Sequence Variation in the Outer-Surface-Protein Genes of *Borrelia burgdorferi*

Diane A. Caporale and Thomas D. Kocher
Department of Zoology and Program in Genetics, University of New Hampshire

Borrelia burgdorferi is a spirochete pathogen transmitted among warm-blooded hosts by ixodid ticks. Frequency-dependent selection for variant outer-surface proteins might be expected to arise in this species, since rare variants are more likely to avoid immune surveillance in previously infected hosts. We sequenced the OspA and OspB genes of nine North American strains and compared them with nine strains previously described. For each gene, the mean number of synonymous substitutions per synonymous site and the mean number of nonsynonymous substitutions per nonsynonymous site show only a twofold excess of silent mutations. Synonymous rates vary widely along the OspB protein. Some regions show a significant excess of silent substitutions, while divergence in other regions is constrained by biased base composition or selection. The presence, in antigenically important regions of the protein, of significant variation among strains, as well as evidence for recombination among strains, should be considered in attempts to develop vaccines against this disease.

Introduction

Lyme disease first gained national attention in the early 1970s and is now the most common arthropod-borne disease in the United States. It was not until the early 1980s that the causative agent was shown to be a spirochete, *Borrelia burgdorferi* (Burgdorfer et al. 1982). Thousands of cases of human infection have been reported in the United States, concentrated in the Northeast (New York, New Jersey, Connecticut, and Massachusetts), the upper Midwest (Minnesota and Wisconsin), and the Pacific Northwest (California and Oregon) (Ciesielski et al. 1988). Human symptoms range from arthritis to neurological and cardiac abnormalities (Steere et al. 1977; Burgdorfer et al. 1982; Prasad 1991).

The recent upsurgence of this disease may be due to the expansion of deer populations into areas which have recently reverted from farmland to deciduous forests (Spielman et al. 1985; Barbour and Fish 1993). Although deer do not transmit *B. burgdorferi* (Telford et al. 1988), they are important in maintaining populations of the *Ixodes* tick vectors. In the United States, the spirochete has been found mainly in *I. dammini* and *I. pacificus*. These ticks parasitize a broad range of vertebrate hosts, including 29 species of mammals and 49 species of birds (Anderson 1988). Immature *I. dammini* tend to concentrate on rodent reservoirs, while adult ticks often feed on deer.

Pathogens such as *B. burgdorferi,* which need to survive long periods in their hosts in order to be transmitted, experience strong selective pressures to develop mechanisms for escaping the immune system (Borst 1983; Hagblom et al. 1985; Plasterk et al. 1985). *Borrelia hermsii,* a causative agent of relapsing fever and a close relative of *B. burgdorferi* (Barbour et al. 1982), successfully evades the immune system by presenting a series of antigenic variants of its outer-surface proteins. During the course of infection, a small number of cells undergo recombination between multiple loci, leading to the expression of a distinct surface protein (Plasterk et al. 1985), which allows these cells to avoid existing immune response. Although *B. burgdorferi* does not have multiple copies of each outer-surface-protein gene type, recombination between outer-surface-protein genes A and B during serial passage in the laboratory has been described (Rosa et al. 1992).

Vaccines being developed today are directed at the major-outer-surface proteins on the spirochete (Fikrig et al. 1990, 1992). The outer surface of *B. burgdorferi* plays an important role in host-parasite interactions and in the ability to establish an infection in a variety of hosts (Schwan et al. 1988). The dominant proteins (denoted "Osp proteins") are encoded by the OspA and

Key words: Lyme disease, *Borrelia burgdorferi,* outer-surface-protein genes, polymerase chain reaction, direct sequencing, positive selection.

Address for correspondence and reprints: Thomas D. Kocher, Department of Zoology and Program in Genetics, University of New Hampshire, Durham, New Hampshire 03824.

Mol. Biol. Evol. 11(1):51–64. 1994.

0737-4038/94/1101-0006$02.00

Table 1
***Borrelia burgdorferi* Isolates Used in Present Study**

Ref Code[a]	Strain	Location	Host	Source
CT1	HB19	Connecticut	Human	A. G. Barbour, University of Texas, San Antonio
CT2	27985	Stamford, Conn.	*Ixodes dammini*	J. F. Anderson, Connecticut Agricultural Experiment Station, New Haven
NY2	19535	Caldor, N.Y.	*Peromyscus leucopus*	J. F. Anderson
NY3	42373	East Hampton, N.Y.	*I. dammini*	J. F. Anderson
MA	41552	Newton, Mass.	*I. dammini*	J. F. Anderson
WI	21343	Fort McCoy, Wis.	*P. leucopus*	J. F. Anderson
CA3	CA3	California	*I. pacificus*	R. S. Lane, University of California, Berkeley
CA7	CA7	California	*I. pacificus*	R. S. Lane
CA8	CA8	California	*I. pacifucus*	R. S. Lane
NY1	B31	Shelter Island, N.Y.	*I. dammini*	Bergstrom et al. 1989
NY4	25015	Millbrook, N.Y.	*I. dammini*	Fikrig et al. 1992
SWED ...	ACA1	Stockholm	Human skin	Jonsson et al. 1992
RUSS	Ip90	Russia	*I. persulcatus*	Jonsson et al. 1992
GER1 ...	ZS7	Freiberg	*I. ricinus*	R. Wallich; GenBank X16467
GER2 ...	PKo	Munich	Human skin	G. Zumstein; GenBank S48322
GER3 ...	ZQ1	Freiberg	*I. ricinus*	R. Wallich; GenBank X66065
GER4 ...	B29	Berlin	*I. ricinus*	W. Fellinger; GenBank M88764
GER5 ...	GO2	Germany		H. Eiffert; GenBank S99475

[a] For simplicity, reference codes have been used to designate to each isolate.

OspB genes, which are tandemly arrayed on linear plasmids and are cotranscribed (Howe et al. 1986; Barbour and Garon 1988; Bergstrom et al. 1989). Osp proteins have been found to vary among strains in their apparent molecular weights and antigenicities (Barbour et al. 1985).

Here we report the DNA sequence of the Osp A and B genes from nine North American isolates of *B. burgdorferi*. Together with the sequences of nine additional strains which have been previously described, these sequences allow us to examine the hypothesis that positive selection acts to accelerate the evolution of outer-surface-protein genes. To test for evidence of positive selection, rates of synonymous and nonsynonymous nucleotide substitution are compared (Hughes and Nei 1988). The finding of significant positive selection for these molecules would have important implications for the development of vaccines against Lyme disease.

The sequences also allow us to investigate the geographic structure of *B. burgdorferi*. Since the dispersal of ticks by their rodent and deer hosts is limited, we expect some geographic structuring of *B. burgdorferi* populations. If, however, birds play a significant role in tick dispersal, then a lack of geographic structuring would be seen among *B. burgdorferi* populations. Although the amplification of *B. burgdorferi* DNA from nymph or adult ticks in museum collections suggests that Lyme disease is not a recent introduction to this continent (Persing et al. 1990), it is still possible that the recent upsurgence of Lyme disease is due to the spread of a particularly virulent bacterial strain. If isolates from across the country are identical in sequence, this fact would provide support for the recent spread of a single strain.

Material and Methods

Bacterial Strains

The bacterial strains used in this study, along with their hosts and sources, are listed in table 1. Nine of these strains had not been previously sequenced and were the focus of our laboratory effort. The strains were maintained in BSK II medium at 34°C as described elsewhere (Barbour 1984). Alterations in Osp DNA sequences have been found in serial passaging (Rosa et al. 1992). To ensure that the DNA sequence obtained would be unmodified from that of the original isolate, the isolates have only been passaged one to seven times from the time of isolation until DNA extraction.

DNA Isolation

Total DNA was purified from a 15-ml stationary-phase borrelial culture, using a standard organic extraction protocol (Sambrook et al. 1989). Spirochetes were pelleted from each culture and were resuspended in 100

μl of TE (10 mM Tris pH 8.0, 1 mM ethylenediaminetetraacetate [EDTA]). A 500-μl amount of extraction buffer (10 mM Tris, 2 mM EDTA pH 8.0, 10 mM NaCl, 1% sodium dodecyl sulfate, 8 mg of dithiothreitol/ml, and 0.4 mg of proteinase K/ml) was added to each sample, and the mixture was incubated at 37°C for ~1 h. The DNA was extracted once with phenol, once with a 1:1 phenol:chloroform mixture, and finally with chloroform alone. The DNA was precipitated with 2 volumes of 100% ethanol, washed with 70% ethanol, and suspended in 500 μl of TE. DNA concentration was determined fluorometrically using 0.1 μg Hoechst dye #33258/ml, in a TK100 fluorometer (Hoeffer).

Oligonucleotide Primers

The targets for PCR amplification were the gene encoding the OspA protein and the gene encoding the OspB protein. The complete nucleotide sequence of the OspA and OspB genes within the B31 strain has been determined elsewhere (Bergstrom et al. 1989). Primer sequences are presented in table 2, and their locations within the OspA/OspB operon are shown in figure 1. All primers were synthesized using an Applied Biosystems 381A DNA synthesizer.

Polymerase Chain Reaction (PCR)

Whole and partial OspA and OspB gene segments were amplified by PCR. The partial segments overlapped by ~300 bases, totaling three segments per gene. Ten nanograms of genomic DNA were added to a 50-μl PCR mixture containing 5 μl of 10 × *Taq* buffer (670 mM Tris pH 8.8, 20 mM $MgCl_2$, 98 mM B-mercaptoethanol, 0.1% Tween-20), 1 μl of each deoxynucleoside triphosphate (10 mM stock), 0.2 μl of *Taq* DNA polymerase (5 units/μl), and 5 μl of each primer (10 μM stock).

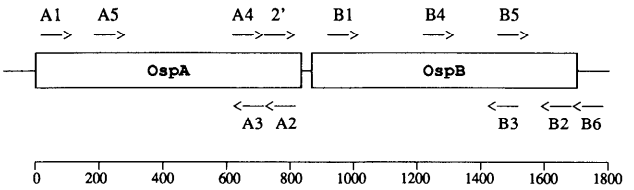


FIG. 1.—Map of the OspA/OspB operon and location of primers used in PCR amplification. The orientation of each primer is shown, with the arrow pointing in the 5'-3' direction.

Reactions were performed in either an Autogene I (Grant) or a Perkin-Elmer model 480 thermal cycler. Components were denatured at 93°C for 30 s, annealed at 50°C for 1 min, and extended at 72°C for 2 min, for a total of 30 cycles. The double-stranded amplification products formed were separated by gel electrophoresis on a 1% NuSieve agarose (FMC Bioproducts) gel in 1 × TAE (Tris-acetate/EDTA).

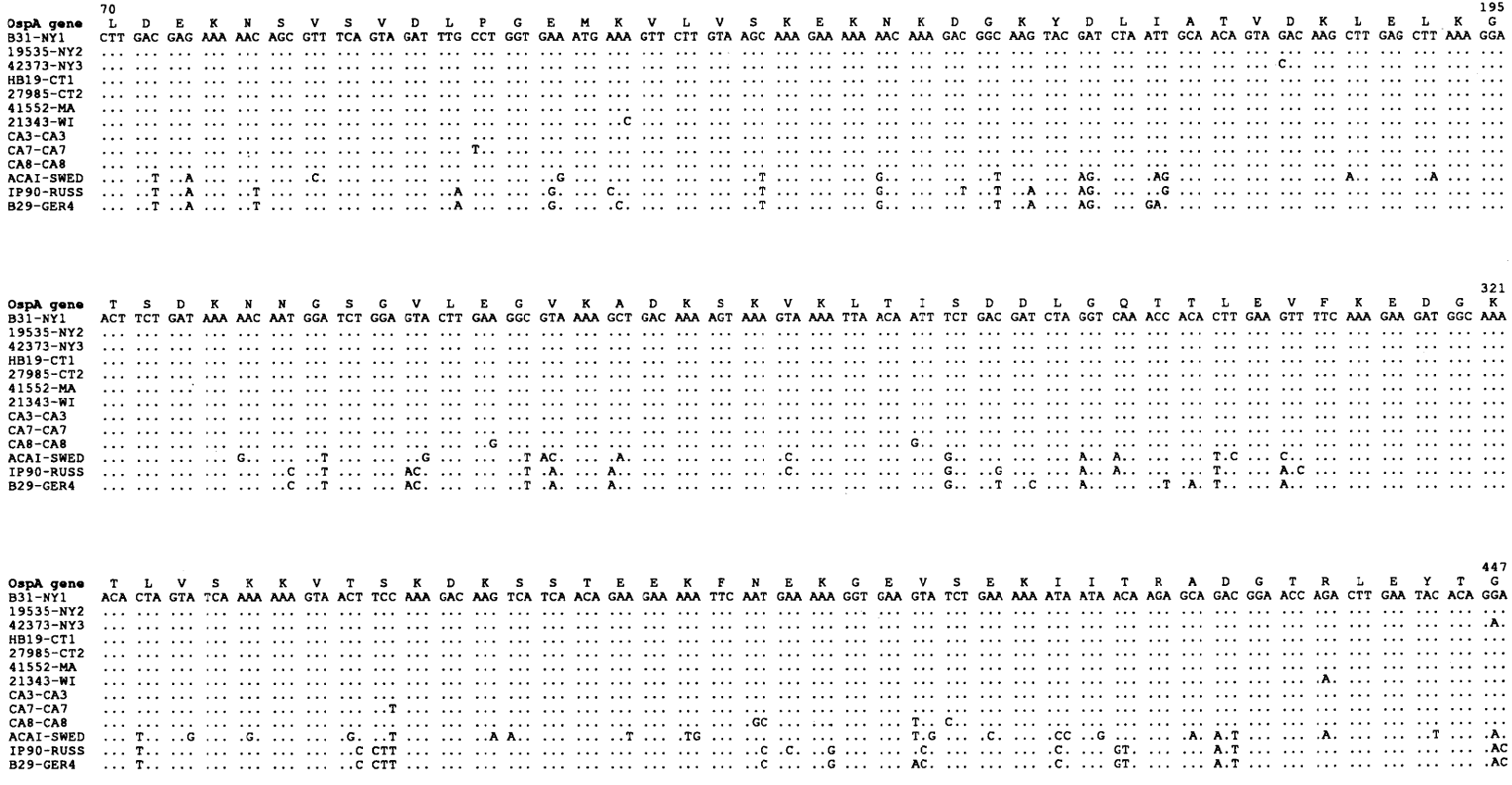
Purification of PCR Products

The sizes of the PCR products were compared with φX 174 RF DNA/*Hae*III fragments in the gel. Products of the expected size were cut from the gel and were prepared for nucleotide sequencing. A gel slice containing 600 ng of PCR product was diluted with 400 μl of TE and then was melted at 65°C for 5 min. NaCl was then added to a final concentration of 0.2 M. Agarose was removed by adding an equal volume of hot phenol to each sample, vortexing, heating at 65°C for 5 min, spinning at 13,000 *g* for 5 min, placing on ice for 5 min, and then spinning at 4°C for 5 more min. The DNA was further extracted by first using a 1:1 phenol:chloroform mixture and then using chloroform alone. $MgCl_2$ was added to a final concentration of 0.01 M, and the

Table 2
Twelve Primers for Amplification and Sequencing of OspA and OspB Genes

Primer[a]	Location	Sequence
OspA-1 ...	54–70	5'-GCAAAATGTTAGCAGCC-3'
OspA-5 ...	196–215	5'-ACTTCTGATAAAAACAATGG-3'
OspA-4 ...	643–662	5'-GCTTGGAATTCAGGCACTTC-3'
OspA-3 ...	743–724	5'-TATTGTTGTACTGTAATTGT-3'
OspA-2 ...	793–770	5'-GTTTTGTAATTTCAACTGCTGACC-3'
Osp-2'	770–793	5'-GGGTCAGCAGTTGAAATTACAAAAC-3'
OspB-1 ...	889–909	5'-GGTGCTGAGTCAATTGGTTCT-3'
OspB-4 ...	1201–1221	5'-TTAGAAGCATTTGATGCCAGC-3'
OspB-5 ...	1420–1445	5'-GTAGTCGGAAAAACAACAGTGGAAAT-3'
OspB-3 ...	1431–1411	5'-TTTTCCGACTACAAGACTTCC-3'
OspB-2 ...	1668–1648	5'-TTCTAGGCTGGTTCCAGCTGT-3'
OspB-6 ...	1760–1741	5'-TACACTAGCTCATGCCTTGT-3'

[a] All primers except for OspA-1 were developed for this study; primer OspA-1 was designed by Calvin Vary, MCRI, South Portland, Maine.


70
195
OspA gene L D E K N S V S V D L P G E M K V L V S K E K N K D G K Y D L I A T V D K L E L K G
B31-NY1 CTT GAC GAG AAA AAC AGC GTT TCA GTA GAT TTG CCT GGT GAA ATG AAA GTT CTT GTA AGC AAA GAA AAA AAC AAA GAC GGC AAG TAC GAT CTA ATT GCA ACA GTA GAC AAG CTT GAG CTT AAA GGA
19535-NY2
42373-NY3
HB19-CT1
27985-CT2
1a 41552-MA
21343-WI
CA3-CA3
CA7-CA7
CA8-CA8
ACAI-SWED
IP90-RUSS
B29-GER4
321
OspA gene T S D K N N G S G V L E G V K A D K S K V K L T I S D D L G Q T T L E V F K E D G K
B31-NY1 ACT TCT GAT AAA AAC AAT GGA TCT GGA GTA CTT GAA GGC GTA AAA GCT GAC AAA AGT AAA GTA AAA TTA ACA ATT TCT GAC GAT CTA GGT CAA ACC ACA CTT GAA GTT TTC AAA GAA GAT GGC AAA
19535-NY2
42373-NY3
HB19-CT1
27985-CT2
2a 41552-MA
21343-WI
CA3-CA3
CA7-CA7
CA8-CA8
ACAI-SWED
IP90-RUSS
B29-GER4
447
OspA gene T L V S K K V T S K D K S S T E E K F N E K G E V S E K I I T R A D G T R L E Y T G
B31-NY1 ACA CTA GTA TCA AAA AAA GTA ACT TCC AAA GAC AAG TCA TCA ACA GAA GAA AAA TTC AAT GAA AAA GGT GAA GTA TCT GAA AAA ATA ATA ACA AGA GCA GAC GGA ACC AGA CTT GAA TAC ACA GGA
19535-NY2
42373-NY3
HB19-CT1
27985-CT2
3a 41552-MA
21343-WI
CA3-CA3
CA7-CA7
CA8-CA8
ACAI-SWED
IP90-RUSS
B29-GER4

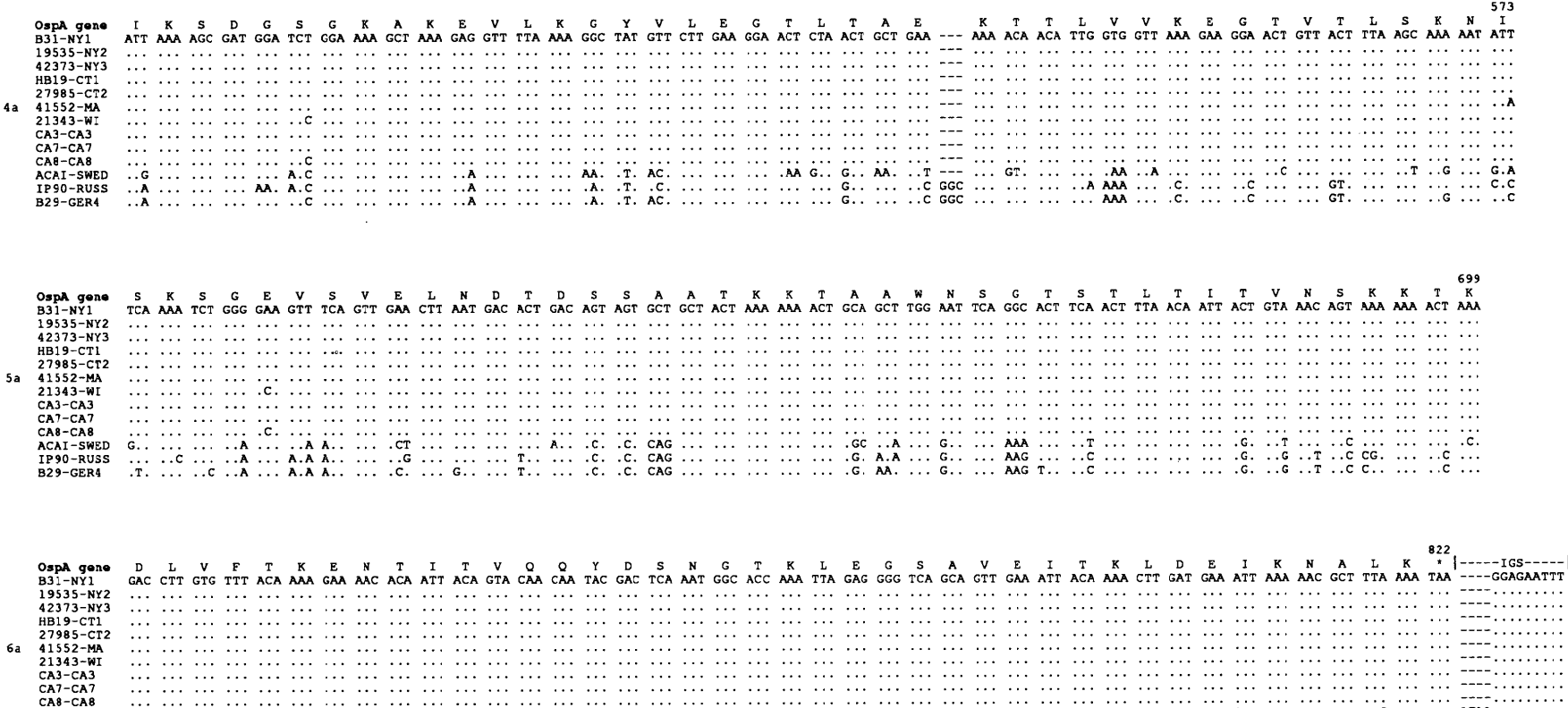
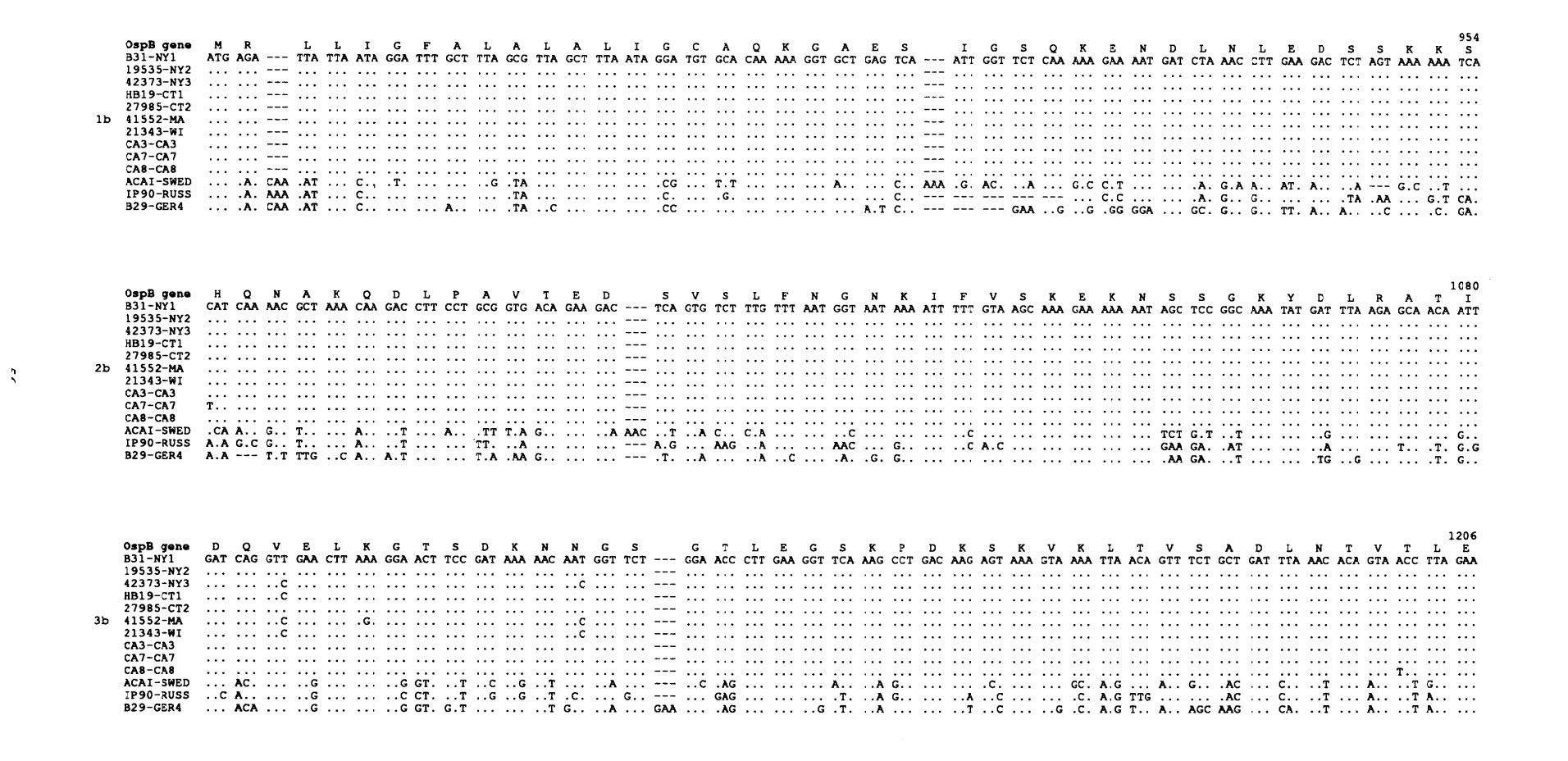


FIG. 2 (*Continued*)


OspB gene
B31-NY1
19535-NY2
42373-NY3
HB19-CT1
27985-CT2
41552-MA
21343-WI
CA3-CA3
CA7-CA7
CA8-CA8
ACAI-SWED
IP90-RUSS
B29-GER4
1b
2b
3b
954
1080
1206

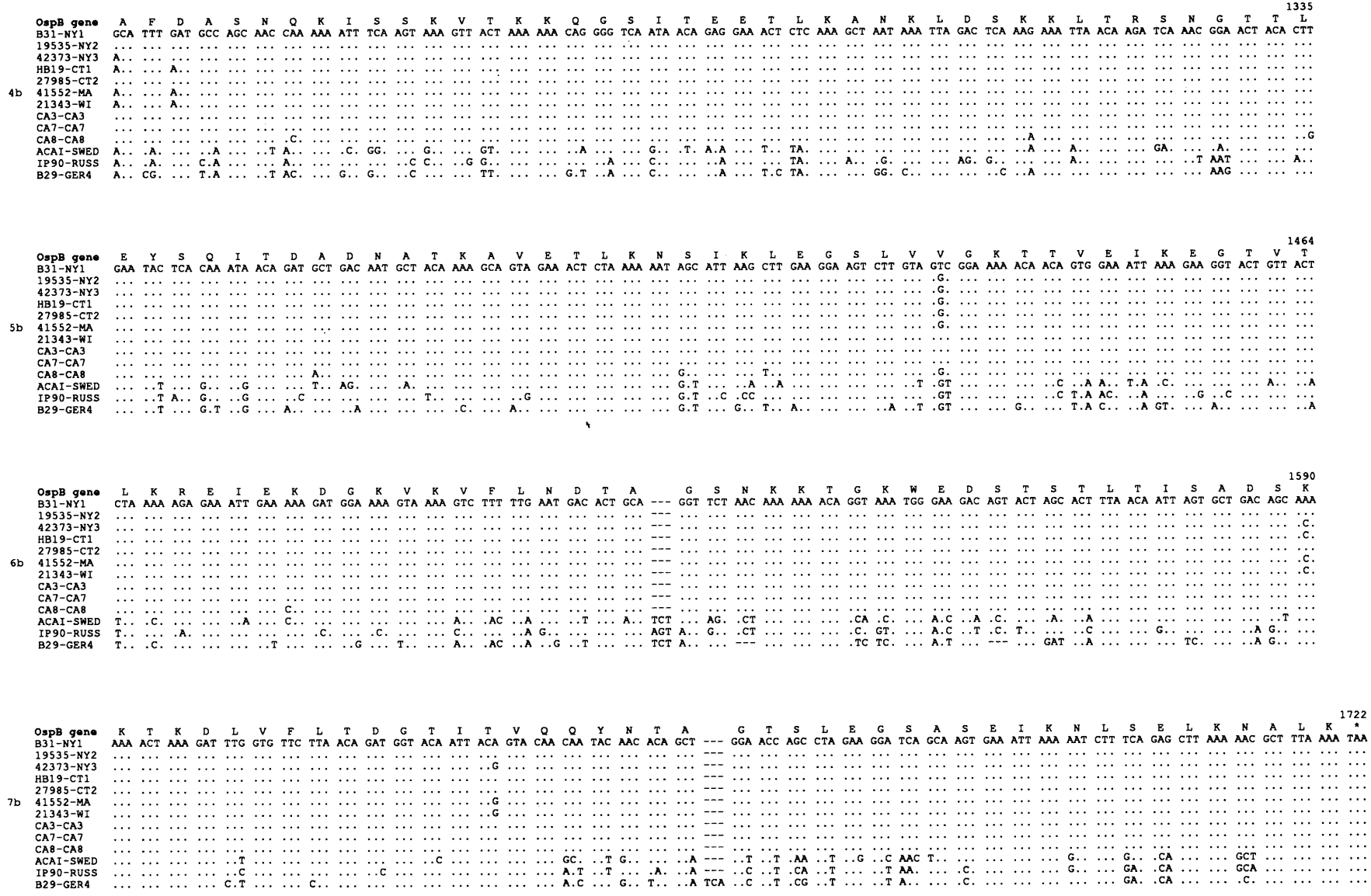


FIG. 2.—Nucleotide sequence of the OspA and OspB gene. The sequence is reported for 1,653–1,672 bases from 13 strains of *Borrelia burgdorferi*. The sequence for B31 strain, beginning at site 70 and ending at site 1722 on the first line, is given as the reference. The dots indicate bases that are the same as those in the B31 sequence. The dashes indicate deletions within the sequence. Regions of the OspA gene are labeled 1a–6a, and regions of the OspB gene are labeled 1b–7b. GenBank accession numbers are as follows: 19535-NY2, L23138; 42373-NY3, L23139; HB19-CT1, L23136; 27985-CT2, L23137; 41552-MA, L23140; 21343-WI, L23141; CA3, L23142; CA7, L23143; and CA8, L23144.

Table 3
Mean d_S and Mean d_N, between OspA and OspB Sequences of Several *Borrelia burgdorferi* Strains

Sequence Region[a]	G+C[b] (%)	Mean d_S ± Standard Error (%)	Mean d_N ± Standard Error (%)	d_S/d_N[c]
OspA:				
1 (70–195)	19.0	16.8 ± 5.8	4.4 ± 1.3	3.82[d]
2 (196–321)	9.5	13.6 ± 4.1	6.3 ± 1.5	2.16
3 (322–447)	9.1	18.5 ± 5.2	8.2 ± 1.6	2.26*
4 (448–573)	20.0	17.2 ± 5.2	9.5 ± 1.9	1.81
5 (574–699)	7.4	21.4 ± 6.4	12.1 ± 2.4	1.77
6 (700–819)	22.2	22.9 ± 6.6	8.6 ± 2.1	2.66[d]
Mean	14.5	18.1 ± 2.2	8.1 ± 0.7	2.23
OspB:				
1 (832–954)	5.0	9.4 ± 3.4	14.6 ± 2.4	0.64
2 (955–1080)	40.0	25.0 ± 6.5	11.4 ± 1.9	2.19[d]
3 (1081–1206) ...	12.0	41.6 ± 11.1	9.7 ± 1.8	4.29[d]
4 (1207–1335) ...	17.4	13.0 ± 4.0	10.3 ± 1.8	1.26
5 (1336–1464) ...	12.0	16.6 ± 4.8	7.1 ± 1.4	2.34*
6 (1465–1590) ...	20.0	15.1 ± 4.8	9.2 ± 1.6	1.64
7 (1591–1719) ...	16.7	18.4 ± 5.0	6.4 ± 1.5	2.88[d]
Mean	17.6	18.5 ± 2.0	9.6 ± 0.7	1.93

[a] Nos. refer to position in the nucleotide sequence shown in fig. 2.
[b] Includes fourfold-degenerate sites only.
[c] A d_S/d_N ratio of 1 indicates a molecule undergoing neutral evolution. Eighteen isolates were included in the OspA analysis, and 13 were included in the OspB analysis.
[d] Significant ($P < 0.05$) difference between mean d_S and mean d_N.
* $P < 0.06$.

DNA was precipitated with 2 volumes of 100% isopropanol. The DNA pellet was dried in a Speed-Vac (Savant) and was suspended in 19 µl of sterile distilled water.

Cycle Sequencing

A cycle sequencing kit (Applied Biosystems) was used for the *Taq* DNA polymerase–mediated incorporation of dye-labeled dideoxy terminators. Reactions were performed in a Perkin-Elmer thermal cycler. Components were denatured at 96°C for 30 s, annealed at 50°C for 15 s, and extended at 60°C for 4 min, for a total of 25 cycles. After thermal cycling, samples were run through Centri-Sep columns (Princeton Separations) to remove unincorporated nucleotides and then were dried in a Speed-Vac for 1 h. To each tube, 4 µl of 5:1 formamide:50 mM EDTA mixture was added. The whole mixture was heated at 90°C for 2 min, placed on ice for 5 min, and then loaded on a 6% polyacrylamide gel in an ABI 373A DNA sequencer (Applied Biosystems). Sequences were analyzed on a MacIntosh computer using the Seq Ed™ 675 DNA sequence editor program (Applied Biosystems). Primer sets used produced PCR products which overlapped 100–300 bases to ensure proper sequence alignment. Sequence analyses were performed on both strands of each gene segment for verification.

Calculation of d_S and d_N

To examine the hypothesis of positive selection for variant outer-surface proteins, d_S (i.e., no. of synonymous substitutions per synonymous site) and d_N (i.e., no. of nonsynonymous substitutions per nonsynonymous site) from the aligned nucleotide sequences were estimated using the method of Nei and Gojobori (1986). All strains listed in table 1 were included in the OspA gene analysis. The OspB nucleotide sequence of strains 25015, ZS7, ZQ1, and PKo have not been reported. Strain GO2 does not express the OspB protein, because of a 79-bp deletion at the 5′ end (Eiffert et al. 1992). Therefore, all except these five strains were included in the OspB gene analysis. The FORTRAN program NAG, version 2.0 (Nei and Jin 1989), was used to calculate the average d_S and d_N among all comparisons. A d_S/d_N ratio close to 1 would indicate that the molecule is undergoing neutral evolution. A d_S value higher than its corresponding d_N value ($d_S/d_N > 1$) would suggest that purifying selection is occurring. A protein experiencing positive selection for amino acid sequence variation would have a d_S value smaller than its d_N value ($d_S/d_N < 1$).

Phylogenetic Tree

Trees describing the phylogenetic history of the variants were constructed using parsimony. The branch-

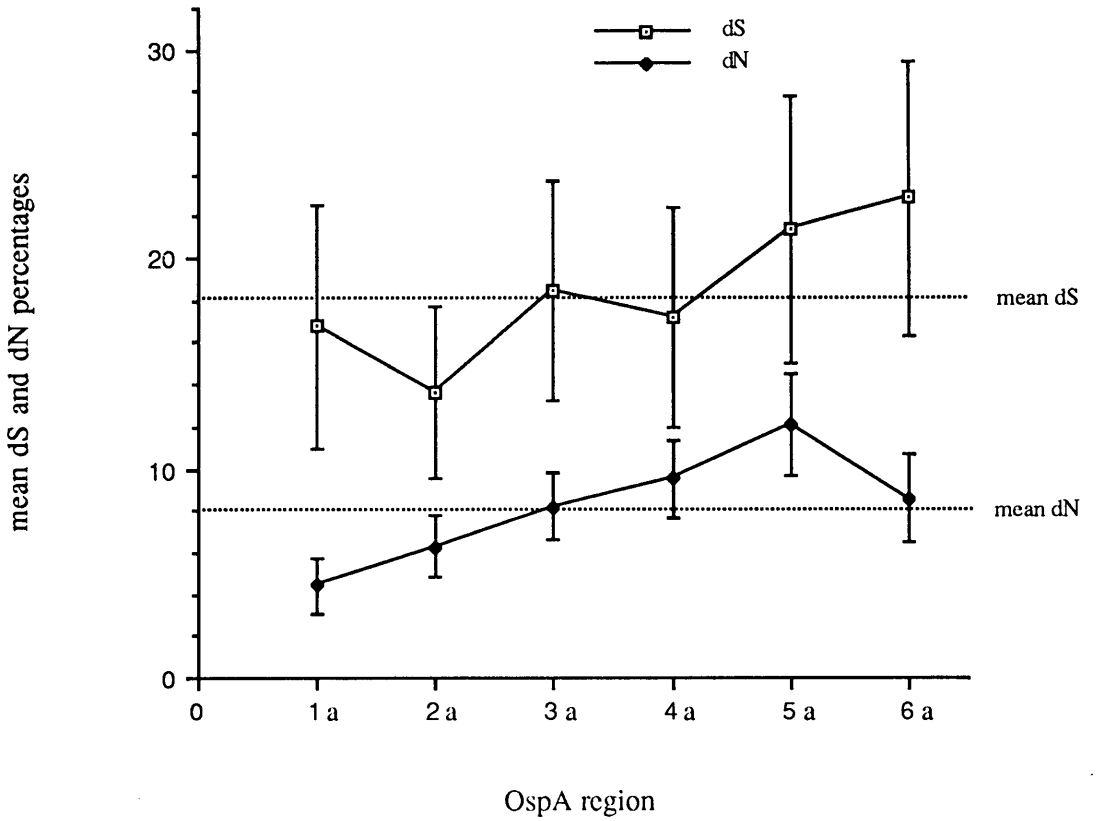


FIG. 3.—Comparisons of d_S and d_N percentages within six regions of the OspA gene

and-bound algorithm of PAUP (Swofford 1990) was used to find the most parsimonious tree. Majority-rule consensus trees were used when multiple equally parsimonious reconstructions were found. One thousand bootstrap replicates were performed to evaluate the reliability of some trees. Distance trees were constructed from the d_S and d_N values using the neighbor-joining algorithm (Saitou and Nei 1987). Parsimony and distance-matrix trees were obtained for the OspA gene, the OspB gene, and both combined.

Results

OspA and OspB Sequences

The sequence of 753 of 822 nt in the OspA gene, the intragenic sequence of 9 nt, and the entire nucleotide sequence of the OspB gene were determined for North American isolates 19535, 42373, HB19, 27985, 41552, 21343, CA3, CA7, and CA8. In figure 2 the sequences are aligned with the sequence of strain B31 determined by Bergstrom et al. (1989). Only one sequence was observed for each isolate, and there were no insertions or deletions among these sequences. Base substitutions were scattered across more than half of the OspA gene, but none were observed within the last quarter. Most of the point mutations observed in the OspB gene were scattered through the second half of the gene, with only one seen in the first quarter of the gene.

Strains ACAI, Ip90, and B29 were also included in the nucleotide sequence alignment (fig. 2), for further comparisons. Of the 187 sites which have mutations within the OspA gene, 56 substitutions were seen in the first position of a codon, 49 in the second position, and 82 in the third position. The OspB gene contained a total of 330 point mutations, 123 of which were in the first position of a codon, 80 in the second, and 127 in the third. Strains IP90 and B29 show a codon insertion relative to the other strains directly after site 522 of the OspA gene, as shown in figure 2 and described by Jonsson et al. (1992). Insertion/deletion of codons has occurred in OspB directly after sites 837, 900, 942, 957, 996, 1125, 1521, 1527, 1554, and 1653. Strains CT1, CT2, NY1, NY2, and CA3 contain the same OspA sequence but vary in their OspB sequences (with the exception of NY2 and CT2 being identical). Within the OspA gene, the pairwise number of nucleotide differences among all isolates was 0–121, and within the OspB gene it was 0–235. The average pairwise difference in the OspA gene was 4.1%, and for the OspB gene it was 6.2%.

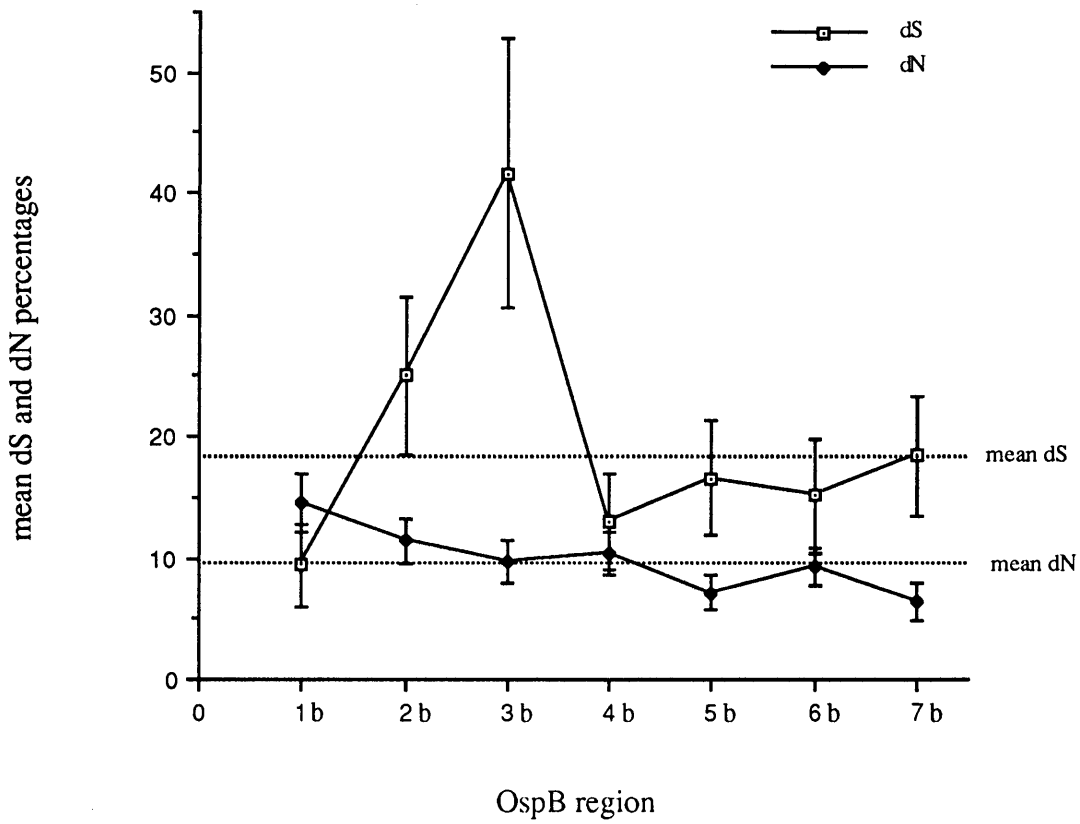


FIG. 4.—Comparisons of d_S and d_N percentages within seven regions of the OspB gene

d_S and d_N Values

The mean percentage of synonymous substitutions per synonymous site within the OspA gene overall was 18.1% ± 2.2%, and that for nonsynonymous sites was 8.1% ± 0.7%. These values indicate a higher rate of silent substitution. In order to detect different patterns of selection from region to region, mean d_S and d_N values were determined from six equally sized regions of the OspA gene (table 3). d_S values appear to increase from region 2a to region 7a, and from region 1a to region 5a the d_N values increase, as seen in figure 3. The d_S and d_N values are significantly different ($P < 0.5$) only in regions 1a and 6a.

Although the mean d_S and d_N values of the OspB gene overall were found to be 18.5% ± 2.0% and 9.6% ± 0.7%, respectively, which suggests that the gene is under weak purifying selection, d_S and d_N values taken from separate regions of the OspB gene suggest a more complex picture. According to table 3, the first OspB region is experiencing a slight bias toward nonsynonymous substitutions or amino acid changes, while OspB regions 2b and 3b are experiencing a large bias toward silent substitutions. Regions 4b–7b contain comparatively lower d_N values. This is further seen in figure 4, as areas of overlap between mean d_S and mean d_N values. Contrary to d_S values, d_N values remain fairly constant throughout the OspB gene.

Phylogenetic Analysis

Trees describing the phylogenetic history of the borrelial variants were constructed by parsimony. ACA1, Ip90, and B29 were found to be the most distantly related from the other isolates and were therefore used to root the network. Distance trees were also determined from the isolates by using the neighbor-joining algorithm. Relationships among the North American strains generally were not resolved in these trees, because of the small number of character differences.

Comparisons between parsimony trees of the OspA gene, the OspB gene, and both genes combined showed inconsistencies in the placement of strains WI and CA8. There appear to be sites of homoplasy between the California strains and the Wisconsin strain. Contiguous groupings of informative sites (table 4) suggest that there may be recombinational events occurring within the OspA/B operon. It is possible that a single crossover event has occurred between strains WI and CA8, within the region 587–1089. Two recombinational events may have occurred between strains WI and CA8, within regions 1207–1424 and 1424–1589. An alternative expla-

nation for the existence of homoplasy may be that these informative sites are hot spots for rapid point-mutational events. Since the possibility of recombination exists, strains WI and CA8 were excluded from the phylogenetic analysis.

The phylogenetic tree, seen in figure 5, shows the CT1 isolate being distantly related to CT2, both of which had come from Connecticut. New York strains 1–3 were also found to be distantly related to one another. In comparison, New York strain 1 and California strain 3 are closely related. There appears to be no correlation between the amount of OspA and OspB sequence similarity and geographic distribution of the North American isolates. It is also interesting to note that OspB sequence variation is present locally, particularly in New York, Connecticut, and California.

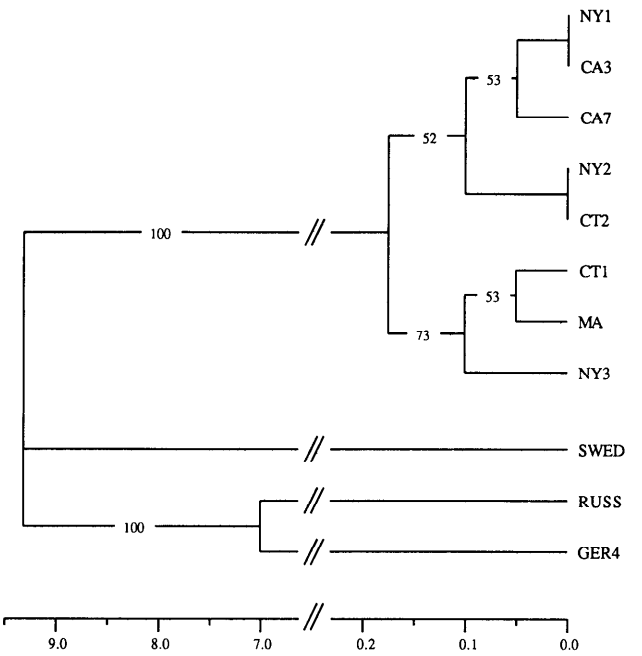


FIG. 5.—Phylogenetic relationships of 11 Osp genes (OspA and OspB combined) of *Borrelia burgdorferi*. The shortest tree was found by using the branch-and-bound algorithm of PAUP, version 3.0. One thousand bootstrap replications were performed, and the 50%-majority-rule consensus tree was constructed. Bootstrap values >50 are indicated on the tree. Branch lengths were drawn proportional to the mean percentages of d_N values of all pairwise comparisons.

Discussion

The OspA and OspB sequences have allowed us to examine the possibility that positive selection acts to accelerate the evolution of the outer-surface-protein genes. Most protein-coding genes exhibit high d_S/d_N ratios, suggestive of purifying selection acting to conserve the structure of the protein. Examples of mammalian genes include histones, actins, globins, and many hormone genes, with their d_S/d_N ratios together averaging ~5 (Li and Graur 1991). Influenza virus genes (e.g., hemagglutinin and neuraminidase) also experience high d_S/d_N ratios (Nei 1987). On the other hand, a low d_S/d_N would suggest positive selection favoring diversity at the amino acid level. Such diversity might be an important component of the immune-avoidance mechanism of the spirochete. An example of positive selection has been observed at major-histocompatibility-complex class II loci (Hughes and Nei 1988, 1989). Evidence of positive selection acting on surface proteins has been seen in malaria parasites (Hughes 1991, 1992). Although in *Borrelia burgdorferi* the observed d_S/d_N ratios for both OspA and B genes were slightly high overall, their ratios are much lower than those in mammalian protein-coding genes. The corresponding d_S and d_N values from most areas of each Osp gene were not statistically different, suggesting that the Osp proteins are under little, if any, purifying selection for maintenance of primary structure.

Although the OspA gene appears to be undergoing little purifying selection, the number of nucleotide substitutions and d_S and d_N values were found to be greater

Table 4
Comparison of North American Isolates at Informative Sites within the OspA/B Operon

Isolate	465	587	1089	1119	1207	1424	1589
CT2	T	A	T	T	G	G	A
NY2	T	A	T	T	G	G	A
NY1	T	A	T	T	G	T	A
CA3	T	A	T	T	G	T	A
CA7	T	A	T	T	G	T	A
CA8	C	C	T	T	G	T	A
WI	C	C	C	C	A	T	C
MA	T	A	C	C	A	G	C
NY3	T	A	C	C	A	G	C
CT1	T	A	C	T	A	G	C

at the carboxyl end than at the amino end of the protein. A progressive increase, in both values, from region 1a to region 6a suggests a general increase in mutation rate along the OspA protein. Human and mouse antibodies recognize epitopes located within the last 60 amino acids of the OspA protein (Sears et al. 1991; Shanafelt et al. 1992). This portion of the protein may have a lower functional constraint, thus accounting for more variation. The d_N values were found to be significantly lower in regions 1a and 2a. This area of the protein is under more functional constraint, since it is the portion which is membrane embedded.

Within the OspB gene, sequence regions 1b and 2b contained virtually no variation among the 10 North American strains reported (with the exception of one nucleotide substitution in the CA7 strain). Similar to the OspA protein, this hydrophobic region of the OspB protein membrane is probably embedded and under functional constraint.

The first sequence block, 126 bases long, contained a significantly low d_S/d_N ratio. Wolfe et al. (1989) have found that a G+C-content bias at the third codon position can reduce d_S values. Since region 1b is very A+T rich at the third position, this may be the cause of the depressed d_S/d_N value. Also, there appears to be a Shine-Delgarno sequence (TAAGGAGA) starting at site 820 and ending 5 nt upstream of the OspB start codon. Since the ribosome spans ~35 nt of mRNA (Lewin 1990), this might introduce a small amount of constraint on region 1b, thus lowering its d_S value.

Sequence regions 4b–7b, located in the second half of the OspB gene, have a low d_S/d_N ratio, compared with regions 2b and 3b, suggesting a higher functional constraint on this portion of the molecule as well. There appears to be an open reading frame (ORF) on the antisense strand ending at site 1083 and continuing through the 3′ end of the OspB gene within the North American strains. If this ORF codes for another protein, then this could cause the latter half of the OspB gene to be under additional functional constraint.

Although there appears to be no overall positive selection, there may be selection favoring an increase in variation in particular parts of the molecule. In circumsporozoite protein gene regions encoding T-cell epitopes of *Plasmodium falciparum,* the rate of nonsynonymous nucleotide substitution was found to be significantly higher than that of synonymous substitution (Hughes 1991). Fikrig et al. (1993) have found a borrelial strain containing a truncated form of OspB (containing a stop codon within region 5b) to evade vaccination immunity. They have also found that epitopes in the C-terminus of the OspB are exposed at the surface and can bind protective antibodies. Regions 4b–6b of the OspB protein may indeed be immunogenic. We did not find positive selection in these regions. However, our discovery of possible recombination within these regions suggests caution for programs to develop vaccines. The amount of OspB variation may be increasing if recombinational events are occurring within these regions. Programs to develop vaccines should include strains containing variation in these regions of the OspB gene, to ensure effectiveness of the vaccine on the range of variation present in natural populations.

Dispersal of spirochetes is expected to be low, since ticks and their major hosts do not move very far. Therefore, some genetic structuring of borrelial populations would be expected. Although the three Eurasian strains differed considerably from the North American strains, phylogenetic analyses of the 10 North American isolates tested showed little geographic structure. What might cause so little variation among isolates hundreds of miles apart? Perhaps the lack of variation between strains taken from Wisconsin and from New York is because birds can carry and disperse infected ticks from state to state (Anderson 1988). Domesticated animals might also be carriers of infected ticks (Bosler et al. 1988). Dispersal by these kinds of hosts would allow recombinational events to take place between isolates from different states.

The average heterozygosity at the nucleotide level (π; Nei 1987) among North American strains was found to be 0.003. Relatively low, this value is comparable to that of the β-globin, growth-hormone, and insulin genes in humans. The average pairwise synonymous divergence among North American strains was found to be 1.0%. The absolute rate of evolution at synonymous sites in bacteria is ~0.7%/Myr (Wilson et al. 1987). This suggests that these North American *B. burgdorferi* strains could have been diverging for ≤1 Myr. In addition, isolates taken from humans (HB19, ACA1, and PKo) appear to be highly divergent, with an average pairwise synonymous divergence of 21.9%. It is not likely that the recent spread of Lyme disease is due to one particular strain.

It is interesting that several isolates taken from the same local region were not identical in sequence. Selection could be favoring the maintenance of several Osp variants within a single population. Variants might be favored during infection of a previously exposed mouse, as well as within a population which had adapted to a particular borrelial strain. Multiple variants might be maintained by a form of frequency-dependent selection. We are exploring these possibilities by sequencing a number of isolates from a single location.

Acknowledgments

R. Zsigray and A. Klein and two anonymous reviewers provided helpful comments on the manuscript.

This work was supported by Hatch project 348. This is New Hampshire Agricultural Experiment Station publication 1832.

LITERATURE CITED

ANDERSON, J. F. 1988. Mammalian and avian reservoirs for *Borrelia burgdorferi.* Ann. N.Y. Acad. Sci. **539**:180–191.

BARBOUR, A. G. 1984. Isolation and cultivation of Lyme disease spirochetes. Yale J. Biol. Med. **57**:521–525.

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. N.Y. Acad. Sci. **539**:144–153.

BARBOUR, A. G., R. A. HEILAND, and T. R. HOWE. 1985. Heterogeneity of major proteins in Lyme disease borreliae: a molecular analysis of North American and European isolates. J. Infect. Dis. **152**:478–484.

BARBOUR, A. G., S. L. TESSIER, and H. G. STOENNER. 1982. Variable major proteins of *Borrelia hermsii.* J. Exp. Med. **156**:1312–1324.

BERGSTROM, S., V. G. BUNDOC, and A. G. BARBOUR. 1989. Molecular analysis of linear plasmid-encoded major surface proteins, OspA and OspB, of the Lyme disease spirochaete *Borrelia burgdorferi.* Mol. Microbiol. **3**:479–486.

BORST, P. 1983. Mobile genetic elements. Academic Press, New York.

BOSLER, E. M., D. P. COHEN, T. L. SCHULTZE, C. OLSEN, W. BERNARD, and B. LISSMAN. 1988. Host responses to *Borrelia burgdorferi* in dogs and horses. Ann. N. Y. Acad. Sci. **539**:221–234.

BURGDORFER, W., A. G. BARBOUR, S. G. HAYES, J. L. BENACH, E. GRUNWALDT, and J. P. DAVIS. 1982. Lyme disease—a tick-borne spirochetosis? Science **216**:1317–1319.

CIESIELSKI, C. A., L. E. MARKOWITZ, R. HORSLEY, A. B. HIGHTOWER, H. RUSSELL, and C. V. BROOME. 1988. The geographic distribution of Lyme disease in the United States. Ann. N. Y. Acad. Sci. **539**:283–288.

EIFFERT, H., A. OHLENBUSCH, W. FEHLING, H. LOTTER, and R. THOMSSEN. 1992. Nucleotide sequence of the OspAB operon of a *Borrelia burgdorferi* strain expressing OspA but not OspB. Infect. Immun. **60**:1864–1868.

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**: 553–555.

FIKRIG, E., S. W. BARTHOLD, D. H. PERSING, X. SUN, F. S. KANTOR, and R. A. FLAVELL. 1992. *Borrelia burgdorferi* strain 25015: characterization of outer surface protein A and vaccination against infection. J. Immun. **148**:2256–2260.

FIKRIG, E., H. TAO, F. S. KANTOR, S. W. BARTHOLD, and R. A. FLAVELL. 1993. Evasion of protective immunity by *Borrelia burgdorferi* by truncation of outer surface protein B. Proc. Natl. Acad. Sci. USA **90**:4092–4096.

HAGBLOM, P., E. SEGEL, E. BILLGARD, and M. SO. 1985. Intragenic recombination leads to pilus antigenic variation in *Neisseria gonorrhoeae.* Nature **315**:156–158.

HOWE, T. R., F. W. LAQUIER, and A. G. BARBOUR. 1986. Organization of genes encoding two outer membrane proteins of the Lyme disease agent *Borrelia burgdorferi* within a single transcriptional unit. Infect. Immun. **54**:207–212.

HUGHES, A. L. 1991. Circumsporozoite protein genes of malaria parasites (*Plasmodium* spp.): evidence for positive selection on immunogenic regions. Genetics **127**:345–363.

———. 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.

———. 1989. Nucleotide substitution at major histocompatibility complex class II loci: evidence for overdominant selection. Proc. Natl. Acad. Sci. USA **86**:958–962.

JONSSON, M., L. NOPPA, A. G. BARBOUR, and S. BERGSTROM. 1992. Heterogeneity of outer membrane proteins in *Borrelia burgdorferi:* comparison of Osp operons of three isolates of different geographic origins. Infect. Immun. **60**:1845–1853.

LEWIN, B. 1990. Genes IV. Oxford University Press, New York.

LI, W. and D. GRAUR. 1991. Rates and patterns of nucleotide substitution. Pp. 67–98 in Fundamentals of molecular evolution. Sinauer, Sunderland, Mass.

NEI, M. 1987. Molecular evolutionary genetics. Columbia University Press, New York.

NEI, M., and T. GOJOBORI. 1986. Simple methods 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.

PERSING, D. H., S. R. TELFORD III, P. N. RYS, D. E. DODGE, T. J. WHITE, S. E. MALAWISTA, and A. SPIELMAN. 1990. Detection of *Borrelia burgdorferi* DNA in museum specimens of *Ixodes dammini* ticks. Science **249**:1420–1423.

PLASTERK, R. H. A., M. I. SIMON, and A. G. BARBOUR. 1985. Transposition of structural genes to an expression sequence on a linear plasmid causes antigenic variation in the bacterium *Borrelia hermsii.* Nature **318**:257–263.

PRASAD, V. 1991. Lyme disease from a physician's point of view. Int. J. Acarol. **17**:37–41.

ROSA, P. A., T. SCHWAN, and D. HOGAN. 1992. Recombination between genes encoding major outer surface proteins A and B of *Borrelia burgdorferi.* Mol. Microbiol. **6**:3031–3040.

SAITOU, N., and M. NEI. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4**:406–425.

SAMBROOK, J., E. F. FRITSCH, and T. MANIATIS. 1989. Molecular cloning: a laboratory manual, 2d ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

SCHWAN, T. G., W. BURGDORFER, and C. F. GARON. 1988. Changes in infectivity and plasmid profile of the Lyme disease spirochete, *Borrelia burgdorferi,* as a result of in vitro cultivation. Infect. Immun. **56**:1831–1836.

SEARS, J. E., E. FIKRIG, T. Y. NAKAGAWA, K. DEPONTE, N. MARCANTONIO, F. S. KANTOR, and R. A. FLAVELL. 1991.

Molecular mapping of OspA mediated immunity against *Borrelia burgdorferi,* the agent of Lyme disease. J. Immunol. **147**:1995–2000.

SHANAFELT, M. C., J. ANZOLA, C. SODERBERG, H. YSSEL, C. W. TURCH, and G. PELTZ. 1992. Epitopes on the outer surface protein A of *Borrelia burgdorferi* recognized by antibodies and T cells of patients with Lyme disease. J. Immunol. **148**:218–224.

SPIELMAN, A., M. L. WILSON, J. F. LEVINE, and J. PIESMAN. 1985. Ecology of *Ixodes dammini*–borne human babesiosis and Lyme disease. Annu. Rev. Entomol. **30**:439–460.

STEERE, A. C., S. E. MALAWISTA, J. A. HARDIN, S. RUSSY, P. W. ACKENASE, and W. A. ANDIMAN. 1977. Erythema chronicum migrans and lyme arthritis: the enlarging clinical spectrum. Ann. Intern. Med. **86**:685–698.

SWOFFORD, D. L. 1990. PAUP: phylogenetic analysis using parsimony, version 3.0, program manual. Illinois Natural History Survey, Champaign.

TELFORD, S. R. III, T. M. MATHER, S. I. MOORE, M. L. WILSON, and A. SPIELMAN. 1988. Incompetence of deer as reservoirs of the Lyme disease spirochete. Am. J. Trop. Med. Hyg. **39**:105–109.

WILSON, A. C., H. OCHMAN, and E. M. PRAGER. 1987. Molecular time scale for evolution. Trends Genet. **3**:241–247.

WOLFE, K. H., P. M. SHARP, and W. H. LI. 1989. Mutation rates differ among regions of the mammalian genome. Nature **337**:283–285.

MASATOSHI NEI, reviewing editor

Received October 5, 1992

Accepted August 30, 1993