identical to Vmp33 (Fig. 1). Amino acid identities were calculated on the basis of this alignment. Of the 48 amino acids completely conserved among the OspC sequences, 25 are also present in Vmp33 (Fig. 1).

```
                    ---------------------->
25015       HTE---LGKQDAQDDDAKKAILRTHNT-KDKGAEELDKLFKAVENLSKAAKEMLSNSVKELTSPVV
272         HAQ---LGIENVTDENAKKAILITDAA-KDKGAAELEKLFKAVENLPKAAKEMLANSVKELTSPIV
297         HAQ---LGIEDVTDENANEAILTTDAA-KDKGAAEREKLFKAVENLSKAAKEMLANSVKELTSPIV
B31         HTD--SFGKEGVTDADAKEAILKTNGT-KTKGAEELGKLFESVEVLSKAAKEMLANSVKELTSPVV
BUR         HAQ---LGIQGVTDENAKKAILEANAAGKDKGVEELEKLSGSLESLSKAAKEMLANSVKELTSPVV
DK1         HAELGVAGNGATTDEDAQKAILKTNAN-NDKGAKELKELFESVESLAKAAKESLTNSVKELTSPVV
DK15        NADLGKHD-A--TDADAKKAILKTDAT-KDKGAKELEELFKSVESLSKAAKEALSNSVKELTSPVV
DK2         HADLGKQD-A--TDDHAKAVILKTHAT-TDKGAKEFKDLFESVEGLLKAAQVALTNSVKELTSPVV
DK21        HADLGKQD-A--TDDHAKAAILKTHAT-TDKGAKEFKDLFESVEGLLKAAQVALTNSVKELTSPVV
DK22        HAELGVAGNGATTDENAQKAIFNTNAN-NDKGAKELKELFESVESLAKAAKEALSNSVKELTSPVV
DK26        HADLGKQD-A--TDDHAKAAILKTHAT-TDKGAKEFKDLFESVEGLLKAAQVALTNSVKELTSPVV
DK27        HAQLGAVG-GAINDDRAKEAILKTHGT-NDKGAKELKELSESVESLAKAAQAALANSVKELTSPVV
DK29        HAQLGAVG-GAINDDRAKEAILKTHGT-NDKGAKELKELSESVESLAKAAQAALANSVKELTSPVV
DK3         HADLAKQD-A--TDDHAKAAILKTHAT-TDKGAKEFKDLFESVEGLLKAAQVALTNSVKELTSPVV
DK32        HAQLGAVG-GAINDDRAKEAILKTHGT-NDKGAKELKELSESVESLAKAAQAALANSVKELTSPVV
DK35        HQDLGTAG-GDATDDHPKAAILKTNAT-DDKGAKELKELFESVESLSKAAKAALANSVKELTSPVV
DK4         HAELGAAN-GATTDENAKAAILKTHGT-KDKGATELGELFKSVESLSKAAKESLTNSVKELTSPVV
DK5         HADLGKQG-VN--DDDAKKAILNTYAD-QTKGAEELGKLFKSVEGLVKAAQEALTNSVKELTSPVV
DK6         HAELGIQS---VQDDNAKKAILKTHGT-KDKGAKELEELFKSLESLSKAAQAALTNSVKELTNPVV
DK7         HAQ---LGIQGVTDENAKKAILKANAAGKDKGVEELEKLSGSLESLSKAAKEMLANSVKSLQS-VV
DK8         HADLGKQD-A--TDDHAKAAILKTHAT-TDKGAKEFKDLFESVEGLLKAAQVALTNSVKELTSPVV
DK9         HADLGKQG-VN--DDDAKKAILKTNAD-KTKGAEELGKLFKSVEGLVKAAQEALTNSVKELTSPVV
DN127       HAELGVAG-GATTDENAKKAILKSNAD-KTKGADELGKLFESVESLAKAAKEMLANSVKELTSPVV
DUN         NAQ---LNQANANDANAKAAILKTHNT-KDKGAEELVKLAESVAGLFKVAQEMLNNSVKELTSPVV
HB19        HTD---LGKEGVTDDNAKKAILKTNN-DKTKGADELEKLFESVKNLSKAAKEMLTNSVKELTSPVV
KIPP        HAQ---LGIENVTDENAKKAILITDAA-KDKGAAELEKLFKAVENLPKAAKEMLANSVKELTSPIV
MUL         HAQ---LGIENVTDENAKKAILITDAA-KDKGAAELEKLFKAVENLPKAAKEMLANSVKELTSPIV
N40         HA---VLGLDNLTDDNAQRAILKKHA-NKDKGAAELEKLFKAVENLSKAAQDTLKNAVKELTSPIV
PBI         HAELGIQS---VQDDNAKKAILKTHGT-KDKGAKELEELFKSLESLSKAAQAALTNSVKELTNPVV
PKO         HADLGKQD-A--TDDHAKAAILKTHAT-TDKGAKEFKDLFESVEGLLKAAQVALTNSVKELTSPVV
SL10        HAQHGVAA-A--TDDHAKEAILKSNPT-KDKGAKELKDLSESVESLAKAAQEALANSVKELTNPVV
SL14        HAELGIQN---VQDDNAKRAILKTHGN-KDKGAKELKELSESLEKLSKAAQAALANSVQELTSPVV
SL20        HAELGIQN---VQDDNAKRAILKTHGN-KDKGAKELKELSESLEKLSKAAQAALANSVQELTSPVV
TXGW        HTD---LGKRGVTDADAKEAILKTNGT-KTKGAEELGKLFESVEVLSKAAKEMLANSVKELTSPVV
            .           *     *         **  *   *      * *.*   * * *  *    *
Vmp33       *           *     *         **  *   *      *  **
```

FIG. 1—*Continued.*

A phylogenetic tree based on the amino acid sequences of the mature proteins illustrates the relatedness of OspC and Vmp33 (Fig. 5). The phylogenetic tree suggests that the six OspC alleles from the sensu stricto strains share a common evolutionary ancestor with Vmp33.

DISCUSSION

Outer surface protein OspC of *B. burgdorferi* has recently received much attention as a diagnostic antigen and as a potential vaccine candidate for Lyme borreliosis. On the basis of

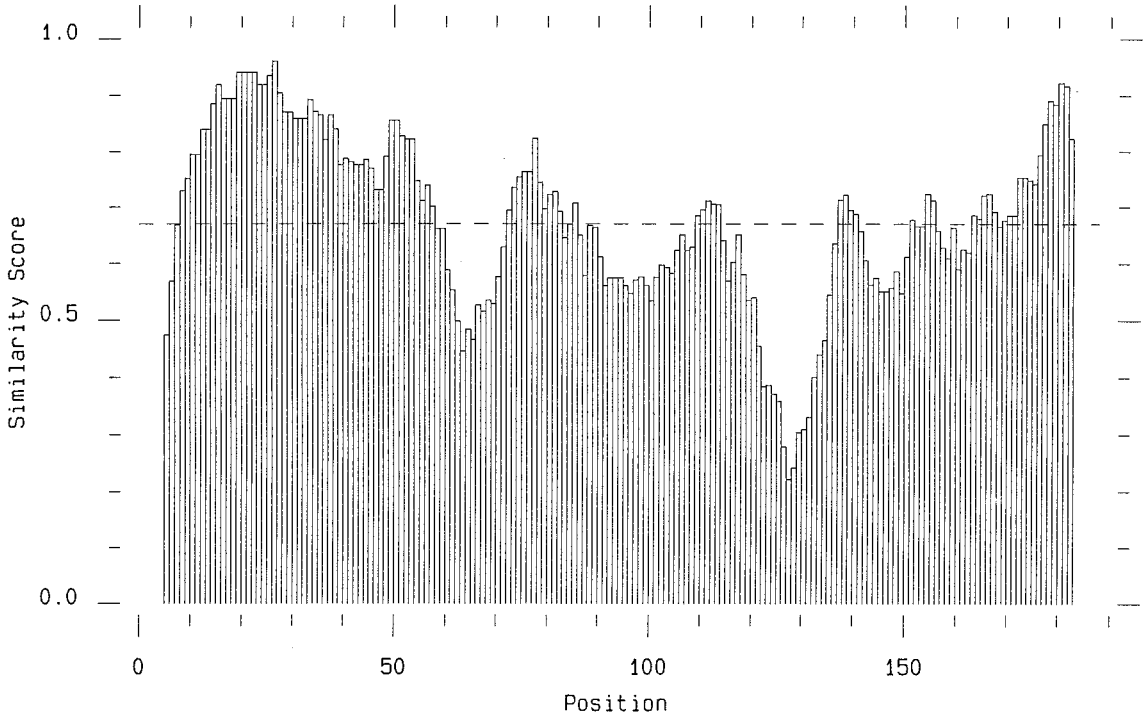

FIG. 2. Plot of the identity score versus alignment position number from the alignment in Fig. 1. The plot was produced by using the program PlotSimilarity with a sliding window of comparison of 10 amino acids. Numbers on the *x* axis indicate codon positions relative to the ninth codon in the mature OspC protein of PKo.

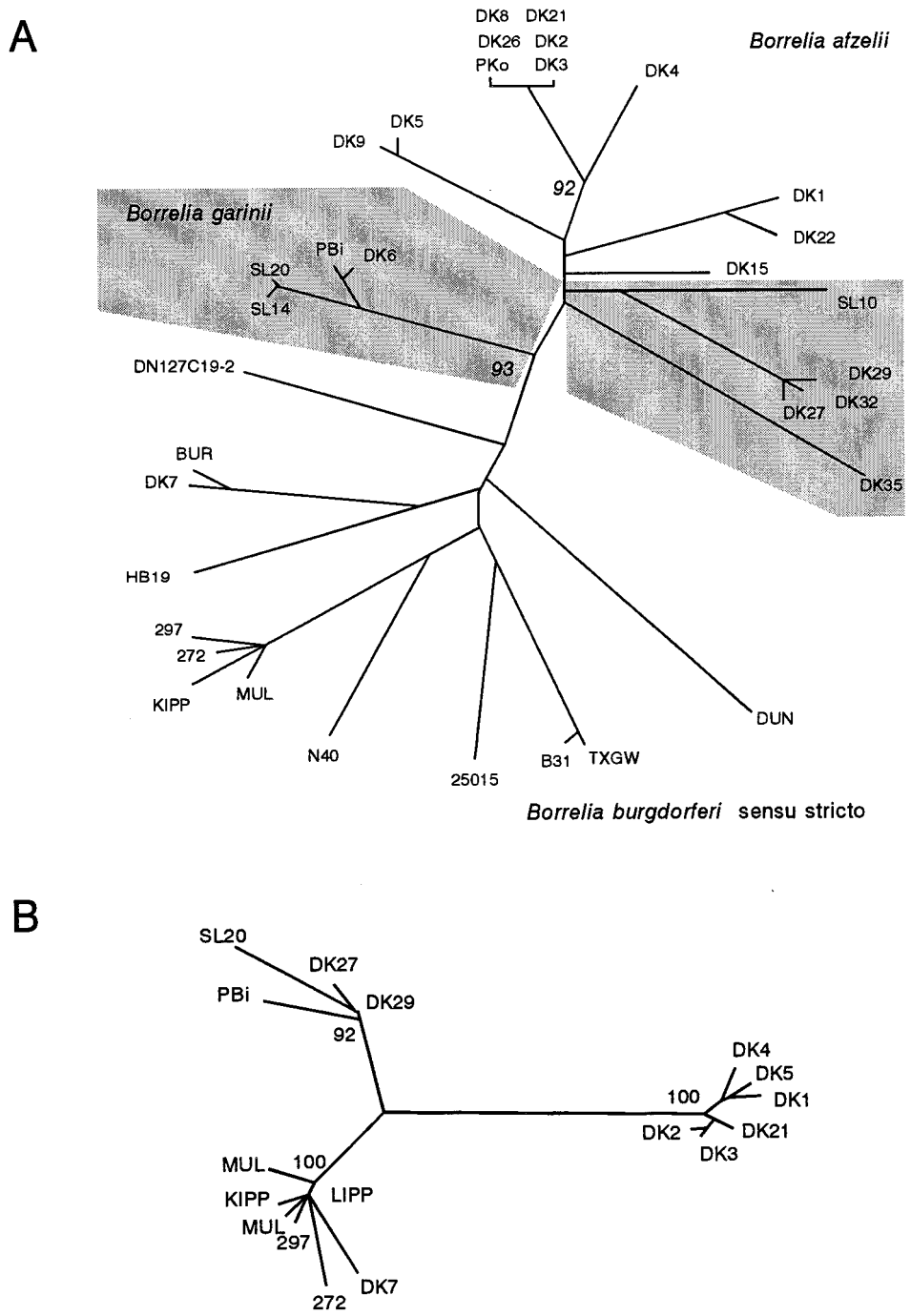

FIG. 3. Evolutionary-distance tree based on *ospC* nucleotide sequences for the 34 *B. burgdorferi* sensu lato strains listed in Table 1 (A) and for 16S rRNA genes from 17 of these *B. burgdorferi* strains (B). Bootstrap values are shown for selected nodes which are supported in more than 75 of 100 trees. Branch lengths reflect evolutionary distances.

comparison of a few *ospC* alleles, a high degree of polymorphism in *ospC* was observed (13, 17, 25). In order to further analyze this polymorphism and the evolution of the *ospC* gene, we have cloned and sequenced 22 *ospC* alleles from *B. burgdorferi* strains isolated from humans and compared them with 12 *ospC* alleles retrieved from the EMBL database. The almost-complete sequences of the 16S rRNA genes from 17 of the strains were also determined and used as an independent phylogenetic marker.

Recently, Dykhuizen et al. suggested that *B. burgdorferi* is clonal, which means that all genes belonging to the same strain share a common evolutionary history in the absence of recombination (6). According to this hypothesis, the *ospC* alleles should have fallen into three monophyletic groups, each corresponding to one of the three genospecies. This was not observed when the near-complete *ospC* sequence was subjected to phylogenetic analysis, as *ospC* alleles from five *B. afzelii* strains fell into a cluster with *ospC* alleles from all of the *B.*

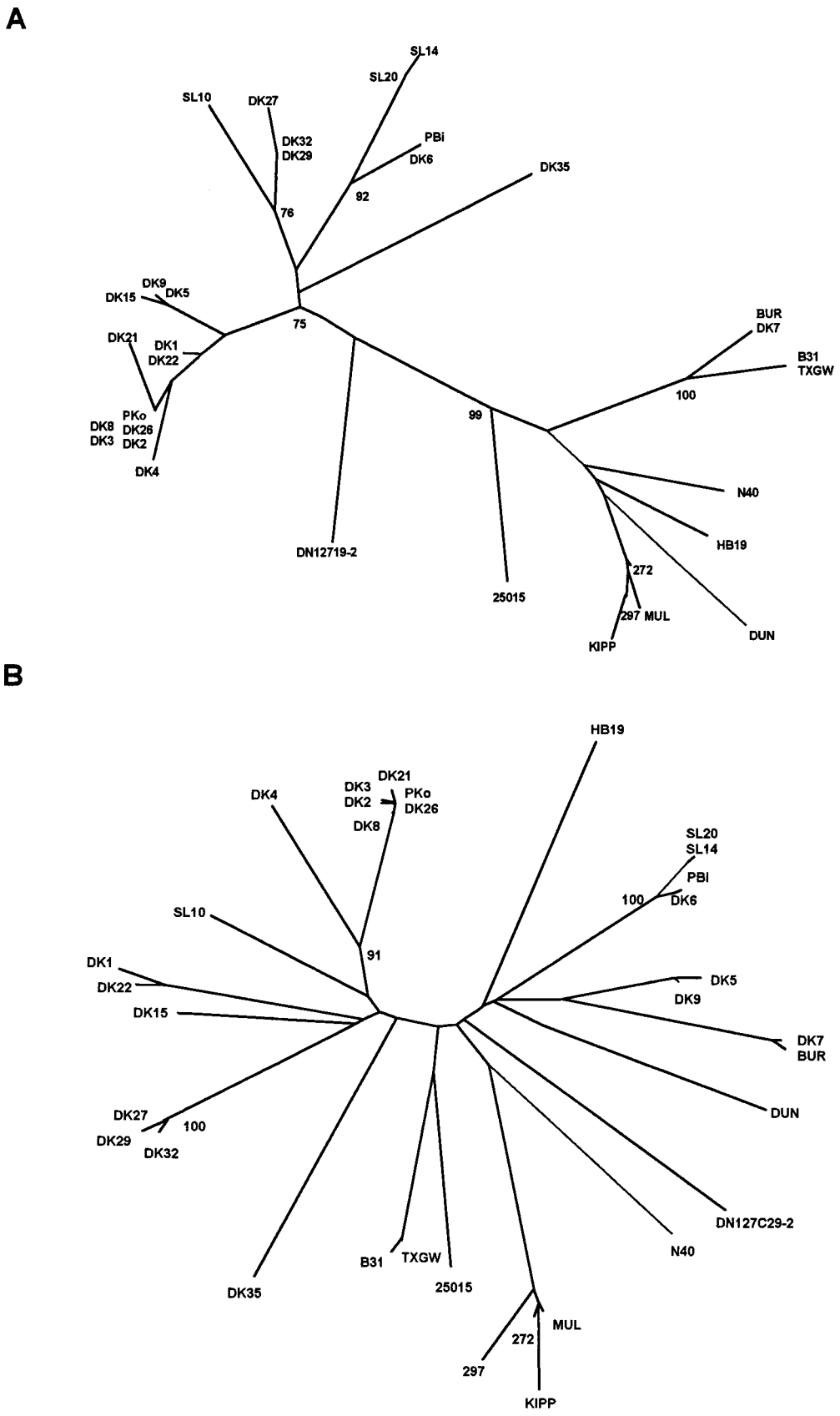

FIG. 4. Distance trees for the conserved region (A) and variable region (B) of the *ospC* variants shown in Fig. 3. Bootstrap values are shown for selected nodes which are supported in more than 75 of 100 trees. Branch lengths reflect evolutionary distances.

TABLE 2. Comparison of d_S and d_N values per site for the conserved and variable regions of *ospC* and for *ospA*[a]

Strain (no. *ospC*/no. *ospA*)[b]	*ospC*				*ospA*	
	Conserved region		Variable region			
	d_S	d_N	d_S	d_N	d_S	d_N
B. burgdorferi sensu stricto (10/3)	0.196 ± 0.055	0.067 ± 0.055	0.384 ± 0.063	0.255 ± 0.030	0.191 ± 0.031	0.114 ± 0.014
B. garinii (9/6)	0.089 ± 0.029	0.072 ± 0.015	0.262 ± 0.050	0.236 ± 0.030	0.073 ± 0.014	0.043 ± 0.006
B. afzelii (10/3)	0.058 ± 0.024	0.022 ± 0.008	0.226 ± 0.047	0.170 ± 0.024	0.011 ± 0.007	0.007 ± 0.003

[a] The numbers of substitutions per site were estimated according to the method of Jukes and Cantor. Values are means ± standard errors of the average (calculated by the method of Ota and Nei [21]).

[b] *ospC* nucleotide sequences from the following strains were used in this analysis: *B. burgdorferi* sensu stricto DK7, BUR DUN, TxGW, 272, 297, LIP, Kipp, and MUL; *B. garinii* DK6, DK27, DK29, DK32, DK35, SL10, SL14, SL20, and PBi; and *B. afzelii* DK1, DK3, DK4, DK5, DK8, DK9, DK15, DK21, DK22, and DK26. The entire *ospA* sequence retrieved from the EMBL databases was analyzed from *B. burgdorferi* sensu stricto PKA1, ZS7, and 19857; *B. garinii* DK29, G25, K48, PBi, PHei, PTROb, and ZQ1; and *B. afzelii* IP3, PGau, and PKo.

garinii strains (Fig. 3A). The topology of the 16S rRNA gene tree, however, confirmed that the strains analyzed fall into groups corresponding to each of the three genospecies. The trees for the conserved region of *ospC* and the 16S rRNA genes share 16 strains (Fig. 4A and 3B), and the topological relationships of these 16 strains are identical. This similarity of topology suggests that these two sets of sequences have evolved together. The apparent contrasting patterns of evolutionary divergence of *ospC* compared with 16S rRNA and other *B. burgdorferi* genes may be explained by unequal rates of evolution. The notion that *ospC* evolves faster than other *B. burgdorferi* genes analyzed is supported by the observation that both d_S and d_N are clearly higher in the variable region of *ospC* than in *ospA* (Table 2). Thus, the similarity among *ospC* alleles from *B. afzelii* and *B. garinii* strains may be due to a high mutation rate resulting in parallel and convergent substitutions.

Another possible explanation for the apparent difference in evolutionary history could be that horizontal transfer and recombination in the variable regions of *ospC* genes from distantly related strains would alter the phylogenetic history of *ospC*. However, horizontal transfer of *ospC* among *B. burgdor-*

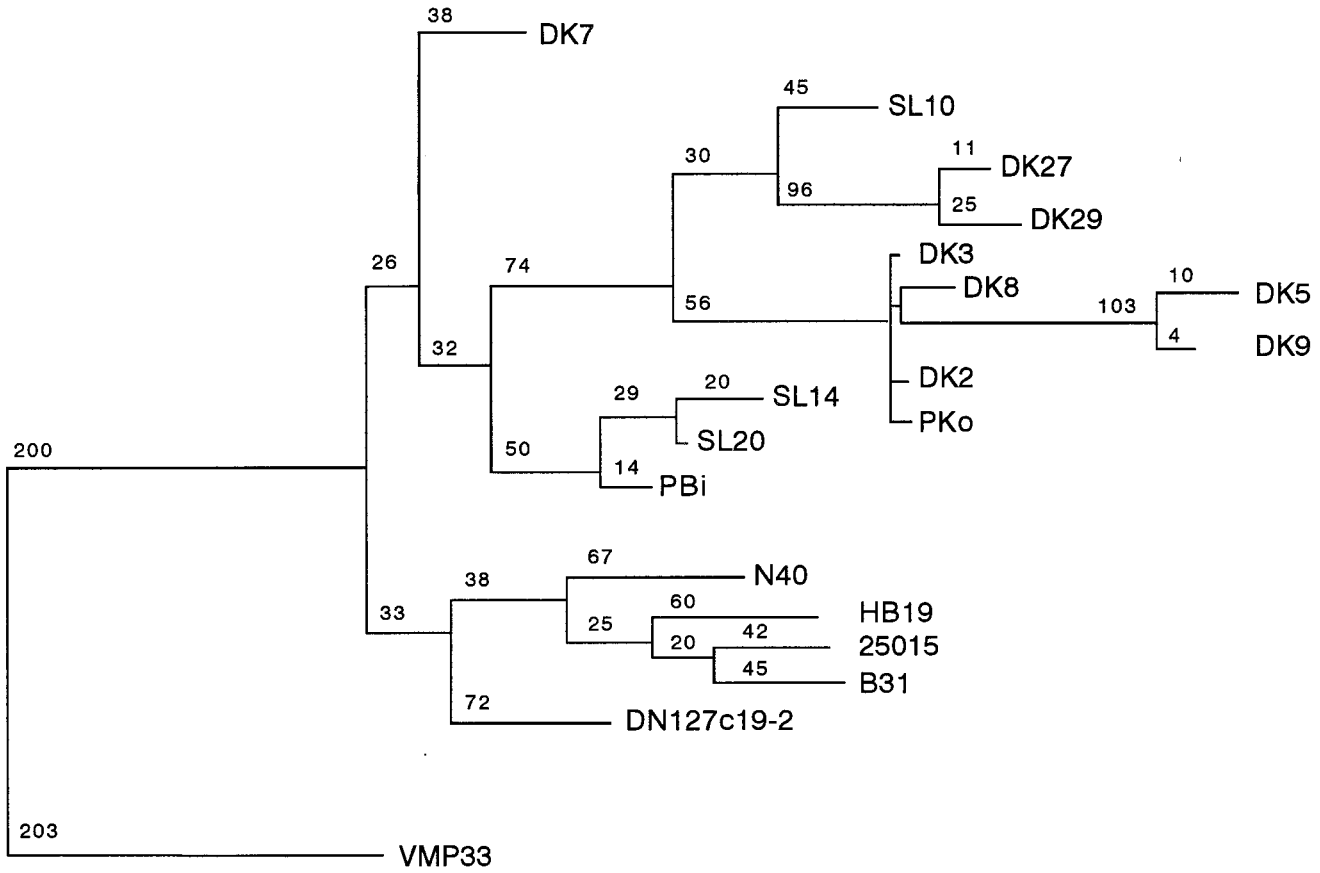

FIG. 5. Phylogenetic tree of *B. hermsii* Vmp33 and OspC variants from *B. burgdorferi*, determined by the method of Hein (10). The Vmp33 sequence was retrieved from the EMBL database. Only the amino acid sequences of the mature proteins were used in this analysis. Branch lengths were determined on the basis of the amino acid alignment; numbers indicate relative evolutionary distances calculated according to the values from the distance matrix. The tree was rooted by midpoint rooting.

feri strains would produce gene pairs with a conspicuously low level of synonymous divergence, but this is generally not the case for the variable region. In those instances in which the d_S value for a particular gene pair is low in the variable region, it is equally low in the conserved region, indicating a close overall relatedness rather than recombination. When calculating d_S and d_N values for individual *ospC* gene pairs among *B. afzelii* strains (DK22, DK5, DK9, DK3, and DK4), *B. garinii* strains (DK6, DK27, DK29, SL10, and SL14), and *B. burgdorferi* sensu stricto strains (BUR, B31, 297, DK7, and HB19), we found values close to the means for each of the groups (not shown). Thus, intraspecies gene exchange does not appear to be an important confounding factor here.

A positive selection for diversity in the variable region of OspC would be indicated by a d_S/d_N ratio of <1. According to Table 2, this was not the case. However, when the d_S/d_N ratios of the conserved and variable regions of OspC from *B. burgdorferi* sensu stricto and *B. afzelii* were compared, a marked decrease was observed. This finding suggests the possibility of positive selection favoring diversity at the amino acid level in the variable region of OspC. It is also noteworthy that in Table 2, the d_S/d_N ratio is approaching 1 in the case of both the conserved and the variable regions for *B. garinii*. This may suggest a stronger evolutionary pressure on both regions of OspC in *B. garinii* than in *B. burgdorferi* sensu stricto and *B. afzelii*. This hypothesis would be compatible with the previous findings of a relatively greater diversity of *B. garinii* OspA epitopes (27) and greater diversity of *B. garinii* 16S rRNA, HSP60, and *ospA* genes (1, 14, 26). The ratios of d_S between the two regions in the *B. garinii* and *B. afzelii* alleles are 2.9 and 3.9, respectively. The simplest explanation for these elevated d_S values would be recombination between distantly related sequences (15, 20).

We have also compared the *B. burgdorferi* 16S rRNA gene sequences with the *B. hermsii* 16S rRNA gene sequence retrieved from the EMBL database. The phylogenetic tree obtained shows that *B. hermsii* is more closely related to *B. afzelii* than to *B. garinii* and *B. burgdorferi* sensu stricto (data not shown). This contrasts with the phylogenetic tree obtained for the OspC and *B. hermsii* Vmp33 sequences (Fig. 5) which suggests that OspC from *B. burgdorferi* sensu stricto and Vmp33 share a common evolutionary ancestor. However, since the 16S rRNA gene is more conserved than Vmp33 and OspC, the phylogeny inferred from the 16S rRNA sequences is more likely to reflect the evolutionary relationship between *B. burgdorferi* and *B. hermsii.*

Since the variable regions of OspC most often coincide with areas with a high antigenicity index, it may be speculated that these regions are recognized by the host immune system and therefore are subjected to selection for diversity. Antigenic variation may allow a population of spirochetes to avoid the immune system and persist in the host, or it could have some adaptive function in permitting reinfection of hosts. Using a mouse infection model, Stevenson et al. failed to detect nucleotide sequence alterations in *ospC* during a chronic infection (24). However, our finding that the d_S/d_N ratio is much smaller for the variable region than the conserved region of OspC suggests the possibility that antibody pressure influences evolution of this variable region.

ACKNOWLEDGMENTS

We thank M. Paulli Andersen, K. Lindbo, and J. Severinsen for technical assistance and J. Hein for critical reading of the manuscript.

This work was supported by the Research Center for Medical Biotechnology under the Danish Biotechnological Research and Development program.

REFERENCES

1. **Baranton, G., D. Postic, I. Saint Girons, P. Boerlin, J.-C. Piffaretti, M. Assous, and P. A. D. Grimont.** 1992. Delineation of *Borrelia burgdorferi* sensu stricto, *Borrelia garinii* sp. nov., and group VS461 associated with Lyme borreliosis. Int. J. Syst. Bacteriol. **42:**378–383.
2. **Barbour, A. G.** 1984. Isolation and cultivation of Lyme disease spirochetes. Yale J. Biol. Med. **57:**521–525.
3. **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.
4. **Carter, C. J., S. Bergström, S. J. Norris, and A. G. Barbour.** 1994. A family of surface-exposed proteins of 20 kilodaltons in the genus *Borrelia*. Infect. Immun. **62:**2792–2799.
5. **Devereux, J., P. Haeberli, and O. Smithies.** 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12:**387–395.
6. **Dykhuizen, D. E., D. S. Polin, J. J. Dunn, B. Wilske, V. Preac-Mursic, R. J. Dattwyler, and B. D. Luft.** 1993. *Borrelia burgdorferi* is clonal: implications for taxonomy and vaccine development. Proc. Natl. Acad. Sci. USA **90:**10163–10167.
7. **Felsenstein, J.** 1985. Confidence limits on phylogenies with a molecular clock. Syst. Zool. **34:**152–161.
8. **Finney, M.** 1994. Molecular cloning of PCR products, p. 15.7.1–15.7.6. *In* Current protocols in molecular biology. John Wiley and Sons, Inc., New York.
9. **Hansen, K., P. Hindersson, and N. S. Pedersen.** 1988. Measurement of antibodies to the *Borrelia burgdorferi* flagellum improves serodiagnosis in Lyme disease. J. Clin. Microbiol. **26:**338–346.
10. **Hein, J.** 1990. Unified approach to alignment and phylogenies. Methods Enzymol. **183:**626–645.
11. **Higgins, D. G., and P. M. Sharp.** 1990. Fast and sensitive multiple sequence alignments on a microcomputer. Comput. Appl. Biosci. **5:**151–153.
12. **Hughes, C. A. N., 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.
13. **Jauris-Heipke, S., R. Fuchs, M. Motz, V. Preac-Mursic, E. Schwab, E. Soutschek, G. Will, and B. Wilske.** 1993. Genetic heterogeneity of the genes coding for the outer surface protein C (OspC) and the flagellin of *Borrelia burgdorferi*. Med. Microbiol. Immunol. **182:**37–50.
14. **Lebech, A.-M., K. Hansen, B. Wilske, and M. Theisen.** 1995. Taxonomic classification of 29 *Borrelia burgdorferi* strains isolated from patients with Lyme borreliosis: a comparison of five different phenotypic and genotypic typing schemes. Med. Microbiol. Immunol. **183:**325–341.
15. **Li, J., K. Nelson, A. C. McWhorter, T. S. Whittam, and R. K. Selander.** 1994. Recombinational basis of serovar diversity in *Salmonella enterica*. Proc. Natl. Acad. Sci. USA **91:**2552–2556.
16. **Marconi, R. T., and C. F. Garon.** 1992. Development of polymerase chain reaction primer sets for diagnosis of Lyme disease and for species-specific identification of Lyme disease isolates by 16S rRNA signature nucleotide analysis. J. Clin. Microbiol. **30:**2830–2834.
17. **Margolis, N., D. Hogan, W. Cieplak, Jr., T. G. Schwan, and P. A. Rosa.** 1994. Homology between *Borrelia burgdorferi* OspC and members of the family of *Borrelia hermsii* variable major proteins. Gene **143:**105–110.
18. **Miller, J. H.** 1972. Experiments in molecular genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
19. **Nei, M., and T. Gojobori.** 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. **3:**418–426.
20. **Ohta, T., and C. J. Basten.** 1992. Gene conversion generates hypervariability at the variable regions of kallikreins and their inhibitors. Mol. Phylogenet. Evol. **1:**87–90.
21. **Ota, T., and M. Nei.** 1994. Variance and covariances of the number of synonymous and nonsynonymous substitutions per site. Mol. Biol. Evol. **11:**613–619.
22. **Preac-Mursic, V., B. Wilske, E. Patsouris, S. Jauris, G. Will, W. E. Soutschek, S. Rainhardt, G. Lehnert, U. Klockmann, and P. Mehraein.** 1992. Active immunization with pC protein of *Borrelia burgdorferi* protects gerbils against *B. burgdorferi* infection. Infection **20:**342–349.
23. **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.
24. **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.
25. **Theisen, M., B. Frederiksen, A.-M. Lebech, J. Vuust, and K. Hansen.** 1993. Polymorphism in the *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.
26. **Wallich, R., C. Helmers, U. E. Schaibel, Y. Lobet, S. E. Moter, M. D.**

Kramer, and M. M. Simon. 1992. Evaluation of genetic divergence among *Borrelia burgdorferi* isolates using OspA, *fla*, HSP60, and HSP70 gene probes. Infect. Immun. **60:**4856–4866.

27. **Wilske, B., V. Preac-Mursic, U. B. Göbel, B. Graf, S. Jauris, E. Soutschek, E. Schwab, and G. Zumstein.** 1993. An OspA serotyping system for *Borrelia burgdorferi* based on reactivity with monoclonal antibodies and OspA sequence analysis. J. Clin. Microbiol. **31:**340–350.
28. **Wilske, B., V. Preac-Mursic, S. Jauris, A. Hofmann, I. Pradel, E. Soutschek, E. Schwab, G. Will, and G. Wanner.** 1993. Immunological and molecular polymorphisms of OspC, an immunodominant major outer surface protein of *Borrelia burgdorferi*. Infect. Immun. **61:**2182–2191.
29. **Wilske, B., V. Preac-Mursic, G. Schierz, G. Liegl, and W. Gueye.** 1989. Detection of IgM and IgG antibodies to *Borrelia burgdorferi* using different strains as antigen. Zentralbl. Bakteriol. Suppl. **18:**299–309.