

Molecular and Evolutionary Characterization of the cp32/18 Family of Supercoiled Plasmids in *Borrelia burgdorferi* 297

MELISSA J. CAIMANO,¹ XIAOFENG YANG,² TAISSIA G. POPOVA,² MICHAEL L. CLAWSON,¹
DARRIN R. AKINS,³ MICHAEL V. NORGARD,² AND JUSTIN D. RADOLF^{1,4,5*}

Center for Microbial Pathogenesis¹ and Departments of Medicine⁴ and Microbiology,⁵ University of Connecticut Health Center, Farmington, Connecticut 06030; Department of Microbiology, University of Texas Southwestern Medical Center, Dallas, Texas 75235;² and Department of Microbiology and Immunology, University of Oklahoma Health Sciences Center, Oklahoma City, Oklahoma 73190³

Received 20 September 1999/Returned for modification 17 November 1999/Accepted 26 November 1999

In this study, we characterized seven members of the cp32/18 family of supercoiled plasmids in *Borrelia burgdorferi* 297. Complete sequence analysis of a 21-kb plasmid (cp18-2) confirmed that the strain 297 plasmids are similar in overall content and organization to their B31 counterparts. Of the 31 open reading frames (ORFs) in cp18-2, only three showed sequence relatedness to proteins with known functions, and only one, a ParA/SopA ortholog, was related to nonborrelial polypeptides. Besides the lipoproteins, none of the ORFs appeared likely to encode a surface-exposed protein. Comparison with the B31 genomic sequence indicated that paralogs for most of the ORFs in cp18-2 can be identified on other genetic elements. cp18-2 was found to lack a 9- to 10-kb fragment present in the 32-kb homologs which, by extrapolation from the B31 cp32 sequences, contains at least 15 genes presumed to be unnecessary for plasmid maintenance. Sequence analysis of the lipoprotein-encoding variable loci provided evidence that recombinatorial processes within these regions may result in the acquisition of exogenous DNA. Pairwise analysis with random shuffling revealed that the multiple lipoproteins (Mlp; formerly designated 2.9 LPs) fall into two distinct homology groups which appear to have arisen by gene fusion events similar to those recently proposed to have generated the three OspE, OspF, and Elp lipoprotein families (D. R. Akins, M. J. Caimano, X. Yang, F. Cerna, M. V. Norgard, and J. D. Radolf, *Infect. Immun.* 67:1526–1532, 1999). Comparative analysis of the variable regions also indicated that recombination within the loci of each plasmid may occur independently. Last, comparison of variable loci revealed that the cp32/18 plasmid complements of the B31 and 297 isolates differ substantially, indicating that the two strains have been subject to divergent adaptive pressures. In addition to providing evidence for two different types of recombinatorial events involving cp32/18 plasmids, these findings underscore the need for genetic analysis of diverse borrelial isolates in order to elucidate the Lyme disease spirochete's complex parasitic strategies.

Borrelia burgdorferi, the most common arthropod vector-borne pathogen in North America, is maintained in a complex enzootic cycle involving *Ixodes scapularis* and a mammalian host, typically rodents (28, 40). *B. burgdorferi* is distinguished from other prokaryotes by the extraordinary complexity of its genome, which is composed of a linear chromosome and a variable plasmid complement consisting of as many as 21 linear and circular plasmids (6, 18, 20). Two lines of evidence implicate plasmid-encoded proteins in the physiological and antigenic changes which underlie the Lyme disease spirochete's adaptation to different host environments as well as virulence expression within the mammalian host. First, most of the differentially expressed borrelial genes identified to date are plasmid encoded (2, 3, 11, 12, 19, 26, 39, 43, 48, 51, 53, 58). Second, loss of plasmids during continuous in vitro cultivation results in diminished infectivity (35, 42, 61, 63). *B. burgdorferi* plasmids are also of interest because they contain multiple paralogous genes which appear to have undergone extensive recombination during their evolution (2, 5, 8, 10, 15, 20, 39, 45, 49, 64).

The publication of the near-complete genome of a *B. burgdorferi* B31 clone (20) was a major advance in our understanding of this pathogen's complex molecular and evolutionary biology. The entire genome of this strain was not published,

however, due to the difficulties associated with the sequencing and assembly of a family of highly homologous 32-kb circular plasmids (10, 52). Recently, complete sequences for seven of these plasmids have become available from the *B. burgdorferi* genome database (BBGD; <http://www.tigr.org>). Each contains a large number of well-conserved open reading frames (ORFs) as well as two variable loci with genes encoding differentially expressed, polymorphic lipoproteins, some of which have been shown to be surface exposed (2, 3, 13, 19, 27, 32, 39, 49, 51, 53–55). Truncated forms of these supercoiled plasmids (designated cp18s) have been described for the N40 and 297 strains but not fully characterized (2, 50).

B. burgdorferi initially was thought to be a single species. In recent years, however, it has become apparent that *B. burgdorferi* sensu lato is composed of multiple distinct species and genomic groups (4, 24, 31, 59) and that this heterogeneity may relate to differences in disease manifestations (57). Although less marked, genetic diversity also has been detected among the *B. burgdorferi* sensu stricto isolates which cause Lyme disease in North America (30, 33), and recent findings suggest that this heterogeneity also translates into differences in invasive potential (29, 44, 60). An important implication of these findings is that genetic characterization of multiple *B. burgdorferi* isolates may be needed in order to understand the relationship(s) between genomic content, borrelial virulence, and disease pathogenesis. To this end, we have extended our previous studies of cp32/18-encoded, differentially expressed lipoproteins (1–3, 39, 62) by comprehensively analyzing the

* Corresponding author. Mailing address: Center for Microbial Pathogenesis, University of Connecticut Health Center, 263 Farmington Ave., Farmington, CT 06030-3710. Phone: (860) 679-8129. Fax: (860) 679-8130. E-mail: JRadolf@up.uchc.edu.

TABLE 1. Oligonucleotide primers and probes used in this study

Designation	Sequence	Purpose
<i>p21</i> -spec	5'-GCAAGTAATAATGAGTTAAAAGTTAAGCAAAG-3'	Library screening
<i>p21</i> -specR+C	5'-CTTTGCTTAACCTTTAACTCATTATTACTTTGC-3'	Long-distance PCR ^d
<i>elpB2</i> 5'	5'-GAGATCAATGGAGATATTGATG-3'	Library screening
<i>elpB2</i> 3'	5'-CACCACCCCAAGTTTGGCGTATAG-3'	Library screening
<i>ospF</i> -LD-F	5'-CATCTATAACTTTGTTTGGCAATTCATTGC-3'	Long-distance PCR ^{b,c}
2.10-LD-F	5'-GGCCAACAACCAGCTCTTGTGTTTATTTCGC-3'	Long-distance PCR ^{b,c}
2.11-LD-F	5'-GCCCCCGTTGGTTTGTCTCTTATTTCGC-3'	Long-distance PCR ^{b,c}
<i>elpA1</i> -LD-F	5'-GAGCTTTGAGCTTTAATTATTTTGGAGCAGCACAAAGTAGA-3'	Long-distance PCR ^{b,c}
<i>elpA2</i> -LD-F	5'-CGGAATTTTAAATATATTTTACGCTGTC-3'	Long-distance PCR ^{b,c}
<i>elpB1</i> -LD-F	5'-CGTGAAGTGAAAAGTAAAAAAGTAGAAGTATCCG-3'	Long-distance PCR ^{b,c}
<i>elpB2</i> -LD-F	5'-CGTGAAGTGAAAAGTAAAAAAGTAGAAGTATCCG-3'	Long-distance PCR ^{b,c}
<i>mlp</i> 5'-cons	5'-AAAGGAGTAACAATGAAAATYATCAAC-3' ^a	Long-distance PCR ^d
<i>mlp1</i> -spec	5'-GAGACGGTAGTGGCAACTTAATAGAGC-3'	Long-distance PCR ^d
<i>mlp3</i> -spec	5'-GTTACTACTCGCGCAATGGTGGTCC-3'	Long-distance PCR ^d
<i>mlp5</i> -spec	5'-AAGCTAATACTCCACAAGATTGTAATAAATAA-3'	Long-distance PCR ^d
<i>mlp7A</i> -spec	5'-CGGTATAGATACTTTCACGGCTCAAAGC-3'	Long-distance PCR ^d
<i>mlp7B</i> -spec	5'-CAATAGTGTCCAGCAAGCAAAGCAG-3'	Long-distance PCR ^d
<i>mlp8</i> -spec	5'-ACAGAGCAGTCAGATAGTACATGTGGTG-3'	Long-distance PCR ^d
<i>mlp9</i> -spec	5'-AATCCAAACAAATGTAATGATAATAAACATGG-3'	Long-distance PCR ^d
<i>mlp10</i> -spec	5'-CCACATACAAAACCTTTAGTTAAAGAATCTCT-3'	Long-distance PCR ^d
cp32-1 BBP41cons(R+C)	5'-TCGCTTCTAACTTTTCTAGCGGTTAACTTCTG-3'	Long-distance PCR ^c
cp32cons (4306)R+C	5'-CTCGCTCCATAACACATATGTCAGTGTGA-3'	Long-distance PCR ^b
cp32cons (8952)R+C	5'-AAGTCTTACCACCTTCAGAACTAAATCCA-3'	Long-distance PCR ^b
cp32cons (11813)R+C	5'-ACAACAAAAGTGTGTCTATGTACTTTAAGA-3'	Long-distance PCR ^b
cp18/32 (8435) 3' LD	5'-TTTCTCTTAAGTATTTAACTCTTAAGCTTTAAACAGC-3'	Long-distance PCR ^b

^a Y, nucleotides C or T.

^b Deletion analysis.

^c Variable-locus sequence analysis.

^d *mlp* and *ospE/ospF/elp* linkage analysis.

cp32/18 family of plasmids in the 297 strain, a human cerebrospinal fluid isolate (46). In addition to providing evidence for two different types of recombination events involving cp32/18 plasmids, these findings underscore the need for genetic analysis of diverse borrelial isolates in order to elucidate the Lyme disease spirochete's complex parasitic strategies.

MATERIALS AND METHODS

Bacterial strains and media. Virulent low-passage *B. burgdorferi* 297, the source of DNA template for all PCR experiments, was passaged and maintained as previously described (1). Electrocompetent *Escherichia coli* DH5 α and DH10B (Gibco/BRL Life Technologies Inc., Gaithersburg, Md.) cells were used for all transformations. Strains and transformants were grown on tryptone-yeast agar or broth supplemented with the appropriate antibiotic.

Nucleotide sequencing. Greater than 85% of the cp18-2 sequence was derived by analyzing overlapping genomic clones from a Lambda Zap II library (Stratagene, La Jolla, Calif.) (3, 39). Oligonucleotides used as probes for library screening or as primers for PCR amplification of duplex probes are shown in Table 1. The region of cp18-2 extending from *mlp9* to *p21* (Fig. 1) was amplified by long-distance PCR (GeneAmp XL-PCR kit; Perkin-Elmer Corp., Branchburg, N.J.) using the forward and reverse primers *mlp* 5'-cons and *p21*-R+C, respectively (2). The PCR parameters consisted of a preincubation at 94°C for 3 min to fully denature the template DNA, 35 cycles of 94°C for 1 min and 58°C for 8 min, followed by a final incubation at 72°C for 10 min; PCR products were gel purified using the GeneClean glass milk procedure according to the instructions of the manufacturer (Bio101 Inc., Vista, Calif.). Nucleotide sequencing was performed with an Applied Biosystems Inc. model 373A automated DNA sequencer and PRISM ready reaction DyeDeoxy terminator cycle sequencing kits as instructed by the manufacturer (Applied Biosystems, Inc., Foster City, Calif.).

PCR amplification of the truncated portion of cp18-2. Long-distance PCR to obtain the deleted portion of cp18-2 was performed as described above, using the forward and reverse primers listed in Table 1. The products were gel purified for sequencing as previously described.

Linkage of *elpA1*, *elpA2*, *bbk2.10*, and *bbk2.11* with *mlp* genes and analysis of downstream variable regions. Long-distance PCR to link *elpA1*, *elpA2*, *bbk2.10*, *ospF*, and *bbk2.11* with *mlp* genes was performed as described above, using the forward and reverse primers listed in Table 1. The variable regions downstream from *elpA1*, *elpA2*, *elpB1*, *ospF*, *bbk2.10*, and *bbk2.11* were obtained either as genomic clones or by PCR amplification using the forward and reverse primers listed in Table 1. PCR products were cloned into the pCR2.1-TOPO vector (Invitrogen, Carlsbad, Calif.) and sequenced as described above.

Computer analysis. Routine sequence analysis and multiple sequence alignments were performed using the MacVector version 6.5 software (Oxford Molecular Group Inc., Campbell, Calif.). ORFs were analyzed for potential signal peptides and cellular localization using PSORT (34). Transmembrane domains were predicted with the TMpred program (23). To generate an unrooted phylogenetic tree, pairwise sequence alignments were first generated using the ClustalW program (56). Sequence alignments were then analyzed using the ProtDist and Neighbor programs included in the PHYLIP program suite (17). The resulting neighbor-joining tree was viewed using TreeView, version 1.5 (36). Binary comparisons with randomizations were generated using the GAP program from the Genetics Computer Group (Madison, Wis.) sequence analysis software package, version 10.0. Shuffle scores were calculated by subtracting the average quality score based on 100 randomizations from the quality score of the pairwise alignment; the adjusted quality score of the pairwise alignment then was divided by the standard deviation of the average quality score of randomizations. Scores are expressed as the number of standard deviations, with scores ≥ 9 indicating statistically significant alignments (41).

Nucleotide sequence accession numbers. The entire sequence of 297 cp18-2 was submitted to GenBank under accession no. AF169008. Revised sequences for the variable *ospE/ospF/elp* loci have GenBank accession no. U18292 (*bbk2.10*), U30617 (*bbk2.11*), U19754 (*ospF*), AF023852 (*ospE/elpB1*), AF023853 (*p21/elpB2*), AF077602 (*elpA1*), and AF077603 (*elpA2*). Accession numbers for genes encoded within the 2.9 loci can be found in references 39 and 62.

RESULTS AND DISCUSSION

Strategy for molecular characterization of seven cp32/18 plasmids in *B. burgdorferi* 297. The availability of complete sequences for seven cp32 plasmids from *B. burgdorferi* B31 (from the BBGD; see above), in concert with our own previous studies (2, 3, 39, 62), enabled us to devise a strategy for a molecular and evolutionary characterization of seven homologous plasmids in *B. burgdorferi* 297. The strategy was based on the assumption that, as in strain B31, each plasmid consists of large, highly conserved stretches with two interspersed variable regions. We initiated this effort by determining the entire sequence of a truncated family member, using a combination of overlapping genomic library clones and long-distance PCR products. In addition to its smaller size, this plasmid, designat-

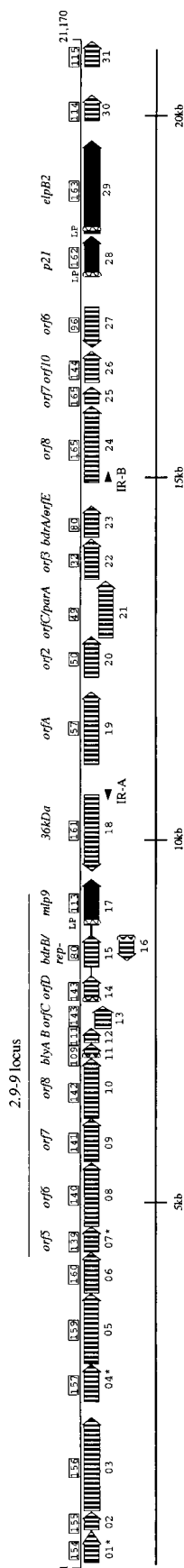


FIG. 1. Schematic linear representation of cp18-2. ORFs are represented as open arrows. Striped arrows represent ORFs belonging to paralogous families in the B31 strain; the boxed numbers above each arrow indicate the *Borrelia* paralogous family to which the ORF belongs (BBGD). Black arrows designate the previously characterized, differentially expressed p21 and Mlp9 lipoproteins. Predicted leader peptides (LP) are represented by curved lines; IRs are indicated as solid arrowheads. The names of previously characterized proteins are given above the corresponding ORFs. Asterisks indicate possible pseudogenes. The approximate distances in base pairs are indicated below.

ed cp18-2 (2), was an attractive candidate for complete sequence analysis for two reasons: (i) its variable loci encode lipoproteins (p21/ElpB2 and Mlp9, respectively) which are selectively expressed during mammalian infection (1, 2, 62) and (ii) loss of this plasmid occurs early during in vitro passage and correlates with attenuation of virulence in the murine model of Lyme disease (D. R. Akins, M. J. Caimano, and J. D. Radolf, unpublished data). Next, to confirm the overall similarity of cp32 plasmids in *B. burgdorferi* B31 and 297, we amplified from other strain 297 cp32 plasmids the fragment presumed to have been deleted from cp18-2. The two unique, variable lipoprotein-encoding loci in each plasmid were then fully sequenced and linked by long-distance PCR. Last, these results enabled us to assess how recombination has influenced the evolution of individual lipoprotein families as well as the variable regions within the plasmids.

Complete sequence of cp18-2. (i) Content and organization. A linear representation of cp18-2 is shown in Fig. 1. The plasmid contains a total of 21,170 bp, 86% of which is coding sequence. cp18-2 is very similar in overall gene content and organization to its untruncated *B. burgdorferi* B31 counterparts. This was expected given the high degree of intra- and interstrain similarity already demonstrated for this family of plasmids. Following the convention previously established by Fraser et al. (20); see also the BBGD website), the 31 ORFs were numbered sequentially; the truncated ORF immediately following the deletion junction (described in detail below) was arbitrarily designated ORF01. Interestingly, all but three ORFs are transcribed in the same direction. The salient features of the 31 genes and corresponding polypeptides are summarized in Table 2. Two genes (ORF04 and ORF07), in addition to ORF01, lack identifiable Shine-Dalgarno sequences and therefore may not be translated. All of the genes except ORF16 (*rep-* in the 2.9 locus [39]) could be assigned to a paralogous family from *B. burgdorferi* B31 (Fig. 1 and Table 2). Paralogs for ORF16 are present in the B31 cp32s but were not annotated (BBGD). The reason for this omission is unclear given that at least one of the *rep-* genes in *B. burgdorferi* 297 has been shown to be transcribed during in vitro growth (39).

(ii) Genomic distribution of paralogous ORFs. The *B. burgdorferi* genome, and its plasmid components in particular, has undergone extensive duplications and rearrangements throughout its evolution (5, 8, 15, 20, 39, 45, 49, 64). Lacking the genomic sequence for strain 297, we used the B31 sequence to examine the genomic distribution of cp18-2 paralogs. This analysis revealed that cp18-2 is a mosaic or patchwork consisting of small, plasmid-specific regions and a number of paralogous sequences distributed to various extents throughout the genome (Fig. 2 and Table 2). The widely distributed ORFs can be grouped into five categories. The first contains two genes (ORF19 and ORF22) with chromosomal, linear, and circular plasmid-encoded paralogs. The second contains two genes (ORF20 and ORF21) with paralogs present on nearly all of the B31 plasmids but not the chromosome. The third contains several ORFs with paralogs located on cp9 and the cp32s. The fourth contains members of 10 paralogous families also represented only on lp54, seven of which (ORF04 to ORF10) form a contiguous stretch on both plasmids (Fig. 2 and Table 2). The fifth group is heterogeneous and contains paralogs present on one or more plasmids (linear and/or circular) other than cp9 and lp54.

(iii) Homologies with other proteins. Database searches revealed that only three ORFs matched proteins with known functions. Two of these (ORF11 and ORF12; previously designated *blyA* and *blyB*, respectively) are both cp32/18 and *B. burgdorferi* specific and encode the hemolytic activity described

TABLE 2. ORFs encoded on *B. burgdorferi* 297 cp18-2

ORF ^a	Alternate name	Family	bp			Orientation ^b	Potential RBS ^c	No. of amino acids	Mol wt	pI	PSORT results			TMpred results		Borrelia paralogous family distribution		
			5'	3'	Leader						Leader	Location	No. of trans-membrane domains	Position	Score	DNA type ^d	Paralog location ^e	
1		154	161	415	255	F	5' truncated	85	9,880	3.98	N	C	0	NA	NA	C/L	A, L, M, N, O, P, Q, R, S	
2		155	415	648	234	F	AAAGAA-11	78	9,343	6.59	N	C	0	NA	NA	C/L	A, L, M, N, O, P, Q, R, S	
3		156	659	1945	1,287	F	TGAGG-5	429	47,854	9.5	N	C ^e	1	134-153	919 ^f	C	L, M, N, O, P, Q, R, S	
4		157	2188	2649	462	F	None	154	17,557	8.93	N	C	0	NA	NA	C/L	A, L, M, N, O, P, Q, R, S	
5		159	2656	3603	948	F	AAAGAA-6	316	36,059	7.96	N	C	0	NA	NA	C/L	A, L, M, N, O, P, Q, R, S	
6		160	3624	4175	552	F	AAAGAA-6	184	20,764	6.23	N	C	0	NA	NA	C/L	A, L, M, N, O, P, Q, R, S	
7	2.9 orf5	139	4208	4537	330	F	None	110	12,848	5.68	N	C	0	NA	NA	C/L	A, L, M, N, O, P, Q, R, S	
8	2.9 orf6	140	4537	5412	876	F	ATGGAGTT-6	292	33,272	5.02	N	C	0	NA	NA	C/L	A, L, M, N, O, P, Q, R, S	
9	2.9 orf7	141	5421	6023	602	F	ATGGATAG-9	201	22,731	9.83	N	C ^e	1	50-67	684 ^f	C/L	A, L, M, N, O, P, Q, R, S	
10	2.9 orf8	142	6036	6845	809	F	AGGGGGG-4	270	30,614	5.64	N	C	0	NA	NA	C/L	A, L, M, N, O, P, Q, R, S	
11	bly4	109	6925	7125	201	F	AAAGGATT-5	67	7,477	9.71	Y, uncleaved	IM	1	18-36	2876	C	L, M, N, O, P, Q, R, S	
12	blyB	111	7132	7476	344	F	AAAGAA-7	115	13,200	5.59	N	C	0	NA	NA	C	L, M, N, O, P, Q, R, S	
13	2.9 orfC	143	7469	7801	333	F	AAAGAA-8	111	12,698	0.18	Y, uncleaved	IM	0	NA	NA	C/L	A, G, L, M, N, O, P, Q, R, S	
14	2.9 orfD	143	7797	8145	349	F	AAAGAA-5	117	13,616	9.22	Y, cleaved ^f	POM ^f	1	5-22	1702	C/L	A, G, L, M, N, O, P, Q, R, S	
15	bhbB	80	8255	8719	465	F	ATAAGGAA-5	155	17,853	4.94	N	IM	1	132-154	2639	C/L	F, G, H, L, M, N, O, P, Q, R, S	
16	rep-	NA	8712	8371	342	R	AAGTGAGG-16	114		10.38	Y, cleaved ^f	IM	3	32-50 3-23	2037 629	NA		
17	mlp9	113	8828	9460	633	F	AAAGGAG-5	211	24,081	5.04	Y, uncleaved ^f	IM ^f	1	72-90	1786	C	L, M, N, O, P, Q, R, S	
18	36kDa	161	10563	9547	1,016	R	AAAGGAGG-6	339	39,970	7.91	N	C	0	NA	NA	C	C, L, M, N, O, P, Q, R, S	
19	orf4	57	10943	12040	1,098	F	AAAGGAGG-6	366	44,309	9.65	N	C	0	NA	NA	C/L/X	X, A, C, D, E, F, G, H, I, L, M, N, O, P, Q, R, S, T, U	
20	orf2	50	12054	12617	563	F	TATGAGG-8	188	22,580	9.66	N	C	0	NA	NA	C/	A, B, C, E, F, G, H, I, J, K, L, M, N, O, P, Q, R, S	
21	orfC/par4	49	12596	13348	753	F	CAAGGA-11	251	29,091	8.8	N	C ^e	1	139-154	711 ^e	C/L	A, B, C, E, F, G, H, I, J, K, L, M, N, O, P, Q, R, S	
22	orf3	32	13393	13950	558	F	TAAGGAG-5	186	22,011	9.51	N	C	0	NA	NA	C/L/X	X, A, B, C, E, F, G, H, I, J, K, L, M, N, O, P, Q, R, S, U	
23	bhl4	80	13963	14484	522	F	TAAGGAGAA-4	174	20,268	5.58	N	IM	1	147-167	2324	C/L	F, G, H, L, M, N, O, P, Q, R, S	
24	orf8	165	15022	16050	1,029	F	AAAGGAGG-6	343	39,666	5.15	N	C	0	NA	NA	C	C, L, M, N, O, P, Q, R, S	
25	orf7	165	16072	16350	278	F	AATGGAGG-9 (L)	93	10,848	9.51	N	C ^e	1	69-89	833 ^e	C	C, L, M, N, O, P, Q, R, S	
26	orf10	144	16405	16827	423	F	AGGAGAGG-6	141	16,745	9.61	N	C ^e	1	52-75	999 ^e	C/L	A, L, M, N, O, P, Q, R, S	
27	orf6	96	17612	16839	774	R	AAAGGAGGAGG-5	258	30,270	9.9	N	C ^e	1	74-90	550 ^e	C/L	C, H, L, M, N, O, P, Q, R, S	
28	p21	162	18043	18597	555	F	TATGGAGG-6	185	20,747	7.19	Y, cleaved	Lipoprotein	1	9-26	1632	C	L, N, P, R	
29	elpB2	163	18628	19845	1,218	F	AAAGGAGGAG-6	406	46,539	5.02	Y, uncleaved	IM ^f	1	6-21	1262	C	F, L, N, O, P, Q	
30		114	19998	20387	390	F	AAAAGG-12 (L)	130	15,661	9.64	Y, uncleaved	IM ^f	1	1-20	803	C	L, N, O, P, Q, R	
31		115	20747	21169	423	F	TTAGGAGGAA-6	141	16,214	9.22	N	C	0	NA	NA	C	L, M, N, O, P, Q, R, S	

^a Potential pseudogenes are indicated in boldface.
^b F, forward; R, reverse.
^c Number following each potential ribosome binding site (RBS) indicates the distance to the AUG start codon. ORFs with putative leucine start codons are indicated by (L).
^d Refers to the type of genetic element on which ORFs belonging to the designated paralogous family have been described in B31: C, circular plasmid; L, linear plasmid; X, chromosome (BBGD).
^e Discrepancy between the presence of a predicted transmembrane domain (TMpred) and cytoplasm localization (PSORT) most likely due to relatively low (<1,000) TMpred score.
^f Experimental studies demonstrated that the 2.9 ORF leader peptide is not cleaved (39).
^g Atypical leader peptide.
^h Mlp leader peptide has been shown to be lipid modified in vitro (39). ElpB2 contains appropriate lipid modification signal sequences.

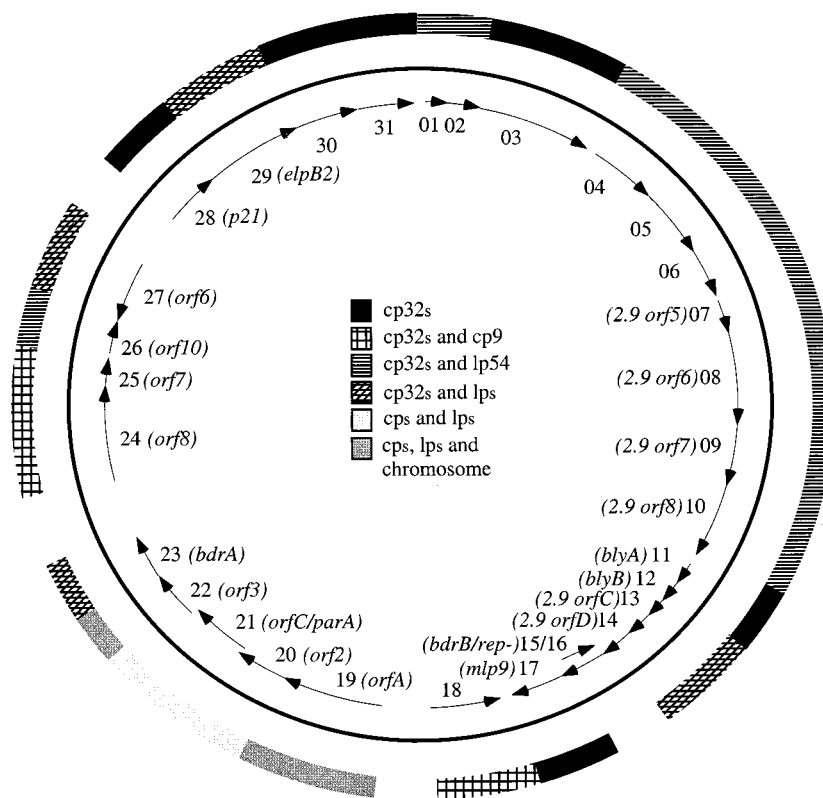


FIG. 2. Distribution of cp18-2 paralogue families within the *B. burgdorferi* B31 genome.

by Guina and Oliver (21). The third gene (ORF21; previously designated *orfC* [64]) belongs to a paralogue family (family 49) whose members recently were shown by Stevenson et al. (49) to be related to the ParA/SopA protein involved in the partitioning of low-copy-number plasmids between daughter cells following cell division (22). As with the B31 paralogs, ORF21 contains the ATP-1 and ATP-2 motifs required for binding and hydrolysis of ATP as well as two motifs postulated to be involved in the binding of accessory proteins or in membrane attachment during partitioning (data not shown) (49). Stevenson and coworkers (49) also proposed that sequence variation among ParA/SopA homologs prevents incompatibility of cp32 homologs within individual borrelial strains; studies are currently under way to examine the sequence heterogeneity of the strain 297 paralogs. In *par*-mediated plasmid partitioning, ParA/SopA interacts with a protein, ParB/SopB, which complexes to a *cis*-acting site, *parS/sopC*, near the origin of plasmid replication (22). The applicability of this scenario to the partitioning of borrelial plasmids, however, is uncertain given that directed searches failed to identify a homolog for ParB or a site resembling *parS/sopC* in cp18-2 or, for that matter, in the *B. burgdorferi* B31 genome. If these functions are encoded on cp18-2, the relevant sequences evidently lack detectable homologies.

(iv) **Export signals and transmembrane domains.** The 31 ORFs were analyzed for N-terminal export signals, transmembrane domains, and cellular location, using the PSORT and TMpred algorithms. A summary of these data are included in Table 2. PSORT identified N-terminal export signals on eight polypeptides (ORF11 [*blyA*], ORF13 [*orfC*], ORF14 [*orfD*], ORF16 [*rep-*], ORF17 [*mlp9*], ORF28 [*p21*], ORF29 [*elpB2*], and ORF30), seven of which have been characterized previously (2, 21, 39, 62). It is noteworthy that PSORT failed to

recognize the signal peptidase II (SPase II) cleavage sites for two of the three lipoproteins (Mlp9 and ElpB2). Although lipid modification of ElpB2 has not been confirmed experimentally, an alkaline phosphatase fusion containing the Mlp9 leader peptide is lipidated when expressed in *E. coli* (39). Of the five nonlipoproteins with predicted leader peptides, only ORF14 and ORF16 were predicted to have SPase I cleavage sites. In both cases, however, these predictions are questionable. The putative leader sequence of ORF16 has three prolines within the hydrophobic core region and a predicted cleavage site 30 amino acids downstream from the end of the hydrophobic stretch. With respect to ORF14, previous analyses using Triton X-114 phase partitioning and *in vivo* lipid labeling of the recombinant protein expressed in *E. coli* indicated that the protein's leader peptide also is not cleaved (39). TMpred identified potential internal transmembrane domains, predictive of a cytoplasmic membrane location, in 10 ORFs, while one (ORF16) was predicted to have two internal transmembrane domains in addition to its leader peptide. PSORT, in contrast, failed to identify the putative transmembrane domains for 6 of these 10 polypeptides, placing them instead within the cytoplasmic compartment. These discrepancies probably reflect the relatively low TMpred scores (<1,000) for all seven membrane-spanning regions. The two proteins predicted by both PSORT and TMpred to possess internal transmembrane domains were ORF15 and ORF23. These two polypeptides are members of the family (family 80) of borrelia direct repeat (Bdr) proteins (formerly *rep*⁺ within the 2.9 loci [39] and *orfE* [64]) which characteristically possess an internal, tandemly repeated motif with K-I-D core sequence and C-terminal cytoplasmic membrane anchors (9, 65). The above analyses clearly highlight the need for caution when interpreting computer-based predictions of protein topology and cellular location.

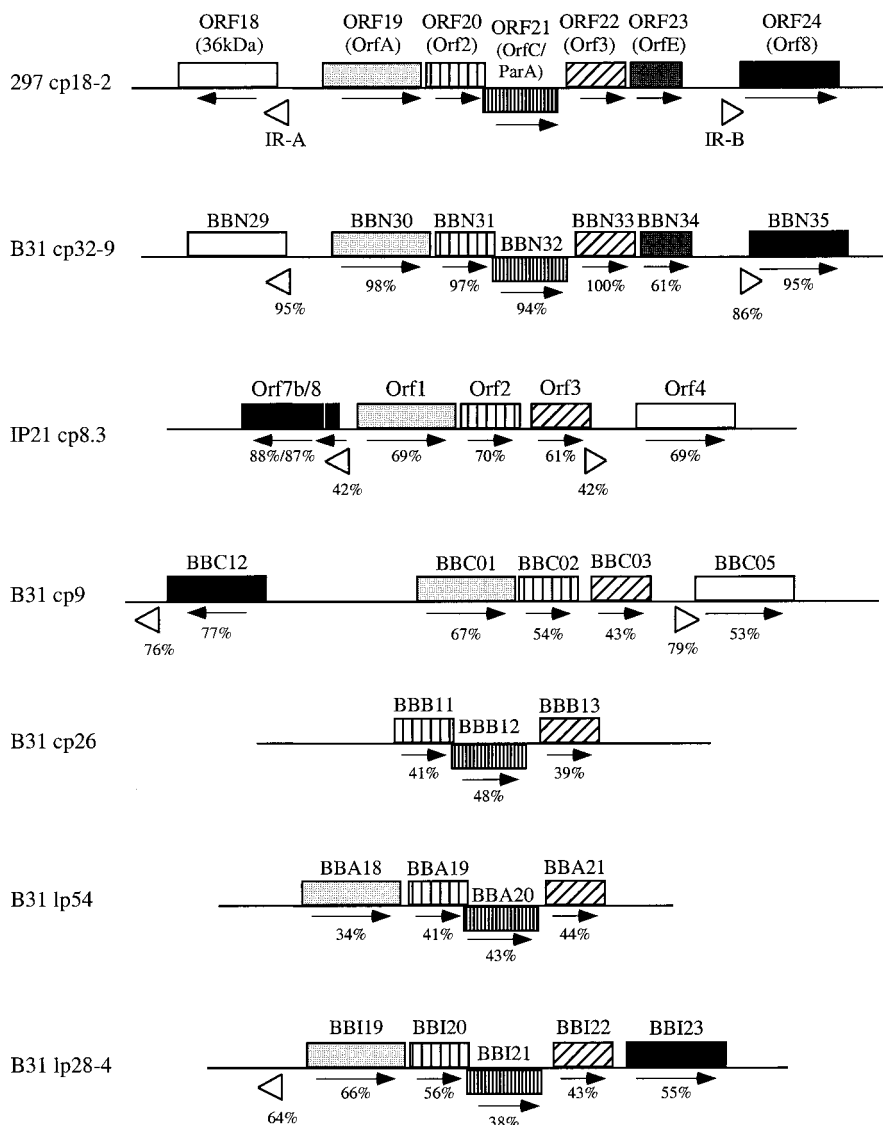


FIG. 3. Organization of IRs and/or related flanking ORFs in selected *B. burgdorferi* plasmids. ORFs are depicted as shaded boxes, with similarly shaded boxes indicating paralogous ORFs; the arrow below each ORF indicates the direction of transcription. Percent similarities between cp18-2 and corresponding paralogs (using amino acid sequences) and IRs are indicated.

Equally important, they indicate that most of the coding capacity of the plasmid is devoted to polypeptides residing within intracellular compartments.

(v) **Inverted repeats.** Flanking the stretch from ORF18 to ORF24 are 185- and 191-bp inverted repeats (IRs) which are 42% identical to the 183- and 184-bp IRs on cp8.3 of *B. burgdorferi* Ip21 (15) (Fig. 1 and 3). The cp18-2 IRs are similar to their cp8.3 counterparts in that they overlap the Shine-Dalgarno sequences and translational starts of the adjacent genes (Fig. 3). They differ, however, in that they are theoretically unable to form stem-loop structures comparable to those which were proposed to interfere with translation of cp8.3 transcripts (15) (data not shown). Three of the genes (ORF19, ORF20, and ORF22) flanked by the IRs in cp18-2 are paralogs for the genes flanked by the IRs in cp8.3; it is noteworthy that the *parA/sopA* homolog is one of the two genes missing from this region in cp8.3 (Fig. 3). A survey of the B31 genome revealed that similar IRs are found only on the cp32 family and cp9 and that the cp32 IRs flank the same five ORFs as in

cp18-2 whereas the cp9 IRs flank the same three ORFs as in cp8.3. (20, BBGD). Variable combinations of these five ORFs are found in similar configurations on B31 linear plasmids which lack the IRs (Fig. 3). One possible interpretation of these observations is that the IR-flanked genes were acquired initially by cp32/18 and/or cp9 plasmids (perhaps via a transpositional insertion) and that various combinations of these genes subsequently recombined into other plasmids. The absence of the *ParA/SopA* homologs in cp8.3 and cp9 raises the intriguing possibility that the replication and partitioning mechanisms used by these small plasmids differ from those of the larger supercoiled plasmids. In this regard, it should be noted that plasmids which replicate as rolling circles typically are less than 10 kb (25) and that a *B. burgdorferi* chromosomal gene (BB0607) encodes a protein related to the DNA helicase involved in rolling circle replication in *Bacillus subtilis* (37).

Identification of the cp18-2 deletion. Stevenson and coworkers (50) showed that the truncated cp32 homolog in *B. burgdorferi* N40 lacks an approximate 12.5-kb fragment. However,

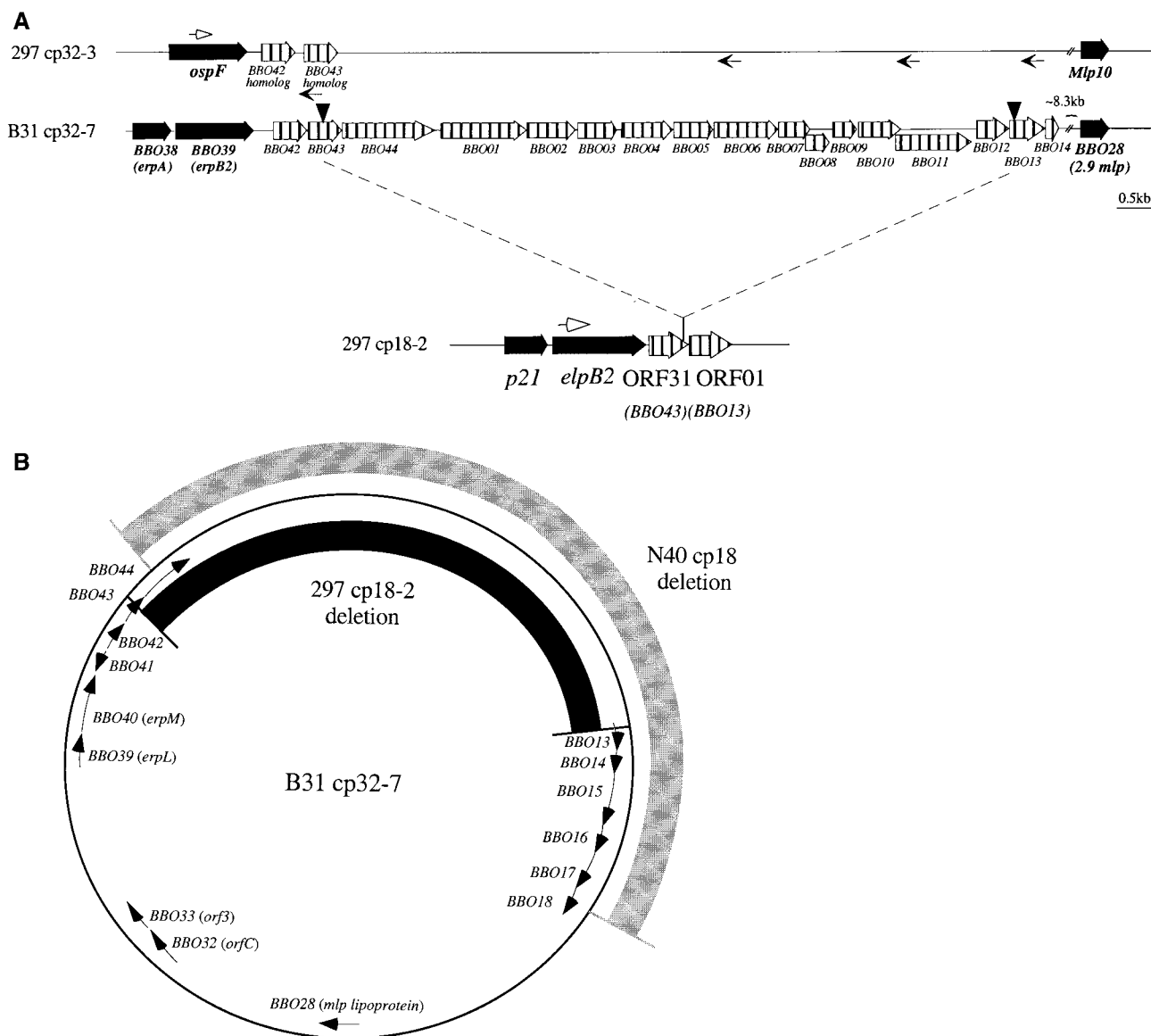


FIG. 4. The cp18-2 deletion. (A) cp32 ORFs predicted to be deleted from cp18-2. Small arrows designate the nested primer pairs used to confirm the similarity between the cp18-2 deletion in a full-length 297 cp32 (represented by cp32-3) and the corresponding segments from the B31 homologs (represented by cp32-7). (B) The deletions in cp18-2 and cp18 of N40 are shown alongside a representative B31 plasmid (cp32-7).

because their study predated the availability of the B31 cp32 sequences, they were unable to identify the missing ORFs. Early in our sequencing effort, it became apparent that the N40 and 297 truncations differed substantially. This observation prompted us to define the deletion in cp18-2. Using 5' primers specific for each of the separately encoded *ospE*, *ospF*, and *elp* genes and a reverse primer located within ORF01 (Table 1), 9- to 10-kb fragments were amplified from strain 297 DNA (Fig. 4A and data not shown). Additional PCRs using the same 5' primers and nested reverse primers (Table 1) generated smaller fragments of the exact sizes predicted from the 297 and B31 sequences (Fig. 4A and data not shown). Together with the cp18-2 sequence, these results established the high degree of similarity between the strain B31 and 297 cp32s.

The cp18-2 deletion junction consists of a fusion between ORF31 and ORF01 (Fig. 4A). As a result, ORF31 lacks the 45 amino acids at the C termini of its family 115 paralogs (repre-

sented by BBO43), possessing seven unique C-terminal residues instead (data not shown). ORF01, on the other side of the deletion junction, lacks both a Shine-Dalgarno sequence and translational start; from its second residue on, ORF01 is identical to its family 154 paralogs (beginning with amino acid 14) but terminates at amino acid 64 due to an independent frameshift (data not shown). A comparison of the cp18-2 and N40 cp18 deletions (Fig. 4B) shows that the N40 deletion begins within the first cp32 ORF deleted from cp18-2 (represented by BBO44) and extends five ORFs further downstream to the family 160 paralog (represented by BBO18). The simplest explanation for these overlapping deletions is that they occurred separately but involved genes which are nonessential for stable plasmid maintenance and/or which can be complemented in *trans* by functions encoded elsewhere. The latter possibility is supported by the presence of highly similar paralogs for each of the deleted ORFs on the full-length cp32/18 plasmids in

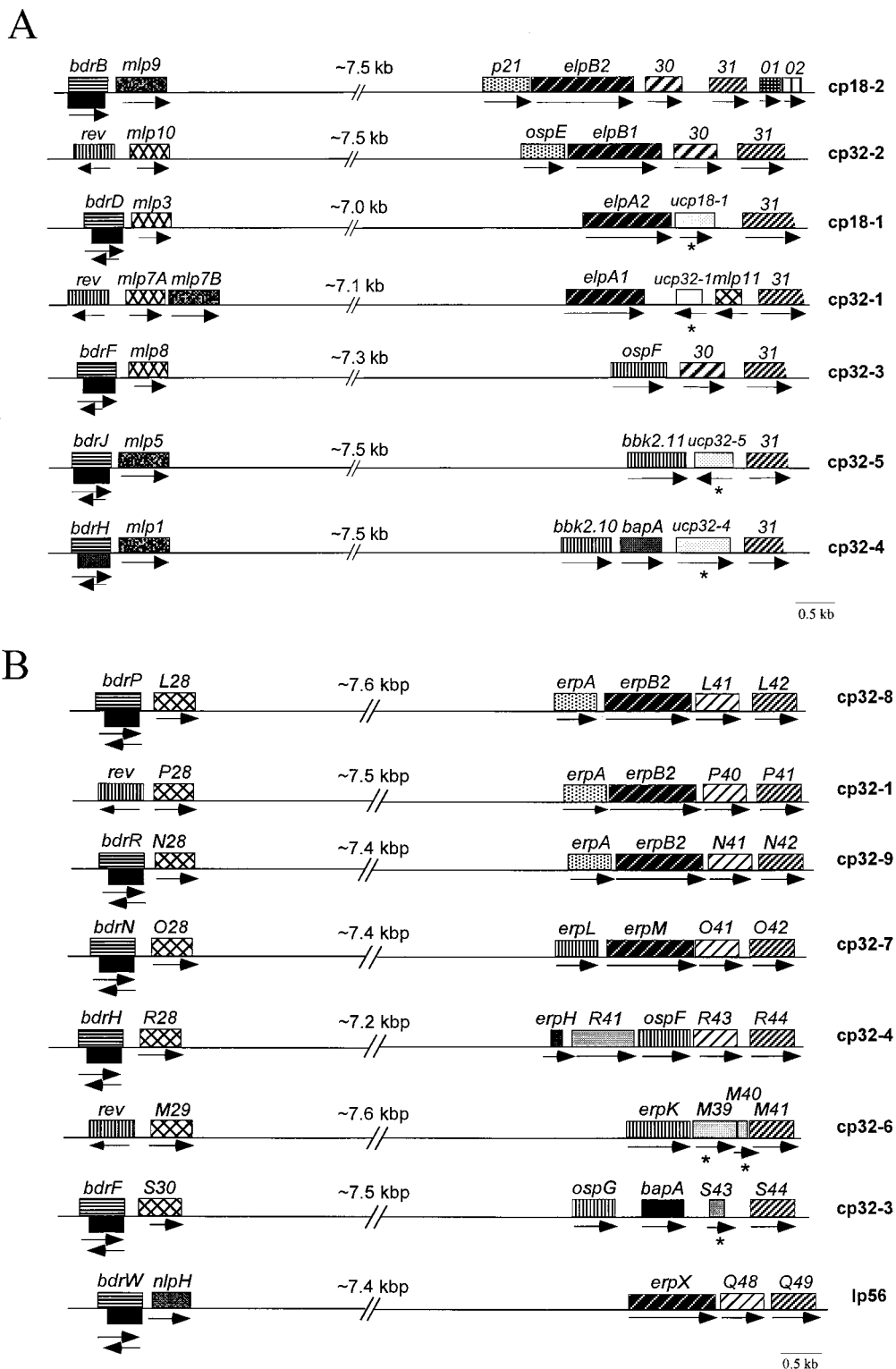


FIG. 5. Schematic representation of the variable loci in *B. burgdorferi* 297 (A) and B31 (B). Similarly shaded boxes are used to indicate paralogous ORFs; the arrow below each ORF indicates the direction of transcription. In panel A, paralogs of cp18-2 ORF30 and ORF31 are designated 30 and 31; nonparalogous ORFs are indicated by asterisks.

strain B31 (BBGD). Database searches of the 15 B31 ORFs presumed to be encoded by the cp18-2 deletion failed to reveal significant homologies with other prokaryotic proteins. PSORT predicted that all 15 genes encode proteins lacking N-terminal

export signals and that 13 of the 15 are cytoplasmic (data not shown), further underscoring the paucity of plasmid coding capacity devoted to surface-exposed proteins. TMpred predicted that three genes (BBO03, BBO08, and BBO44) encode pro-

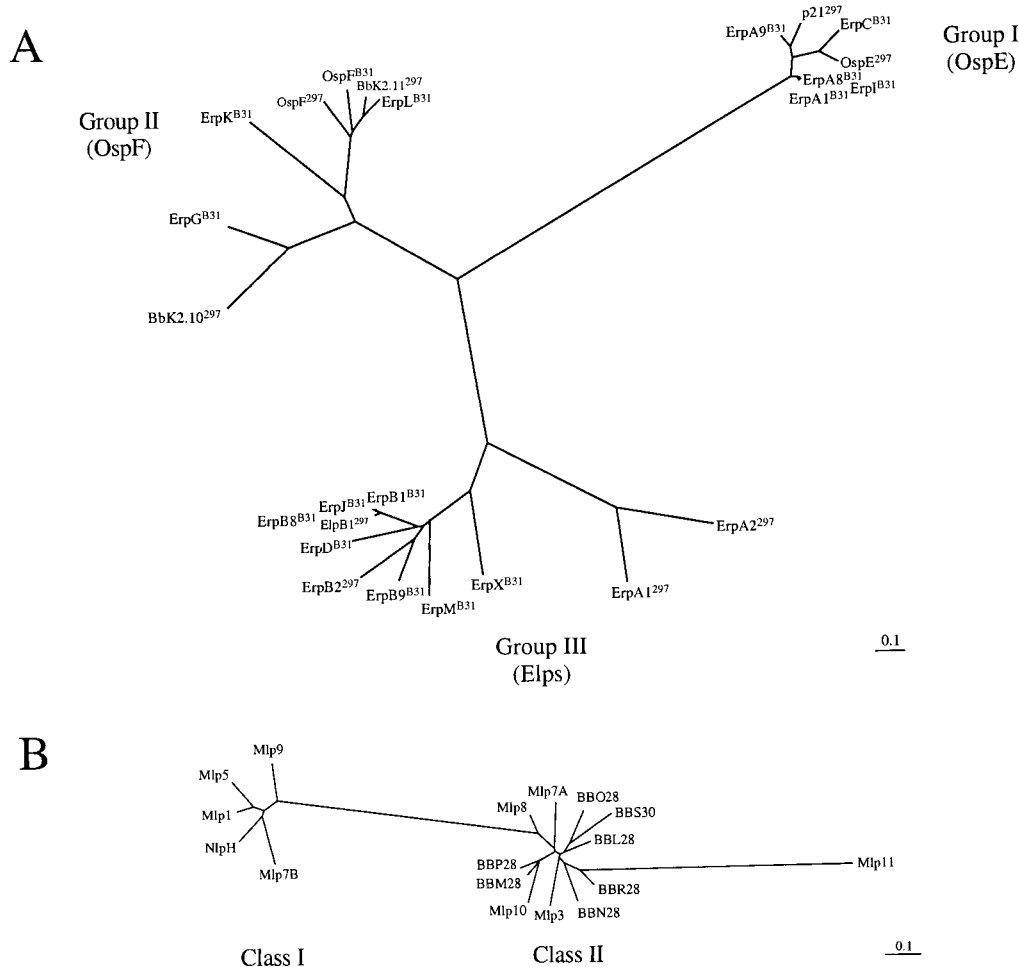


FIG. 6. Unrooted neighbor-joining phylogenetic trees of OspE, OspF, and Elp (A) and Mlp (B) homologs in *B. burgdorferi* 297 and B31 cp32/18 plasmids. Scale bars represent the base pair substitutions per site.

teins with a single internal transmembrane domain (data not shown). As before, the discrepancies between the two programs probably reflect the relatively low TMpred scores for these three putative membrane-spanning regions.

Evolutionary analysis of the cp32/18-specific, variable loci. Sequence divergence among cp32/18 plasmids can be attributed primarily to variable regions defined by the genes encoding the Mlp (formerly 2.9-LP [39]) and OspE/OspF/Elp lipoprotein antigens (2). A characterization of these plasmid-specific loci, therefore, was essential for definition of the *B. burgdorferi* 297 homologs. Sequences for the loci encoding the Mlp lipoproteins were presented in two prior studies (39, 62). In a third report (2), we provided partial sequences for the seven OspE/OspF/Elp-encoding regions and established linkage between the loci containing the *ospF*, *p21/elpB2*, and *ospE/elpB1* genes and 2.9 loci containing the *mlp8*, *mlp9*, and *mlp10* genes, respectively (2). To complete this phase of the study, we sequenced all remaining gaps and linked the four unpaired *ospE/ospF/elp* loci with *mlp* loci by long-distance PCR.

Schematics for the variable loci of all seven plasmids are shown in Fig. 5A. Though corresponding variable loci are clearly similar in configuration, a variety of rearrangements, insertions, and deletions, believed to be reflective of past recombination events (2, 49), are readily apparent. Two discoveries relating to the *ospE/ospF/elp* loci are of particular importance because they suggest that these localized recombination

events may involve exogenously acquired DNAs. The first is that a *mlp* paralog (*mlp11*) resides downstream from *elpA1* in cp32-1; to the best of our knowledge, this is the only instance in which a *mlp* gene has been identified outside of a 2.9 locus. Phylogenetic analysis (Fig. 6B) and sequence alignments (not shown) demonstrate that this lipoprotein is an outlier from the other 297 Mlps, leading us to speculate that it was acquired from another borrelial strain. The second is that cp18-1, cp32-1, cp32-4, and cp32-5 from 297 contain nonparalogous ORFs in place of the cp18-2 ORF30 paralogs present in the other three plasmids. Because these unique circular plasmid-encoded (*ucp*) genes lack homologies with other sequences within the B31 genome, it seems highly plausible that they were acquired as the result of some form of horizontal gene transfer. In light of the relatively small proportion of total cp32/18 coding capacity devoted to exported proteins, it was of interest to note that three *ucp* genes (*ucp32-1*, *ucp32-4*, and *ucp18-1*) were predicted by PSORT to encode proteins with N-terminal export signals and that the signal sequences of two (*ucp32-1* and *ucp32-4*) have putative SPase I cleavage sites.

Recently, we presented evidence that the OspE, OspF, and Elp homologs within the 297 and B31 strains arose as fusions between a common N terminus and three unrelated ancestral proteins (2). The phylogram in Fig. 6A supports this by demonstrating that these proteins segregate into three distinct homology groups. In light of the evidence for past rearrange-



FIG. 7. Sequence relatedness among the Mlp lipoproteins. The sequences of class I (A), class II (B), and the entire complement of proteins (C) were aligned using the ClustalW algorithm. Shaded residues represent identical or conserved matches.

ments at the 2.9 loci (Fig. 5A), it seemed logical to consider the possibility that analogous fusion events contributed to the evolution of the Mlps. We began this line of inquiry by first assessing the phylogenetic relationships among Mlps; the phylograms in Fig. 6B show that the Mlps fall into two divergent classes. Sequence alignments show that the proteins in each class exhibit extensive sequence similarity across their entire lengths (Fig. 7A and B); sequence relatedness for the entire complement, in contrast, falls off strikingly after position 51 and is highly dependent on the introduction of large gaps in order to accommodate the smaller class II proteins (Fig. 7C). A similar degradation in sequence relatedness among the OspE, OspF, and Elp homologs was an important clue to the potential contribution of gene fusion to the evolutionary histories of these lipoproteins (2). Consequently, we examined the Mlps for common ancestry by calculating binary comparison scores for pairwise alignments with 100 randomizations (41). The resulting score is expressed in standard deviations,

with a score equal to or greater than 9 standard deviations being strong evidence that the sequences in question arose from the same ancestral gene. As shown in Table 3, the binary comparison scores for full-length proteins were all significant and extremely so among proteins of the same class. These scores fell off markedly, and in many cases became insignificant, for interclass comparisons of proteins without their leader sequences (data not shown). When these comparisons were repeated with the first 51 amino acids of each protein omitted, significant scores were obtained only for lipoproteins within the same class (Table 3). The most straightforward interpretation of these results is that the two Mlp classes evolved from chimeric progenitors in which shared N termini were joined to unrelated proteins. One implication of this conclusion is that the tandemly arranged genes in cp32-1 (*mip7A/mip7B*) encoding class I and class II Mlp lipoproteins had to have arisen by a different mechanism than the gene duplication events thought to have generated other tandemly

TABLE 3. Binary comparison scores for *B. burgdorferi* 297 Mlp lipoproteins^a

Protein	Mlp3	Mlp5	Mlp7A	Mlp7B	Mlp8	Mlp9	Mlp10
Mlp1	18 [0]	154 [113]	18 [0]	102 [91]	26 [0]	128 [76]	29 [0]
Mlp3		12 [0]	88 [63]	18 [0]	75 [58]	16 [0]	90 [56]
Mlp5			14 [0]	144 [107]	28 [0]	145 [94]	30 [0]
Mlp7A				20 [0]	87 [61]	17 [0]	83 [62]
Mlp7B					25 [0]	129 [121]	20 [0]
Mlp8						25 [0]	86 [65]
Mlp9							31 [1]

^a Binary comparison scores are indicated in standard deviations for pairwise alignments performed using either full-length proteins or proteins without the first 51 amino acids (bracketed). Statistically significant alignments (≥ 9 standard deviations) are in boldface.

arrayed genes encoding lipoproteins (e.g., OspA and OspB) with sequence homology along their entire lengths.

To complete this analysis, we used the lipoprotein genes and other ORFs as physically linked markers to assess the interdependence of recombinatorial processes within the variable loci. The data in Fig. 5A and 6, taken together, argue strongly that recombination events occur independently at each locus. For example, whereas the *p21/elpB2* and *ospE/elpB1* loci of cp18-2 and cp32-2 are highly similar, they differ significantly at the 2.9 loci with respect to both the class of Mlp and the presence of *bdr/rep-* (cp18-2) as opposed to *rev* (cp32-2) genes. Similarly, cp32-3 and cp32-5 encode closely related *ospF* paralogs but have different downstream ORFs at this locus (an ORF30 paralog versus *ucp32-5*) as well as *mlp* genes of different classes at their 2.9 loci. Plasmids cp18-1 and cp32-1 both have *elpA* genes but otherwise differ markedly at both loci. Last, although the two truncated plasmids cp18-1 and cp18-2 both contain distantly related *elp* genes, one is bicistronic and paired with a lipoprotein gene (*p21*) of a different class, indicating a subsequent downstream insertion; at the other locus the two plasmids contain *mlp* genes of different classes. Interestingly, the importance of recombination as a mechanism for generating sequence diversity at the variable loci is less apparent if one repeats this analysis using the B31 cp32 plasmids. The combined data in Fig. 5B and 6 reveal the two principal reasons for this. The first is that there is less sequence diversity among the B31 variable-locus lipoproteins; this is particularly true for the Mlp-like paralogs which, with a single exception (NlpH on lp56), fall into a single homology group. The other is the greater redundancy of the *erp* loci, most notably the three virtually identical *erpA/erpB2* pairings. A direct comparison of the 297 and B31 variable loci (Fig. 5A and B, respectively) is also noteworthy because it reveals numerous other differences in configuration and lipoprotein gene content, as well as the presence of unrelated nonparalogous genes, which, as a whole, may be indicative of divergent adaptive pressures acting on the two isolates.

Summary and conclusions. In the present study, we exploited existing sequence information and the highly conserved nature of the cp32/18 plasmids in order to characterize seven homologs from *B. burgdorferi* 297. Though clearly sequence intensive, a major advantage of our strategy was that it enabled us to delineate the principal features of an entire family of plasmids without having to completely sequence each member. Inherent in this approach, however, is the need to extrapolate from heterologous genomic sequences. Nevertheless, we believe that any potential inaccuracies due to genomic polymorphisms between strains 297 and B31 are unlikely to alter our major findings and are easily counterbalanced by the plethora of new genetic information pertaining to a highly virulent,

neurotropic clinical isolate (46). The need for such information is highlighted by our finding, using both needle and tick inoculation in mouse infectivity tests, that the B31 clone used for the genomic sequence is less virulent than either the parental B31 isolate or *B. burgdorferi* 297 (M. J. Caimano, D. R. Akins, S. K. Wikel, and J. D. Radolf, unpublished data). While these differences could reflect the loss of unrelated sequences during cloning, it is our working hypothesis that the greater virulence of *B. burgdorferi* 297 reflects, at least in part, the greater sequence diversity of its cp32/18 plasmids. Pending the development of efficient mutagenesis techniques for virulent borrelial isolates, it may not be possible to directly evaluate this conjecture in the near future. Nevertheless, the work reported here sets the stage for a variety of studies ranging from expression mapping and cellular localization of the cp32/18-encoded gene products in spirochetes cultivated under various in vitro and in vivo conditions to molecular epidemiological correlations of spirochetal genetic diversity and disease expression.

It is now well established that recombinatorial processes have conferred a remarkable degree of plasticity upon the *B. burgdorferi* genome, presumably in the context of enhancing its adaption to diverse ecological niches. Based on the analyses reported here, we now propose that two different types of recombinatorial events have occurred within the cp32/18 family. The first involved exchanges between these plasmids and the bacterium's other genetic elements. This process appears to have been relatively unrestricted in that it utilized diverse donor sequences and recipient sites; its principal constraint was that it was limited to sequences already present within the borrelial genome. One of its primary functions appears to have been plasmid building, that is, generating the scaffolding or framework needed to assemble novel genetic elements. The second type of recombinatorial event, in contrast, was confined to relatively small regions of the cp32/18 plasmids but was less restrictive in that it utilized nonparalogous DNAs, including exogenous genes. A major consequence of this second type of event was to individualize plasmids via the variable regions. While the nonparalogous ORFs present in both the B31 *erp* and 297 *ospE/ospF/elp* variable loci are the most tangible evidence for this second process, we believe that it also was involved in the importation of the novel sequences which fused to form the progenitors of the lipoprotein families within the variable loci. Some evidence for this is provided by the *mlp* outlier in the *elpA1* locus; in fact, the proximity of *ucp32-1* and *mlp11* in cp32-1 (Fig. 5A) suggests that they were introduced on the same recombining DNA fragment. A final question concerns the milieu which serves as the incubator for lateral gene exchange. With the exception of the recently described *vls* locus (63), it has not been possible to demonstrate genetic rearrangements in Lyme disease spirochetes during the mammalian phase of infection (16, 47). Because of the low spirochetal burdens in *B. burgdorferi*-infected tissues (7), it seems more likely that such exchanges occur during and shortly after the tick feeding phase(s) of the enzootic cycle, when spirochetes are in close proximity, highly metabolically active, and actively replicating (14, 38) and that novel antigenic phenotypes are subsequently selected within the mammalian host.

ACKNOWLEDGMENTS

We gratefully acknowledge funding for this work provided by grants AI-29735 and AI-45538 from the Lyme disease program of the National Institute of Allergy and Infectious Diseases and by grants from the Arthritis Foundation (via the Egalitcheff Research Fund) and the Centers for Disease Control and Prevention (U50/CCU614875). M.J.C. was supported in part by Molecular Microbiology Training Grant AI-07520 (NIAID).

ADDENDUM IN PROOF

Recent work by Richard Marconi and coworkers has suggested that the *bdr* paralogs belong to three subfamilies, designated *bdrD*, *bdrE*, and *bdrF*. To clarify the relatedness of these genes both within and among *Borrelia* isolates, a new nomenclature has been suggested. To that end, we have redesignated the *bdr* paralogs in strain 297 as follows: *bdrA* to *bdrD1*, *bdrB* to *bdrE7* (cp18-2), *bdrD* to *bdrE3* (cp18-1), *bdrF* to *bdrE6* (cp32-3), *bdrH* to *bdrE1* (cp32-4), and *bdrJ* to *bdrE5* (cp32-5).

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. Investig.* **101**:2240–2250.
- Akins, D. R., M. J. Caimano, X. Yang, F. Cerna, M. V. Norgard, and J. D. Radolf. 1999. Molecular and evolutionary analysis of *Borrelia burgdorferi* 297 circular plasmid-encoded lipoproteins with OspE- and OspF-like leader peptides. *Infect. Immun.* **67**:1526–1532.
- 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.
- Baranton, G., D. Postic, I. Saint Girons, P. Boerlin, J. C. Piffaretti, M. Assous, and P. A. 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.
- Barbour, A. G., C. J. Carter, V. Bundoc, and J. Hinnebusch. 1996. The nucleotide sequence of a linear plasmid of *Borrelia burgdorferi* reveals similarities to those of circular plasmids of other prokaryotes. *J. Bacteriol.* **178**:6635–6639.
- Barbour, A. G., and C. F. Garon. 1987. Linear plasmids of the bacterium *Borrelia burgdorferi* have covalently closed ends. *Science* **237**:409–411.
- Barthold, S. W. 1996. Lyme borreliosis in the laboratory mouse. *J. Spirochet. Tick-Borne Dis.* **3**:22–44.
- 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.
- Carlyon, J. A., and R. T. Marconi. 1998. Cloning and molecular characterization of a multicopy, linear plasmid-carried, repeat motif-containing gene from *Borrelia turicatae*, a causative agent of relapsing fever. *J. Bacteriol.* **180**:4974–4981.
- 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.
- Cassatt, D. R., N. K. Patel, N. D. Ulbrandt, and M. S. Hanson. 1998. DbpA, but not OspA, is expressed by *Borrelia burgdorferi* during spirochetemia and is a target for protective antibodies. *Infect. Immun.* **66**:5379–5387.
- 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.
- Das, S., S. W. Barthold, S. S. Giles, R. R. Montgomery, S. R. Telford III, and E. Fikrig. 1997. Temporal pattern of *Borrelia burgdorferi* *p21* expression in ticks and mammalian host. *J. Clin. Investig.* **99**:987–995.
- de Silva, A. M., and E. Fikrig. 1995. Growth and migration of *Borrelia burgdorferi* in Ixodes ticks during blood feeding. *Am. J. Trop. Med. Hyg.* **53**:397–404.
- 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.
- El Hage, N., L. D. Lieto, and B. Stevenson. 1999. Stability of *erp* loci during *Borrelia burgdorferi* infection: recombination is not required for chronic infection of immunocompetent mice. *Infect. Immun.* **67**:3146–3150.
- Felsenstein, J. 1989. PHYLIP—Phylogeny Inference Package (version 3.2). *Cladistics* **5**:164–166.
- Ferdows, M. S., and A. G. Barbour. 1989. Megabase-sized linear DNA in the bacterium *Borrelia burgdorferi*, the Lyme disease agent. *Proc. Natl. Acad. Sci. USA* **86**:5969–5973.
- Fikrig, E., M. Chen, S. W. Barthold, J. Anguita, W. Feng, S. R. Telford III, and R. A. Flavell. 1999. *Borrelia burgdorferi* *erpT* expression in the arthropod vector and murine host. *Mol. Microbiol.* **31**:281–280.
- Fraser, C. M., S. Casjens, W. H. 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. Wattney, 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.
- Guina, T., and D. B. Oliver. 1997. Cloning and analysis of a *Borrelia burgdorferi* membrane-interactive protein exhibiting haemolytic activity. *Mol. Microbiol.* **24**:1201–1213.
- Helinski, D. R., A. E. Toukdarian, and R. P. Novick. 1996. Replication control and other stable maintenance mechanisms of plasmids, p. 2295–2324. *In* R. Curtiss III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella*: cellular and molecular biology, 2nd ed. ASM Press, Washington, D.C.
- Hofmann, K., and W. Stoffel. 1993. Tmbase, a database of membrane spanning protein segments. *Biol. Chem. Hoppe-Seyler* **347**:166–166.
- Kawabata, H., T. Masuzawa, and Y. Yanagihara. 1993. Genomic analysis of *Borrelia japonica* sp. nov. isolated from *Ixodes ovatus* in Japan. *Microbiol. Immunol.* **37**:843–848.
- Khan, S. A. 1997. Rolling-circle replication of bacterial plasmids. *Microbiol. Mol. Biol. Rev.* **61**:442–455.
- 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.
- Liveris, D., S. Varde, R. Iyer, S. Koenig, S. Bittker, D. Cooper, D. McKenna, J. Nowakowski, R. B. Nadelman, G. P. Wormser, and I. Schwartz. 1999. Genetic diversity of *Borrelia burgdorferi* in Lyme disease patients as determined by culture versus direct PCR with clinical specimens. *J. Clin. Microbiol.* **37**:565–569.
- Liveris, D., G. P. Wormser, J. Nowakowski, R. Nadelman, S. Bittker, D. Cooper, F. H. Moy, G. Forster, C. S. Pavia, and I. Schwartz. 1996. Molecular typing of *Borrelia burgdorferi* from Lyme disease patients by PCR-restriction fragment length polymorphism analysis. *J. Clin. Microbiol.* **34**:1306–1309.
- Marconi, R. T., and C. F. Garon. 1992. Identification of a third genomic group of *Borrelia burgdorferi* through signature nucleotide analysis and 16S rRNA sequence determination. *J. Gen. Microbiol.* **138**:533–536.
- Marconi, R. T., M. E. Konkel, and C. F. Garon. 1993. Variability of *osp* genes and gene products among species of Lyme disease spirochetes. *Infect. Immun.* **61**:2611–2617.
- Mathiesen, D. A., J. H. Oliver, Jr., C. P. Kolbert, E. D. Tullson, B. J. B. Johnson, G. L. Campbell, P. D. Mitchell, K. D. Reed, S. R. Telford III, J. F. Anderson, R. S. Lane, and D. H. Persing. 1997. Genetic heterogeneity of *Borrelia burgdorferi* in the United States. *J. Infect. Dis.* **175**:98–107.
- Nakai, K., and M. Kanehisa. 1991. Expert system for predicting protein localization sites in gram-negative bacteria. *Proteins* **11**:95–110.
- Norris, S. J., J. K. Howell, S. A. Garza, M. S. Ferdows, and A. G. Barbour. 1995. High- and low-infectivity phenotypes of clonal populations of in vitro-cultured *Borrelia burgdorferi*. *Infect. Immun.* **63**:2206–2212.
- Page, R. D. M. 1996. TREEVIEW: an application to display phylogenetic trees on personal computers. *Comput. Appl. Biosci.* **12**:357–358.
- Petit, M.-A., E. Dervyn, M. Rose, K.-D. Entian, S. McGovern, S. D. Ehrlich, and C. Bruand. 1998. PcrA is an essential DNA helicase of *Bacillus subtilis* fulfilling functions both in repair and rolling circle replication. *Mol. Microbiol.* **29**:261–273.
- Piesman, J., J. R. Oliver, and R. J. Sinsky. 1990. Growth kinetics of the Lyme disease spirochete (*Borrelia burgdorferi*) in vector ticks (*Ixodes dammini*). *Am. J. Trop. Med. Hyg.* **42**:352–357.
- 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.
- Ryan, J. R., J. F. Levine, C. S. Apperson, L. Lubke, R. A. Wirtz, P. A. Spears, and P. E. Orndorff. 1998. An experimental chain of infection reveals that distinct *Borrelia burgdorferi* populations are selected in arthropod and mammalian hosts. *Mol. Microbiol.* **30**:365–379.
- Saier, M. H., Jr. 1994. Computer-aided analyses of transport protein sequences: gleaming evidence concerning function, structure, biogenesis, and evolution. *Microbiol. Rev.* **58**:71–93.
- 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.
- 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.
- Seinost, G., D. E. Dykhuizen, R. J. Dattwyler, W. T. Golde, J. J. Dunn, I.-N. Wang, G. P. Wormser, M. E. Schriefer, and B. J. Luft. 1999. Four clones of

- Borrelia burgdorferi* sensu stricto cause invasive infection in humans. Infect. Immun. **67**:3518–3524.
45. 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.
 46. 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.
 47. 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.
 48. Stevenson, B., J. L. Bono, T. G. Schwan, and P. A. Rosa. 1998. *Borrelia burgdorferi* Erp proteins are immunogenic in mammals infected by tick bite, and their synthesis is inducible in cultured bacteria. Infect. Immun. **66**:2648–2654.
 49. Stevenson, B., S. Casjens, and P. Rosa. 1998. Evidence of past recombination events among the genes encoding Erp antigens of *Borrelia burgdorferi*. Microbiology **144**:1869–1879.
 50. 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.
 51. 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.
 52. 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.
 53. 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.
 54. 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.
 55. Theisen, M. 1996. Molecular cloning and characterization of *nlpH*, encoding a novel, surface-exposed, polymorphic, plasmid-encoded 33-kilodalton lipoprotein of *Borrelia afzelii*. J. Bacteriol. **178**:6435–6442.
 56. Thompson, J. D., D. G. Higgins, and T. J. Gibson. 1994. Clustal W: improving the sensitivity of progressive multiple sequence alignment through weighting positions-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22**:4673–4680.
 57. van Dam, A. P., H. Kuiper, K. Vos, A. Widjojokusumo, B. M. de Jongh, L. Spanjaard, A. C. P. Ramselaar, M. D. Kramer, and J. Dankert. 1993. Different genospecies of *Borrelia burgdorferi* are associated with distinct clinical manifestations of Lyme borreliosis. Clin. Infect. Dis. **17**:708–717.
 58. 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.
 59. Welsh, J., C. Pretzman, D. Postic, I. Saint Girons, G. Baranton, and M. McClelland. 1992. Genomic fingerprinting by arbitrarily primed polymerase chain reaction resolves *Borrelia burgdorferi* into three distinct phyletic groups. Int. J. Syst. Bacteriol. **42**:370–377.
 60. Wormser, G. P., D. Liveris, J. Nowakowski, R. B. Nadelman, L. F. Cavaliere, D. McKenna, D. Holmgren, and I. Schwartz. 1999. Association of specific subtypes of *Borrelia burgdorferi* with hematogenous dissemination in early Lyme disease. J. Infect. Dis. **180**:720–725.
 61. Xu, Y., C. Kodner, L. Coleman, and R. C. Johnson. 1996. Correlation of plasmids with infectivity of *Borrelia burgdorferi* sensu stricto type strain B31. Infect. Immun. **64**:3870–3876.
 62. Yang, X., T. G. Popova, K. E. Hagman, S. K. Wikel, G. G. Schoeler, M. J. Caimano, J. D. Radolf, and M. V. Norgard. 1999. Identification, characterization, and expression of three new members of the *Borrelia burgdorferi* Mlp (2.9) lipoprotein gene family. Infect. Immun. **67**:6008–6018.
 63. Zhang, J. R., J. M. Hardham, A. G. Barbour, and S. J. Norris. 1997. Antigenic variation in Lyme disease *Borreliae* by promiscuous recombination of *Vmp*-like sequence cassettes. Cell **89**:275–285.
 64. 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.
 65. Zückert, W. R., J. Meyer, and A. G. Barbour. 1999. Comparative analysis and immunological characterization of the *Borrelia* Bdr protein family. Infect. Immun. **67**:3257–3266.

Editor: A. D. O'Brien