

# A bacterial genome in flux: the twelve linear and nine circular extrachromosomal DNAs in an infectious isolate of the Lyme disease spirochete *Borrelia burgdorferi*

Sherwood Casjens,<sup>1\*</sup> Nanette Palmer,<sup>1</sup>  
René van Vugt,<sup>1</sup> Wai Mun Huang,<sup>1</sup> Brian Stevenson,<sup>2</sup>  
Patricia Rosa,<sup>3</sup> Raju Lathigra,<sup>4</sup> Granger Sutton,<sup>5</sup>  
Jeremy Peterson,<sup>5</sup> Robert J. Dodson,<sup>5</sup> Daniel Haft,<sup>5</sup>  
Erin Hickey,<sup>5</sup> Michelle Gwinn,<sup>5</sup> Owen White<sup>5</sup> and  
Claire M. Fraser<sup>5</sup>

<sup>1</sup>Division of Molecular Biology and Genetics,  
Department of Oncological Sciences, University of Utah  
Medical School, Salt Lake City, UT 84132, USA.

<sup>2</sup>Department of Microbiology and Immunology, University  
of Kentucky College of Medicine, Lexington, KY 40536,  
USA.

<sup>3</sup>Laboratory of Human Bacterial Pathogenesis,  
Rocky Mountain Laboratory, NIAID, NIH, Hamilton, MT  
59840, USA.

<sup>4</sup>MedImmune Inc., 35 West Watkins Mill Road,  
Gaithersburg, MD 20878, USA.

<sup>5</sup>The Institute for Genomic Research, 9712 Medical  
Center Drive, Rockville, MD 20850, USA.

## Summary

**We have determined that *Borrelia burgdorferi* strain B31 MI carries 21 extrachromosomal DNA elements, the largest number known for any bacterium. Among these are 12 linear and nine circular plasmids, whose sequences total 610 694 bp. We report here the nucleotide sequence of three linear and seven circular plasmids (comprising 290 546 bp) in this infectious isolate. This completes the genome sequencing project for this organism; its genome size is 1 521 419 bp (plus about 2000 bp of undetermined telomeric sequences). Analysis of the sequence implies that there has been extensive and sometimes rather recent DNA rearrangement among a number of the linear plasmids. Many of these events appear to have been mediated by recombinational processes that formed duplications. These many regions of similarity are reflected in the fact that most plasmid genes are members of one of the genome's 161 paralogous gene families; 107 of these gene families, which vary in size from two to 41 members, contain at least one plasmid gene. These**

rearrangements appear to have contributed to a surprisingly large number of apparently non-functional pseudogenes, a very unusual feature for a prokaryotic genome. The presence of these damaged genes suggests that some of the plasmids may be in a period of rapid evolution. The sequence predicts 535 plasmid genes  $\geq 300$  bp in length that may be intact and 167 apparently mutationally damaged and/or unexpressed genes (pseudogenes). The large majority, over 90%, of genes on these plasmids have no convincing similarity to genes outside *Borrelia*, suggesting that they perform specialized functions.

## Introduction

Spirochetes of the genus *Borrelia* are unique among bacteria in that they have linear chromosomes and carry a large number of linear and circular plasmids. Their linear chromosomes range from 900 to 920 kbp in length (Baril *et al.*, 1989; Ferdows and Barbour, 1989; Davidson *et al.*, 1992; Casjens and Huang, 1993; Ojaimi *et al.*, 1994; Casjens *et al.*, 1995; Fraser *et al.*, 1997) [the known range of bacterial chromosome sizes is 580–9300 kbp (Casjens, 1998)]. Barbour and co-workers originally found that *Borrelia* isolates carry multiple linear extrachromosomal elements (Plasterk *et al.*, 1985; Barbour and Garon, 1987; Barbour, 1988), and all natural isolates that have been examined since then have non-identical, but similar, complements of such DNAs (Simpson *et al.*, 1990a,b; Stalhammar-Carlemalm *et al.*, 1990; Hughes *et al.*, 1992; Sadziene *et al.*, 1993a; Samuels *et al.*, 1993; Busch *et al.*, 1995; Casjens *et al.*, 1995, 1997a; Xu and Johnson, 1995; Marconi *et al.*, 1996a). In the few cases that have been examined, the individual *Borrelia* extrachromosomal DNAs are present in approximately the same numbers of molecules per cell as the chromosome (Hinnebusch and Barbour, 1992; Kitten and Barbour, 1992; Casjens and Huang, 1993), although a small circular plasmid of isolate lp90 appears to have a higher copy number (Dunn *et al.*, 1994). The linear DNAs have covalently closed hairpin telomeres (Barbour and Garon, 1987; Hinnebusch *et al.*, 1990; Hinnebusch and Barbour, 1991; Casjens *et al.*, 1997b; Fraser *et al.*, 1997). Most of the plasmids can be lost and are not required for propagation of the bacteria in culture, but loss of infectivity in mice often parallels plasmid

Received 19 May, 1999; revised 27 September, 1999; accepted 4 October, 1999. \*For correspondence. E-mail sherwood.casjens@hci.utah.edu; Tel. (+1) 801 581 5980; Fax (+1) 801 581 3607.

loss (for example Barbour, 1988; Hyde and Johnson, 1988; Schwan *et al.*, 1988; Norris *et al.*, 1992; Sadziene *et al.*, 1993b; Persing *et al.*, 1994; Xu *et al.*, 1996; Zhang *et al.*, 1997; but see Casjens *et al.*, 1997a). For brevity, we will refer to the *Borrelia* extrachromosomal DNA elements as 'plasmids', even though some of them may be universally present and are probably essential in nature (see Marconi *et al.*, 1996a; Casjens *et al.*, 1997b; Tilly *et al.*, 1997), carry genes that may be metabolically important (Margolis *et al.*, 1994) or have never been lost in culture. Perhaps some would more correctly be referred to as 'mini-chromosomes' (Barbour, 1993).

*Borreliae* were found to be the aetiological agent of Lyme disease in the USA in 1982 (Burgdorfer *et al.*, 1982; Steere *et al.*, 1983). Lyme disease is currently the most prevalent tick-borne disease in the USA (Walker, 1998) and is known to be caused by at least three different named bacterial species, *Borrelia burgdorferi*, *Borrelia garinii* and *Borrelia afzelii*, in North America and Europe. These are members of a cluster of very closely related species that also currently includes *Borrelia andersonii*, *Borrelia japonica*, *Borrelia valaisiana*, *Borrelia lusitaniae*, *Borrelia turdae*, *Borrelia tanukii*, *Borrelia bissettii* sp. nov. and several other as yet unnamed types (see, for example, Casjens *et al.*, 1995; Fukunaga *et al.*, 1996; Le Fleche *et al.*, 1997; Wang *et al.*, 1997a; 1998; Postic *et al.*, 1998). Together, this cluster of bacteria is referred to as the Lyme agent group or *Borrelia burgdorferi* (*sensu lato*).

The *B. burgdorferi* isolate characterized in this report, strain B31 culture MI (Casjens *et al.*, 1997a; Fraser *et al.*, 1997), was isolated from an *Ixodes scapularis* tick on Shelter Island, NY, in 1982 (Burgdorfer *et al.*, 1982; Johnson *et al.*, 1984). In our first report on the project to sequence completely the *B. burgdorferi* genome (Fraser *et al.*, 1997), we showed that the random DNA clone sequencing strategy gave contiguous sequence blocks that unambiguously assembled into the large chromosome, nine linear plasmids and two circular plasmids. At that time there were approximately 300 kbp of sequence data that could not be assembled unambiguously. We have since refined the TIGR ASSEMBLER software and now report the nucleotide sequences of seven additional circular and three additional linear plasmids, which completes the sequence of the genome of *B. burgdorferi* strain B31.

## Results and discussion

### *Sequence determination of 10 additional B. burgdorferi B31 plasmids*

**Sequence assembly.** In the *B. burgdorferi* isolate B31 MI genome sequencing project described by Fraser *et al.* (1997), the initial assembly of the whole-genome random nucleotide sequence data resulted in contiguous blocks

(contigs) of nucleotide sequence that correspond to the chromosome, two circular plasmids and nine linear plasmids. The remaining sequence data assembled ambiguously. In order to determine the nucleotide sequence of the remainder of the genome, the TIGR SEQUENCE ASSEMBLER computer program was modified (see *Experimental procedures*). After this modification, the previously unassembled raw sequence assembled uniquely into an additional seven circular and three linear contigs, corresponding to the following plasmids: cp32-1, cp32-3, cp32-4, cp32-6, cp32-7, cp32-8, cp32-9, lp5, lp21 and lp56 (named 'cp' for circular and 'lp' for linear plasmids and according to their approximate size in kbp. Previously utilized names were not changed when the actual length did not correspond precisely to those numbers). Plasmids lp5, lp21, cp32-8 and cp32-9 did not have previously used names, although each had been previously observed: lp5 (B. Stevenson, unpublished); lp21 (Barbour, 1988; P. Rosa and S. Casjens, unpublished); cp32-8 and cp32-9 (Casjens *et al.*, 1997a). Casjens *et al.* (1997a) have described two additional circular plasmids, cp32-2 and cp32-5, in other cultures of isolate B31 that are not present in B31 culture MI. These 10 new plasmid DNA sequences, along with those previously published in Fraser *et al.* (1997), account for all of the random sequence generated by this genome sequencing project.

Because of the difficulties encountered in the sequence assembly process, it was necessary to confirm the accuracy of the assembly of the plasmid sequences. Restriction maps of six plasmids from strain B31 MI have been described in Casjens *et al.* (1997a) and Tilly *et al.* (1997), and, in this study, we determined the restriction maps of 13 of the remaining 15 plasmids (344 total sites mapped and correctly predicted on 19 plasmids; N. Palmer, R. van Vugt and S. Casjens, unpublished). Only the cp9 and lp17 assemblies were not confirmed in this way because: (i) they assembled unambiguously, even with the original less stringent TIGR ASSEMBLER; (ii) Barbour *et al.* (1996) previously reported the complete sequence of B31 lp17; and (iii) Dunn *et al.* (1994) previously reported the very similar sequence of a cp9-like plasmid from a related isolate. Assembly of the sequences of the cp32s and the closely related portion of lp56 were particularly difficult. Nonetheless, they are likely to be correct because all of their restriction maps are predicted perfectly by the nucleotide sequences, which were assembled without knowledge of the restriction maps, and all of the 19 blocks of sequence that had been previously mapped to individual cp32s (Zuckert and Meyer, 1996; Casjens *et al.*, 1997a; Stevenson *et al.*, 1998a) are present in the correct cp32 at the experimentally determined location. [We note that the pOMB25 sequence that was attributed without mapping data to cp32-1 (Zuckert and Meyer, 1996) is actually in the cp32-3 sequence.] Assembly of the lp21 sequence had a special problem in that it contains a long tract of

176 (plus one partial) tandem copies of a 63 bp sequence (11 004 bp total). There are no unique, unrelated sequences interspersed among the 63 bp repeats, but not all of the repeats are identical (as is indicated experimentally by the small number of *Tsp509I* sites within the tract, see below). This non-identity made assembly from random sequencing runs possible, and experimental determination of the repeat tract length confirmed the predicted tract length (see below). Thus, the sequences of all 21 of the B31 plasmids are strongly supported either by physical maps that are correctly predicted by the sequence or by independent sequence determinations.

**Sequence accuracy and changes during growth in culture.** In general, our sequence agrees with all previously published nucleotide sequences from the strain B31 plasmids. We will discuss only a few long sections that have been previously sequenced. Our 16 823 bp sequence of lp17 has 26 nucleotide differences (at 23 locations, all unambiguous with multiple runs in each direction in our data) from the previously published complete sequence of this plasmid by Barbour *et al.* (1996). Thirteen of these are frameshift differences, one of which lengthens orfH (our BBD11) of Barbour *et al.* (1996). The reported lp28-1 8574 bp sequence of the silent *vsE* cassette region (BBF32) (Zhang *et al.*, 1997) has one difference from that reported here in the leftmost cassette and 14 differences in the ~300 bp of known sequence between the cassettes and the *vsE* expression site (13 in one 35 bp region!). An unknown mechanism rapidly moves sequences to the *vsE* expression site from the silent cassettes when the bacteria are in a mouse, and it is of interest to note that our B31 culture was passed through a mouse independently from that of Zhang *et al.* (1997) so that the extreme similarity of the cassette regions in the two sequences indicates that this movement is essentially unidirectional in that it does not rapidly exchange sequences among the silent 'genes' or from the expression site to the cassettes (see also Zhang and Norris, 1998). We reanalysed our previously reported 16 810 bp of sequence from cp32-1, cp32-3 and cp32-4 from high-passage B31 [clones e-1, e-2 and their parent high-passage culture (Stevenson *et al.*, 1996; Casjens *et al.*, 1997a)] and found 11 substantiated differences from the B31 MI sequence reported here. In each of these 11 instances, as well as in the lp28-1 sequence (J. Zhang and S. Norris, personal communication), the data are unambiguous; there are multiple sequencing runs in agreement from each source. Thus, the cp32 differences between the B31 high-passage and MI (low passage) cultures appear to be mutational changes that have accumulated during long-term growth of several thousand generations in culture [most are missense changes, but frameshifts truncate genes BBP38 (*erpB*) and BBR38 in

the high-passage culture]. Curiously, six of these differences are in one 31 bp region in gene BBP36 of cp32-1. This and the group of differences in lp28-1 suggest that such changes can be made in clusters; in the cp32-1 cluster, most of the changes that occurred in BBP36 during propagation do not appear to be simply derived by recombination from paralogous sequences because none of the seven BBP36 paralogues (see below) in B31 MI contains all these changes.

**The 'complete' *B. burgdorferi* genome nucleotide sequence.**

Sequence remains unknown for a few nucleotides at the tips of the linear plasmid telomeres because the DNA library used for sequencing did not contain cloned terminal fragments (Fraser *et al.*, 1997). Each of the six B31 telomere sequences that have previously been reported uniquely overlap one terminus among our library-generated linear plasmid and chromosome sequences; these terminal sequences show that the following numbers of bp are missing from the cognate ends of our random library-derived sequences: lp17 left end, 29 bp; lp17 right end, 78 bp; lp28-1 right end, ~1300 bp; lp56 right end, 25 bp [this sequence, called TL49, was reported to be an lp54 telomeric sequence at a time when the existence of lp56 was not known (Hinnebusch *et al.*, 1990)]; chromosome left end, 106 bp; chromosome right end, 72 bp (Fraser *et al.*, 1997). As between 25 and 106 bp were missing from five of these telomeres, we suspect that, unless an unclonable region is positioned within 1–2 kbp of a telomere, on average less than 100 terminal bp are likely to be missing from the sequences determined in this project. At one telomere, the right end of lp28-1, a short unclonable region apparently kept the terminal 1300 bp from being present in our library (Zhang *et al.*, 1997; J. Zhang and S. Norris, personal communication). Our measurements of whole plasmid sizes and terminal restriction fragment sizes supports the idea that unsequenced regions at most plasmid telomeres are <1 kbp; in the case where we analysed terminal fragment lengths most accurately, both terminal fragments of lp5 extend ≤150 bp beyond the ends of the nucleotide sequence (data not shown).

We conclude that at the 20 unsequenced telomeres a total of 2000 bp or less of telomeric sequences and few, if any, protein coding regions are likely to be missing from the sequence of the *B. burgdorferi* B31 genome. The complete genome thus includes the 910 725 bp chromosome, 249 330 bp in nine circular plasmids and 361 364 bp in 12 linear plasmids for a total genome size of 1521 419 bp (plus ≤2000 bp of unsequenced linear plasmid termini).

*Twenty-two replicons in one bacterium?*

Although the B31 MI culture whose genome was sequenced had not been grown from a single bacterium, there is no

evidence for macrorestriction fragment length heterogeneity in its genome (R. van Vugt and S. Casjens, unpublished), and we have found that nearly all of the 21 plasmids found in B31 MI can coexist in an individual bacterium. First, we found that 25 clones derived from B31 MI had linear plasmid patterns in CHEF (contour-clamped homogeneous electric field) electrophoresis gels that were indistinguishable from the uncloned parent whose DNA was sequenced (data not shown). In addition, we examined the parallel clonal culture B31 4a in detail using DNA probes specific for each plasmid in Southern analyses. After the isolation of clone 4a from a solid agar colony, passage through a BALB/c mouse for 4 weeks and re-isolation from the mouse (see Casjens *et al.*, 1997a), it carried all of the plasmids whose sequences are known except for cp9, lp5, lp28-3 and lp28-4 (data not shown). The plasmids missing in clone 4a may well have been lost during the cloning/mouse passage procedure because *Borrelia* strains have often been found to lose one or more plasmids during laboratory propagation and cloning procedures (for example Barbour, 1988; Schwan *et al.*, 1988; Persing *et al.*, 1994; Norris *et al.*, 1995; Xu *et al.*, 1996).

In addition to the chromosome and 21 plasmids in B31 MI, two additional cp32 relatives, cp32-2 and cp32-5, have been reported to be present in other subcultures of the original *B. burgdorferi* B31 isolate (cp32-5 is present in clone 4a above) (Stevenson *et al.*, 1996; Zuckert and Meyer, 1996; Casjens *et al.*, 1997a). Because B31 MI is infectious in mice, cp32-2 and cp32-5 must not be required for this process. It is not known whether cp32-2 and cp32-5 are absent from culture MI because they were lost during propagation of an originally clonal isolate or whether the original isolate was a mixture of closely related bacteria carrying slightly different plasmid complements (Casjens *et al.*, 1997a; Stevenson *et al.*, 1998a).

This analysis proves that at least 17 of B31 MI's 21 plasmids are present in the only clonal B31 subculture that has been completely analysed, and it is probable that as many as 23 plasmids existed in the original B31 isolate. Clearly, the existence of so many replicons in one bacterium raises issues concerning replication specificity, compatibility and segregation that remain to be addressed.

#### *Features of the B. burgdorferi* plasmid nucleotide sequences

**Nucleotide distribution.** The overall G+C contents of the B31 plasmids vary from 20.7% to 31.6% (cf. 28.6% in the long chromosome; Table 1). Plots of G+C content by position show a few notable features: (i) as has been previously noted by Zhang *et al.* (1997), the *vlsE* gene and its related pseudogene cassettes (BBF32) have a G+C content of about 50%, which is strikingly higher than

the remainder of the plasmid where the local G+C content is mostly between 25% and 20%; (ii) the middle 15 kbp of lp28-2 has a relatively high G+C content of about 35%; (iii) the very low G+C content of lp21 is as a result largely of the ~18.5% G+C content of the long 63 bp repeat tract; (iv) the G+C content of lp17 is very low at 23%; and (v) in lp56, the cp32-like sequence (see below) is about 29% G+C, whereas the remainder is mostly between 21% and 25% G+C. These variations from uniformity could be indications of recent arrival of the lp28-1 and lp28-2 higher G+C regions by horizontal transfer (Lawrence and Ochman, 1997). In addition, it may be that the very low values for the parts of lp17, lp28-1 and lp56 mentioned above are so low because they no longer encode functional proteins and are largely in a state of mutational decay (see below). It has been proposed that genomes have different G+C contents because of inherent species-specific directionality of mutation and/or repair systems (Sueoka, 1993), and one might imagine that *Borrelia*, whose chromosome is 28.6% G+C, is approaching its lower 'limit' in that most new changes towards even lower G+C values would be selected against. However, when selection for function is lifted in a particular region (indicated here by the presence of pseudogenes), G+C content there may continue to drift to even lower values. GC skew  $[(G - C)/(G + C)]$  analysis of the plasmids (data not shown) shows that a number of the plasmids, especially lp54, lp28-2 and the cp32s, show a significant skew sign change adjacent to the 'partition gene cluster' (see below), providing a weak indication of possible divergent replication and hence an origin in those regions (McLean *et al.*, 1998 and references therein). However, gene orientation may contribute significantly to GC skew on these DNAs.

**Direct tandem repeats.** Tracts of short, tandemly repeated sequences are not abundant or well understood in bacteria. However, in known cases, they often occur in association with 'contingency genes' because the hypermutability of such sequences, due to changes in the number of repeat units during slipped-strand replication and/or recombination, can lead to switching between on and off expression states (phase variation) of the associated genes at either a transcriptional or translational level (Moxon *et al.*, 1994; Saunders *et al.*, 1998).

By far the most extensive short sequence repeat in the *B. burgdorferi* B31 genome is the 11 kbp tract of 63 bp repeats in lp21. Each repeat has stop codons in all six frames. There are about one and a half copies of this repeat between 1630 and 1780 bp on lp28-3, where gene BBH05 terminates within the repeat, and less well-conserved partial copies about 200 bp from the right ends of lp28-4 and lp36 where they do not overlap predicted open reading frames. No other matches to the 63 bp unit were found in the current sequence data base. Its function is unknown,

**Table 1.** The 22 *B. burgdorferi* B31 replicons.

Replicon	Geometry	Size (bp)	G+C (%)	R sites <sup>a</sup>	Coding <sup>b</sup> (%)	Genes <sup>c</sup> > 300 bp	Genes <sup>c</sup> ≤ 300 bp	Pseudogenes <sup>d</sup>	Fraction pseudogenes <sup>e</sup>	Fraction pseudogenes <sup>f</sup>	LP <sup>g</sup>
Chromosome <sup>h</sup>	Linear	910 725	28.6	32	93 (93)	769	73	1	0.001	0.001	35/38/16
cp9	Circular	9386	23.9	ND <sup>i</sup>	75 (75)	9	2	0	0.00	0.00	1/1/0
cp26	Circular	26 498	26.5	16	88 (88)	26	3	0	0.00	0.00	7/7/1
cp32-1	Circular	30 750	29.4	35	92 (92)	40	2	0	0.00	0.00	4/4/1
cp32-3	Circular	30 223	28.9	30	92 (92)	40	4	0	0.00	0.10	2/2/1
cp32-4	Circular	30 299	29.3	29	92 (86)	37	2	4	0.09	0.10	2(1)/2(1)/1
cp32-6	Circular	29 838	29.3	21	92 (92)	39	2	0	0.00	0.00	3/3/1
cp32-7	Circular	30 800	29.1	19	93 (93)	40	2	0	0.00	0.00	3/3/1
cp32-8	Circular	30 885	29.1	12	92 (92)	40	3	0	0.00	0.00	2/3/1
cp32-9	Circular	30 651	29.3	9	92 (71)	31	2	9	0.21	0.23	3/3/1
lp5 <sup>j</sup>	Linear	5228	23.8	7	73 (47)	3	0	4	0.57	0.57	0/0/0
lp17 <sup>i</sup>	Linear	16 928 <sup>k</sup>	23.1	ND	64 (45)	9	10	9	0.32	0.50	2/3/0
lp21 <sup>i</sup>	Linear	18 901	20.7	13	32 (21)	5	1	6	0.50	0.55	0/0/0
lp25 <sup>i</sup>	Linear	24 177	23.4	8	66 (47)	10	15	13	0.34	0.57	4/4/0
lp28-1 <sup>i</sup>	Linear	28 250 <sup>k</sup>	32.3	12	79 (25)	10	3	36 <sup>m</sup>	0.73	0.78	1(2)/1(2)/1
lp28-2	Linear	29 766	31.6	20	92 (85)	29	2	3	0.09	0.09	2/3/0
lp28-3 <sup>i</sup>	Linear	28 601	25.0	8	66 (46)	12	11	17	0.41	0.57	3/4/1
lp28-4 <sup>i</sup>	Linear	27 323	24.5	12	62 (51)	16	13	12	0.29	0.43	6(1)/8(1)/0
lp36 <sup>j</sup>	Linear	36 849	26.9	17	76 (54)	21	18	11	0.22	0.34	6(1)/9(1)/2
lp38 <sup>j</sup>	Linear	38 829	26.1	20	67 (49)	21	11	15	0.32	0.42	3(1)/4(1)/1
lp54	Linear	53 541	28.2	23	82 (81)	53	21	2	0.03	0.04	19/21/2
lp56 (cp32) <sup>j</sup>	Linear	30 349	29.0	14	93 (85)	36	2	4	0.10	0.10	2/2/2
lp56 (other) <sup>ij</sup>	Linear	22 622 <sup>k</sup>	25.0	9	78 (22)	8	6	22	0.61	0.76	3(1)/3(1)/0
Pseudogene plasmid <sup>i</sup> total		247 708	25.6	106	67 (42)	116	88	145	0.42	0.56	28(6)/36(6)/5
Other plasmid total		362 986	28.9	238	90 (86)	419	47	22	0.04	0.05	50(1)/54(1)/12
All plasmid total		610 694	27.6	344	81 (68)	535	135	167	0.20	0.24	78(7)/90(7)/17

**a.** The number of experimentally determined restriction site locations. These were all correctly predicted by the sequence. In all plasmids, the restriction sites were scattered across the full length of the plasmid. Six apparent discrepancies between the published cp32-1, -3, -4 and -6 maps (made with B31 e-1 and B31 clone p4 DNAs; Casjens *et al.*, 1997a) were resolved by additional mapping experiments. In each case, our reported sequence was verified in strain B31 MI DNA. The confirmed results are as follows: cp32-1, *SacII* site at 15.0 kbp and *SacI* at 17.6 kbp; cp32-3, *SacII* at 15.0 kbp; cp32-4, *SacII* at 22.5 kbp and there is no *PvuII* site at 31 kbp; cp32-6, *A/wNI* at 13.6 kbp.

**b.** Per cent of plasmid occupied by putative genes plus pseudogenes; putative intact genes alone in parentheses.

**c.** Predicted potentially intact genes which have no substantially larger paralogues (the 61 'questionable' genes discussed in the text are not included). This is a best estimate of genes that are likely to be functional, however the functionality of most *Borrelia* genes is unknown so there are many uncertainties. In the 10 plasmids noted in footnote i, the fraction of ≤ 300 bp genes is high, and they are not tightly packed with neighbouring genes, so it seems likely that many of these may not be real genes (see text).

**d.** DNA regions with sequence similarity to a *Borrelia* gene, but which do not appear to contain a complete open reading frame (see text).

**e.** Pseudogene fraction of all gene-like entities: number of pseudogenes/(number of all non-pseudogenes + number of pseudogenes).

**f.** Pseudogene fraction if genes ≤ 300 bp are ignored: number of pseudogenes/(number of non-pseudogenes > 300 bp + number of pseudogenes).

**g.** Number of predicted lipoprotein-encoding genes (pseudogenes in parentheses): genes whose products contain the 'stringent' [L,A,V,I,F,T,M]-[L,A,V,I,F,S]-X-[G,A,S,N]-C lipidation consensus/potential lipoprotein genes from our analysis (see text)/genes just below our lipidation prediction cut-off.

**h.** Does not include the rightmost 7.2 kbp because this, unlike the 'constant portion' of the chromosome (genes BB0001 to BB0843), has a plasmid-like character in that it contains mostly pseudogenes. About 40% of the chromosomal '≤ 300 bp genes' are homologues of similar small genes with known function in other bacteria and, unlike the plasmid '≤ 300 bp genes', they usually are closely packed with neighbouring genes.

**i.** The 10 plasmids or parts thereof that contain ≥ 22% pseudogenes in column 10 (lp5, lp17, lp21, lp25, lp28-1, lp28-3, lp28-4, lp36, lp38 and the non-cp32-like portion of lp56).

**j.** For demonstration purposes, we have separated the cp32-like and non-cp32-like parts of the linear plasmid lp56 (see text).

**k.** These plasmid sizes include the known terminal sequences that were not determined in this study; Barbour *et al.* (1996) reported the terminal 29 bp left end and 78 bp right end for lp17; Zhang *et al.* (1997) reported an additional 1227 bp that lie beyond (about) 100 bp of unclonable DNA (J. Zhang and S. Norris, personal communication) at the right end of our lp28-1 sequence which is 26 921 bp in length. Hinnebusch *et al.* (1990) reported a plasmid telomere sequence that corresponds to the right terminal 25 bp of lp56. Short regions remain unsequenced at all the other plasmid telomeres (see text).

**l.** ND, not determined.

**m.** Includes 15 silent *vlsE* gene cassettes (Zhang *et al.*, 1997).

but to our knowledge it is the largest such repeat tract to have been found in a prokaryote. In the reported sequence, there are 34 distinct repeat types; 27 of these types (128 total repeat units) are 63 bp long and seven types (48 units) are 61 bp long. The maximum number of adjacent identical units is three, and there are two large exact repeats within the tract, suggesting possible recent duplications; units 2–19 are identical to units 129–146, and units 20–30 are identical to units 31–41. In order to experimentally confirm and characterize this repeat tract further, we used Southern analyses to measure the sizes of restriction fragments that contain the repeats. Electrophoresis gels of B31 MI DNA cleaved with *MseI* (which cleaves at TTAA, and so cleaves the 71.8% A+T *Borrelia* DNA extremely frequently), *DraI*, *Asel*, *HindIII*, *EcoRI*, *StuI*, *BsrGI*, *XbaI* and *EcoO109I* (all of which are predicted not to cleave within the repeat tract) gave single DNA bands that hybridize to a 63 bp repeat DNA probe. Calculations from the resulting data gave an experimentally determined repeat tract length value of  $11.9 \pm 1.0$  kbp. In addition, *Tsp509I* is predicted to cleave the repeat region twice and gave three repeat-containing bands close to the expected sizes. We conclude that the assembly of this repeat region is likely to be accurate.

Several other smaller tandem repeat tracts (7–12 repeats with repeat unit sizes of 21, 17, 7 and 11 bp lie on plasmids lp17, lp38, lp38 and lp54 respectively) do not appear to be within genes and their functions are also unknown. The lp38 17 bp repeat (just 5' of the *ospD* gene BBJ09) and the lp17 21 bp repeat have also been sequenced from B31 derivatives with different propagation histories (Norris *et al.*, 1992; Barbour *et al.*, 1996) and in each case the number of repeats was the same as our determination, suggesting that the repeat numbers are not in extremely rapid fluctuation even during passage through a mouse. Marconi *et al.* (1994) found that the number of the lp38 17 bp repeats varied from 1 to 12 in other *B. burgdorferi* (*sensu stricto*) isolates, suggesting a longer-term instability in repeat number in that case. In addition, a number of predicted plasmid genes include tandem direct repeats. The paralogous family 80 genes (the *bdr* genes; see below) carry within them 6–18 copies of related 33 and/or 54 bp repeats (the latter include the 33 bp repeat with 21 additional bp; Porcella *et al.*, 1996; Zuckert and Meyer, 1996; Zuckert *et al.*, 1999). These repeats are variable in number among the genes within the paralogous family and often contain imprecise or fragmented repeats, but in all cases the repeats are translated in the same frame so that the proteins have related amino acid repeats. The N- and C-terminal non-repeated parts of these proteins are not uniform within the family, but are present as several types. In addition, BBI16, a putative lipoprotein, has 21.5 repeats of a nine codon unit, and BBQ47 (*erpX*) has five repeats of a five-codon unit; neither of these repeats are

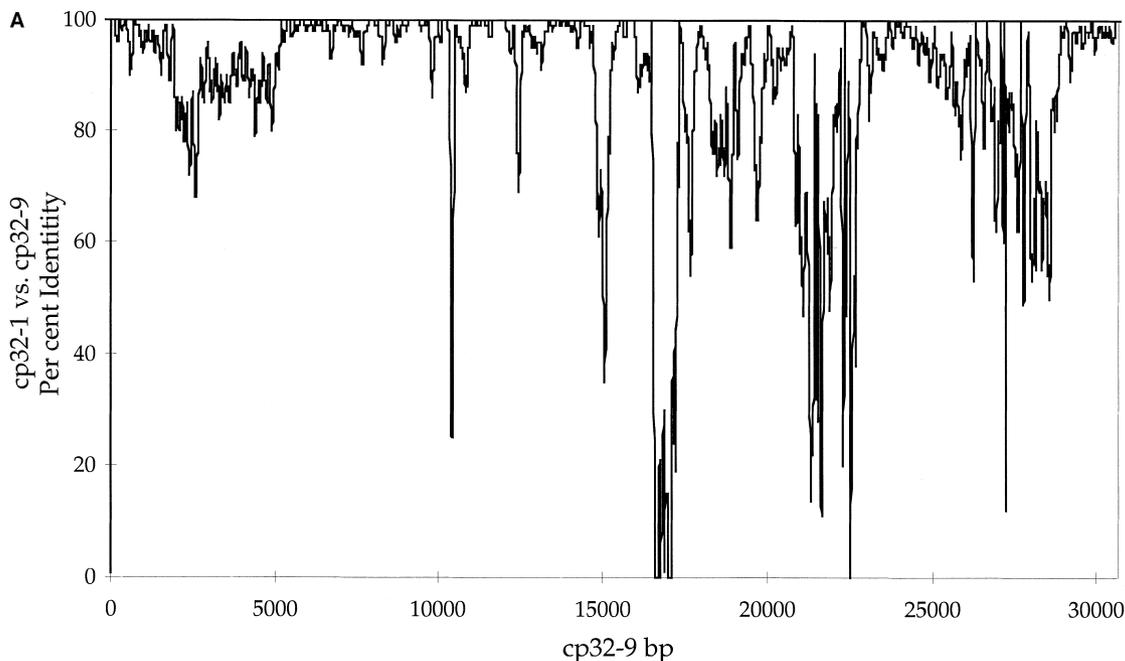
found in other members of their paralogous families. In all these intragenic repeats, the numbers of nucleotides in the repeat units are multiples of three, so that changes in repeat number would not cause early termination of translation and so are unlikely to be involved in phase variation of their expression, although variation of the proteins' properties is possible. The function of all these proteins and their repeats remains unknown.

**Inverted repeats.** On the cp32 plasmids, two of the largest intergenic gaps bracket the BBP30–BBP34 gene cluster and each of its paralogues. These gaps contain ~180 bp inverted repeats (which contain smaller inverted repeats) that were previously noted for cp32-1 and lp56 by Zuckert and Meyer (1996); similar inverted repeats surround the related BBC01–BBC03 cluster on cp9 (see also plasmid cp8.3 in Dunn *et al.*, 1994). The function of these repeats is unknown, but each of them contains an ATG with an associated GGAG possible ribosome binding site that appears to be the most likely translation start for the paralogous family 161 and 165 genes (for example BBP29 and BBP35 respectively) that extend outward from the inverted repeats (see *Supplementary materials* in the *Experimental procedures*). As a result, these divergent genes all have very similar upstream regulatory regions (for co-ordinate regulation?), and the N-terminal 15–17 amino acids are predicted to be nearly identical in all members of these two protein families.

Finally, all the *B. burgdorferi* telomeres that have been studied have very similar ~25 bp sequences at their tips, and so constitute inverted repeats at the two ends of the linear replicons (Hinnebusch and Barbour, 1991; Casjens *et al.*, 1997b; Fraser *et al.*, 1997; Zhang *et al.*, 1997; reviewed by Casjens, 1999). Because so little is known about *Borrelia* molecular biology, we have not analysed the plasmid sequences for smaller inverted repeats that might indicate regulatory protein binding sites.

#### *Overall evolutionary relationships among the B. burgdorferi plasmids*

One of the most striking observations concerning the sequences of the plasmids in *B. burgdorferi* B31 is their high degree of apparent genetic redundancy. Several previous studies have shown the presence of multiple, similar copies of various sequences in this species (Simpson *et al.*, 1990a,b; Stalhammar-Carlemalm *et al.*, 1990; Hinnebusch and Barbour, 1991; Marconi *et al.*, 1996b; Porcella *et al.*, 1996; Stevenson *et al.*, 1996; Zuckert and Meyer, 1996; Casjens *et al.*, 1997a; Misonne *et al.*, 1997; Carlyon *et al.*, 1998). We find that all of these previously identified 'repeated' sequences and many additional sequence similarities lie on the plasmids. We discuss some of these relationships below.



**Fig. 1.** The cp32 circular plasmids.

A. Quantitative comparison of cp32-1 and cp32-9 nucleotide sequences.

B. Qualitative comparison of the seven sequenced B31 cp32 plasmids.

Large multicoloured bars represent the maps of the seven cp32 plasmids and the cp32-like portion of lp56. Different colours indicate sequence groups that are more than 10% different from one another (i.e. the same colour indicates a transitive set of sequences, each of which is  $\geq 90\%$  identical to some other member of the set). Sequence groups and the comparison algorithm used are defined in *Experimental procedures*. Small solid bars above each map indicate the predicted open reading frames. In each map, the genes translated from left to right are above those translated right to left; selected gene names assigned in this study are given above each map and previously named genes and new *erp*, *bdr* and *mlp* gene names are given below (*bdr* and *mlp* gene names according to W. Zuckert and S. Porcella, respectively, personal communications). Blue bars indicate genes that have paralogues in all eight sequences, and red, orange and green bars indicate genes that do not have paralogues in all eight sequences. Some small genes, such as BBS32, BBO35 and BBQ36, that were found by our gene recognition method may be questionable because similar genes were not recognized in paralogous sequence in other cp32s. Slanted white separations within gene bars indicate shifts in reading frame or in frame stop codons relative to the other cp32s. Horizontal black and crosshatched bars within the maps indicate previously sequenced regions that had, or had not, been mapped to particular cp32 plasmids respectively (Zuckert and Meyer, 1996; Casjens *et al.*, 1997a; Gilmore *et al.*, 1997; Guina and Oliver, 1997; Stevenson *et al.*, 1998a). The numbers beside some of these bars indicate the cp32 plasmid that has an identical sequence in that region. The short arrows represent the  $\sim 180$  bp inverted repeat (see text).

**cp32 plasmid family.** The most extensive set of multiple, highly similar sequences in isolate B31 is the cp32 plasmids. It was previously known that this isolate carries multiple circular plasmids in the 29–32 kbp range that have a high degree of similarity with one another (Stevenson *et al.*, 1996, 1998a; Zuckert and Meyer, 1996; Casjens *et al.*, 1997a), and we show here that B31 MI in fact contains seven such plasmids that are homologous nearly throughout their lengths. Figure 1A shows a typical comparison of two cp32 plasmids, cp32-1 and cp32-9, indicating that there are substantial regions of very high similarity (including sections of near identity up to several kbp in length) and regions of lower similarity. Figure 1B shows the overall sequence relationships among the seven cp32s. There are three regions that are more diverse than the bulk of the cp32 DNAs, centred at about 17, 22 and 27 kbp, which correspond to the *mlp* lipoprotein, putative segregation gene cluster (previously ORF-1 to ORF-3; see below and

also Zuckert and Meyer, 1996; Casjens *et al.*, 1997a; Stevenson *et al.*, 1998a) and *erp* lipoprotein gene regions respectively. Surface lipoprotein genes could be under selection to maintain or increase diversity, and plasmid-specific partitioning functions might be expected to be at least somewhat different on each plasmid (see also Stevenson *et al.*, 1998a). Most of the cp32 genes have homologues present on every cp32 but there are a few exceptions, such as the *bdr* (family 80) or *rev* (family 63) alternatives at  $\sim 17$  kbp and the family 114, BBS41 and BBM39 alternatives at  $\sim 28$  kbp.

Stevenson *et al.* (1998a) and Akins *et al.* (1999) have pointed out that inter-cp32 recombination might have occurred in strain B31 and 297 progenitors, and the complete sequence data strongly supports this. For example, there are two types of sequences in the 2–5 kbp region, with cp32-1, -3, -7 and -8 forming one type, and there are two types of sequence at  $\sim 17$  kbp where cp32-1 and

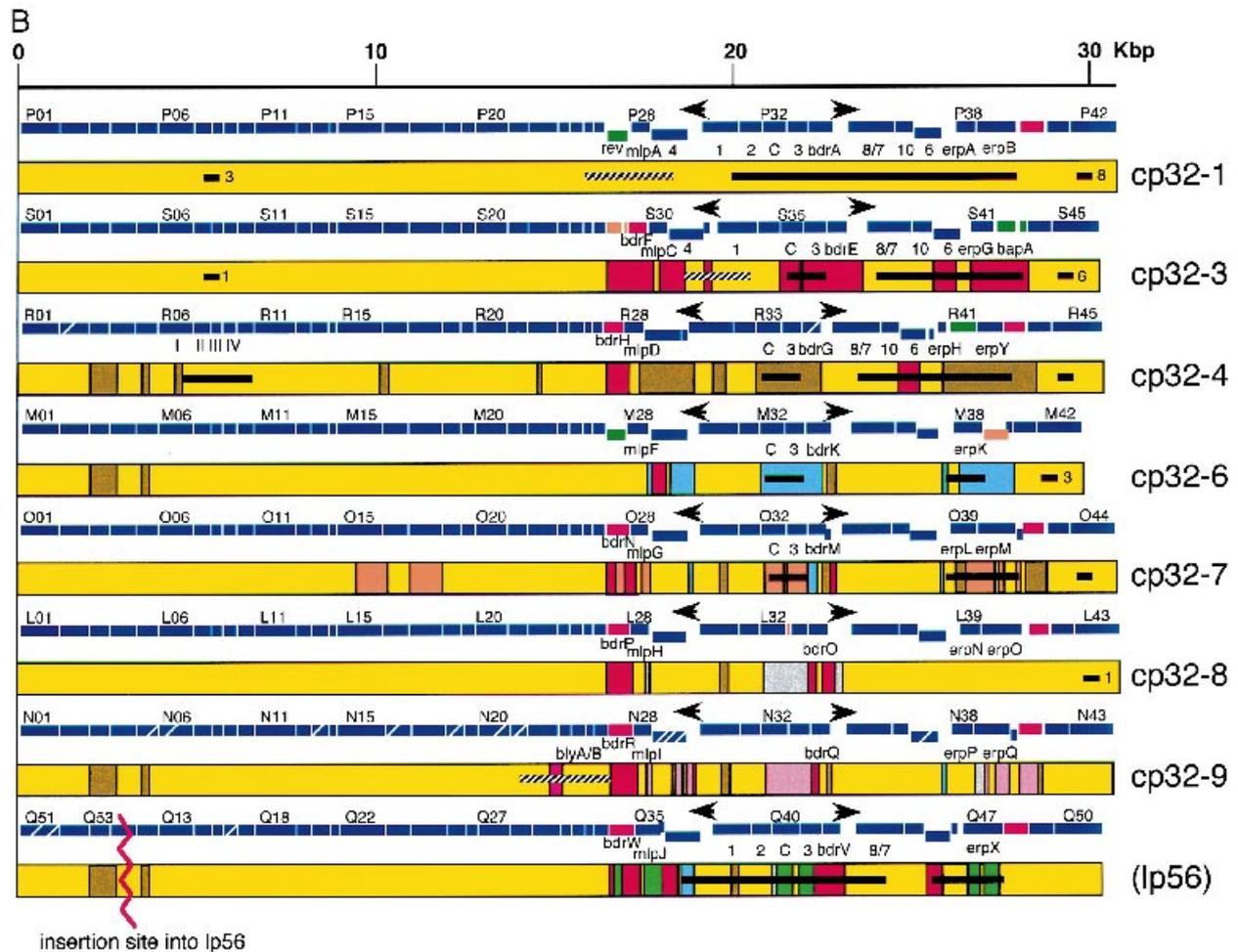


Fig. 1. Continued.

cp32-6 have a *rev* gene and the others have a *bdr* gene (Fig. 2B). Recombination is the simplest way to imagine generating situations such as this in which all four possible combinations of 'alleles' are present in the cp32s: A–B in cp32-1; A–b in cp32-3, cp32-7 and cp32-8; a–B in cp32-6; and a–b cp32-4, cp32-9 and lp56 (upper- and lower-case letters represent the two types of sequence in each of the two regions). Given their extensive similarity, it would be surprising if recombination did not occur among the cp32s, although it is perhaps remarkable that no such recombination has been observed in the laboratory (El Hage *et al.*, 1999; R. van Vugt and S. Casjens, unpublished).

The high variability of the Erp and Mlp lipoproteins has contributed to speculation that they might be involved in presenting different surface antigenicities to the host (Porcella *et al.*, 1996; Stevenson *et al.*, 1996, 1998a; Casjens *et al.*, 1997a). It is thus possible that the cp32s are 'only' complex mechanisms for disseminating and controlling the expression of these and perhaps other cp32 genes

that encode possible host interaction proteins, such as the Rev lipoprotein or BlyAB haemolysin (Guina and Oliver, 1997; Gilmore and Mbow, 1998). We have previously speculated that these plasmids could be prophages because: (i) bacteriophage-like particles have been produced by several *B. burgdorferi* strains (Hayes *et al.*, 1983; Neubert *et al.*, 1993; Schaller and Neubert, 1994); (ii) sequence relationships among the cp32s are reminiscent of temperate bacteriophage families (Casjens *et al.*, 1992, 1997a); and (iii) prophages often express genes that affect bacteria–host interactions (Cheetham and Katz, 1995). Our finding here that cp32-1 gene BBP42 and its family 145 paralogues are similar to a putative *Streptococcus thermophilus* phage  $\phi$ O1205 morphogenetic gene (Stanley *et al.*, 1997) lends additional credence to this hypothesis.

**Similarity between lp56 and the cp32s.** The linear plasmid lp56 contains within it an essentially intact copy of a cp32-like plasmid. This region of lp56 is not identical to

any of the seven circular cp32 plasmids, but represents yet another member of this family of sequences. Figure 2 shows that lp56 most probably arose through integration of a circular cp32 into a linear progenitor, and now lies between nucleotides 6585 and 36935 in lp56. The progenitor cp32 was apparently identical to cp32-4, -6 and -9 in the integration region, and integration must have occurred by opening the cp32 circle between nucleotides equivalent to, for example, 3125 and 3126 of cp32-4. It is likely that this integration event happened relatively recently, since, in spite of the fact that the event disrupted predicted protein coding regions on both progenitor plasmids and so probably destroyed their function (Fig. 2B), no nucleotides appear to have been lost in the integration region since that time (Fig. 2C). The gene on the cp32-like progenitor that was disrupted would have encoded a protein that is 96.9% identical to the product of gene BBR04 of cp32-4; although it is now present in two parts, its original reading frame is intact as parts of 'genes' BBQ54 and BBQ11. On the linear plasmid parent, the integration event separated a family 62 gene into two parts, now the C-terminal portion of BBQ10 and N-terminal portion of BBQ55; its original reading frame also appears to be intact, although the putative original start codon is altered.

There is no long sequence similarity between the integration sites on the putative linear and cp32 progenitors (only 2 bp are the same at the integration site; Fig. 2C). This, with the lack of a substantial inverted repeat surrounding either site, makes it seem unlikely that this recombination event happened either through homologous recombination or through an integrase-like reaction. There is also no evidence for generation of a terminal duplication during the integration event, suggesting it was not transposase mediated.

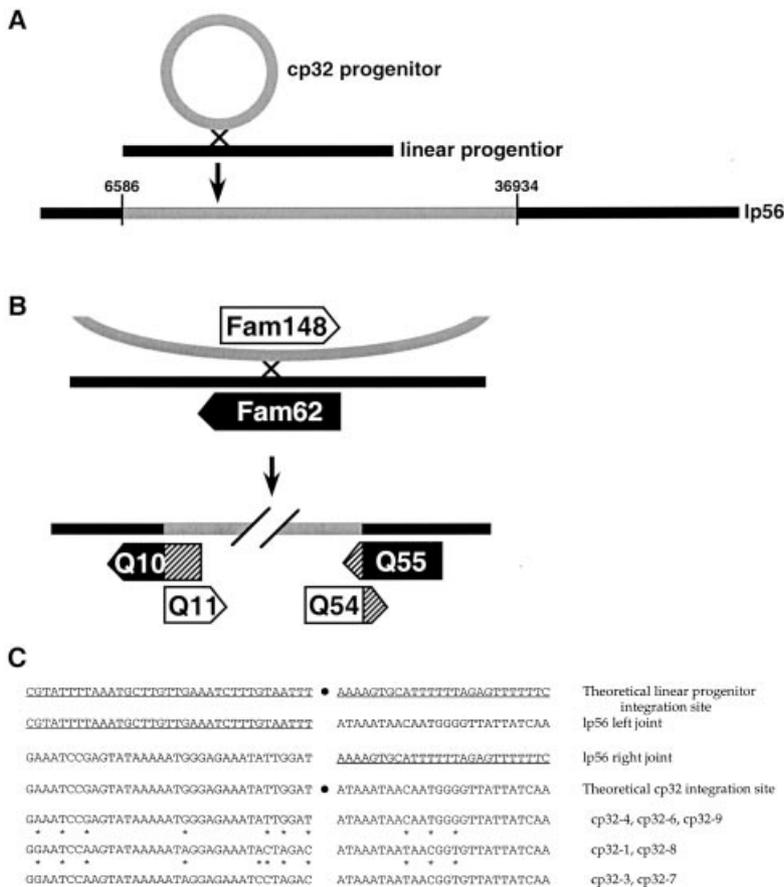
**lp54 and cp9 relationships to cp32s.** The linear plasmid lp54 contains nine blocks of homology with the cp32 plasmids that are in the same orientation and order on the two plasmids (Fig. 3), so that 26 of lp54's 76 predicted genes have cp32 paralogues. It seems very likely that this indicates another event in which a cp32-like plasmid integrated into a linear plasmid, opening the cp32 between, for example, the cp32-1 genes BBP18 and BBP19. However, unlike the lp56 event, in lp54 there have been insertions (for example between lp54 genes BBA51 and BBA55) and replacements (for example between lp54 BBA31 and BBA38) of the DNA between the cp32-like genes. These differences suggest that this integration is more ancient than the lp56 event. It is curious that the positions of two major cp32 lipoprotein genes, *mlp* and *erp*, are occupied in lp54 by the non-homologous lipoprotein gene pairs *ospAB* and *dbpAB* respectively.

cp9 is also very similar to a section of the cp32s (Fig. 3). cp9 could have been derived from a *rev* gene-containing cp32 (such as cp32-1) through: (i) deletions of the 21 kbp of contiguous DNA outside the BBP27–BBP37 interval as well as genes BBP32, BBP34 and BBP36; (ii) inversions of the BBP30–BBP33 gene region (possibly mediated by the inverted repeats surrounding these genes) and the BBP27 gene; and (iii) replacement of the cp32 BBP28 (*mlpA*) gene by BBC06 through BBC08. As in lp54, it is interesting to note that surface protein genes *epmA* (BBC06) and *mlp* occupy similar positions in cp9 and the cp32s respectively. The cp32 sites that were opened during the putative integration events that formed lp54 and lp56 are not near one another, nor are the deletion end-points that formed cp9 near to these sites. However, one of the deletion end-points that formed the shortened plasmid cp18 from a cp32 in isolate N40 (Stevenson *et al.*, 1997) is indistinguishable at current resolution from the lp54 integration site, near the C-terminus of BBP18 on cp32-1.

**lp5 relationship to lp21.** The linear plasmids lp5 and lp21 also have extensive similarity. lp5 has only one open reading frame (BBT02) that does not have a paralogue in lp21. lp21 contains the 11 kbp 63 bp direct repeat tract (above) between genes BBU05 and BBU06. These three elements together constitute an ~12 800 bp insertion relative to the otherwise very similar lp5. Open reading frames BBT01, BBT03 and BBT05 on lp5 and BBU01, BBU02, BBU09 and BBU10 on lp21 appear to be fragments of genes found on other plasmids (see below), and genes BBU07 and BBU08 appear to be the result of rather recent duplications of fragments of larger genes that appear elsewhere on both plasmids, suggesting that several, perhaps illegitimate, recombination events have occurred on the progenitors of these plasmids.

**Mosaic relationships among the linear plasmid telomeric regions and other recent recombination events among the linear plasmids.**

In the cases in which sequence is known to the ends of the linear DNAs in *Borrelia*, a conserved ~25 bp sequence is present at the tip (Casjens *et al.*, 1997b and references therein). However, there are a number of additional, more lengthy similarities among the plasmid termini. Within each of the following seven groups, telomeric sequences are similar for at least several hundred bp near the ends: (i) the left ends of lp5, lp17, lp21, lp28-1, lp28-3 and lp28-4, the right ends of lp25, lp28-2 and lp56, and the right end of the chromosome (Fig. 4 shows the complex relationships among this group of telomeres); (ii) the left ends of lp25 and lp36; (iii) the left ends of lp28-2 and lp36; (iv) the left end of lp56 and the right ends of lp28-4 and lp36; (v) the right ends of lp5 and lp21; (vi) the left end of lp54 and the right end of



**Fig. 2.** Integration of a cp32-like circular DNA into a linear progenitor plasmid to form linear plasmid lp56.

**A.** Location of putative cp32 integration event. Numbers indicate the lp56 nucleotides immediately outside the integrated cp32. **B.** Genes affected by the cp32 integration event in lp56. The gene names above are the paralogue families of the putative parental genes, and below the products are shown. Crosshatched fill indicates regions fortuitously in frame with the parental genes. **C.** Nucleotide sequences surrounding the putative integration event. Underlining indicates putative linear progenitor sequence, asterisks above the cp32 sequences indicate differences from cp32-4, and bullets (●) indicate the sites where the progenitors are thought to have recombined.

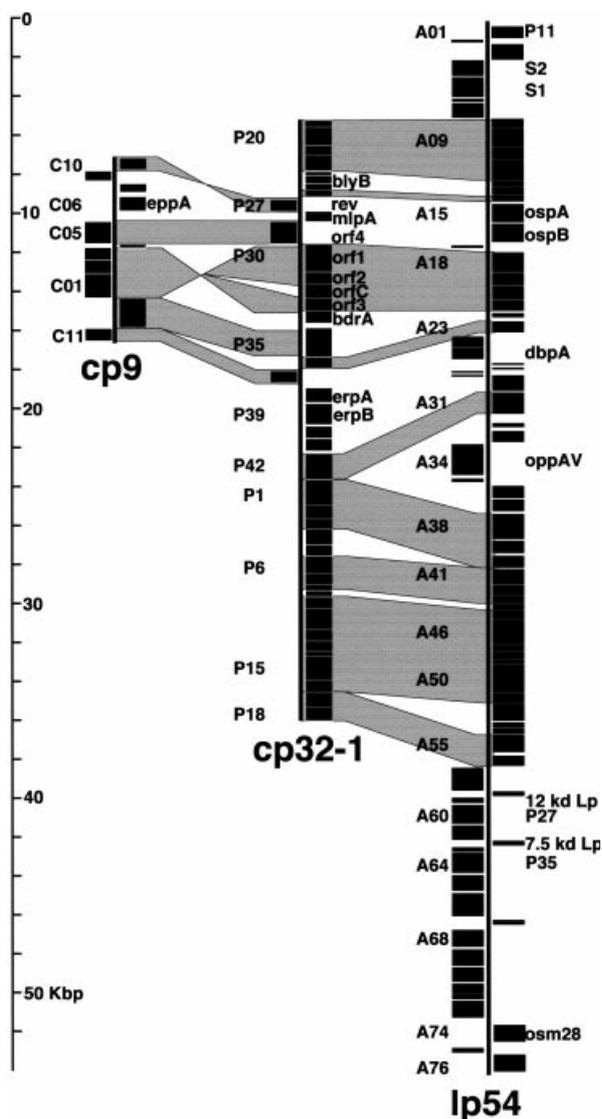
lp28-3; and (vii) the right ends of lp28-1 and lp38. Only the right end of lp54 does not have some substantial similarity with another B31 telomere. It is not known whether these related sequences contribute to telomere function. The strikingly mosaic relationships within each of the above groups can be explained if the termini grew by successive partial replacements and/or additions derived from the telomeric regions of other plasmids. Some of these inter-telomere region similarities are quite extensive and are nearly identical. The two most similar pairs are: (i) the left end of lp17 and the right end of lp56, where the terminal 2655 bp [including the terminal sequences of these two DNA molecules reported by Hinnebusch *et al.* (1990)] are 99.8% identical (three single bp differences and deletions of 1, 4 and 30 bp in lp17); and (ii) the right end of lp36 and the left end of lp56, which are 96% identical for 2392 bp if a perfect 902 bp inversion of one relative to the other is taken into account. Such high similarity suggests rather recent duplications of these telomeric regions.

There are many additional non-telomeric regions of nucleotide similarity among the plasmids. These regions of homology usually have abrupt boundaries and vary from, for example, the region between 3321 and 5167 bp of lp28-1 (genes BBF06–BBF10), which is 97.7% identical to a section of lp36 (genes BBK41–BBK45) and a

2535 bp non-tandem duplication on lp38 (genes BBJ29–BBJ32 and BBJ43–BBJ45.1) in which the repeated sections are 99.1% identical, to regions whose similarity is only recognizable in the sequence of the encoded proteins. All the sequence relationships described above combine to give a picture in which recombination events, especially ones that cause duplications of sequence, appear to have been rather common among the *Borrelia* linear plasmids.

**Possible DNA exchange between the chromosome and plasmids.**

Although the chromosomes of different *B. burgdorferi* isolates are in general very similar, some isolates are known to have between 7 and 20 kbp of DNA added to the chromosomal right end (Casjens *et al.*, 1995). B31 has 7.2 kbp of extra right-end DNA when compared with the isolates with the shortest chromosomes. We previously noted that DNA probes from near the right end of the B31 chromosome hybridize with plasmids in many other *B. burgdorferi* isolates (Casjens *et al.*, 1997b), and that at least some sequences within 7.2 kbp of the right end are similar to sequences on plasmids (Fraser *et al.*, 1997). From this, we suggested that the rightmost 7.2 kbp of chromosomal DNA was likely to be plasmid derived. The sequences of all the B31 plasmids allow a much more

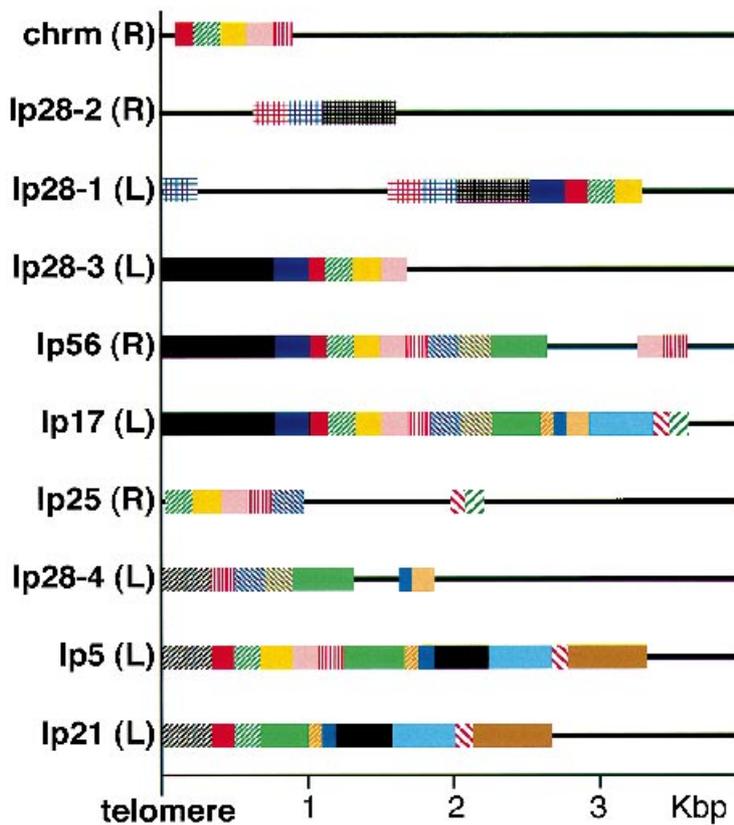


**Fig. 3.** Structural similarities among cp9, cp32 and lp54. Maps of plasmids lp54, cp32 and cp9 indicate their predicted genes with black rectangles (to right of line translated top to bottom and to left in opposite direction). Selected gene names are given to the left of each map and names previously used in the literature are given to the right. Grey connections between the three plasmids indicate regions of sequence similarity.

complete analysis of the relationship between the chromosome and the plasmids, and Fig. 5A shows that nearly all of the DNA in the rightmost 7.2 kbp is in fact similar to sequences on the B31 plasmids because all of the genes and pseudogenes in this region are members of paralogous families made up of largely plasmid genes. The region between genes BB0844 and BB0852 is the largest non-rRNA region without substantial open reading frames on the chromosome (BB0845.1 to BB0849.1 as determined by our gene analysis protocol are unlikely to be functional genes; see below). In a FASTA (Pearson, 1990)

comparison of all the B31 plasmid sequences with the entire B31 genome (requiring >62% identity and no gap >64 bp), 4739 patches of non-self similarity > 100 bp were recognized. Of these, 4668 were between two plasmids, and only 71 were similarities between a plasmid and the chromosome. Fifty of the latter 71 similarities were in the rightmost 7.2 kbp of the chromosome. The remaining 21 plasmid–chromosome similarities were with the following chromosomal sequences: eight transporter genes, two *S*-adenosylhomocysteine nucleosidase genes, two with BB0223 and BB0224, seven with BB0733 and BB0734, and two small lp38 fragments with BB0003 (no potential functions have been deduced for the last five chromosomal genes). This analysis demonstrates that plasmid-like sequences are much more likely to be found very near the right end than elsewhere on the chromosome. Combined with the numerous similarities among plasmid telomeres (above), these findings support the notion that there appears to have been frequent exchanges of terminal regions among the linear replicons of *Borrelia*. Curiously, there is little similarity between the linear plasmids and the left end of the B31 chromosome. Two small plasmid-like sections in gene BB0003, about 2 kbp from the left chromosomal end, are the only current indication of plasmid-like sequences near the left chromosomal telomere. We do not know why this exchange is limited to the right end of the chromosome. Except for a similar phenomenon that may be limited to the left end of the *B. japonica* chromosome, evidence for terminal plasmid–chromosome exchanges has not yet been found in other *Borrelia*s (Casjens *et al.*, 1995; M. Murphy and S. Casjens, unpublished).

**Plasmid rearrangements and pseudogenes in other *Borrelia* isolates?** Are rearrangements and pseudogenes unique to the isolate that was sequenced by the *Borrelia* genome project? Almost certainly not. For example, apparent changes in *Borrelia* plasmid size or geometry with propagation (Sadziene *et al.*, 1992; Munderloh *et al.*, 1993; Fikrig *et al.*, 1995a,b; Ferdows *et al.*, 1996; Tilly *et al.*, 1997; Ryan *et al.*, 1998) and differences in the sizes of related plasmids or regions in different bacterial isolates have been observed (Hyde and Johnson, 1988; Feng *et al.*, 1996, 1998; Marconi *et al.*, 1996a; N. Palmer and S. Casjens, unpublished). In addition, several plasmid genes sequenced from other isolates appear to contain open reading frame-disrupting mutations, deletions or other rearrangements when compared with the B31 sequence (Rosa *et al.*, 1992; Marconi *et al.*, 1993a; Restrepo and Barbour, 1994; Wang *et al.*, 1997b; Kornacki and Oliver, 1998; Akins *et al.*, 1999). Finally, several phylogenetic analyses of plasmid genes have shown some evidence for lateral transfer of these plasmid genes (Dykhuisen *et al.*, 1993; Marconi *et al.*, 1994; Stevenson and Barthold,



**Fig. 4.** Some sequence relationships among the telomeres of the strain B31 linear replicons. Eight linear plasmid ends and the right chromosomal end are shown with their telomeres on the left. Similarly coloured blocks indicate blocks of similar sequence (>65% identity), and thinner black lines indicate sequences that have no paralogue within the regions shown. Each colour boundary indicates a sequence break compared with one of the other telomeres in the figure.

1994; Jauris-Heipke *et al.*, 1995; Livey *et al.*, 1995; Will *et al.*, 1995), so it seems likely that such rearrangements and pseudogenes are not unique to B31 plasmids.

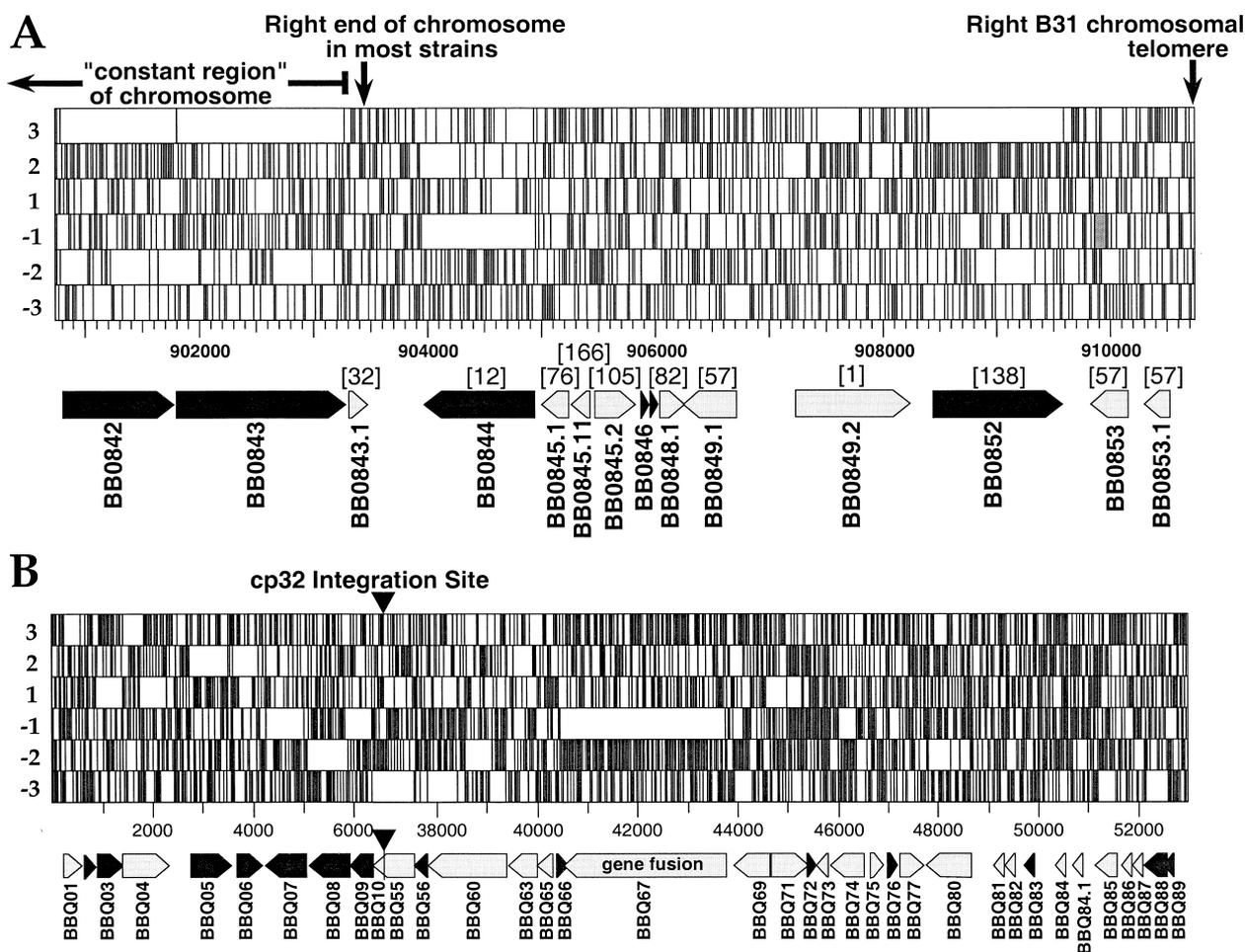
#### *The open reading frames of the B. burgdorferi plasmids*

**Some plasmids carry numerous pseudogenes.** There are several very unusual aspects of the protein coding potential of the B31 plasmids. Unlike the 'constant portion' of the chromosome (genes BB0001 to BB0843), a number of the B31 plasmids have: (i) an apparent protein coding density that is <70%, a value that is substantially less than the *B. burgdorferi* chromosome or other bacterial genomes; (ii) a surprisingly large fraction of small open reading frames ( $\leq 100$  codons); and (iii) a large number of predicted 'genes' that are truncated or have damaged reading frames relative to other members of their paralogous gene families. For ease of discussion below, we define a 'pseudogene' as any region of DNA that is similar in sequence to a paralogous *Borrelia* predicted gene or to a gene from another organism, but which is obviously truncated and/or does not have full open reading frames relative to its homologues. These mostly appear to be mutationally damaged genes that include frameshift changes, in frame stop codons, and fused or truncated genes. We suspect that most of them may not currently

encode functional polypeptides. However, in any given instance, we cannot rule out *in vivo* synthesis or even a biologically important function of a protein 'fragment'.

We initially identified 731 putative (non-pseudo)genes and 167 pseudogenes on the 21 plasmids, and their names and locations are shown in Fig. 6. Putative genes were identified according to Salzberg *et al.* (1998), and pseudogenes not found as truncated members of paralogous gene families by that procedure were identified by DNA similarities; see *Experimental procedures*. Among the 731 potential genes, 61 are probable false identifications because they lie inside another gene or pseudogene, or because they are very small and were not identified in paralogous sequence elsewhere in the genome (these 'questionable genes' are ignored in the remainder of this discussion, but are shown in Fig. 6 and are noted in the complete predicted gene list in *Supplementary information*; see *Experimental procedures*). Thus, our current best estimate is that there are 670 potentially functional genes and 167 pseudogenes on the B31 plasmids (Table 1).

Ten of the B31 plasmids (lp5, lp17, lp21, lp25, lp28-1, lp28-3, lp28-4, lp36, lp38 and the non-cp32-like portion of lp56) contain 87% of the pseudogenes and have a total non-pseudogene protein coding capacity of only 41%, and a very large fraction (43%) of these predicted genes



**Fig. 5.** Lack of long open reading frames in lp56 and the plasmid-like sequences near right end of strain B31 chromosome.

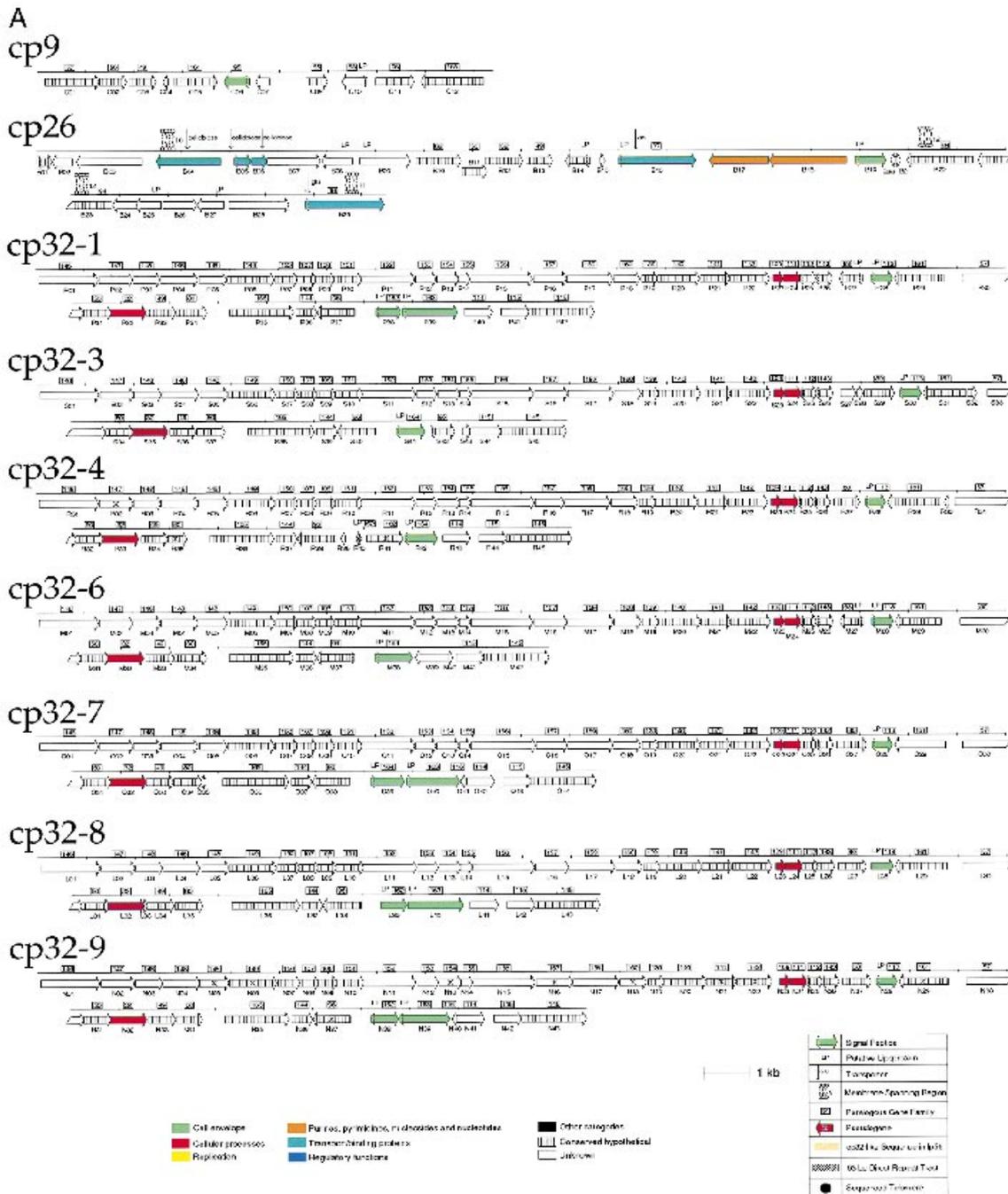
**A.** A reading frame diagram for the rightmost 10 000 bp of the B31 chromosome. All six reading frames (1, 2 and 3 reading from left to right; -1, -2 and -3 reading right to left) are indicated with stop codon locations marked by vertical lines in each frame. Below, arrows indicate genes (black) and pseudogenes (grey) from our analysis; the paralogous gene families to which they belong are indicated above each arrow and the gene names are given below.

**B.** A reading frame diagram for the non-cp32-like portion of lp56. The six reading frames and putative genes and pseudogenes are displayed as in part A; the black triangle indicates where the cp32-like sequences were removed (see text).

are  $\leq 300$  bp in length (Table 1). The remaining 'low-pseudogene' plasmids and the constant portion of the chromosome contain 10% and 11% putative genes that are  $\leq 300$  bp in length (where genes average about 750 and 1000 bp in size) respectively. Putative protein-encoding genes are nearly always tightly packed on these latter 'well-behaved' DNAs. Although some pseudogenes (for example family 57 members) tend to be located near the ends of these plasmids, pseudogenes are found scattered across the 10 'high-pseudogene' plasmids (Fig. 6; see Fig. 8 below for the distributions of several gene families and their pseudogenes). In addition, the  $\leq 300$  bp genes on these plasmids are typically not in regions of tightly packed genes (see, for example, the regions between lp25 genes BBE09 and BBE16 and between lp28-4 genes BBI16 and BBI19 in Fig. 6). The fact that such regions,

which contain only small widely scattered open reading frames, exist only on the 'high-pseudogene' plasmids suggests that they too may no longer have a useful function. Of course, the functionality of any given small open reading frame is unknown, but many of the non-tightly packed small putative genes on these plasmids may be the result of spurious gene prediction in regions where functional genes no longer exist. Thus, 670 'intact' plasmid genes is likely to be an overestimate.

The plasmids lp28-2, lp54, cp9, cp26, the seven cp32s and the cp32-like portion of lp56 appear to carry mainly apparently 'intact' genes that are arranged in a tightly packed fashion. These 11 plasmids plus the cp32-like region of lp56 are predicted to carry 87% protein-encoding sequences (90% if 'simple frameshifted' pseudogenes are included), a value that is similar to most other completed



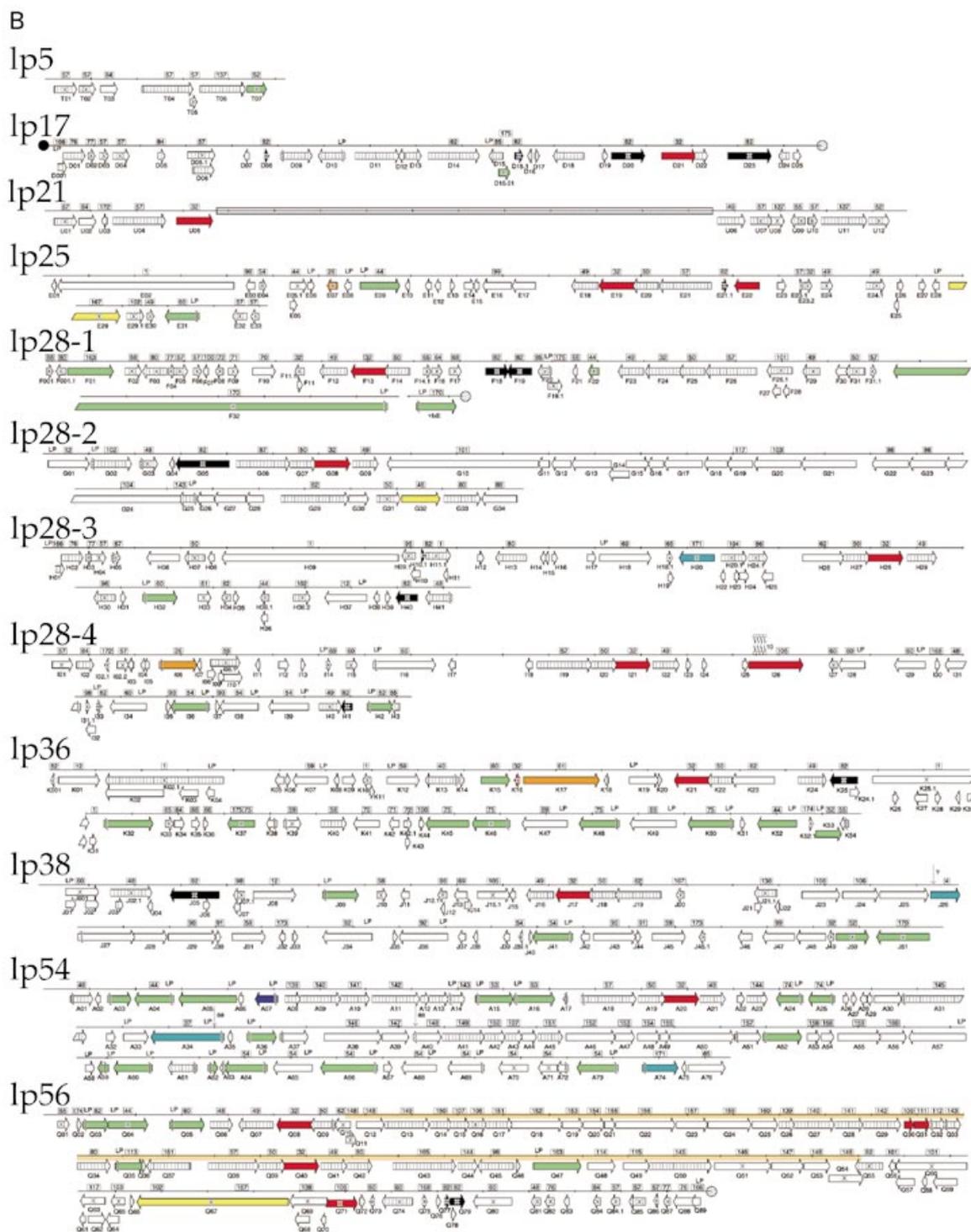


Fig. 6. Continued.

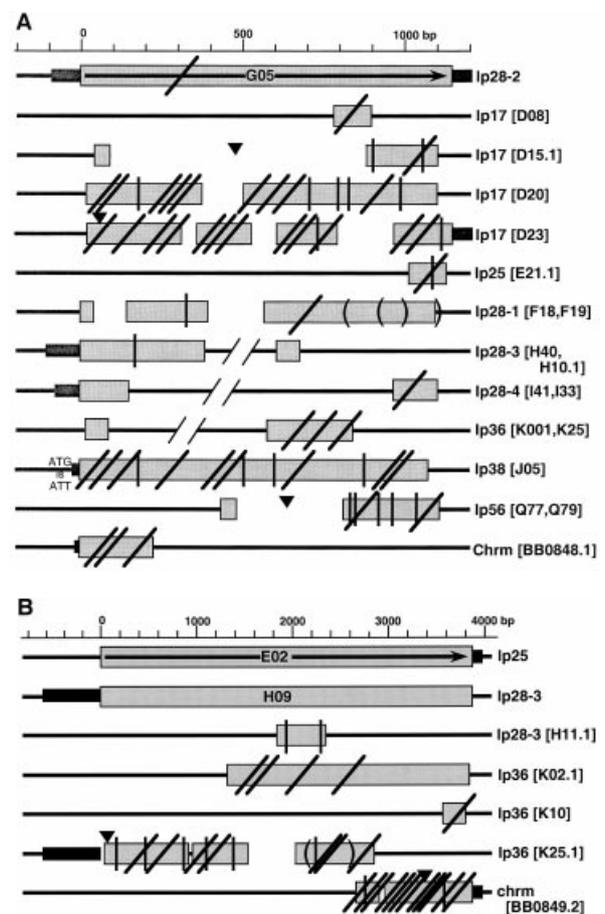
prokaryote genomes; for example, *B. burgdorferi* chromosome, 93% (Fraser *et al.*, 1997); *Mycoplasma genitalium*, 88% (Fraser *et al.*, 1995); *Escherichia coli*, 89% (Blattner *et al.*, 1997); *Treponema pallidum*, 93% (Fraser *et al.*, 1998); *Mycobacterium tuberculosis*, 91% (Cole *et al.*,

1998). There are, nonetheless, a few apparently inactivated genes even in these 'well-behaved' B31 plasmids (Table 1). The most damaged of these plasmids is cp32-9, in which nine of its 42 genes have been inactivated by simple frame-disrupting (mostly point) mutations. It is not clear

why cp32-9 contains so many mutations of this type; perhaps all or parts of it have become superfluous and have begun to decay. The other 'well-behaved' plasmids carry a small number of more dramatic rearrangements, e.g. the apparent insertion of the 5' portion of a BBP29 (family 161) homologue into a precursor *erp*-like (family 162) gene to create two new genes on cp32-4, a severely truncated *erpH* gene (BBR40) and an in frame fusion between the N-terminus of the family 161 member and the C-terminus of the precursor *erp* gene to form gene BBR41. Although BBR41 may well be expressed because it carries the putative translation start signal of the family 161 gene, it seems unlikely that this fusion protein is functional as the family 161 portion is severely truncated and the *erp*-like portion has lost its lipidation signal. In a recent analysis of the *erp* genes of strain 297, Akins *et al.* (1999) did not find a fusion gene analogous to BBR41, suggesting it may have arisen recently. It is not clear why some plasmids should carry so many pseudogenes and others have few or none; perhaps those with many have undergone recent rearrangements events that may have damaged genes directly and/or made various regions redundant.

**The nature of the plasmid pseudogenes.** The least-damaged pseudogenes contain only one or a few simple frameshifts relative to their homologues. We find a number of such apparently lightly damaged genes in the B31 plasmids that contain only one or a few in frame stop codons and/or frameshifts, e.g. BBG05 in lp28-2; BBQ04, BBQ16 and BBQ51 in lp56; BBR02 and BBR35 in cp32-4; BBN05, BBN06, BBN13, BBN16, BBN19, BBN21, BBN22, BBN29 and BBN37 in cp32-9.

Most of the pseudogenes are much more badly damaged. As an example of the nature of these pseudogenes, Fig. 7 shows all the regions that are similar to two putative linear plasmid genes BBG05 and BBE02. BBG05 is homologous to a putative transposase gene family originally discovered in *Anabena*, *Saccharopolyspora*, *Salmonella* and the thermophilic bacterium PS3 (Bancroft and Wolk, 1989; Krause *et al.*, 1991; Gulig *et al.*, 1992; Donadio and Staver, 1993; Murai *et al.*, 1995). BBG05, which lies near the left end of plasmid lp28-2, has convincing similarity throughout its length to these putative transposase genes. However, BBJ05, which is the most intact member of this family in B31, has a single frameshift relative to its homologues and the nucleotide sequence surrounding the frameshift is not similar to known programmed frameshift sites (Gesteland and Atkins, 1996), so we suspect that BBJ05 has been functionally inactivated by a frameshift mutation. There are 15 regions of similarity to BBG05 elsewhere in the genome, 14 on other linear plasmids and one near the right end of the large chromosome. None is a complete gene and all appear to have been severely damaged by mutation (Fig. 7A). In addition to deletions, they have



**Fig. 7.** Two families of pseudogenes on the *Borrelia* linear plasmids.

A. BBG05 and its related pseudogenes.

B. BBE02 and BBH09 and their related pseudogenes.

Pseudogenes were located as described in the *Experimental procedures*. Horizontal grey bars below the full genes indicate the extent of the other regions of similarity (similar sequence is aligned vertically). Gaps between such bars indicate deletions, black triangles indicate the locations of insertions, vertical lines are in frame stop codons, slanted lines are frameshifts, and parentheses indicate inversions. On the right, the location is given with the gene name.

suffered insertions, inversions, frameshift mutations and mutation to in frame stop codons. The second most intact version, BBJ05, is missing its C-terminus, the ATG start codon is changed to an ATT, and it contains at least nine frameshifts and four in frame stop codons; it does not contain the particular frameshift that exists in BBG05, suggesting that the pseudogenes may be derived from a non-frameshift-containing progenitor of BBG05. It is possible that this paralogous family's multiplicity is a result of past transposition events, but most pseudogenes on the B31 plasmids are not likely to have resulted from transposition. For example, gene family 1 has no known homologues outside *Borrelia*, and it contains a 'typical' set of pseudogenes (Fig. 7B). BBE02 and BBH09, on plasmids

lp25 and lp28-3, are thought to be intact as they are large and have very similar open reading frames. There are four badly damaged paralogues elsewhere on the linear plasmids, and one, near the right end of the large chromosome, that has suffered two deletions, an insertion, 12 frameshifts, and one in frame stop codon in about 1300 bp of remaining DNA.

Some regions of the plasmids appear to be particularly rich in pseudogenes. The non-cp32-like portion of lp56 contains one of the highest fractions of pseudogenes among the B31 plasmids (Table 1). Of the 36 genes and pseudogenes there, seven are short putative genes  $\leq 300$  bp long that have no homologues, and 22 appear to be pseudogenes (most of them severely damaged). Figure 5B shows the nearly complete lack of substantial open reading frames in long sections of this DNA. Interestingly, the largest lp56 open reading frame BBQ67 is a bipartite gene in which the N-terminal 80% is convincingly similar to full-length adenine DNA methyltransferase genes (best match *Helicobacter pylori* HP1354) and the C-terminal 20% is very similar to BBG02, a putative lipoprotein-encoding gene of unknown function on lp28-2 (but whose lipidation consensus was removed by the postulated BBQ67 fusion). Thus, even the larger genes on the B31 plasmids may have been recently altered by DNA rearrangements. Also noteworthy is a section of lp56 DNA that is similar to the BBI26–BBI34 region of lp28-4. All of the lp28-4-like pseudogenes in this region of lp56 have accumulated serious mutational damage, and a transposase BBG05 pseudogene between BBQ75 and BBQ80 suggests that transposition may have contributed to the damage. Curiously, only three of the lp56 cp32-like progenitor's 41 genes are damaged; the gene broken by the insertion event and two that contain a small number of frameshift mutations. It is tempting to speculate that, after the cp32-like plasmid integrated into lp56's linear progenitor, many of the linear plasmid's genes became superfluous. In addition, plasmid-like sequences in the rightmost 7.2 kbp of the chromosome also appear to be largely decaying (Fig. 5A).

Relatively few pseudogenes have been found in other bacteria, and these have been rare exceptions when compared with the number of functional genes. Genes with one or a few frameshifts, in frame stop codons or inactivated control regions have been found in a few anecdotal cases (for example Hall *et al.*, 1983; Morris *et al.*, 1995; Fsihi *et al.*, 1996; Lai *et al.*, 1996). The number of such genes in the completely sequenced bacterial genomes is low, e.g. only 1.3% of the genes (23 out of 1758) in the complete genome of *Haemophilus influenzae* contain substantiated in frame stop codons or frameshifts, and similar values of 0.9%, 0.6% and 1.4% are found for the chromosomes of *T. pallidum*, *M. genitalium* and *H. pylori* respectively (some of these may be in the initial stages of evolutionary

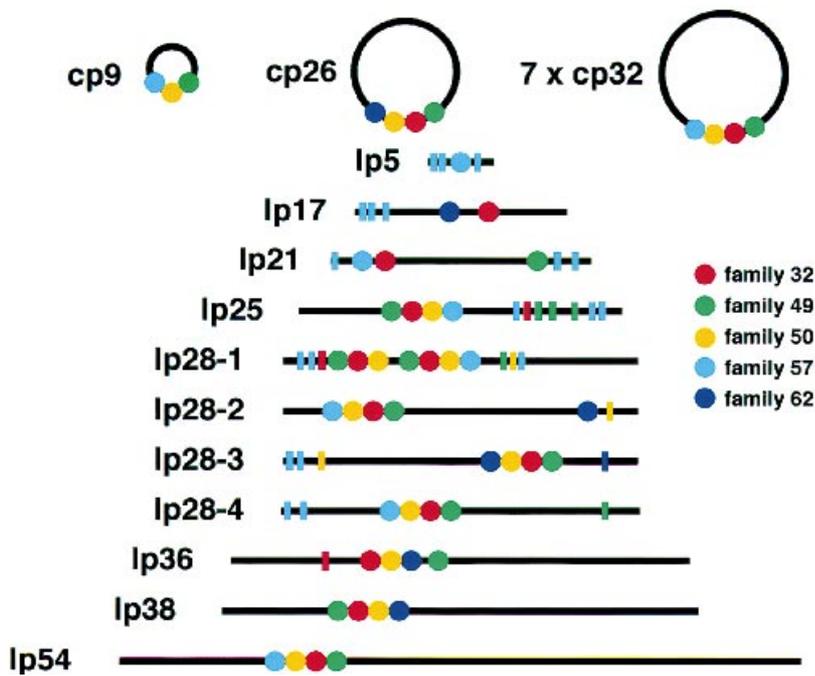
inactivation whereas others could be phase variable genes in the 'off' state). Only one pseudogene, BB119 which contains a single simple frameshift, has been identified among the 843 genes of the 'constant portion' of the *Borrelia* chromosome (Fraser *et al.*, 1997). These values may be underestimates because the status of genes of unknown function that have no homologues cannot presently be assessed, but as related genomes are sequenced other instances of damaged currently hypothetical genes may become recognizable (the comparison of two *H. pylori* isolates has allowed recognition of a few additional apparently damaged genes; Alm *et al.*, 1999). Truncated gene fragments have been observed even less frequently in bacteria, although a few have been reported especially in association with defective prophages (see, for example, Xiang *et al.*, 1994; Skamrov *et al.*, 1995; Casjens, 1998). Smith *et al.* (1997) have reported a substantial number of damaged genes in the chromosome of *Mycobacterium leprae* (3.7% pseudogenes in a 1.5 Mbp region), and Andersson *et al.* (1998) observed that 12 of the 846 genes in the complete genome sequence of *Rickettsia prowazekii* appear to be mutationally damaged and, in addition, find that only 75.4% of its genome appears to encode proteins. All of these pseudogene frequencies are very much less than those found on some of the *Borrelia* linear plasmids (over 50% in lp5, lp21, lp28-1 and the non-cp32-like portion of lp56). Andersson *et al.* (1998) interpreted pseudogenes in bacteria to be the remnants of genes that are no longer useful but have not yet been completely eliminated from the genome. This is a reasonable hypothesis, but at present we do not know why the *Borrelia* plasmids should have so many damaged but not yet eliminated genes.

#### *Highly paralogous nature of plasmid genes*

As has been mentioned above, the genes and pseudogenes on the B31 plasmids form a large number of paralogous gene families. The complete genome contains 161 families of paralogous genes, 107 of which contain at least one plasmid-borne member. Family sizes range from 41 members (including pseudogenes) in family 57 to families with only two members; 83% of the  $\geq 300$ -bp-long plasmid non-pseudogenes are members of such families. The very high fraction of genes with plasmid-borne paralogues may reflect some as yet unknown advantage in carrying multiple similar but slightly different copies of these genes. The family membership of each plasmid gene is indicated in Fig. 6, and the members of each gene family can be found in the *Supplementary information* section (see *Experimental procedures*).

#### *Predicted functions of plasmid open reading frames*

Each open reading frame on the plasmids was compared



**Fig. 8.** Putative replication and partition genes are the most widely distributed paralogues on the 21 plasmids. Candidates for plasmid replication and partition genes (see text) are represented by coloured circles on maps of the B31 plasmids. These putative genes are members of paralogous gene families as follows: 32, red; 49, green; 50, yellow; 57, light blue; and 62, dark blue. Pseudogene relatives are indicated by rectangles of the same colours.

with the extant sequence database as previously described (Fleischmann *et al.*, 1995; Fraser *et al.*, 1997). Of the 670 potentially intact genes on the 21 B31 plasmids, only 39 (5.8%) and 14 (2.1%) were found to be convincingly similar to previously sequenced genes of known and unknown function outside *Borrelia* respectively. The predicted functional categories for these genes, as deduced from these similarities, are indicated in Fig. 6. More detailed information on these database matches can be found in the *Supplementary information* section of the *Experimental procedures*.

**Possible replication and partition genes.** Among the genes present on the plasmids are families of putative genes that, because of their similarity to proteins of known function in other bacteria, have been suggested to encode proteins involved in plasmid DNA partitioning and replication. Zuckert and Meyer (1996) noticed that family 32 genes are similar to the *parA* gene of *E. coli* bacteriophage P1, which is a member of a large family of genes implicated in bacterial plasmid partitioning. The first *Borrelia* family 50 genes that were analysed were reported to contain a short motif found in proteins involved in initiation of replication in rolling circle plasmids (Dunn *et al.*, 1994; Zuckert and Meyer, 1996), however this motif is poorly represented in many members of this paralogous gene family, lowering the credence of this idea somewhat. Members of three other paralogous gene families, 57 and 49, which are similar to ORF-1 and ORF-3 of Dunn *et al.* (1994), respectively, and gene family 62, all of which have no known homologues outside *Borrelia*, are also present

on most plasmids, and when they are present they are nearly always tightly clustered with family 32 and 50 genes. When present, the family 62 genes typically replace the family 57 gene in this gene cluster, and analysis with PSI-BLAST (Altschul *et al.*, 1997) shows a weak but significant similarity between these two families and that these two families are each other's closest relatives in *B. burgdorferi*; perhaps they have similar functions. Because of their universal presence, these five gene families (32, 49, 50, 57, 62) are all reasonable candidates for functioning in plasmid replication and partitioning. No other putative genes are nearly so widely distributed on the plasmids, and the sequences of the various proteins within each family are not identical and thus could have the possibly required plasmid-specific properties (see also Stevenson *et al.*, 1998a). Figure 8 shows the presence and locations of these gene families on the B31 plasmids. It is notable that the smaller plasmids do not carry the full set of four genes; all of the  $\geq 25$  kbp plasmids carry all four genes, family 32, 49, 50 and 57/62, whereas lp21 and cp9 carry three of them, lp17 has two and lp5 has only one. Every plasmid carries an apparently intact family 57 or 62 gene. Several plasmids, lp28-1, lp28-2 and lp56, carry two 'intact' members of some of these gene families; it is possible that these are the result of recent interplasmid recombination events.

**Possible lipoprotein genes.** Previous study of the Lyme disease spirochetes has focused largely on their outer surface proteins because these proteins are important in vaccine development, in bacterial detection

and in the interaction of these bacteria with their arthropod and vertebrate hosts. Most of the outer surface proteins that have been characterized are lipoproteins (Rosa, 1997). We previously noted that there were 105 genes on the chromosome and the 11 previously published plasmids that encode proteins that contain a type II signal sequence in which a positively charged N-terminal region is followed by a hydrophobic signal sequence and the lipidation consensus [L,A,V,I,F,T,M]-[L,A,V,I,F,S]-X-[G,A,S,N]-C (Fraser *et al.*, 1997). In an effort to develop more quantitative criteria for identifying potential *Borrelia* lipoprotein genes, this lipidation consensus sequence (which was deduced from data on other bacterial species; Sutcliffe and Russell, 1995), information from the *Borrelia* lipoprotein literature, and the assumption that proteins with membership in paralogous families with *bona fide* lipoprotein members and only conservative changes from the above consensus are likely to be lipoproteins were used to perform a hidden Markov model (HMM) analysis to find all of the B31 genes that might encode lipoproteins (see *Experimental procedures*). Some of the predicted proteins that fit these criteria include the following (mostly conservative) additions to the above lipidation consensus: G, N and S in position -1; M and T in position -2; and L, T and I in position -4 (for example BBA68, BBA69, BBI36, BBI38, BBI39 and BBJ41 which all have a T in position -2 and an I in position -4 and are members of a gene family that includes proteins that contain the more stringent consensus). These slightly expanded criteria suggest that there may be as many as 91 plasmid lipoprotein-encoding genes [ignoring eight pseudogenes and a questionable gene (BBI32) that meet the criteria] and 41 apparently intact chromosomal lipoprotein genes. The genes that encode these proteins fall into 27 of the B31 paralogous families, nine of which have only lipoprotein members. Most of these predicted lipoproteins are likely to be surface proteins because nearly all of the previously characterized plasmid-encoded putative lipoproteins are surface localized, although periplasmic proteins may also be lipidated in *Borrelia* (Bono *et al.*, 1998; Kornacki and Oliver, 1998). This may still be an underestimate of the lipoprotein-coding potential of *Borrelia* because there are 32 additional genes (15 chromosomal, 17 plasmid) that appear to have a properly placed Cys and a reasonable HMM score, but whose characteristics did not quite meet our criteria (including the known surface protein gene BBK32 and genes such as BBK45, BB0329 and BB0330 that belong to paralogous families which contain other lipoprotein genes). These are included in a complete list of predicted lipoprotein genes in the *Supplementary information* section.

*Borrelia* appears to have an especially large fraction of its genome devoted to lipoprotein production. Putative lipoprotein genes represent about 4.9% of the chromosomal genes, a value somewhat higher than but similar

to other completely sequenced bacterial genomes such as *H. pylori* (1.3%; Tomb *et al.*, 1997) or *T. pallidum* (2.1%; Fraser *et al.*, 1998). The putative lipoprotein genes on the *B. burgdorferi* plasmids represent 14.5% of the 'intact' plasmid genes (17% if the 'near cut-off' genes above are included). This high fraction of genes that encode potentially outer surface proteins correlates with the observation in other bacterial parasites that proteins involved in interaction with the host are often plasmid encoded.

Each of the seven cp32 plasmids, and the cp32 inserted into lp56, carry two types of putative lipoprotein genes, an *mfp*-type gene (family 113) (characterized in isolate 297 by Porcella *et al.*, 1996) and one or two adjacent *erp* genes (families 162, 163 and 164; Lam *et al.*, 1994; Stevenson *et al.*, 1996, 1998a,b; Casjens *et al.*, 1997a). The five new *erp* genes discovered here on plasmids cp32-4, cp32-8 and cp32-9 are named *erpY*, *erpN/O* and *erpP/Q* respectively (see Fig. 1B). The ErpY protein is rather similar to ErpL, the ErpP and Q proteins are quite similar to Erps A and B, respectively, and Erps N and O are identical to Erps A and B respectively. cp32-1 and cp32-6 each encode an additional putative lipoprotein (family 63/Rev; Gilmore and Mbow, 1998) that are identical to one another.

#### **Other possible functions for plasmid genes.**

In addition to the putative partitioning proteins and lipoproteins discussed above, other plasmid-encoded functions that have been characterized are a porin on lp54 (BBA74; Skare *et al.*, 1996), a fibronectin binding protein on lp36 (BBK32; Probert and Johnson, 1998) and two haemolysin genes *blyA* and *blyB* on a cp32 (Guina and Oliver, 1997; we find the *bly* genes they characterized on cp32-9 – BBN23 and BBN24 – but each of the cp32s and lp56 carries paralogues of both of these genes). Other putative functions convincingly predicted by similarity to genes of known function in other organisms for intact genes on the plasmids are two genes for GMP synthesis on cp26 (BBB17, BBB18; Margolis *et al.*, 1994; Zhou *et al.*, 1997), four genes for sugar transport on cp26 (BBB04, BBB05, BBB06 and BBB29; Fraser *et al.*, 1997), a nicotinamidase on lp25 (BBE22; Fraser *et al.*, 1997), a DNA helicase on lp28-2 (BBG32; Fraser *et al.*, 1997), a 5'-methylthioadenosine/*S*-adenosylhomocysteine nucleosidase and a multi-drug transporter on lp28-4 (BBI06, BBI16; Fraser *et al.*, 1997; Cornell and Riscoe, 1998), ABC transporter component genes on cp26, lp38 and lp54 (BBB16, BBJ26 and BBA34; Fraser *et al.*, 1997; Bono *et al.*, 1998), an adenine deaminase on lp36 (BBK17; Fraser *et al.*, 1997), and a possible DNA methylase on lp56 (BBQ67). (BBB01 and BBA76 have weak similarities to acyl phosphatase of *E. coli* and thymidylate synthase-complementing protein of *Dictyostelium discoideum* respectively.) In addition, only

BBG02, BBK13 and members of paralogous families 94, 137 and 145 have convincing homologues among hypothetical genes of unknown function from another bacteria.

*B31-like plasmids present in other Lyme Borrelia natural isolates*

All natural isolates of *B. burgdorferi* (sensu stricto) and related species that have been examined carry numerous linear plasmids in the 5–56 kbp size range as well as multiple circular plasmids. Do all isolates carry a set of extrachromosomal elements that are similar to those found in this study of isolate B31, or does each bacterium carry a complement of plasmids chosen from a much larger menu? And how similar are similar-sized plasmids in different isolates?

**Linear plasmids.** *B. burgdorferi* isolates Sh-2-82 and CA-11.2A have lp54s that have gene orders and restriction maps that are, at relatively low resolution, similar to lp54 of B31 (Marconi *et al.*, 1996a; Fraser *et al.*, 1997; R. van Vugt and S. Casjens, unpublished). These findings, combined with the observations that BBA15 and BBA16 (*ospA* and *ospB*) are almost universally present in *B. burgdorferi* isolates and when examined are always on a 50–55 kbp linear plasmid, strongly suggest that a plasmid similar to B31 lp54 is present in almost all *B. burgdorferi* bacteria in nature. (Out of hundreds of isolates analysed, less than 10 have been found that may lack the *ospA/B*-containing lp54 plasmid, and those few could, in theory, have lost it during isolation or only lost the genes used as probes, e.g. Samuels *et al.*, 1993; Casjens *et al.*, 1995; Anderson *et al.*, 1996; Guttman *et al.*, 1996; Mathiesen *et al.*, 1997.) However, Feng *et al.* (1998) reported that, unlike B31, in isolate N40 the *ospA* and *dbpA* genes are on different linear plasmids, suggesting that this plasmid may not be completely constant among different isolates.

Isolated B31 lp17 DNA hybridized only to a similar-sized linear plasmid in three of four isolates previously examined (Hinnebusch and Barbour, 1991). A *vsE* gene similar to that on B31 lp28-1 was found on an ~20 kbp linear plasmid in isolate 297 (Kawabata *et al.*, 1998), and two different sequences cloned from a 29 kbp linear plasmid of *B. burgdorferi* isolate 297 hybridize to a 28 kbp plasmid in B31 and in 12 out of 16 other isolates examined (Xu and Johnson, 1995). Finally, an *ospD* gene probe from B31 lp38 hybridizes with a similar-sized plasmid in all cases in which the *ospD* gene was present (18 of 24 isolates examined; Norris *et al.*, 1992; Marconi *et al.*, 1994). In a preliminary systematic study, we have found that hybridization targets for multiple, unique probes from each of the 12 B31 linear plasmids are present in at least some members (usually more than half) of a panel of 15 North American *B. burgdorferi* (sensu stricto) isolates, and when present

almost always reside on a linear plasmid similar in size to the cognate B31 plasmid. In addition, multiple probes from any particular plasmid of B31 usually hybridize to a single plasmid in most other isolates (N. Palmer and S. Casjens, unpublished).

**Circular plasmids.** Tilly *et al.* (1997) showed that the B31 cp26 has a restriction map that is very similar to cp26s of three other independent geographically diverse *B. burgdorferi* isolates. In addition, cp26 gene BBB19 (*ospC*) has been found in all isolates examined (Marconi *et al.*, 1993b; Theisen *et al.*, 1993; Stevenson and Barthold, 1994; Fukunaga and Hamase, 1995; Jauris-Heipke *et al.*, 1995; Livey *et al.*, 1995; Wilske *et al.*, 1995, 1996) and, when studied, it was found to be present on a 26–28 kbp circular plasmid, so B31 cp26-like DNA elements appear to be universally present in *B. burgdorferi*.

Circular plasmids approximately 9 kbp in size have been found in several Lyme agent isolates (Hyde and Johnson, 1988; Simpson *et al.*, 1990a; Stalhammar-Carlemalm *et al.*, 1990; Champion *et al.*, 1994; Dunn *et al.*, 1994). The relationships among most of these are not known, but two, cp8.3 in *B. afzelii* isolate lp21 (Dunn *et al.*, 1994) and cp9 in B31 (Fraser *et al.*, 1997), have been completely sequenced. Their structures are sufficiently similar (both have BBC08 homologues and inversions of the BBC01–BBC03 regions relative to the cp32s; see Fig. 3) that it seems very likely that they are derived from the same small progenitor, rather than having independently arisen from a cp32-like plasmid. There are only two major differences: cp9 putative lipoprotein (*rev* paralogue) gene BBC10 is missing in cp8.3, and in cp9 BBC06 (*epaA*) and BBC07 replace part of cp8.3.

Multiple related ~30 kbp plasmid sequences (very probably circular cp32 homologues) are present in all *B. burgdorferi* isolates that have been examined (Simpson *et al.*, 1990b; Marconi *et al.*, 1996b; Porcella *et al.*, 1996; Theisen, 1996; Zuckert and Meyer, 1996; Casjens *et al.*, 1997a; Akins *et al.*, 1999). Porcella *et al.* (1996) and Akins *et al.* (1999) found seven different homologues of cp32-borne paralogue family 113 genes in isolate 297, and here we characterize seven cp32 plasmids and the related lp56 in isolate B31, and other B31 cultures have been observed to contain two additional cp32s (Casjens *et al.*, 1997a). It is not known whether this is an indication that all natural isolates will carry 7–10 specific cp32 plasmids, or whether each will carry different numbers of a large variety of cp32 family plasmids. The cp32 sequences presented here will allow this question to be addressed in the future.

**Unrelated extrachromosomal DNAs in other *B. burgdorferi* isolates?** Has the genome project identified all or most of the types of extrachromosomal elements present in *B. burgdorferi* in nature? There are no reports

of *B. burgdorferi* linear plasmids whose sizes are not close to members of the B31 complement of linear plasmids or of plasmid-borne genes that are not present in the B31 plasmid sequences. It is, however, noteworthy that closely related species *B. andersonii*, *B. afzelii*, *B. garinii*, *B. valaisiana*, *B. hermsii* and *B. turicatae* have been reported to carry linear plasmids in the 90–180 kbp range, which are not known to be related to the known B31 plasmids (Casjens *et al.*, 1995; Busch *et al.*, 1996; Marconi *et al.*, 1996a), so it is quite possible that additional linear plasmid types will be found in *B. burgdorferi* in the future. Circular plasmid distributions have been surveyed much less thoroughly. Although a uniquely sized 18 kbp circular plasmid, cp18, was found in *B. burgdorferi* N40, it appears to be a cp32-like plasmid with a single 14 kbp deletion (Stevenson *et al.*, 1997). Larger 50–70 kbp circular plasmids have been reported in *B. garinii*, *B. bissettii* and *B. burgdorferi* isolates (Simpson *et al.*, 1990a; Casjens *et al.*, 1995; Carlyon *et al.*, 1998).

All these findings combine to strongly indicate that a very substantial overlap exists among the plasmid types carried by various natural *B. burgdorferi* (*sensu stricto*) isolates and that a significant fraction of the plasmid types present in this species in nature are likely to have been characterized in this study. The plasmid sequences reported here will provide a necessary knowledge base for deciphering similarities and differences in the plasmid complements carried by other Lyme agent isolates and perhaps even the non-Lyme agent *Borrelia*s.

## Conclusions

The nucleotide sequences of the 21 plasmids in *B. burgdorferi* isolate B31 are now known. A surprisingly small fraction (8%) of their putative genes have similarity to genes in other genera, and these similarities are not to known bacterial virulence genes. As many parasitic and pathogenic bacteria carry host interaction genes on plasmids, this suggests that *B. burgdorferi* interacts with its hosts in basically different ways than the more well-studied bacterial pathogens. This is perhaps not surprising because the spirochetes are only very distantly related to those proteobacteria and Gram-positive bacteria. The complete *B. burgdorferi* genome sequence will serve as a critical resource in the unravelling of the molecular pathogenesis of Lyme disease.

The most unusual aspects of the *Borrelia* genome are the presence of: (i) linear replicons; (ii) more than 20 replicons in a single bacterium; (iii) large tracts of directly repeated short DNA sequences; (iv) a substantial fraction of plasmid DNA that appears to be in a state of evolutionary decay; and (v) evidence for numerous and sometimes recent exchanges of DNA sequences among the plasmids. These rearrangements probably contributed to the

presence of a large number of pseudogenes on some of the linear plasmids. It appears that *B. burgdorferi* B31 is in the throes of a rapid evolutionary spurt in terms of the physical arrangements of the linear plasmids. Whether this process is 'finished' (i.e. among the many rearrangements are ones that gave selective advantage) but not yet ameliorated (messy 'loose ends' not yet made physically tidy) or is a continuing process in all extant Lyme agent lineages is not yet known; however, Restrepo and Barbour (1994) noted several pseudogenes on a *B. hermsii* plasmid that suggest a possible independent pseudogene generation there. Given the presumed ongoing evolutionary sparring between *B. burgdorferi* and the defence mechanisms of its hosts, it is tempting to speculate that the rearrangements might be the product of a relatively new, as yet untidy, diversity generation mechanism.

No potential mechanism for the plasmid DNA rearrangements is decipherable from our current knowledge, however the events appear to include non-homologous recombination (for example integration of a cp32-like plasmid into lp56 and the 902 bp inversion on lp56 relative to lp36). Zhang *et al.* (1997) have described, on plasmid lp28-1, a contiguous set of 15 silent gene-like cassettes ('gene' BBF32) whose sequence information can be moved into an expression site, the *vlsE* gene on the same plasmid, to generate diversity in the VlsE outer surface protein. This probably homologous recombination appears to be quite active under some circumstances, and its enzymatic machinery could possibly be responsible for some of the other rearrangements. In other bacteria, dispersed paralogous gene fragments may serve as silent cassettes for a 'dispersed cassette' diversity generation mechanism (for example, the pilin gene in *Neisseria gonorrhoeae*; Koomey, 1994). However, this seems a rather unlikely function for most pseudogenes described here because the many frameshifts and in frame stop codons contained within them would block expression if they were moved into an expression site.

Perhaps the simplest scenario for generation of the current situation is: (i) that many DNA sequences have been rapidly transferred among the plasmids, at least sometimes through non-homologous and duplicative rearrangements; and (ii) that many of the duplicated and/or truncated genes thus generated were no longer under selection for function and so have begun to decay through random mutational events. This model, however, gives rise to several unanswered questions as follows. If illegitimate recombination is in fact frequent among the linear plasmids, how do they maintain the apparently rather constant plasmid sizes as has been observed in comparisons of multiple independent isolates? Is plasmid spread fast compared with the rate of rearrangements? Is there a mechanism for maintaining plasmid sizes? Why have only certain plasmids and only the extreme right end of the chromosome

participated in the rearrangements? Is there an underlying advantage to allowing such apparently disorderly DNA rearrangements? These questions are not easily answered at present, but study of plasmids from other isolates and more knowledge of the biology of *Borrelia* should lead to a better understanding.

## Experimental procedures

### Sequence determination and DNA methodology

The whole-genome random nucleotide sequencing methodologies that were used are described by Fraser *et al.* (1997) and references therein. A summary of the features of the improved TIGR ASSEMBLER program, which was used to assemble the sequences described here, can be found in the *Supplementary information* deposited at the *Molecular Microbiology* web site (see below). Southern analysis and restriction map construction were performed as described by Casjens and Huang (1993) and Casjens *et al.* (1997a).

### Reading frame analysis

Potential protein coding genes were initially identified using GLIMMER (Salzberg *et al.*, 1998). To find additional pseudogenes, a modified version of FASTA (Pearson, 1990) was used to find nucleotide sequence similarities between plasmid genes and regions where no open reading frames were initially found. A set of nucleotide sequences containing 330 plasmid genes (including all unique genes and at least the longest member of each paralogous family with a plasmid-borne member) was used to probe a set of target sequences that contained the 79 longest plasmid 'intergenic regions' (as defined by the original gene search when only genes  $\geq 300$  bp were considered) for sequence similarities. FASTA parameters were tuned so that about 600 matches were returned. Lowering of the stringency of this search resulted in additional matches that were nearly always short (<20 bp) stretches of very high similarity in otherwise not convincingly similar sequence, so we believe that most regions of similarity  $\geq 100$  bp were found. Each of the resulting matches, as well as all truncated members of paralogous gene families were evaluated manually, and matrix comparison plots of the two regions (by DNA STRIDER; Douglas, 1994) were used to determine whether the match was part of a longer region of similarity. Each resulting patch of similarity between one gene in the probe set and a region of the target set were considered to be pseudogenes. It is often difficult to determine precisely where similarity ends in such pseudogenes, so their boundaries are less precise than putative gene ends.

Searches for similarities between putative plasmid-encoded proteins and putative proteins in the extant sequence data bases were performed with BLAST (Altschul *et al.*, 1997) as previously described (Fleischmann *et al.*, 1995; Fraser *et al.*, 1997). Possible B31 encoded lipoproteins were identified by generating a preliminary list by rules derived from other bacterial species (Sutcliffe and Russell, 1995), and using this alignment in a hidden Markov model analysis of the N-terminal region of all predicted B31 proteins was constructed

using the HMMER 1.8.4 package (S. Eddy, personal communication; Eddy, 1998).

### Sequence comparisons

A 'sequence type' in Fig. 1B is defined to be a set of sequences where a path of  $\geq 90\%$  identity matches can be traced from any member to any other member (perhaps through other members), but in which any two members do not have to be  $\geq 90\%$  identical to each other (transitive closure). No group member is  $\geq 90\%$  identical to any non-group member. This transitive closure was applied to a set of pairwise comparison data as follows. First, a multiple sequence alignment of the seven cp32s and the cp32 sequence in lp56 was performed with a modified FASTA that provided a common structure and co-ordinate system. Each of the 28 pairwise comparisons in this structure was analysed for per cent identity for window lengths of 25, 50, 75...750 bp. Each 25 bp window was then marked as a potential member of a  $\geq 90\%$  identical transitive closure set if any of the windows spanning that 25 bp was  $\geq 90\%$  identical. Next, in each of the pairwise comparisons, all  $\geq 150$  bp regions that were bounded by <90% identical 25 bp windows and whose set of overlapping 100 bp windows were all <90% was marked as <90% identical regions. If a  $\geq 90\%$  region that was spanned by a window (of any of the 25–750 bp sizes) that was <90% identical and if the  $\geq 90\%$  region was <150 bp, it was marked as <90% identical. This procedure smoothes over some small features and effectively, at the pairwise level, shows features that are  $\geq 150$  bp. In this way, the  $\geq 90\%$  identity transitive closure sets shown in Fig. 2B were deduced for each of the 25 bp windows in each cp32 sequence.

### Accession numbers and annotation

The nucleotide sequences have been deposited with GeneBank under the following accession numbers: cp32-1, AE001575; cp32-3, AE001576; cp32-4, AE001577; cp32-6, AE001578; cp32-7, AE001579; cp32-8, AE001580; cp32-9, AE001581; lp5, AE001583; lp21, AE001582; lp56, AE001584. There are 14 ambiguous nucleotides in the 21 B31 plasmids (see *Supplementary information*); these are positions that we interpret to be genuinely heterogeneous in the population of DNA clones that was sequenced.

### Supplementary information

Supplementary information has been deposited on the *Molecular Microbiology* web site (<http://www.blackwell-science.com/mmi>). It contains: (i) a list of all of the *Borrelia* B31 plasmid genes and annotates them according to location, data base hits, predicted pseudogene status, previous names and references, etc.; (ii) a cross-referenced table of paralogous gene families; (iii) an explanation of our lipoprotein prediction analysis and annotation of potential plasmid-encoded lipoproteins; (iv) a list of reasons why each of the 167 putative pseudogenes on the plasmids is thought to be a pseudogene; (v) an analysis of the tandemly repeated sequences on the plasmids and the cp32 inverted repeats; (vi) locations of the 14 ambiguous nucleotides in the 21 strain B31 plasmids; and

(vii) methodological information on our sequence assembly techniques. In addition, the (searchable and downloadable) *B. burgdorferi* B31 nucleotide sequences, gene list with predicted gene functions, as well as paralogue and homologue alignments are available at the TIGR *Borrelia* web site (<http://www.tigr.org/tdb/mdb/bbdb/bbdb.html>).

### Note added in proof

Additional circumstantial evidence for cp32 plasmids being prophage DNAs has been provided by Eggers and Samuels (1999) *J Bacteriol* **181**: 7308–7313, who found cp32 DNA within the capsids of bacteriophage-like particles released from *B. burgdorferi* strain CA-11.2A.

### Acknowledgements

We thank the members of the TIGR sequencing group for excellent technical assistance, Jeff Lawrence and David Ussery for help with GC skew analyses, and Tom Schwan for mouse passage and demonstration of the infectivity of *B. burgdorferi* clone 4a. This work was supported by a grant from the G. Harold and Leila Y. Mathers Charitable Foundation.

### References

- Akins, D.R., Caimano, M.J., Yang, X., Cerna, F., Norgard, M.V., and Radolf, J.D. (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.
- Alm, A., Ling, L., Moir, D., King, B., Brown, E., Doig, P., et al. (1999) Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen *Helicobacter pylori*. *Nature* **397**: 176–180.
- Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**: 3389–3402.
- Anderson, J.F., Flavell, R.A., Magnarelli, L.A., Barthold, S.W., Kantor, F.S., Wallich, R., et al. (1996) Novel *Borrelia burgdorferi* isolates from *Ixodes scapularis* and *Ixodes dentatus* ticks feeding on humans. *J Clin Microbiol* **34**: 524–529.
- Andersson, S.G., Zomorodipour, A., Andersson, J.O., Sichert-Ponten, T., Alsmark, U.C., Podowski, R.M., et al. (1998) The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. *Nature* **396**: 133–140.
- Bancroft, I., and Wolk, C.P. (1989) Characterization of an insertion sequence (IS891) of novel structure from the cyanobacterium *Anabaena* sp. strain M-131. *J Bacteriol* **171**: 5949–5954.
- Barbour, A.G. (1988) Plasmid analysis of *Borrelia burgdorferi*, the Lyme disease agent. *J Clin Microbiol* **26**: 475–478.
- Barbour, A.G. (1993) Linear DNA of *Borrelia* species and antigenic variation. *Trends Microbiol* **1**: 236–239.
- Barbour, A.G., and Garon, C.F. (1987) Linear plasmids of the bacterium *Borrelia burgdorferi* have covalently closed ends. *Science* **237**: 409–411.
- Barbour, A.G., Carter, C.J., Bundoc, V., and 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., and Saint Girons, I.S. (1989) Linear chromosome of *Borrelia burgdorferi*. *Res Microbiol* **140**: 507–516.
- Blattner, F.R., Plunkett, G., Bloch, C.A., Perna, N.T., Burland, V., Riley, M., et al. (1997) The complete genome sequence of *Escherichia coli* K-12. *Science* **277**: 1453–1474.
- Bono, J.L., Tilly, K., Stevenson, B., Hogan, D., and Rosa, P. (1998) Oligopeptide permease in *Borrelia burgdorferi*: putative peptide-binding components encoded by both chromosomal and plasmid loci. *Microbiology* **144**: 1033–1044.
- Burgdorfer, W., Barbour, A.G., Hayes, S.F., Benach, J.L., Grunwaldt, E., and Davis, J.P. (1982) Lyme disease – a tick-borne spirochetosis? *Science* **216**: 1317–1319.
- Busch, U., Teufel, C.H., Boehmer, R., Wilske, B., and Preac-Mursic, V. (1995) Molecular characterization of *Borrelia burgdorferi* sensu lato strains by pulsed-field gel electrophoresis. *Electrophoresis* **16**: 744–747.
- Busch, U., Hizo-Teufel, C., Boehmer, R., Fingerle, V., Nitschko, H., Wilske, B., et al. (1996) Three species of *Borrelia burgdorferi* sensu lato (*B. burgdorferi* sensu stricto, *B. afzelii*, and *B. garinii*) identified from cerebrospinal fluid isolates by pulsed-field gel electrophoresis and PCR. *J Clin Microbiol* **34**: 1072–1078.
- Carlyon, J.A., LaVoie, C., Sung, S.Y., and Marconi, R.T. (1998) Analysis of the organization of multicopy linear- and circular-plasmid-carried open reading frames in *Borrelia burgdorferi* sensu lato isolates. *Infect Immun* **66**: 1149–1158.
- Casjens, S. (1998) The diverse and dynamic structure of bacterial genomes. *Annu Rev Genet* **32**: 339–377.
- Casjens, S. (1999) Evolution of the linear DNA replicons of the *Borrelia* spirochetes. *Curr Opin Microbiol* **2**: 529–534.
- Casjens, S., Hatfull, G., and Hendrix, R. (1992) Evolution of dsDNA tailed-bacteriophage genomes. *Semin Virol* **3**: 383–397.
- Casjens, S., and Huang, W.M. (1993) Linear chromosomal physical and genetic map of *Borrelia burgdorferi*, the Lyme disease agent. *Mol Microbiol* **8**: 967–980.
- Casjens, S., Delange, M., Ley, 3rd, H.L., Rosa, P., and Huang, W.M. (1995) Linear chromosomes of Lyme disease agent spirochetes: genetic diversity and conservation of gene order. *J Bacteriol* **177**: 2769–2780.
- Casjens, S., van Vugt, R., Tilly, K., Rosa, P.A., and Stevenson, B. (1997a) Homology throughout the multiple 32-kilobase circular plasmids present in Lyme disease spirochetes. *J Bacteriol* **179**: 217–227.
- Casjens, S., Murphy, M., DeLange, M., Sampson, L., van Vugt, R., and Huang, W.M. (1997b) Telomeres of the linear chromosomes of Lyme disease spirochaetes: nucleotide sequence and possible exchange with linear plasmid telomeres. *Mol Microbiol* **26**: 581–596.
- Champion, C.I., Blanco, D.R., Skare, J.T., Haake, D.A., Giladi, M., Foley, D., et al. (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.
- Cheetham, B.F., and Katz, M.E. (1995) A role for bacteriophages in the evolution and transfer of bacterial virulence determinants. *Mol Microbiol* **18**: 201–208.

- Claros, M.G., and von Heijne, G. (1994) TOPPED II: an improved software for membrane protein structure predictions. *Comput Appl Biosci* **10**: 685–686.
- Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., *et al.* (1998) Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* **393**: 537–544.
- Cornell, K.A., and Riscoe, M.K. (1998) Cloning and expression of *Escherichia coli* 5'-methylthioadenosine/S-adenosylhomocysteine nucleosidase: identification of the *pfs* gene product. *Biochim Biophys Acta* **1396**: 8–14.
- Davidson, B.E., MacDougall, J., and 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.
- Donadio, S., and Staver, M.J. (1993) IS1136, an insertion element in the erythromycin gene cluster of *Saccharopolyspora erythraea*. *Gene* **126**: 147–151.
- Douglas, S.E. (1994) DNA STRIDER. A Macintosh program for handling protein and nucleic acid sequences. *Methods Mol Biol* **25**: 181–194.
- Dunn, J.J., Buchstein, S.R., Butler, L.L., Fisenne, S., Polin, D.S., Lade, B.N., *et al.* (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.J., Wilske, B., Preac-Mursic, V., Dattwyler, R.J., *et al.* (1993) *Borrelia burgdorferi* is clonal: implications for taxonomy and vaccine development. *Proc Natl Acad Sci USA* **90**: 10163–10167.
- Eddy, S.R. (1998) Profile hidden Markov models. *Bioinformatics* **14**: 755–763.
- El Hage, N., Lieto, L.D. and Stevenson, B. (1999) Stability of *erp* loci during *Borrelia burgdorferi* infection: recombination is not required for chronic infection of immunocompetent mice. *Infect Immun* **67**: 3146–3150.
- Feng, S., Das, S., Barthold, S.W., and Fikrig, E. (1996) Characterization of two genes, p11 and p5, on the *Borrelia burgdorferi* 49-kilo base linear plasmid. *Biochim Biophys Acta* **1307**: 270–272.
- Feng, S., Hodzic, E., Stevenson, B., and Barthold, S.W. (1998) Humoral immunity to *Borrelia burgdorferi* N40 decorin binding proteins during infection of laboratory mice. *Infect Immun* **66**: 2827–2835.
- Ferdows, M.S., and 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.
- Ferdows, M.S., Serwer, P., Griess, G.A., Norris, S.J., and Barbour, A.G. (1996) Conversion of a linear to a circular plasmid in the relapsing fever agent *Borrelia hermsii*. *J Bacteriol* **178**: 793–800.
- Fikrig, E., Liu, B., Fu, L.L., Das, S., Smallwood, J.I., Flavell, R.A., *et al.* (1995a) An *ospA* frame shift, identified from DNA in Lyme arthritis synovial fluid, results in an outer surface protein A that does not bind protective antibodies. *J Immunol* **155**: 5700–5704.
- Fikrig, E., Tao, H., Barthold, S.W., and Flavell, R.A. (1995b) Selection of variant *Borrelia burgdorferi* isolates from mice immunized with outer surface protein A or B. *Infect Immun* **63**: 1658–1662.
- Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A., Kirkness, E.F., Kerlavage, A.R., *et al.* (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269**: 496–512.
- Fraser, C.M., Gocayne, J.D., White, O., Adams, M.D., Clayton, R.A., Fleischmann, R.D., *et al.* (1995) The minimal gene complement of *Mycoplasma genitalium*. *Science* **270**: 397–403.
- Fraser, C.M., Casjens, S., Huang, W.M., Sutton, G.G., Clayton, R., Lathigra, R., *et al.* (1997) Genomic sequence of a Lyme disease spirochaete, *Borrelia burgdorferi*. *Nature* **390**: 580–586.
- Fraser, C.M., Norris, S.J., Weinstock, G.M., White, O., Sutton, G.G., Dodson, R., *et al.* (1998) Complete genome sequence of *Treponema pallidum*, the syphilis spirochete. *Science* **281**: 375–388.
- Fsihi, H., De Rossi, E., Salazar, L., Cantoni, R., Labo, M., Riccardi, G., *et al.* (1996) Gene arrangement and organization in a approximately 76 kb fragment encompassing the *oriC* region of the chromosome of *Mycobacterium leprae*. *Microbiology* **142**: 3147–3161.
- Fukunaga, M., and Hamase, A. (1995) Outer surface protein C gene sequence analysis of *Borrelia burgdorferi* sensu lato isolates from Japan. *J Clin Microbiol* **33**: 2415–2420.
- Fukunaga, M., Hamase, A., Okada, K., and Nakao, M. (1996) *Borrelia tanukii* sp. nov. & *Borrelia turdae* sp. nov. found from ixodid ticks in Japan: rapid species identification by 16S rRNA gene-targeted PCR analysis. *Microbiol Immunol* **40**: 877–881.
- Gesteland, R.F., and Atkins, J.F. (1996) Recoding: dynamic reprogramming of translation. *Annu Rev Biochem* **65**: 741–768.
- Gilmore, Jr, R.D., and Mbow, M.L. (1998) A monoclonal antibody generated by antigen inoculation via tick bite is reactive to the *Borrelia burgdorferi* Rev protein, a member of the 2.9 gene family locus. *Infect Immun* **66**: 980–986.
- Gilmore, Jr, R.D., Kappel, K.J., and Johnson, B.J. (1997) Molecular characterization of a 35-kilodalton protein of *Borrelia burgdorferi*, an antigen of diagnostic importance in early Lyme disease. *J Clin Microbiol* **35**: 86–91.
- Guina, T., and Oliver, D.B. (1997) Cloning and analysis of a *Borrelia burgdorferi* membrane-interactive protein exhibiting haemolytic activity. *Mol Microbiol* **24**: 1201–1213.
- Gulig, P.A., Caldwell, A.L., and Chiodo, V.A. (1992) Identification, genetic analysis and DNA sequence of a 7.8 kb virulence region of the *Salmonella typhimurium* virulence plasmid. *Mol Microbiol* **6**: 1395–1411.
- Guttman, D.S., Wang, P.W., Wang, I.N., Bosler, E.M., Luft, B.J., and Dykhuizen, D.E. (1996) Multiple infections of *Ixodes scapularis* ticks by *Borrelia burgdorferi* as revealed by single-strand conformation polymorphism analysis. *J Clin Microbiol* **34**: 652–656.
- Hall, B.G., Yokoyama, S., and Calhoun, D.H. (1983) Role of cryptic genes in microbial evolution. *Mol Biol Evol* **1**: 109–124.
- Hayes, S.F., Burgdorfer, W., and Barbour, A.G. (1983) Bacteriophage in the *Ixodes dammini* spirochete, etiological agent of Lyme disease. *J Bacteriol* **154**: 1436–1439.
- Hinnebusch, J., and Barbour, A.G. (1991) Linear plasmids of *Borrelia burgdorferi* have a telomeric structure and sequence similar to those of a eukaryotic virus. *J Bacteriol* **173**: 7233–7239.

- Hinnebusch, J., and Barbour, A.G. (1992) Linear- and circular-plasmid copy numbers in *Borrelia burgdorferi*. *J Bacteriol* **174**: 5251–5257.
- Hinnebusch, J., Bergstrom, S., and Barbour, A.G. (1990) Cloning and sequence analysis of linear plasmid telomeres of the bacterium *Borrelia burgdorferi*. *Mol Microbiol* **4**: 811–820.
- Hughes, C.A., Kodner, C.B., and Johnson, R.C. (1992) DNA analysis of *Borrelia burgdorferi* NCH-1, the first north-central U.S. human Lyme disease isolate. *J Clin Microbiol* **30**: 698–703.
- Hyde, F.W., and Johnson, R.C. (1988) Characterization of a circular plasmid from *Borrelia burgdorferi*, etiologic agent of Lyme disease. *J Clin Microbiol* **26**: 2203–2205.
- Jauris-Heipke, S., Liegl, G., Preac-Mursic, V., Rossler, D., Schwab, E., Soutschek, E., et al. (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.
- Johnson, R.C., Hyde, F.W., and Rumpel, C.M. (1984) Taxonomy of the Lyme disease spirochetes. *Yale J Biol Med* **57**: 529–537.
- Kawabata, H., Myouga, F., Inagaki, Y., Murai, N., and Watanabe, H. (1998) Genetic and immunological analyses of Vls (VMP-like sequences) of *Borrelia burgdorferi*. *Microb Pathog* **24**: 155–166.
- Kitten, T., and Barbour, A.G. (1992) The relapsing fever agent *Borrelia hermsii* has multiple copies of its chromosome and linear plasmids. *Genetics* **132**: 311–324.
- Koomey, M. (1994) Mechanisms of pilus antigenic variation in *Neisseria gonorrhoeae*. In *Molecular Genetics of Bacterial Pathogenesis*. Miller, V., Kaper, J., Prtnoy, D., and Isberg, R. (eds). Washington, DC: American Society for Microbiology.
- Kornacki, J.A., and Oliver, D.B. (1998) Lyme disease-causing *Borrelia* species encode multiple lipoproteins homologous to peptide-binding proteins of ABC-type transporters. *Infect Immun* **66**: 4115–4122.
- Krause, M., Harwood, J., Fierer, J., and Guiney, D. (1991) Genetic analysis of homology between the virulence plasmids of *Salmonella dublin* and *Yersinia pseudotuberculosis*. *Infect Immun* **59**: 1860–1863.
- Lai, C.Y., Baumann, P., and Moran, N. (1996) The endosymbiont (*Buchnera* sp.) of the aphid *Diuraphis noxia* contains plasmids consisting of *trpEGG* and tandem repeats of *trpEG* pseudogenes. *Appl Environ Microbiol* **62**: 332–339.
- Lam, T.T., Nguyen, T.P., Montgomery, R.R., Kantor, F.S., Fikrig, E., and Flavell, R.A. (1994) Outer surface proteins E and F of *Borrelia burgdorferi*, the agent of Lyme disease. *Infect Immun* **62**: 290–298.
- Lawrence, J.G., and Ochman, H. (1997) Amelioration of bacterial genomes: rates of change and exchange. *J Mol Evol* **44**: 383–397.
- Le Fleche, A., Postic, D., Girardet, K., Peter, O., and Baranton, G. (1997) Characterization of *Borrelia lusitaniae* sp. nov. by 16S ribosomal DNA sequence analysis. *Int J Syst Bacteriol* **47**: 921–925.
- Lively, I., Gibbs, C.P., Schuster, R., and Dorner, F. (1995) Evidence for lateral transfer and recombination in OspC variation in Lyme disease *Borrelia*. *Mol Microbiol* **18**: 257–269.
- McLean, M.J., Wolfe, K.H. and Devine, K.M. (1998) Base composition skews, replication orientation, and gene orientation in 12 prokaryote genomes. *J Mol Evol* **47**: 691–696.
- Marconi, R.T., Konkel, M.E., and Garon, C.F. (1993a) Variability of *osp* genes and gene products among species of Lyme disease spirochetes. *Infect Immun* **61**: 2611–2617.
- Marconi, R.T., Samuels, D.S., Schwan, T.G., and Garon, C.F. (1993b) Identification of a protein in several *Borrelia* species which is related to OspC of the Lyme disease spirochetes. *J Clin Microbiol* **31**: 2577–2583.
- Marconi, R.T., Samuels, D.S., Landry, R.K., and Garon, C.F. (1994) Analysis of the distribution and molecular heterogeneity of the *ospD* gene among the Lyme disease spirochetes: evidence for lateral gene exchange. *J Bacteriol* **176**: 4572–4582.
- Marconi, R.T., Casjens, S., Munderloh, U.G., and Samuels, D.S. (1996a) Analysis of linear plasmid dimers in *Borrelia burgdorferi* sensu lato isolates: implications concerning the potential mechanism of linear plasmid replication. *J Bacteriol* **178**: 3357–3361.
- Marconi, R.T., Sung, S.Y., Hughes, C.A., and Carlyon, J.A. (1996b) 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.
- Margolis, N., Hogan, D., Tilly, K., and Rosa, P.A. (1994) Plasmid location of *Borrelia* purine biosynthesis gene homologs. *J Bacteriol* **176**: 6427–6432.
- Mathiesen, D.A., Oliver, Jr, J.H., Kolbert, C.P., Tullson, E.D., Johnson, B.J., Campbell, G.L., et al. (1997) Genetic heterogeneity of *Borrelia burgdorferi* in the United States. *J Infect Dis* **175**: 98–107.
- Misonne, M.C., Schuttler, M., Dernelle, J.M., De Kesel, M., and Hoet, P.P. (1997) Cloning and sequencing of a species-specific nucleotide fragment of *Borrelia burgdorferi* sensu stricto, which is repeated in several plasmids of the species. *FEMS Microbiol Lett* **150**: 157–164.
- Morris, D.D., Reeves, R.A., Gibbs, M.D., Saul, D.J., and Bergquist, P.L. (1995) Correction of the beta-mannanase domain of the *celC* pseudogene from *Caldocellulosiruptor saccharolyticus* and activity of the gene product on kraft pulp. *Appl Environ Microbiol* **61**: 2262–2269.
- Moxon, E.R., Rainey, P.B., Nowak, M.A., and Lenski, R.E. (1994) Adaptive evolution of highly mutable loci in pathogenic bacteria. *Curr Biol* **4**: 24–33.
- Munderloh, U.G., Park, Y.J., Dioh, J.M., Fallon, A.M., and Kurtti, T.J. (1993) Plasmid modifications in a tick-borne pathogen, *Borrelia burgdorferi*, cocultured with tick cells. *Insect Mol Biol* **1**: 195–203.
- Murai, N., Kamata, H., Nagashima, Y., Yagisawa, H., and Hirata, H. (1995) A novel insertion sequence (IS)-like element of the thermophilic bacterium PS3 promotes expression of the alanine carrier protein-encoding gene. *Gene* **163**: 103–107.
- Neubert, U., Schaller, M., Januschke, E., Stolz, W., and Schmieger, H. (1993) Bacteriophages induced by ciprofloxacin in a *Borrelia burgdorferi* skin isolate. *Zentralbl Bacteriol* **279**: 307–315.
- Norris, S.J., Carter, C.J., Howell, J.K., and Barbour, A.G. (1992) Low-passage-associated proteins of *Borrelia burgdorferi*

- B31: characterization and molecular cloning of OspD, a surface-exposed, plasmid-encoded lipoprotein. *Infect Immun* **60**: 4662–4672.
- Norris, S.J., Howell, J.K., Garza, S.A., Ferdows, M.S., and Barbour, A.G. (1995) High- and low-infectivity phenotypes of clonal populations of in vitro-cultured *Borrelia burgdorferi*. *Infect Immun* **63**: 2206–2212.
- Ojaimi, C., Davidson, B.E., Saint Girons, I., and Old, I.G. (1994) Conservation of gene arrangement and an unusual organization of rRNA genes in the linear chromosomes of the Lyme disease spirochaetes *Borrelia burgdorferi*, *B. garinii* and *B. afzelii*. *Microbiology* **140**: 2931–2940.
- Pearson, W.R. (1990) Rapid and sensitive sequence comparison with FASTP and FASTA. *Methods Enzymol* **183**: 63–98.
- Persing, D.H., Mathiesen, D., Podzorski, D., and Barthold, S.W. (1994) Genetic stability of *Borrelia burgdorferi* recovered from chronically infected immunocompetent mice. *Infect Immun* **62**: 3521–3527.
- Plasterk, R.H., Simon, M.I., and Barbour, A.G. (1985) Transposition of structural genes to an expression sequence on a linear plasmid causes antigenic variation in the bacterium *Borrelia hermsii*. *Nature* **318**: 257–263.
- Porcella, S.F., Popova, T.G., Akins, D.R., Li, M., Radolf, J.D., and Norgard, M.V. (1996) *Borrelia burgdorferi* supercoiled plasmids encode multicopy tandem open reading frames and a lipoprotein gene family. *J Bacteriol* **178**: 3293–3307.
- Postic, D., Ras, N.M., Lane, R.S., Henderson, M., and Baranton, G. (1998) Expanded diversity among Californian *borrelia* isolates and description of *Borrelia bissettii* sp. nov. (formerly *Borrelia* group DN127). *J Clin Microbiol* **36**: 3497–3504.
- Probert, W., and Johnson, B. (1998) Identification of a 47 kd fibronectin-binding protein expressed by *Borrelia burgdorferi* isolate B31. *Mol Microbiol* **30**: 1003–1015.
- Restrepo, B.I., and Barbour, A.G. (1994) Antigen diversity in the bacterium *B. hermsii* through 'somatic' mutations in rearranged *vmp* genes. *Cell* **78**: 867–876.
- Rosa, P.A. (1997) Microbiology of *Borrelia burgdorferi*. *Semin Neurol* **17**: 5–10.
- Rosa, P.A., Schwan, T., and Hogan, D. (1992) Recombination between genes encoding major outer surface proteins A and B of *Borrelia burgdorferi*. *Mol Microbiol* **6**: 3031–3040.
- Ryan, J.R., Levine, J.F., Apperson, C.S., Lubke, L., Wirtz, R.A., Spears, P.A., et al. (1998) An experimental chain of infection reveals that distinct *Borrelia burgdorferi* populations are selected in arthropod and mammalian hosts. *Mol Microbiol* **30**: 365–379.
- Sadziene, A., Rosa, P.A., Thompson, P.A., Hogan, D.M., and Barbour, A.G. (1992) Antibody-resistant mutants of *Borrelia burgdorferi*: in vitro selection and characterization. *J Exp Med* **176**: 799–809.
- Sadziene, A., Wilske, B., Ferdows, M.S., and Barbour, A.G. (1993a) The cryptic *ospC* gene of *Borrelia burgdorferi* B31 is located on a circular plasmid. *Infect Immun* **61**: 2192–2195.
- Sadziene, A., Barbour, A.G., Rosa, P.A., and Thomas, D.D. (1993b) An OspB mutant of *Borrelia burgdorferi* has reduced invasiveness in vitro and reduced infectivity in vivo. *Infect Immun* **61**: 3590–3596.
- Salzberg, S.L., Delcher, A.L., Kasif, S., and White, O. (1998) Microbial gene identification using interpolated Markov models. *Nucleic Acids Res* **26**: 544–548.
- Samuels, D.S., Marconi, R.T., and Garon, C.F. (1993) Variation in the size of the *ospA*-containing linear plasmid, but not the linear chromosome, among the three *Borrelia* species associated with Lyme disease. *J Gen Microbiol* **139**: 2445–2449.
- Saunders, N.J., Peden, J.F., Hood, D.W., and Moxon, E.R. (1998) Simple sequence repeats in the *Helicobacter pylori* genome. *Mol Microbiol* **27**: 1091–1098.
- Schaller, M., and Neubert, U. (1994) Bacteriophages and ultrastructural alterations of *Borrelia burgdorferi* induced by ciprofloxacin. *J Spirochetal Tick Borne Dis* **1**: 37–40.
- Schwan, T.G., Burgdorfer, W., and Garon, C.F. (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.
- Simpson, W.J., Garon, C.F., and 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., and Schwan, T.G. (1990b) *Borrelia burgdorferi* contains repeated DNA sequences that are species specific and plasmid associated. *Infect Immun* **58**: 847–853.
- Skamrov, A., Goldman, M., Klasova, J., and Beabealashvili, R. (1995) *Mycoplasma gallisepticum* 16S rRNA genes. *FEMS Microbiol Lett* **128**: 321–325.
- Skare, J.T., Champion, C.I., Mirzabekov, T.A., Shang, E.S., Blanco, D.R., Erdjument-Bromage, H., et al. (1996) Porin activity of the native and recombinant outer membrane protein Oms28 of *Borrelia burgdorferi*. *J Bacteriol* **178**: 4909–4918.
- Smith, D., Richterich, P., Rubenfield, M., Rice, P., Butler, C., Lee, H., et al. (1997) Multiplex sequencing of 1.5 Mb of the *Mycobacterium leprae* genome. *Genome Res* **7**: 802–819.
- Stalhammar-Carlemalm, M., Jenny, E., Gern, L., Aeschlimann, A., and Meyer, J. (1990) Plasmid analysis and restriction fragment length polymorphisms of chromosomal DNA allow a distinction between *Borrelia burgdorferi* strains. *Zentralbl Bakteriol* **274**: 28–39.
- Stanley, E., Fitzgerald, G.F., Le Marrec, C., Fayard, B., and van Sinderen, D. (1997) Sequence analysis and characterization of phiO1205, a temperate bacteriophage infecting *Streptococcus thermophilus* CNRZ1205. *Microbiology* **143**: 3417–3429.
- Steere, A.C., Grodzicki, R.L., Kornblatt, A.N., Craft, J.E., Barbour, A.G., Burgdorfer, W., et al. (1983) The spirochetal etiology of Lyme disease. *N Engl J Med* **308**: 733–740.
- Stevenson, B., and 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., Tilly, K., and 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., et al. (1997) Characterization of cp18, a

- naturally truncated member of the cp32 family of *Borrelia burgdorferi* plasmids. *J Bacteriol* **179**: 4285–4291.
- Stevenson, B., Casjens, S., and Rosa, P. (1998a) Evidence of past recombination events among the genes encoding the Erp antigens of *Borrelia burgdorferi*. *Microbiology* **144**: 1869–1879.
- Stevenson, B., Bono, J.L., Schwan, T.G., and Rosa, P. (1998b) *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.
- Sueoka, N. (1993) Directional mutation pressure, mutator mutations, and dynamics of molecular evolution. *J Mol Evol* **37**: 137–153.
- Sutcliffe, I., and Russell, R. (1995) Lipoproteins of Gram-positive bacteria. *J Bacteriol* **177**: 1123–1128.
- 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.
- Theisen, M., Frederiksen, B., Lebech, A.M., Vuust, J., and 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.
- Tilly, K., Casjens, S., Stevenson, B., Bono, J.L., Samuels, D.S., Hogan, D., et al. (1997) The *Borrelia burgdorferi* circular plasmid cp26: conservation of plasmid structure and targeted inactivation of the *ospC* gene. *Mol Microbiol* **25**: 361–373.
- Tomb, J.F., White, O., Kerlavage, A.R., Clayton, R.A., Sutton, G.G., Fleischmann, R.D., et al. (1997) The complete genome sequence of the gastric pathogen *Helicobacter pylori*. *Nature* **388**: 539–547.
- Walker, D.H. (1998) Tick-transmitted infectious diseases in the United States. *Annu Rev Public Health* **19**: 237–269.
- Wang, G., van Dam, A.P., Le Fleche, A., Postic, D., Peter, O., Baranton, G., et al. (1997a) Genetic and phenotypic analysis of *Borrelia valaisiana* sp. nov. (*Borrelia* genomic groups VS116 and M19). *Int J Syst Bacteriol* **47**: 926–932.
- Wang, J., Masuzawa, T., Li, M., and Yanagihara, Y. (1997b) Deletion in the genes encoding outer surface proteins OspA and OspB of *Borrelia garinii* isolated from patients in Japan. *Microbiol Immunol* **41**: 673–679.
- Wang, G., van Dam, A.P., Spanjaard, L., and Dankert, J. (1998) Molecular typing of *Borrelia burgdorferi* sensu lato by randomly amplified polymorphic DNA fingerprinting analysis. *J Clin Microbiol* **36**: 768–776.
- Will, G., Jauris-Heipke, S., Schwab, E., Busch, U., Rossler, D., Soutschek, E., et al. (1995) Sequence analysis of *ospA* genes shows homogeneity within *Borrelia burgdorferi* sensu stricto and *Borrelia afzelii* strains but reveals major subgroups within the *Borrelia garinii* species. *Med Microbiol Immunol (Berl)* **184**: 73–80.
- Wilske, B., Jauris-Heipke, S., Lobentanzer, R., Pradel, I., Preac-Mursic, V., Rossler, D., et al. (1995) Phenotypic analysis of outer surface protein C (OspC) of *Borrelia burgdorferi* sensu lato by monoclonal antibodies: relationship to genospecies and OspA serotype. *J Clin Microbiol* **33**: 103–109.
- Wilske, B., Busch, U., Eiffert, H., Fingerle, V., Pfister, H.W., Rossler, D., et al. (1996) Diversity of OspA and OspC among cerebrospinal fluid isolates of *Borrelia burgdorferi* sensu lato from patients with neuroborreliosis in Germany. *Med Microbiol Immunol (Berl)* **184**: 195–201.
- Xiang, S.H., Hobbs, M., and Reeves, P.R. (1994) Molecular analysis of the *rfb* gene cluster of a group D2 *Salmonella enterica* strain: evidence for its origin from an insertion sequence-mediated recombination event between group E and D1 strains. *J Bacteriol* **176**: 4357–4365.
- Xu, Y., and Johnson, R.C. (1995) Analysis and comparison of plasmid profiles of *Borrelia burgdorferi* sensu lato strains. *J Clin Microbiol* **33**: 2679–2685.
- Xu, Y., Kodner, C., Coleman, L., and Johnson, R.C. (1996) Correlation of plasmids with infectivity of *Borrelia burgdorferi* sensu stricto type strain B31. *Infect Immun* **64**: 3870–3876.
- Zhang, J.R., and Norris, S.J. (1998) Genetic variation of the *Borrelia burgdorferi* gene *vsE* involves cassette-specific, segmental gene conversion. *Infect Immun* **66**: 3698–3704.
- Zhang, J.R., Hardham, J.M., Barbour, A.G., and Norris, S.J. (1997) Antigenic variation in Lyme disease *borreliae* by promiscuous recombination of VMP-like sequence cassettes. *Cell* **89**: 275–285.
- Zhou, X., Cahoon, M., Rosa, P., and Hedstrom, L. (1997) Expression, purification, and characterization of inosine 5'-monophosphate dehydrogenase from *Borrelia burgdorferi*. *J Biol Chem* **272**: 21977–21981.
- Zuckert, W.R., and 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.
- Zuckert, W.R., Meyer, J., and Barbour, A.G. (1999) Comparative analysis and immunological characterization of the *Borrelia* Bdr protein family. *Infect Immun* **67**: 3257–3266.