Whole genome sequence and comparative analysis of *Borrelia burgdorferi* MM1

Neda Jabbari, Gustavo Glusman, and Christopher G. Lausted*

Institute for Systems Biology, Seattle, Washington, United States of America

*Corresponding author:
E-mail: clausted@systemsbiology.org (CGL)

Keywords: Borrelia burgdorferi; De novo assembly; Genomics; Linear plasmid; Lyme disease; Spirochetes

Abstract

Lyme disease is caused by spirochaetes of the *Borrelia burgdorferi* sensu lato (*B. burgdorferi* s.l.). Complete genome assemblies are available for fewer than ten strains of *B. burgdorferi* sensu stricto (s.s.), the primary cause of Lyme disease in North America. MM1 is a strain of *B. burgdorferi* s.s. that was originally isolated in the midwestern United States. Except for a few genes, the complete genome sequence of this strain has not been reported.

Here we present the complete genome sequence of *B. burgdorferi* MM1 in relation to other *B. burgdorferi* strains, and in terms of its Multi Locus Sequence Typing (MLST). Our results indicate that *B. burgdorferi* MM1 is a new Sequence Type that contains a conserved main chromosome and 15 plasmids. Our results include the first contiguous 28.5 kbp assembly of lp28-8, a linear plasmid carrying the *vls* antigenic variation system, from a *B. burgdorferi* s.s. strain.

Introduction

Lyme disease is the most prevalent tick-borne disease in North America. The most frequent initial manifestation is erythema migrans, a characteristic expanding ring-shaped cutaneous lesion[1,2]. While some patients experience erythema migrans with or without flu-like symptoms, others may develop extracutaneous manifestations that can affect the nervous system (neuroborreliosis), the heart (Lyme carditis) or the joints (Lyme arthritis)[3–5]. What role pathogen genetic diversity plays in the producing varying symptoms is unknown. Currently, genomic sequence of forty seven strains from the *B. burgdorferi* s.s., the primary clinically relevant species in North America[6], are available from NIH GenBank, of which only fewer than ten have complete assembly. Given the complexity of Lyme manifestations and the scarcity of sequenced genomes, sequencing and characterization of more strains is vital for the understanding of *Borrelia* genetic diversity and disease manifestations.

B. burgdorferi MM1 was isolated from the kidney of a white-footed mouse, *Peromyscus leucopus*, in Minnesota [7,8]. It was used in early studies of Lyme disease [7,9–12] and it appears in patent applications for the development of *B. burgdorferi* detection assays and recombinant vaccines against Lyme disease [13,14]. However, its whole genome sequence has not yet been reported. We present here the full genome sequence of *B. burgdorferi* MM1, and provide comparative analysis, MLST and annotation of its genome. Most notably, we report on the presence of the lp28-8 plasmid which contains its *vls* antigenic variation system.

Materials and methods

Growth conditions and DNA preparation

B. burgdorferi MM1 spirochete was purchased from ATCC and cultured in BSK-H with 10% rabbit serum (Sigma-Aldrich). Total DNA was isolated from pelleted spirochetes of the fourth passage of the ATCC 51990 using DNeasy blood and tissue kit (Qiagen). RNase A (Qiagen) was used to digest the RNA content as recommended by the manufacturer's protocol. DNA was purified using the Agencourt AMPure XP method (Beckman Coulter) and quantified with the Quant-iT PicoGreen dsDNA Assay Kit (ThermoFisher Scientific) and a NanoDrop ND-1000 spectrophotometer. Integrity of the isolated DNA was checked on a 0.6 % agarose gel as well as with an Agilent DNA 12000 Kit (Agilent Technologies, Inc.).

Genome sequencing

DNA sequence data was primarily obtained using the Pacific Biosciences (PacBio) Single Molecule, Real-Time (SMRT) system with DNA/Polymerase Binding Kit P6 v2, MagBead Kit v1 and DNA Sequencing Reagent Kit 4.0 v2 at University of Washington PacBio Sequencing Services. Long DNA reads were collected in one SMRT cell on a PacBio RSII instrument. Additional long reads were obtained using the Oxford Nanopore Technologies (ONT) MinION system. The MinION genomic DNA library was prepared according to the ONT Nanopore Sequencing Kit SQK-MAP006 protocol and quantified with Quant-iT PicoGreen dsDNA Assay Kit (ThermoFisher Scientific). Sequencing reads were generated on Flow Cells (Flo-MAP003) with MinKNOW 0.51.1.62 and base-called with Metrichor 2.39. "2D pass" reads were converted into FASTA format with poretools v. 0.5.1. These reads are the highest quality

reads, selected from reads where both strands of a double-stranded fragment were analyzed together in one pore.

Genome assembly

De novo assembly of PacBio reads was performed with Hierarchical Genome Assembly Process (HGAP v2) using a minimum seed read length of 7800 bp. Contigs were merged or circularized with Circlator (v 1.5.0) [15]. In order to acquire the final consensus sequence, raw reads were filtered and mapped to the *B.burgdorferi* MM1 assembly with the resequencing protocol as implemented in SMRT Analysis Software v2.3.0.

Sequence characterization and visualisation

Tandem repeats were defined with Unipro UGENE v1.9.8 as follows: 1–9 bp, 10–100 bp and >100 bp unit sizes were termed micro-, mini- and macro-satellite, respectively, and overlapping tandem repeats were excluded [16]. BLAST alignments of the sequences along with annotations were plotted using Easyfig (v 2.2.2) [17] and the Artemis Comparison Tool (ACT; Release 13.0.0) [17,18].

B. burgdorferi MM1 plasmid identification and nomenclature

The plasmid content of MM1 was first evaluated visually with respect to the standard type strain B31. Reads were mapped to the B31 genome with the "resequencing" protocol included in PacBio's SMRT Analysis software v2.3.0 (S1 Fig). Next, MM1 plasmid identities were confirmed by the sequence type of the *Borrelia* paralogous family (PFam) 32 protein encoded by each plasmid as defined by Casjens et al. [19]. Specifically, the protein sequences of

95 MM1 and three fully-sequenced *B.burgdorferi* s.s. strains (B31, N40 and JD1; accessions
96 GCA_000008685.2, GCA_000166635.2 and GCA_000166655.2, respectively) were aligned
97 with BLAST using an E-value cutoff of 0.001. These alignments were then clustered using
98 Spectral Clustering of Protein Sequences (SCPS) [20] to infer homology relations between
99 sequences based on pairwise similarity scores. Members of the four protein families, PFam32,
100 PFam49, PFam50 and PFam57/62, as reasonable candidates for functioning in plasmid
101 replication and partitioning [21], were identified on each contig (S1 Table). Plasmids were
102 assigned names according to the PFam32 type matching their encoding PFam32 member (>95%
103 amino acid sequence identity). Two exceptions were made to this method: cp9 and lp28-8. The
104 smallest plasmid, cp9, does not encode PFam32, so it was identified by the sequence type of its
105 PFam49, PFam50 and PFam57/62 members (S1 Table). Other than cp9, short contigs with no
106 identifiable PFam32 type were not considered. The PFam32 protein from what would eventually
107 be identified as the MM1 lp28-8 plasmid shared only a 60% amino acid sequence identity with
108 PFam32 sequence type from B31 and JD1 cp32-3 plasmids. Meanwhile a protein BLAST against
109 the NCBI non-redundant protein database, revealed it as fully identical to PFam32 type on *B.*
110 *burgdorferi* 94a lp28-8 plasmid.
111 The MM1 lp54 plasmid subtype was identified following the criteria by Casjens et
112 al.[22]. Dot plot analysis was performed to identify large (>400 bp) insertions, inversions or
113 deletions in MM1 lp54 plasmid relative to lp54 subtypes. Nucleotide BLAST sequence
114 comparisons and overall identical plasmid organization was investigated with Artemis
115 Comparison Tool (ACT; Release 13.0.0) [18].

116 ## Functional annotation

117 Genome features were annotated with RAST v2.0 [23] with the initial NCBI Taxonomy
118 ID set to *B. burgdorferi* (139). To create a comparative annotation map of main chromosomes,
119 FASTA sequences of all *B. burgdorferi* strains with complete genome assemblies were
120 download from NCBI GenBank. B31, CA283, JD1, N40, PABe, PAli, and ZS7 (accession
121 numbers AE000783.1, CP005925.1, CP002312.1, CP002228.1, CP019916.1, CP019844.1, and
122 CP001205.1, respectively). B31, CA283, JD1, N40, and ZS7 were re-annotated with RAST
123 annotation server using the same settings used for MM1 genome annotation. GenBank
124 annotations for PAbe and PAli were used directly.

125 ## Phylogenetic analysis

126 A phylogenetic tree was constructed to evaluate the relationship of MM1 with respect to
127 seven prominent laboratory *B. burgdorferi* sensu stricto strains. The tree was inferred from the
128 whole main chromosome sequence data of sensu stricto strains MM1, B31, CA283, JD1, N40,
129 PABe, PAli, and ZS7 using REALPHY v.1.12[24] with outgroup *B. bissettii* DN127
130 (CP002746.1). The strains were ordered for plotting based on genomic distance from MM1 as
131 determined by GGDC v.2.1[25] with its default settings for formula two. The tree was plotted
132 using TreeGraph2 v2.13[26].

133 The eight chromosomal housekeeping genes *clpA*, *clpX*, *nifS*, *pepX*, *pyrG*, *recG*, *rplB* and
134 *uvrA* of the Multi Locus Sequence Typing (MLST) scheme [27] were used to determine MM1
135 Sequence Type (ST). Corresponding locus sequences in MM1 were queried against the *Borrelia*
136 MLST database [28]. Alleles with perfect matches to the database were assigned type numbers.

137 Where perfect matches did not exist, single-base mismatches were noted and confirmed using
138 Oxford Nanopore data. For confirmation, the novel alleles were searched against all MinION
139 filtered 2D reads using FASTA v.36[29]. Hits were aligned with Clustal Omega v.1.2.1[30] and
140 visually inspected.

141 Results and discussion

142 Overall description of the *B. burgdorferi* MM1 genome

143 The genome of the sequenced culture of *B. burgdorferi* MM1 was determined to be
144 1,280,240 bp in size with an average coverage depth of 737× and ranging from 351× to 2312×
145 across the chromosome and plasmids (Table 1). It consists of a main linear chromosome
146 representing 71% of the genome size, and a plasmid content of 371,728 bp in size. According to
147 RAST analysis, *B. burgdorferi* MM1 genome carries 1338 annotated genes with 865 (65%) on
148 the main chromosome including 36 RNA coding genes and 473 plasmid features (Table 1). Of
149 all the annotated features, 37% are connected to a subsystem which are mainly (92%) on the
150 chromosome (Table 1). From among the subsystem categories present in MM1 genome, Protein
151 Metabolism and Sulfur Metabolism have the highest and lowest subsystem feature counts,
152 respectively. Interestingly, all the subsystem categories except for "Phages, Prophages,
153 Transposable elements, Plasmids" category are also present in *B. burgdorferi* B31 genome
154 (Table 2).

155 Given that the plasmid profile of *B. burgdorferi* may undergo changes as a result of
156 prolonged *in vitro* cultivation [31–33], our sequenced low-passage culture of *B. burgdorferi*

157 MM1 (see Materials and methods), like other sequenced *Borrelia,* might still be no longer

158 homogeneous with some lost plasmids during culturing.

159

160 **Table 1. GenBank, coverage and Annotation Details for *B. burgdorferi* MM1 Genome**

Replicon[a]	Accession	Base Coverage	Annotations[o]
Chromosome		782	865
lp54		804	74
lp17		832	17
lp25		637	25
lp28-3		550	25
lp28-4		655	31
lp28-8		589	29
lp36		789	34
cp26		725	27
cp9		2312	12
cp32-1		555	53
cp32-4		435	20
cp32-5		376	31
cp32-6		741	28
cp32-7		351	20
cp32-9		662	47

161 [a] Other GenBank details are as follows: BioProject: , BioSample: , Assembly:

162 [o]Of all annotated features, 456 chromosomal and 39 total plasmid features are linked to a subsystem.

163

164 **Table 2. Subsystem Categories Present in MM1 Genome in Relation to B31**

	Subsystem feature counts	
Category	**MM1**	**B31**
Protein Metabolism	171	164
RNA Metabolism	68	82
Motility and Chemotaxis	62	58
DNA Metabolism	55	55

Carbohydrates	48	49
Nucleosides and Nucleotides	38	34
Fatty Acids, Lipids, and Isoprenoids	37	35
Cell Wall and Capsule	36	41
Amino Acids and Derivatives	32	32
Cell Division and Cell Cycle	26	26
Virulence, Disease and Defense	22	23
Stress Response	21	15
Membrane Transport	18	15
Cofactors, Vitamins, Prosthetic Groups, Pigments	17	21
Phosphorus Metabolism	10	14
Respiration	9	9
Regulation and Cell signaling	7	12
Potassium metabolism	6	4
Phages, Prophages, Transposable elements, Plasmids	6	_
Miscellaneous	3	10
Dormancy and Sporulation	2	2
Sulfur Metabolism	1	1

The main, linear chromosome of *B. burgdorferi* MM1

B. burgdorferi MM1 carries a main linear chromosome of 908512 bps in length with a GC content of 28.6% within the size and GC content range (902191-922801 bps and 28.5-28.6% respectively) of main chromosome in *B. burgdorferi* s.s. strains that have a complete genome assembly (Table 3). Consistently, *B. burgdorferi* MM1 chromosome shares a 99% nucleotide sequence identity and a similar tandem repeat profile (e.g., microsatellite, minisatellite and macrosatellite) with the main chromosome in *B. burgdorferi* s.s. strains (Table 2).

Table 3. *B. burgdorferi* MM1 Chromosome in Relation to Representative *B. burgdorferi* Strains

Strain	Length (bp)	GC%	Tandem Repeats		
			Micro-satellite	Mini-satellite	Macro-satellite
B. burgdorferi MM1	908512	28.6	933	1	1
B. burgdorferi JD1	922801	28.6	951	2	0
B. burgdorferi PAbe	910728	28.6	939	2	1
B. burgdorferi B31	910724	28.6	940	2	2
B. burgdorferi CA382	910736	28.6	944	2	1
B. burgdorferi PAli	909921	28.6	940	2	0
B. burgdorferi ZS7	906707	28.5	947	2	0
B. burgdorferi N40	902191	28.6	931	0	0

With the exception of the right end extension, the chromosomal gene contents of sequenced *B. burgdorferi* s.s. strains are essentially identical [34,35]. Accordingly, comparative annotation map of MM1 and *B. burgdorferi* s.s. strains suggests a consistent overall gene synteny over the entire chromosome with some variation in right end extension (Fig 1).

Fig 1. Comparative Annotation Map of the Main Chromosome of *B. burgdorferi* s.s. Strains with Complete Genome Assembly. Rectangles represent annotated genes, color-coded by type: CDS in light blue, predicted CDS in dark blue, hypothetical CDS in red, rRNA in green and tRNA in black. Rectangle heights denote gene length; genes shorter than 1 kb are displayed enlarged for visibility. Features above or below the midline for each strain represent genes transcribed on the top or bottom strand, respectively. The phylogenetic tree was inferred from the whole main chromosome sequence data and rooted using *B. bissettii* DN127 as outgroup.

B. burgdorferi MM1 plasmid content and variation from *B.burgdorferi* strains

An initial mapping of raw sequence reads from *B.burgdorferi* MM1 genome to the type strain B31 (see Materials and methods) indicated less than 26% base calling for B31 plasmids lp21, lp28-1, lp28-2, lp38, lp56 and lp5 suggesting their absence in MM1 genome. However, the presence/absence of other B31 plasmids and, in particular, the cp32s that are homologous nearly

195 throughout their lengths in *B. burgdorferi [34]* remained ambiguous. The observed depth of
196 coverage across B31 plasmids is depicted in Fig. S1.

197 According to the PFam32 nomenclature scheme [19], the sequenced culture of *B.*
198 *burgdorferi* MM1 is determined to carry fifteen plasmids encoding 7 linear and 7 circular
199 plasmid PFam32 types. The small cp9 plasmid that does not encode a PFam32 protein defines an
200 additional plasmid compatibility type in MM1 genome. While some plasmids might be
201 spontaneously lost between original isolation and the low passage culture used for sequencing,
202 the number of identified plasmids in MM1 is within the previously defined range in a set of 14 *B.*
203 *burgdorferi* strains (e.g. 6-11 and 6-12 circular and linear sequenced plasmids, respectively) [22].

204 A comparison of variation in plasmid profile, size and GC content among MM1, B31 ,
205 JD1 and N40 strains, is shown in Table 4. Similar to the type strain B31, MM1 carries linear
206 plasmids lp54, lp17, lp25, lp28-3, lp28-4, lp36 and circular plasmids cp9, cp26, cp32-1, cp32-4,
207 cp32-6, cp32-7, cp32-9. Among the non-cp32 plasmids, lp17, lp25, and lp36 appear to have
208 major structural differences between the two strains (Table 4, Fig 2). MM1 genome carries the
209 linear plasmid PFam32 type, lp28-8, that is not present in the sequenced B31. In addition, MM1
210 carries cp32-5 PFam32 type which is not present among encoded by B31 plasmids but is found
211 among encoded by the circular plasmids of N40 and present in strains JD1 in fusion with cp32-5
212 as cp32-1+5 (Table 4) [34].

213
214 **Table 4. *B. burgdorferi* MM1 plasmid content in relation to *B. burgdorferi* B31, JD1 and N40 strains.**

Replicon[a]	Code	MM1		B31		JD1		N40	
		Length	GC%	Length	GC%	Length	GC%	Length	GC%

lp54	**A**	53.8	28.1	53.7	28.1	52.9	28.2	53.8	28.1
lp17	**D**	14.4	23.7	16.8	23.1	17.5	23.9	20.7	24.3
lp21	**U**	_	_	18.8	20.7	_	_	_	_
lp25	**E**	24.2	25.4	24.2	23.4	22.8	25.4	23.4	23.3
lp28-1	**F**	_	_	28.2	32.8	24.4	32.7	_	_
lp28-2	**G**	_	_	29.8	31.6	_	_	29.5	31.7
lp28-3	**H**	26.8	25.0	28.6	25.0	28.8	25.0	_	_
lp28-4	**I**	27.5	24.5	27.3	24.5	30.1	24.7	29.4	24.6
lp28-5	**Y**	_	_	_	_	24.6	23.9	26.6	25.3
lp28-6	**Z**	_	_	_	_	27.0	32.4	_	_
lp28-7	**AA**	_	_	_	_	30.0	31.6	_	_
lp28-8º	**RC**	28.5	33.2	_	_	_	_	_	_
lp36	**K**	28.0	25.5	36.8	26.9	22.8	25.6	29.2	31.6
lp38	**J**	_	_	38.8	26.1	27.6	22.9	37.9	26.0
lp5	**T**	_	_	5.2	23.8	_	_	_	_
lp56	**Q**	_	_	53.0	27.3	_	_	_	_
cp26	**B**	26.5	26.3	26.5	26.3	26.5	26.2	26.5	26.3
cp9	**C**	11.4	24.3	9.4	23.7	_	_	8.7	24.3
cp32-1	**P**	30.9	29.3	30.8	29.4	60.7[n]	29.2	_	_
cp32-3	**S**	_	_	30.2	28.9	29.7	28.9	_	_
cp32-4	**R**	13.3	26.8	30.3	29.3	_	_	15.9	27.0
cp32-5	**V**	21.9	28.5	_	_	60.7[n]	29.2	29.6	29.2
cp32-6	**M**	20.0	27.7	29.8	29.3	30.8	29.0	_	_
cp32-7	**O**	14.0	26.3	30.8	29.1	_	_	17.5	27.2
cp32-8	**L**	_	_	30.9	29.1	31.1	29.1	_	_
cp32-9	**N**	30.6	29.1	30.7	29.3	30.7	29.2	30.7	29.0
cp32-10	**Q**	_	_	_	_	30.3	29.3	30.2	29.2
cp32-11	**W**	_	_	_	_	29.9	29.0	_	_
cp32-12	**X**	_	_	_	_	30.0	29.3	27.8	28.1
Total plasmids		15 : 7 (L), 8 (C)		21: 12 (L), 9 (C)		20: 11 (L), 9 (C)		16: 8 (L), 8 (C)	

[a]Length is listed in kbps.

[o]In strains of *B. afzelii*, where lp28-8 is documented (e.g. K78 and PKo), code "AC" is initially given to plasmid lp28-8 and later revised to "RS", whereas, in others (e.g. *B. burgdorferi* strain 94a, *B. valaisiana* VS116, *B. spielmanii* A14S), code N is initially given and later revised to "RS" [22]. Consistently, we designate "RS" as the code for MM1 lp28-8 plasmid.
[n]JD1cp32-1 and cp32-5 are described as fused into 60.7 kbps circular cp32-1+5 plasmid [34]. □In B31 cp32-10 is integrated into lp56 plasmid [34]. L, linear; C, Circular.

Fig 2. A Diagrammatic Depiction of the Differences in the Organization of the Non-cp32 Plasmids Present in *B. burgdorferi* MM1 and B31 strains. Annotated genes with or without CDSs are represented as grey arrows. PFam organization and synteny is displayed on each plasmid in MM1 (top) and B31 (bottom) strains as red (PFam32), green (PFam49), yellow (PFam50) and blue (PFam57/62) arrows. Among B31the 4 PFams in B31, genes that appear to be disrupted and pseudogene relatives are not color coded [21] and only PFam types with consecutive numbers in MM1 are color coded. BLAST alignments between plasmid pairs (E-value <0.001) are indicated by blocks ranging in color from red (80% sequence identity) to dark blue (100% sequence identity). GC content is calculated in a window of 400 bps and displayed as a blue histogram bar of a GC content lower than 50%. Sequences are shown in full and drawn to scale with scale bars representing 5 kbps.

The plasmids lp54 and cp26 in *B. burgdorferi* MM1

The linear plasmid lp54 and the circular plasmid cp26 have for the most part conserved synteny among the *B. burgdorferi* s.l. genomes [36,37]. Among the lp54 plasmid genes with low sequence identity across *Borrelia* species, are the genes encoding decorin binding proteins A (DbpA) and B (DbpB) that are critical for the overall virulence of *B. burgdorferi [38]*. DbpA is highly heterogeneous between and within *B. burgdorferi* s.l. species with as low as 70% sequence identity within *B. burgdorferi* s.s., whereas DbpB is more conserved (>96% identity) within *B. burgdorferi* s.s. [39–41]. Consistent with the substantial difference in their sequence heterogeneity among *Borrelia* isolates, DbpA and DbpB proteins of MM1 (corresponding gene locus tags BbuMM1_A250 and BbuMM1_A260, respectively) share 79% and 100% sequence identity to B31 (Fig 3).

Fig 3. Synteny Map of Lp54 Containing the PFam54 Gene Array in MM1, JD1, B31 and N40 Strains. Left: The lp54 plasmid of MM1 *B. burgdorferi* in comparison to JD1, B31 and N40 strains.The

strains are indicated in the left side of the map along with organizational subtypes indicated by Roman numerals in parenthesis. Plasmid is represented by a horizontal white bar labeled with positions with annotations on light grey bars. BLAST alignments between the lp54 assemblies are indicated by blocks ranging in color from green (95% sequence identity) to dark blue (100% sequence identity) with a minimum Score Cutoff of 28. Annotated features are indicated in dark grey and orange (dbpA and dbpB genes, PFam54 gene array). Right: A zoomed view of PFam54 gene array of lp54 plasmid in JD1, B31, MM1 and N40 strains. PFam54 variable region is sandwiched by *bba64*, *bba65*, and *bba66* on one side, and *bba73* on the other side. Among the 4 strains, MM1 variable region is highly conserved with N40.

Near one terminus of the lp54 plasmid, we identified a PFam54 gene array of close homologs of complement regulator-acquiring surface protein 1, CRASP-1, that has the most variable gene content caused by gene duplication, loss and sequence diversification [42,43]. This variable region is between the genes homologous to *bba66* and *bba73* in a set of 10 *Borrelia* lp54 plasmids [42]. The gene order within MM1 lp54 variable region is very different from B31, in that MM1 carries Bbu_MM1A660 located between genes homologous to *bba66* and *bba68*, while B31 carries *bba70* gene that is absent in MM1 (Fig 3). A comparison of the lp54 plasmid variable region in MM1 and N40 strains reveals highly conserved synteny and sequence, with 99% sequence identity between the Bbu_MM1A660 and BbuN40_A67a genes (Fig 3). MM1 and N40 lp54 plasmids have conserved gene order throughout their length. They do not carry insertions, inversions or deletions larger than 400 bps and inter-plasmid DNA exchanges relative to one another [22], therefore, MM1 lp54 plasmid appears to be a type II lp54.

B. burgdorferi cp26 plasmid is present in all natural isolates and encodes functions critical to bacterial viability [44]. The highly diverse outer-surface protein C (OspC) against which mammals develop protective immunity is encoded by the *ospC* gene on the cp26 plasmid [45]**.** MM1 strain carries a type U *ospC* allele with 100% nucleotide sequence identity to *ospC*

alleles U in Northern US *B. burgdorferi* s.s. strains 94a and CS5[46] 84% sequence identity with type A *ospC* allele in B31 strain.

B. burgdorferi MM1 plasmid lp28-8

MM1 plasmid lp28-8 encoding PFam32 protein is 100% and 90% identical in amino acid sequence to the lp28-8 encoding PFam32 proteins in *B. burgdorferi* 94a and *B. valaisiana* VS116, respectively. MM1 lp28-8 carries a typical cluster of PFam32, PFam49, PFam50, PFam57/62 that is 100% identical to the corresponding region of lp28-8 in *B. burgdorferi* 94a [22] and 86% identical to the corresponding region of lp28-8 in *B. afzelii* strains PKo and K78, *B. valaisiana* VS116 and *B. spielmanii* A14S (Fig 4).

Fig 4. A Diagrammatic Depiction of the Differences in the Organization of the Lp28-8 Plasmids in the *B. burgdorferi* MM1 and Other *B. burgdorferi* s.l. species. Annotations are represented as grey arrows. MM1 PFam organization and synteny is displayed on each alignment as red (PFam32), green (PFam49), yellow (PFam50) and blue (PFam57/62) arrows. Only PFam types with consecutive numbers are color coded. Blast sequence comparison (E-value <0.001) represents a similarity range from 80% (red) to 100% (blue) among plasmid pairs. GC content is calculated in a window of 400 bps and displayed as a blue histogram bar of a GC content less or higher than 50%. The size of lp28-8 in *B. burgdorferi* s.s. 94a (largest contig), *B. afzelii* strain 78, *B. afzelii* strain PKo, *B. valaisiana* VS 116, and *B. spielmanii* A14S is 21,295, 28,638 , 20715, 18022, and 24828 bps respectively. Sequences are shown in full and scale bars represent 5 kbps.

This assembly of the MM1 lp28-8 plasmid is the most complete for a B. burgdorferi s.s. to date. Previous contigs had not approached 28 kb [22,34,47]. In fact, it was recently described as a new PFam32 type present only in one *B. burgdorferi* s.s. strain, 94a, from among a set of 14 characterized *B. burgdorferi* s.s. isolates. Due to difficulties in assembling long repetitive tracts, the sequence of lp28-8 plasmid of 94a remains unfinished and reported as contigs of 21,295, 4,910 and 393 base pairs [22]. Our sequenced *B. burgdorferi* MM1 lp28-8 is a continuous contig

301 of 28,515 bps in length that is 99% identical to lp28-8 in *B. burgdorferi* 94a over 95% of its
302 sequence, and its leftmost 13 kbp shares a 99% identity to plasmid lp38 subtype V. A
303 comparison of sequence and organization of lp28-8 in MM1 to lp28-8 in *B. burgdorferi* 94a and
304 *B. burgdorferi* s.l. species such as *B. afzelii [48,49]*, *B. valaisiana* and *B. spielmanii* [50] is
305 depicted in Fig 4.

306 MM1 lp28-8 plasmid carries a *vls* locus in one end with a characteristic high GC content
307 [51]. The *vls* locus is required for long-term survival of Lyme *Borrelia* in infected mammals and
308 is characterized with a *cis* location of the expression site *vlsE* and a contiguous array of *vls* silent
309 cassettes [51]. Consistently, the average GC content of annotations within the *vls* locus of MM1
310 is 47% leading to an elevated average GC content of 33.2% for plasmid lp28-8, while the GC
311 content range for the rest of MM1 genome is 23.7% to 29.3% (Table 4, Fig 4). MM1 lp28-8
312 includes an array of 18 sequences with at least 85% sequence identity to *vlsE* (Fig 5).

313
314 **Fig 5. The *Vls* Antigenic Variation System in *B. burgdorferi* MM1.** This locus on lp28-8 plasmid
315 contains the expression site *vlsE* as well as a contiguous array of 18 sequences, that are at least 85%
316 identical in DNA sequence to *vlsE* (BbuMM1_RS290). BLAST alignments between the MM1 *vlsE* locus
317 and lp28-8 plasmid are indicated by blocks ranging in color from green (85% sequence identity) to dark
318 blue (100% sequence identity) with a minimum Score Cutoff of 85. The leftmost 13134 bps of lp28-8
319 plasmid is 99% identical to lp38 subtype V.
320

321 The cp32 plasmids of *B. burgdorferi* MM1

322 The cp32 plasmid family members are described to have 12 PFam32 types in a set of 14
323 fully analyzed *B. burgdorferi* s.s. genome sequences [22]. *B. burgdorferi* MM1 genome contains
324 six members of the cp32 family, cp32-1, cp32-4, cp32-5, cp32-6, cp32-7, and cp32-9 (Table 4).
325 Based on SCPS analysis, the MM1 cp32 plasmids consistently share an amino acid sequence

326 identity of more than 99% with their PFam32 genes in either B31 or N40 strains (S1 Table).
327 Among the six, cp32-5 plasmid is absent in *B. burgdorferi* B31.

328 MLST reveals *B. burgdorferi* MM1 strain is a new sequence type

329 Based on the MLST scheme, no isolate in *Borrelia* MLST database [28] was found to
330 have the same allelic profile as MM1, therefore, MM1 is characterized as a unique sequence type
331 (ST). Of the eight loci used in the MLST scheme, six loci (*clpA*, *clpX*, *pyrG*, *recG*, *rplB* and *uvrA*
332 excluding *nifS* and *pepX*) in MM1 genome were the same allelic type as the corresponding loci in
333 ST37 and ST18 *B. burgdorferi* s.s. isolates within the *Borrelia* MLST database (Table 5). *NifS*
334 and *pepX* loci were partially matched (different in only 1 nt) to allelic types 12 and 1,
335 respectively. The closest allelic profile to MM1 was observed in ST37 isolates with 1 bp
336 mismatch in *nifS* loci. Interestingly, ST37 and ST18 isolates belong to Northeastern US or
337 Eastern/East-central Canada, while, MM1 is a midwestern US isolate.

338

339 **Table 5. MM1 MLST in relation to isolates (ST 37 and ST 18) with 6 out of 8 allelic types identical to MM1**

isolate	country[a]	source[o]	allelic type							
			clpA	***clpX***	***nifS***	***pepX***	***pyrG***	***recG***	***rplB***	***uvrA***
MM1[o]	US	M	7	6	_	_	1	5	5	5
ST37	US (19), CA (2)	H (2), T(19)	7	6	12	1	1	5	5	5
ST18	US (20), CA (5)	H (7), T (18)	7	6	6	1	1	5	5	5

340 [a]All US locations are Northeastern US and Eastern/East-central Canada locations respectively. M, mouse; H, human,
341 T, tick. [o]MM1 *nifS* and *pepX* loci match partially (1 nt difference) to *nifS*-12 and *pepX*-1, respectively.
342

343 # Conclusions

344 We reported the full genome sequence and organization of *B. burgdorferi* MM1. Like
345 other *B. burgdorferi* s.s. strains, MM1 appears to have a quite evolutionarily stable chromosome
346 with little variation in size and content [19,34,35]. Our sequenced culture of MM1 has 15
347 plasmids of which 7 are linear and 8 circular. Based on the unique allelic profile of MM1 in
348 MLST analysis, our study revealed the existence of a new ST among northern US *B. burgdorferi*
349 s.s. isolates. Except for one *B. burgdorferi* s.s. strain (94a), the lp28-8 plasmid had not been
350 detected in sequenced cultures of *B. burgdorferi* s.s. strains. We report on the presence of lp28-8
351 plasmid in MM1 as another strain, in addition to strain 94a, among all *B. burgdorferi* s.s. strains
352 that are fully studied and reported to date. MM1 carries *vls* antigenic variation system on its
353 lp28-8 plasmid which consists of the expression site *vlsE* and an array of 18 cassettes. The
354 characterization of MM1 as a midwestern US isolate and a well-established *B. burgdorferi* s.s.
355 strain in Lyme research and vaccine studies, contributes to understanding *B.burgdorferi* s.s.
356 diversity, and will facilitate the development of more specific diagnostics and vaccines.

357 # Acknowledgments

358 We thank the Moritz Lab for culturing the bacterial cells used in this study. This work
359 was supported by The Wilke Family Foundation, The Steven and Alexandra Cohen Foundation,
360 Jeff and MacKenzie Bezos, Global Lyme Alliance, Bay Area Lyme Foundation and the NIGMS
361 National Centers for Systems Biology Grant 2P50GM076547-06A1.

362
363

References

1. Swedish Council on Health Technology Assessment. Treatment Duration for Lyme Disease. Stockholm: Swedish Council on Health Technology Assessment (SBU); 2016.

2. Lyme disease. Recognising and treating erythema migrans. Prescrire Int. 2015;24: 247–249.

3. Murray TS, Shapiro ED. Lyme disease. Clin Lab Med. 2010;30: 311–328.

4. Hengge UR, Tannapfel A, Tyring SK, Erbel R, Arendt G, Ruzicka T. Lyme borreliosis. Lancet Infect Dis. 2003;3: 489–500.

5. Robinson ML, Kobayashi T, Higgins Y, Calkins H, Melia MT. Lyme carditis. Infect Dis Clin North Am. 2015;29: 255–268.

6. Feder HM Jr, Johnson BJB, O'Connell S, Shapiro ED, Steere AC, Wormser GP, et al. A critical appraisal of "chronic Lyme disease." N Engl J Med. 2007;357: 1422–1430.

7. Xu Y, Johnson RC. Analysis and comparison of plasmid profiles of Borrelia burgdorferi sensu lato strains. J Clin Microbiol. 1995;33: 2679–2685.

8. Loken KI, Wu CC, Johnson RC, Bey RF. Isolation of the Lyme disease spirochete from mammals in Minnesota. Proc Soc Exp Biol Med. 1985;179: 300–302.

9. Lee S-H, Lee J-H, Park H-S, Jang W-J, Koh S-E, Yang Y-M, et al. Differentiation of Borrelia burgdorferi sensu lato through groEL gene analysis. FEMS Microbiol Lett. 2003;222: 51–57.

10. Hughes CA, Johnson RC. Methylated DNA in Borrelia species. J Bacteriol. 1990;172: 6602–6604.

11. Miller JC, Stevenson B. Immunological and genetic characterization of Borrelia burgdorferi BapA and EppA proteins. Microbiology. 2003;149: 1113–1125.

12. Xu Y, Kodner C, Coleman L, Johnson RC. Correlation of plasmids with infectivity of Borrelia burgdorferi sensu stricto type strain B31. Infect Immun. 1996;64: 3870–3876.

13. Anthony C. Caputa, Russell F. Bey, Michael P. Murtaugh. Recombinant vaccine against Lyme disease. US Patent. 5554371 A, 1996.

14. Roger N. Picken HCA. Dna probes and primers for detection of b. burgdorferi using the polymerase chain reaction. European Patent. 0487709 B1, 1996.

394 15. Hunt M, Silva ND, Otto TD, Parkhill J, Keane JA, Harris SR. Circlator: automated
395 circularization of genome assemblies using long sequencing reads. Genome Biol. 2015;16:
396 294.

397 16. Zhou K, Aertsen A, Michiels CW. The role of variable DNA tandem repeats in bacterial
398 adaptation. FEMS Microbiol Rev. 2014;38: 119–141.

399 17. Sullivan MJ, Petty NK, Beatson SA. Easyfig: a genome comparison visualizer.
400 Bioinformatics. 2011;27: 1009–1010.

401 18. Carver TJ, Rutherford KM, Berriman M, Rajandream M-A, Barrell BG, Parkhill J. ACT: the
402 Artemis Comparison Tool. Bioinformatics. 2005;21: 3422–3423.

403 19. Sherwood R. Casjens CHEAIS. Borrelia genomics: chromosome, plasmids, bacteriophages
404 and genetic variation. In: D. Scott Samuels and Justin D. Radolf, editor. Borrelia: molecular
405 biology, host interaction and pathogenesis. Caister Academic Press; pp. 27–54.

406 20. Nepusz T, Sasidharan R, Paccanaro A. SCPS: a fast implementation of a spectral method for
407 detecting protein families on a genome-wide scale. BMC Bioinformatics. 2010;11: 120.

408 21. Casjens S, Palmer N, van Vugt R, Huang WM, Stevenson B, Rosa P, et al. A bacterial
409 genome in flux: the twelve linear and nine circular extrachromosomal DNAs in an infectious
410 isolate of the Lyme disease spirochete Borrelia burgdorferi. Mol Microbiol. 2000;35: 490–
411 516.

412 22. Casjens SR, Gilcrease EB, Vujadinovic M, Mongodin EF, Luft BJ, Schutzer SE, et al.
413 Plasmid diversity and phylogenetic consistency in the Lyme disease agent Borrelia
414 burgdorferi. BMC Genomics. 2017;18: 165.

415 23. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, et al. The RAST Server:
416 rapid annotations using subsystems technology. BMC Genomics. 2008;9: 75.

417 24. Bertels F, Silander OK, Pachkov M, Rainey PB, van Nimwegen E. Automated
418 reconstruction of whole-genome phylogenies from short-sequence reads. Mol Biol Evol.
419 2014;31: 1077–1088.

420 25. Auch AF, Klenk H-P, Göker M. Standard operating procedure for calculating genome-to-
421 genome distances based on high-scoring segment pairs. Stand Genomic Sci. 2010;2: 142–
422 148.

423 26. Stöver BC, Müller KF. TreeGraph 2: combining and visualizing evidence from different
424 phylogenetic analyses. BMC Bioinformatics. 2010;11: 7.

425 27. Margos G, Gatewood AG, Aanensen DM, Hanincová K, Terekhova D, Vollmer SA, et al.

426 MLST of housekeeping genes captures geographic population structure and suggests a
427 European origin of Borrelia burgdorferi. Proc Natl Acad Sci U S A. 2008;105: 8730–8735.

428 28. Jolley KA, Maiden MCJ. BIGSdb: Scalable analysis of bacterial genome variation at the
429 population level. BMC Bioinformatics. 2010;11: 595.

430 29. Pearson WR, Lipman DJ. Improved tools for biological sequence comparison. Proc Natl
431 Acad Sci U S A. 1988;85: 2444–2448.

432 30. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation
433 of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol.
434 2011;7: 539.

435 31. Schwan TG, Burgdorfer W, Garon CF. Changes in infectivity and plasmid profile of the
436 Lyme disease spirochete, Borrelia burgdorferi, as a result of in vitro cultivation. Infect
437 Immun. 1988;56: 1831–1836.

438 32. Grimm D, Elias AF, Tilly K, Rosa PA. Plasmid stability during in vitro propagation of
439 Borrelia burgdorferi assessed at a clonal level. Infect Immun. 2003;71: 3138–3145.

440 33. Biškup UG, Strle F, Ružić-Sabljić E. Loss of plasmids of Borrelia burgdorferi sensu lato
441 during prolonged in vitro cultivation. Plasmid. 2011;66: 1–6.

442 34. Casjens SR, Mongodin EF, Qiu W-G, Luft BJ, Schutzer SE, Gilcrease EB, et al. Genome
443 stability of Lyme disease spirochetes: comparative genomics of Borrelia burgdorferi
444 plasmids. PLoS One. 2012;7: e33280.

445 35. Schutzer SE, Fraser-Liggett CM, Casjens SR, Qiu W-G, Dunn JJ, Mongodin EF, et al.
446 Whole-genome sequences of thirteen isolates of Borrelia burgdorferi. J Bacteriol. 2011;193:
447 1018–1020.

448 36. Haven J, Vargas LC, Mongodin EF, Xue V, Hernandez Y, Pagan P, et al. Pervasive
449 recombination and sympatric genome diversification driven by frequency-dependent
450 selection in Borrelia burgdorferi, the Lyme disease bacterium. Genetics. 2011;189: 951–966.

451 37. Sherwood R. Casjens, Wai Mun Huang, Eddie B. Gilcrease, Weigang Qiu, William D.
452 Mccaig, Benjamin J. Luft, Steven E. Schutzer, Claire M. Fraser. Comparative Genomics of
453 Borrelia burgdorferi. In: Felipe C. Cabello, Dagmar Hulinska, Henry P. Godfrey, editor.
454 Molecular Biology of Spirochetes. 2006. pp. 79–95.

455 38. Shi Y, Xu Q, McShan K, Liang FT. Both decorin-binding proteins A and B are critical for
456 the overall virulence of Borrelia burgdorferi. Infect Immun. 2008;76: 1239–1246.

457 39. Roberts WC, Mullikin BA, Lathigra R, Hanson MS. Molecular analysis of sequence

458 heterogeneity among genes encoding decorin binding proteins A and B of Borrelia
459 burgdorferi sensu lato. Infect Immun. 1998;66: 5275–5285.

460 40. Heikkilä T, Seppälä I, Saxen H, Panelius J, Peltomaa M, Huppertz H-I, et al. Cloning of the
461 gene encoding the decorin-binding protein B (DbpB) in Borrelia burgdorferi sensu lato and
462 characterisation of the antibody responses to DbpB in Lyme borreliosis. J Med Microbiol.
463 2002;51: 641–648.

464 41. Heikkilä T, Seppälä I, Saxen H, Panelius J, Yrjänäinen H, Lahdenne P. Species-specific
465 serodiagnosis of Lyme arthritis and neuroborreliosis due to Borrelia burgdorferi sensu
466 stricto, B. afzelii, and B. garinii by using decorin binding protein A. J Clin Microbiol.
467 2002;40: 453–460.

468 42. Wywial E, Haven J, Casjens SR, Hernandez YA, Singh S, Mongodin EF, et al. Fast,
469 adaptive evolution at a bacterial host-resistance locus: the PFam54 gene array in Borrelia
470 burgdorferi. Gene. 2009;445: 26–37.

471 43. Qiu W-G, Schutzer SE, Bruno JF, Attie O, Xu Y, Dunn JJ, et al. Genetic exchange and
472 plasmid transfers in Borrelia burgdorferi sensu stricto revealed by three-way genome
473 comparisons and multilocus sequence typing. Proc Natl Acad Sci U S A. 2004;101: 14150–
474 14155.

475 44. Byram R, Stewart PE, Rosa P. The essential nature of the ubiquitous 26-kilobase circular
476 replicon of Borrelia burgdorferi. J Bacteriol. 2004;186: 3561–3569.

477 45. Barbour AG, Travinsky B. Evolution and distribution of the ospC Gene, a transferable
478 serotype determinant of Borrelia burgdorferi. MBio. 2010;1. doi:10.1128/mBio.00153-10

479 46. Attie O, Bruno JF, Xu Y, Qiu D, Luft BJ, Qiu W-G. Co-evolution of the outer surface
480 protein C gene (ospC) and intraspecific lineages of Borrelia burgdorferi sensu stricto in the
481 northeastern United States. Infect Genet Evol. 2007;7: 1–12.

482 47. Margos G, Hepner S, Mang C, Marosevic D, Reynolds SE, Krebs S, et al. Lost in plasmids:
483 next generation sequencing and the complex genome of the tick-borne pathogen Borrelia
484 burgdorferi. BMC Genomics. 2017;18: 422.

485 48. Casjens SR, Mongodin EF, Qiu W-G, Dunn JJ, Luft BJ, Fraser-Liggett CM, et al. Whole-
486 genome sequences of two Borrelia afzelii and two Borrelia garinii Lyme disease agent
487 isolates. J Bacteriol. 2011;193: 6995–6996.

488 49. Schüler W, Bunikis I, Weber-Lehman J, Comstedt P, Kutschan-Bunikis S, Stanek G, et al.
489 Complete genome sequence of Borrelia afzelii K78 and comparative genome analysis. PLoS
490 One. 2015;10: e0120548.

491 50. Schutzer SE, Fraser-Liggett CM, Qiu W-G, Kraiczy P, Mongodin EF, Dunn JJ, et al.
492 Whole-genome sequences of Borrelia bissettii, Borrelia valaisiana, and Borrelia spielmanii.
493 J Bacteriol. 2012;194: 545–546.

494 51. Norris SJ. vls Antigenic Variation Systems of Lyme Disease Borrelia: Eluding Host
495 Immunity through both Random, Segmental Gene Conversion and Framework
496 Heterogeneity. Microbiol Spectr. 2014;2. doi:10.1128/microbiolspec.MDNA3-0038-2014

497

Supporting Information

S1 Fig. Coverage of MM1 Sequence Reads Across *B. burgdorferi* B31 Plasmids. An initial mapping of MM1 reads to the type strain B31 suggested lack of plasmids lp21, lp28-1, lp28-2, lp38, lp56 and lp5 in MM1 genome, while the presence/absence status of other plasmids in MM1 remained ambiguous. B31 plasmids are shown in each plot with X-axis and Y-axis representing Reference Start Position and Coverage respectively.

S2 Fig. Nanopore sequencing reads confirm the novel SNPs in *nifS* and *pepX*. Top: The de novo assembly of the *nifS* gene matches the MLST *nifS* type 12 except for a C->T polymorphism. Eleven nanopore reads align to this region and all contain the T variant. Bottom: The de novo assembly of the *pepX* gene matches the MLST *pepX* type 1 except for a T->C polymorphism. Nine nanopore reads align to this region and eight contain the C variant. The two SNPs of interest are noted with a red asterisk (*).

S1 Table. MM1 gene members of PFam32, PFam49, PFam50 and PFam57/62 identified by spectral clustering of protein sequences. The table lists the results of SCPS analysis for the 4 PFams 32, 49, 50 and 57/62 (see Materials and methods) in MM1, N40, JD1 and B31 strains. Genes located on replicons of the same type are arranged in nearby rows within each paralogous gene family. Dashes mark missing entries, which may fall below cut-off or may be missing in the respective strain.

Figure 1

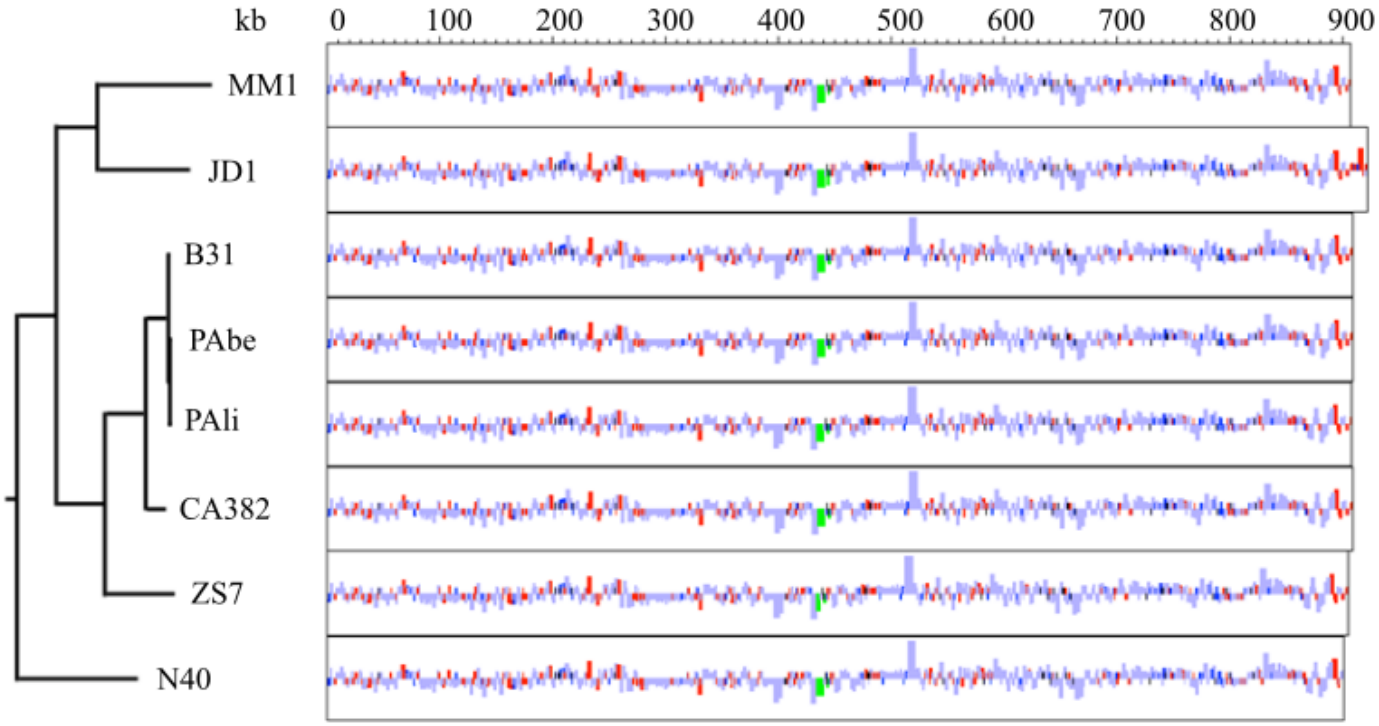

kb
0
100
200
300
400
500
600
700
800
900
MM1
JD1
B31
PAbe
PAli
CA382
ZS7
N40

Figure 2

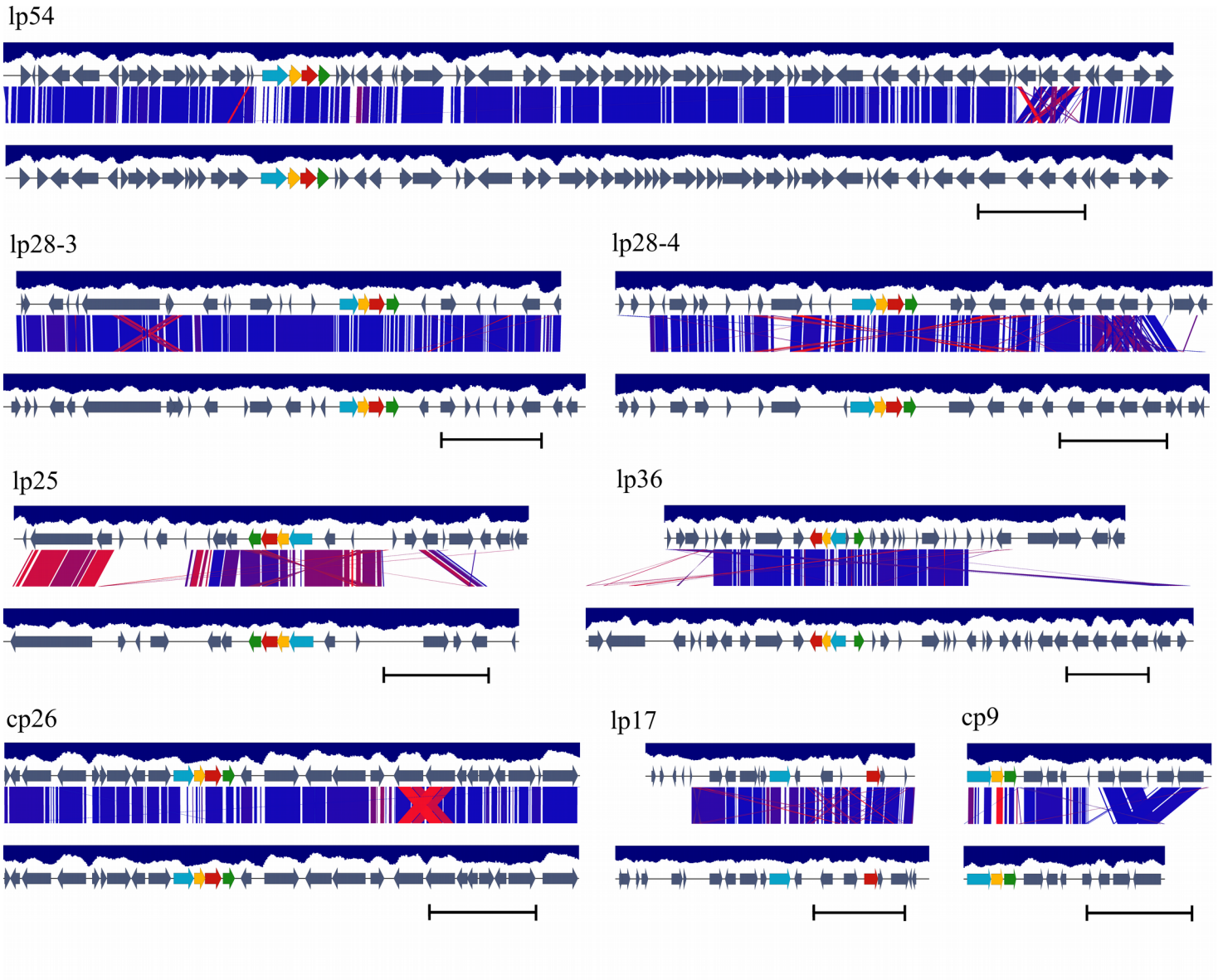


Figure 3

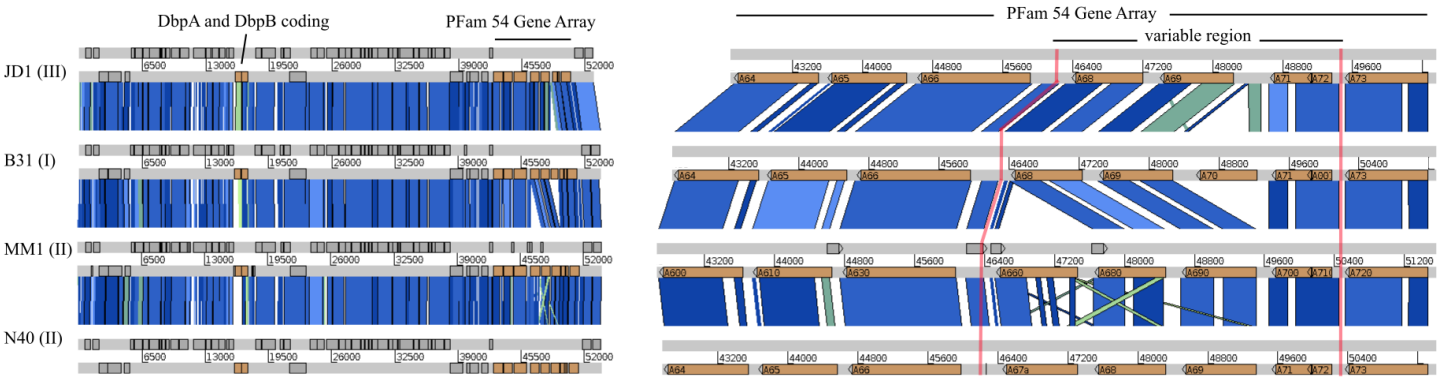

DbpA and DbpB coding
PFam 54 Gene Array
JD1 (III)
B31 (I)
MM1 (II)
N40 (II)
PFam 54 Gene Array
variable region

Figure 4

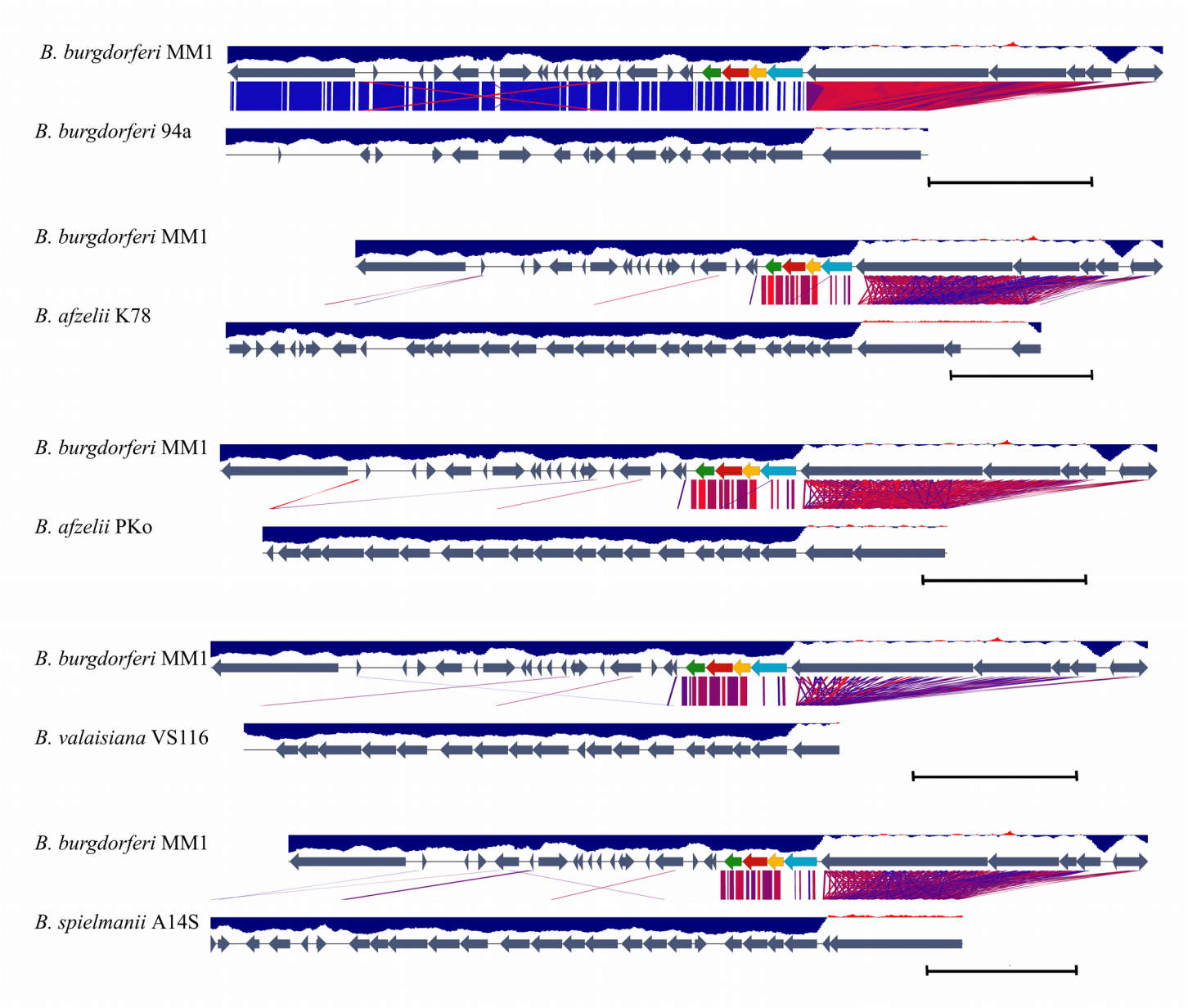
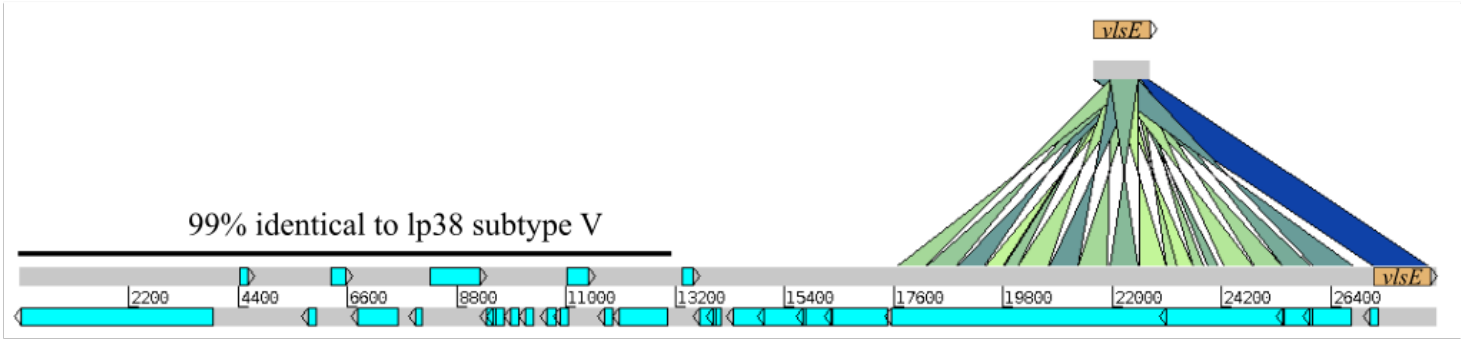

B. burgdorferi MM1
B. burgdorferi 94a
B. burgdorferi MM1
B. afzelii K78
B. burgdorferi MM1
B. afzelii PKo
B. burgdorferi MM1
B. valaisiana VS116
B. burgdorferi MM1
B. spielmanii A14S

Figure 5


vlsE
99% identical to lp38 subtype V
vlsE
2200
4400
6600
8800
11000
13200
15400
17600
19800
22000
24200
26400