

Genome Stability of Lyme Disease Spirochetes: Comparative Genomics of *Borrelia burgdorferi* Plasmids

Sherwood R. Casjens^{1*}, Emmanuel F. Mongodin², Wei-Gang Qiu³, Benjamin J. Luft⁴, Steven E. Schutzer⁵, Eddie B. Gilcrease¹, Wai Mun Huang¹, Marija Vujadinovic¹, John K. Aron¹, Levy C. Vargas³, Sam Freeman³, Diana Radune⁶, Janice F. Weidman^{6^{aa}}, George I. Dimitrov^{6^{ab}}, Hoda M. Khouri^{6^{ac}}, Julia E. Sosa⁶, Rebecca A. Halpin⁶, John J. Dunn⁷, Claire M. Fraser²

1 Department of Pathology, University of Utah School of Medicine, Salt Lake City, Utah, United States of America, **2** Department of Medicine and Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore, Maryland, United States of America, **3** Department of Biological Sciences, Hunter College of the City University of New York, New York City, New York, United States of America, **4** Department of Medicine, Health Science Center, Stony Brook University, Stony Brook, New York, United States of America, **5** Department of Medicine, New Jersey Medical School, Newark, New Jersey, United States of America, **6** J. Craig Venter Institute, Rockville, Maryland, United States of America, **7** Biology Department, Brookhaven National Laboratory, Upton, New York, United States of America

Abstract

Lyme disease is the most common tick-borne human illness in North America. In order to understand the molecular pathogenesis, natural diversity, population structure and epizootic spread of the North American Lyme agent, *Borrelia burgdorferi* sensu stricto, a much better understanding of the natural diversity of its genome will be required. Towards this end we present a comparative analysis of the nucleotide sequences of the numerous plasmids of *B. burgdorferi* isolates B31, N40, JD1 and 297. These strains were chosen because they include the three most commonly studied laboratory strains, and because they represent different major genetic lineages and so are informative regarding the genetic diversity and evolution of this organism. A unique feature of *Borrelia* genomes is that they carry a large number of linear and circular plasmids, and this work shows that strains N40, JD1, 297 and B31 carry related but non-identical sets of 16, 20, 19 and 21 plasmids, respectively, that comprise 33–40% of their genomes. We deduce that there are at least 28 plasmid compatibility types among the four strains. The *B. burgdorferi* ~900 Kbp linear chromosomes are evolutionarily exceptionally stable, except for a short ≤ 20 Kbp plasmid-like section at the right end. A few of the plasmids, including the linear lp54 and circular cp26, are also very stable. We show here that the other plasmids, especially the linear ones, are considerably more variable. Nearly all of the linear plasmids have undergone one or more substantial inter-plasmid rearrangements since their last common ancestor. In spite of these rearrangements and differences in plasmid contents, the overall gene complement of the different isolates has remained relatively constant.

Citation: Casjens SR, Mongodin EF, Qiu W-G, Luft BJ, Schutzer SE, et al. (2012) Genome Stability of Lyme Disease Spirochetes: Comparative Genomics of *Borrelia burgdorferi* Plasmids. PLoS ONE 7(3): e33280. doi:10.1371/journal.pone.0033280

Editor: Brian Stevenson, University of Kentucky College of Medicine, United States of America

Received: December 16, 2011; **Accepted:** February 6, 2012; **Published:** March 14, 2012

Copyright: © 2012 Casjens et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by grants AI49003 (SRC), AI37256 (BJL), GM083722 (WQG) and AI47553/AI077156 (SES) from the National Institutes of Health, and by a grant from the Lyme Disease Association, Inc., Jackson, NJ. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: sherwood.casjens@path.utah.edu

^{aa} Current address: Virginia, United States of America

^{ab} Current address: Center for Genomic Sciences, United States Army Medical Research Institute of Infectious Diseases, Ft. Detrick, Maryland, United States of America

^{ac} Current address: Bethesda, Maryland, United States of America

Introduction

Bacteria in the spirochete genus *Borrelia* cause arthropod-borne human diseases such as Lyme disease and relapsing fever, as well as a number of diseases of veterinary importance [1–6]. They are obligate parasites that are only found in their vertebrate or arthropod hosts and are rather difficult to study in the laboratory. Only quite recently have their biology, genetics and molecular pathogenesis begun to become accessible to experimentation [7–9]. The determination and analysis of the first *Borrelia* genome sequence, that of the *Borrelia burgdorferi* type strain B31, stimulated significant progress in this arena. Its unusual genome was found to comprise a 910 Kbp linear chromosome and twenty-one (twelve linear and nine circular) plasmids that contain over 600 Kbp of

DNA [10,11] (two additional plasmids are now thought to have been lost between the isolation of strain B31 and its genome sequence determination [12,13]). This work confirmed Barbour's [14] original observations, and many other studies have shown that *Borrelia* isolates universally harbor numerous linear and circular plasmids (e.g., [15–27]). The B31 chromosome carries 815 predicted genes (our re-annotation, below) that encode largely housekeeping functions. These functions include a minimal metabolic capability that cannot synthesize amino acids, nucleotides or lipids *de novo* [11].

One circular plasmid, cp26, carries genes that encode several nucleotide metabolism enzymes [28], small molecule transporters [29,30] and the enzyme that creates the unique closed hairpin telomeres present on the *Borrelia* linear replicons [31–34]. The

other plasmids have very few metabolic or housekeeping genes, but do encode numerous lipoproteins, many of which have been shown to be present on the cell surface when they are expressed (*e. g.*, [35–37] and references therein). The plasmids have a number of interesting features in addition to bearing lipoprotein genes. (i) A number of the linear plasmids have an unusually low protein coding density for prokaryotic DNA and carry numerous pseudogenes that appear to be in a state of genetic decay [10,38]. (ii) Several of the circular plasmids in strain B31 (the cp32s) are homologous nearly throughout their lengths [10,12]. (iii) There are unusually large numbers of paralogous genes on the plasmids. The vast majority of strain B31 plasmid genes have plasmid-borne paralogs, and in strain B31 107 of the paralogous gene families (PFams) include mostly plasmid genes. (vi) The highly paralogous nature of the plasmids, along with the apparent mutational decay of some members of PFams, suggests a history of duplicative rearrangements followed by decay of damaged and redundant genes [10,38]. (v) Up to eleven of the B31 plasmids appear to be prophages or are prophage-related [39–41]. (vi) Most of the plasmids, probably all but cp26, can be lost without affecting growth in culture (*e. g.*, [42,43]). (vii) And finally, several plasmids have been shown to be required for growth in mice or in *Ixodes* ticks, and/or encode proteins that interact with host components (details below). Thus, the plasmids appear to be largely involved in the interactions of *Borrelia* with its hosts.

All members of the *Borrelia* genus that have been analyzed carry linear chromosomes that are similar in size to the strain B31 chromosome. These chromosomes appear to be quite evolutionarily stable, since their sizes do not vary greatly and recent sequences of the chromosomes of additional Lyme agent *B. burgdorferi* sensu stricto species [44] and related species *B. garinii*, *B. afzelii*, *B. “bavariensis”*, *B. “finlandensis”*, *B. valaisiana*, *B. spielmanii* and *B. bissettii* [45–49], show that they are all essentially co-linear with the chromosome of *B. burgdorferi* B31, and that there are only a very small number of chromosomal gene content differences among these species (with the exception of *B. burgdorferi* extreme right-end differences [50,51] and the larger but still relatively modest differences between Lyme agent and relapsing fever *Borrelia* species [52]). Directed analyses have shown that *B. burgdorferi* plasmids cp26 [27], lp54 [20] and the cp32s [12] have largely conserved structures and are present in all isolates that have been studied. Other plasmids appear to have conserved structures but are only present in a subset of strains (*e. g.*, B31-like cp9 [10,53,54] and lp38 [21]), while still others such as lp5, lp21, lp36, and lp56 are less frequently present and/or have highly

variable sizes and presumably variable structures [19,21,55,56]. The similar sizes of different plasmids (which are not separable in electrophoresis gels) and the highly paralogous nature of the plasmids has made unambiguous assembly and analysis of plasmid sequences complex and difficult [10,47]. Thus, studies of bacteria in the *Borrelia* genus are in an unenviable position in which determination of all the plasmids present in any new isolate requires that a complete (non-draft) genome sequence be determined.

Comparison of whole genome nucleotide sequences both within and between species is a powerful and critical part of gaining a true understanding of the organization, diversity and evolution of bacterial genomes. This strategy reveals the invariant features of the compared genomes and allows discovery of more variable sequences that (i) correlate with specific host disease features, (ii) permit tracking of sub-types within species, and (iii) give critical insight into evolutionary mechanisms. In addition, comparison of closely related genomes can often illuminate inaccuracies in the prediction of genes and other features in genomes. In this report we discuss the plasmids present in the *B. burgdorferi* genomes of isolates N40, JD1 and 297 and compare their genetic contents and organizations with the previously known strain B31 genome. More global and less gene oriented comparisons of the twenty-two *B. burgdorferi* sensu lato genomes that we have sequenced [10,11,44–46,49] will be presented in subsequent publications.

Results and Discussion

B. burgdorferi whole genome sequence determination

In order to begin to address questions about *B. burgdorferi* and population structure, the genetic basis of virulence, possible exchange of genetic information among individual bacteria in the wild, as well as natural diversity and evolutionary mechanisms of the Lyme disease *Borrelia* species, we determined and annotated the complete sequences of the whole genomes of *B. burgdorferi* strains N40 and JD1 and of the plasmids of strain 297 [44]. These strains and strain B31 were chosen for the analysis presented here because (i) space prevents such a detailed analysis of many more isolates, (ii) they include the three most commonly used laboratory strains (B31, N40 and 297), (iii) they represent four different lineages by rRNA spacer sequence [57–60], pulsed-field gel DNA pattern [16,17,61], OspC (outer surface protein C) [62–65] and multilocus sequence type [66–69] (Table 1), and, importantly, (iv) the accuracy of the computer assembly of the plasmid sequences has only been confirmed experimentally for only these four

Table 1. *B. burgdorferi* isolates in this study.

Isolate	PFGE type ^a	rRNA IGS1 lineage ^b	OspC type ^c	MLST type ^d	Number of plasmids		Biological source	Ref.
					Linear ^e	Circular ^e		
B31	B	1	A	1	12	11 ^f	<i>Ixodes</i> tick/NY	[70]
N40	E	9	E	19	9 ^f	8	<i>Ixodes</i> tick/NY	[71]
JD1	C	5	C	11	11	9	<i>Ixodes</i> tick/MA	[72]
297	A	2	K	3	10 ^f	10	Human/CT	[73]

^aNumerous *B. burgdorferi* chromosomal pulsed-field gel electrophoresis (PFGE) types have been identified (*e. g.*, [16,17,61]).

^bAt least nine *B. burgdorferi* rRNA intragenic spacer (IGS) lineages have been defined [57,58,74].

^cOver 20 *B. burgdorferi sensu stricto* OspC types have been identified [57,62].

^dMultilocus sequence typing (MLST) of Margos *et al.* [67].

^eNumber of plasmids present (the values for N40, JD1 and 297 were determined in this study and for B31 by Casjens *et al.* [10]).

^fIncludes the following plasmids known to have lost from the cultures whose genomes were sequenced: B31, cp9-2 and cp32-5; 297, lp25; and N40, lp28-3 (see also text and footnote c of table 2).

doi:10.1371/journal.pone.0033280.t001

isolates. Strain 297 was isolated from a human with Lyme disease in CT, while JD1, N40 and B31 are *Ixodes scapularis* tick isolates from MA, NY and NY, respectively. All four come from the northeastern part of United States of America, a region with a high frequency of human Lyme disease.

Whole genomic DNAs from *B. burgdorferi* strains N40 and JD1 and isolated plasmid DNA from strain 297 were sequenced by previously described Sanger sequencing random shotgun and closure methods (Materials and Methods). In each genome, DNA library “shotgun” sequencing followed by closure of sequence gaps by sequencing PCR amplicons or additional DNA clones resulted in the accumulation of multiple, unconnected sequence contigs. Our previous experience with the *B. burgdorferi* B31 genome suggested that such contigs most likely represent the large chromosome and the multiple plasmids that all *Borrelia* cells carry, and that repeated sequences on plasmids can cause incorrect sequence assembly. To confirm that the shorter contigs are plasmid-derived and to check the correctness of the assembly of the sequencing runs into contigs, whole cell DNAs, uncloned and cleaved with strategically chosen restriction enzymes, were displayed in pulsed-field electrophoresis agarose gels and analyzed by Southern hybridization using unique probes prepared by PCR amplification from each of the putative plasmids contigs in N40, JD1 and 297 (data not shown). This restriction mapping ensured that improper assemblies were corrected, that the linear or circular nature of each plasmid was independently determined, and that the plasmid sequences accurately reflect the true *in vivo* situation. Since the covalently closed, hairpin-ended terminal fragments from the *Borrelia* linear replicons do not ligate into the vectors used to make circular plasmid sequencing libraries [10,11], sequence determined in this way is expected to be missing some bp from each end of the linear replicons. In the process of confirming the sequence assembly, the native plasmid sizes and sizes of terminal restriction fragments from most linear plasmid ends were measured, and the approximate number of unsequenced bp at the DNA ends were estimated; these values are given in table S1. The nucleotide sequence currently determined for the plasmids of *B. burgdorferi* strains B31, N40, 297 and JD1 are 612108, 437361, 508697 and 608486 bp, and represent about 40%, 33%, 40% and 37% of these total genome sequences, respectively. The accession numbers of the chromosome and plasmids of these four strains were reported previously [10,11,44] (and are also listed in table S1). Some preliminary and/or specific gene findings regarding these genome sequences at their incomplete “draft” stages have been reported elsewhere [66,75–79].

During this work we found that the strain “N40” *ospC* gene sequence deposited in GenBank under accession number AF416430 is in fact *not* from the authentic N40 which was originally isolated and described by Barthold *et al.* [71] and whose genome we sequenced (the correct N40 *ospC* sequence has been previously deposited in GenBank three times under the accession Nos. U04240, DQ437463 and AY275221). We have confirmed that the “AF416430 strain”, which we call strain “M” in this report (see below), is still masquerading as “N40” in some laboratories (see [80] for additional details), and caution is recommended in assuming that the N40 genome sequence discussed here is from the same isolate as all strains previously reported under “N40” name.

The *B. burgdorferi* sensu stricto chromosome

The chromosome “constant portion”. Restriction mapping, anecdotal sequencing, and DNA array analysis have indicated that the chromosomes of different isolates of *B. burgdorferi* sensu stricto are quite similar (*e. g.*, [17,61,66,67]). The genome

sequences show that the B31, N40 and JD1 chromosomes are indeed essentially completely syntenic, with the only major length variation among *B. burgdorferi* sensu stricto chromosomes being different amounts of plasmid-like DNA attached at their right ends (see below). The “constant region” includes the left 903 Kbp that carries strain B31 genes *b31_0001* through *b31_0843* (see Materials and Methods for gene nomenclature). B31-JD1, B31-N40 and JD1-N40 pair wise comparisons show that the constant regions of the chromosomes are 99.5%, 99.4% and 99.4% identical, respectively.

In an effort to improve the accuracy of the annotation of the *B. burgdorferi* genes, we annotated the JD1, N40 and 297 genome sequences in parallel, and updated the B31 genome annotation, as described in Materials and Methods. ORFs ≤ 50 codons were removed from the prediction, and those in the 51–100 range were not predicted unless they were intact and had homologs in all four of the genome sequences discussed here; two putative chromosomal genes, *b31_0771a* and *b31_0838a*, were identified that were not recognized in the original annotation of the B31 chromosome. Many of the short ORFs were previously suspected to be spurious gene identifications or nonfunctional genes [10], and these are not included in the present analysis unless they meet the above criteria. We thus identify 815 putative chromosomal protein-coding genes which occupy 93.5% of the 903 Kbp constant region in the B31, N40 and JD1 chromosomes. These genes as well as the tRNA, tmRNA and rRNA genes are all present and in identical locations in all three *B. burgdorferi* chromosomes, and there are no large indels or rearrangements among the three sequenced chromosome constant regions. The constant regions of the chromosome will be compared in detail elsewhere and will not be discussed further in this report.

The *B. burgdorferi* chromosomes have significant gene content differences only in the variable region at their right ends. We previously identified three different lengths of extensions beyond the right end of the constant region in a panel of 31 isolates from North America [17,50,51]. Strain N40 represents a chromosome that has no long “extra” DNA extension at its right end, and its rightmost gene (*n40_0843*, which encodes a probable arginine-ornithine antiporter) is less than two hundred bp from the right telomere by our measurement of terminal restriction fragment lengths (data not shown). B31, 297 and JD1, have additional sequences that extend about 7, 19 and 20 Kbp beyond the N40 right end position [51]. The B31 extension is “plasmid-like” in that it contains only genes that are similar to those on B31’s plasmids [10]. Strain 297’s right end is extremely similar to that of strain Sh-2-82, and much of the latter’s chromosome extension was determined to be 99% identical in sequence to the B31 linear plasmid lp21 [51]. The exact sources of the B31 and JD1 extensions were not known.

The B31 chromosome extension can now be identified as >98% identical to the version of linear plasmid lp28-1 that is present in strain 297 (plasmid nomenclature is discussed in the following section), and three contiguous sections of the JD1 extension are 99.2%, 99.6% and 99.6% identical to sections of 297 lp28-1, B31 lp28-1, and N40 lp28-5 linear plasmids, respectively (Figures 1 and S1). The extremely high similarity between the chromosomal extensions and these plasmids strongly supports the notion that these chromosomal right-end extensions and linear plasmids have had very recent common ancestors. The mechanism of joining plasmid sequences to the chromosome is not known. It does not appear to be fusion at the telomeres, since all the plasmid-like sequences at the chromosome ends are not near the ends of plasmids except the right chromosomal telomere, which appears to be the plasmid telomere when there is an

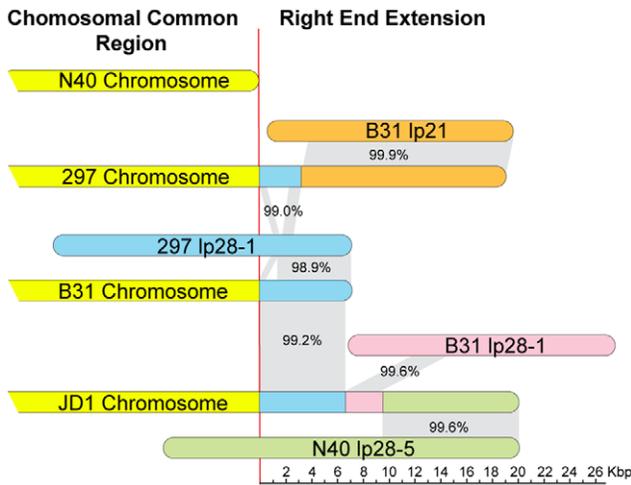


Figure 1. Length variation of *B. burgdorferi* chromosomes. The relationships among the right end, plasmid-like, chromosomal extensions relative to known plasmids are indicated by gray shading; plasmid sizes are not drawn exactly to scale. There is a 1053 bp deletion and a 17 bp insertion in the 297 lp28-1 plasmid relative to the B31 extension. It is assumed that the 297 chromosome is essentially identical to that strain Sh-2-82 (see text and [51]).
doi:10.1371/journal.pone.0033280.g001

extension. We have argued previously that, since (i) closely related species such as *B. bissettii*, *B. garinii* and *B. afzelii* do not appear to have such chromosomal extensions, and (ii) genes are not likely to be rebuilt perfectly by non-homologous recombination events, these represent accretions of DNA onto the chromosome from linear plasmids and thus represent a mechanism by which *Borrelia* recruits plasmid genes onto the chromosome [10,50,51].

Plasmid types, occurrence and nomenclature

Plasmid content of sequenced genomes. Our results show that the sequenced cultures of *B. burgdorferi* strains B31, N40, JD1 and 297 carry 12, 8, 11 and 9 linear plasmids and 9, 8, 9 and 10 circular plasmids, respectively. Thus, these isolates each carry between 16 and 21 plasmids. Table 2 lists the plasmids present in each genome, and table S1 gives their experimentally measured and sequence contig sizes. With the updated re-annotation of the B31 genome described above, the number of annotated genes on sequenced plasmids is 706, 463, 677 and 585 in B31, N40, JD1 and 297, respectively. A difficulty in working with *Borrelia* strains in the laboratory is that plasmids are often spontaneously lost, and, although there is evidence that strain 297 carries a homolog of B31 linear plasmid lp25 [81,82] and N40 carries a homolog of lp28-3 [78], these two plasmids were not present in the genome sequences determined in this study. They must have been lost between the original isolation and growth of the culture for DNA isolation and genome sequencing; strain B31 is also known to have lost two or three plasmids before sequencing [12,13]. The bp position numbers of the linear plasmids of B31 used in this report are those of the original GenBank annotations of Casjens *et al.* [10] that do not include the more recently determined terminal sequences reported by Tourand *et al.* [83].

***B. burgdorferi* plasmid types and nomenclature.** *Borrelia* plasmids were originally named according to their DNA topology (linear “lp” and circular “cp”) and approximate size in Kbp. Strain B31 linear plasmid lp54, for example, is 53678 bp in length. To continue to name the plasmids in all strains according to their size, however, has several difficulties: (i) a majority of the linear

Table 2. *B. burgdorferi* plasmids present in four isolates.

	B31	N40	JD1	297
Linear plasmids	12	8	11	9
Circular plasmids	9	8	9	10
Total plasmids	21	16	20	19
lp5	+	–	–	–
lp17	+	+	+	+
lp21	+	–	–	–
lp25	+	+	+	(+) ^a
lp28-1	+	–	+	+
lp28-2	+	+	–	–
lp28-3	+	(+) ^a	+	+
lp28-4	+	+	+	+
lp28-5	–	+	+	+
lp28-6	–	–	+	+
lp28-7	–	–	+	–
lp36	+	+	+	+
lp38	+	+	+	+
lp54	+	+	+	+
lp56	+	–	–	–
cp9-1	+	+	–	–
cp9-2	(+) ^a	–	–	–
cp26	+	+	+	+
cp32-1	+	–	+(fused) ^b	+
cp32-3	+	–	+	+
cp32-4	+	+(trunc) ^b	–	+
cp32-5	(+) ^a	+	+(fused) ^b	+
cp32-6	+	–	+	+
cp32-7 ^c	+	+(trunc) ^b	–	+
cp32-8	+	–	+	–
cp32-9	+	+	+	+(trunc) ^b
cp32-10	+(int) ^b	+	+	–
cp32-11	–	–	+	+
cp32-12	–	+	+	+

^aPlasmids known to be in some cultures of the indicated strain, but which were not present in the sequenced culture. Their sequences remain undetermined.

^bStructural differences from otherwise organizationally similar plasmids are indicated as follows: trunc, truncated compared to other homologous plasmids; int, B31 cp32-10 is integrated into plasmid lp56; fused, JD1 cp32-1 and cp32-5 are fused into one large circular “cp32-1+5” plasmid.

^cIn some B31 cultures the plasmid cp32-7 is replaced by cp32-2 [12]. These two plasmids have the same apparent compatibility and appear to be prophage DNAs. Since it seems unlikely that they can exist in the same cell, and they are expected to be able to move between strains, one of them may have been inadvertently introduced in the laboratory. We use cp32-7 for this compatibility type since it is the one that is present in the completely sequenced B31 genome.

doi:10.1371/journal.pone.0033280.t002

plasmids are in the 24 to 30 Kbp size range, so different names based only on size are limited, (ii) we find numerous significant organizational and size differences among the strains (*e. g.*, the “lp36’s” present in the four strains considered here range from 23 to 36 Kbp in length; see below), and (iii) such names have no biological significance. To give plasmids names that correlate with at least some biological feature, we [75,77] and Stevenson and Miller [84] have suggested that, when possible, names be given to

Borrelia plasmids according to their type of partitioning genes, and in particular the type of paralogous family (PFam) 32 protein that they encode (PFams defined by Casjens *et al.* [10]; see also below). The PFam32 proteins are homologs of the ParA plasmid partitioning proteins in other better understood systems [85,86], and experiments in *Borrelia* indicate that PFam32 protein “sequence types” correlate with plasmid compatibility types (see references cited in Casjens *et al.* [75,77]). The PFam32 proteins encoded by the four genomes fall into twenty-six easily distinguishable sequence types, fourteen on linear plasmids and twelve on circular plasmids [75]. Additional compatibilities include plasmids cp9 and lp5, which carry no PFam32 gene and so cannot be categorized in this way. The compatibility properties of lp5 and cp9 are not understood; no lp5-like plasmid is present in the other isolates analyzed here, and it is not known if the cp9’s in B31 and N40 are compatible (see below).

In the three new genome sequences three new linear plasmid PFam32 types are identified that are not present among the previously known B31 plasmids. These are named “lp28-5” (present in strains N40, JD1 and 297), “lp28-6” (JD1 and 297) and “lp28-7” (JD1) (Table 2). The sizes of these six new “lp28” linear plasmids range between 27 and 31 Kbp. We chose “lp28-X” names, since they carry genes that are largely from the same set of PFams as the four “lp28” plasmids present in strain B31. In addition, two PFam32 types not present in B31 are found among the circular plasmids of N40, JD1 and 297. Stevenson and Miller [84] independently discovered both of these types and named them cp32-11 and cp32-12.

Plasmid variation among *B. burgdorferi* strains

Plasmid organizational and gene content variation. Comparison of strains B31, N40, JD1 and 297 reveals major structural

differences in a number of the plasmid PFam32 types, in addition to complete plasmid presence-or-absence differences. There appear to have been numerous rearrangements between and within plasmids since their last common ancestors. These rearrangements are largely restricted to the linear plasmids and result in a patchwork or mosaic relationship when two “cognate” (in same PFam32 group) plasmids are compared. A given plasmid can have long patches of very high sequence similarity (often several Kbp >99% identical) with a plasmid in another strain, and yet have immediately adjacent sequences that are different in the two plasmids and that are either (i) nearly identical to part of a different plasmid, (ii) homologous to, but less highly related to sections of other plasmids, or (iii) not present in the other three strains analyzed here. Such *inter-strain* linear plasmid relationships are reminiscent of the relationships among the linear plasmids *within* each individual strain, and we previously argued from analysis of the strain B31 plasmids that the intra-strain mosaic relationships were apparently generated by duplicative rearrangements and perhaps also horizontal transfer processes [10,38]. The presence of such inter-strain differences agrees with previous studies which have shown that particular plasmid sequences (used as hybridization probes) are not present on identically sized plasmids in all *B. burgdorferi* isolates (*e. g.*, [19,21]). Because of these complex relationships, in this report we use a conservative definition of “orthologous” to mean identical syntenic positions on plasmids of the same compatibility type, and *not* for the most closely related genes when two strains are compared or for genes that lie in small regions of synteny on different plasmid types.

Figure 2 shows a diagrammatic depiction of the differences in organization among the linear plasmids of strains B31, N40, JD1 and 297 in which identical colors represent nucleotide sequences that are >94% identical in the different genomes. (The >94% cutoff is arbitrary; however, we note that essentially all of the

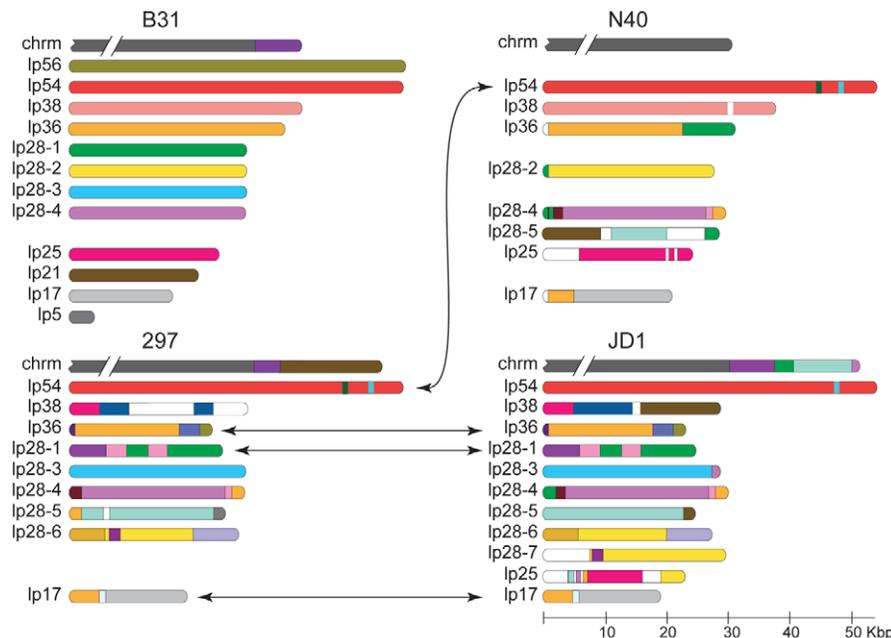


Figure 2. Linear plasmid contents of *B. burgdorferi* strains B31, N40, 297 and JD1. The linear plasmids and right end plasmid-like chromosomal extensions of are shown as horizontal bars with rounded ends. Identical colors indicate regions of nucleotide sequence that is $\geq 94\%$ identical, and white denotes regions that are $< 94\%$ identical to other sequences in the diagram. Each of the B31 plasmids was first defined with a different color and additional colors were added to the other plasmid sets as necessary. Arrows connect plasmids that have identical overall organization and high sequence similarity (ignoring small polymorphisms and indels < 500 bp). Strain 297 plasmids lp28-3 and lp28-4 have not been sequenced to their termini (see table S1), so it is not known whether they are organizationally the same as their B31 or JD1 and N40 or JD1 cognates, respectively.

doi:10.1371/journal.pone.0033280.g002

locally syntenic and thus orthologous regions are >94% identical among these strains. If lower cutoffs are used many more regions that are homologous but not orthologous would be given the same color in the figure.) It is evident from this overview that although a few organizationally identical linear plasmids are found in the different strains, none of the linear plasmid PFam32 types are organizationally identical across the four strains. Linear plasmids, lp54, lp28-2 and lp28-3 have relatively small (~1 Kbp or less) differences among the four strains, but lp17, lp25, lp28-5, lp36 and lp38 each have three very different versions that have significantly different gene contents. There is relatively little strain-specific DNA that is unique to any of the four strains; *i.e.*, only a very small fraction of any of these genomes' total linear plasmid sequence is unrelated to sequence present somewhere in the other three sequenced *B. burgdorferi* genomes; we note that nearly all of the white regions in Figure 2 actually have homology to plasmid sequences in one of the other strains, but it is less than the 94% identical cutoff used in the figure. The following sections discuss the inter-strain relationships of each plasmid type in more detail.

Variation within each plasmid type in different *B. burgdorferi* isolates

cp9. B31 cp9 is not required for mouse or tick infection in the laboratory and is rather easily lost in culture [87–89]. Cp9 plasmids have been previously sequenced from *B. burgdorferi* strains B31 [11] and N40 [53] and *B. afzelii* strain IP21 [54] and found to be rather similar, but not identical in these three cases (see figure 6 in [77]). Of the new genomes compared here, only N40 carries a cp9 plasmid, and it is very similar to B31 cp9-1. Our 8722 bp N40 cp9 sequence has seven single bp differences from the previously reported N40 cp9 sequence [53]. These two plasmids have identical gene organizations, except that one putative gene (*b31_c10* or *revB*) and part of another gene (*b31_c08*) of B31 cp9-1 are replaced in N40 cp9 by several hundred bp of apparently non-protein-coding DNA (Figure S2B). The cognate B31 and N40 cp9-1 genes range from 73.8% to 99.1% identical. These plasmids carry no PFam32 gene, but they do encode PFam57 proteins that are about 90% identical, and members of this family have been shown to be required for proper plasmid replication/partitioning [90]. It is not known if these two plasmids are compatible.

cp26. Circular plasmid cp26 is universally present in *B. burgdorferi* isolates (*e. g.*, [27,55]), is required for virulence in mice [42,91], and is the only plasmid that is known to be essential for growth in culture [27,91–94]. The B31 cp26 carries genes involved in GMP synthesis [28], chitobiose import [30], host integrin binding [95], oligopeptide import [29], and the telomere hairpin formation [32,33]. It also encodes one of the important surface antigens expressed in the mammalian host, OspC protein [57,96–100]. The three additional complete cp26 sequences in the genomes analyzed here all have identical gene content and organization to the B31 cp26 reported by Fraser *et al.* [11] (Figure S2C). The *ospC* genes are especially variable and range from 82.6% to 86.5% identical in the four strains. If *ospC* is removed from the comparison, the remainder of the plasmids range from 98.4% to 99.2% identical in pair wise comparisons, close to the average similarity of the chromosomes (~99.4% identity, above). A more detailed analysis of single-nucleotide polymorphisms in cp26s will be reported elsewhere (E. Mongodin, W. Qiu, B. Luft, S. Schutzer, C. Fraser-Liggett and S. Casjens, unpublished).

cp32s. Members of this family of circular plasmids are present in all *B. burgdorferi* isolates analyzed to date and are thought to be prophages [39–41]. In addition to putative phage virion assembly genes, the cp32s carry genes that encode a number of proteins that have been studied [35,101]. These phage “lysogenic conversion”

(host modification) genes include the *rev* genes whose surface lipoprotein products bind fibronectin [102], the *mlp* encoded surface lipoproteins [103,104], the *bdr* (*Borrelia* direct repeat) genes whose functions are unknown, and the complex family of the *erp* (also called *ospEF* or *elp*) genes whose various members have been shown encode surface lipoproteins that bind to plasminogen [105], laminin [106] and factor H complement regulatory factor binding protein [107–109].

The sequenced genomes of B31, N40, JD1 and 297 contain nine, six, nine and nine members of the cp32 family, respectively. Like the B31 cp32s, the cp32 plasmids in the other three strains are homologous throughout nearly their entire lengths, and the 23 new complete cp32 sequences all have gene arrangements that are very similar to those of the B31 cp32s (ORF maps of these plasmids are shown in Figure S2D–G). Among the thirteen known cp32 compatibility types [84], only plasmids with cp32-9 partition genes are present in all four genomes (if the cp32-5 which was lost before the B31 genome was sequenced [10,12] is included, then it too was present in all four isolates).

The variations in overall organization of these cp32s include the integration of one into a linear plasmid in B31 [10], two are fused into one large “cp32-1+5” plasmid in JD1 that is made up of two different full-length cp32s fused together to form a 60.7 Kbp circular plasmid (this fusion of cp32-1 and cp32-5 plasmids was confirmed by Southern DNA restriction enzyme cleavage analysis, data not shown), two in N40 are truncated (cp32-4 and -7, which have approximately 14 and 13 Kbp deletions, respectively) and two in 297 are truncated (cp32-7 and -9, which both have ~9 Kbp deletions) (Figures S2E and G); several of these deletions have been noted previously, where N40 cp32-7 was called cp18, and 297 cp32-7 and cp32-9 were called cp18-1 and cp18-2, respectively [84,110,111]. In addition, there is an approximately 5.6 Kbp inversion in N40 cp32-5 (Figure S2E). The four deletions and the inversion affect only the putative virion assembly gene region of these prophage plasmids [40,41], so the plasmid partitioning and lysogenic conversion genes of these plasmids appear to remain intact. There are also a small number of other gene content differences among the cp32s, such as the presence or absence of a *revA* gene (*e. g.*, the complement of cp32s in JD1 carries no *revA* gene) and several variably present genes immediately transcriptionally downstream of the *erp* gene region. Only two putative gene types are present in the newly sequenced cp32s that do not have homologues on cp32 plasmids in the B31 genome; these are *297_w45* (a PFam55 gene; members of this family are present on four linear plasmids and cp9-1 in B31) in 297 cp32-11, and *jd1_q42* and *297_m41* found on JD1 cp32-10 and in 297 cp32-7, respectively (a fragment of this gene family lies at the same location in B31 cp32-3). In addition, more than one *mlp* gene is present in JD1 cp32-12 and 297 cp32-4. The *erp/elp/ospEF* gene group diversity will be discussed in detail elsewhere (B. Stevenson, B. Jutras & S. Casjens, unpublished). There appears to have been considerable homologous recombination among the cp32 plasmids, and this is discussed in more detail below.

lp5, lp21 and lp56. The N40, JD1 and 297 genome sequences do not contain plasmids with partition genes similar to B31 lp5, lp21 or the non-cp32-like portion of lp56. However, we have previously found that a 16 Kbp region that is very similar to B31 lp21 sequences is present at the right end of the strain 297 chromosome [51], and we note below that JD1 lp38 also carries a section that is very similar to a major part of B31 lp21.

lp17. Lp17's roles in pathogenesis are unclear, but it encodes protein D18 that regulates OspC expression from cp26 (above) [112]. There are three organizationally different versions of lp17 in the four genomes, the previously characterized 17 Kbp B31

plasmid, a 21 Kbp N40 plasmid, and 19 Kbp plasmids in JD1 and 297 (Figure 3). The rightmost ~13 Kbp regions of these four plasmids are >97% identical; however, they have three very different left ends as follows: (i) N40 lp17 carries four apparently intact genes that are >98% identical to B31 lp36 genes *b31_k45*, *b31_k46*, *b31_k47* and *b31_k50*; orthologs of these four genes are not present in the JD1 or 297 plasmid sequences, although other members of their paralogous families are present. Fikrig *et al.* [113] showed that the *n40_d02* (the *b31_k50* homolog on lp17) protein elicits protective immunity in mice. N40 does not contain orthologs of *b31_k48* and *b31_k49*, and it seems likely that the *b31_k48* and *b31_k49*-like genes were removed from the N40 lp17 by a homologous recombination event between *b31_k47* and *b31_k49*-like genes in a progenitor in which this region resembled B31 lp36 in organization. Xu *et al.*'s [114] PCR amplifications suggested that their set of OspC type E strains all have a similar lp17 arrangement to that of N40. (ii) To the left of their homology with B31 lp17, the very similar JD1 and 297 lp17s have a short 200 bp sequence that is about 84% identical to a fragment the PFam145 genes of the cp32 plasmids, about 900 bp that are not similar to any other sequence in these four genomes, and about 3.7 Kbp that are 94% identical to genes *b31_k13*, *b31_k15* and *b31_k17* (*adeC*) of B31 lp36. The *b31_k15* and *b31_k17* homologs in JD1 and 297 have frame-disrupting mutations, but there are intact versions of both on lp36 in JD1 and in 297. (iii) Finally, the

left end of B31 lp17 has 1496 bp that is unique to B31 and contains no gene of known function.

The 2–3% difference between the orthologous sequences on these plasmids (Figure 3) does not affect the reading frames of any of the putative lp17 genes; however, there have been a few small rearrangements in the different lineages. As an example of the types of such differences that are present between orthologous plasmid sequences in strains B31, N40, JD1 and 297 in general, these are indicated in Figure 3. They include deletions (identical 241 bp deletions relative to the other two at about bp 11100 in JD1 and 9100 in 297), inversions (of 259 bp at bp ~8000 in N40 and of 101 bp in the *d20* pseudogene in JD1), and differing numbers of the tandem 21 bp repeats present in *b31_d20* and its orthologs (8, 18, 8 and 13 copies in B31, N40, JD1 and 297, respectively). All of the repeats in N40 lp17 have the same two single bp differences relative to the other lp17 plasmids, perhaps implying substantial contraction and expansion of the array since the N40 sequence has been evolutionarily separated from the others.

lp25. This linear plasmid in strain B31 has been shown to be essential for infection of mice and ticks [81,89,115,116], and it harbors several important genes, *pncA* (*b31_e22*), which encodes a nicotinamidase [117], *bptA* (*b31_e16*, a surface lipoprotein that is required for persistence of strain 297 in the tick vector [82]), and *b31_e02*, a member of PFam01 that encodes a DNA restriction/

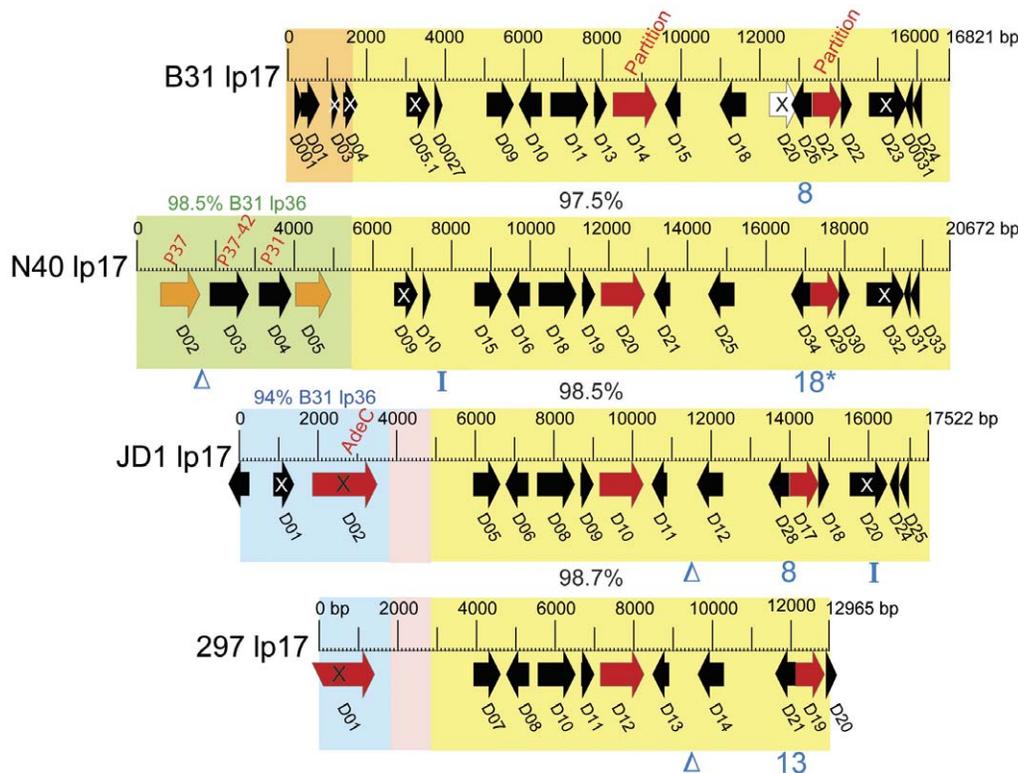


Figure 3. Organizational and open reading frame relationships among four lp17 plasmids. The four lp17s are aligned vertically, and identical colored background rectangles indicate very similar sequence (the 297 plasmid is the same size as JD1 lp17 but is missing several Kbp of sequence from its termini; see text and table S1). Rectangles of the same color denote homologous sequences, and the percentage nucleotide sequence identity of parallel yellow sections are shown between the maps. The arrows denote annotated predicted genes, where red arrows have a predicted function, black have unknown function, orange are known antigens, and white is a pseudogene not annotated except in the B31 plasmid; alternate gene names or predicted function are noted in red text in figure. An "X" indicates that a gene is truncated or has a frame disruption relative to a known homolog. The blue "Δ" indicates a short deletion relative to orthologous sequence in another lp17 plasmid(s); blue numbers indicate the number of short tandem repeats present at that location; an asterisk (*) notes that the repeat sequence is not identical to that of the other lp17s; an (I) marks the locations of short inversions relative to the other lp17s. doi:10.1371/journal.pone.0033280.g003

modification protein whose presence lowers the efficiency of genetic transformation of B31 cells [118–120]. N40 and JD1 lp25s each carry genes that are orthologs of *pncA* and *bptA*, and a homolog of *b31_e02*. Although infectious strain 297 carries an lp25-like plasmid [81,82], it was unfortunately lost from the culture whose genome was sequenced.

Figure 4 shows that the N40 lp25 is very similar to B31 lp25 with only a small number of short indel differences in regions that are not predicted to affect intact genes. The rightmost 18 Kbp regions of the two plasmids are over 98% identical, while the leftmost approximately 6 Kbp is about 91% identical, and a few hundred bp at the extreme left ends are unrelated in the B31 and N40 lp25 plasmids. JD1 lp25 represents a different subtype of this plasmid. The central 9 Kbp of JD1 lp25 is more divergent, with about 94% identity and two several hundred bp indels relative to the B31 and N40 lp25s. JD1's intact *bptA* and *pncA* genes lie in this region, and it also has a PFam01 restriction/modification gene at its left end and a PFam60 putative lipoprotein gene in the right central region that are about 90% identical to the orthologous genes in this position in B31 and N40 lp25. Finally, JD1 lp25 has several sections that are not present in the B31/N40 type lp25s; these “mosaic” patches have 98.6% and 98.8% identity to parts of 297 lp28-5 and B31 lp28-2, respectively, as well as weaker similarities to other *B. burgdorferi* plasmids (Figures 4 and S2H). These latter differences mean that JD1 carries no true ortholog to the antigenic B31 E09 (PFam44) putative lipoprotein, and carries only members of PFam52 (*jd1_e04*) and PFam102 (*jd1_e27*) that are rather divergent from their B31 counterparts.

lp28-1. The B31 plasmid lp28-1 has received considerable attention because it carries two genes, *arp* (*b31_f01*) and *vsE* (near the left and right ends, respectively), that are important in the mouse model of Lyme borreliosis. Loss of lp28-1 severely reduces strain B31 infectivity in mice but not in ticks [81,89,115,116,121–124], and antibodies against the B31 Arp protein cause resolution of *B. burgdorferi* induced arthritis in mice [125–127]. N40 does not carry a plasmid with an lp28-1 PFam32 gene, and the JD1 and 297 lp28-1 plasmids are very similar to each other (99.5% identical over the nearly 15 Kbp), but are quite different from B31 lp28-1. These two lp28-1 types only have the partition genes and *vsE/vsE* region in common (Figures 5 and S2I). Between these two regions of the JD1 and 297 plasmids lies about 2.6 Kbp of DNA that contains a PFam106 gene (*jd1_f23* and *297_f25*) that is homologous to B31 lp38 genes *b31_j23* and *b31_j24*. This JD1 protein is only about 40% identical to B31_J23 protein. At their left ends, JD1 and 297 lp28-1s have about 6 Kbp that is about 99% identical to the right end extension of the B31 chromosome (above) and which contains an apparently intact PFam138 gene (*b31_0852* in B31) and about 2.7 Kbp that has no ortholog in B31 and no convincing intact genes. The *arp* gene is not present on lp28-1 in the three new genomes discussed here. In N40 it is near the right end of lp28-5 (gene *n40_y16*), and in JD1 it is near the left end of lp28-4 (*jd1_i37*; and perhaps also in 297, although the sequence of the parallel region of its lp28-4 was not determined). The B31 and N40 Arp proteins are identical, and the JD1 homolog is 99.1% identical to them, so this movement of the *arp* gene among these different plasmids happened quite recently.

The *vsE* gene at the right end of B31 lp28-1 encodes a major outer surface protein and is unique in *Borrelia* in that during a mouse infection genetic information from fifteen tandem, unexpressed *vsE* cassettes can be copied into the *vsE* expression locus, presumably to present the host immune system with a “moving target” which can in theory have millions of different amino acid sequences [124,128–132]. The *vsE* gene was not present in the DNA libraries used in the original sequencing of the

strain B31 genome, and similarly was not present in any of the libraries used in sequencing the three genomes presented here. Zhang *et al.* [128] cloned the B31 *vsE* expression locus and showed that it lies very close to the right telomere of lp28-1 in strain B31. More recently, Hudson *et al.* [133] and Bykowski *et al.* [134] have shown that there are seven bp (5'-TTCTCTC; see accession No. DQ275473) between the bulk of the lp28-1 sequence (accession No. AE000794) and the *vsE* expression locus sequence (accession No. BBU76405). Possible *vsE* expression loci sequences for strains 297 and N40 have been PCR amplified (297, accession Nos. U76405 and AB011063; N40, X. Wang and J. Weis, personal communication), but attempts at PCR amplification between these sequences and the N40 and 297 cassette regions from N40 or 297 DNA were unsuccessful. However, similar PCR amplification using primers designed from the right end of the JD1 lp28-1 sequence cassette region and the reported 297 expression locus allowed extension of the JD1 lp28-1 sequence to include most of its *vsE* locus. The sequence of the *vsE*-*vsE* region will be examined in more detail elsewhere (S. Norris, D. Edmundson, T. Lin, G. Chaconas and S. Casjens, unpublished).

The *vsE* cassette region is also present on plasmids with lp28-1 compatibility in B31, JD1 and 297, but in N40 they reside on the plasmid that has an lp36 type PFam32 gene. In the N40 plasmid, the joint between B31 lp28-1 and B31 lp36-like sequences is close to the left end of the cassettes, so little other B31, JD1 or 297 lp28-1-like genetic material is present on N40 lp36. Like the parallel region in B31, the *vsE* cassette region of N40, JD1 and 297 all have extremely high (for *Borrelia*) G+C contents of about 50% (Figure 5), and B31, JD1 and N40 each have a dip to a much lower and more nearly normal *Borrelia* G+C content between the cassettes and the *vsE* expression site near the right end of the plasmid (the 297 sequence does not extend this close the plasmid's right end). The presence of this G+C dip and a lack of similarity with the *vsE* gene near the right end of the N40 lp36 sequence suggest that all of the N40 cassettes are likely represented in the reported sequence. Although the cassette regions are similar in size (about 8, 7.2 and 8 Kbp, respectively, in B31, N40 and JD1), there are significant differences between these cassette regions. B31 has 15 cassettes, JD1 has 14 and N40 has 19. The B31 and JD1 cassettes are quite constant in size (about 570 bp with a few that are up to 90 bp shorter; Figure S3). The N40 cassettes are somewhat more variable in size and range from about 200 to over 600 bp in length (average is 395 bp; Figure S3). The cassette sequences are much more divergent than all other clearly orthologous sequences in these isolates. The 6 Kbp of the JD1 and 297 cassette regions that are sequenced in both genomes contains eleven cassettes that are up to 93% identical, but numerous less similar sections make the whole region only about 81% identical between the two strains. All the other pair wise comparisons of the four cassette regions are less similar (*e. g.*, about 65% between B31 and JD1, based on alignments created by DNA Strider [135] and by inspection of diagonal matrix similarity plots, but it is difficult to obtain accurate sequence alignments). Thus the *vsE* cassettes are present in these four strains in three approximately equidistantly related versions represented by B31, N40 and JD1/297.

lp28-3. Plasmids with lp28-3 type PFam32 genes are present in B31, JD1 and 297. N40 has been reported to carry such a plasmid [78], but it was absent from the N40 culture whose DNA was sequenced. These three plasmids are very similar (over 99% identical in pair wise comparisons; Figure S2J), except in JD1 where about 1.2 Kbp at the right end contains a PFam52 gene (*jd1_h47*) that is not homologous to the parallel region of B31 lp28-3 (where a PFam48 gene lies in this region); the 297 lp28-3 sequence does not extend near enough to the right end, so it is not

B31 Ip25

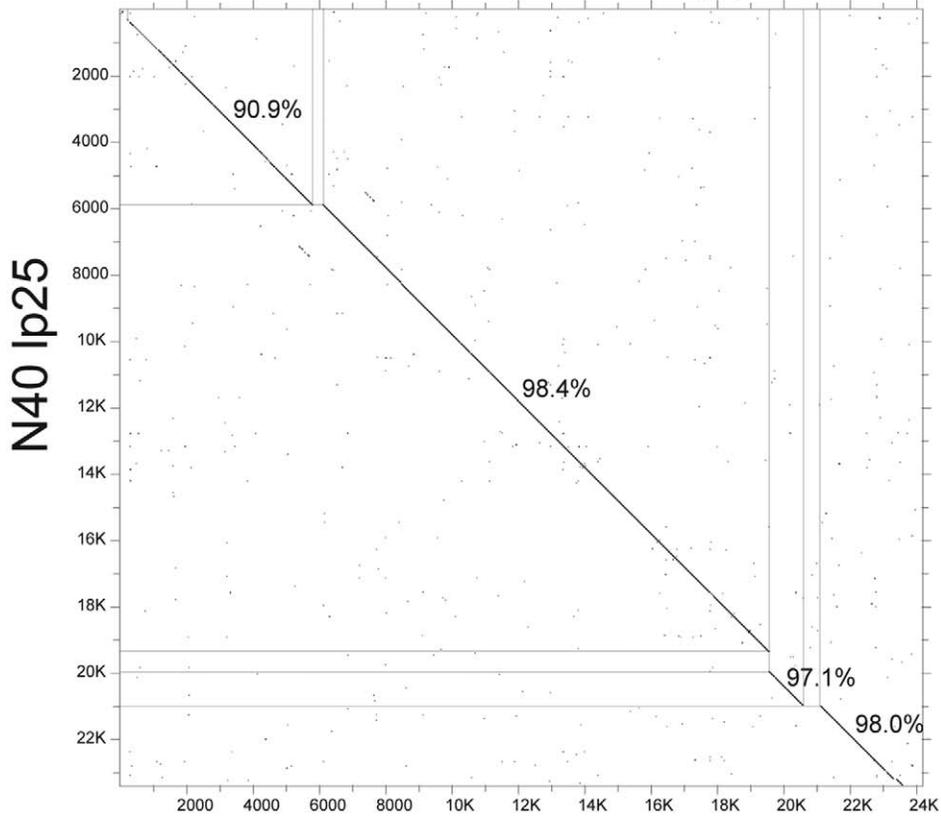
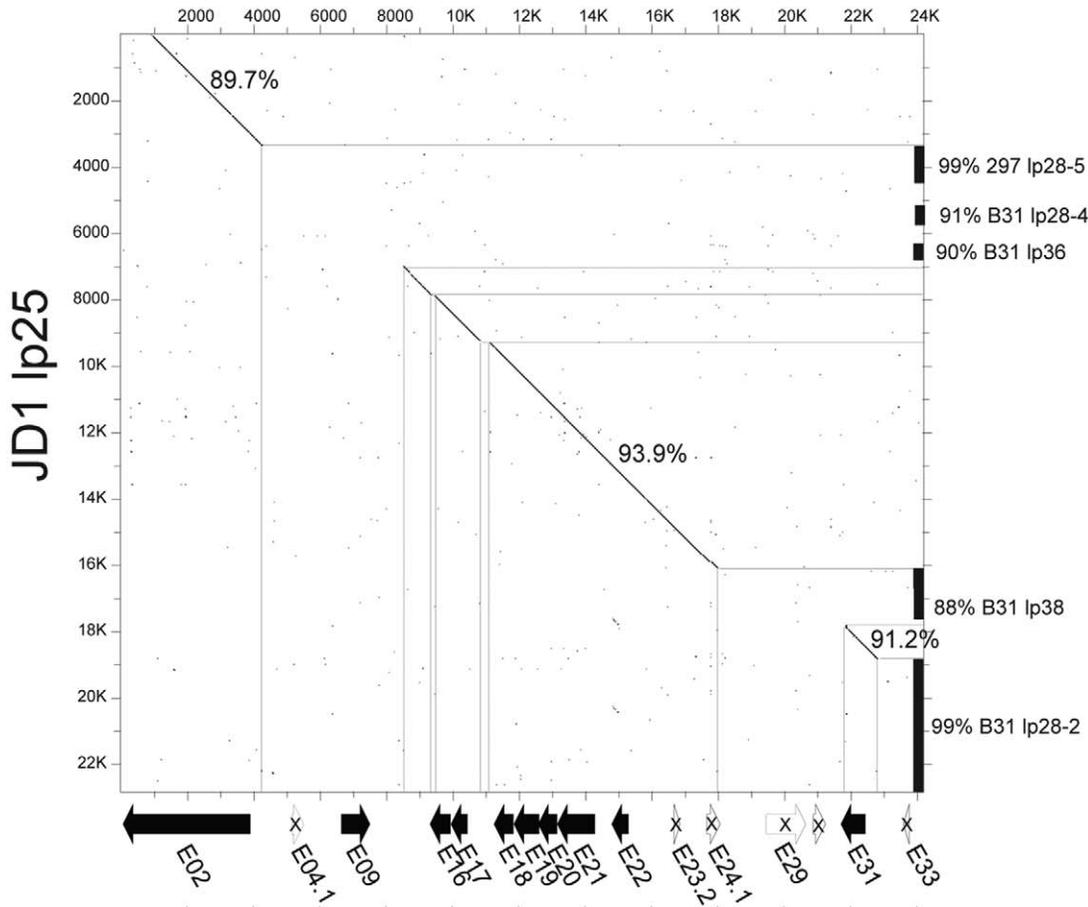


Figure 4. Comparison of three lp25 plasmids. Matrix plots with a 19 identities/23 bp window were created by DNA Strider [135]. Percent identities of nucleotide sequences are indicated near the diagonal identity line for most orthologous regions. The predicted genes for B31 lp25 are shown between the two plots (open arrows with "X"s are putative pseudogenes), and regions of high similarity to other plasmids are noted on the right. doi:10.1371/journal.pone.0033280.g004

known if it is similar to JD1 or B31 in this regard. The JD1 1.2 Kbp right end (above) is 99.5% identical to sequence at the right end of B31 lp36. The only B31 lp28-3 gene that has been studied, the *cspZ* gene (*b31_h06*) encodes the human factor H complement regulatory factor binding protein CRASP-2 [78,136] and lies in the common region of the three lp28-3s.

lp28-4. The B31 lp28-4 is required for the ability to infect the tick gut [116], and it carries several genes of current interest including *b31_i06* which encodes a surface localized nucleotidase [137], *b31_i16* (*vraA*; variable strain-associated repetitive antigen A) which confers partial protection as a vaccine [138], *b31_i26* which encodes a possible multidrug efflux protein, and *b31_i38* and *b31_i39* which encode PFam54 surface proteins [37,139]. The four lp28-4s are very similar (99.2 to 99.9% identical in pair wise comparisons) over most of their lengths, but the B31 plasmid has several Kbp of different, non-homologous DNA at both ends (Figure S2K). The right terminal 2 Kbp of the N40 and JD1 plasmids are nearly identical to each other and are 99.3% identical the left terminus of B31 lp36; this region harbors a PFam12 (*jd1_i36*) gene and a PFam01 fragment (*jd1_i47*) that are not present on B31 lp28-4. The leftmost 1.4 Kbp of B31 lp28-4 is replaced in the JD1 and N40 plasmids by different B31 lp28-1-like sequences; in the JD1 plasmid this DNA contains the *arp* gene (see above).

Although the central orthologous regions of the four lp28-4s are very similar, there are a few differences that could have significant effects on the function of possibly important genes on these plasmids. The *jd1_i19* (putative multidrug efflux pump) has an in-frame stop codon relative to the other three strains, and B31 and JD1 PFam60 orthologs *b31_i28* and *jd1_i22* have different single frameshifts relative to the cognate N40 and 297 genes. Finally, B31 lp28-4 has a set of three tandem paralogous PFam54 genes (*b31_i36*, *b31_i38* and *b31_i39*), while the other three lp28-4s have

two such genes in this position. Since the B31 genes *b31_i36* and *b31_i38* are 99.4% identical, it seems likely that these are the result of a recent gene duplication in the B31 lp28-4 lineage. Thus, although the orthologous lp28-4 regions are over 99% identical, several mutational changes have occurred in them that could have functional importance.

lp28-5. N40, JD1 and 297 each carry a plasmid with a previously unknown type of PFam32 gene that has been named lp28-5 (Figure 6). These represent a new *Borrelia* plasmid PFam32 compatibility type (see PFam32 tree in figure 3 of [75]). A substantial fraction of N40 lp28-5 is not closely related to any of the previously characterized B31 *B. burgdorferi* plasmids; nonetheless, it largely encodes more distant homologs of known B31 genes. The JD1 and 297 lp28-5 partition proteins are 94% identical to their N40 orthologs. Like most of the other linear plasmids, the three lp28-5s also have significant differences. The JD1 and 297 lp28-5 plasmids are quite similar to one another and carry genes that are mostly homologs of the B31 paralogous families (*e. g.*, PFams01, 12, 54, 60; Figure 6); however, they have only the partition gene cluster and one additional 3.4 Kbp section in common with N40 lp28-5. The common 3.4 Kbp regions (>97% identical) contain a PFam44 gene (*jd1_y11*, *297_y02* and *n40_y03*) and a fragment of a PFam01 gene. Each of the three lp28-5 plasmids also carries between 6 and 26 tandem repeats of a 133 bp sequence of unknown function (see below).

There appear to have been a number of recent inter-plasmid DNA transfer events that involved the lp28-5s. The rightmost 2.1 Kbp of N40 lp28-5 is 99.6% identical to the left end of B31 lp28-1 (and contains an *arp* gene; above), while the right ends of JD1 and 297 lp28-5s are different from that of N40 and one another; N40 lp28-5's best matches are 98.5% identical to the left end of B31 lp21. The JD1 lp28-5 leftmost 4.4 Kbp is 99.5% identical to a portion of the lp36 found in strain WI91-23 [44].

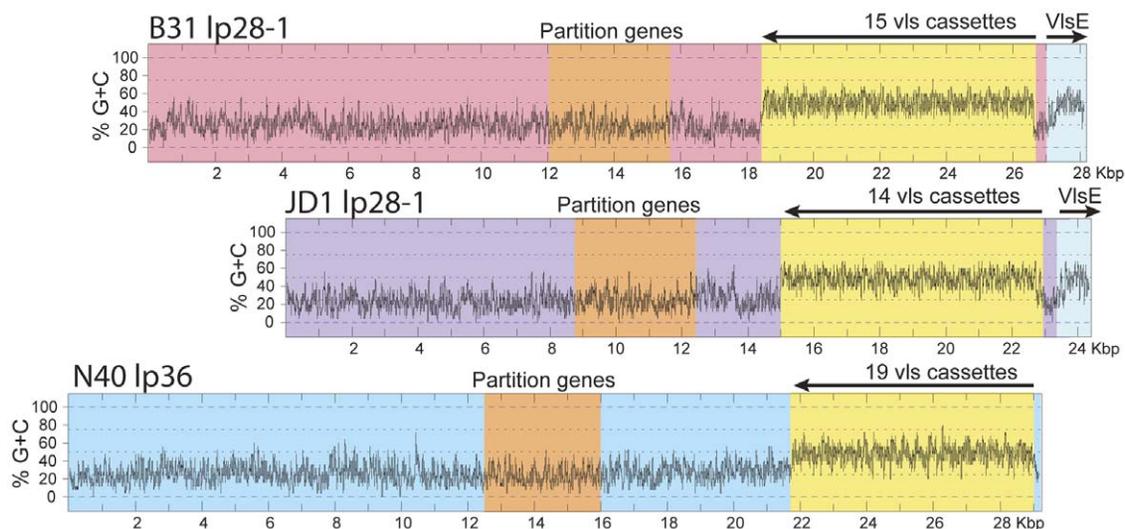


Figure 5. Comparison of lp28-1 plasmids and the vs cassette and vsE loci. Percent G+C plots for the plasmids were created by DNA Strider [135]. Different background color indicates very different sequence in the different plasmids (note that the partition gene regions in the two lp28-1 plasmids are homologous, but moderately divergent from those of lp36; see text). doi:10.1371/journal.pone.0033280.g005

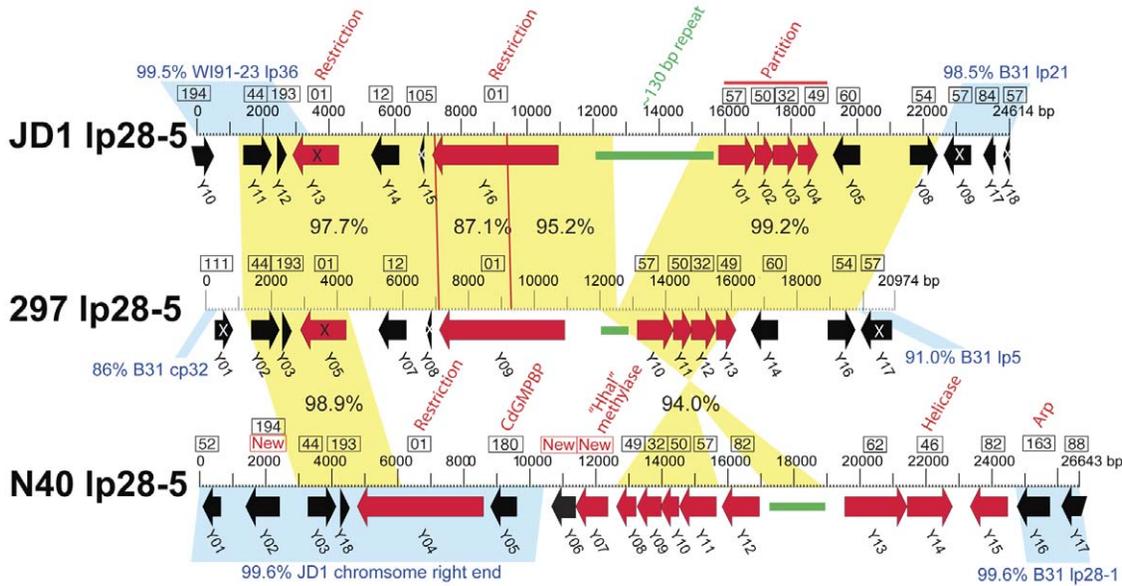


Figure 6. Organizational and open reading frame relationships among three lp28-5 plasmids. Maps are labeled as in Figure 3. Yellow background between maps joins regions of homologous sequence in adjacent maps; paralogous family numbers (table S2) are indicated in black boxes above each putative gene; red boxes marked “new” indicate genes for which there is no homolog in the strain B31 genome; green bars mark the 133 bp repeat regions (see text); CdgMPBP, cyclic-di-GMP binding protein. Blue background marks regions of high similarity to regions in other *B. burgdorferi* genomes.
doi:10.1371/journal.pone.0033280.g006

N40 lp28-5’s leftmost 10.5 Kbp is 99.6% identical to the right end of the JD1 chromosome (above), but the left ends the JD1 and 297 lp28-5s are not highly similar to any sequences in the other genomes compared here. The 297 lp28-5 left end contains a short ~200 bp section that is 86% identical to part of the *blyB* (*b31_r24*) gene of B31 cp32-4. Sequences that appear to have been transferred between *Borrelia*’s circular and linear plasmids are not common, but in addition to this instance, we find that JD1 lp28-7 (below) carries two *bapA/epaA* family (PFam95) genes that are typically found in some cp32s and cp9-1s.

N40 lp28-5 also carries the following predicted genes which have no B31, JD1 or 297 plasmid homologs: (i) *n40_y02* which encodes a novel putative lipoprotein, (ii) *n40_y05* which is predicted to encode a cyclic-di-GMP-binding protein [140,141] and is a homolog of a chromosomal gene that is present in all sequenced chromosomes (e.g., *b31_b0733/plzA*), (iii) *n40_y14* which is a putative helicase, (iv) *n40_y06* which has no known homologs, and (v) *n40_y07* whose predicted protein product is 48% identical in amino acid sequence to HhaI cytosine DNA methyltransferase encoded by *Haemophilus haemolyticus* ([142] and references therein). To test the predicted *in vivo* HhaI-like DNA methyltransferase activity of the N40_Y07 protein, we treated *B. burgdorferi* DNAs with restriction endonucleases SfoI and HhaI (New England Biolabs, Ipswich, MA), since HhaI cytosine methyltransferase methylates GCGC to create GC^{mC}GC, and DNA cleavage by both SfoI and HhaI are blocked by this methylation [143]. Endonuclease cleavage was monitored by CHEF pulsed field gel electrophoresis display of the resulting fragments. Figure 7 shows that SfoI fails to cleave N40 DNA, while it does cleave the DNA of strains M (above), B31, JD1 and 297 (data not shown for the last three). The same is true for restriction endonuclease HhaI (data not shown), suggesting that the *n40_y07* encodes an active cytosine DNA methyltransferase with the above specificity.

lp28-2, lp28-6 and lp28-7. Plasmids of these three compatibility types are considered together because, although

they comprise three different PFam32 types, they are otherwise quite similar (Figure S2L). Near the left end of all five of these plasmids are partition gene clusters whose PFam32 proteins are of three different types (see tree of PFam32 proteins in figure 3 of [75]). We named these two new types lp28-6 and lp28-7. All five of these plasmids have long central syntenic regions that include several genes with homology to bacteriophage virion assembly genes (typified by *b31_g21*, terminase; *b31_g20*, portal protein; *b31_g10* tail tape measure protein), suggesting that this region may be the virion assembly operon of a prophage [40,144]. B31 and N40 carry mostly syntenic lp28-2 plasmids that are 96.1% identical in their homologous regions but carry a few hundred bp of non-homologous sequences at their extreme left ends. The central 12 Kbp regions in the other three plasmids have >99% identity punctuated with shorter stretches of less similar but still homologous sequences. For example, JD1 and 297 lp28-6 central regions are perfectly syntenic and 99.5% identical except for a 1.8 Kbp patch of homology with about 60% identity (Figure S2L).

At the right ends of these five plasmids are 7 to 8 Kbp that are homologous, but are less similar than the central region; here the lp28-6s of JD1 and 297 are over 99% identical, but they are only 75–85% identical (with several indels) to the parallel portion of JD1 lp28-7 and the two lp28-2s in Figure S2L. These five plasmids carry three different types of lipoprotein genes at their extreme left ends; (i) B31 and N40 lp28-2 have PFam12 (e.g., *b31_g01*) and PFam102 (*b31_g02*) putative lipoprotein genes whose roles are not known. (ii) JD1 lp28-7 encodes two *bapA* (PFam95) proteins AA37 and AA38 that are about 50% identical to previously known members of this family [87,145,146]. (iii) The leftmost gene on JD1 lp28-6, *jd1_ε01*, is a rather distantly related homolog of the B31 *vsE* gene. It is less similar to the JD1 lp28-1 *vs* cassettes than is the JD1 *vsE* gene, so it seems that it has not recently procured genetic information from the *vs* cassettes (which are present on lp28-1 in JD1, above); it is about the same length as *vsE* and appears to have an intact lipidation amino acid sequence

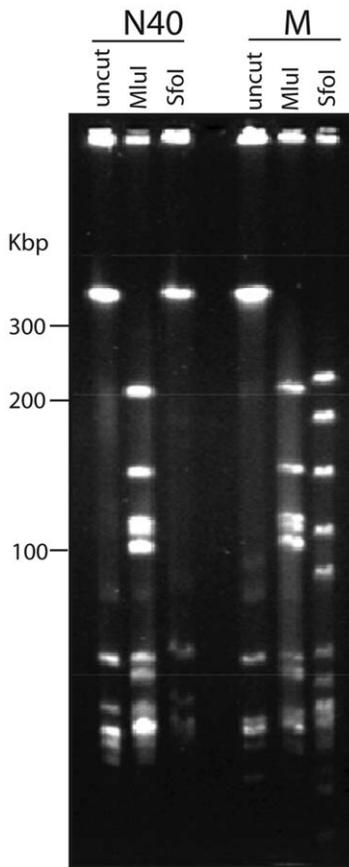


Figure 7. N40 DNA is not cut by restriction endonuclease *SfoI*.

DNAs were prepared in agarose blocks, cleaved with the indicated restriction endonuclease, and subjected to agarose gel pulsed-field agarose electrophoresis and stained with ethidium bromide as previously described [17,147]. Strain M is described in the text. Identical results to those with strain M were obtained with strains B31, JD1 and 297 (data not shown).

doi:10.1371/journal.pone.0033280.g007

consensus. The other three *B. burgdorferi* sequences do not contain an intact gene of this type (B31 and N40 have related *b31_j51* and *n40_j34* pseudogenes at their lp38 right termini); the terminal portion of the possibly syntenic lp28-6 plasmid in 297 was not sequenced.

A surprising feature of the JD1 genome sequence determination is that its lp28-6 was *very* highly represented in the DNA libraries used for sequencing. While the remainder of the genome coverage was in the 9- to 30-fold range (including the other plasmids), lp28-6 sequence coverage was 180-fold, indicating an approximately 10-fold higher copy number than the other *Borrelia* plasmids, which are present in the 1 to 3 per chromosome range when it has been measured [147,148]. This feature is not a general property of the lp28-6 type plasmids, since the very closely related 297 lp28-6 was not over-represented. Interestingly, the JD1 and 297 lp28-6s have extremely similar partition gene clusters; there are only three differences between the nucleotide sequences of this region, single bp differences just upstream of the PFam62 gene (*jd1_z06*) and in the PFam50 gene (*jd1_z05*) that do not change the amino acid sequence of the encoded proteins, and a different number of AAAGAA repeats within the PFam49 gene (Figure S4). There are six tandem copies of this sequence in the JD1 gene (*jd1_z03*) and eight in the 297 ortholog (*297_z01*). It is possible that the upstream

mutations affect expression levels or that the altered number of repeats (each of which encodes Lys-Glu) affects the function of the PFam49 protein to increase the copy number.

lp36. B31 lp36 carries two genes whose products have been studied, an adenine deaminase encoded by *b31_k17* (*adeC*) that is important in mouse infectivity [149] and a fibronectin binding protein encoded by *b31_k32* [150–154]; both of these are present on the lp36 plasmids of all four strains compared here. In addition, lp36 encoded lipoprotein K07 is an immunodominant antigen in human infection [155–157], and its gene is also present in all four lp36s; however, the N40 ortholog (*n40_k04*) has a frame disrupting mutation. B31 lp36 was previously known to be unusual in that DNA probes from it hybridize with plasmids that are particularly variable in size in different strains [21]. The lp36 plasmid sequences from B31, N40, JD1 and 297 were measured in agarose gels to be 36, 31.5, 24 and 24 Kbp long, respectively (Table S1; data not shown). The JD1 and 297 lp36's are 99.5% identical, with only one indel where 234 bp are missing at bp 16524 of the JD1 plasmid; no long ORF is affected by this indel (Figure 8). These two plasmids are shorter versions of B31 lp36 in which about 4 Kbp of the B31 plasmid's left end is replaced by about 1 Kbp of sequence that is 87% identical to the B31 lp28-4 left end, and sections of about 2 and 12.5 Kbp of B31 lp36 are not present in the JD1 and 297 plasmids (Figure 8); in the latter two, the 12.5 Kbp region is replaced by about 4.5 Kbp of which 3 Kbp is closely related to part of B31 lp28-1. N40 lp36 also has left end differences from the other three lp36 plasmids, and as discussed above, the N40 lp36 carries the *uls* cassette region (on lp28-1 plasmids in the other three strains) at its right end (Figure 8). Thus, the lp36 plasmids appear to have undergone several major rearrangements since their last common ancestor and are present as three quite different organizational subtypes in the four strains analyzed here.

lp38. Highly related linear plasmids in the 37 to 39 Kbp size range are present in 25 of the 56 *B. burgdorferi* sensu stricto isolates that have been examined [19,21,158,159], but it is not required for the tick-mouse infectious cycle in the laboratory [160]. Where it has been studied these ~38 Kbp plasmids carry parallel sets of genes (which include the outer surface protein OspD gene, *b31_j09* [21,158,159,161,162]). This suggested a potentially invariant organization for these plasmids [21]. B31 and N40 are among the strains previously known to carry a plasmid in this size range, and 297 and JD1 are among those known not have one [17,21,159], and the sequences reported here confirms those findings. The N40 genome contains a linear plasmid (37903 sequenced bp) that is very similar to B31 lp38 (Figure 9). The B31 and N40 lp38 plasmids are about 99% identical in nucleotide sequence, with only three indels larger than 25 bp as follows: (i) B31 is missing 351 bp at its bp 9571 within a PFam115 pseudogene (*b31_j15.1*), (ii) at bp 10293 of B31 lp38 there are 12 tandem repeats of the heptamer AATAGTT (between genes *b31_j15.1* and *b31_j16*), whereas in the N40 sequence it is repeated 119 times, and (iii) N40 is missing 1093 B31 bp at its bp 30326 that includes most of genes *b31_j41* and *b31_j42* (*b31_j41* is a PFam54 gene that has been shown to encode a membrane protein [37]).

Plasmids that carry lp38-related PFam32 partition genes are nonetheless present in the JD1 and 297 genomes, but these plasmids are quite different from those in B31 and N40 (Figure 9). Although the N40/B31 and JD1/297 lp38 type PFam32 protein sequences are robustly clustered, they form two distinct subgroups within this branch of the PFam32 tree (see figure 3 in [75]), and it is not known if these differences could give rise to compatibility differences. JD1 and 297 lp38's were measured by Southern

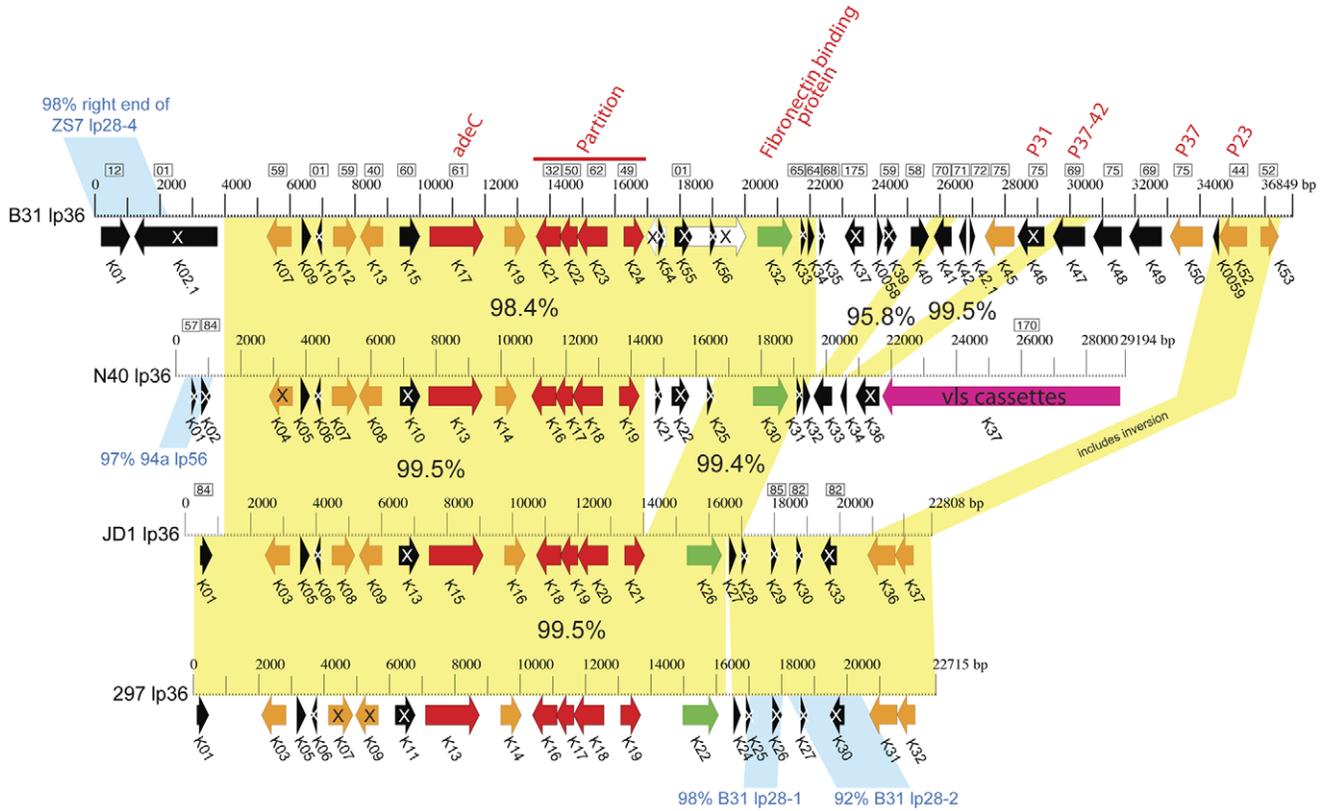


Figure 8. Organizational and open reading frame relationships among four lp36 plasmids. Maps of the four lp36s are labeled as described in Figure 6; green genes encode predicted and proven surface lipoproteins; red, plasmid partitioning and other DNA and nucleotide metabolism proteins; magenta, and *vls* cassette region (see text). Yellow shading between maps marks regions of nucleotide sequence similarity (percent identity values in black text). doi:10.1371/journal.pone.0033280.g008

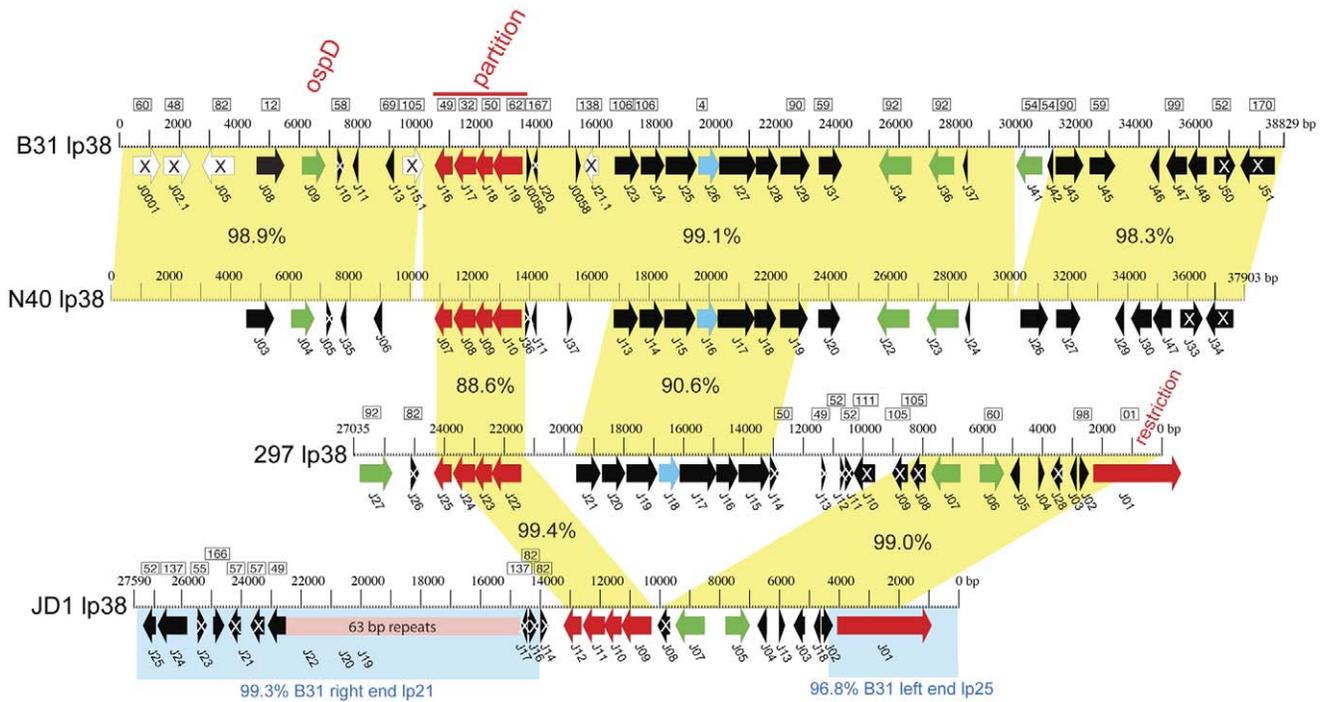


Figure 9. Organizational and open reading frame relationships among four lp38 plasmids. Maps of the four lp38s are labeled as described in Figure 8. Yellow shading between maps marks regions of high nucleotide sequence similarity (percent identity values in black text). The pink horizontal bar indicates the region of 63 bp repeats in JD1 lp38; and blue arrows represent predicted transporter genes. doi:10.1371/journal.pone.0033280.g009

pulsed-field gel analysis to be about 29 and 31 Kbp in length, respectively (data not shown), and both have rather complex relationships to other known *Borrelia* plasmids. In addition to nearly identical 3 Kbp regions that contain the four gene partitioning cluster, 297 and JD1 lp38s contain about 8 Kbp of common sequence; of this 8 Kbp about one third is 97% identical to the left end of B31 lp25 (Figure 9), and the remaining part has no very closely related homolog present elsewhere in these four genomes. The latter region contains (i) genes (*297_j06* and *jd1_j05*) that encode closely related PFam60 putative lipoproteins that are only ~55% identical to their closest relative in B31, Q05 protein, and (ii) genes *jd1_j07* and *297_j07* which have no B31 homolog and encode closely related putative lipoproteins that are ~75% identical to plasmid-encoded proteins PGP088 and BAPKO_6042 of *B. garinii* PBi and *B. afzelii* PKo, respectively [46–48]. In addition, JD1 lp38 contains about 13 Kbp at its left end that is >99% identical to the right half of B31 lp21 (including 7.8 Kbp of the lp21 63 bp tandem repeat tract). The 297 lp38 carries ~9 Kbp of DNA (in three patches) that has no very closely related sequence elsewhere in the four strains (Figure 9); however, it encodes, for example, a 71% identical homolog of the PFam92 *b31_j27*. Neither JD1 nor 297 carries (on lp38 or elsewhere in the genome) a homolog of the B31 *ospD* gene.

lp54. Previous studies have shown that plasmids similar in size and gene content to B31 lp54 are universally present in *B. burgdorferi* sensu lato isolates (*e. g.*, [17,19–22,55,92]). The B31, N40 and 297 lp54s are known to encode the well-studied outer surface proteins OspA and OspB, as well as decorin binding proteins, a thymidylate synthase, the complement factor H binding protein CRASP-1, and a number of less well-characterized surface and antigenic proteins (*e. g.*, [163–169]). Some of these have been found to be important in the tick-mouse infectious cycle [169–182]. The three new lp54 sequences all have gene contents that are nearly identical to B31 lp54 (Figure S2M), and have overall sequence identities between 98.9% and 99.4% (excluding the few differences discussed below). The only translational reading frame-breaking difference found among the genes of the four lp54s is a stop at codon 207 of the *ospB* gene in strain 297 (*297_a16*). Curiously, Probert *et al.* [183] found a stop codon in the 297 *ospB* at codon 199. Apparently these two subcultures of 297 have suffered independent *ospB* mutations. These may have been selected in different laboratory mouse passages, since OspB loss has been reported to be involved in interaction with host immunity [184].

Analysis of the lp54 sequences indicated that the homologs of *b31_a24* and the *b31_a68-b31_a70* cluster are the most variable parts of this plasmid [66] (in the lp54s of the four strains considered here, orthologous genes have identical locus_tag numbers). B31 lp54 carries a cluster of nine tandemly arranged PFam54 genes of which *b31_a64*, *b31_a65*, *b31_a66*, *b31_a68*, *b31_a69*, *b31_a70* and *b31_a73* are apparently intact, and *b31_a71* and *b31_a72* are truncated. Among these genes only the function of B31_A24 and B31_A68 (CRASP-1) proteins have been studied, and they are required during tick-to-mouse transmission [171,175] and bind human complement factor H [166], respectively. N40 and 297 lp54s are missing the *b31_a70* ortholog and have a novel member of the family (called *n40_a67.5* and *297_a67.5*, respectively) inserted into the cluster. JD1 is also missing a *b31_a70* ortholog but has no additional gene. These relationships and their evolutionary significance were discussed in more detail in Wywiał *et al.* [76]. The putative deletion that removed the *b31_a70* ortholog in JD1 has different endpoints from the one that removed it in N40 (which is identical to that of 297), suggesting that independent deletions occurred in these strains. The N40 deletion

appears to have been generated by a homologous recombination event between similar paralogous sequences but, curiously, the JD1 deletion appears to have been a non-homologous event (see below). The other highly sequence-variable location on the lp54s includes the region orthologous to the 3'-terminal portion of *b31_a24* (which encodes decorin binding protein B [185,186]) and the *b31_a23-b31_a24* intergenic region, where B31, JD1 and N40 have quite different sequences, but JD1 and 297 have more closely related sequences. This latter relationship is similar to the larger picture in which the 297 and JD1 linear plasmids are organizationally the most closely related strain pair (summarized in Figure 2); however, the close relationship between N40 and 297 in the *n40_a68-n40_a73* region does not agree with this history, and suggests that this rearrangement (if it happened only once in this exact fashion, which seems likely) has moved horizontally relative to other parts of lp54.

Plasmid genes and paralog families

Paralogous protein families. The sequence of the strain B31 genome indicated that its genome contains a large number of paralogous gene families that lie largely on the plasmids [10], and this is also true of strains 297, N40 and JD1. Our analysis of the PFams in each strain and, where possible, the orthology relationships among the strains within PFams are presented in table S2 (see Materials and Methods and legend of table S2 for methods used). This analysis points out that divergence and rearrangements can make the true inter-strain orthology relationships of genes on the *Borrelia* plasmids difficult to discern (*e. g.*, PFam01 example below).

This analysis found 160 paralogous families in the four genomes analyzed here (note that the PFam numbers go higher than 160 because some originally defined PFams have subsequently been merged under one number). The large majority of these PFams have members in all four strains. Of the 109 plasmid gene-containing PFams, only seventeen do not have representatives present in all of the four genome sequences compared here; and five of these (PFam65, 76, 78, 102 and 192) are likely explained by the missing lp25 sequence and terminal plasmid sequences in strain 297 (above), two (PFam72 and 175) are represented only by putative pseudogenes in B31, and PFam137 has its 297 member near the right end of the chromosome [51]. Thus representatives of only nine intact gene-containing families (PFam63, 68, 70, 71, 76, 88, 90, 193 and 194) appear to legitimately be missing from one or more of the four genomes, and these are largely the result of the failure of B31 to carry an lp28-5 plasmid and the large differences among the lp28-1s and lp36s of the different strains (Figures S2I and 8). Thus, JD1 carries no PFam63 gene (*revA*, normally on a cp32 plasmid in the other strains) and no PFam90 (on lp38 in the other strains) genes, and B31 has no PFam193 or 194 genes (on lp28-5 in other strains). These four PFams contain relatively large and so likely genuine genes; however, their roles have not been studied, except for PFam63 [102]. Thus, in spite of the plasmid content differences and numerous plasmid rearrangements discussed above, there are relatively few examples of PFams that are not present in all four strains, and the genome contents are in fact rather constant.

The number of paralogs present in individual PFams are, however, often variable (table S2); for example PFam01 contains restriction proteins and several related pseudogenes, and B31, 297, JD1 and N40 harbor at least 2, 2, 4 and 2 *apparently* intact members (297 and N40 might each have an additional intact member on their unsequenced lp25 and lp28-3, respectively). These proteins form two major sequence types (JD1_Y04/JD1_0905 and the other eight) that are about 18% different. In

the larger group, B31_E02, B31_H09 (and very similar JD1_H09 and 297_H03), 297_Y09, JD1_E01, JD1_Y16, and N40_E01 proteins all differ from one another by $11 \pm 2\%$. It is not known if such sequence differences could cause restriction target specificity differences. The proteins B31_H09, JD1_H09 and 297_H03, which are encoded at identical positions on lp28-3s, are nearly identical; however, members within the two other syntenic groups B31_E02, JD1_E01 and N40_E01 on lp25 (Figure S2H) and JD1_Y16 and 297_Y09 on lp28-5 (Figures 6, S2H and S2J) differ by about 10%. These differences among syntenic orthologs suggest that there has been (presumably homologous) recombinational “shuffling” of sequences among the PFam01 paralogs.

Another family of note is PFam82, a putative IS605 type transposase [10,187]. There are numerous PFam82 fragments in all four of the genomes compared here. In B31, 297 and JD1 no PFam82 member appears to be intact, but the N40 genome contains several *apparently* intact transposase genes (*n40_g05*, *n40_y12* and *n40_y15*). Of these, *n40_g05* is an ortholog of the frameshifted *b31_g05* gene, and *n40_y12* appears to have recently hopped into the partition gene cluster since the separation of the N40 lp28-5 from its common ancestor with JD1 and 297 (Figure 6).

Novel plasmid-encoded proteins. Only a few percent of the plasmid sequences of B31, 297, N40 and JD1 have no recognizable homologs in the other three strains, and as noted above only very few previously unknown *B. burgdorferi* plasmid gene types were identified in the three new sequences. This small number of “new” genes that encode previously unknown proteins are (i) *n40_y07* which encodes a DNA restriction methylase (above), (ii) homologs *n40_y02*, *jd1_y10* and *jd1_0909* that encode PFam194 putative lipoproteins, (iii) *n40_y06* with no predicted function, (iv) orthologs *297_j07* and *jd1_j07* that encode a putative lipoprotein of unknown function that have plasmid-borne homologs in the *B. afzelii* and *B. garinii* genomes (above), and (v) *jd1_z02* which encodes a putative PFam68 lipoprotein which has only apparently disrupted homologs in B31 and 297 and no homolog in N40.

Plasmid DNA rearrangements

Clearly the patchwork of sequences with near identity to other sequences in the *B. burgdorferi* linear plasmids (Figures 2 and S2) indicates that there have been rearrangements sufficiently recently that there has been little time for extensive divergence of the sequences involved. Non-homologous rearrangements create novel sequence joints and thus novel juxtapositions of genes, and new combinations of such novel sequence joints can be created by homologous recombination among plasmid patches with similar sequence.

Non-homologous recombination. When mosaic sequences were analyzed *within* strain B31, we concluded that those rearrangements were most likely generated by non-homologous recombination [10,38]. This conclusion is strongly reinforced by comparison of the plasmids in the four strains in this study. For example, near its right end JD1 lp28-3 has a novel sequence joint compared to B31 lp28-3. It joins sequences that are 99.6% identical to B31 lp28-3 and 99.5% identical to B31 lp36 (Figure 8). If it is assumed that the direction of this rearrangement is the formation of JD1 lp28-3 by a recombination between parents similar to these two B31 plasmids, the recombination point can be deduced to be at an exact bp in both putative parents, and there is no sequence similarity at all at this location in the putative parental plasmids (Figure S5A); also shown in Figures S5B–D are three other typical examples of such apparently non-homologous recombination events, an inversion in N40 cp32-5, the creation of a novel joint in N40 lp17 by recombination between B31 lp17-

like and lp36-like plasmids, and the putative deletion that removed the putative *B31_a70* ortholog from JD1 lp54 (above). Thus, many of the rearrangements that gave rise to the organizational differences among the plasmids of the four strains appear to be the consequence of such non-homologous recombination.

Homologous recombination. Homologous recombination could occur between any highly similar plasmid sequences, and the event that likely deleted the *n40_a70* gene from the N40 lp54 is shown in Figure S5E. But in most cases homologous recombination can only be recognized by reassortment of outside markers. The group of lp28-2, -6 and -7 linear plasmids shows a particularly clear example of such reassortment of outside markers. B31 lp28-2, N40 lp28-2 and JD1 lp28-6, have PFam101 *b31_g10* and cognate homologs that are locally syntenic and >98% identical, and plasmids 297 lp28-6 and JD1 lp28-7 have parallel PFam101 genes (*297_z06* and *jd1_aa10*) that are 99.8% identical. However, the first three genes contain nearly identical internal 1300 bp patches that are only ~60% identical to the second two (which are also nearly identical). Figure 10 shows this relationship (within gene *jd1_aa10*) between JD1 lp28-6 and lp28-7. The most likely explanation for the existence of such abrupt changes in relatedness within similar sequences is recombination between two homologous but diverged sequences. Thus, the five parallel PFam101 genes are present as two “sequence types”, called A and B, in the five related plasmids (Figure S2L). Another example of such a relationship on these same plasmids is the 99.5% identical right end regions of JD1 lp28-6 and 297 lp28-6 (*jd1_z24-jd1_z28* and parallel 297 region) that are only about 75% identical to a homologous but divergent version on the other three plasmids (the two versions are called C and D in Figure S2L). N40 lp28-2, 297 lp28-6, JD1 lp28-6 and JD1 lp28-7 carry all four of the possible combinations of these sequence type alleles, AC, BD, AD, and BC, respectively. (Other, probably non-homologous, recombination events have occurred nearer to the left ends of these plasmids, but this does not impact these conclusions). It is extremely unlikely that localized mutational divergence can account for such patches of different but uniform relatedness, so two A alleles, for example, cannot be similar by virtue of separate but parallel divergence from a B ancestor. The presence of all four allele combinations in such a situation cannot arise through simple linear evolutionary descent [188], and at least one of them must have arisen by a recombination event between the other combinations. This event was almost certainly homologous recombination within the 12 Kbp of very highly similar (all $\geq 98.7\%$ identical in pair wise comparisons) sequence between the A/B and C/D loci.

Homologous recombination has also occurred among the circular plasmids, and the cp32s are a robust example of such reassortment. Comparison of their sequences indicates that the most variable genes on the largely homologous cp32 plasmids lie in four regions as follows (see Figures S6A and S6B): region 1, *b31_s27*, *rev*, *mlp* and *bdr* genes [110,189–195]; region 2, the partition genes including the PFam32 genes (*e. g.*, [75]); region 3, the *erp/elp/ospEF/p21* type surface protein genes [84]; and region 4, several alternative genes of unknown function immediately to the right of region 3 [145,146,196]. Each region’s contents can be parsed into a small number of robust sequence types (alleles) by neighbor-joining tree analysis, and trees of each of the encoded homologous proteins are shown in Figure S6C–F. In most cases these sequence types are unambiguously very different from one another, for example the three Mlp protein types are more than 50% different from each other on branches with very high bootstrap support, as are the three PFam114 types (Figure S6C and F, and see figure 4 of Casjens *et al.* [75] for the PFam32

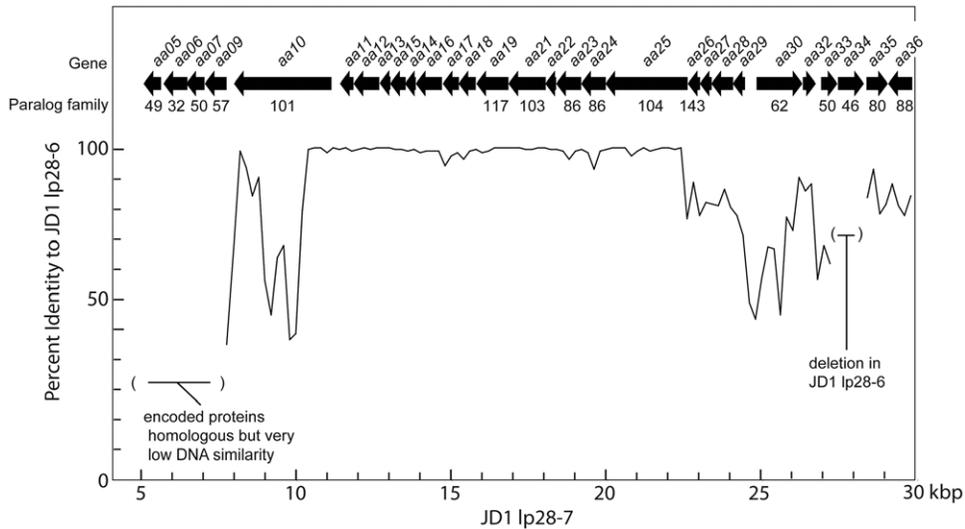


Figure 10. Comparison of JD1 and 297 lp28-6 plasmids. The DNAs of JD1 plasmids lp28-6 and lp28-7 were aligned with DNA strider [135], and the percent identity was computed for sequential 200 bp windows across the region shown in the figure. These two plasmids have two regions of near identity from about 8 to 9 Kbp and 10.5 to 22.5 Kbp, which abut regions with lower similarity. The JD1 lp28-7 open reading frames and their paralogous family relationships are shown above.

doi:10.1371/journal.pone.0033280.g010

partition protein tree). Although it does not impact the conclusions drawn here, the evolution of the *erp* genes is oversimplified in this type of analysis, and will be discussed in more detail in a future publication (B. Stevenson, B. Jutras and S. Casjens, unpublished).

For each gene in the cp32 variable regions, the sequence types thus determined were assigned arbitrary numbers (see Figure S6C–F), and the findings are summarized in table 3, which lists the allele types present at each of the four variable regions on the thirty-two different cp32 sequences present in the four genomes (counting the fused “dimer” cp32 in JD1 as two such sequences arbitrarily divided as shown in Figure S2F). The sequence alleles present in these four regions of the cp32 plasmids have clearly not diversified only by linear evolutionary descent, and considerable reassortment of these different alleles is required to explain their current distribution. In fact, among these thirty-two cp32s, no two plasmids have the same constellation of variable region alleles. For example, the N40, JD1 and 297 cp32-12’s each have the same region 2 PFam32 compatibility allele (by definition), but carry three different sets of alleles at each of the other three variable regions. On the other hand, although movement toward randomization has occurred, the alleles at these four variable positions appear not to have been shuffled to the point of complete loss of linkage in all cases. For example all four cp32-9s have the same alleles in region 3. In addition, many of the genes *within* the variable regions also appear to have been re-assorted, for example, *mlp* allele 2 is found in association with all four *bdr* alleles in region 1. Because the cp32 plasmids retain the same gene order and the vast majority of their genes have not been broken by non-homologous recombination events, this reassortment almost certainly occurred through numerous homologous recombination events.

Mutational decay and pseudogenes

The constant regions of the *Borrelia* chromosomes have very few obvious pseudogenes (unlike the plasmids and plasmid-like right end chromosomal extensions). However, our previous analysis of the B31 plasmids identified over 150 pseudogenes as regions of nucleotide sequence similarity to intact genes but whose reading

frame is disrupted and/or truncated [10]. A large majority of these pseudogenes reside on the “rapidly evolving”/organizationally variable subset of the linear plasmids (probably all linear plasmids except lp54), most of which also carry extensive regions that contain no long open reading frames. The latter appear to represent decaying, useless DNA [38]. Many of these B31 non-coding regions are present in extremely similar orthologous sequences of the plasmids of the other three strains. A typical example of this resides on the lp28-4 plasmids. Here the three largest tracts of apparently non-coding DNA (including >6 Kbp that includes one recognizable ~650 bp degenerate pseudogene *b31_i08.1*) are all present in all four strains and more than 99% identical to one another (Figure S7). Here and elsewhere, the B31, N40, JD1 and 297 linear plasmids carry many of the same pseudogenes that have changed little since their last common ancestor. In addition, the linear plasmid sequences present in N40, JD1 and 297 that have no clear orthologs in B31 have similar densities of apparent pseudogenes; for example, Figure 9 shows that the regions of 297 lp38 that have no B31 orthology harbor a number of pseudogenes (*e. g.*, *297_j08-j14*) which are related to, but not orthologous to other known *Borrelia* genes. In a second example, the JD1 and 297 lp28-5s, which are not present in B31, carry substantial regions of apparently noncoding DNA, which include apparently truncated genes (*e. g.*, *jd1_y13* and *297_y05*) and regions that have no convincing open reading frames (*e. g.*, ~1.5 Kbp between *jd1_y05* and *jd1_y08* and orthologous DNA in 297; Figure 6).

Mutational inactivation of plasmid genes is apparently ongoing, since among the strains studied here there are a number of plasmid orthologs in which only a subset is disrupted. Some examples are as follows:

(i) Mutations in the JD1 lineage. The putative lipoprotein-encoding gene *jd1_h20* in plasmid lp28-3 has a frame-breaking insertion of a single T at 12731 relative to its orthologs in B31 (*b31_h18*) and 297 (*297_h12*); these three orthologs are >99% identical. Also in JD1, the homologs of *b31_i26* and *b31_i34* have an in-frame stop codon and frameshift, respectively, compared to the orthologous regions of the other three lp28-4s (Figure S7).

Table 3. Homologous recombination among cp32 plasmids.

strain	Region 1 ^a				Region 2 ^a	Region 3 ^a		Region 4 ^a		
	S27	Rev	Bdr	Mlp	cp32	A	B	bapA	PF114	other
297	-	-	1	1	1	4	1	-	1	-
297	1	-	3	2	3	2	-	-	3	-
297	-	1	-	1+2	4	3	-	-	-	1
297	-	-	3	2	5	4	1	-	1	-
297	-	-	3	1	6	2	-	1	-	4
297	-	-	1	2	7	3	-	-	-	5
297	-	-	2	1	9	4	1	-	-	-
297	-	-	1	1	11	2	-	-	-	3
297	-	1	-	2	12	4	1	-	1	-
JD1	1	-	1	2	1	2	-	-	-	3
JD1	-	-	3	2	3	2	-	-	3	-
JD1	-	-	1	2	5	1	-	-	1	-
JD1	-	-	3	2	6	4	2	-	3	-
JD1	-	-	3	2	8	4	1	-	3	-
JD1	-	-	2	2	9	4	1	-	3	-
JD1	-	-	3	2	10	2	-	1	-	4
JD1	-	-	1	1	11	2	-	-	3	-
JD1	-	-	1	2	12	3	-	-	-	1
B31	-	1	-	2	1	4	1	-	1	-
B31	1	-	1	2	3	2	-	1	-	-
B31	-	-	4	2	4	4	2	-	3	-
B31	-	1	-	2	6	2	-	-	-	2
B31	-	-	1	2	7	2	1	-	3	-
B31	-	-	1	2	8	4	1	-	1	-
B31	-	-	2	2	9	4	1	-	2	-
B31	-	-	3	1	10	1	-	-	1	-
N40	-	-	1	1	4	2	1	-	-	-
N40	-	-	1	2	5	2	-	-	-	2
N40	-	-	1	2	7	4	2	-	3	-
N40	-	-	4	1	9	4	1	-	1	-
N40	-	1	-	1	10	1	-	-	3	-
N40	-	-	-	-	12	2	-	-	-	-

^aClustalX2 0.3 [214] was used to create a neighbor-joining tree for proteins encoded in the genes the four cp32 “variable regions”. The variable regions were identified by inspection of matrix comparison plots (Figure S6A), and their positions are shown in Figure S6B. Robust branches (amino acid “sequence types”) were identified for the proteins encoded by the genes in each region, and these different protein types were given arbitrary numbers. The trees and definitions of sequence types are shown in figures S6C–F. Where a region 3 has two genes, the left (transcriptionally upstream) gene is in column “a” and the right gene is in column “b”. A dash “-” indicates no gene is present; PF, paralogous family. The genes in the rightmost region 4 “other” position are as follows: 1, *mip* type 3 gene; 2, homolog of *b31_m39*; 3, homologs of *297_w43* (very weak similarity to *b31_m39*) and *297_w44* (PFam55); 4, homolog of *297_m41*; 5, unique *297_o29*. The different B31_S27, Rev and BapA proteins are all so similar that subtypes were not evident.
doi:10.1371/journal.pone.0033280.t003

(ii) Mutations in the B31 lineage. The lp28-2 genes *b31_g03* (PFam48) and *b31_g05* (PFam82; putative transposase) were originally suggested to be pseudogenes that contain single frameshifting mutations on the basis of comparison with paralogs in strain B31 and/or genes in other bacteria [10,38]. The reading

frames of orthologs of both of these genes (*n40_g03* and *n40_g05*) appear to be intact in N40, confirming this interpretation. The nucleotide changes that shifted the frames in these two genes in B31 are a tandem duplication of a TGGAG (N40 bp 2327-2331) and run of 4 T’s in N40 is lengthened to 5 T’s in B31 (B31 bp 3675), respectively.

(iii) Mutations in the N40 lineage. The lp28-2 genes *n40_g17* and *n40_g22* have frameshifting mutations relative to their B31 orthologs (Figure S2L). The fact that paralogous sequence on JD1 lp28-6, 297 lp28-6, and parallel *B. garinii* PBI (accession No. AY722917), and *B. bissettii* DN127 (accession No. CP002760) plasmids carry the longer form of both genes strongly suggests that these N40 genes are inactivated by mutation. Here a run of 8 T’s in B31 (B31 bp 15923-30) is shortened to 7 in N40 and a run of 4 C’s in B31 (B31 bp 20677-80) is lengthened to 5 in N40, respectively.

(iv) Mutations in the JD1 and 297 lineages. JD1 and 297 carry a version of lp17 that has sequences similar to B31 lp36 at one end (above). This region carries *adeC* (*b31_k17*) and *b31_k15* homologs that contain frameshift mutations relative to the lp36 homologs that are present in all four genomes. The JD1 and 297 lp17 *adeC* genes (*jd1_d02* and *297_d01*) are 99.3% identical to one another, and their sets of disrupting mutations are overlapping but not identical, suggesting that their decay began before the separation of the lp17 plasmids in the JD1 and 297 lineages and has continued since that time. In these lp17 *adeC* pseudogenes, a run of six A’s is extended to seven at position 1111 of JD1 lp17, a run of nine T’s in the intact genes is shortened to eight in both genes (at 2163 in JD1), and a run of two C’s in the intact genes is three and four long in 297 and JD1, respectively (at 3049 in JD1). Many of the recent, frame-disrupting changes in plasmid genes appear to be slippage mutations where runs of a single nucleotide are shortened or lengthened.

Tandem direct repeat sequences

Since they can expand and contract relatively rapidly on an evolutionary scale, the number of repeats in “variable number tandem repeat” (VNTR) tracts is often used for separation and identification of closely related lineages of any organism, and bacteria are no exception [197]. Such methods have been applied to other *Borrelia* species [198,199] and such repeat tracts have been previously noted on lp21 (63 bp repeat) [10], lp38 (17 bp repeat) [158,159] and lp28-4 (27 bp repeat in *vraA* gene) [138]. In addition, members of the *bdr* gene family (PFam80) contain more complex imperfect direct repeats [191,200]. Although Zuckert and Barbour [192] did not observe changes in the number of repeats in the *bdrT* (*b31_g33* on lp28-2) gene they examined during growth in culture, we find, for example, that the *bdrT* gene contains about 9 repeats in B31 and its otherwise 99.4% identical ortholog in N40 has about 6 repeats (imprecision is due to the presence of partial and overlapping repeats). So strains from different rRNA [57–60] or OspC [62–65] lineages can have different numbers of *bdr* gene repeats.

Tandem Repeats Finder ([201] and <http://tandem.bu.edu/trf/trf.html>) was used to identify tandem repeats of short sequences in B31, N40, JD1 and 297, and table S3 lists twelve such tracts that show substantial variation among these four strains. These include repeat tracts within three chromosomal genes, *b31_0210*, *b31_0546* and *b31_0801* and their orthologs in the other strains, as well as tracts on lp17, lp28-4 and lp54, all of which are present in essentially all *B. burgdorferi* strains tested and so could be useful in tracking closely related members of this species. Of these chromosomal genes, *b31_0210* (*lmp1*) is required for persistence in murine tissues [202], and *b31_0801* is predicted to encode a

translational initiation factor. The B31 *vraA* (*b31_i16*) lp28-4 gene encodes a surface lipoprotein [138] that contains 21 perfect repeats of the highly charged nine amino acid sequence KKKQEEEL; the *vraA* gene is a member of PFam60, but the other members of this family do not contain this nine amino acid sequence. The N40, JD1 and 297 *vraA* orthologs have 31, 24 and 35 repeats, respectively. These strains belong to different rRNA and OspC lineages of *B. burgdorferi*; however, the *vraA* genes in two other strains from the B31 rRNA lineage 1 (ZS7 and Bol26) and one from its sister rRNA lineage 3 (strain 64a) contain 3, 1 and 11 repeats, respectively [44], showing that the length of at least this VNTR can apparently change quite rapidly.

Measurement of precise lengths of very long VNTRs is not simple, so they are likely to be less useful in lineage tracking, but strain B31 plasmid lp21, the homologous chromosome region in strain 297, and JD1 lp38 carry about 11, 11 and 8 Kbp of imperfect 63 bp tandem repeats [10,51]. The lp28-5 plasmids carry between 5 and 26 imperfect repeats of an unrelated 133 bp sequence. All six translational reading frames are blocked in each of the 133 bp repeats. Since accurate assembly of sequencing runs in such regions is difficult, we experimentally confirmed the approximate length of the block of tandem 133 bp direct repeats to about 1.8 Kbp in N40 lp28-5 by measuring the size of DNA restriction fragments that contain mostly the repeat region (data not shown). The roles of such long repeat tracts are unknown, and they are quite unusual in prokaryote genomes.

Conclusions

Overall genome relationships. The *B. burgdorferi* genome contains an approximately 903 Kbp chromosome “constant region” and plasmids cp26 and lp54, which are quite evolutionarily stable. Comparisons among the four genome sequences analyzed in this study show that these three replicons are nearly completely syntenic and more than 98% identical in nucleotide sequence among strains from different rRNA/OspC lineages. In addition to these highly conserved regions, the genomes of this species also contain a large number of much more variable plasmids, and the majority of *B. burgdorferi* isolates carry 7–20 Kbp of variable, plasmid-like sequences at the right end of the otherwise genetically stable chromosome. Much of this more variable portion of the genome is also very highly related in sequence among strains, but it has suffered numerous rearrangements. For example, the orthologous parts of variable plasmids lp28-3, lp28-4 and lp36 (which constitute the majority of the sequence of each of these plasmids) are each $\geq 99\%$ identical among the cognate plasmids in the four strains. In most of the rearrangements found in the plasmids, non-homologous DNA has apparently replaced previously existing sequences. The very high identity ($>99\%$) of the sequences in two or more related but rearranged plasmid versions indicates that these rearrangements happened rather recently on an evolutionary time scale.

Some of the variable regions appear to have suffered multiple sequential or parallel replacements by different non-homologous plasmid sequences. For example, (i) three lp17s have three different left ends but have the same right end; (ii) three lp28-4s have two different right ends and three different left ends; (iii) the three chromosomes with right end extensions have three largely non-homologous extensions; and (iv) the *arp* gene and *vs* cassette region each lie on several different plasmid types in the different genomes. These and other observations strongly indicate that the linear plasmid rearrangement process is ongoing, and that such events have happened independently in different *B. burgdorferi* lineages. Yet, in spite of the many organizational and plasmid content differences, the gene content of the four strains remains

relatively constant. Although there is some variation in the number of members of the different paralogous gene families each strain carries, a large majority of such families are represented in all four strains. In addition, only a few “new” previously unknown *B. burgdorferi* gene types were identified in the three new genome sequences.

Rates of horizontal exchange?. The presence of such a large number of plasmids, some of which appear to be prophages, suggests that horizontal exchange of these DNAs could be frequent [10,40,144,203]. Indeed, previous analysis of several plasmid genes has suggested that there has been “extensive” horizontal exchange among *B. burgdorferi* lineages (*e. g.*, [66,79,84,158,204–206]), and Eggers and Samuels [39] have demonstrated that in the laboratory cp32 plasmids can transfer between strains as phage virions.

Among of the forty sequenced linear plasmids in the four strains, there are only four pairs of cognate linear plasmids in which we found no organizational differences between strains. These are lp17, lp28-1 and lp36 in 297 and JD1, and lp54 in 297 and N40 (noted by arrows in Figure 2; because of undetermined sequence at the ends of 297 lp28-3 and lp28-4, it is not known whether they might be organizationally the same as their B31 or JD1 and N40 or JD1 cognates, respectively). In addition, the cognate plasmid pairs of lp28-5, lp28-6 and lp38 are more similar to each other in JD1 and 297 than to the cognate plasmids in the other two strains. Although there are some substantial differences between the JD1 and 297 plasmids, these two linear plasmid sets appear to be more like one another than the other pair wise combinations. Although the JD1 and 297 rRNA IGS/chromosomal MLST/OspC lineages do not appear to be particularly closely related [57,58,67], the similarity of their plasmid contents might indicate that they are in fact more closely related than previously suspected, and the very different plasmids that are present in JD1 and 297 (*e. g.*, their lp38s) could be examples of horizontal transfer of plasmids, but study of more isolates will be required to determine if this is true.

Previous work and the sequences analyzed here show that the cp32 plasmids in different rRNA/OspC *B. burgdorferi* lineages are similar in overall structure, but can have considerable differences at the four variable positions discussed above. On the other hand B. Stevenson and co-workers (personal communication) have shown, by sequencing several of the variable regions, that strains B31 and BL206 (both rRNA IGS lineage 1, OspC type A [58,62]) appear have very similar cp32 complements, as do strains 297 and Sh-2-82 which are both rRNA lineage 2, OspC type K (Sh-2-82 our unpublished results). Thus, although the complete BL206 and Sh-2-82 genome sequences have not been determined, it appears that different members of the same rRNA/OspC lineages can have highly similar cp32 contents, implying that transfer between these lineages may not be so rapid in the wild that plasmid contents are randomized. None of the completely sequenced cp32 sets present presented here (from members of four different rRNA lineages) are as highly related as the above strain pairs (table 3), but they will provide a robust basis for the future determination of whether the cp32 contents of all or most isolates within rRNA/OspC lineages are indeed similar. This is especially intriguing because the cp32 plasmid prophages could be prone to particularly rapid horizontal transfer [10,40,144,203].

Plasmid types. Have all extant *B. burgdorferi* sensu stricto plasmid “compatibility types” been identified? In addition to the 26 PFam32 types mentioned above that are present in the four strains, a 27th PFam32 type has been reported for “cp32-13” in California isolate CA15 ([84] and our unpublished results for other strains). In addition, strain B31 linear plasmid lp28-1 carries two

apparently intact PFam32 genes. One of them, *b31_f13*, represents a novel PFam32 type that lies in a partition gene cluster that is missing its PFam57/62 member gene and so was ignored in previous analyses. It seems quite possible that plasmids of this compatibility type (a 28th type) will be found in other strains. If PFam32 types are distributed randomly in natural isolates, saturation has probably not been reached by analyzing only four strains, although the number of undiscovered compatibility types is likely not large since the N40, JD1 and 297 plasmid sequences add only five “new” types (lp28-5, lp28-6, lp28-7, cp32-11 and cp32-12) to the 21 types previously known in strain B31. We note that the four sequenced isolates are all from a geographically rather restricted region (southern New England and New York), and it will be interesting to determine whether *B. burgdorferi* isolates from other locations have different plasmid types.

The organizational variation *within* each PFam32 type in the four strains studied here suggests that the overall number of *B. burgdorferi* linear plasmid “organizational types” may not be small. Nonetheless, the facts that (i) some pairs of organizationally identical cognate plasmids exist among these four strains, and that (ii) there are numerous novel sequence joints that are present in more than one strain, suggest that a limited number of such variants exists for each plasmid compatibility type in nature. Our analyses also indicate that a many of the rearrangements that formed the different organizational types occurred so recently that the sequences involved have diverged substantially less than one percent since the rearrangement, yet the process is not so fast that every *B. burgdorferi* isolate has a completely different set of plasmid organizations. Since no such rearrangement has been observed in the laboratory, such events are in fact quite rare, at least under laboratory conditions.

Future directions. Many important unanswered questions regarding Lyme *Borrelia* genomics and population structure remain, including the following: Do other *B. burgdorferi* isolates harbor additional plasmid compatibility types? How many organizational subtypes within each plasmid compatibility type exist in *B. burgdorferi* in nature? Are any plasmids or plasmid subtypes restricted to particular *B. burgdorferi* chromosomal lineages or geographic areas? What are the relationships among plasmids of different *B. burgdorferi* sensu lato species? Are plasmids transferred between strains as whole entities or as fragments, and what are the rates of transfer in nature? Are plasmids transferred only within species or between closely related *Borrelia* species in nature? What gene set constitutes the *B. burgdorferi* pangenome? We have recently sequenced nine additional *B. burgdorferi* sensu stricto genomes [44] and eight genomes of related species [45,46,49] in order to begin to extend our knowledge in all of these areas.

Materials and Methods

Strains and DNA preparation

Low passage cultures of *B. burgdorferi* isolates B31, N40, JD1 and 297 were the kind gifts of Drs. Alan Barbour, Martin Schriefer, Tom Schwan and Justin Radolf, respectively. In this study, low passage cultures of N40, JD1 and 297 were propagated in complete BSK-II medium (Sigma, St. Louis, MO) at 34°C without isolation through a single colony or passage through a mouse in order to minimize loss of plasmids. For isolation of whole genomic DNA, 1 liter of log-phase bacteria ($\sim 4 \times 10^7$ bacteria/ml) were harvested by centrifugation at 10,000 rpm for 30 min at 4°C. The bacterial pellet was washed twice with 10 mM Tris pH 7.5, 100 mM NaCl buffer, and resuspended in 430 μ l TES (10 mM Tris pH 7.5, 100 mM NaCl, 10 mM EDTA). Subsequently, 10 μ l

of freshly prepared lysozyme (50 mg/ml), 50 μ l Sarkosyl (10%), and 10 μ l proteinase K (10 mg/ml; Sigma, St. Louis, MO) were added, and the mixture was incubated at 50°C overnight prior to RNase treatment. DNA was then extracted with phenol/chloroform and chloroform, precipitated with ethanol, and finally resuspended in TE buffer (1 mM Tris pH 7.5, 1 mM EDTA). Strain 297 plasmids were isolated with a QiaGen (Valencia, CA) Plasmid Midi-100 Kit according to the manufacturer’s recommendations.

Sequencing and sequence analysis

Sequencing, assembly, and gap closure. Sanger shotgun sequencing and assembly were performed as described previously for genomes sequenced at TIGR/JCVI [207]. All three genomes were sequenced to closure. Briefly, small-insert and medium-insert plasmid libraries were generated by random nebulization and cloning of genomic DNA. The following libraries were generated: N40, one 3–4 kb small-insert and one 6–8 kb medium-insert libraries; JD1, one 2–3 kb small-insert, one 3–4 kb small-insert and one 8–12 kb medium-insert libraries; 297, one 3–4 kb small-insert and one 10–12 kb medium-insert libraries. In the random sequencing phase, at least a 9-fold coverage across the genome was achieved from the shotgun sequencing libraries generated for each strain. More specifically, a total of 19971, 59714 and 11126 Sanger sequencing reads were generated during the random sequencing phase for N40, JD1 and 297 respectively. The sequences were assembled using the TIGR Assembler (www.jcvi.org/cms/research/software/) and the Celera Assembler (<http://sourceforge.net/projects/wgs-assembler/>), and the scaffolds constructed using TIGR BAMBUS [208]. All sequence and physical gaps were closed by editing the ends of sequence traces, primer walking or transposon-primed sequencing on plasmid clones, and combinatorial PCR followed by sequencing of the PCR product.

A number of the termini of bulk-determined sequence contigs were extended by sequencing DNAs from inverse PCR or by direct PCR amplification using outside primers designed from sequence predicted to be orthologous by comparison with plasmids from one of the other strains. About 8, 19, and 5 Kbp were determined by these directed methods in the N40, JD1 and 297 plasmid sequences. The lengths of the sequence contigs and plasmid sizes (determined by pulsed-field electrophoresis and Southern analysis as in Casjens *et al.* [17]), as well as lengths of the “missing” unsequenced terminal regions (calculated from the sizes of terminal restriction fragments) are given in Table S1. The linear 297 plasmid sequence contigs are often missing from 2000 to 2500 terminal bp; this was a poorly understood property of the DNA libraries, not sequencing depth. The GenBank accession numbers for the sequences determined in this study were reported in Schutzer *et al.* [44] and are included in Table S1.

Sequence annotation. For the N40, 297 and JD1 genomes, an initial set of ORFs likely to encode proteins was identified by GLIMMER (<http://cbcb.umd.edu/software/glimmer/>). This first set of open reading frames (ORFs) was then manually curated so that ORFs equal to or less than 50 codons long (not counting the stop codon) were removed unless they are homologs of a similarly sized gene of known function, and ORFs in the 51–100 codon range were only included if their reading frame is intact in cognate sequence in all of the strains that carry the sequence. ORFs that overlapped were inspected visually and, in some cases, removed. ORFs were searched against a nonredundant protein database as described previously for all TIGR genomes. Frameshifts and point mutations were detected, checked and corrected where appropriate. Remaining frameshifts and point mutations are

considered authentic, and corresponding regions were annotated as “authentic frameshift” or “authentic point mutation,” respectively. Two sets of hidden Markov models (HMMs) were used to determine ORF membership in families and superfamilies. These included 10,340 HMMs from PFAM version 23.0 (<http://pfam.sanger.ac.uk/>) and 3,603 HMMs from TIGRFam version 8.0 (www.jvarkit.org/cms/research/projects/tigrfam/overview/). TOPPRED [209] was used to identify membrane-spanning domains in proteins. For ease of comparison, the genome of strain B31 was also re-annotated by this pipeline, and this reannotation can be found in the original accession numbers of the B31 replicons (Table S1 and [10,11]).

Some of the plasmids carry full-length, degenerate pseudogene paralogs of other intact plasmid genes. These were not annotated (as they were in the original B31 annotation). The automated ORF searches identified some smaller ORFs within these pseudogenes, and since they could theoretically be expressed they were kept in the predicted ORF list. Translation frameshift and in-frame stop differences among the strains sequenced here were compared to homologs in *B. garinii* PBi [47], *B. afzelii* PKo ([48] and our unpublished results) and *B. bissettii* DN127 (our unpublished results) to determine which state is most likely functional.

Open reading frame nomenclature. *Borrelia* researchers have usually used the “locus tags” of the strain B31 genome GenBank annotation [10] as names for genes and their encoded proteins. Thus, according to bacterial convention, the B31 chromosomal genes have often been named “*bb0xxx*” (lower case and italicized) in ascending order from *bb0001* upward across the chromosome. The B31 plasmid locus tag names are similar but have the form “*bb\$xx*” in which “\$” is a letter code denoting which plasmid type carries the gene (e. g., *bba74* encodes protein B31_A74 and lies on lp54, *bbs09* lies on cp32-3, etc.). Increased genome sequencing forces the use of more complex locus tags such as, Bbujd1_Axx for strain JD1 plasmid lp54. To avoid very long gene names now that multiple genomes have been sequenced, we suggest the use the form “strain name_locus tag number only” for gene names so that their strain source is included (e. g., *b31_0843* for a B31 chromosomal gene, and “*jd1_a34*” for JD1 plasmid lp54 reading frame 34 with plasmid letter code lower case “a”), and we follow these conventions here. Table S4 lists the locus tag letters with their corresponding plasmids for all the plasmids in the four current genome sequences as well as for our additional unpublished sequences. In the different genomes, the same locus tag numbers in the B31, N40, 297 and JD1 chromosome, cp26 and lp54 indicate orthology of the corresponding genes; however, organizational differences in the other plasmids made this system unworkable, so the same locus tag numbers on these replicons do not indicate orthology.

Methods of ortholog/paralog analysis. We identified orthologous plasmids by inspection and by using NUCMER [210] and BLASTn [211]. For each set of orthologous replicons, we identified orthologous ORF sets by first finding all homologs of each ORF using all-against-all BLASTn [211]. Homologous ORFs were clustered using the MCL algorithm [212]. Within each homolog cluster, orthologs were distinguished from paralogs by visual inspection of gene orders displayed by the authors’ unpublished synteny browser and by matrix comparison with DNA Strider [135]. Percent identity of DNA and protein

sequences was calculated by DNA Strider using alignments created by that program. Protein multiple sequence alignments were constructed using ClustalW 1.83 [213] and ClustalX2 0.3 [214]. Codon alignments were derived from protein alignment templates using PERL scripts.

Supporting Information

Figure S1 The right end of the B. Burgdorferi JD1 chromosome.

(PDF)

Figure S2 Open reading frame maps for plasmids carried by B. burgdorferi strains B31, N40, JD1 and 297.

(PDF)

Figure S3 Vls cassette regions.

(PDF)

Figure S4 Ip28-6 copy number.

(PDF)

Figure S5 Recombination points.

(PDF)

Figure S6 Neighbor-joining trees of proteins encoded by B. burgdorferi cp32 variable regions.

(PDF)

Figure S7 Orthologous non-coding DNA in Ip28-4.

(PDF)

Table S1 B31, JD1, N40 and 297 replicon sizes and accession numbers.

(PDF)

Table S2 Paralogous protein families in four B. burgdorferi genomes.

(PDF)

Table S3 Major tandem repeat tracts.

(PDF)

Table S4 Lyme agent Borrelia plasmid letter appellations for locus tags.

(PDF)

Acknowledgments

We thank Drs. Martin Schriefer, Tom Schwan, Janis Weis, Steve Barthold and Justin Radolf for *B. burgdorferi* strains. We also thank Dan Qiu, Xiaohua Yang, Yun Xu and John Bruno for DNA preparation and access to their unpublished strain typing information, Kevin Holden, Xiaohui Wang and Janis Weis for access to the N40 *arp* and putative *vsE* gene sequences before release, Brian Stevenson for access to unpublished information, and Lauren Brinkac, Anthony Durkin, Heather Huot-Creasy and Sean Daugherty for help with genome sequence annotation.

Author Contributions

Conceived and designed the experiments: SRC WMH WGQ BJL JJD SES CMF. Performed the experiments: EFM EBG MV DR JKA LCV SF JFW GID HMK JES RAH SRC. Analyzed the data: SRC EBG MV LCV SF WGQ EFM. Contributed reagents/materials/analysis tools: SRC BJL WGQ EFM CMF. Wrote the paper: SRC.

References

1. Steere AC (2001) Lyme disease. *N Engl J Med* 345: 115–125.
2. Steere AC, Coburn J, Glickstein L (2004) The emergence of Lyme disease. *J Clin Invest* 113: 1093–1101.
3. Steere AC (2006) Lyme borreliosis in 2005, 30 years after initial observations in Lyme Connecticut. *Wien Klin Wochenschr* 118: 625–633.
4. Feder HM, Jr., Johnson BJ, O’Connell S, Shapiro ED, Steere AC, et al. (2007) A critical appraisal of “chronic Lyme disease”. *N Engl J Med* 357: 1422–1430.
5. Kurtenbach K, Hanincova K, Tsao JI, Margos G, Fish D, et al. (2006) Fundamental processes in the evolutionary ecology of Lyme borreliosis. *Nat Rev Microbiol* 4: 660–669.

6. Dworkin MS, Schwan TG, Anderson DE, Jr., Borchardt SM (2008) Tick-borne relapsing fever. *Infect Dis Clin North Am* 22: 449–468, viii.
7. Stewart PE, Byram R, Grimm D, Tilly K, Rosa PA (2005) The plasmids of *Borrelia burgdorferi*: essential genetic elements of a pathogen. *Plasmid* 53: 1–13.
8. Tilly K, Rosa PA, Stewart PE (2008) Biology of infection with *Borrelia burgdorferi*. *Infect Dis Clin North Am* 22: 217–234, v.
9. Radolf JD, Caimano MJ, Stevenson B, Hu LT (2012) Of ticks, mice and men: understanding the dual-host lifestyle of Lyme disease spirochaetes. *Nat Rev Microbiol* 10: 87–99.
10. Casjens S, Palmer N, van Vugt R, Huang WM, Stevenson B, et al. (2000) 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* 35: 490–516.
11. Fraser CM, Casjens S, Huang WM, Sutton GG, Clayton R, et al. (1997) Genomic sequence of a Lyme disease spirochaete, *Borrelia burgdorferi*. *Nature* 390: 580–586.
12. Casjens S, van Vugt R, Tilly K, Rosa PA, Stevenson B (1997) Homology throughout the multiple 32-kilobase circular plasmids present in Lyme disease spirochetes. *J Bacteriol* 179: 217–227.
13. Miller J, Bono J, Babb K, El-Hage N, Casjens S, et al. (2000) A second allele of *ospA* in *Borrelia burgdorferi* strain B31 is located on the previously undetected circular plasmid cp9-2. *J Bacteriol* 182: 6254–6258.
14. Barbour AG, Garon CF (1987) Linear plasmids of the bacterium *Borrelia burgdorferi* have covalently closed ends. *Science* 237: 409–411.
15. Barbour AG (1988) Plasmid analysis of *Borrelia burgdorferi*, the Lyme disease agent. *J Clin Microbiol* 26: 475–478.
16. Busch U, Teufel CH, Boehmer R, Wilske B, Preac-Mursic V (1995) Molecular characterization of *Borrelia burgdorferi* sensu lato strains by pulsed-field gel electrophoresis. *Electrophoresis* 16: 744–747.
17. Casjens S, Delange M, Ley HL, Rosa P, Huang WM (1995) Linear chromosomes of Lyme disease agent spirochetes: genetic diversity and conservation of gene order. *J Bacteriol* 177: 2769–2780.
18. Hughes CA, Kodner CB, Johnson RC (1992) DNA analysis of *Borrelia burgdorferi* NCH-1, the first northcentral U.S. human Lyme disease isolate. *J Clin Microbiol* 30: 698–703.
19. Iyer R, Kalu O, Purser J, Norris S, Stevenson B, et al. (2003) Linear and circular plasmid content in *Borrelia burgdorferi* clinical isolates. *Infect Immun* 71: 3699–3706.
20. Marconi RT, Casjens S, Munderloh UG, Samuels DS (1996) Analysis of linear plasmid dimers in *Borrelia burgdorferi* sensu lato isolates: implications concerning the potential mechanism of linear plasmid replication. *J Bacteriol* 178: 3357–3361.
21. Palmer N, Fraser C, Casjens S (2000) Distribution of twelve linear extrachromosomal DNAs in natural isolates of the Lyme disease spirochetes. *J Bacteriol* 182: 2476–2480.
22. Samuels D, Marconi R, Garon C (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(Pt 10): 2445–2449.
23. Simpson WJ, Garon CF, Schwan TG (1990) *Borrelia burgdorferi* contains repeated DNA sequences that are species specific and plasmid associated. *Infect Immun* 58: 847–853.
24. Simpson WJ, Garon CF, Schwan TG (1990) Analysis of supercoiled circular plasmids in infectious and non-infectious *Borrelia burgdorferi*. *Microb Pathog* 8: 109–118.
25. Stalhammar-Carlemalm M, Jenny E, Gern L, Aeschlimann A, Meyer J (1990) Plasmid analysis and restriction fragment length polymorphisms of chromosomal DNA allow a distinction between *Borrelia burgdorferi* strains. *Int J Med Microbiol* 274: 28–39.
26. Xu Y, Johnson RC (1995) Analysis and comparison of plasmid profiles of *Borrelia burgdorferi* sensu lato strains. *J Clin Microbiol* 33: 2679–2685.
27. Tilly K, Casjens S, Stevenson B, Bono JL, Samuels DS, 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.
28. Margolis N, Hogan D, Tilly K, Rosa PA (1994) Plasmid location of *Borrelia* purine biosynthesis gene homologs. *J Bacteriol* 176: 6427–6432.
29. Bono JL, Tilly K, Stevenson B, Hogan D, Rosa P (1998) Oligopeptide permease in *Borrelia burgdorferi*: putative peptide-binding components encoded by both chromosomal and plasmid loci. *Microbiology* 144: 1033–1044.
30. Tilly K, Grimm D, Bueschel DM, Krum JG, Rosa P (2004) Infectious cycle analysis of a *Borrelia burgdorferi* mutant defective in transport of chitobiose, a tick cuticle component. *Vector Borne Zoonotic Dis* 4: 159–168.
31. Rybchin VN, Svarechevsky AN (1999) The plasmid prophage N15: a linear DNA with covalently closed ends. *Mol Microbiol* 33: 895–903.
32. Ravin V, Ravin N, Casjens S, Ford ME, Hatfull GF, et al. (2000) Genomic sequence and analysis of the atypical temperate bacteriophage N15. *J Mol Biol* 299: 53–73.
33. Kobryn K, Chaconas G (2002) ResT, a telomere resolvase encoded by the Lyme disease spirochete. *Mol Cell* 9: 195–201.
34. Chaconas G, Stewart PE, Tilly K, Bono JL, Rosa P (2001) Telomere resolution in the Lyme disease spirochete. *Embo J* 20: 3229–3237.
35. Bergström S, Zückert W (2010) Structure, function and biogenesis of the *Borrelia* cell envelope. In: Samuels DS, Radolf J, eds. *Borrelia* molecular biology, host interaction and pathogenesis. Norfolk: Caister Academic Press. pp 138–166.
36. Brooks CS, Vuppala SR, Jett AM, Akins DR (2006) Identification of *Borrelia burgdorferi* outer surface proteins. *Infect Immun* 74: 296–304.
37. Nowalk AJ, Nolder C, Clifton DR, Carroll JA (2006) Comparative proteome analysis of subcellular fractions from *Borrelia burgdorferi* by NEPHGE and IPG. *Proteomics* 6: 2121–2134.
38. Casjens S (2000) *Borrelia* genomes in the year 2000. *J Mol Microbiol Biotechnol* 2: 401–410.
39. Eggers CH, Samuels DS (1999) Molecular evidence for a new bacteriophage of *Borrelia burgdorferi*. *J Bacteriol* 181: 7308–7313.
40. Eggers CH, Casjens S, Samuels DS (2001) Bacteriophages of *Borrelia burgdorferi* and other spirochetes. In: Saier M, Garcia-Lara J, eds. *The spirochetes: molecular and cellular biology*. Wymondham, United Kingdom: Horizon Press. pp 35–44.
41. Zhang H, Marconi RT (2005) Demonstration of cotranscription and 1-methyl-3-nitroso-nitroguanidine induction of a 30-gene operon of *Borrelia burgdorferi*: evidence that the 32-kilobase circular plasmids are prophages. *J Bacteriol* 187: 7985–7995.
42. Byram R, Stewart PE, Rosa P (2004) The essential nature of the ubiquitous 26-kilobase circular replicon of *Borrelia burgdorferi*. *J Bacteriol* 186: 3561–3569.
43. Sadziene A, Barbour A, Rosa P, Thomas D (1993) An *OspB* mutant of *Borrelia burgdorferi* has reduced invasiveness in vitro and reduced infectivity in vivo. *Infect Immun* 61: 3590–3596.
44. Schutzer SE, Fraser-Liggett CM, Casjens SR, Qiu WG, Dunn JJ, et al. (2011) Whole genome sequences of thirteen isolates of *Borrelia burgdorferi*. *J Bacteriol* 193: 1018–1020.
45. Casjens SR, Fraser-Liggett CM, Mongodin EF, Qiu WG, Dunn JJ, et al. (2011) Whole genome sequence of an unusual *Borrelia burgdorferi* sensu lato isolate. *J Bacteriol* 193: 1489–1490.
46. Casjens S, Mongodin E, Qiu W-G, Dunn J, Luft B, et al. (2011) Whole genome sequences of two *Borrelia afzelii* and two *Borrelia garinii* Lyme disease agent isolates. *J Bacteriol* 193: 6695–6696.
47. Glöckner G, Lehmann R, Romualdi A, Pradella S, Schulte-Spechtel U, et al. (2004) Comparative analysis of the *Borrelia garinii* genome. *Nucleic Acids Res* 32: 6038–6046.
48. Glöckner G, Schulte-Spechtel U, Schilhabel M, Felder M, Suhnel J, et al. (2006) Comparative genome analysis: selection pressure on the *Borrelia* *vis* cassettes is essential for infectivity. *BMC Genomics* 7: 211.
49. Schutzer S, Fraser-Liggett C, Qiu W, Kraiczay P, Mongodin E, et al. (2012) Whole genome sequences of *Borrelia bisettii*, *Borrelia valaisiana* and *Borrelia spielmanni*. *J Bacteriol* 194: 545–546.
50. Casjens S, Murphy M, DeLange M, Sampson L, van Vugt R, et al. (1997) Telomeres of the linear chromosomes of Lyme disease spirochaetes: nucleotide sequence and possible exchange with linear plasmid telomeres. *Mol Microbiol* 26: 581–596.
51. Huang WM, Robertson M, Aron J, Casjens S (2004) Telomere exchange between linear replicons of *Borrelia burgdorferi*. *J Bacteriol* 186: 4134–4141.
52. Lescot M, Audic S, Robert C, Nguyen TT, Blanc G, et al. (2008) The genome of *Borrelia recurrentis*, the agent of deadly louse-borne relapsing fever, is a degraded subset of tick-borne *Borrelia duttoni*. *PLoS Genet* 4: e1000185.
53. Stewart PE, Thalken R, Bono JL, Rosa P (2001) Isolation of a circular plasmid region sufficient for autonomous replication and transformation of infectious *Borrelia burgdorferi*. *Mol Microbiol* 39: 714–721.
54. Dunn JJ, Buchstein SR, Butler LL, Fisman S, Polin DS, et al. (1994) Complete nucleotide sequence of a circular plasmid from the Lyme disease spirochete, *Borrelia burgdorferi*. *J Bacteriol* 176: 2706–2717.
55. Terekhova D, Iyer R, Wormser GP, Schwartz I (2006) Comparative genome hybridization reveals substantial variation among clinical isolates of *Borrelia burgdorferi* sensu stricto with different pathogenic properties. *J Bacteriol* 188: 6124–6134.
56. Norris SJ, Howell JK, Odeh EA, Lin T, Gao L, et al. (2011) High-throughput plasmid content analysis of *Borrelia burgdorferi* B31 by using Luminex multiplex technology. *Appl Environ Microbiol* 77: 1483–1492.
57. Attie O, Bruno JF, Xu Y, Qiu D, Luft BJ, et al. (2007) Co-evolution of the outer surface protein C gene (*ospC*) and intraspecific lineages of *Borrelia burgdorferi* sensu stricto in the northeastern United States. *Infect Genet Evol* 7: 1–12.
58. Bunikis J, Garpino U, Tsao J, Berglund J, Fish D, et al. (2004) Sequence typing reveals extensive strain diversity of the Lyme borreliosis agents *Borrelia burgdorferi* in North America and *Borrelia afzelii* in Europe. *Microbiology* 150: 1741–1755.
59. Qiu WG, Bruno JF, McCaig WD, Xu Y, Livey I, et al. (2008) Wide distribution of a high-virulence *Borrelia burgdorferi* clone in Europe and North America. *Emerg Infect Dis* 14: 1097–1104.
60. Travinsky B, Bunikis J, Barbour AG (2010) Geographic differences in genetic locus linkages for *Borrelia burgdorferi*. *Emerg Infect Dis* 16: 1147–1150.
61. Mathiesen DA, Oliver JH, Jr., Kolbert CP, Tullson ED, Johnson BJ, et al. (1997) Genetic heterogeneity of *Borrelia burgdorferi* in the United States. *J Infect Dis* 175: 98–107.
62. Wang I, Dykhuizen D, Qiu W, Dunn J, Bosler E, et al. (1999) Genetic diversity of *ospC* in a local population of *Borrelia burgdorferi* sensu stricto. *Genetics* 151: 15–30.



63. Qiu WG, Bosler EM, Campbell JR, Uguine GD, Wang IN, et al. (1997) A population genetic study of *Borrelia burgdorferi* sensu stricto from eastern Long Island, New York, suggested frequency-dependent selection, gene flow and host adaptation. *Heredity* 127: 203–216.
64. Qiu WG, Dykhuizen DE, Acosta MS, Luft BJ (2002) Geographic uniformity of the Lyme disease spirochete (*Borrelia burgdorferi*) and its shared history with tick vector (*Ixodes scapularis*) in the Northeastern United States. *Genetics* 160: 833–849.
65. Barbour AG, Travinsky B (2010) Evolution and distribution of the *ospC* gene, a transferable serotype determinant of *Borrelia burgdorferi*. *MBio* 1: e00153–00110.
66. Qiu WG, Schutzer SE, Bruno JF, Attie O, Xu Y, et al. (2004) Genetic exchange and plasmid transfers in *Borrelia burgdorferi* sensu stricto revealed by three-way genome comparisons and multilocus sequence typing. *Proc Natl Acad Sci U S A* 101: 14150–14155.
67. Margos G, Gatewood A, Aaenens D, Hanincova K, Terekova D, et al. (2008) MLST of housekeeping genes captures geographic population structure and suggests a European origin of *Borrelia burgdorferi*. *Proc Natl Acad Sci U S A* 105: 8730–8735.
68. Margos G, Vollmer SA, Cornet M, Garnier M, Fingerle V, et al. (2009) A new *Borrelia* species defined by multilocus sequence analysis of housekeeping genes. *Appl Environ Microbiol* 75: 5410–5416.
69. Hoen AG, Margos G, Bent SJ, Diuk-Wasser MA, Barbour A, et al. (2009) Phylogeography of *Borrelia burgdorferi* in the eastern United States reflects multiple independent Lyme disease emergence events. *Proc Natl Acad Sci U S A* 106: 15013–15018.
70. Burgdorfer W, Barbour AG, Hayes SF, Benach JL, Grunwaldt E, et al. (1982) Lyme disease—a tick-borne spirochetosis? *Science* 216: 1317–1319.
71. Barthold SW, Moody KD, Terwilliger GA, Duray PH, Jacoby RO, et al. (1988) Experimental Lyme arthritis in rats infected with *Borrelia burgdorferi*. *J Infect Dis* 157: 842–846.
72. Piesman J, Mather TN, Sinsky RJ, Spielman A (1987) Duration of tick attachment and *Borrelia burgdorferi* transmission. *J Clin Microbiol* 25: 557–558.
73. Steere AC, Grodzicki RL, Kornblatt AN, Craft JE, Barbour AG, et al. (1983) The spirochetal etiology of Lyme disease. *N Engl J Med* 308: 733–740.
74. Wormser GP, Liveris D, Nowakowski J, Nadelman RB, Cavaliere LF, et al. (1999) Association of specific subtypes of *Borrelia burgdorferi* with hematogenous dissemination in early Lyme disease. *J Infect Dis* 180: 720–725.
75. Casjens SR, Huang WM, Gilcrease EB, Qiu WG, McCaig WD, et al. (2006) Comparative genomics of *Borrelia burgdorferi*. In: Cabello FC, Huliniska D, Godfrey HP, eds. *Molecular Biology of Spirochetes*. Amsterdam: IOS Press. pp 79–95.
76. Wywiał E, Haven J, Casjens SR, Hernandez YA, Singh S, et al. (2009) Fast, adaptive evolution at a bacterial host-resistance locus: The Pfam54 gene array in *Borrelia burgdorferi*. *Gene* 445: 26–37.
77. Casjens S, Eggers CH, Schwartz I (2010) *Borrelia* genomics: chromosome, plasmids, bacteriophages and genetic variation. In: Samuels DS, Radolf J, eds. *Borrelia* molecular biology, host interaction and pathogenesis. Norfolk: Caister Academic Press. pp 27–53.
78. Kraiczky P, Seling A, Brissette CA, Rossmann E, Hunfeld KP, et al. (2008) *Borrelia burgdorferi* complement regulator-acquiring surface protein 2 (CspZ) as a serological marker of human Lyme disease. *Clin Vaccine Immunol* 15: 484–491.
79. Haven J, Vargas L, Mongodin E, Fraser-Liggett C, Schutzer S, et al. (2011) How bacterial genomes diverge under recombination: Frequency-dependent selection in *Borrelia burgdorferi*, the Lyme disease bacterium. *Genetics* 189: 951–966.
80. Chan K, Casjens S, Parveen N (2012) Detection of established virulence genes and plasmids to differentiate *Borrelia burgdorferi* strains. *Infect Immun* in press.
81. Grimm D, Eggers CH, Caimano MJ, Tilly K, Stewart PE, et al. (2004) Experimental assessment of the roles of linear plasmids lp25 and lp28-1 of *Borrelia burgdorferi* throughout the infectious cycle. *Infect Immun* 72: 5938–5946.
82. Revel AT, Blevins JS, Almazan C, Neil L, Kocan KM, et al. (2005) *bpLA* (*bbe16*) is essential for the persistence of the Lyme disease spirochete, *Borrelia burgdorferi*, in its natural tick vector. *Proc Natl Acad Sci U S A* 102: 6972–6977.
83. Tourand Y, Deneke J, Moriarty TJ, Chaconas G (2009) Characterization and in vitro reaction properties of 19 unique hairpin telomeres from the linear plasmids of the Lyme disease spirochete. *J Biol Chem* 284: 7264–7272.
84. Stevenson B, Miller JC (2003) Intra- and interbacterial genetic exchange of Lyme disease spirochete *erp* genes generates sequence identity amidst diversity. *J Mol Evol* 57: 309–324.
85. Leonard TA, Moller-Jensen J, Lowe J (2005) Towards understanding the molecular basis of bacterial DNA segregation. *Philos Trans R Soc Lond B Biol Sci* 360: 523–535.
86. Deneke J, Chaconas G (2008) Purification and properties of the plasmid maintenance proteins from the *Borrelia burgdorferi* linear plasmid lp17. *J Bacteriol* 190: 3992–4000.
87. Champion CI, Blanco DR, Skare JT, Haake DA, Giladi M, 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.
88. Elias AF, Stewart PE, Grimm D, Caimano MJ, Eggers CH, et al. (2002) Clonal polymorphism of *Borrelia burgdorferi* strain B31 MI: implications for mutagenesis in an infectious strain background. *Infect Immun* 70: 2139–2150.
89. Purser JE, Norris SJ (2000) Correlation between plasmid content and infectivity in *Borrelia burgdorferi*. *Proc Natl Acad Sci U S A* 97: 13865–13870.
90. Eggers CH, Caimano MJ, Clawson ML, Miller WG, Samuels DS, et al. (2002) Identification of loci critical for replication and compatibility of a *Borrelia burgdorferi* cp32 plasmid and use of a cp32-based shuttle vector for the expression of fluorescent reporters in the Lyme disease spirochete. *Mol Microbiol* 43: 281–295.
91. Jewett MW, Byram R, Bestor A, Tilly K, Lawrence K, et al. (2007) Genetic basis for retention of a critical virulence plasmid of *Borrelia burgdorferi*. *Mol Microbiol* 66: 975–990.
92. Barbour AG (1993) Linear DNA of *Borrelia* species and antigenic variation. *Trends Microbiol* 1: 236–239.
93. Sadziene A, Rosa PA, Thompson PA, Hogan DM, Barbour AG (1992) Antibody-resistant mutants of *Borrelia burgdorferi*: in vitro selection and characterization. *J Exp Med* 176: 799–809.
94. Tilly K, Lubke L, Rosa P (1998) Characterization of circular plasmid dimers in *Borrelia burgdorferi*. *J Bacteriol* 180: 5676–5681.
95. Behera AK, Durand E, Cugini G, Antonara S, Bourassa L, et al. (2008) *Borrelia burgdorferi* BBB07 interaction with integrin $_{\alpha 5\beta 1}$ stimulates production of pro-inflammatory mediators in primary human chondrocytes. *Cell Microbiol* 10: 320–331.
96. Marconi RT, Samuels DS, Schwan TG, Garon CF (1993) Identification of a protein in several *Borrelia* species which is related to OspC of the Lyme disease spirochetes. *J Clin Microbiol* 31: 2577–2583.
97. Sadziene A, Wilske B, Ferdows M, Barbour A (1993) The cryptic *ospC* gene of *Borrelia burgdorferi* B31 is located on a circular plasmid. *Infect Immun* 61: 2192–2195.
98. Wilske B, Preac-Mursic V, Schierz G, Busch KV (1986) Immunochemical and immunological analysis of European *Borrelia burgdorferi* strains. *Zentralbl Bakteriol Mikrobiol Hyg [A]* 263: 92–102.
99. Tilly K, Krum JG, Bestor A, Jewett MW, Grimm D, et al. (2006) *Borrelia burgdorferi* OspC protein required exclusively in a crucial early stage of mammalian infection. *Infect Immun* 74: 3554–3564.
100. Radolf JD, Caimano MJ (2008) The long strange trip of *Borrelia burgdorferi* outer-surface protein C. *Mol Microbiol* 69: 1–4.
101. Brissette CA, Cooley AE, Burns LH, Riley SP, Verma A, et al. (2008) Lyme borreliosis spirochete Erp proteins, their known host ligands, and potential roles in mammalian infection. *Int J Med Microbiol* 298 Suppl 1: 257–267.
102. Brissette CA, Bykowski T, Cooley AE, Bowman A, Stevenson B (2009) *Borrelia burgdorferi* RevA antigen binds host fibronectin. *Infect Immun* 77: 2802–2812.
103. 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.
104. Yang X, Popova TG, Hagman KE, Wikel SK, Schoeler GB, et al. (1999) Identification, characterization, and expression of three new members of the *Borrelia burgdorferi* Mlp (2.9) lipoprotein gene family. *Infect Immun* 67: 6008–6018.
105. Brissette CA, Haupt K, Barthel D, Cooley AE, Bowman A, et al. (2009) *Borrelia burgdorferi* infection-associated surface proteins ErpP, ErpA, and ErpC bind human plasminogen. *Infect Immun* 77: 300–306.
106. Brissette CA, Verma A, Bowman A, Cooley AE, Stevenson B (2009) The *Borrelia burgdorferi* outer-surface protein ErpX binds mammalian laminin. *Microbiology* 155: 863–872.
107. Alitalo A, Meri T, Lankinen H, Seppala I, Lahdenne P, et al. (2002) Complement inhibitor factor H binding to Lyme disease spirochetes is mediated by inducible expression of multiple plasmid-encoded outer surface protein E paralogs. *J Immunol* 169: 3847–3853.
108. Kraiczky P, Hartmann K, Hellwege J, Skerka C, Kirschfink M, et al. (2004) Immunological characterization of the complement regulator factor H-binding CRASP and Erp proteins of *Borrelia burgdorferi*. *Int J Med Microbiol* 293 Suppl 37: 152–157.
109. Metts MS, McDowell JV, Theisen M, Hansen PR, Marconi RT (2003) Analysis of the OspE determinants involved in binding of factor H and OspE-targeting antibodies elicited during *Borrelia burgdorferi* infection in mice. *Infect Immun* 71: 3587–3596.
110. Caimano MJ, Yang X, Popova TG, Clawson ML, Akins DR, et al. (2000) Molecular and evolutionary characterization of the cp32/18 family of supercoiled plasmids in *Borrelia burgdorferi* 297. *Infect Immun* 68: 1574–1586.
111. Stevenson B, Casjens S, van Vugt R, Porcella SF, Tilly K, et al. (1997) Characterization of cp18, a naturally truncated member of the cp32 family of *Borrelia burgdorferi* plasmids. *J Bacteriol* 179: 4285–4291.
112. Sarkar A, Hayes BM, Dulebohn DP, Rosa PA (2011) Regulation of the virulence determinant OspC by *bbd18* on linear plasmid lp17 of *Borrelia burgdorferi*. *J Bacteriol* 193: 5365–5373.
113. Fikrig E, Barthold SW, Sun W, Feng W, Telford SR, 3rd, et al. (1997) *Borrelia burgdorferi* P35 and P37 proteins, expressed *in vivo*, elicit protective immunity. *Immunity* 6: 531–539.
114. Xu Y, Bruno JF, Luft BJ (2003) Detection of genetic diversity in linear plasmids 28-3 and 36 in *Borrelia burgdorferi* sensu stricto isolates by subtractive hybridization. *Microb Pathog* 35: 269–278.
115. Labandeira-Rey M, Seshu J, Skare JT (2003) The absence of linear plasmid 25 or 28-1 of *Borrelia burgdorferi* dramatically alters the kinetics of experimental infection via distinct mechanisms. *Infect Immun* 71: 4608–4613.
116. Strother KO, Broadwater A, De Silva A (2005) Plasmid requirements for infection of ticks by *Borrelia burgdorferi*. *Vector Borne Zoonotic Dis* 5: 237–245.

117. Purser JE, Lawrenz MB, Caimano MJ, Howell JK, Radolf JD, et al. (2003) A plasmid-encoded nicotinamidase (PncA) is essential for infectivity of *Borrelia burgdorferi* in a mammalian host. *Mol Microbiol* 48: 753–764.

118. Kawabata H, Norris SJ, Watanabe H (2004) BBE02 disruption mutants of *Borrelia burgdorferi* B31 have a highly transformable, infectious phenotype. *Infect Immun* 72: 7147–7154.

119. Lawrenz MB, Kawabata H, Purser JE, Norris SJ (2002) Decreased electroporation efficiency in *Borrelia burgdorferi* containing linear plasmids lp25 and lp56: impact on transformation of infectious *B. burgdorferi*. *Infect Immun* 70: 4798–4804.

120. Rego RO, Bestor A, Rosa PA (2011) Defining the plasmid-borne restriction-modification systems of the Lyme disease spirochete *Borrelia burgdorferi*. *J Bacteriol* 193: 1161–1171.

121. Grimm D, Tilly K, Bueschel DM, Fisher MA, Policastro PF, et al. (2005) Defining plasmids required by *Borrelia burgdorferi* for colonization of tick vector *Ixodes scapularis* (Acari: Ixodidae). *J Med Entomol* 42: 676–684.

122. Labandeira-Rey M, Skare JT (2001) Decreased infectivity in *Borrelia burgdorferi* strain B31 is associated with loss of linear plasmid 25 or 28-1. *Infect Immun* 69: 446–455.

123. Embers ME, Alvarez X, Ooms T, Philipp MT (2008) The failure of immune response evasion by linear plasmid 28-1-deficient *Borrelia burgdorferi* is attributable to persistent expression of an outer surface protein. *Infect Immun* 76: 3984–3991.

124. Botkin DJ, Abbott AN, Stewart PE, Rosa PA, Kawabata H, et al. (2006) Identification of potential virulence determinants by Himar1 transposition of infectious *Borrelia burgdorferi* B31. *Infect Immun* 74: 6690–6699.

125. Barthold SW, Hodzic E, Tunev S, Feng S (2006) Antibody-mediated disease remission in the mouse model of Lyme borreliosis. *Infect Immun* 74: 4817–4825.

126. Feng S, Hodzic E, Barthold SW (2000) Lyme arthritis resolution with antiserum to a 37-kilodalton *Borrelia burgdorferi* protein. *Infect Immun* 68: 4169–4173.

127. Feng S, Hodzic E, Freet K, Barthold SW (2003) Immunogenicity of *Borrelia burgdorferi* arthritis-related protein. *Infect Immun* 71: 7211–7214.

128. Zhang JR, Hardham JM, Barbour AG, Norris SJ (1997) Antigenic variation in Lyme disease borreliae by promiscuous recombination of VMP-like sequence cassettes. *Cell* 89: 275–285.

129. Zhang JR, Norris SJ (1998) Genetic variation of the *Borrelia burgdorferi* gene *vlsE* involves cassette-specific, segmental gene conversion. *Infect Immun* 66: 3698–3704.

130. Iyer R, Hardham JM, Wormser GP, Schwartz I, Norris SJ (2000) Conservation and heterogeneity of *vlsE* among human and tick isolates of *Borrelia burgdorferi*. *Infect Immun* 68: 1714–1718.

131. Bankhead T, Chaconas G (2007) The role of *VlsE* antigenic variation in the Lyme disease spirochete: persistence through a mechanism that differs from other pathogens. *Mol Microbiol* 65: 1547–1558.

132. Coutte L, Botkin DJ, Gao L, Norris SJ (2009) Detailed analysis of sequence changes occurring during *vlsE* antigenic variation in the mouse model of *Borrelia burgdorferi* infection. *PLoS Pathog* 5: e1000293.

133. Hudson CR, Frye JG, Quinn FD, Gherardini FC (2001) Increased expression of *Borrelia burgdorferi vlsE* in response to human endothelial cell membranes. *Mol Microbiol* 41: 229–239.

134. Bykowski T, Babb K, von Lackum K, Riley SP, Norris SJ, et al. (2006) Transcriptional regulation of the *Borrelia burgdorferi* antigenically variable *VlsE* surface protein. *J Bacteriol* 188: 4879–4889.

135. Douglas SE (1994) DNA Strider: A Macintosh program for handling protein and nucleic acid sequences. *Methods Mol Biol* 25: 181–194.

136. Hartmann K, Corvey C, Skerka C, Kirschfink M, Karas M, et al. (2006) Functional characterization of BbCRASP-2, a distinct outer membrane protein of *Borrelia burgdorferi* that binds host complement regulators factor H and FHL-1. *Mol Microbiol* 61: 1220–1236.

137. Parveen N, Cornell KA, Bono JL, Chamberland C, Rosa P, et al. (2006) Bgp, a secreted glycosaminoglycan-binding protein of *Borrelia burgdorferi* strain N40, displays nucleosidase activity and is not essential for infection of immunodeficient mice. *Infect Immun* 74: 3016–3020.

138. Labandeira-Rey M, Baker EA, Skare JT (2001) VraA (BBI16) protein of *Borrelia burgdorferi* is a surface-exposed antigen with a repetitive motif that confers partial protection against experimental Lyme borreliosis. *Infect Immun* 69: 1409–1419.

139. Nowalk AJ, Gilmore RD, Jr., Carroll JA (2006) Serologic proteome analysis of *Borrelia burgdorferi* membrane-associated proteins. *Infect Immun* 74: 3864–3873.

140. Freedman JC, Rogers EA, Kostick JL, Zhang H, Iyer R, et al. (2009) Identification and molecular characterization of a cyclic-di-GMP effector protein, PlzA (BB0733): additional evidence for the existence of a functional cyclic-di-GMP regulatory network in the Lyme disease spirochete, *Borrelia burgdorferi*. *FEMS Immunol Med Microbiol* 58: 285–294.

141. Pitzer JE, Sultan SZ, Hayakawa Y, Hobbs G, Miller MR, et al. (2011) Analysis of the *Borrelia burgdorferi* cyclic-di-GMP-binding protein PlzA reveals a role in motility and virulence. *Infect Immun* 79: 1815–1825.

142. Zhang X, Bruce TC (2006) The mechanism of M.HhaI DNA C5 cytosine methyltransferase enzyme: a quantum mechanics/molecular mechanics approach. *Proc Natl Acad Sci U S A* 103: 6148–6153.

143. Gowers DM, Bellamy SR, Halford SE (2004) One recognition sequence, seven restriction enzymes, five reaction mechanisms. *Nucleic Acids Res* 32: 3469–3479.

144. Eggers CH, Casjens S, Hayes SF, Garon CF, Damman CJ, et al. (2000) Bacteriophages of spirochetes. *J Mol Microbiol Biotechnol* 2: 365–373.

145. Wallich R, Brenner K, Kramer M, Simon M (1995) Molecular cloning and immunological characterization of a novel linear-plasmid-encoded gene, pG, of *Borrelia burgdorferi* expressed only *in vivo*. *Infect Immun* 63: 3327–3335.

146. Miller JC, Stevenson B (2003) Immunological and genetic characterization of *Borrelia burgdorferi* BapA and EppA proteins. *Microbiology* 149: 1113–1125.

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

148. Hinnebusch J, Barbour AG (1992) Linear- and circular-plasmid copy numbers in *Borrelia burgdorferi*. *J Bacteriol* 174: 5251–5257.

149. Jewett MW, Lawrence K, Bestor AC, Tilly K, Grimm D, et al. (2007) The critical role of the linear plasmid lp36 in the infectious cycle of *Borrelia burgdorferi*. *Mol Microbiol* 64: 1358–1374.

150. Fikrig E, Feng W, Barthold SW, Telford SR, 3rd, Flavell RA (2000) Arthropod- and host-specific *Borrelia burgdorferi bbk32* expression and the inhibition of spirochete transmission. *J Immunol* 164: 5344–5351.

151. Probert W, Johnson B (1998) Identification of a 47 kd fibronectin-binding protein expressed by *Borrelia burgdorferi* isolate B31. *Molecular Microbiology* 30: 1003–1015.

152. Fischer JR, LeBlanc KT, Leong JM (2006) Fibronectin binding protein BBK32 of the Lyme disease spirochete promotes bacterial attachment to glycosaminoglycans. *Infect Immun* 74: 435–441.

153. Li X, Liu X, Beck DS, Kantor FS, Fikrig E (2006) *Borrelia burgdorferi* lacking BBK32, a fibronectin-binding protein, retains full pathogenicity. *Infect Immun* 74: 3305–3313.

154. Seshu J, Esteve-Gassent MD, Labandeira-Rey M, Kim JH, Trzeciakowski JP, et al. (2006) Inactivation of the fibronectin-binding adhesin gene *bbk32* significantly attenuates the infectivity potential of *Borrelia burgdorferi*. *Mol Microbiol* 59: 1591–1601.

155. Barbour AG, Jasinskas A, Kayala MA, Davies DH, Steere AC, et al. (2008) A genome-wide proteome array reveals a limited set of immunogens in natural infections of humans and white-footed mice with *Borrelia burgdorferi*. *Infect Immun* 76: 3374–3389.

156. Coleman AS, Pal U (2009) BBK07, a dominant *in vivo* antigen of *Borrelia burgdorferi*, is a potential marker for serodiagnosis of Lyme disease. *Clin Vaccine Immunol* 16: 1569–1575.

157. Coleman AS, Rossmann E, Yang X, Song H, Lamichhane CM, et al. (2011) BBK07 immunodominant peptides as serodiagnostic markers of Lyme disease. *Clin Vaccine Immunol* 18: 406–413.

158. Marconi RT, Samuels DS, Landry RK, Garon CF (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.

159. Norris SJ, Carter CJ, Howell JK, Barbour AG (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.

160. Dulebohn DP, Bestor A, Rego RO, Stewart PE, Rosa PA (2011) The *Borrelia burgdorferi* linear plasmid lp38 is dispensable for completion of the mouse-tick infectious cycle. *Infect Immun* 79: 3510–3517.

161. Stewart PE, Bestor A, Cullen JN, Rosa PA (2008) A tightly regulated surface protein of *Borrelia burgdorferi* is not essential to the mouse-tick infectious cycle. *Infect Immun* 76: 1970–1978.

162. Li X, Neelakanta G, Liu X, Beck DS, Kantor FS, et al. (2007) Role of outer surface protein D in the *Borrelia burgdorferi* life cycle. *Infect Immun* 75: 4237–4244.

163. Bergstrom S, Bundoc VG, Barbour AG (1989) Molecular analysis of linear plasmid-encoded major surface proteins, *OspA* and *OspB*, of the Lyme disease spirochete *Borrelia burgdorferi*. *Mol Microbiol* 3: 479–486.

164. Li H, Dunn JJ, Luft BJ, Lawson CL (1997) Crystal structure of Lyme disease antigen outer surface protein A complexed with an Fab. *Proc Natl Acad Sci U S A* 94: 3584–3589.

165. Neelakanta G, Li X, Pal U, Liu X, Beck DS, et al. (2007) Outer surface protein B is critical for *Borrelia burgdorferi* adherence and survival within *Ixodes* ticks. *PLoS Pathog* 3: e33.

166. Kraiczky P, Hellwege J, Skerka C, Becker H, Kirschfink M, et al. (2004) Complement resistance of *Borrelia burgdorferi* correlates with the expression of BbCRASP-1, a novel linear plasmid-encoded surface protein that interacts with human factor H and FHL-1 and is unrelated to Erp proteins. *J Biol Chem* 279: 2421–2429.

167. Zhong J, Skouloubris S, Dai Q, Mylykallio H, Barbour AG (2006) Function and evolution of plasmid-borne genes for pyrimidine biosynthesis in *Borrelia* spp. *J Bacteriol* 188: 909–918.

168. Feng S, Das S, Lam T, Flavell RA, Fikrig E (1995) A 55-kilodalton antigen encoded by a gene on a *Borrelia burgdorferi* 49-kilobase plasmid is recognized by antibodies in sera from patients with Lyme disease. *Infect Immun* 63: 3459–3466.

169. Promnares K, Kumar M, Shroder DY, Zhang X, Anderson JF, et al. (2009) *Borrelia burgdorferi* small lipoprotein Lp6.6 is a member of multiple protein

complexes in the outer membrane and facilitates pathogen transmission from ticks to mice. *Mol Microbiol* 74: 112–125.

170. Bestor A, Stewart PE, Jewett MW, Sarkar A, Tilly K, et al. (2010) Use of the Cre-lox recombination system to investigate the lp54 gene requirement in the infectious cycle of *Borrelia burgdorferi*. *Infect Immun* 78: 2397–2407.
171. Gilmore RD, Jr., Howison RR, Dietrich G, Patton TG, Clifton DR, et al. (2010) The *bbA64* gene of *Borrelia burgdorferi*, the Lyme disease agent, is critical for mammalian infection via tick bite transmission. *Proc Natl Acad Sci U S A* 107: 7515–7520.
172. Kumar M, Yang X, Coleman AS, Pal U (2010) BBA52 facilitates *Borrelia burgdorferi* transmission from feeding ticks to murine hosts. *J Infect Dis* 201: 1084–1095.
173. Maruskova M, Seshu J (2008) Deletion of BBA64, BBA65, and BBA66 loci does not alter the infectivity of *Borrelia burgdorferi* in the murine model of Lyme disease. *Infect Immun* 76: 5274–5284.
174. Maruskova M, Esteve-Gassent MD, Sexton VL, Seshu J (2008) Role of the BBA64 locus of *Borrelia burgdorferi* in early stages of infectivity in a murine model of Lyme disease. *Infect Immun* 76: 391–402.
175. Patton TG, Dietrich G, Dolan MC, Piesman J, Carroll JA, et al. (2011) Functional analysis of the *Borrelia burgdorferi* *bbA64* gene product in murine infection via tick infestation. *PLoS One* 6: e19536.
176. Xu H, He M, Pang X, Xu ZC, Piesman J, et al. (2010) Characterization of the highly regulated antigen BBA05 in the enzootic cycle of *Borrelia burgdorferi*. *Infect Immun* 78: 100–107.
177. Xu H, He M, He JJ, Yang XF (2010) Role of the surface lipoprotein BBA07 in the enzootic cycle of *Borrelia burgdorferi*. *Infect Immun* 78: 2910–2918.
178. Benoit VM, Fischer JR, Lin YP, Parveen N, Leong JM (2011) Allelic variation of the Lyme disease spirochete adhesin DbpA influences spirochetal binding to decorin, dermatan sulfate, and mammalian cells. *Infect Immun* 79: 3501–3509.
179. He M, Oman T, Xu H, Blevins J, Norgard MV, et al. (2008) Abrogation of *ospAB* constitutively activates the Rrp2-RpoN-RpoS pathway (sigmaN-sigmaS cascade) in *Borrelia burgdorferi*. *Mol Microbiol* 70: 1453–1464.
180. Raju BV, Esteve-Gassent MD, Karma SL, Miller CL, Van Laar TA, et al. (2011) Oligopeptide permease A5 (OppA5, BBA34) modulates vertebrate host-specific adaptation of *Borrelia burgdorferi*. *Infect Immun* 79: 3407–3420.
181. Shi Y, Xu Q, McShan K, Liang FT (2008) Both decorin-binding proteins A and B are critical for the overall virulence of *Borrelia burgdorferi*. *Infect Immun* 76: 1239–1246.
182. Weening EH, Parveen N, Trzeciakowski JP, Leong JM, Hook M, et al. (2008) *Borrelia burgdorferi* lacking DbpBA exhibits an early survival defect during experimental infection. *Infect Immun* 76: 5694–5705.
183. Probert WS, Crawford M, LeFebvre RB (1997) Antibodies to OspB prevent infection of C3H mice challenged with *Borrelia burgdorferi* isolates expressing truncated OspB antigens. *Vaccine* 15: 15–19.
184. Fikrig E, Tao H, Kantor FS, Barthold SW, Flavell RA (1993) Evasion of protective immunity by *Borrelia burgdorferi* by truncation of outer surface protein B. *Proc Natl Acad Sci U S A* 90: 4092–4096.
185. Guo BP, Brown EL, Dorward DW, Rosenberg LC, Hook M (1998) Decorin-binding adhesins from *Borrelia burgdorferi*. *Molecular Microbiology* 30: 711–723.
186. Hagman KE, Lahdenne P, Popova TG, Porcella SF, Akins DR, et al. (1998) Decorin-binding protein of *Borrelia burgdorferi* is encoded within a two-gene operon and is protective in the murine model of Lyme borreliosis. *Infect Immun* 66: 2674–2683.
187. Delilhas N (2009) Stem loop sequences specific to transposable element IS605 are found linked to lipoprotein genes in *Borrelia* plasmids. *PLoS ONE* 4: e7941.
188. Hudson RR, Kaplan NL (1985) Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics* 111: 147–164.
189. Porcella SF, Fitzpatrick CA, Bono JL (2000) Expression and immunological analysis of the plasmid-borne *mip* genes of *Borrelia burgdorferi* strain B31. *Infect Immun* 68: 4992–5001.
190. Yang XF, Hubner A, Popova TG, Hagman KE, Norgard MV (2003) Regulation of expression of the paralogous Mip family in *Borrelia burgdorferi*. *Infect Immun* 71: 5012–5020.
191. Zuckert WR, Meyer J, Barbour AG (1999) Comparative analysis and immunological characterization of the *Borrelia* Bdr protein family. *Infect Immun* 67: 3257–3266.
192. Zuckert WR, Barbour AG (2000) Stability of *Borrelia burgdorferi* *bdr* loci *in vitro* and *in vivo*. *Infect Immun* 68: 1727–1730.
193. Roberts DM, Carlyon JA, Theisen M, Marconi RT (2000) The *bdr* gene families of the Lyme disease and relapsing fever spirochetes: potential influence on biology, pathogenesis, and evolution. *Emerg Infect Dis* 6: 110–122.
194. Roberts DM, Caimano M, McDowell J, Theisen M, Holm A, et al. (2002) Environmental regulation and differential production of members of the Bdr protein family of *Borrelia burgdorferi*. *Infect Immun* 70: 7033–7041.
195. Gilmore RD, Jr., Mbow ML (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.
196. Bauer Y, Hofmann H, Jahraus O, Mytilineos J, Simon MM, et al. (2001) Prominent T cell response to a selectively *in vivo* expressed *Borrelia burgdorferi* outer surface protein (pG) in patients with Lyme disease. *Eur J Immunol* 31: 767–776.
197. Vergnaud G, Pourcel C (2009) Multiple locus variable number of tandem repeats analysis. *Methods Mol Biol* 551: 141–158.
198. Farlow J, Postic D, Smith KL, Jay Z, Baranton G, et al. (2002) Strain typing of *Borrelia burgdorferi*, *Borrelia afzelii*, and *Borrelia garinii* by using multiple-locus variable-number tandem repeat analysis. *J Clin Microbiol* 40: 4612–4618.
199. Guyard C, Chester EM, Raffel SJ, Schrupf ME, Policastro PF, et al. (2005) Relapsing fever spirochetes contain chromosomal genes with unique direct tandemly repeated sequences. *Infect Immun* 73: 3025–3037.
200. Carlyon JA, Roberts DM, Marconi RT (2000) Evolutionary and molecular analyses of the *Borrelia bdr* super gene family: delineation of distinct sub-families and demonstration of the genus wide conservation of putative functional domains, structural properties and repeat motifs. *Microb Pathog* 28: 89–105.
201. Benson G (1999) Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res* 27: 573–580.
202. Yang X, Coleman AS, Anguita J, Pal U (2009) A chromosomally encoded virulence factor protects the Lyme disease pathogen against host-adaptive immunity. *PLoS Pathog* 5: e1000326.
203. Eggers CH, Kimmel BJ, Bono JL, Elias AF, Rosa P, et al. (2001) Transduction by ϕ BB-1, a bacteriophage of *Borrelia burgdorferi*. *J Bacteriol* 183: 4771–4778.
204. Schulte-Spechtel U, Fingerle V, Goettner G, Rogge S, Wilske B (2006) Molecular analysis of decorin-binding protein A (DbpA) reveals five major groups among European *Borrelia burgdorferi* sensu lato strains with impact for the development of serological assays and indicates lateral gene transfer of the *dbpA* gene. *Int J Med Microbiol* 296 Suppl 40: 250–266.
205. Livey I, Gibbs CP, Schuster R, Dorner F (1995) Evidence for lateral transfer and recombination in *OspC* variation in Lyme disease *Borrelia*. *Mol Microbiol* 18: 257–269.
206. Wang G, van Dam AP, Dankert J (1999) Evidence for frequent *OspC* gene transfer between *Borrelia valaisiana* sp. nov. and other Lyme disease spirochetes. *FEMS Microbiol Lett* 177: 289–296.
207. Mongodin EF, Shapir N, Daugherty SC, DeBoy RT, Emerson JB, et al. (2006) Secrets of soil survival revealed by the genome sequence of *Arthrobacter aurescens* TC1. *PLoS Genet* 2: e214.
208. Pop M, Kosack DS, Salzberg SL (2004) Hierarchical scaffolding with Bambus. *Genome Res* 14: 149–159.
209. Claros MG, von Heijne G (1994) TopPred II: an improved software for membrane protein structure predictions. *Comput Appl Biosci* 10: 685–686.
210. Delcher AL, Kasif S, Fleischmann RD, Peterson J, White O, et al. (1999) Alignment of whole genomes. *Nucleic Acids Res* 27: 2369–2376.
211. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25: 3389–3402.
212. Enright AJ, Van Dongen S, Ouzounis CA (2002) An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res* 30: 1575–1584.
213. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22: 4673–4680.
214. Jeanmougin F, Thompson JD, Gouy M, Higgins DG, Gibson TJ (1998) Multiple sequence alignment with Clustal X. *Trends Biochem Sci* 23: 403–405.