

Supplementary Material to Casjens *et al.*

Table of Contents - SUPPLEMENTARY MATERIAL for Casjens *et al.*

Figure	page
Figure S1. <i>Borrelia burgdorferi</i> plasmid PFam32 protein tree	2-3
Figure S2. Comparative maps of linear plasmids in fourteen <i>B. burgdorferi</i> isolates	4-19
Panel A, lp5	5
Panel B, lp17	6
Panel C, lp21	7
Panel D, lp25	8
Panel E, lp28-1	9
Panel F, lp28-2, lp28-7, lp28-9 and one lp28-6	10
Panel G, lp28-3	11
Panel H, lp28-4	12
Panel I, lp28-5	13
Panel J, lp28-6	14
Panel K, lp32-3	15
Panel L, lp36	16
Panel M, lp38	17
Panel N, lp54	18
Panel O, lp56	19
Figure S3. Deletions in cp32 circular plasmids	20
Figure S4. Comparative cp9 maps	21
Figure S5. All possible recombination products	22
Figure S6. Past transfer of genetic material from the <i>B. burgdorferi</i> chromosome right end to a linear plasmid	24
Table S1. Linear plasmid location of selected genes	25-26

Figure S1. *Borrelia burgdorferi* plasmid PFam32 protein tree.

PFam32 amino acid sequences were aligned and an unrooted maximum likelihood tree was created by RAxML using the PROTGAMMAWAG model (Stamatakis A, RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 2014, 30:1312-1313); bootstrap values from 100 trials are shown for the nodes that define the PFam groups. A scale bar for number of amino acid substitutions per aligned site is shown at the left. At the tip of each branch the names of the plasmids that encode the PFam32 proteins are indicated in large bold text, and *Borrelia* isolates carrying them are indicated in smaller text ("Sh" indicates strain Sh-2-82; see text). The colored boxes highlight the new plasmid types discussed in this report. A neighbor-joining Clustal X (Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* 2007, 23:2947-2948) tree gave a very similar and robust tree with identical groupings at the branch tips (not shown).

Figure S1

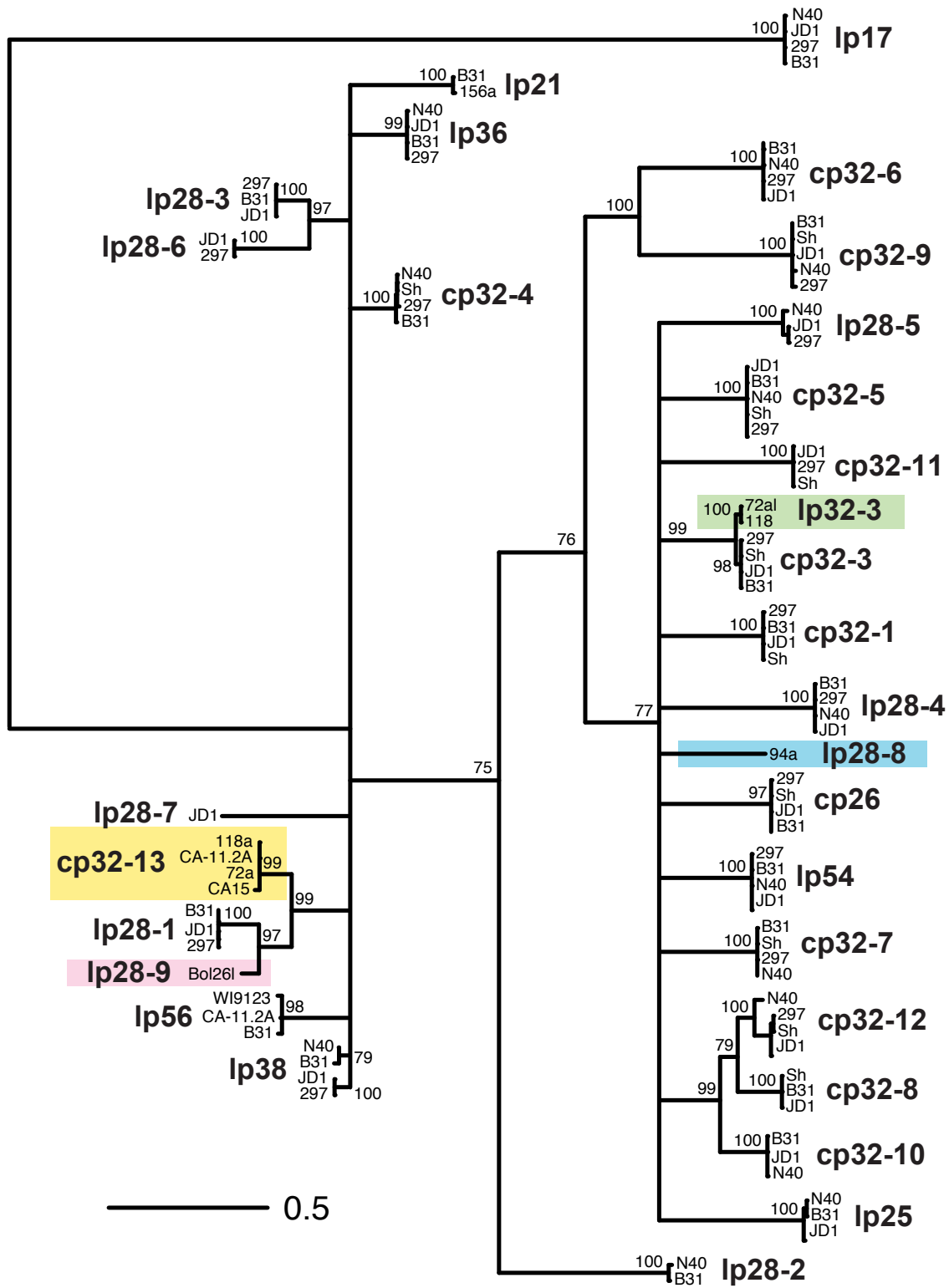


Figure S2. Comparative maps of linear plasmids in fourteen *B. burgdorferi* isolates.

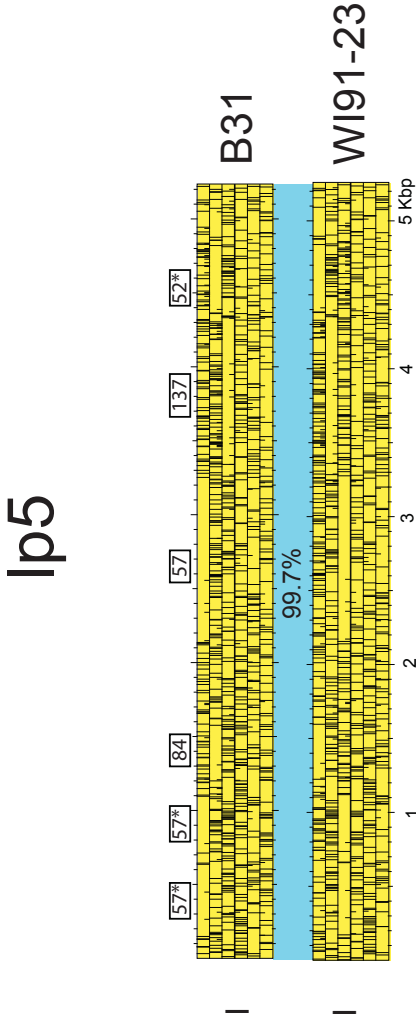
Linear plasmid reading frame maps are shown with the six possible reading frames (top three rightward frames and bottom three leftward) with stop codons marked as vertical lines that span the frame rectangle, and potential start codons are indicated by short vertical lines. The maps were created with DNA Strider (Douglas SE: DNA Strider. A Macintosh program for handling protein and nucleic acid sequences. *Methods Mol Biol* 1994, 25:181-194). Plasmid names are shown at the top of each panel and strain names at the right. Above the maps some gene names are indicated and some protein family names are shown in boxes are shown in boxes (protein families defined by Casjens SR *et al.*: Genome stability of Lyme disease spirochetes: comparative genomics of *Borrelia burgdorferi* plasmids. *PLoS One* 2012, 7:e33280). Plasmid subtype Roman numerals are indicated at the left. Blue shading between maps indicates homologous regions and percent identity is indicated for selected regions on the shading. In each panel the same shading color on the maps indicates homologous regions.

Maps of the following linear plasmids are shown in the following panels: A, lp5; B, lp17; C, lp21; D, lp25; E, lp28-1; F, lp28-2, lp28-7, lp28-9 and one lp28-6; G, lp28-3; H, lp28-4; I, lp28-5; J, lp28-6; K, lp32-3; L, lp36; M, lp38; N, lp54; O, lp56.

Panel S2B

In panel S2B numbers in blue circles with black numbers mark the eleven novel sequence junctions present in the lp17 plasmids. Red "C", "D" and "E" above the maps indicate three indel alleles, and red numbers above the maps at about 11.6 kbp of the bottom scale are the number of repeats present in this tandem repeat array.

Figure S2A



Supplementary Material to Casjens *et al.*

In panel S2B numbers in blue circles with black numbers mark the eleven novel sequence junctions present in the Ip17 plasmids. The larger genes are colored red and large pseudogenes are colored green.

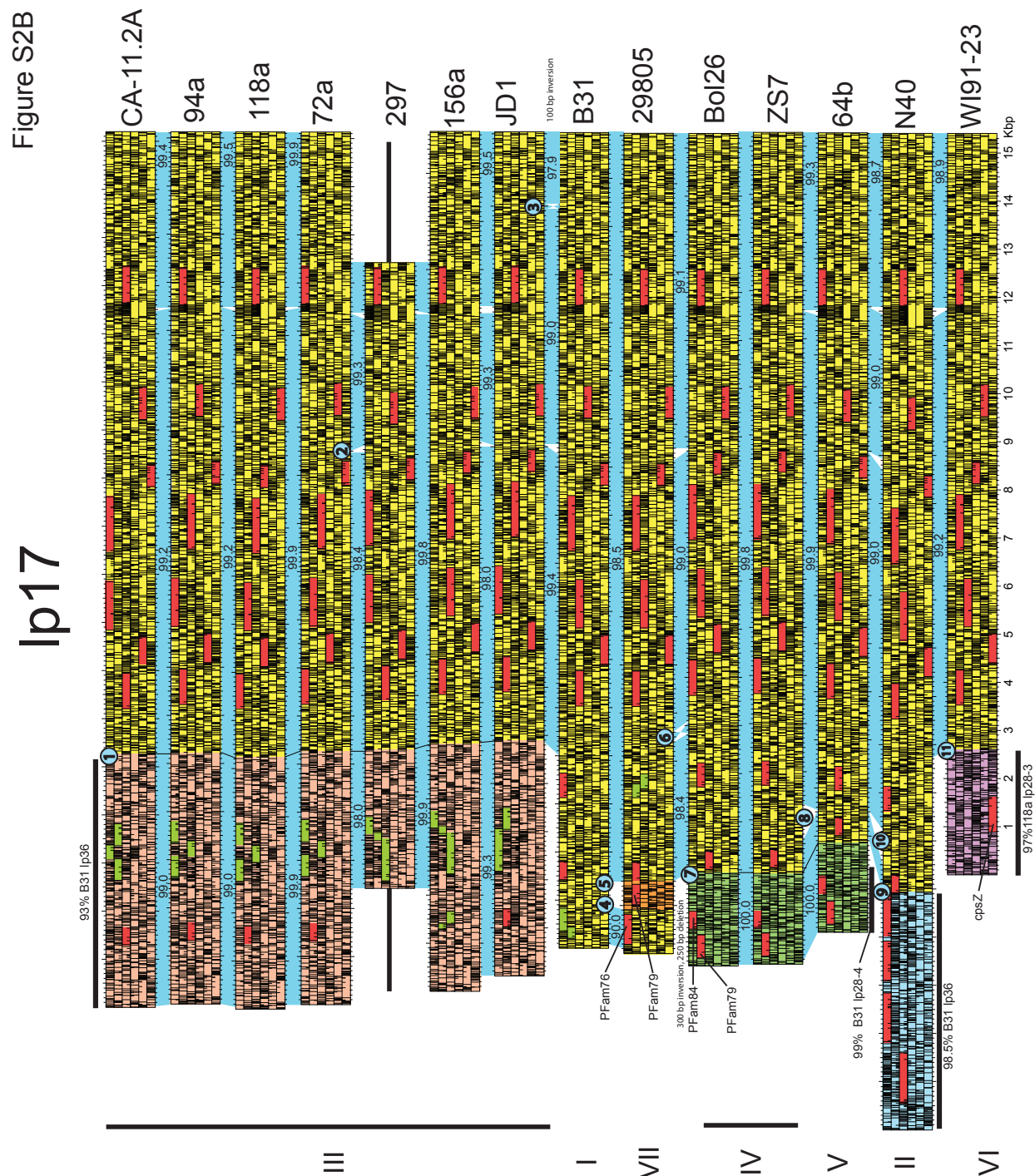


Figure S2C

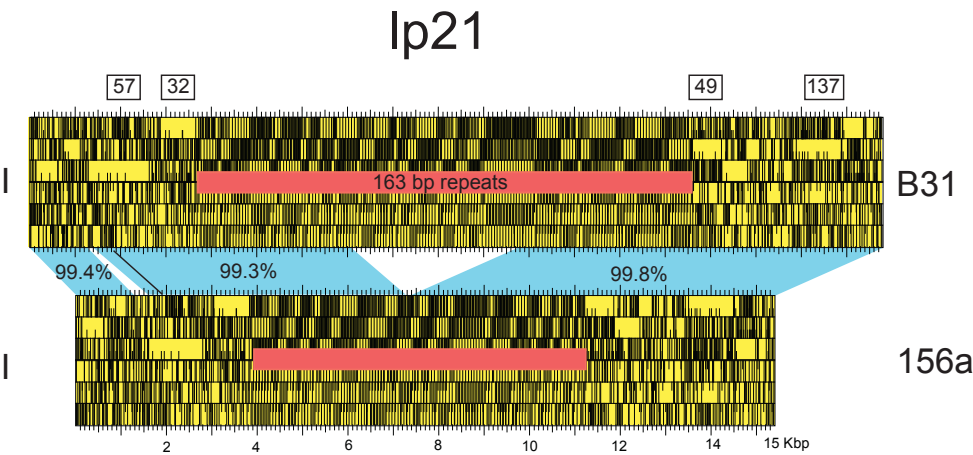


Figure S2D

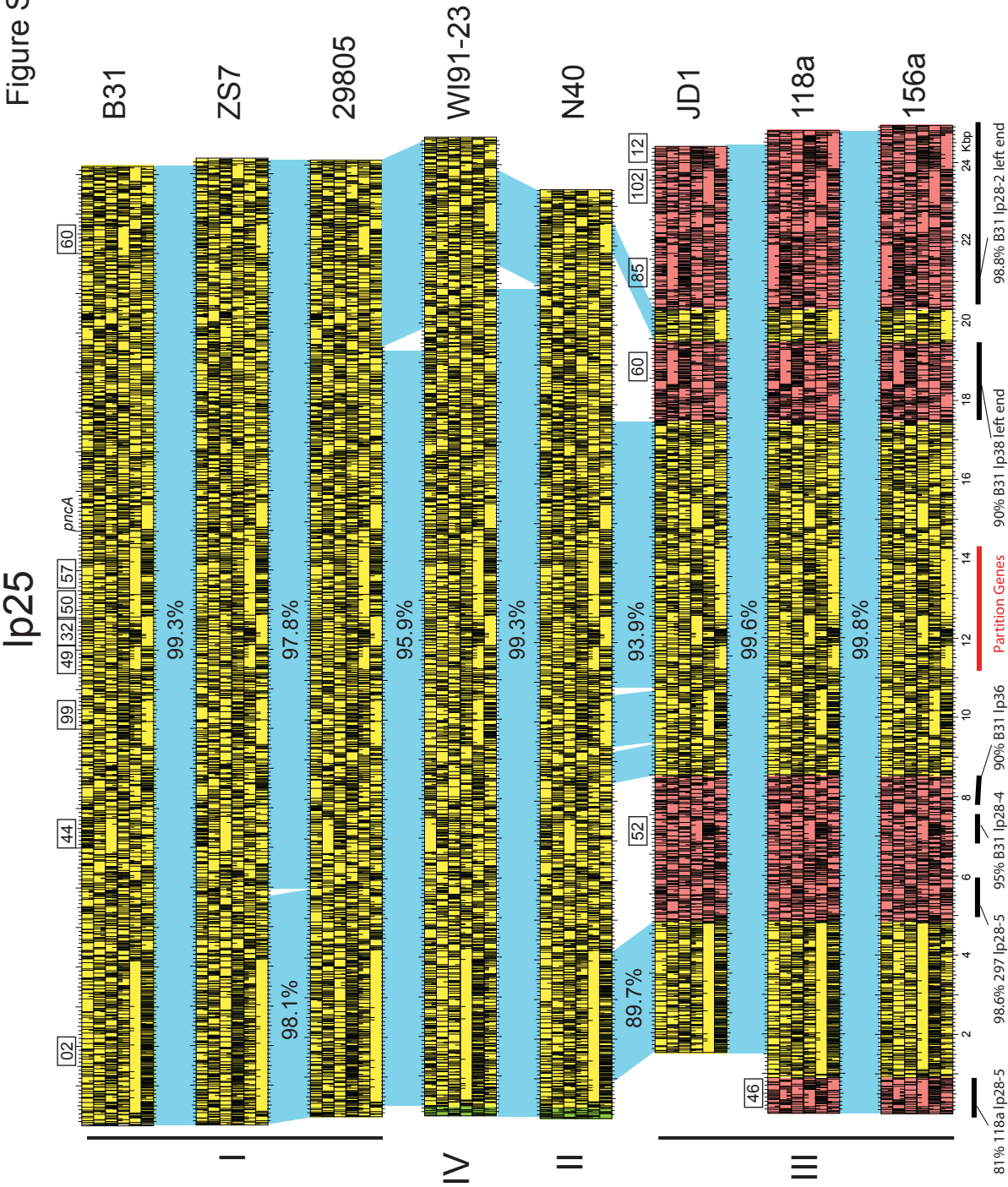
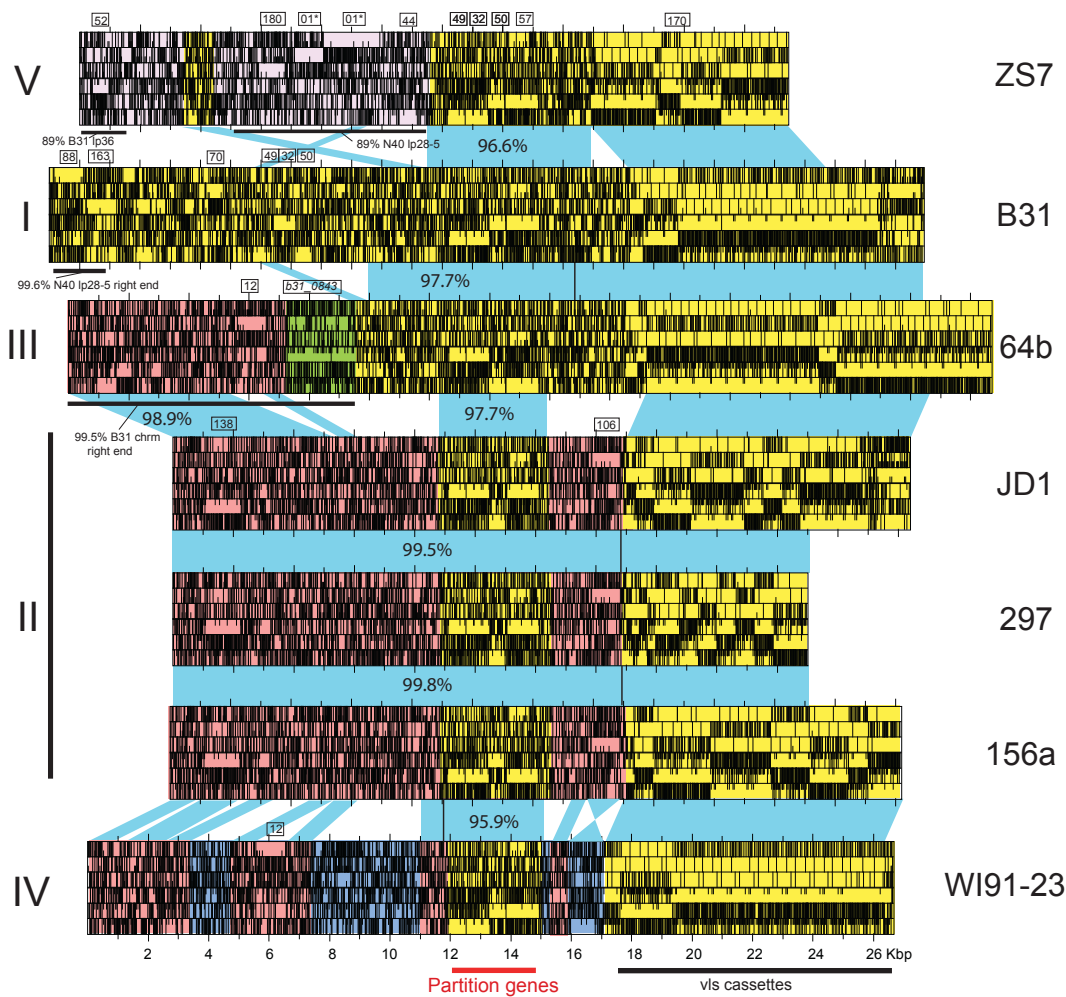


Figure S2E

lp28-1



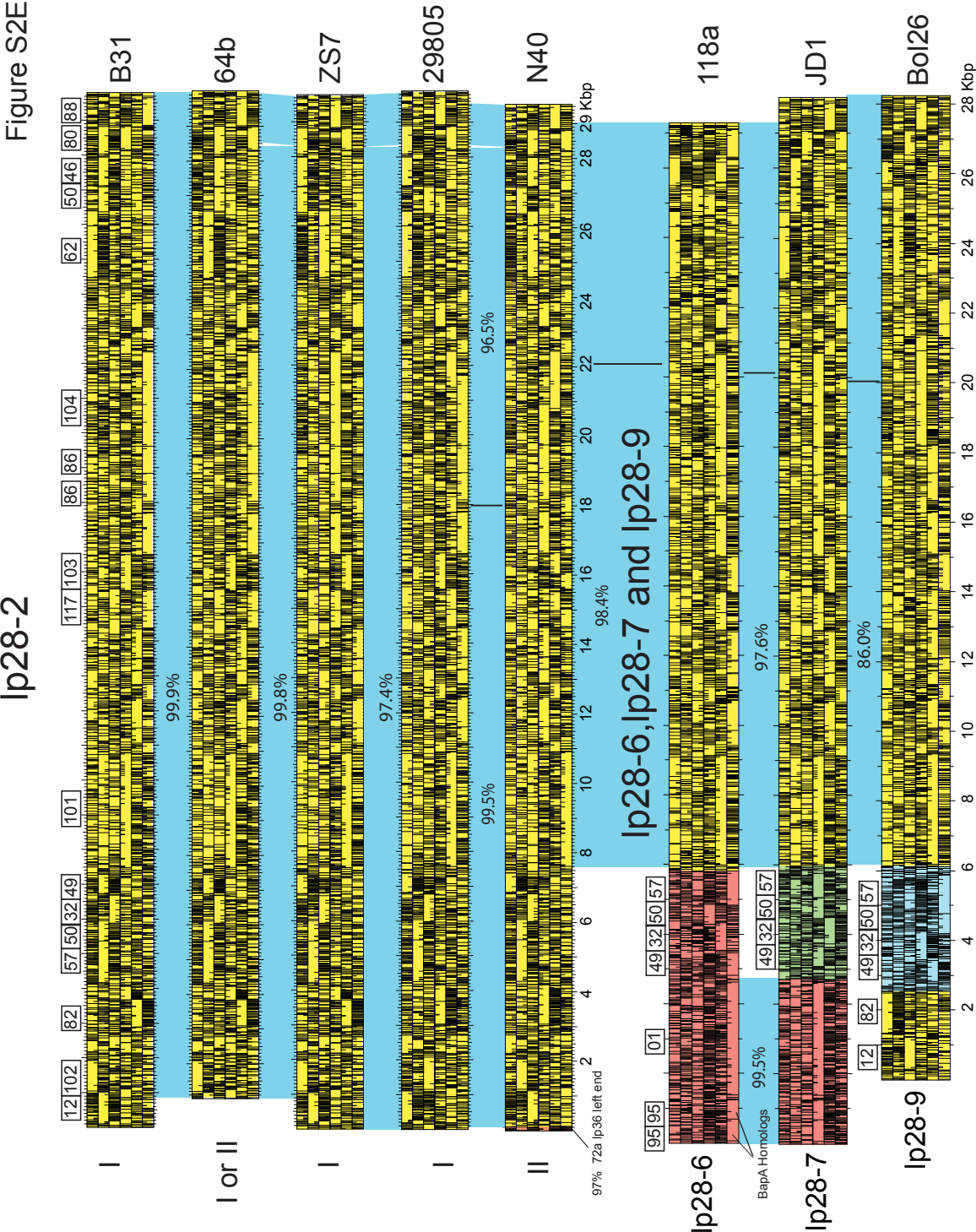
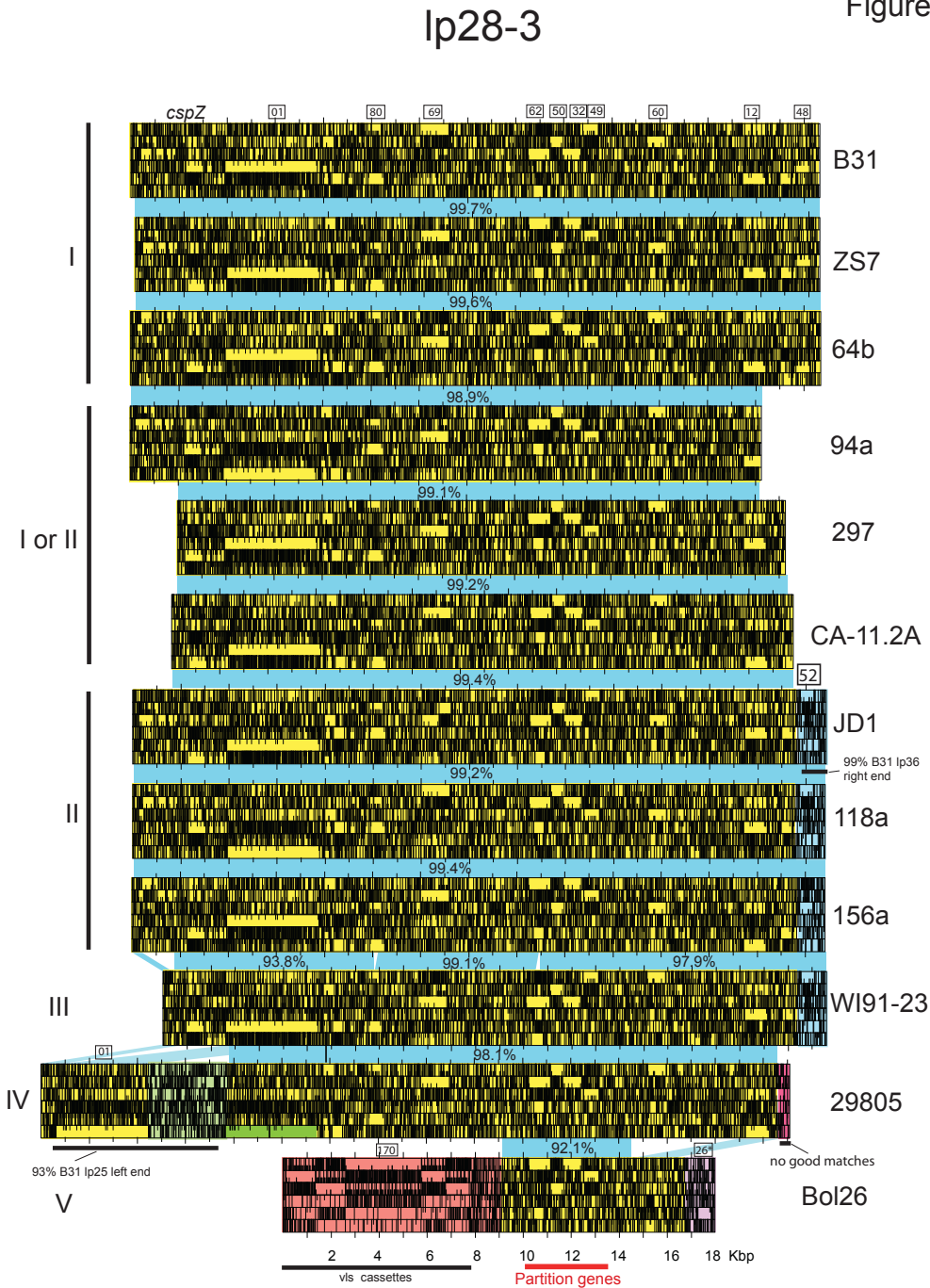
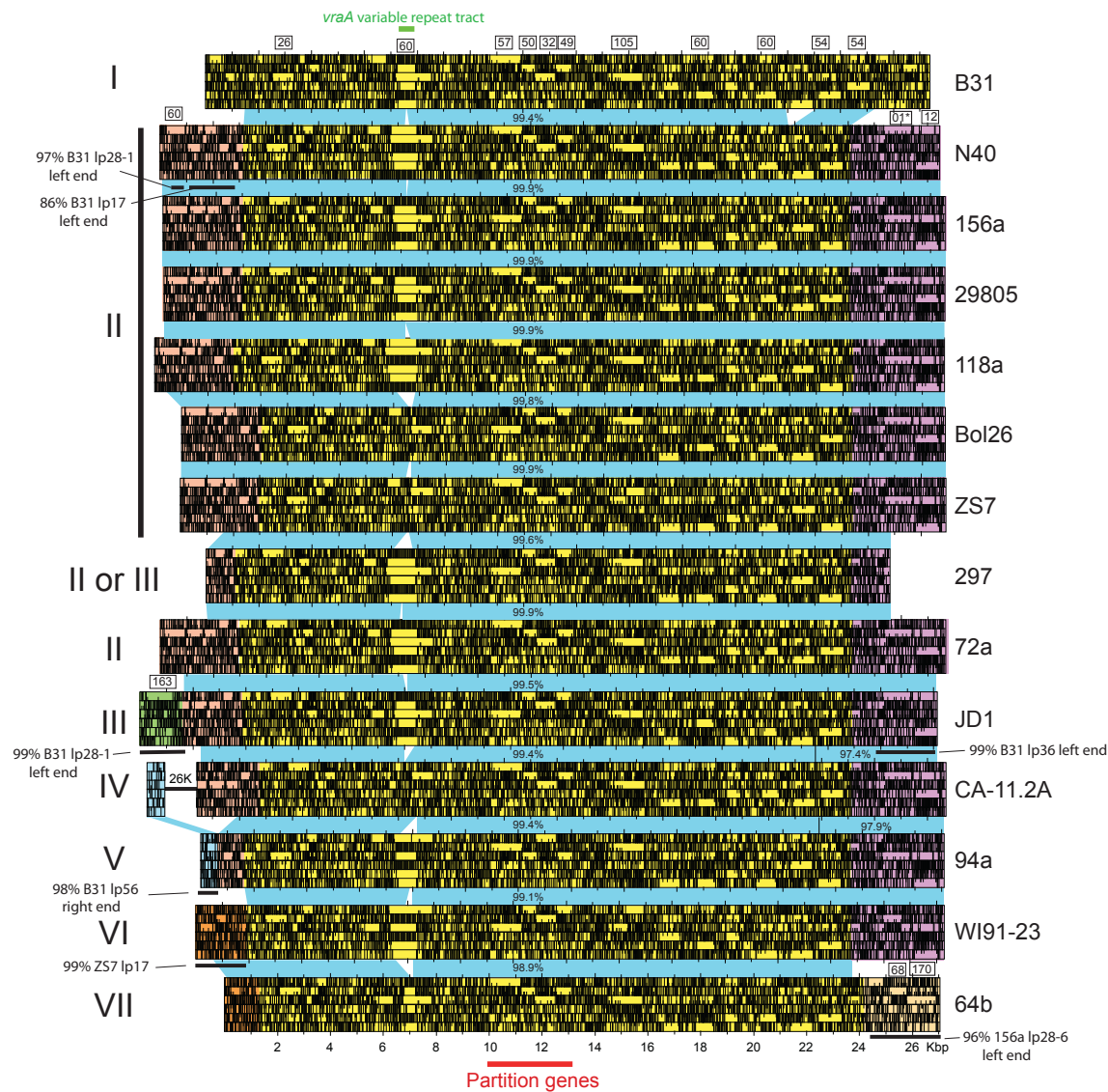


Figure S2G



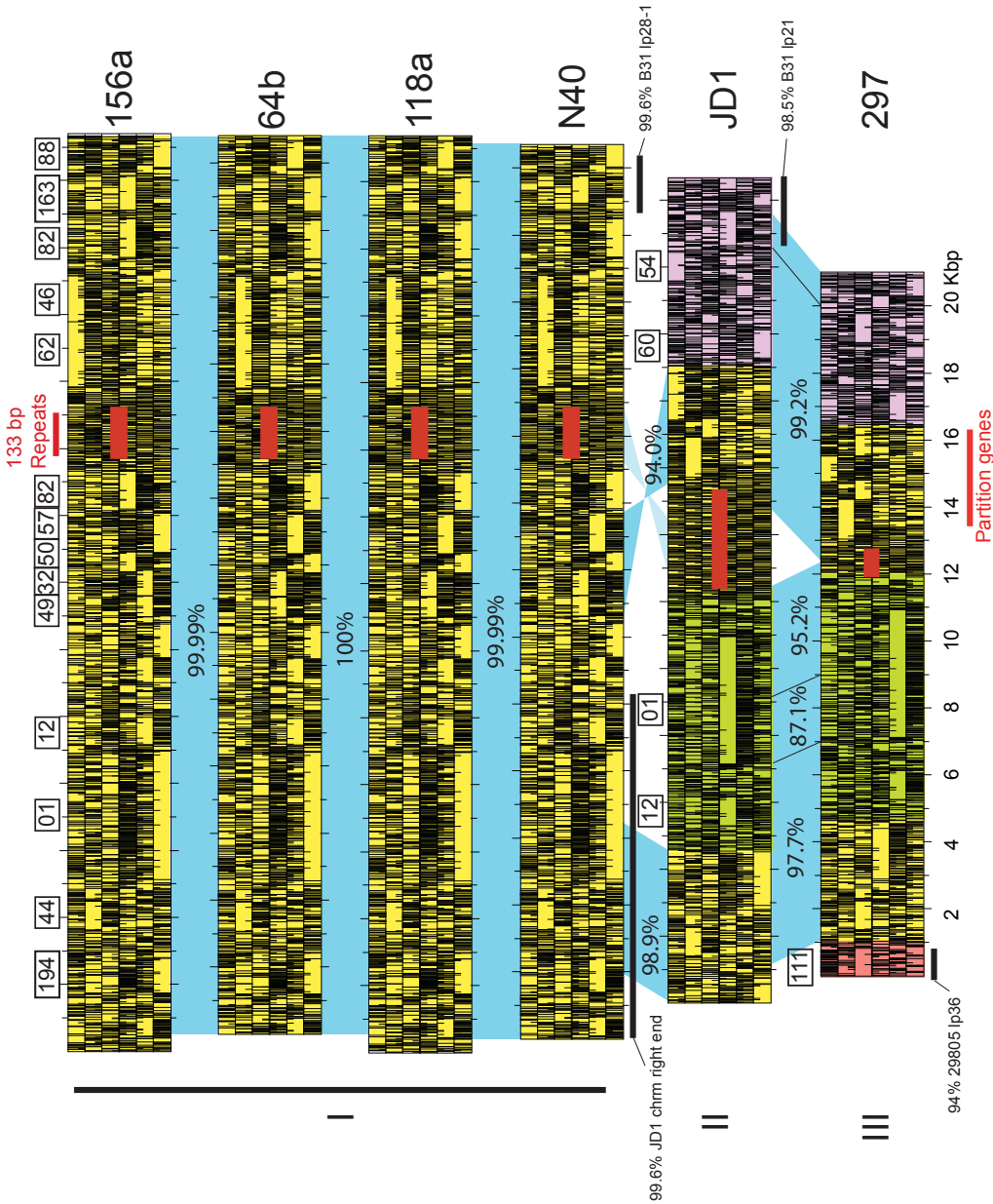
Ip28-4

Figure S2H



lp28-5

Figure S2I



Ip28-6

Figure S2J

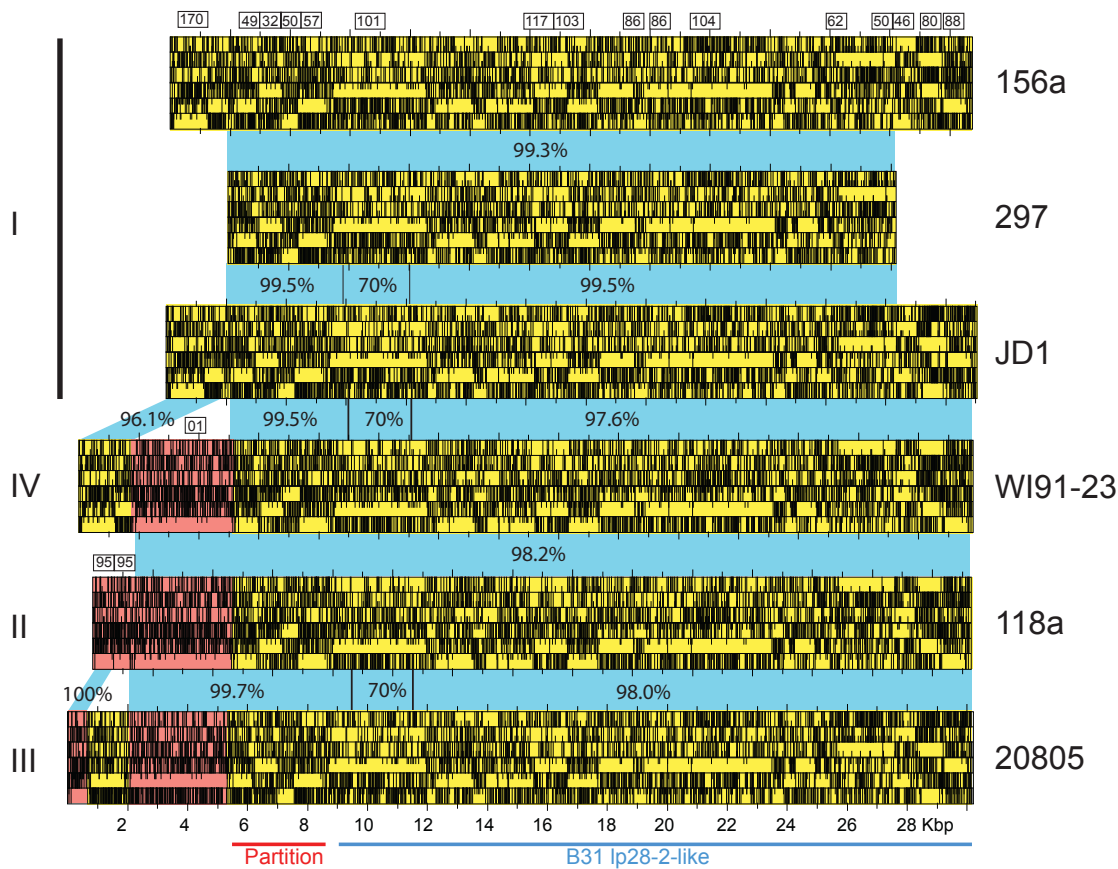
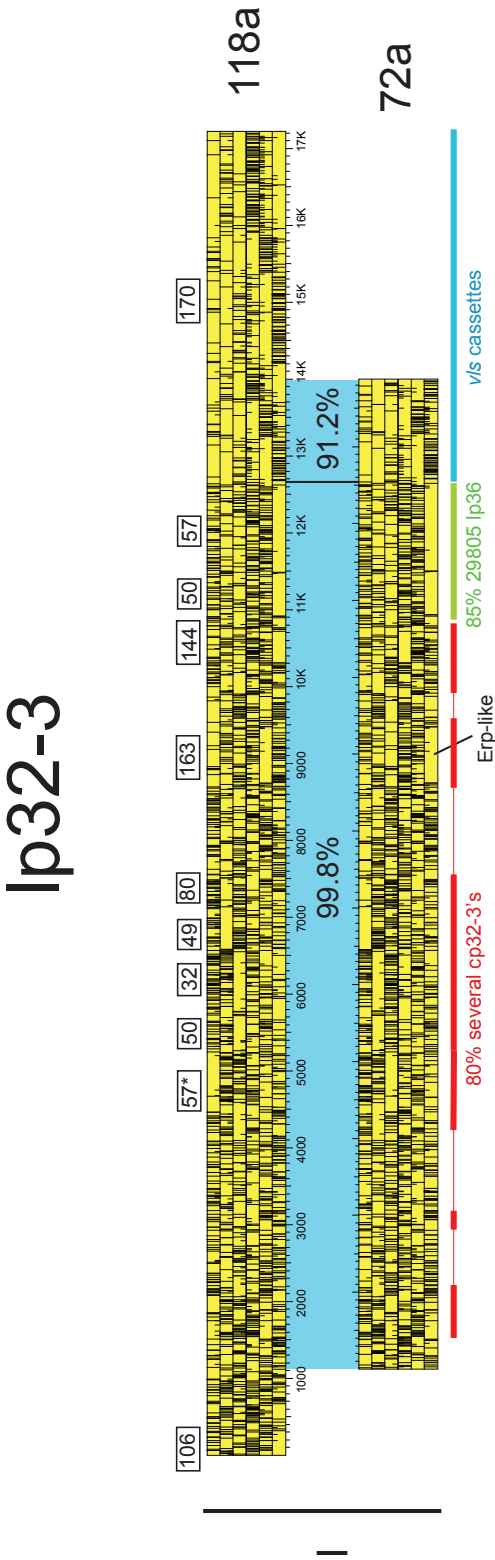


Figure S2K



Supplementary Material to Casjens *et al.*

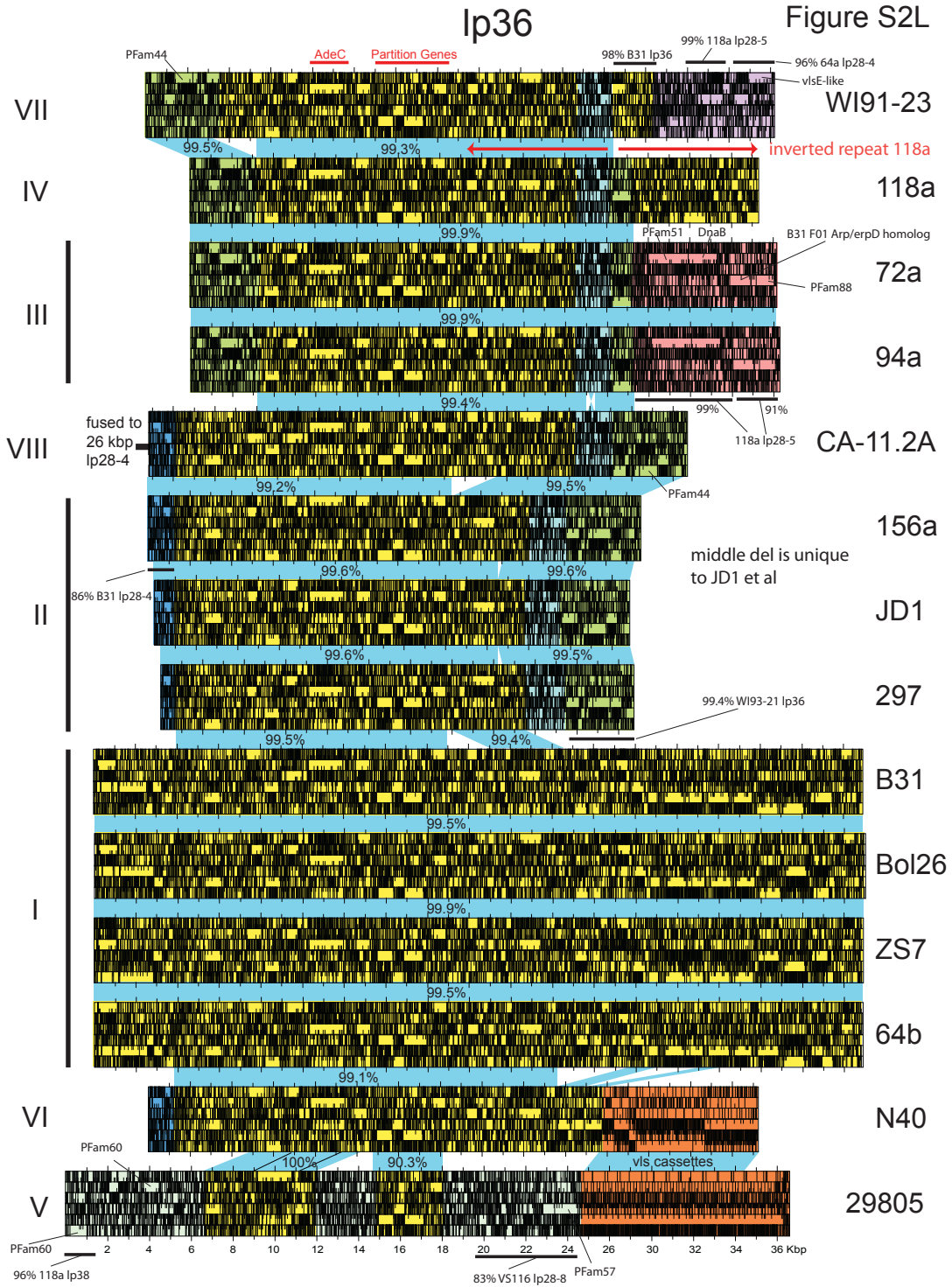
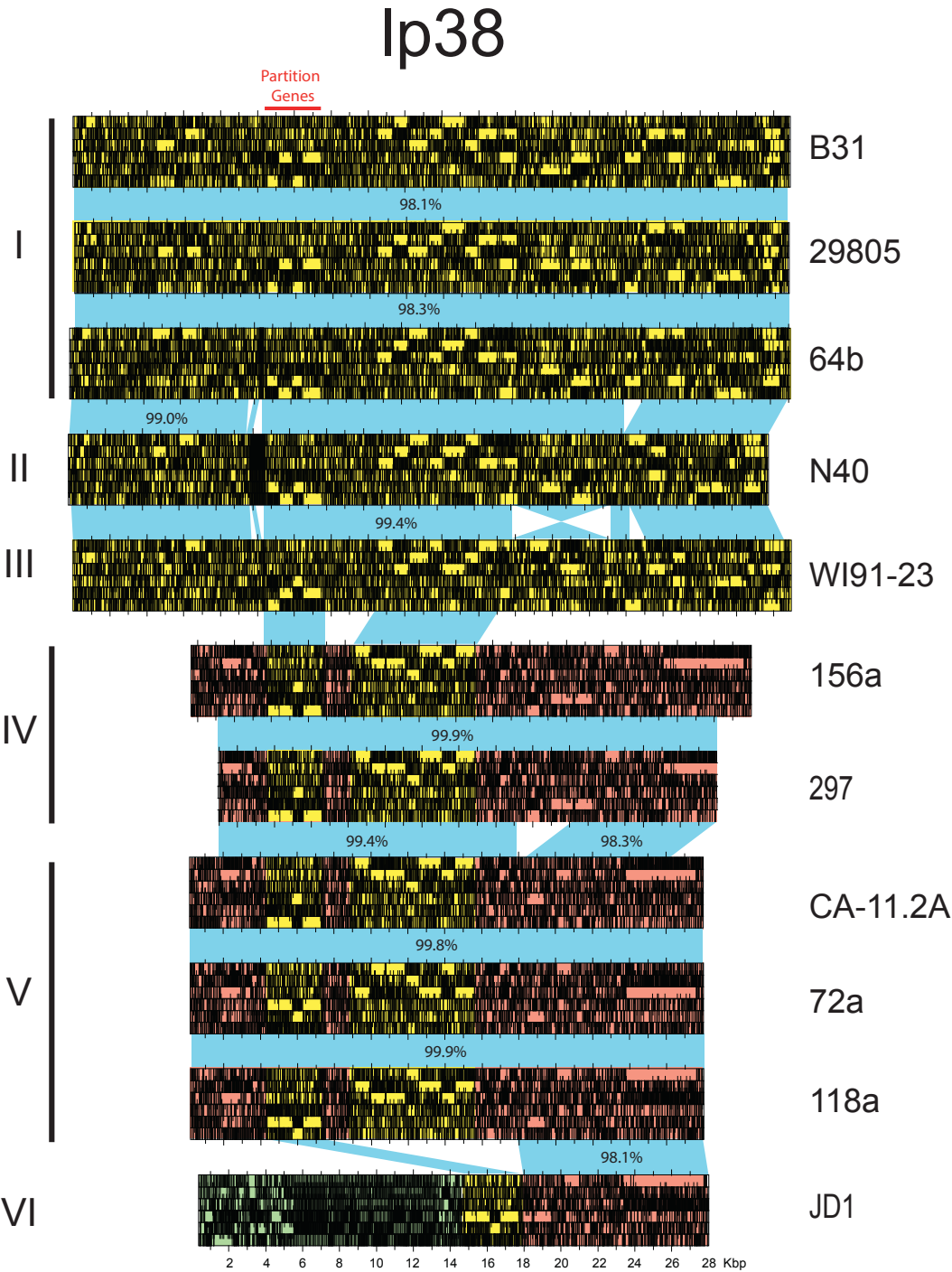


Figure S2M



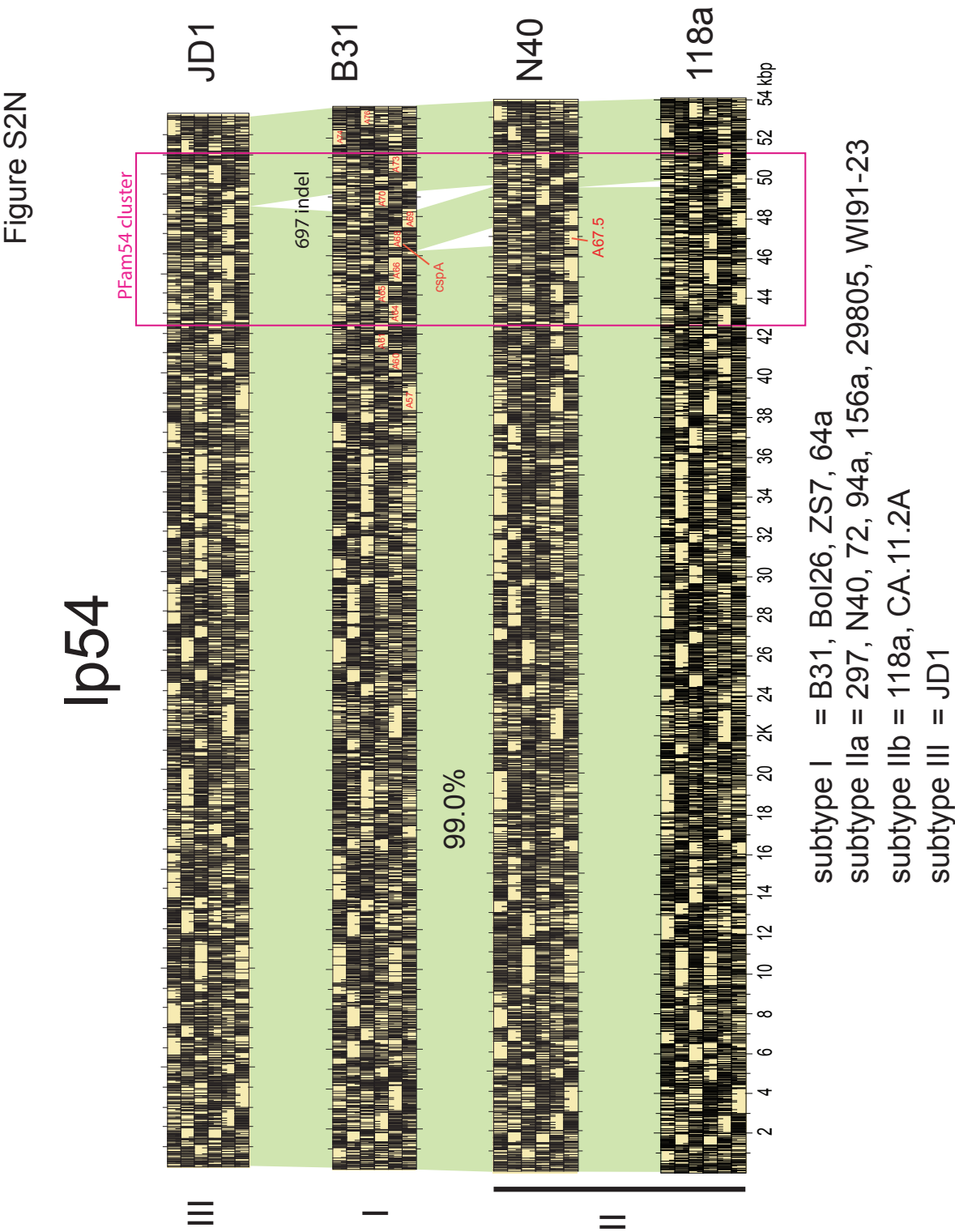


Figure S20

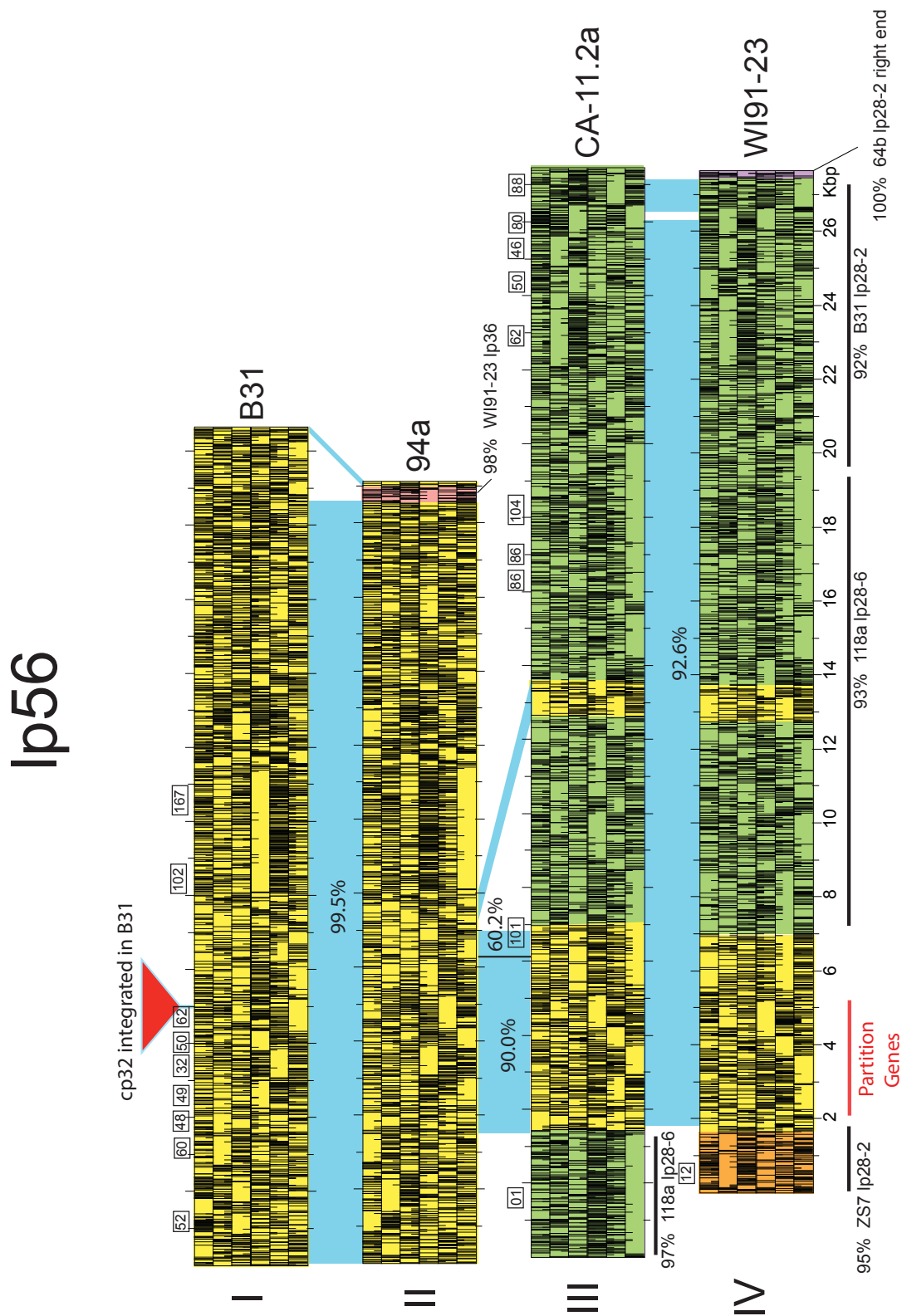


Figure S3. Deletions in cp32 circular plasmids.

The long deletions in cp32 plasmids are indicated by thin lines with the deletion names (also see Figure 1 of text). A scale in kbp is shown below, and the plasmid sequences are oriented parallel to the previously reported strain B31 cp32 plasmids (Casjens *S, et al*: A bacterial genome in flux: the twelve linear and nine circular extrachromosomal DNAs in an infectious isolate of the Lyme disease spirochete *Borrelia burgdorferi*. *Mol Microbiol* 2000, 35:490-516).

Figure S3

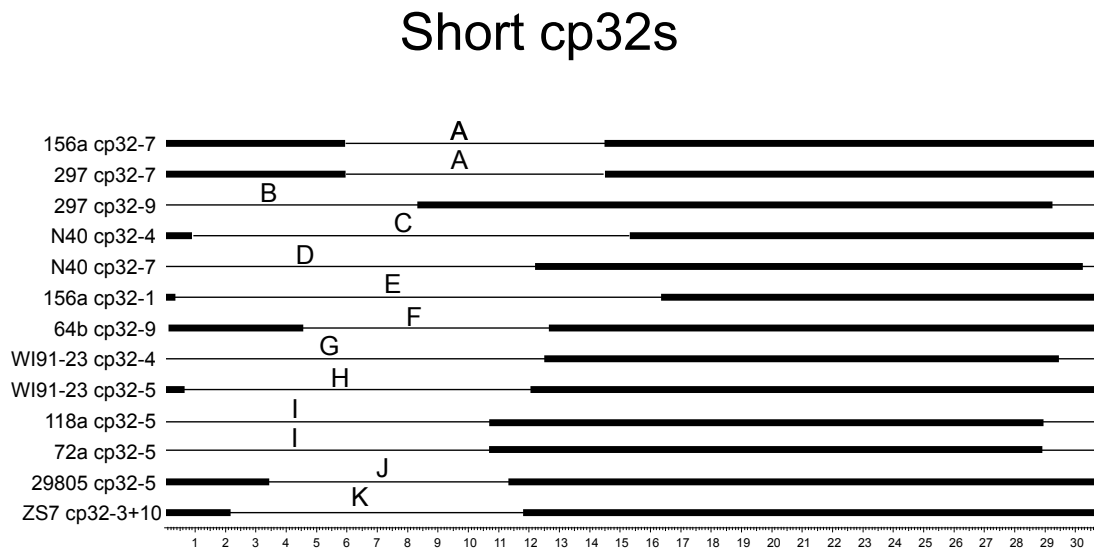


Figure S4. Comparative cp9 maps.

Cp9 plasmid maps are displayed as described in Supplementary figure S2. Plasmid WI91-23#B has remained in three contigs due to sequence assembly difficulties.

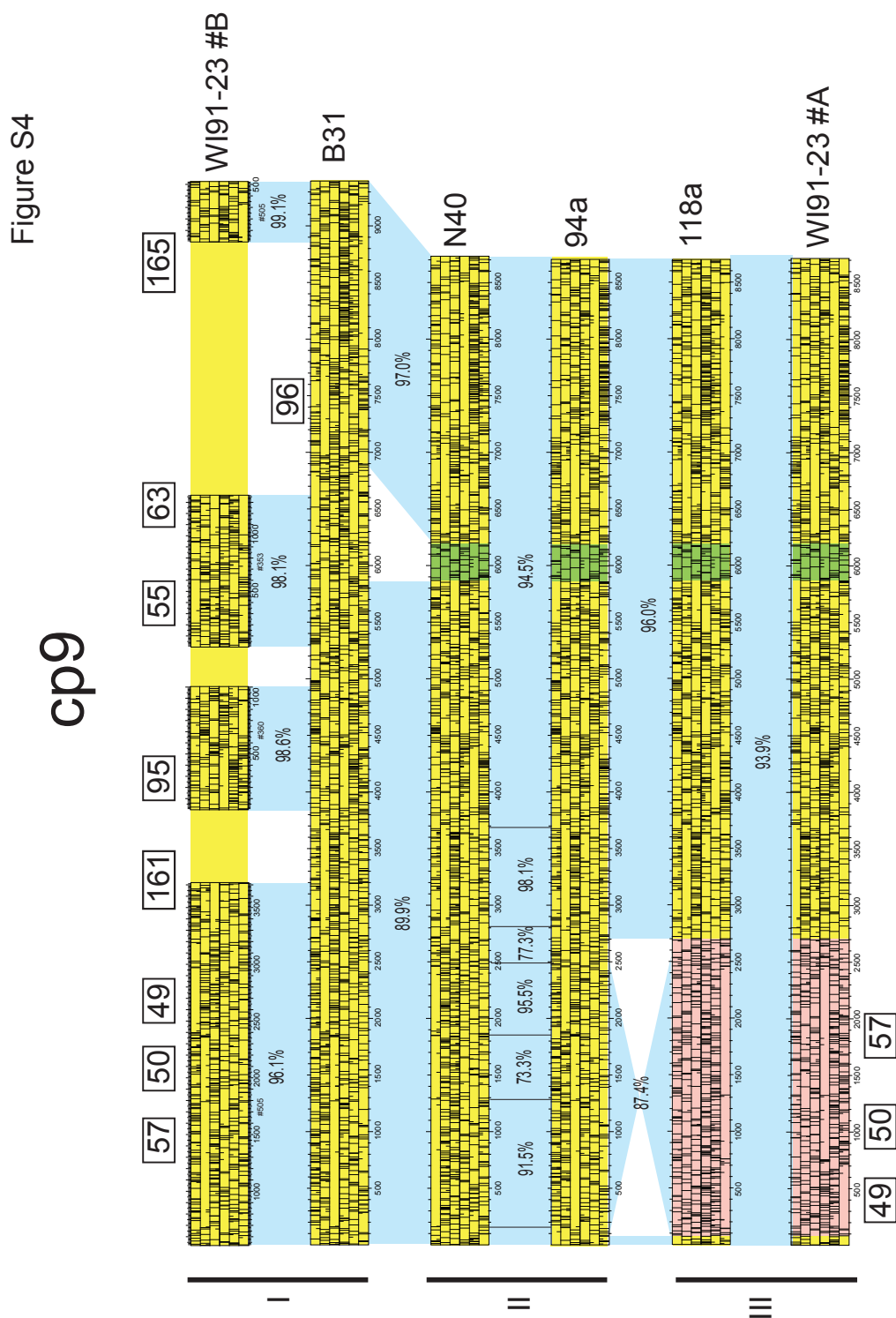


Figure S5. All possible recombination products.

The number of possible ways to randomly (non-homologously) recombine two identical sequences in a given orientation is equal to the number of bp squared, since each bp in one sequence can abut any of the bp in the other sequence. However, there are several possible orientations in which such an operation can be performed.

Panel A shows that two identical sequences represented by black and red arrows can be oriented head-to-head (above) or head-to-tail (below). Each of these can give rise to two sets of reciprocal products; in the figure each product represents all possible ways to recombine them in the orientations shown. If the sequences represented by the red and black arrows are identical, the bottom two sets of reciprocal products - "heads outward" and "tails outward" - are clearly different sequences. However, the top two product sets are identical since the red and black sequences are the same sequence. Thus there are three different sets of products, the number of which in each case is the number of bp in the sequence squared; hence, the total number of possible recombinants is $= 3 \times \text{\#bp}^2$.

Panel B shows this by example. The four boxes list all possible ways in which two identical double-stranded four bp sequences can be joined together so that every recombinant has at least one bp from both parents. There are 16 (4^2) ways to join any 4 bp sequence to another 4 bp sequence, and there are four possible orientations to join them shown in the four boxes (red and black sequences show the orientations that are recombined in each box). However, the recombined sequences in the upper-left and lower-right boxes are the reverse complements of one another when the two sequences are identical (the top strand in one is equivalent to the bottom strand in the other and *vice versa*). The recombined sequences in the upper-right and lower-left boxes are unique. Therefore, the number of possible different recombinant sequences is $3 \times 4^2 = 48$ in this example.

Figure S5

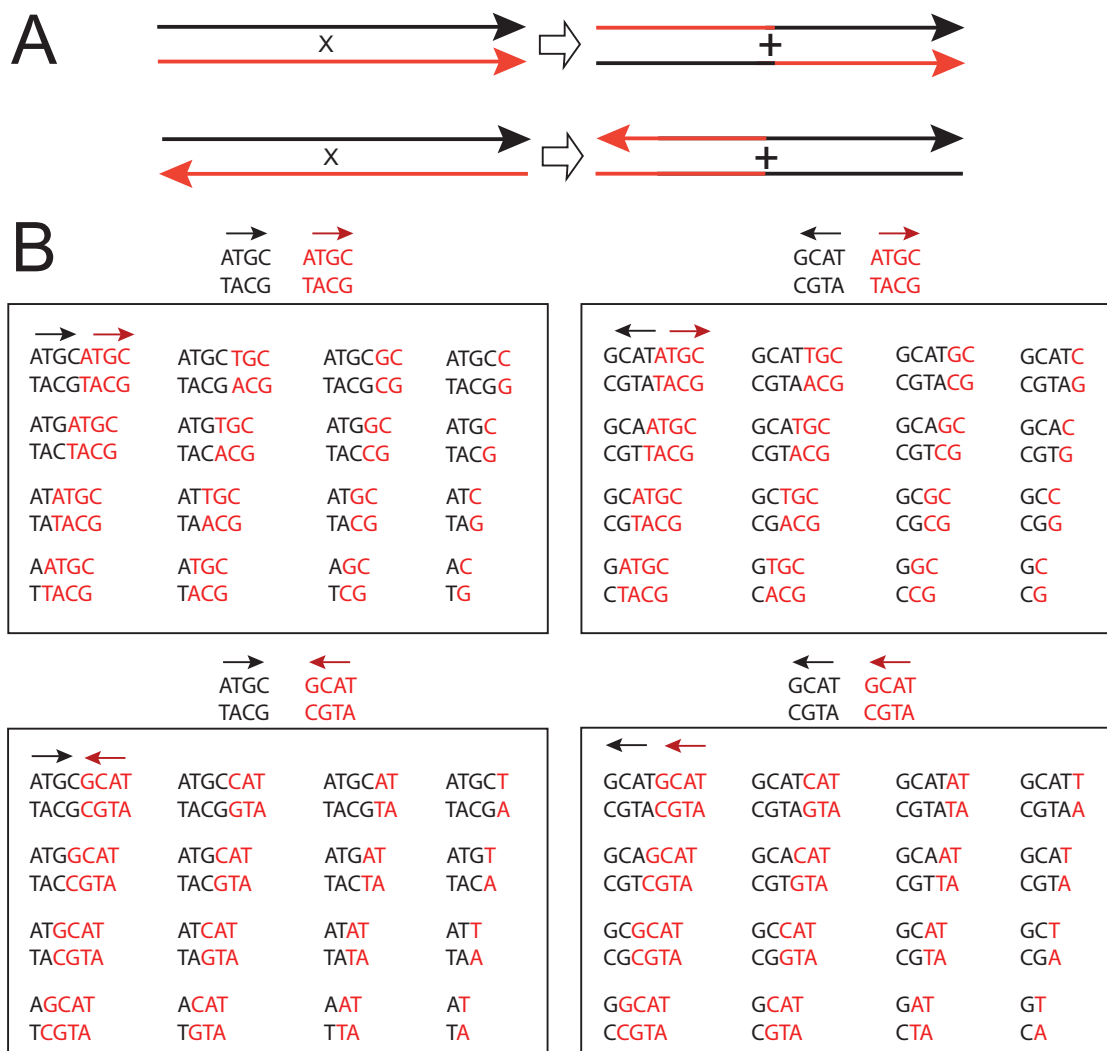


Figure S6. Past transfer of genetic material from the *B. burgdorferi* chromosome right end to a linear plasmid.

A dot matrix plot compares strain 64b linear plasmid lp28-1 to the right end of the strain B31 chromosome, and diagrammatic maps are shown at the top and left of these two DNAs, respectively. Map regions of the same color have very similar sequence and the same putative origin, and gene names are indicated on the map. The yellow plasmid region appears to have been obtained through recombination with a B31-like chromosome (see text). The plot was constructed by DNA Strider (Douglas SE: DNA Strider. A Macintosh program for handling protein and nucleic acid sequences. *Methods Mol Biol* 1994, 25:181-194).

Figure S6

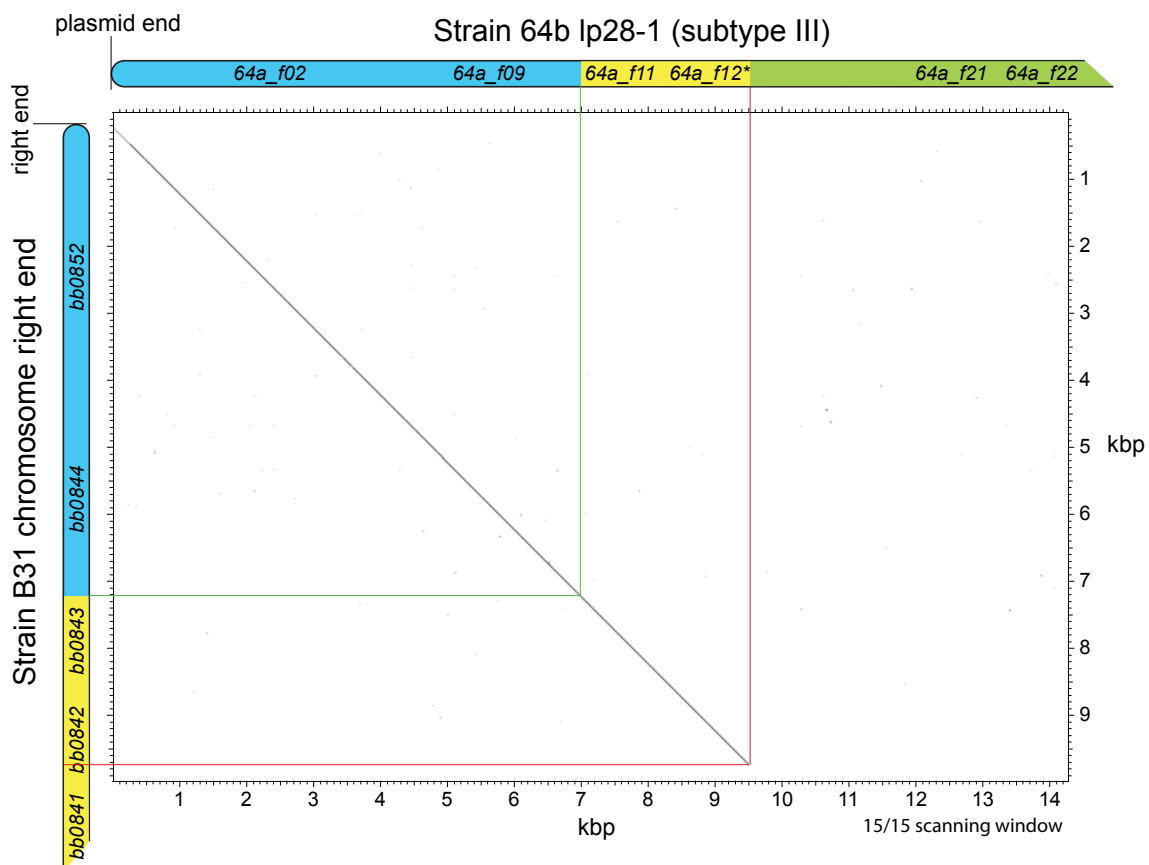


Table S1. Linear plasmid location of selected genes.

The linear plasmids that carry selected genes each of in the fourteen *B. burgdorferi* genome sequences are indicated in the table. Gene names used here are as follows: *adeC*, adenine deaminase; *arp*, arthritis related protein; *cpsZ*, complement regulatory protein Z; *dbpAB*, decorin binding proteins A and B; *fbp*, fibronectin protein; *plzB*, cyclic di-GMP binding protein; *hhaI*, HhaI-like restriction endonuclease; *ospAB*, outer surface proteins A and B; *ospD*, outer surface protein D; *pncA*, nicotinamidase; *res-mod1*, restriction-modification protein Fam01; *res-mod2*, restriction-modification protein Fam102/167; *vls*, variable outer surface protein cassettes.

Footnotes for Table S1

- a. Gene is likely missing due to missing linear plasmid terminal sequence.
- b. Relevant plasmid is known to have been originally present but was lost from the sequenced culture (see article text).
- c. Lp54 gene content is quite constant so only a few examples this plasmid's many important genes are shown.
- d. Gene is to be truncated relative to the longest paralogues; it may be non-functional.
- e. Chromosomal single nucleotide polymorphism (SNP) type (Mongodin, E, *et al.*: Inter- and intra-specific pan-genomes of *Borrelia burgdorferi sensu lato*: Adaptive radiation and phylogenetic predictions. *BMC Genomics* 2013, 14:693).
- f. lp28-4 and lp36 are apparently fused end-to-end in strain CA-11.2A; *fbp* is in the lp36-like portion.

Table S1

Linear plasmid locations of selected genes

<i>Borrelia burgdorferi</i> sensu stricto														
gene (a)	B31	Bo126	ZS7	64b	N40	WI91- 23	29805	94a	72a	CA-11.2A	118a	297	156a	JD1
<i>adeC</i>	lp36	lp36	lp36	lp36	lp36	lp36	lp36	lp36	lp36	lp36	lp36	lp36	lp36	lp36
<i>arp</i>	lp28-1			lp28-5	lp28-5			lp36	lp28-4		lp28-5		lp28-5	lp28-4
<i>cpsZ</i>	lp28-3		lp28-3	lp28-3	(b)	lp17, lp28-3		lp28-3		lp28-3	lp28-3	lp28-3	lp28-3	lp28-3
<i>dbpAB</i>	lp54(c)	lp54	lp54	lp54	lp54	lp54	lp54	lp54	lp54	lp54	lp54	lp54	lp54	lp54
<i>fbp</i>	lp36	lp36	lp36	lp36	lp36	lp36		lp36	lp36	lp36 (f)	lp36	lp36	lp36	lp36
<i>plzB</i>			lp28-1	lp28-5	lp28-5						lp28-5		lp28-5	chrM
<i>hhal</i>				lp28-5	lp28-5						lp28-5		lp28-5	
<i>ospAB</i>	lp54(c)	lp54	lp54	lp54	lp54	lp54	lp54	lp54	lp54	lp54	lp54	lp54	lp54	lp54
<i>ospD</i>	lp38			lp38	lp38	lp38	lp38							
<i>pncA</i>	lp25		lp25		lp25	lp25	lp25				lp25	(b)	lp25	lp25
<i>res-mod1</i>	lp25, lp28-3		lp25, lp28-3	lp28-3, lp28-5	lp25, lp28-5	lp25, lp28-3, lp28-6	lp25, lp28-3, lp28-3(d), lp28-6(d)	lp28-3, lp28-8	lp38	lp28-3, lp38, lp56(d)	lp25, lp28-3, lp28-5, lp28-6(d), lp38	lp28-3, lp28-5, lp38(d)	lp25, lp28-3, lp28-5, lp38	lp25, lp28-3, lp28-5, lp38
<i>res-mod2</i>	lp56							lp56						chromosome
<i>vis</i>	lp28-1	lp28-3	lp28-1	lp28-1	lp36	lp28-1	lp36	lp28-3	lp32-3		lp32-3	lp28-1	lp28-1	lp28-1
Chromosome	A	A	A	A	B	B	B	C	C	C	C	D	D	D
SNP group (e)														