Microbiology (1998), **144**, 1869–1879 Printed in Great Britain

Evidence of past recombination events among the genes encoding the Erp antigens of *Borrelia burgdorferi*

Brian Stevenson,[1]† Sherwood Casjens[2] and Patricia Rosa[1]

Author for correspondence: Brian Stevenson. Tel: +1 606 323 8967. Fax: +1 606 257 8994.
e-mail: lkspic00@pop.uky.edu

[1] Laboratory of Microbial Structure and Function, Rocky Mountain Laboratories, National Institute of Allergy and Infectious Diseases, NIH, Hamilton, MT 59840, USA

[2] Division of Molecular Biology and Genetics, Department of Oncological Sciences, University of Utah, Salt Lake City, UT 84132, USA

A single *Borrelia burgdorferi* bacterium may contain six or more different 32 kb circular plasmids (cp32s). Although these plasmids are homologous throughout much of their sequences, two loci have been identified at which they can vary significantly. The cp32 plasmids and their relatives each contain two adjacent genes, *orfC* and *orf3*, that vary in sequence between plasmids found within clones of individual bacteria. The *orfC* gene product is homologous to proteins involved in partitioning of bacterial plasmids, and the differences at this locus between plasmids may account for their compatibility. The *orfC–orf3* loci are located approximately 5 kb from another variable locus called *erp*. The *orfC–orf3* loci were used as physically linked markers to assess genetic rearrangements in the *erp* loci; this revealed examples of recombination involving both individual genes and entire *erp* loci. Recombination of the genes encoding the Erp antigens might contribute to the evasion of the mammalian immune response and could play roles in the establishment and persistence of *B. burgdorferi* infections in mammalian hosts.

Keywords: *Borrelia burgdorferi*, Lyme disease, *erp* genes, recombination, plasmids

INTRODUCTION

The Lyme disease spirochaete *Borrelia burgdorferi* contains an unusual genome composed of a linear chromosome (Baril *et al*., 1989; Casjens & Huang, 1993; Davidson *et al*., 1992; Ferdows & Barbour, 1989) and a large number of linear and circular plasmids (Barbour, 1988; Simpson *et al*., 1990a; Xu & Johnson, 1995). Among the DNA species found in these bacteria is a family of circular plasmids of approximately 32 kb in size (cp32-1 through cp32-7 in isolate B31) (Amouriaux *et al*., 1993; Porcella *et al*., 1996; Simpson *et al*., 1990b; Stevenson *et al*., 1996; Zückert *et al*., 1994; Zückert & Meyer, 1996). Restriction site mapping and partial sequencing of these plasmids has indicated that they contain similar nucleotide sequences and gene orders throughout much of their lengths (Casjens *et al*., 1997; Stevenson *et al*., 1997). Variants of cp32 plasmids have also been described, including an 18 kb truncated form (cp18) and a 56 kb linear plasmid that contains a large segment of DNA homologous to the cp32 plasmids (lp56) (Casjens *et al*., 1997; Porcella *et al*., 1996; Simpson *et al*., 1990b; Stevenson *et al*., 1997; Zückert & Meyer, 1996). Two 8–10 kb truncated and rearranged members of the cp32 family (cp8·3 and cp9) have also been described (Dunn *et al*., 1994; Fraser *et al*., 1997), although they lack many of the genes discussed in this work. Clonal cultures of *B. burgdorferi* containing as many as seven different members of the cp32 plasmid family have been isolated (Casjens *et al*., 1997), and it is possible that other clones of isolate B31 may contain fewer plasmids or additional, uncharacterized members of the cp32 family. Due to the extensive sequence similarities of the cp32 plasmids, they could not be confidently assembled by the *B. burgdorferi* B31 genome sequencing project of the Institute for Genomic Research (TIGR) (Fraser *et al*., 1997), and their complete sequences have not yet been published.

Many of the cp32 family of plasmids are compatible with each other and would therefore be expected to utilize different segregation proteins (Novick, 1987).

†**Present address:** Department of Microbiology and Immunology, MS 415 UKMC, University of Kentucky, Lexington, KY 40536, USA.

The GenBank accession numbers for the *orfC–orf3* loci of B31 plasmids cp32-2, cp32-3, cp32-4, cp32-6 and cp32-7 are AF022479, AF022480, AF022481, AF022482 and AF022483, respectively.

Two putative genes, *orfC* and *orf3*, were found to be adjacent to each other on two of the cp32s and the lp56 of *B. burgdorferi* isolate B31 (Casjens *et al.*, 1997; Zückert & Meyer, 1996) and on the cp18 of isolate N40 (Stevenson *et al.*, 1997), and these four *orfC–orf3* loci are heterogeneous in sequence (Stevenson *et al.*, 1997; Zückert & Meyer, 1996). The deduced Orf3 proteins are not homologous to any other previously characterized protein. It has been previously noted that the *orfC* genes are similar to proteins that are required for efficient partitioning of low-copy-number plasmids of other bacteria (Barbour *et al.*, 1996; Zückert & Meyer, 1996), such as the ParA and SopA proteins encoded by the *Escherichia coli* plasmids P1 and F, respectively (Abeles *et al.*, 1984; Ogura & Hiraga, 1983). Due to the predicted similarities of the OrfC proteins with these plasmid segregation proteins, we postulate that each cp32 family member within a single bacterium might carry different *orfC* genes and we have now characterized the *orfC–orf3* loci of all the known cp32 plasmids that can be carried by *B. burgdorferi* isolate B31.

The cp32 plasmids also encode members of a family of surface-exposed, membrane-bound lipoproteins we have called Erp proteins (Casjens *et al.*, 1997; Stevenson *et al.*, 1996). Shortly after mammals become infected with *B. burgdorferi*, antibodies are produced against a small number of bacterial proteins (Craft *et al.*, 1986; Engstrom *et al.*, 1995; Wilske *et al.*, 1986), including homologues of the Erp proteins (Akins *et al.*, 1995; Nguyen *et al.*, 1994; Stevenson *et al.*, 1995, 1998; Suk *et al.*, 1995; Wallich *et al.*, 1995). Individual spirochaetes may be capable of producing a large number of different Erp proteins (Casjens *et al.*, 1997; Stevenson *et al.*, 1996). Genes encoding Erp homologues from several *B. burgdorferi* isolates have been sequenced (Akins *et al.*, 1995; Lam *et al.*, 1994; Marconi *et al.*, 1996; Stevenson *et al.*, 1996, 1998; Suk *et al.*, 1995; Wallich *et al.*, 1995), and often show a variability of sequence suggestive of past recombination events among these genes. However, the multiplicity and similarity of the cp32 plasmids have made it impossible to determine whether the variation seen among members of the *erp* family is due to recombination events or reflects separately evolving gene sequences on a very large number of homologous plasmids. In this study we used the distinctive *orfC–orf3* loci of the cp32 plasmids as physically linked markers to analyse the relationships between *erp* loci and we found evidence of historical recombination events among the *erp* gene family.

METHODS

Bacterial strains and culture. *B. burgdorferi* were grown at 34 °C in liquid BSK-H medium (Sigma) supplemented with 6% rabbit serum (Sigma). Isolate B31 is the type strain of *B. burgdorferi* (ATCC 35210), originally isolated from an infected tick collected on Shelter Island, New York (Burgdorfer *et al.*, 1982). Two previously described clones of B31 were used in this work: B31-4a, cloned from a low-passage culture and infectious in laboratory mice, and B31-e1, cloned from a culture that has been maintained in the laboratory for several years and is no longer infectious in mice (Casjens *et al.*, 1997). Isolate N40 was originally cultured from an infected tick collected in Westchester County, New York (Barthold *et al.*, 1993).

DNA cloning and sequencing. Plasmid DNA from 100 ml cultures of *B. burgdorferi* or *E. coli*, grown in BSK-H or terrific broth (Maniatis *et al.*, 1982), respectively, were prepared using Qiagen Midi plasmid purification kits following the manufacturer's recommendations. A previously described recombinant *E. coli* plasmid containing an *Eco*RI fragment of the B31 cp32-3 (Casjens *et al.*, 1997) was used to determine the *orfC* and *orf3* sequences of cp32-3.

The *orfC–orf3* loci of the B31 cp32-2, cp32-4, cp32-6 and cp32-7 were amplified by PCR from purified B31 plasmids using an oligonucleotide complementary to the conserved *orf2* gene in combination with an oligonucleotide primer complementary to a specific DNA sequence within the *erp* locus of each plasmid (Table 1). DNA fragments of cp32-2 and cp32-4 were amplified from clone B31-e1, and fragments of cp32-6 and cp32-7 were amplified from clone B31-4a. PCR conditions in a DNA Thermal Cycler (Perkin-Elmer) consisted of 25 cycles of 94 °C for 1 min, 50 °C for 1 min and 65 °C for 6 min. PCR reaction products [both uncloned and cloned into pCR2.1 (Invitrogen)] were sequenced using a model 370A automated sequencer (Applied Biosystems). DNA and predicted protein alignments were executed using the program CLUSTAL V (Higgins *et al.*, 1992) and phylogenetic trees were constructed using the program MEGALIGN (DNAStar).

Southern blot analysis. DNA electrophoresis, blotting and hybridization with radiolabelled probes was carried out as described previously (Casjens & Huang, 1993; Casjens *et al.*, 1997). Using the oligonucleotides listed in Table 1, probes specific for the *orfC* genes of plasmids cp32-2, cp32-3, cp32-4, cp32-6 and cp32-7 were produced by PCR amplification from cloned DNA fragments that included each gene. Reaction conditions in a DNA Thermal Cycler (Perkin-Elmer) consisted of 25 cycles of 94 °C for 1 min, 50 °C for 30 s and 72 °C for 1 min. PCR reaction products were diluted 1:100 in distilled water and subjected to a second round of PCR amplification.

Accession numbers of previously described genes and proteins. The GenBank accession numbers of the B31 *erpAB2*, *erpCD*, *erpG*, *erpH*, *erpIJ*, *erpK*, *erpLM* and *erpX* loci are U78764, U44914, U42598, U44913, U72996, U72997, U72998 and AF020657, respectively (Casjens *et al.*, 1997; Stevenson *et al.*, 1996, 1998). The accession numbers of the *orfC–orf3* loci of the B31 plasmids cp32-1, cp32-5 and lp56 are X87127, X87202 and X87201, respectively (Casjens *et al.*, 1997; Zückert & Meyer, 1996). The accession numbers of *ospE*, *ospF* and the *orfC–orf3* locus of the N40 plasmid cp18 are L13924, L13925 and U42599, respectively (Lam *et al.*, 1994; Stevenson *et al.*, 1997). The accession number of the *bbk2.11* locus of *B. burgdorferi* isolate 297 is U30617 (Akins *et al.*, 1995). The accession numbers of the P1 ParA and F SopA proteins are X02954 and U26464, respectively.

RESULTS AND DISCUSSION

Characterization of the *orfC* and *orf3* genes from the B31 cp32 plasmids

We have previously identified eight plasmids in cultures of *B. burgdorferi* isolate B31 (seven different cp32s and lp56) that are largely homologous to one another (Casjens *et al.*, 1997). Since the previously sequenced

Table 1. Oligonucleotides used in this work

Use	Our designation	Sequence (5′ to 3′)
Amplification of *orf2* to *erp* locus		
cp32-2	B-1	CAACAAAGTTTTATTTAGTATG
	E-190	TTTTCTATTTTTGTTATTATCTCC
cp32-4	B-1	CAACAAAGTTTTATTTAGTATG
	E-404	ATTCATTCTTAGGGTTTTCATATC
cp32-6	B-3	TTTATGGGAAAAAATACCCGG
	E-504	CACTTTTAGGATCCGACTTCTC
cp32-7	B-3	TTTATGGGAAAAAATACCCGG
	E-704	CGCCTTGAGTTTTCCCAGTTGAAG
Construction of *orfC*-specific probes		
cp32-2 and cp32-7	E-311	GGAATGTATTAATTGATAATTCAG
	E-328	GCGAAATAAATAGTGCCTTATGGG
cp32-3	CP3-1	TTACGAAAAAATAGAAAAACTAGG
	CP3-2	TTTCCACTGCCCACTTTTCAGCCG
cp32-4	CP4-1	AGATCCTCAAAATAGTTTAACCAG
	CP4-2	TTAATATTGGCAGAGAGTCTACAG
cp32-6	CP6-1	GACTTTACATAGTATAAATGCTTTTGG
	CP6-2	TCTCGTTATTATAAAATAAGTAGG

Table 2. The known members of the cp32 plasmid family of isolates B31 and N40 and the designations of their *erp/ospEF* loci

Isolate	Plasmid	Gene(s)
B31	cp32-1	*erpA*, *erpB2*
	cp32-2	*erpC*, *erpD*
	cp32-3	*erpG*
	cp32-4	*erpH*
	cp32-5	*erpI*, *erpJ*
	cp32-6	*erpK*
	cp32-7	*erpL*, *erpM*
	lp56	*erpX*
N40	cp18	*ospE*, *ospF*

orfC–orf3 genes of cp32-1 and cp32-5 (Zückert & Meyer, 1996) are located approximately 5 kb 5′ of the *erp* loci of these two plasmids (Casjens *et al*., 1997), we reasoned that the remaining cp32 plasmids would probably contain *orfC* and *orf3* homologues in similar locations. Consistent with our hypothesis, we were able to identify *orfC–orf3* loci in DNA fragments that included each of the other *erp* loci (Table 2).

A previously cloned *Eco*RI fragment of cp32-3 isolated from a library of *B. burgdorferi* clone B31-e1 plasmid fragments encompassed the region predicted to contain the *orfC–orf3* locus (Casjens *et al*., 1997), which we confirmed by sequencing. The remaining *orfC–orf3* loci, from plasmids cp32-2, cp32-4, cp32-6 and cp32-7, were obtained by PCR amplification from B31 plasmid DNA. Another putative gene, *orf2*, is located immediately 5′ of the *orfC* gene on cp32-1, cp32-3, cp32-5 and lp56 (see Fig. 3) (this work; Zückert & Meyer, 1996). Unlike *orfC* and *orf3*, the previously known *orf2* genes have more than 95% identical nucleotide sequences (Zückert & Meyer, 1996). We used oligonucleotides based on conserved sequences of the *orf2* genes, in combination with oligonucleotides complementary to specific sequences located within each known B31 *erp* locus, to amplify the DNA spanning the region between the *orf2* and *erp* genes from cp32-2, -4, -6 and -7. These DNA fragments were partially sequenced, which revealed that each contained the appropriate *erp* locus on one end and an *orfC–orf3* locus on the other end.

We previously constructed detailed maps for each member of the B31 cp32 plasmid family that locate the cleavage sites of 20 different restriction endonucleases, and have found that each cp32 plasmid and lp56 has a unique cleavage pattern (S. Casjens, unpublished results; Casjens *et al*., 1997) To confirm that we had identified the correct *orfC–orf3* genes from each cp32 plasmid, we used probes specific for each *orfC–orf3* locus in Southern hybridization studies. In all cases, the restriction patterns obtained using the *orfC–orf3* probes were identical to those found previously using other probes specific for each plasmid. The same results were obtained when hybridizing to DNA purified from either the high-passage, non-infectious clone B31-e1, the low-passage, infectious clone B31-4a or B31-MI, the

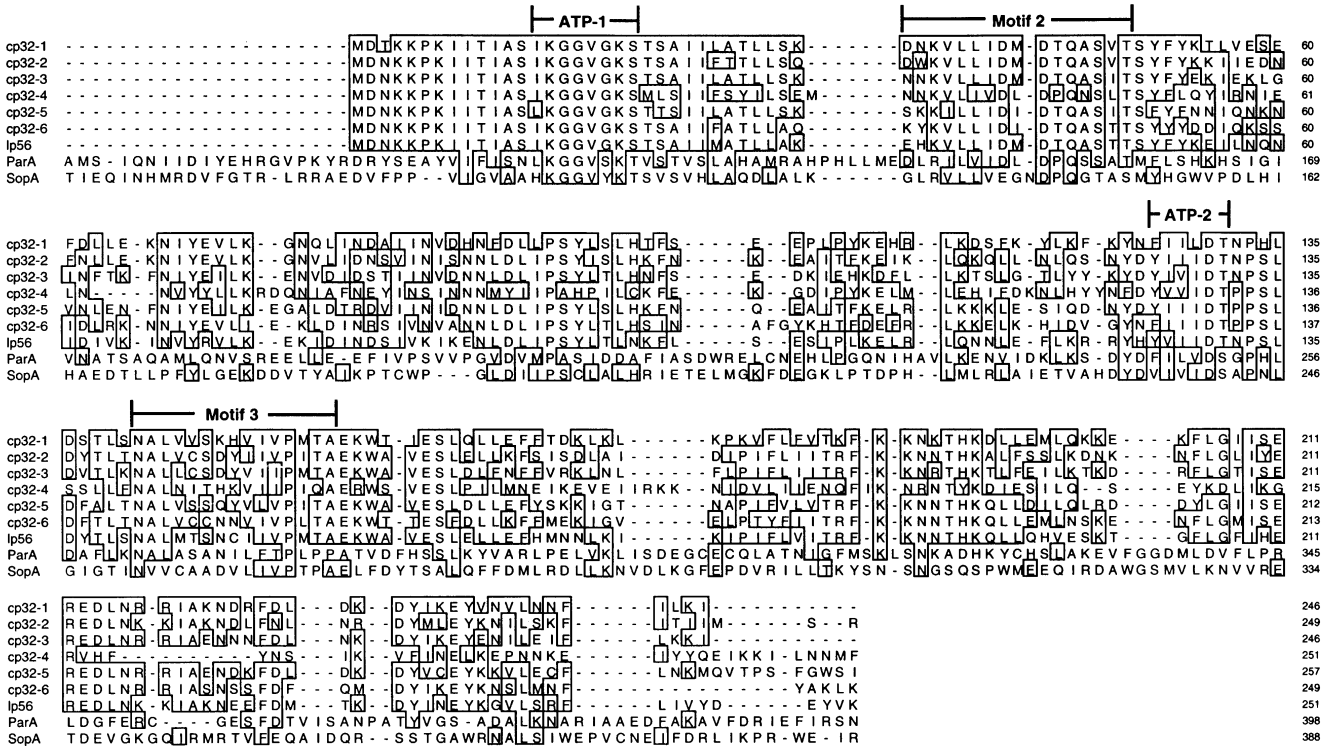

Fig. 1. Comparisons of the predicted amino acid sequences of OrfC proteins encoded on the B31 cp32 and lp56 plasmids and of the ParA and SopA proteins encoded on the *E. coli* plasmids P1 and F, respectively. The predicted OrfC protein sequence of cp32-7 is identical to that of cp32-2. The putative ATP-binding sites (ATP-1 and ATP-2) and Motifs 2 and 3 are as described by Motallebi-Veshareh *et al.* (1990). For brevity the first 82 amino acid residues of both ParA and SopA have not been included.

uncloned cuture that has been partially sequenced by TIGR (Fraser *et al*., 1997). We conclude that the PCR fragments we obtained accurately reflect the *orfC–orf3* and *erp* locus arrangements on the *B. burgdorferi* B31 cp32 plasmids.

Comparing the predicted OrfC proteins of plasmids cp32-1 through cp32-6 and of lp56 indicated that these proteins all contain regions of similarity with the ParA and SopA proteins of *E. coli* plasmids P1 and F (Fig. 1). The ParA and SopA proteins are essential for accurate segregation of these low-copy-number plasmids between daughter cells following bacterial cell divison (Abeles *et al*., 1984; Austin & Abeles, 1983a, b; Martin *et al*., 1991; Motallebi-Veshareh *et al*., 1990). All the predicted OrfC proteins contain sequences homologous to the conserved ATP-1 and ATP-2 motifs that are involved in ATP binding and hydrolysis by the plasmid segregation proteins (Motallebi-Veshareh *et al*., 1990; Watanabe *et al*., 1992). Homology is also found to the semi-variable motifs (Motifs 2 and 3, Fig. 1) that are postulated to be involved in binding accessory proteins or membrane attachment sites (Motallebi-Veshareh *et al*., 1990). The predicted OrfC proteins of the *B. burgdorferi* plasmids also exhibit a large degree of sequence variation (Fig. 1, Table 3) that is consistent with that observed among the segregation proteins of the P1, F and other, similar plasmids (Fig. 1) (Motallebi-Veshareh *et al*., 1990). This divergence presumably allows specific recognition of other plasmid segregation proteins (Motallebi-Veshareh *et al*., 1990). If the OrfC proteins function in plasmid partition, the variation among them may account for the compatibility of the numerous cp32s and related plasmids found in isolate B31.

In contrast to the heterogeneity seen among the other plasmids of B31, cp32-7 contains *orfC* and *orf3* genes that are nearly identical to those located on cp32-2, differing by only 4 bp over a stretch of 1311 bp (Table 3). These four differences do not alter the amino acid sequences, so the OrfC and Orf3 proteins of cp32-2 and cp32-7 are predicted to be indistinguishable (Table 3). If our hypothesis about the OrfC proteins being involved in plasmid partitioning is correct, then cp32-2 and cp32-7 would be incompatible with one another. Consistent with this idea, none of the clonal derivatives of B31 we have obtained contain both cp32-2 and cp32-7 (Casjens *et al*., 1997). It has been suggested that the original B31 isolate may have been a mixed population of closely related bacteria (Saint Girons & Davidson, 1992), and some of these may have contained cp32-2 while others contained cp32-7. Additional studies will be required to conclusively identify the compatibility determinants of the many *B. burgdorferi* plasmids.

Table 3. Similarities between *orfC–orf3* loci and predicted OrfC and Orf3 proteins of each cp32 and lp56 plasmid of *B. burgdorferi* B31 and cp18 of *B. burgdorferi* N40

The three values given for each comparison represent the percentage identity between the indicated pairs of sequences in the following order: *orfC–orf3* locus nucleotides/OrfC amino acids/Orf3 amino acids.

	B31							N40
	cp32-2	**cp32-3**	**cp32-4**	**cp32-5**	**cp32-6**	**cp32-7**	**lp56**	**cp18**
cp32-1	65/62/62	65/64/63	44/31/41	61/58/60	62/59/57	65/62/62	61/60/58	65/62/62
cp32-2		64/65/60	43/33/40	61/61/58	59/57/53	99·7/100/100	64/63/59	99·7/100/100
cp32-3			46/32/39	64/62/64	62/61/56	64/65/60	63/60/63	64/65/60
cp32-4				46/32/47	46/28/42	43/33/40	45/31/37	43/33/40
cp32-5					60/58/63	61/61/58	58/56/53	61/61/58
cp32-6						59/57/53	59/58/50	59/57/53
cp32-7							64/63/59	100/100/100
lp56								64/63/59

Table 4. Similarities between the B31 *erp* genes and their protein products and the related N40 *ospE* and *ospF* and 297 *bbk2.11* genes and proteins

Similarities are expressed as percentage identical nucleotides of the genes/identical amino acids of the proteins.

	B31								N40		297
	erpB2/J	***erpC***	***erpD***	***erpG***	***erpK***	***erpL***	***erpM***	***erpX***	***ospE***	***ospF***	***bbk2.11***
erpA/I	35/19	83/77	34/21	34/21	35/19	36/23	37/19	41/24	85/81	34/18	36/23
erpB2/J		34/20	71/62	49/32	40/26	38/22	71/58	56/38	34/18	39/23	38/21
erpC			35/20	34/19	36/20	38/22	37/18	36/21	88/77	33/19	37/23
erpD				49/33	42/25	37/24	67/58	53/35	32/20	38/25	38/23
erpG					51/36	50/34	48/33	45/28	32/20	48/30	52/33
erpK						55/39	41/28	41/25	34/18	58/43	64/49
erpL							39/17	40/24	35/21	76/67	83/80
erpM								61/47	35/19	38/22	40/22
erpX									38/22	40/17	41/25
ospE										34/17	35/23
ospF											76/63

Related *orfC–orf3* loci of another *B. burgdorferi* isolate

The *orfC* and *orf3* genes of cp32-7 are identical to those of cp18, an 18 kb truncated member of the cp32 family found in *B. burgdorferi* isolate N40 that contains the *ospEF* locus, a homologue of the B31 *erp* loci (Table 2) (Lam *et al*., 1994; Stevenson *et al*., 1997). Since the OrfC and Orf3 proteins of cp18 are predicted to be identical to those of cp32-2 and cp32-7, we conclude that at least this portion of these three plasmids shares a recent common ancestor. We have previously found that all *B. burgdorferi sensu stricto* isolates examined contained plasmids of the cp32 family (Casjens *et al*., 1997). It is unknown whether other bacteria besides N40 contain any such close relatives of the B31 cp32 plasmids. B31 was isolated on Shelter Island, New York (Burgdorfer *et al*., 1982) while N40 was isolated several years later in Westchester County on the New York mainland (Barthold *et al*., 1993). Analyses of the cp32 *orfC–orf3* loci of different isolates could serve as markers of how widespread these plasmids are among Lyme disease bacteria and whether they can be transmitted horizontally.

orfC–orf3 genes as markers of genetic drift versus recombination

Most of the B31 *erp* genes vary considerably in their sequences, as do other members of this gene family that have been identified from other isolates of *B. burgdorferi* (Akins *et al*., 1995; Casjens *et al*., 1997; Lam *et al*., 1994; Marconi *et al*., 1996; Stevenson *et al*., 1996, 1998; Suk *et al*., 1995; Wallich *et al*., 1995). While this high degree of variation may indicate recombination among these genes, it might also be argued that these differences are due to gradual genetic drift of plasmid sequences that

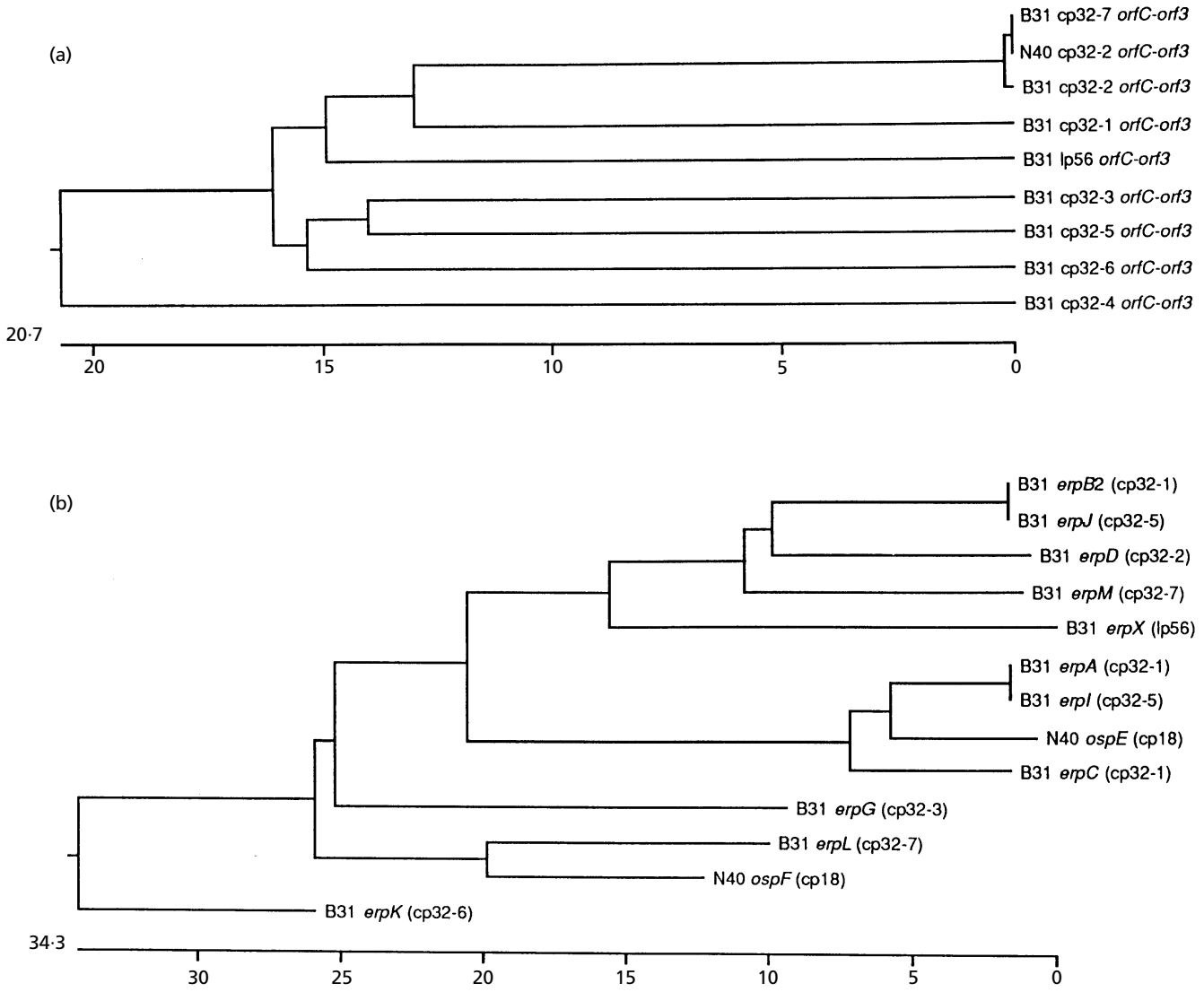

Fig. 2. Unbalanced phylogenetic tree obtained from CLUSTAL V (Higgins *et al.*, 1992) analysis of genes located on the cp32 and lp56 plasmids of B31 and the cp18 plasmid of N40. (a) *orfC–orf3* genes. (b) B31 *erp* and N40 *ospE* and *ospF* genes. The B31 *erpH* gene (located on cp32-4) was not included, since this gene is truncated and is significantly shorter than the other *erp* genes (Stevenson *et al.*, 1996). The scale represents the number of nucleotide substitutions expressed as a percentage of the number of compared nucleotides.

are no longer in genetic contact with one another. The physically linked *orfC–orf3* and *erp* loci provide a framework for determining whether the *erp* loci have diverged in isolation or if there have been recent recombination events among the cp32 plasmids.

As an example, plasmid cp32-3 contains *erpG*, and cp32-6 contains *erpK* (Table 2), which share 51% identical nucleotides, with the two deduced proteins predicted to contain 36% identical amino acids (Table 4) (Casjens *et al.*, 1997; Stevenson *et al.*, 1996). The physically linked *orfC–orf3* loci of cp32-3 and cp32-6 are also different between the two plasmids, sharing 62% identical nucleotides (Table 3). Dendrograms based on *orfC–orf3* and *erp* gene sequences also illustrate the distant relationships of these loci (Fig. 2). The data suggest that cp32-3 and cp32-6 are not closely related members of the cp32 plasmid family, and it is therefore impossible to determine whether variation between *erpG* and *erpK* is due to recombination or to random mutations that have accumulated since these two plasmids diverged from a presumed common ancestral cp32.

The *orfC–orf3* loci of plasmids cp32-1 and cp32-5 are also different from each other, again suggesting a distant relationship (Table 3, Figs 2a and 3a). Restriction site maps of these two plasmids indicate that they also contain unique sequences at other locations throughout

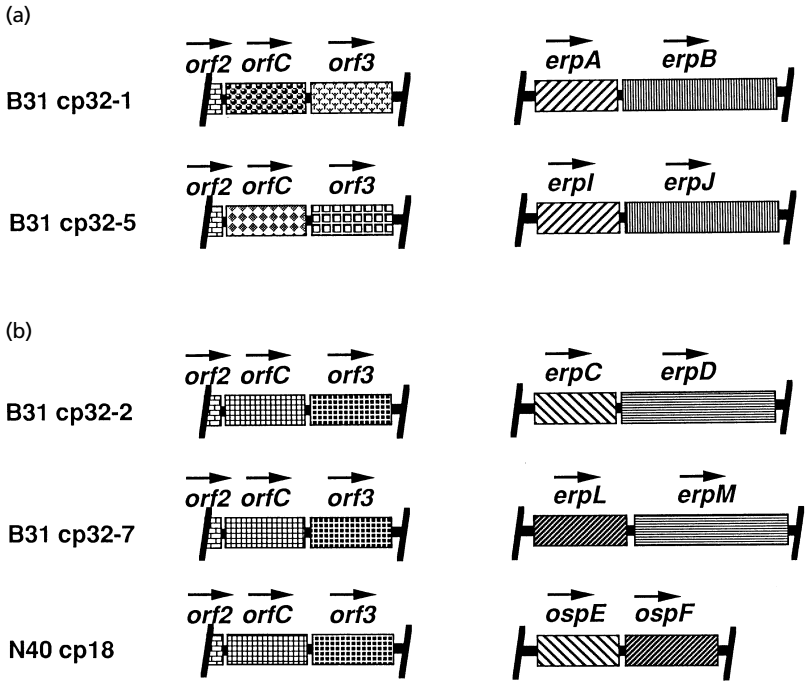

Fig. 3. Diagrams of conserved and non-conserved DNA sequences found on plasmids of the cp32 family, with similar fill patterns indicating similar gene sequences. Arrows indicate the direction of transcription of each gene. (a) B31 plasmids cp32-1 and cp32-5 contain identical *erp* loci but very different *orfC–orf3* loci. (b) B31 plasmids cp32-2 and cp32-7 and N40 plasmid cp18 contain nearly identical *orfC–orf3* loci but, in some cases, very different *erp* homologues.

the plasmids (Casjens *et al*., 1997). In contrast, the *erpAB*2 and *erpIJ* loci of cp32-1 and cp32-5, respectively, have identical nucleotide sequences and therefore must be extremely close relatives (Figs 2b and 3a) (Casjens *et al*., 1997; Stevenson *et al*., 1998). These observations strongly suggest that a recent recombination event gave rise to the current forms of these two plasmids, with a region of DNA that included *erpAB*2/*erpIJ* replacing the predecessor *erp* locus of one plasmid.

Recombination within *erp* loci on closely related plasmids

The near identity of the *orfC–orf3* loci found on the B31 cp32-2 and cp32-7 plasmids and the N40 cp18 plasmid indicates that at least these regions of the three plasmids are closely related, therefore permitting the comparison of the *erpCD*, *erpLM* and *ospEF* loci for evidence of past recombination events (Fig. 3b). Several genes in these three loci are similar to one another, as would be expected for genes on closely related plasmids. The *erpC* and *ospE* genes, each the first gene of bicistronic operons, share 88% identical nucleotides and 77% identical amino acids (Table 4, Figs 2b and 4a). Similarly, the *erpD* and *erpM* genes, the second genes of bicistronic operons, share 67% identical nucleotides and 58% identical amino acids (Table 4, Figs 2b and 4b). Comparisons of these gene and protein pairs indicate extended regions of similarity with interspersed regions of variation that may represent drift due to gradual accumulation of random mutations (Fig. 4a, b) (Dykhuizen *et al*., 1993). There are, however, suggestions of past insertion and/or deletion events within these pairs, since in each pair, short lengths of sequence are present in one member of the pair that are absent in the other (Fig. 4a, b).

In contrast, the *erpL* gene, the first gene of the *erpLM* bicistronic operon, is significantly different from either *erpC* or *ospE*, sharing only 38% and 35% identical nucleotides, respectively, and little amino acid identity (Table 4, Figs 2b and 4a). Likewise, *ospF*, the second gene of the *ospEF* operon, shares little sequence with either *erpD* or *erpM* (38% identity with both genes) (Table 4, Figs 2b and 4b). *erpL* and *ospF*, however, share 76% identical nucleotides and 67% identical amino acids (Table 4), with extended stretches of identity throughout their lengths (Fig. 4c). This similarity suggests that *erpL* and *ospF* may be descendants of a common ancestor that recombined into the first or second gene, respectively, of primordial *erpCD*-like loci. The first 452 bp of *erpL* are also identical to the corresponding region of *bbk2.11*, a monocistronic *erp* homologue of *B. burgdorferi* isolate 297 (Akins *et al*., 1995) (Table 4, Fig. 4c). The location of the *bbk2.11* locus in the 297 genome has not yet been determined (Akins *et al*., 1995), so its other relationships with the B31 and N40 plasmids and *erp* homologues cannot be analysed at this time.

Although recombination has apparently occurred in natural settings, we have not detected any cp32 rearrangements within laboratory cultures, since both *erp* gene and cp32 plasmid sequences were found to be virtually identical among cultures of B31 that have been separated for thousands of generations of cultivation (B. Stevenson, unpublished results; Casjens *et al*., 1997). The only known variation that has occurred in the B31 *erp* loci during laboratory cultivation is a single base in the *erpB* gene of a high-passage, non-infectious culture

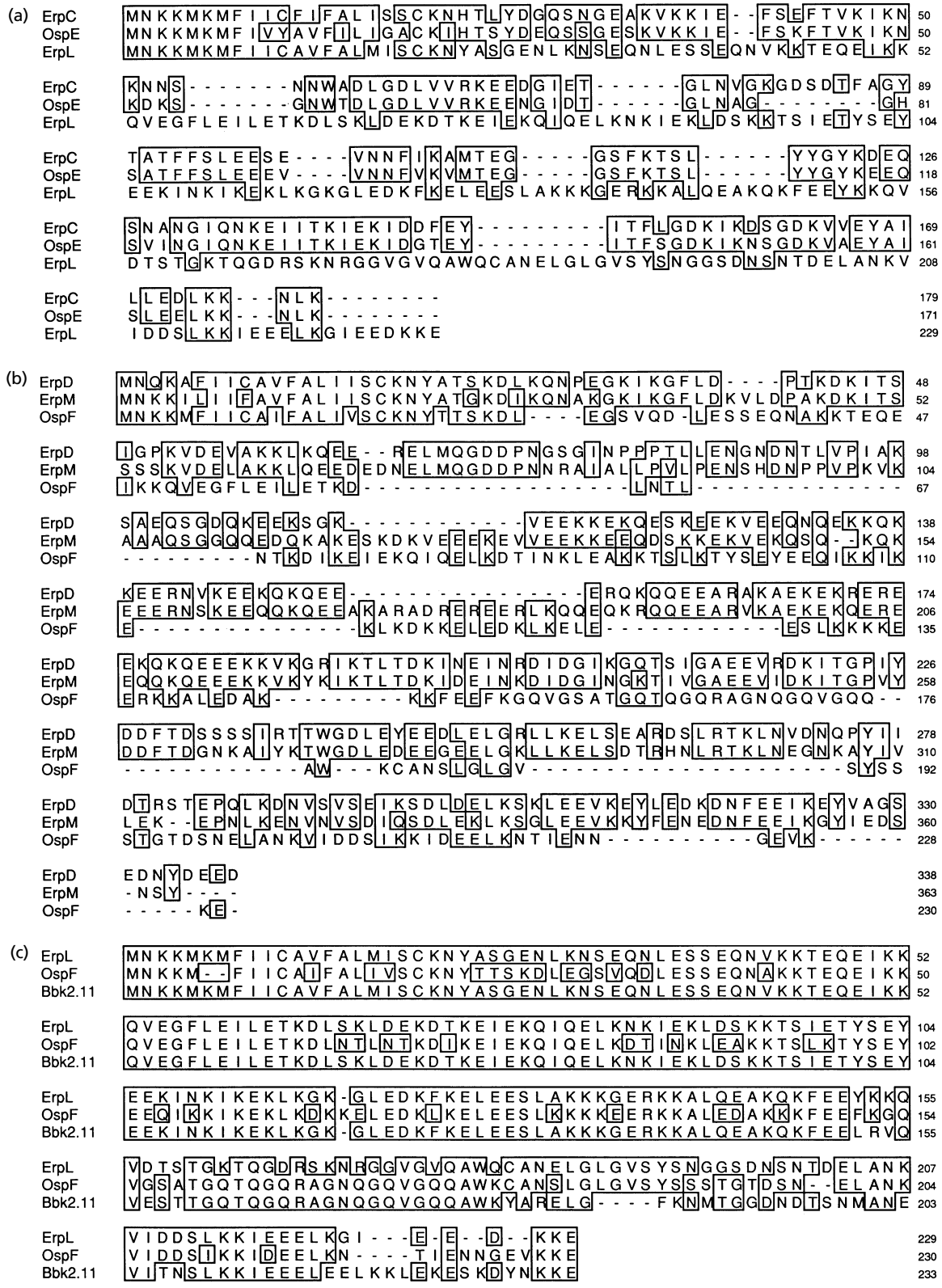

Fig. 4. Comparisons of the predicted amino acid sequences of Erp homologues found on plasmids with nearly identical *orfC–orf3* loci. (a) ErpC, OspE and ErpL, each encoded by the first gene of a bicistronic operon. (b) ErpD, ErpM and OspF, each encoded by the second gene of a bicistronic operon. (c) Comparison of ErpL, OspF and the Bbk2.11 protein of isolate 297. The sequence of the *orfC–orf3* locus that is presumably linked to *bbk2.11* is unknown.

that has mutated from that of both infectious cultures we have studied (B31-4a and B31-MI) (Casjens *et al*., 1997; Stevenson *et al*., 1996, 1998). The recombinational rearrangement of the *erp* genes that we have detected may have taken place within infected mammals, perhaps as part of a mechanism of evading clearance by the immune system. Antibodies directed against the Erp proteins are produced within the first few weeks of mammalian infection (Akins *et al*., 1995; Das *et al*., 1997; Nguyen *et al*., 1994; Stevenson *et al*., 1995, 1998; Wallich *et al*., 1995), indicating that they are expressed by *B. burgdorferi* in the early stages of infection. The gene encoding the antigenic *B. burgdorferi* VlsE lipoprotein appears to vary rapidly within infected mice by a homologous recombination mechanism (Zhang *et al*., 1997). In contrast, the *ospC* gene, which encodes another antigenic *B. burgdorferi* lipoprotein (Fuchs *et al*., 1992), also varies substantially between different isolates of *B. burgdorferi* and exhibits evidence of genetic exchange and recombination (Gibbs *et al*., 1996; Jauris-Heipke *et al*., 1995; Livey *et al*., 1995; Stevenson & Barthold, 1994; Theisen *et al*., 1993, 1995; Tilly *et al*., 1997), but the *ospC* gene is stable in chronically infected laboratory mice, indicating that variation of this gene is not necessary for long-term mammalian infection (Stevenson *et al*., 1994). Bacteria reisolated from long-term infected animals appear to be antigenically stable at several other loci as well (Barthold, 1993; Persing *et al*., 1994). Further experiments will determine whether the sequences of the *erp* genes also undergo recombination within mammals or are stably maintained during chronic mammalian infections.

Conclusions

We have presented evidence of past recombination events within the plasmids carrying *erp* genes, using the physically linked *orfC–orf3* locus as a marker of plasmid relatedness. Further analyses may reveal evidence of additional genetic rearrangements. Since a single bacterium may contain several different cp32 plasmids, each possibly carrying a unique *erp* locus (Casjens *et al*., 1997), we cannot speculate on whether these recombination events involved DNA already within a bacterium or if the DNA came from external sources. Homologues of the B31 *erp* genes have been identified in several other isolates of *B. burgdorferi* (Akins *et al*., 1995; Lam *et al*., 1994; Marconi *et al*., 1996; Suk *et al*., 1995; Wallich *et al*., 1995) and characterization of the *orfC–orf3* loci that are physically linked to these genes would enable a more detailed analysis of *erp* gene family evolution.

ACKNOWLEDGEMENTS

We thank Nanette Palmer for technical assistance, Daniel Hogan and Stephen Porcella for assistance with the DNA sequencing machinery and tree construction, Kit Tilly, James Bono, Abdallah Elias, Michael Chaussee, Joseph Hinnebusch and Stephen Porcella for constructive comments on the manuscript, and Kelly Matteson and Carole Smaus for secretarial assistance.

REFERENCES

Abeles, A. L., Snyder, K. M. & Chattoraj, D. K. (1984). P1 plasmid replication: replicon structure. *J Mol Biol* **173**, 307–324.

Akins, D. R., Porcella, S. F., Popova, T. G., Shevchenko, D., Baker, S. I., Li, M., Norgard, M. V. & Radolf, J. D. (1995). Evidence for in vivo but not in vitro expression of a *Borrelia burgdorferi* outer surface protein F (OspF) homologue. *Mol Microbiol* **18**, 507–520.

Amouriaux, P., Assous, M., Margarita, D., Baranton, G. & Saint Girons, I. (1993). Polymerase chain reaction with the 30-kb circular plasmid of *Borrelia burgdorferi* B31 as a target for detection of the Lyme borreliosis agents in cerebrospinal fluid. *Res Microbiol* **144**, 211–219.

Austin, S. & Abeles, A. (1983a). Partition of unit-copy miniplasmids to daughter cells. I. P1 and F miniplasmids contain discrete, interchangeable sequences sufficient to promote equipartition. *J Mol Biol* **169**, 353–372.

Austin, S. & Abeles, A. (1983b). Partition of unit-copy miniplasmids to daughter cells. II. The partition region of miniplasmid P1 encodes an essential protein and a centromere-like site at which it acts. *J Mol Biol* **169**, 373–387.

Barbour, A. G. (1988). Plasmid analysis of *Borrelia burgdorferi*, the Lyme disease agent. *J Clin Microbiol* **26**, 475–478.

Barbour, A. G., Carter, C. J., Bundoc, V. & Hinnebusch, J. (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.

Baril, C., Richaud, C., Baranton, G. & Saint Girons, I. (1989). Linear chromosome of *Borrelia burgdorferi*. *Res Microbiol* **140**, 507–516.

Barthold, S. W. (1993). Antigenic stability of *Borrelia burgdorferi* during chronic infections of immunocompetent mice. *Infect Immun* **61**, 4955–4961.

Barthold, S. W., de Souza, M. S., Janotka, J. L., Smith, A. L. & Persing, D. H. (1993). Chronic Lyme borreliosis in the laboratory mouse. *Am J Pathol* **143**, 959–972.

Burgdorfer, W., Barbour, A. G., Hayes, S. F., Benach, J. L., Grunwaldt, E. & Davis, J. P. (1982). Lyme disease – a tick-borne spirochetosis? *Science* **216**, 1317–1319.

Casjens, S. & Huang, W. M. (1993). Linear chromosomal physical and genetic map of *Borrelia burgdorferi*, the Lyme disease agent. *Mol Microbiol* **8**, 967–980.

Casjens, S., van Vugt, R., Tilly, K., Rosa, P. A. & Stevenson, B. (1997). Homology throughout the multiple 32-kilobase circular plasmids present in Lyme disease spirochetes. *J Bacteriol* **179**, 217–227.

Craft, J. E., Fischer, D. K., Shimamoto, G. T. & Steere, A. C. (1986). Antigens of *Borrelia burgdorferi* recognized during Lyme disease: appearance of a new immunoglobulin M response and expansion of the immunoglobulin G response late in the illness. *J Clin Invest* **78**, 934–939.

Das, S., Barthold, S. W., Stocker Giles, S., Montgomery, R. R., Telford, S. R. & Fikrig, E. (1997). Temporal pattern of *Borrelia burgdorferi p21* expression in ticks and the mammalian host. *J Clin Invest* **99**, 987–995.

Davidson, B. E., MacDougall, J. & Saint Girons, I. (1992). Physical map of the linear chromosome of the bacterium *Borrelia burgdorferi* 212, a causative agent of Lyme disease, and localization of rRNA genes. *J Bacteriol* **174**, 3766–3774.

Dunn, J. J., Buchstein, S. R., Butler, L.-L., Fisenne, S., Polin, D. S., Lade, B. N. & Luft, B. J. (1994). Complete nucleotide sequence of a circular plasmid from the Lyme disease spirochete, *Borrelia burgdorferi*. *J Bacteriol* **176**, 2706–2717.

Dykhuizen, D. E., Polin, D. S., Dunn, J., Wilske, B., Preac-Mursic, V., Dattwyler, R. J. & Luft, B. J. (1993). *Borrelia burgdorferi* is clonal: implications for taxonomy and vaccine development. *Proc Natl Acad Sci USA* **90**, 10163–10167.

Engstrom, S. M., Shoop, E. & Johnson, R. C. (1995). Immunoblot interpretation criteria for serodiagnosis of early Lyme disease. *J Clin Microbiol* **33**, 419–427.

Ferdows, M. S. & Barbour, A. G. (1989). Megabase-sized linear DNA in the bacterium *Borrelia burgdorferi*, the Lyme disease agent. *Proc Natl Acad Sci USA* **86**, 5969–5973.

Fraser, C. M., Casjens, S., Huang, W. M. & 35 other authors (1997). Genomic sequence of a Lyme disease spirochaete, *Borrelia burgdorferi*. *Nature* **390**, 580–586.

Fuchs, R., Jauris, S., Lottspeich, F., Preac-Mursic, V., Wilske, B. & Soutschek, E. (1992). Molecular analysis and expression of a *Borrelia burgdorferi* gene encoding a 22 kDa protein (pC) in *Escherichia coli*. *Mol Microbiol* **6**, 503–509.

Gibbs, C. P., Livey, I. & Dorner, F. (1996). The role of recombination in OspC variation in Lyme disease Borrelia. *Acta Dermatovenereol Alpina Pannonica Adriatica* **5**, 179–183.

Higgins, D. G., Bleasby, A. J. & Fuchs, R. (1992). CLUSTAL V: improved software for multiple sequence alignment. *Comput Appl Biosci* **8**, 189–191.

Jauris-Heipke, S., Liegl, G., Preac-Mursic, V., Robler, D., Schwab, E., Soutschek, E., Will, G. & Wilske, B. (1995). Molecular analysis of genes encoding outer surface protein C (OspC) of *Borrelia burgdorferi* sensu lato: relationship to *ospA* genotype and evidence of lateral gene exchange of *ospC*. *J Clin Microbiol* **33**, 1860–1866.

Lam, T. T., Nguyen, T.-P. K., Montgomery, R. R., Kantor, F. S., Fikrig, E. & Flavell, R. A. (1994). Outer surface proteins E and F of *Borrelia burgdorferi*, the agent of Lyme disease. *Infect Immun* **62**, 290–298.

Livey, I., Gibbs, C. P., Schuster, R. & Dorner, F. (1995). Evidence for lateral transfer and recombination in OspC variation in Lyme disease *Borrelia*. *Mol Microbiol* **18**, 257–269.

Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982). *Molecular Cloning: a Laboratory Manual*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Marconi, R. T., Sung, S. Y., Norton Hughes, C. A. & Carlyon, J. A. (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.

Martin, K. A., Davis, M. A. & Austin, S. (1991). Fine-structure analysis of the P1 plasmid partition site. *J Bacteriol* **173**, 3630–3634.

Motallebi-Veshareh, M., Rouch, D. A. & Thomas, C. M. (1990). A family of ATPases involved in active partitioning of diverse bacterial plasmids. *Mol Microbiol* **4**, 1455–1463.

Nguyen, T.-P. K., Lam, T. T., Barthold, S. W., Telford, S. R., III, Flavell, R. A. & Fikrig, E. (1994). Partial destruction of *Borrelia burgdorferi* within ticks that engorged on OspE- or OspF-immunized mice. *Infect Immun* **62**, 2079–2084.

Novick, R. P. (1987). Plasmid incompatibility. *Microbiol Rev* **51**, 381–395.

Ogura, T. & Hiraga, S. (1983). Partition mechanism of F plasmid: two plasmid gene-encoded products and a *cis*-acting region are involved in partition. *Cell* **32**, 351–360.

Persing, D. H., Mathiesen, D., Podzorski, D. & Barthold, S. W. (1994). Genetic stability of *Borrelia burgdorferi* recovered from chronically infected immunocompetent mice. *Infect Immun* **62**, 3521–3527.

Porcella, S. F., Popova, T. G., Akins, D. R., Li, M., Radolf, J. D. & Norgard, M. V. (1996). *Borrelia burgdorferi* supercoiled plasmids encode multi-copy tandem open reading frames and a lipoprotein gene family. *J Bacteriol* **178**, 3293–3307.

Saint Girons, I. & Davidson, B. E. (1992). Genome organization of *Borrelia burgdorferi*. In *Lyme Disease: Molecular and Immunologic Approaches*, pp. 111–118. Edited by J. Ingilis & J. W. Sitkowski. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Simpson, W. J., Garon, C. F. & Schwan, T. G. (1990a). Analysis of supercoiled circular plasmids in infectious and non-infectious *Borrelia burgdorferi*. *Microb Pathog* **8**, 109–118.

Simpson, W. J., Garon, C. F. & Schwan, T. G. (1990b). *Borrelia burgdorferi* contains repeated DNA sequences that are species specific and plasmid associated. *Infect Immun* **58**, 847–853.

Stevenson, B. & Barthold, S. W. (1994). Expression and sequence of outer surface protein C among North American isolates of *Borrelia burgdorferi*. *FEMS Microbiol Lett* **124**, 367–372.

Stevenson, B., Bockenstedt, L. K. & Barthold, S. W. (1994). Expression and gene sequence of outer surface protein C of *Borrelia burgdorferi* reisolated from chronically infected mice. *Infect Immun* **62**, 3568–3571.

Stevenson, B., Schwan, T. G. & Rosa, P. A. (1995). Temperature-related differential expression of antigens in the Lyme disease spirochete, *Borrelia burgdorferi*. *Infect Immun* **63**, 4535–4539.

Stevenson, B., Tilly, K. & Rosa, P. A. (1996). A family of genes located on four separate 32-kilobase circular plasmids in *Borrelia burgdorferi* B31. *J Bacteriol* **178**, 3508–3516.

Stevenson, B., Casjens, S., van Vugt, R., Porcella, S. F., Tilly, K., Bono, J. L. & Rosa, P. (1997). Characterization of cp18, a naturally truncated member of the cp32 family of *Borrelia burgdorferi* plasmids. *J Bacteriol* **179**, 4285–4291.

Stevenson, B., Bono, J. L., Schwan, T. G. & Rosa, P. (1998). The *Borrelia burgdorferi* Erp proteins are immunogenic in tick-bite-infected mammals and their synthesis is inducible in cultured bacteria. *Infect Immun* (in press).

Suk, K., Das, S., Sun, W., Jwang, B., Barthold, S. W., Flavell, R. A. & Fikrig, E. (1995). *Borrelia burgdorferi* genes selectively expressed in the infected host. *Proc Natl Acad Sci USA* **92**, 4269–4273.

Theisen, M., Frederiksen, B., Lebech, A.-M., Vuust, J. & Hansen, K. (1993). Polymorphism in *ospC* gene of *Borrelia burgdorferi* and immunoreactivity of OspC protein: implications for taxonomy and for use of OspC protein as a diagnostic antigen. *J Clin Microbiol* **31**, 2570–2576.

Theisen, M., Borre, M., Mathiesen, M. J., Mikkelsen, B., Lebech, A.-M. & Hansen, K. (1995). Evolution of the *Borrelia burgdorferi* outer surface protein OspC. *J Bacteriol* **177**, 3036–3044.

Tilly, K., Casjens, S., Stevenson, B., Bono, J. L., Samuels, D. S., Hogan, D. & Rosa, P. (1997). The *Borrelia burgdorferi* circular plasmid cp26: conservation of plasmid structure and targeted inactivation of the *ospC* gene. *Mol Microbiol* **25**, 361–373.

Wallich, R., Brenner, C., Kramer, M. D. & Simon, M. M. (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.

Watanabe, E., Wachi, M., Yamasaki, M. & Nagai, K. (1992). ATPase activity of SopA, a protein essential for active partitioning of F plasmid. *Mol Gen Genet* **234**, 346–352.

Wilske, B., Preac-Mursic, V., Schierz, G. & Busch, K. V. (1986). Immunochemical and immunological analysis of European *Borrelia burgdorferi* strains. *Zentralbl Bakteriol Hyg A* **263**, 92–102.

Xu, Y. & Johnson, R. C. (1995). Analysis and comparison of plasmid profiles of *Borrelia burgdorferi* sensu lato strains. *J Clin Microbiol* **33**, 2679–2685.

Zhang, J.-R., Hardham, J. M., Barbour, A. G. & Norris, S. J. (1997). Antigenic variation in Lyme disease borreliae by promiscuous recombination of VMP-like sequence cassettes. *Cell* **89**, 1–20.

Zückert, W. R. & Meyer, J. (1996). Circular and linear plasmids of Lyme disease spirochetes have extensive homology: characterization of a repeated DNA element. *J Bacteriol* **178**, 2287–2298.

Zückert, W. R., Filipuzzi-Jenny, E., Meister-Turner, J., Stålhammar-Carlemalm, M. & Meyer, J. (1994). Repeated DNA sequences on circular and linear plasmids of *Borrelia burgdorferi* sensu lato. In *Lyme Borreliosis*, pp. 253–260. Edited by J. S. Axford & D. H. E. Rees. New York: Plenum Press.

Received 19 November 1997; revised 20 February 1998; accepted 2 March 1998.