

Molecular and Evolutionary Analysis of *Borrelia burgdorferi* 297 Circular Plasmid-Encoded Lipoproteins with OspE- and OspF-Like Leader Peptides

DARRIN R. AKINS,¹† MELISSA J. CAIMANO,¹ XIAOFENG YANG,² FELIX CERNA,¹‡
MICHAEL V. NORGARD,² AND JUSTIN D. RADOLF^{1,2*}

*Departments of Internal Medicine¹ and Microbiology,² University of Texas
Southwestern Medical Center, Dallas, Texas 75235*

Received 6 November 1998/Returned for modification 8 December 1998/Accepted 16 December 1998

We previously described two OspE and three OspF homologs in *Borrelia burgdorferi* 297 (D. R. Akins, S. F. Porcella, T. G. Popova, D. Shevchenko, S. I. Baker, M. Li, M. V. Norgard, and J. D. Radolf, *Mol. Microbiol.* 18:507–520, 1995; D. R. Akins, K. W. Bourell, M. J. Caimano, M. V. Norgard, and J. D. Radolf, *J. Clin. Investig.* 101:2240–2250, 1998). In this study, we characterized four additional lipoproteins with OspE/F-like leader peptides (Elps) and demonstrated that all are encoded on plasmids homologous to cp32 and cp18 from the B31 and N40 strains, respectively. Statistical analysis of sequence similarities using the binary comparison algorithm revealed that the nine lipoproteins from strain 297, as well as the OspE, OspF, and Erp proteins from the N40 and B31 strains, fall into three distinct families. Based upon the observation that these lipoproteins all contain highly conserved leader peptides, we now propose that the ancestors of each of the three families arose from gene fusion events which joined a common N terminus to unrelated proteins. Additionally, further sequence analysis of the strain 297 circular plasmids revealed that rearrangements appear to have played an important role in generating sequence diversity among the members of these three families and that recombinational events in the downstream flanking regions appear to have occurred independently of those within the lipoprotein-encoding genes. The association of hypervariable regions with genes which are differentially expressed and/or subject to immunological pressures suggests that the Lyme disease spirochete has exploited recombinatorial processes to foster its parasitic strategy and enhance its immunoevasiveness.

Lyme disease, the most common tickborne infection in the United States, is a chronic, multisystem disorder caused by spirochetes of the *Borrelia burgdorferi* sensu lato complex (4, 27). Lyme disease spirochetes are maintained in nature via an enzootic cycle which typically involves wild rodents and *Ixodes* ticks (22, 34). To sustain this cycle, *B. burgdorferi* must adapt to two markedly different host environments. Furthermore, during the mammalian phase of infection, the bacterium presumably expresses virulence determinants which facilitate its ability to cause disease and establish persistent infection. There is now a substantial body of evidence that *B. burgdorferi* meets these environmental challenges by altering its antigenic composition (12). Our present understanding of differential gene expression by *B. burgdorferi* was initiated by the finding that outer surface protein A (OspA) and OspC undergo reciprocal changes in expression during tick feeding (10, 13, 26, 32). It subsequently was extended by the discovery of numerous differentially expressed borrelial proteins, many of which are homologous to the circular-plasmid-encoded OspE and OspF lipoproteins (1, 2, 9, 11, 17, 20, 38, 40, 44). The presence of multiple, differentially expressed OspE and OspF homologs in various *B. burgdorferi* isolates raises a number of intriguing questions about their molecular evolution, their physiological function(s) during the spirochete's enzootic cycle, their poten-

tial involvement in immune evasion and protective immunity, and the genetic mechanisms and environmental signals which regulate their expression. To begin to address these questions, we have extended our prior work (1, 2) with *B. burgdorferi* 297, a human cerebrospinal fluid isolate (35), and characterized four novel lipoproteins with sequence relatedness to OspE and OspF. These studies have greatly extended our overall insight into the evolutionary relationships between these lipoproteins and, equally important, have underscored the importance of recombination as a mechanism for generating sequence diversity at these evolutionarily unstable loci.

Identification of novel *B. burgdorferi* 297 genes with ospE/F-like leader peptides. Employing a strategy developed earlier to identify the multiple *ospE*- and *ospF*-related loci in other borrelial isolates (8, 24, 39, 41), we generated an oligonucleotide (US-47, Table 1) specific for a highly conserved region upstream of the known *ospE* and *ospF* homologs in *B. burgdorferi* 297 (1, 2). Using this probe in Southern hybridization analyses, we identified two *Hind*III fragments (0.5 and 2.5 kb) which did not correlate with any of the previously identified strain 297 *ospE* or *ospF* homologs. Cloning and nucleotide sequence analysis revealed open reading frames (ORFs) encoding closely related 34.8- and 41.3-kDa polypeptides. These proteins were designated ElpA1 and ElpA2 to indicate that they contained OspE/F-like leader peptides but, as discussed below, were otherwise unrelated to OspE and OspF. *ospE* homologs in the N40 and B31 strains have been shown to be contiguous to other lipoprotein-encoding genes (8, 21, 39). Further sequencing of the regions downstream of the strain 297 *ospE* and *p21* genes revealed that after short (27 bp), identical noncoding regions, both genes are followed by ORFs coding for polypeptides of 43 and 46.5 kDa, respectively, which also were found

* Corresponding author. Mailing address: Division of Infectious Diseases, U.T. Southwestern Medical Center, Dallas, TX 75235-9113. Phone: (214) 648-6896. Fax: (214) 648-5476. E-mail: Jradolf@mednet.swmed.edu.

† Present address: The University of Oklahoma Health Sciences Center, Oklahoma City, Okla.

‡ Present address: Lovelace Health Systems, Albuquerque, N.Mex.

TABLE 1. Oligonucleotide primers and probes used in this study

Designation	Sequence	Purpose
US-47	5'-GAATGTAACAAAATTATATATTTAAATCTTTGAAAAAT-3'	Probe for <i>ospE/F</i> loci
<i>fla5'</i>	5'-GATTATATCAATCATAATACATCAGC-3'	Specific probe
<i>ospC</i>	5'-CTTATTAAGTTTTTTTTGGACTTTCTGC-3'	Specific probe
<i>ospF</i>	5'-TACTATAGAAAATAATGAAAAAG-3'	Specific probe
<i>bbk2.11</i>	5'-ATCTAAATTAGATGAAAAAG-3'	Specific probe; PCR
2.10-5'	5'-AAAAAGAAGAGTTGGTTGGTGGTTT-3'	<i>bbk2.10</i> -specific primer
2.10-3'	5'-AACACCTTGAGCCACTTGCTC-3'	<i>bbk2.10</i> -specific primer
<i>p21</i>	5'-GCAAAGTAATAATGAGTTAAAAGTTAAGCAAAG-3'	Specific probe; PCR
<i>ospE</i>	5'-GGAGACTCTGATACATTTCGAG-3'	Specific probe; PCR
<i>elpA1</i>	5'-CAATAGTAGGCGAAATAGCAAAG-3'	Specific probe; PCR
<i>elpA2</i>	5'-GAAAAGAAATAAAAATGTTGTAGGG-3'	Specific probe; PCR
<i>ospA</i>	5'-TATGTTCTTGAAGGAACCTAACTGCTGAAAAACAACATT-3'	Specific probe
Cons-DS	5'-CCCTTACTATTTTTGTAGTTGCTTTTTCG-3'	Downstream PCR primer
2.9-5'consL	5'-AAAGGAGTAACAATGAAAATYATCAAC-3' ^a	Long-distance PCR
<i>ospE</i> -R+C	5'-CCTGCGAATGTATCAGAGTCTCC-3'	Long-distance PCR
<i>p21</i> -R+C	5'-CCCGGAATCCTTAATTTTATTTC-3'	Long-distance PCR
<i>ospF</i> -LD	5'-CATCTATAACTTTGTTTGCCAATTCATTGC-3'	Long-distance PCR

^a Y represents a C or T nucleotide.

to contain OspE/F-like leader peptides. These proteins were designated ElpB1 and ElpB2. The seven plasmid-encoded loci containing the five *ospE* and *ospF* homologs and four *elp* genes characterized to date are diagrammatically represented in Fig. 1.

The strain 297 *OspE*, *OspF*, and *Elp* proteins fall into three distinct homology groups. Dendrogram analysis of the strain

297 *OspE* and *OspF* homologs and the four newly identified polypeptides revealed that they clustered into three major groups: (i) *OspE* and *p21*, (ii) *OspF*, *Bbk2.11*, and *Bbk2.10*, and (iii) the four *Elp* proteins (data not shown). Subsequently, BLAST searches (3) were conducted with the nine *B. burgdorferi* 297 polypeptides to identify similar proteins in the GenBank databases. Matches were obtained for known *OspE* and

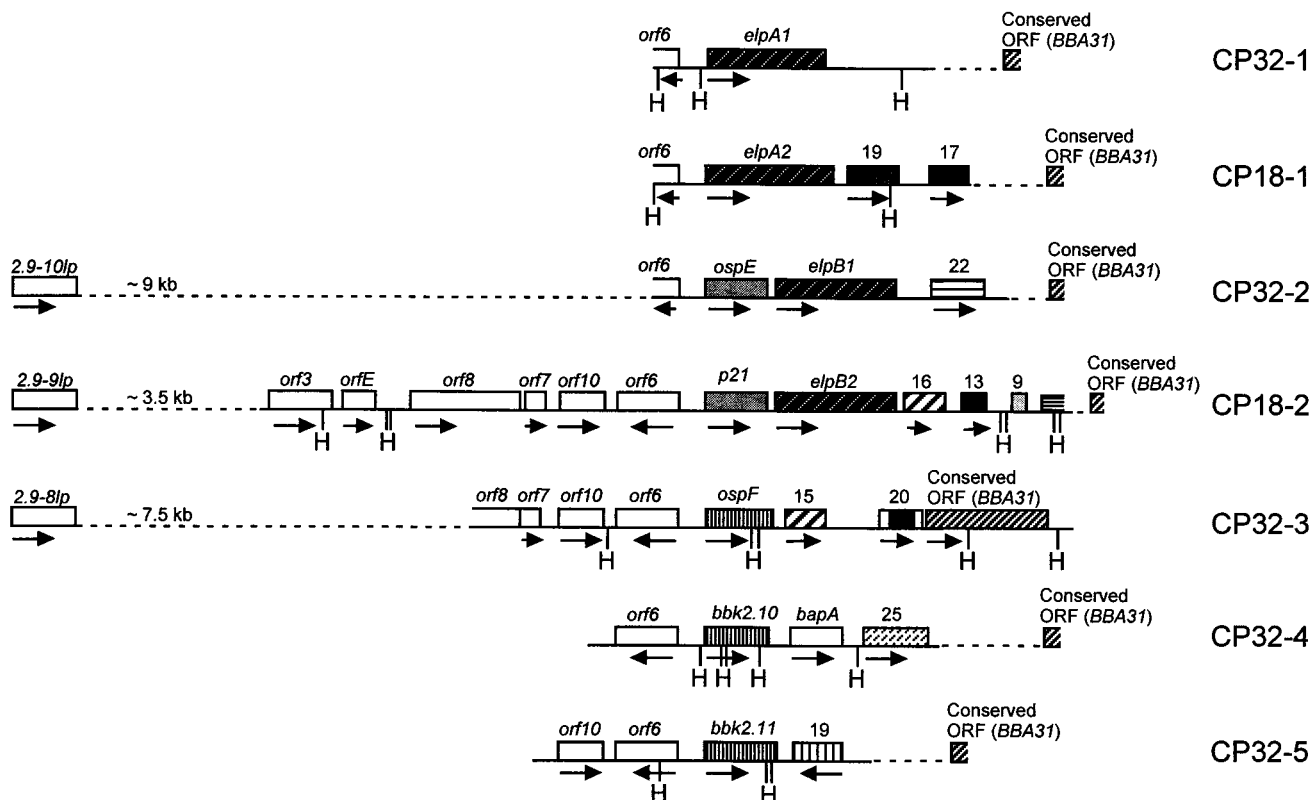


FIG. 1. Schematic representation of the seven circular plasmid loci characterized in *B. burgdorferi* 297 containing *ospE* and *ospF* homologs and *elp* genes. ORFs are shown as boxed regions. The numbers above the ORFs in the downstream regions flanking the lipoprotein-encoding genes indicate the deduced sizes of the encoded polypeptides. The arrows below the ORFs indicate the directions of transcription. H indicates the unique *Hind*III sites for each locus. The dashed regions indicate unsequenced regions. The *OspE*, *p21*, *OspF*, *Bbk2.10*, *Bbk2.11*, and *Elp* proteins determined to be evolutionarily related are shaded the same. ORFs identified downstream of the lipoprotein-encoding loci which are highly similar between the different circular plasmids also are shaded the same.

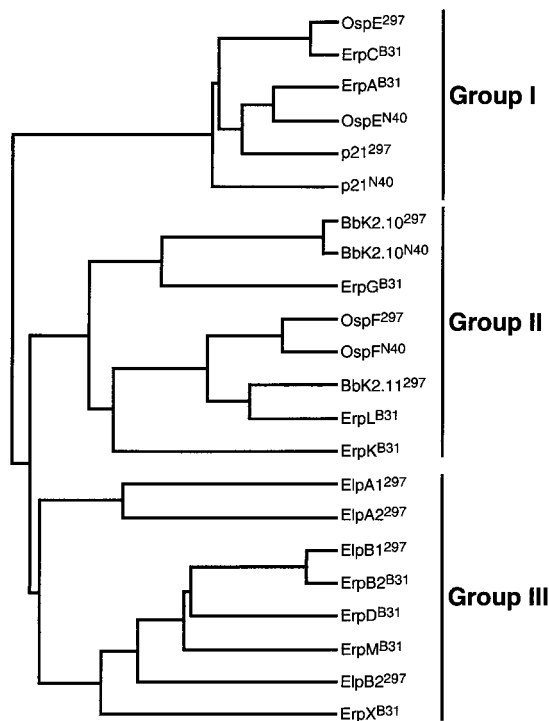


FIG. 2. Dendrogram analysis of *B. burgdorferi* 297, B31, and N40 proteins with OspE/F-like leader peptides. Dendrogram analysis of the *B. burgdorferi* 297, B31, and N40 lipoproteins identified by BLAST searches using the *B. burgdorferi* 297 OspE/F homologs and Elp proteins. The dendrogram was generated by using the PILEUP program from the University of Wisconsin GCG sequence analysis software package, version 8.1. The *B. burgdorferi* B31 ErpI and ErpJ lipoproteins and the strain ZS7 pG lipoprotein were not included in the analysis because they are equivalent to ErpA, ErpB2, and ErpG, respectively (8, 39).

OspF homologs from other borrelial strains, including several which have been partially sequenced, as well as the Erp proteins from the B31 strain (8, 21, 39, 41, 44). We next constructed a dendrogram using all of the completely sequenced proteins identified by the BLAST searches, as well as the 297 polypeptides. Similar to the analysis of the 297 proteins alone, the expanded tree revealed three major groups comprised of OspE-related proteins (group I), OspF-related proteins (group II), and the four Elp proteins plus ErpB2, ErpD, ErpM, and ErpX (group III) (Fig. 2). Multiple sequence alignment of these proteins revealed that the only regions of extensive sequence homology among the three groups are their respective leader peptides (data not shown), suggesting that the mature forms of these proteins might not be evolutionarily related.

Common ancestry of two sequences is typically determined from pairwise alignments by calculating a binary comparison score with a large number (e.g., 100) of random shuffles between the sequences in question (31). The resulting score is expressed in standard deviations with a score equal to or greater than nine standard deviations indicating that the sequences evolved from the same ancestral gene (31). Binary comparison scores were calculated to statistically analyze the percent similarities for all of the proteins in the dendrogram (Table 2). The binary comparison scores within groups were significant (most extremely so), whereas the large majority (138 of 160) of scores between groups were not. It was particularly noteworthy that none of the comparisons between groups I and II yielded significant scores, thereby firmly establishing that the OspE and OspF homologs are evolutionarily

TABLE 2. Binary comparison scores for *B. burgdorferi* 297, N40, and B31 proteins containing OspE/F-like leader peptides^a

Protein	Binary comparison score																								
	ErpC ^{B31}	ErpA ^{B31}	OspE ^{N40}	p21 ²⁹⁷	p21 ^{N40}	BbK2.10 ²⁹⁷	BbK2.10 ^{N40}	ErpG ^{B31}	OspE ²⁹⁷	OspF ^{N40}	BbK2.11 ²⁹⁷	ErpL ^{B31}	ErpK ^{B31}	ElpA1 ²⁹⁷	ElpA2 ²⁹⁷	ElpB1 ²⁹⁷	ErpB2 ^{B31}	ErpD ^{B31}	ErpM ^{B31}	ElpB2 ²⁹⁷	ErpX ^{B31}				
OspE ²⁹⁷																									
ErpC ^{B31}	57 (50)																								
ErpA ^{B31}	48 (44)	47 (46)																							
OspE ^{N40}	55 (43)	44 (49)	45 (39)																						
p21 ²⁹⁷	46 (44)	52 (49)	41 (44)	49 (41)																					
p21 ^{N40}	49 (41)	47 (42)	40 (35)																						
BbK2.10 ²⁹⁷	3 (0)	3 (0)	3 (0)	3 (0)	5 (0)	5 (0)	5 (0)	5 (0)	5 (0)	5 (0)	5 (0)	5 (0)	5 (0)	5 (0)	5 (0)	5 (0)	5 (0)	5 (0)	5 (0)	5 (0)	5 (0)	5 (0)	5 (0)	5 (0)	5 (0)
BbK2.10 ^{N40}	78 (65)	34 (27)	23 (20)	23 (16)	23 (20)	23 (20)	26 (22)	26 (22)	26 (22)	26 (22)	26 (22)	26 (22)	26 (22)	26 (22)	26 (22)	26 (22)	26 (22)	26 (22)	26 (22)	26 (22)	26 (22)	26 (22)	26 (22)	26 (22)	26 (22)
ErpG ^{B31}	32 (29)	32 (29)	23 (18)	23 (16)	23 (18)	23 (16)	21 (18)	21 (18)	21 (18)	21 (18)	21 (18)	21 (18)	21 (18)	21 (18)	21 (18)	21 (18)	21 (18)	21 (18)	21 (18)	21 (18)	21 (18)	21 (18)	21 (18)	21 (18)	21 (18)
OspE ²⁹⁷	19 (16)	19 (16)	19 (16)	21 (19)	18 (13)	18 (13)	18 (16)	19 (18)	19 (18)	19 (18)	19 (18)	19 (18)	19 (18)	19 (18)	19 (18)	19 (18)	19 (18)	19 (18)	19 (18)	19 (18)	19 (18)	19 (18)	19 (18)	19 (18)	19 (18)
OspF ^{N40}	43 (40)	43 (40)	43 (40)	60 (56)	49 (43)	49 (43)	43 (40)	53 (45)	53 (45)	53 (45)	53 (45)	53 (45)	53 (45)	53 (45)	53 (45)	53 (45)	53 (45)	53 (45)	53 (45)	53 (45)	53 (45)	53 (45)	53 (45)	53 (45)	53 (45)
BbK2.11 ²⁹⁷	47 (44)	47 (44)	47 (44)	47 (44)	47 (44)	47 (44)	47 (44)	47 (44)	47 (44)	47 (44)	47 (44)	47 (44)	47 (44)	47 (44)	47 (44)	47 (44)	47 (44)	47 (44)	47 (44)	47 (44)	47 (44)	47 (44)	47 (44)	47 (44)	47 (44)
ErpL ^{B31}	58 (55)	58 (55)	58 (55)	58 (55)	58 (55)	58 (55)	58 (55)	58 (55)	58 (55)	58 (55)	58 (55)	58 (55)	58 (55)	58 (55)	58 (55)	58 (55)	58 (55)	58 (55)	58 (55)	58 (55)	58 (55)	58 (55)	58 (55)	58 (55)	58 (55)
ErpK ^{B31}	29 (19)	29 (19)	29 (19)	29 (19)	29 (19)	29 (19)	29 (19)	29 (19)	29 (19)	29 (19)	29 (19)	29 (19)	29 (19)	29 (19)	29 (19)	29 (19)	29 (19)	29 (19)	29 (19)	29 (19)	29 (19)	29 (19)	29 (19)	29 (19)	29 (19)
ElpA1 ²⁹⁷	13 (10)	13 (10)	13 (10)	13 (10)	13 (10)	13 (10)	13 (10)	13 (10)	13 (10)	13 (10)	13 (10)	13 (10)	13 (10)	13 (10)	13 (10)	13 (10)	13 (10)	13 (10)	13 (10)	13 (10)	13 (10)	13 (10)	13 (10)	13 (10)	13 (10)
ElpA2 ²⁹⁷	17 (17)	17 (17)	17 (17)	17 (17)	17 (17)	17 (17)	17 (17)	17 (17)	17 (17)	17 (17)	17 (17)	17 (17)	17 (17)	17 (17)	17 (17)	17 (17)	17 (17)	17 (17)	17 (17)	17 (17)	17 (17)	17 (17)	17 (17)	17 (17)	17 (17)
ElpB1 ²⁹⁷	80 (81)	80 (81)	80 (81)	80 (81)	80 (81)	80 (81)	80 (81)	80 (81)	80 (81)	80 (81)	80 (81)	80 (81)	80 (81)	80 (81)	80 (81)	80 (81)	80 (81)	80 (81)	80 (81)	80 (81)	80 (81)	80 (81)	80 (81)	80 (81)	80 (81)
ErpB2 ^{B31}	48 (45)	48 (45)	48 (45)	48 (45)	48 (45)	48 (45)	48 (45)	48 (45)	48 (45)	48 (45)	48 (45)	48 (45)	48 (45)	48 (45)	48 (45)	48 (45)	48 (45)	48 (45)	48 (45)	48 (45)	48 (45)	48 (45)	48 (45)	48 (45)	48 (45)
ErpD ^{B31}	52 (44)	52 (44)	52 (44)	52 (44)	52 (44)	52 (44)	52 (44)	52 (44)	52 (44)	52 (44)	52 (44)	52 (44)	52 (44)	52 (44)	52 (44)	52 (44)	52 (44)	52 (44)	52 (44)	52 (44)	52 (44)	52 (44)	52 (44)	52 (44)	52 (44)
ErpM ^{B31}	43 (43)	43 (43)	43 (43)	43 (43)	43 (43)	43 (43)	43 (43)	43 (43)	43 (43)	43 (43)	43 (43)	43 (43)	43 (43)	43 (43)	43 (43)	43 (43)	43 (43)	43 (43)	43 (43)	43 (43)	43 (43)	43 (43)	43 (43)	43 (43)	43 (43)
ElpB2 ²⁹⁷	50 (41)	50 (41)	50 (41)	50 (41)	50 (41)	50 (41)	50 (41)	50 (41)	50 (41)	50 (41)	50 (41)	50 (41)	50 (41)	50 (41)	50 (41)	50 (41)	50 (41)	50 (41)	50 (41)	50 (41)	50 (41)	50 (41)	50 (41)	50 (41)	50 (41)
ErpX ^{B31}	26 (26)	26 (26)	26 (26)	26 (26)	26 (26)	26 (26)	26 (26)	26 (26)	26 (26)	26 (26)	26 (26)	26 (26)	26 (26)	26 (26)	26 (26)	26 (26)	26 (26)	26 (26)	26 (26)	26 (26)	26 (26)	26 (26)	26 (26)	26 (26)	26 (26)

^a Binary comparison scores are indicated in standard deviations for pairwise alignments performed with or without (in parentheses) leader peptides. Statistically significant alignments (≥ 9 standard deviations) are in boldface. Comparisons between polypeptides within the same group are outlined.

unrelated. All of the comparisons between groups which yielded significant scores involved proteins in groups II and III. When these comparisons were repeated using protein sequences without the leader peptides, only the scores for ErpG remained significant (Table 2), supporting our classification of group III proteins as a separate family. The above comparisons between mature portions of lipoproteins were based upon the assumption that evolutionary relationships among exported proteins should be demonstrable regardless of whether the analysis was performed on complete protein sequences or just the biologically relevant (i.e., processed) portion of the protein. To validate this idea, we calculated binary comparison scores for four groups of *B. burgdorferi* lipoprotein homologs (OspA/B, DbpA/B, BmpA-D, and OppA/A1-4). As expected, for each group, the scores were highly significant with or without the leader peptides (data not shown).

The strain 297 *ospE* and *ospF* homologs and *elp* genes are located on cp32 and cp18 homologs. Specific probes for all seven loci (Table 1) were used for Southern analyses of circular plasmids purified from *B. burgdorferi* 297 (29, 33). Each probe hybridized with a plasmid of approximately 32 or 18 kb (Fig. 1). Three lines of evidence indicated that these plasmids are cp32 and cp18 homologs. First, sequence analysis of the regions upstream from each of the lipoprotein genes revealed ORFs (*orf6*, *orf10*, *orf7*, etc.) identical to those previously described upstream from the *erp* genes (8) and the *ospE/F* operon (37) (Fig. 1). Second, between 1.5 and 2.5 kb downstream from each of the lipoprotein-encoding genes is a highly conserved sequence previously used by Casjens et al. (8) as an anchor for PCR mapping of the *erp* loci in the strain B31 cp32s. Finally, consistent with the recent report that at least one of the multicopy 2.9 loci is positioned approximately 10 kb upstream of the *ospE/F* operon in the N40 strain (29, 37), we mapped three of the 2.9 loci to the same position on three separate plasmids in the 297 strain utilizing long-distance PCR with a consensus 2.9 primer and specific primers for the *ospE*, *p21*, and *ospF* genes (Fig. 1 and Table 1).

Evidence for genetic rearrangements at the cp18 and cp32 loci defined by the *ospE* and *ospF* homologs and *elp* genes. Gene fusion and recombination is one possible explanation for the generation of three evolutionarily unrelated families of proteins which contain virtually identical leader peptides. Indeed, in support of this, nucleotide and protein alignments provided evidence that genetic rearrangements have contributed to the sequence diversity of the nine strain 297 lipoproteins. When OspE and p21 were compared, three rearrangements were apparent (Fig. 3A). The first is a 13-amino-acid segment near the N terminus of p21 (the region from 22 to 34) which is a nearly exact duplication of a contiguous downstream stretch (the region from 35 to 47), while the second, an 8-amino-acid stretch in OspE (region 93 to 100), is not present in p21 but is, however, an exact match for a sequence in a corresponding region of ErpC from the B31 strain. Interestingly, these rearrangements are within regions which correspond to recently identified OspE variable domains 1 and 2, respectively (41). The third, which is located in a region of the primary sequence where recombination has not previously been identified, is an apparent seven-amino-acid deletion near the C terminus of OspE which is present in p21 (the region from 169 to 175), as well as the OspE variants from B31 and N40. OspF and Bbk2.11 differ mainly as a result of amino acid substitutions and small one- to four-residue insertions-deletions; they also contain a 17-amino acid stretch near their N termini (amino acids 31 to 47) deleted from Bbk2.10 (Fig. 3B). ElpB2 contains a 39-amino-acid stretch (the region from 172 to 210) which is not present in the other three Elps; however, se-

quences with approximately 75% similarity were identified in ErpB2, ErpD, and ErpM from *B. burgdorferi* B31 (Fig. 3C). Immediately downstream from this potential insertion are five full and one partial tandemly repeated copies of a 10-amino-acid motif, EEE(Q or R)ORRAKE, which is not present in either the other Elps or other proteins in the databases (Fig. 3C). A 12-amino-acid insertion in ElpB1 (residues 78 to 89) is an exact match for a sequence in ErpB2 from the B31 strain. Interestingly, the Kyte-Doolittle, Emini, and Jameson-Wolf algorithms predicted that many of the variable regions identified are hydrophilic, surface exposed, and antigenic, respectively (14). It is tempting to speculate that the variability identified in these regions results from the immunological pressures *B. burgdorferi* encounters within the mammalian host during infection.

Similar to the above analyses, that of Stevenson et al. (36) recently provided evidence that past recombination events have resulted in the variability observed in the genes encoding the Erp proteins of *B. burgdorferi* B31. However, this study did not examine the regions downstream of the lipoprotein-encoding genes. Analysis of these regions for all seven loci in the 297 strain revealed that genetic rearrangements have occurred independently of those involving the contiguous lipoprotein genes (please refer to Fig. 1). For example, unique sequences commence immediately downstream of *ospF*, *bbk2.10*, and *bbk2.11* on plasmids cp32-3, cp32-4, and cp32-5, respectively. Furthermore, the sequences downstream from the *ospE/elpB1* and *p21/elpB2* operons (cp 32-2 and cp18-2, respectively) diverge after only approximately 200 bp, while the region downstream of *elpA1* (cp32-1) appears to be entirely different from those downstream of the other *elp* genes (cp18-1, cp32-2, and cp18-2). Lastly, the second ORF downstream from *elpB2* on cp18-2 is found within a larger ORF which encodes a putative 20-kDa polypeptide downstream of *ospF* on cp32-3.

Implications for *B. burgdorferi* evolution and Lyme disease pathogenesis. It has previously been reported that OspE and OspF homologs represent a closely related family of lipoproteins (8, 24, 39). In fact, Stevenson et al. (39) coined the term Erp (OspE/F-related protein) to reflect this presumptive close relationship among the strain B31 polypeptides. However, when we performed a more rigorous evolutionary analysis, a different phylogenetic picture emerged which clearly demonstrates that the strain 297, N40, and B31 lipoproteins comprise three evolutionarily distinct families which have the same N-terminal leader peptide. The biological function of these and other lipoproteins is unknown, and their role in Lyme disease pathogenesis is poorly understood. However, a major implication of our findings is that the OspE, OspF, and Elp lipoprotein families are likely to perform distinct physiological roles during the borrelial enzootic cycle.

Although lipoprotein leader peptides are constrained by the fact that they must possess a positive charge at the N terminus, a hydrophobic core region, and a signal peptidase II cleavage site, the amino acids which make up these three regions can undergo numerous substitutions without loss of functionality (43). The most likely explanation, therefore, for the high degree of sequence conservation among the leader peptides in all three families is that these N-terminal domains were derived from the same ancestral lipoprotein. On the other hand, this idea seems paradoxical in light of our contention that these proteins comprise three distinct families. These two notions can be reconciled, however, by proposing that the progenitors for each of the families arose from gene fusion events which joined unrelated proteins to the same N terminus. Three lines of evidence support this hypothesis. First, it is now well established that gene rearrangements are widespread in borrelial isolates and are intrinsic to the genomic variability of Lyme

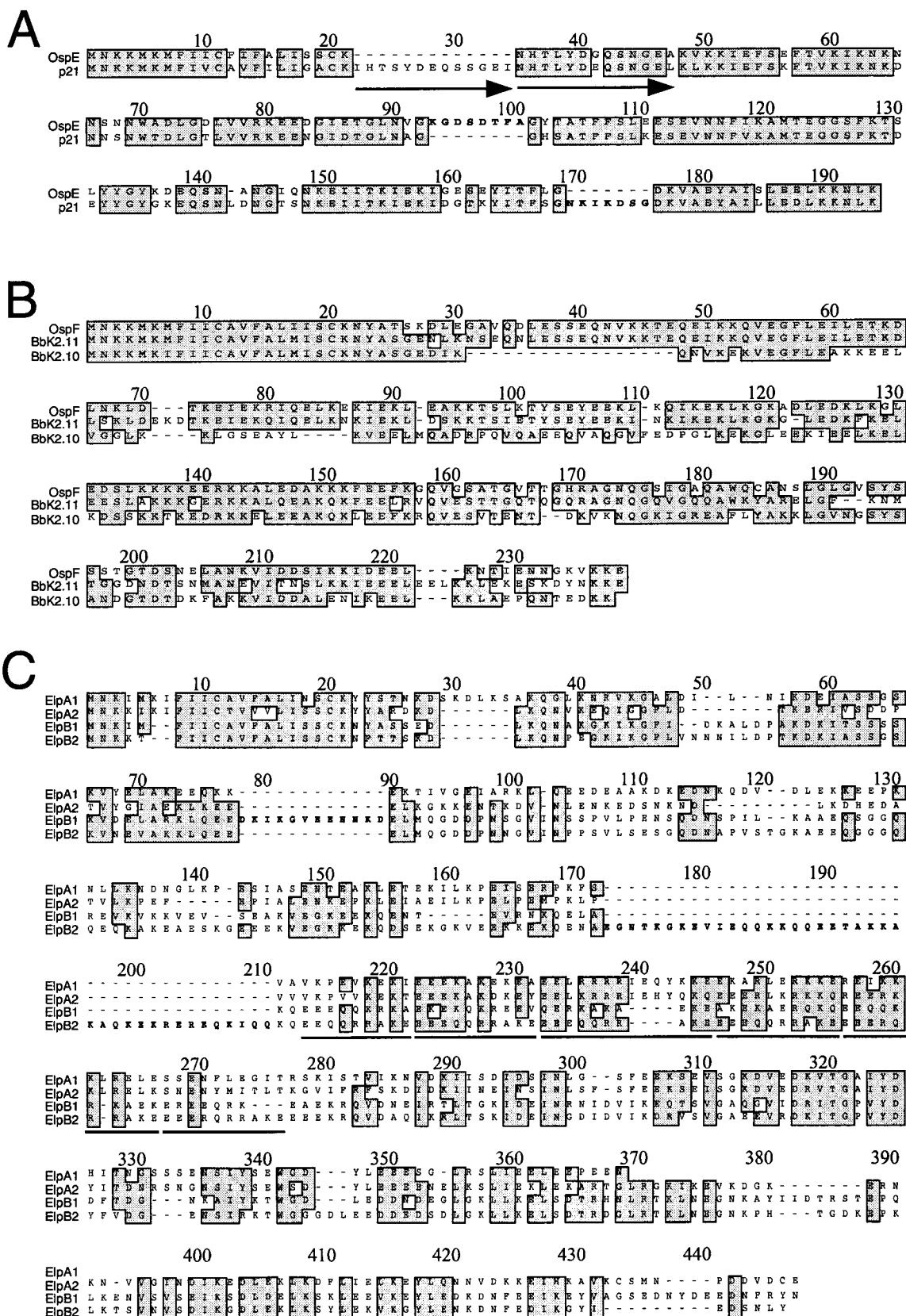


FIG. 3. Evidence for possible genetic rearrangements revealed by sequence alignments of the *B. burgdorferi* 297 OspE and OspF homologs and Elp proteins. ClustalW alignments of OspE and p21 (A); OspF, BbK2.10, and BbK2.11 (B); and ElpA1, ElpA2, ElpB1, and ElpB2 (C). In panel A, arrows and boldface type indicate tandem duplication and possible insertion-deletion, respectively. In panel C, boldface sequences indicate potential insertions-deletions described in text, while tandem repeats in ElpB2 are underlined. In all of the panels, identical and similar residues are shaded.

disease spirochetes, particularly their plasmid component (7, 23, 30, 33, 45). Second, there are well-described precedents in other prokaryotes in which protein families evolved via cassette transfers or gene fusions which resulted in the acquisition of functional domains, including export signals (6, 19, 42). Lastly, and possibly of greatest importance, we and others have shown that genetic rearrangements have played an important role in the generation of sequence diversity within the genes encoding these lipoproteins (36, 41).

Tandemly arranged lipoproteins in *B. burgdorferi* typically are thought to have arisen from gene duplication and subsequent divergence (5, 18). A major implication of our evolutionary analysis is that the *ospE/elpB1* and *p21/elpB2* bicistronic operons had to have arisen via a different mechanism. Based upon the findings that (i) *elp* genes can exist independently and (ii) the regions downstream from the lipoprotein-encoding genes appear to be particularly prone to rearrangements involving relatively large DNA segments, we propose that the bicistronic operons arose by insertion of *elp* genes downstream from the two *ospE* homologs. A final intriguing question is why certain regions of these plasmids appear to be genetically unstable. Circular and linear plasmids of Lyme disease spirochetes are thought to have a common origin and to replicate by a rolling-circle mechanism (15, 16, 23) which is known to be associated with high frequencies of both homologous and illegitimate recombination (25, 28). One can postulate that the hypervariable loci contain a high density of structural features (e.g., palindromes) which impede the progress of the replication fork and increase the likelihood of strand slippage and recombination (28). Regardless of the precise mechanism, the association of hypervariable regions with lipoproteins which are differentially expressed and/or subject to immunological pressures suggests that the Lyme disease spirochete has exploited recombinatorial processes to influence host-pathogen interactions.

Nucleotide sequence accession numbers. Additional sequences have been added to the *B. burgdorferi* 297 *bbk2.10*, *bbk2.11*, *ospF*, *ospE/elpB1*, and *p21/elpB2* loci under GenBank accession no. U18292, U30617, U19754, AF023852, and AF023853, respectively. The *elpA1*, *elpA2*, *2.9-8lp*, *2.9-9lp*, and *2.9-10lp* gene sequences have been given GenBank accession no. AF077602, AF077603, AF046998, AF046999, and AF047000, respectively. The GenBank accession numbers for the strain N40 proteins analyzed in this study are as follows: *OspE*, L13924, *p21*, L32797; *Bbk2.10*, U19105, *OspF*, L13925; *DbpA/B*, AF042796. Those for the strain B31 proteins analyzed are as follows: *ErpA/ErpB2*, U78764; *ErpC/ErpD*, U44914; *ErpL/ErpM*, U72998; *ErpG*, U42598; *ErpK*, U72997; *ErpX*, AF020657; *OppA*, AE000792; *OppA1-3*, AE001139; *OppA4*, AE000790; *OspA/B*, X14407. Those for the strain 297 proteins analyzed are as follows: *DbpA*, U75866; *DbpB*, U75867.

We are indebted to Nick Grishin, Dan Dykhuizen, Milton Saier, and Richard Marconi for helpful discussions.

This work was supported in part by Public Health Service grant AI-29735 from the National Institute of Allergy and Infectious Diseases.

REFERENCES

- Akins, D. R., K. W. Bourell, M. J. Caimano, M. V. Norgard, and J. D. Radolf. 1998. A new animal model for studying Lyme disease spirochetes in a mammalian host-adapted state. *J. Clin. Invest.* **101**:2240–2250.
- Akins, D. R., S. F. Porcella, T. G. Popova, D. Shevchenko, S. I. Baker, M. Li, M. V. Norgard, and J. D. Radolf. 1995. Evidence for *in vivo* but not *in vitro* expression of a *Borrelia burgdorferi* outer surface protein F (*OspF*) homolog. *Mol. Microbiol.* **18**:507–520.
- Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. *J. Mol. Biol.* **215**:403–410.
- 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.
- 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.
- Boesen, T., J. Emmersen, L. T. Jensen, S. A. Ladefoged, P. Thorsen, S. Birke-lund, and G. Christiansen. 1998. The *Mycoplasma hominis* *vaa* gene displays a mosaic gene structure. *Mol. Microbiol.* **29**:97–110.
- Carlyon, J. A., C. LaVoie, S.-Y. Sung, and R. T. Marconi. 1998. Analysis of the organization of multicopy linear- and circular-plasmid-carried open reading frames in *Borrelia burgdorferi* sensu lato isolates. *Infect. Immun.* **66**:1149–1158.
- Casjens, S., R. van Vugt, K. Tilly, P. A. Rosa, and B. Stevenson. 1997. Homology throughout the multiple 32-kilobase circular plasmids present in Lyme disease spirochetes. *J. Bacteriol.* **179**:217–227.
- Champion, C. I., D. R. Blanco, J. T. Skare, D. A. Haake, M. Giladi, D. Foley, J. N. Miller, and M. A. Lovett. 1994. A 9.0-kilobase-pair circular plasmid of *Borrelia burgdorferi* encodes an exported protein: evidence for expression only during infection. *Infect. Immun.* **62**:2653–2661.
- Coleman, J. L., J. A. Gebbia, J. Piesman, J. L. Degen, T. H. Bugge, and J. L. Benach. 1997. Plasminogen is required for efficient dissemination of *B. burgdorferi* in ticks and for enhancement of spirochetemia in mice. *Cell* **89**:1111–1119.
- Das, S., S. W. Barthold, S. S. Giles, R. R. Montgomery, S. R. Telford, and E. Fikrig. 1997. Temporal pattern of *Borrelia burgdorferi* *p21* expression in ticks and the mammalian host. *J. Clin. Invest.* **99**:987–995.
- de Silva, A. M., and E. Fikrig. 1997. Arthropod- and host-specific gene expression by *Borrelia burgdorferi*. *J. Clin. Invest.* **99**:377–379.
- de Silva, A. M., S. R. Telford, L. R. Brunet, S. W. Barthold, and E. Fikrig. 1996. *Borrelia burgdorferi* *OspA* is an arthropod-specific transmission-blocking Lyme disease vaccine. *J. Exp. Med.* **183**:271–275.
- Devereux, J., P. Haerberli, and O. Smithies. 1984. A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.* **12**:387–395.
- Dunn, J. J., S. R. Buchstein, L.-L. Butler, S. Fisenne, D. S. Polin, B. N. Lade, and B. J. Luft. 1994. Complete nucleotide sequence of a circular plasmid from the Lyme disease spirochete, *Borrelia burgdorferi*. *J. Bacteriol.* **176**:2706–2717.
- Ferdows, M. S., P. Serwer, G. A. Griess, S. J. Norris, and A. G. Barbour. 1996. Conversion of a linear to a circular plasmid in the relapsing fever agent *Borrelia hermsii*. *J. Bacteriol.* **178**:793–800.
- Fikrig, E., S. W. Barthold, W. Sun, W. Feng, S. R. Telford, and R. A. Flavell. 1997. *Borrelia burgdorferi* *p35* and *p37* proteins, expressed *in vivo*, elicit protective immunity. *Immunity* **6**:531–539.
- Fraser, C. M., S. Casjens, W. M. Huang, G. G. Sutton, R. Clayton, R. Lathigra, O. White, K. A. Ketchum, R. Dodson, E. K. Hickey, M. Gwinn, B. Dougherty, J.-F. Tomb, R. D. Fleischmann, D. Richardson, J. Peterson, A. R. Kerlavage, J. Quackenbush, S. Salzberg, M. Hanson, R. van Vugt, N. Palmer, M. D. Adams, J. Gocayne, J. Weidman, T. Utterback, L. Watthey, L. McDonald, P. Artiach, C. Bowman, S. Garland, C. Fujii, M. D. Cotton, K. Horst, K. Roberts, B. Hatch, H. O. Smith, and J. C. Venter. 1997. Genomic sequence of a Lyme disease spirochaete, *Borrelia burgdorferi*. *Nature* **390**:580–586.
- Goward, C. R., M. D. Scawen, J. P. Murphy, and T. Atkinson. 1993. Molecular evolution of bacterial cell-surface proteins. *Trends Biochem. Sci.* **18**:136–140.
- Lahdenne, P., S. F. Porcella, K. E. Hagman, D. R. Akins, T. G. Popova, D. L. Cox, J. D. Radolf, and M. V. Norgard. 1997. Molecular characterization of a 6.6-kilodalton *Borrelia burgdorferi* outer membrane-associated lipoprotein (*lp6.6*) which appears to be downregulated during mammalian infection. *Infect. Immun.* **65**:412–421.
- Lam, T. T., T. P. K. Nguyen, R. R. Montgomery, F. S. Kantor, E. Fikrig, and R. A. Flavell. 1994. Outer surface proteins E and F of *Borrelia burgdorferi*, the agent of Lyme disease. *Infect. Immun.* **62**:290–298.
- Lane, R. S., J. Piesman, and W. Burgdorfer. 1991. Lyme borreliosis: relation of its causative agent to its vectors and hosts in North America and Europe. *Annu. Rev. Entomol.* **36**:587–609.
- Marconi, R. T., S. Casjens, U. G. Munderloh, and S. D. Samuels. 1996. Analysis of linear plasmid dimers in *Borrelia burgdorferi* sensu lato isolates: implications concerning the potential mechanism of linear plasmid replication. *J. Bacteriol.* **178**:3357–3361.
- Marconi, R. T., S. Y. Sung, C. A. Norton Hughes, and J. A. Carlyon. 1996. Molecular and evolutionary analyses of a variable series of genes in *Borrelia burgdorferi* that are related to *ospE* and *ospF*, constitute a gene family, and share a common upstream homology box. *J. Bacteriol.* **178**:5615–5626.
- Michel, B., and S. D. Ehrlich. 1986. Illegitimate recombination occurs between the replication origin of the plasmid *pC194* and a progressing replication fork. *EMBO J.* **5**:3691–3696.
- Montgomery, R. K., S. E. Malawista, K. J. M. Feen, and L. K. Bockenstedt. 1996. Direct demonstration of antigenic substitution of *Borrelia burgdorferi* *ex vivo*: exploration of the paradox of the early immune response to outer surface proteins A and C in Lyme disease. *J. Exp. Med.* **183**:261–269.

27. **Nocton, J. J., and A. C. Steere.** 1995. Lyme disease. *Adv. Intern. Med.* **40**:69–117.
28. **Pinder, D. J., C. E. Blake, J. C. Lindsey, and D. R. F. Leach.** 1998. Replication strand preference for deletions associated with DNA palindromes. *Mol. Microbiol.* **28**:719–727.
29. **Porcella, S. F., T. G. Popova, D. R. Akins, M. Li, J. D. Radolf, and M. V. Norgard.** 1996. *Borrelia burgdorferi* supercoiled plasmids encode multiple tandem open reading frames and a lipoprotein gene family. *J. Bacteriol.* **178**:3293–3307.
30. **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.
31. **Saier, M. H., Jr.** 1994. Computer-aided analyses of transport protein sequences: gleaned evidence concerning function, structure, biogenesis, and evolution. *Microbiol. Rev.* **58**:71–93.
32. **Schwan, T. G., J. Piesman, W. T. Golde, M. C. Dolan, and P. A. Rosa.** 1995. Induction of an outer surface protein on *Borrelia burgdorferi* during tick feeding. *Proc. Natl. Acad. Sci. USA* **92**:2909–2913.
33. **Simpson, W. J., C. F. Garon, and T. G. Schwan.** 1990. *Borrelia burgdorferi* contains repeated DNA sequences that are species specific and plasmid associated. *Infect. Immun.* **58**:847–853.
34. **Smith, R. P., Jr., P. W. Rand, E. H. Lacombe, S. R. Telford, S. M. Rich, J. Piesman, and A. Spielman.** 1993. Norway rats as reservoir hosts for Lyme disease spirochetes on Monhegan Island, Maine. *J. Infect. Dis.* **168**:687–691.
35. **Steere, A. C., R. L. Grodzicki, J. E. Craft, M. Shrestha, A. N. Kornblatt, and S. E. Malawista.** 1984. Recovery of Lyme disease spirochetes from patients. *Yale J. Biol. Med.* **57**:557–560.
36. **Stevenson, B., S. Casjens, and P. Rosa.** 1998. Evidence of past recombination events among the genes encoding the Erp antigens of *Borrelia burgdorferi*. *Microbiology* **144**:1869–1879.
37. **Stevenson, B., S. Casjens, R. van Vugt, S. F. Porcella, K. Tilly, J. L. Bono, and P. A. Rosa.** 1997. Characterization of cp18, a naturally truncated member of the cp32 family of *Borrelia burgdorferi* plasmids. *J. Bacteriol.* **179**:4285–4291.
38. **Stevenson, B., T. G. Schwan, and P. A. Rosa.** 1995. Temperature-related differential expression of antigens in the Lyme disease spirochete, *Borrelia burgdorferi*. *Infect. Immun.* **63**:4535–4539.
39. **Stevenson, B., K. Tilly, and P. A. Rosa.** 1996. A family of genes located on four separate 32-kilobase circular plasmids in *Borrelia burgdorferi*. *J. Bacteriol.* **178**:3508–3516.
40. **Suk, K., S. Das, W. Sun, B. Jwang, S. W. Barthold, R. A. Flavell, and E. Fikrig.** 1995. *Borrelia burgdorferi* genes selectively expressed in the infected host. *Proc. Natl. Acad. Sci. USA* **92**:4269–4273.
41. **Sung, S.-Y., C. P. Lavoie, J. A. Carlyon, and R. T. Marconi.** 1998. Genetic divergence and evolutionary instability in *ospE*-related members of the upstream homology box gene family in *Borrelia burgdorferi* sensu lato complex isolates. *Infect. Immun.* **66**:4656–4668.
42. **Vartak, N. B., J. Reizer, A. Reizer, J. T. Gripp, E. A. Groisman, L. F. Wu, J. M. Tomich, and M. H. Saier, Jr.** 1991. Sequence and evolution of the FruR protein of *Salmonella typhimurium*: a pleiotropic transcriptional regulatory protein possessing both activator and repressor functions which is homologous to the periplasmic ribose-binding protein. *Res. Microbiol.* **142**:951–963.
43. **von Heijne, G.** 1985. Signal sequences. The limits of variation. *J. Mol. Biol.* **184**:99–105.
44. **Wallich, R., C. Brenner, M. D. Kramer, and M. M. Simon.** 1995. Molecular cloning and immunological characterization of a novel linear-plasmid-encoded gene, *pG*, of *Borrelia burgdorferi* expressed only in vivo. *Infect. Immun.* **63**:3327–3335.
45. **Zückert, W. R., and J. Meyer.** 1996. Circular and linear plasmids of Lyme disease spirochetes share extensive homology: characterization of a repeated DNA element. *J. Bacteriol.* **178**:2287–2298.

Editor: J. T. Barbieri