Multipartite Genome of Lyme Disease *Borrelia*: Structure, Variation and Prophages

Ira Schwartz[1]*, Gabriele Margos[2], Sherwood R. Casjens[3], Wei-Gang Qiu[4] and Christian H. Eggers[5]

[1]Department of Microbiology and Immunology, New York Medical College, Valhalla, NY USA, [2]National Reference Centre for *Borrelia* and Bavarian Health and Food Safety Authority, Oberschleissheim, Germany, [3]Division of Microbiology and Immunology, Pathology Department, University of Utah School of Medicine, Salt Lake City, UT USA,
[4]Department of Biological Sciences, Hunter College of the City University of New York, New York, NY USA
[5]Department of Biomedical Sciences, Quinnipiac University, Hamden, CT, USA
*Corresponding author: ira_schwartz@nymc.edu

DOI: https://doi.org/10.21775/cimb.042.409

Abstract
All members of the *Borrelia* genus that have been examined harbour a linear chromosome that is about 900 kbp in length, as well as a plethora of both linear and circular plasmids in the 5-220 kbp size range. Genome sequences for 27 Lyme disease *Borrelia* isolates have been determined since the elucidation of the *B. burgdorferi* B31 genome sequence in 1997. The chromosomes, which carry the vast majority of the housekeeping genes, appear to be very constant in gene content and organization across all Lyme disease *Borrelia* species. The content of the plasmids, which carry most of the genes that encode the differentially expressed surface proteins that interact with the spirochete's arthropod and vertebrate hosts, is much more variable. Lyme disease *Borrelia* isolates carry between 7-21 different plasmids, ranging in size from 5-84 kbp. All strains analyzed to date harbor three plasmids, cp26, lp54 and lp17. The plasmids are unusual, as compared to most bacterial plasmids, in that they contain many paralogous sequences, a large number of pseudogenes, and, in some cases, essential genes. In addition, a number of the plasmids have features indicating that they are prophages. Numerous methods have been developed for Lyme disease *Borrelia* strain typing. These have proven valuable for clinical and epidemiological studies, as well as phylogenomic and population genetic analyses. Increasingly, these approaches have been displaced by whole genome sequencing techniques. Some correlations between genome content and pathogenicity have been deduced, and comparative whole genome analyses promise future progress in this arena.

Introduction
The genus *Borrelia* forms a deeply separated lineage within the *Spirochaetes* branch of the bacterial tree of life (Paster et al., 1991). The organisms are genomically unique and not closely related to any other bacteria, including the other *Spirochaetes*. *Borrelia burgdorferi* and *Borrelia hermsii* are representative of the two major branches within the *Borrelia* genus, wherein *B. burgdorferi* typifies the Lyme disease *Borrelia* branch (abbreviated herein as LB) and *B. hermsii* typifies the relapsing fever agent branch (Paster et al., 1991; Ras et al., 1996; Schwan et al., 2007). The term '*B. burgdorferi*' denotes *B. burgdorferi sensu stricto* throughout this review. By analysing the genomes of both *B. burgdorferi* and *B. hermsii,* Barbour and coworkers originally showed that both have a topologically unusual genome with linear chromosomes and multiple linear plasmids (Plasterk et al., 1985; Barbour, 1988; Ferdows and Barbour, 1989; Hinnebusch et al., 1990; Kitten and Barbour, 1990; Hinnebusch and Barbour, 1991) (Figure 1).

Subsequent work has shown this to be a property of all members of the genus *Borrelia* that have been examined. This review will focus on the Lyme Disease *Borrelia* (abbreviated throughout as LB) branch of the genus; the genomics of the relapsing

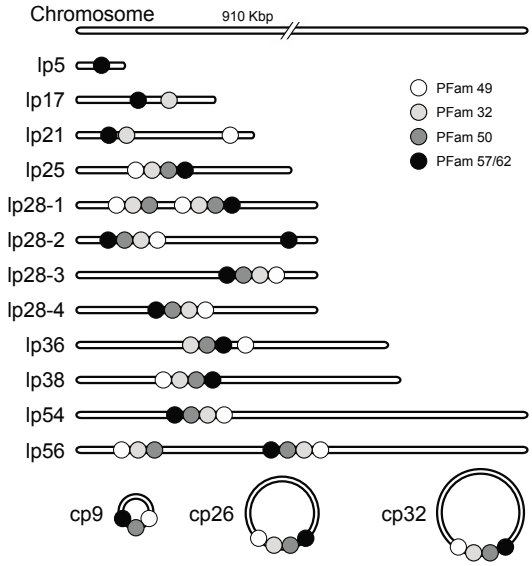

Figure 1. The genome of *B. burgdorferi* type strain B31. The linear and circular plasmids are shown with the genes thought to be involved in plasmid replication, partitioning and compatibility (see text) indicated by small circles with different shading. Different B31 cultures have been found to carry one or two related cp9 plasmids and up to eight different cp32 plasmids. The figure is modified from figure 8 of Casjens *et al.* (2000).

fever branch is covered elsewhere (Radolf and Samuels, 2021).

The *B. burgdorferi* genome was among the first bacterial chromosomes to be sequenced (Fraser et al., 1997). Since then 27 complete LB genomes (including chromosome and plasmids) have been elucidated and numerous other less complete genome sequences have been deposited in the NCBI Genomes database (as of 5/12/20). Each LB species carries a linear chromosome about 900 kbp in length and multiple circular and linear low copy number plasmids that are usually, but not always, in the 5-84 kbp range. These features have been demonstrated directly for many of the defined LB species, including *Borrelia afzelii*, *Borrelia andersonii*, *Borrelia bavariensis, Borrelia bissettiae*, *B. burgdorferi, Borrelia filandensis*, *Borrelia garinii*, *Borrelia japonica, Borrelia lusitaniae*, *Borrelia maritima, Borrelia mayonii, Borrelia spielmanii*, *Borrelia turdi* and *Borrelia valaisiana* (Ferdows and Barbour, 1989; Davidson et al., 1992; Casjens and Huang, 1993; Ojaimi et al., 1994; Casjens et al., 1995; Vitorino et al., 2010; Mongodin et al., 2013; Kingry et al., 2016; Casjens et al., 2018; Margos et al., 2019a; Margos et al., 2019b; Margos et al., 2020) and all LB species likely will have genomes of this type. The chromosomes of these bacteria all have quite similar gene contents, whereas the plasmids are more variable (details below).

In light of their importance for disease pathogenesis and maintenance of the enzootic cycle (see below and Radolf and Samuels, 2021), the number and diversity of the plasmids carried by LB species have been the focus of much research. This enterprise, however, has been hampered by their unusual features. First, individual LB cells harbor a more diverse plasmid complement than that of any other bacterium - the sequenced genome of *B. burgdorferi* type strain B31 has 21 plasmids, and several additional plasmids appear to have been lost between isolation of the strain in 1982 and the completion of the complete genome sequence (Casjens et al., 2000). Second, a number of the plasmids have very similar sizes, and thus cannot be separated by pulsed-field gel electrophoresis. Consequently, mapping a gene to a particular plasmid by Southern hybridization can be difficult. Third, different plasmids in the same cell often carry paralogous sequences that can be extremely similar (Casjens, 2000; Casjens et al., 2000). This can and does confuse both plasmid identification and assembly of plasmid sequences. Thus, the complete plasmid content of a particular LB strain is difficult to ascertain without complete genomic sequence analysis. Fortunately, advances in sequencing technologies has made such analysis simpler and is leading to a much better understanding of LB species plasmid content (Casjens et al., 2012; Mongodin et al., 2013; Margos et al., 2017b; Casjens et al., 2018). Finally, assessing the true plasmid content of an LB cell is further complicated by the fact that plasmids can be lost during primary isolation; knowing whether native plasmids have been lost from any strain isolated in the laboratory is essentially impossible at this time. Furthermore, many of the *Borrelia* plasmids are readily lost during passage or manipulation *in vitro*, and many clones of any given strain will have lost one or more (Schwan et al., 1988; Xu et al., 1996; Purser and Norris, 2000; Elias et al., 2002; Grimm et al., 2003).

Overall, the 900-kbp chromosomes (often called the “large” or “main” chromosome by workers in this field)

carry the great majority of the genes that encode metabolic enzymes, and the 400 to 650 kbp (in different isolates) of plasmid DNAs carry the bulk of the surface lipoprotein encoding genes. With the notable exception of cp26 (see below), the plasmids are not required for growth in culture (Sadziene et al., 1993; Casjens et al., 1997b), but have often been found to be required for mouse infectivity or tick transmission in the laboratory (Schwan et al., 1988; Xu et al., 1996; Purser and Norris, 2000; Purser et al., 2003; Grimm et al., 2005; Revel et al., 2005; Stewart et al., 2005; Strother et al., 2005; Strother and de Silva, 2005; Lin et al., 2012; Radolf et al., 2012). The chromosome carries tightly packed genes, as is typical of bacteria, while many of the linear plasmids have substantially lower gene densities and many apparently decaying pseudogenes (Casjens, 2000; Casjens et al., 2000). Currently, the general impression is that of an evolutionarily quite stable chromosome that encodes the machinery required for existence as a bacterial cell along with a rather large menu of much more evolutionarily variable plasmids that encode most of the proteins that interact with the vertebrate and arthropod host environments encountered by the bacteria (there are, however, a small number of important metabolic enzymes encoded by the plasmids and surface-exposed proteins encoded by the chromosome; see below).

We note that references cited in this review were chosen to allow the reader access to the latest literature and not necessarily to credit discoverers, the originators of ideas, or the first to obtain particular kinds of data. Thus, cited articles are sometimes more recent articles instead of those in which the original discoveries were first reported.

Lyme Disease *Borrelia* Chromosomes

Borrelia burgdorferi B31 Chromosome

The complete 910,725-bp sequence of the isolate B31 (the type strain of *B. burgdorferi*) linear chromosome was published in 1997 (Fraser et al., 1997). Its nucleotide composition is 28.6% G+C,and it is predicted to contain 803 protein-encoding genes and 17 pseudogenes. The chromosomal genes have been shown to be under intense purifying selection (Tyler et al., 2018). The fraction of mutationally inactivated chromosomal genes is among the lowest of the analyzed bacterial genomes. Although smaller genomes of free-living bacteria are known, this chromosome size is near the small end of the spectrum (*e.g.* Casjens, 1998)). The protein-encoding genes occupy 93% of the chromosome, a typical value for a bacterial genome not undergoing current reduction in size (Lynch, 2006). About 67% of the genes are oriented such that they are transcribed away from the center of the chromosome. GC skew analysis and experimental evidence strongly suggest the presence of a replication origin near the center of the linear chromosome (Picardeau et al., 1999). The chromosome of B31 carries five rRNA genes (two 23S, two 5S and one 16S) and 32 tRNA genes. The former are clustered very near the center of the chromosome. The tRNAs, predicted to encode individuals specific for all 20 amino acids, are scattered across the chromosome in seven clusters and 13 single genes (Schwartz et al., 1992; Fraser et al., 1997). Variable numbers of tandem repeats were found in three genes: BB_0210, BB_0546 and BB_0801 (Mongodin et al., 2013). BB_0210 encodes a surface protein, Lmp1, which has been observed to hamper chromosome sequence assembly from short-read sequences such as Illumina likely due to difficulties in assembling the tandem repeats (Tyler et al., 2018).

Since the complete genome of *B. burgdorferi* strain B31 was sequenced (Fraser et al., 1997; Casjens et al., 2000), the *Borrelia* research community has usually used the 'locus tags' in its GenBank annotation as names for its genes and their encoded proteins (and their homologues discovered in other isolates). Thus, according to bacterial convention, chromosomal genes are named '*bb0xxx*' (lower case and italicized) in ascending order from *bb0001* through *bb0853* across the chromosome, and chromosomally encoded proteins are named 'BB_0xxx' (upper case and not italicized, *e.g.* BB_0364). The B31 plasmid locus tag names are similar but have the form 'BB$xx' in which '$' indicates a letter, A through U, denoting the plasmid which carries the gene (*e.g. bba74* encodes protein BBA74 and lies on lp54; *bbs09* lies on cp32-3, etc.; see LB plasmid section below).

Depending on the annotation pipeline, 60-85% of the chromosome's predicted genes have some similarity to a gene in another organism whose role or function is at least partly understood; about 10% are similar to known genes in other organisms whose roles are unknown; the remainder are unique to *Borrelia* and have unknown functions (Fraser et al., 1997; Margos et al., 2017c). The chromosome carries what appears to be a rather minimal set of genes required for cell maintenance and replication. These include genes for cell wall biosynthesis (but not synthesis of lipopolysaccharide); protein export and lipidation; DNA,

RNA and protein biosynthesis; DNA repair; nucleotide metabolism (but not *de novo* synthesis); membrane lipid and phospholipid biosynthesis; glycolysis and a few enzymes that provide substrates for the glycolytic pathway, the bacterium's sole mode for generating ATP. A large and complete set of the genes known to be required for motility and chemotaxis also is present, as are a number of genes for transport of small molecules across the cytoplasmic membrane. A major conclusion from this bioinformatic analysis is that the biosynthetic and intermediary metabolic capacity of *Borrelia* is very limited. Genes that encode enzymes that perform functions in respiration, amino acid synthesis, nucleotide synthesis, lipid synthesis and enzyme cofactor synthesis are almost completely lacking, consistent with the many and fastidious requirements for growth in culture, as well as its restricted enzootic, host-associated lifestyle. Although originally no requirement for iron or iron transporter genes were predicted (Posey and Gherardini, 2000), in a subsequent paper a ferritin-like Dps (DNA-binding protein from starved bacteria) protein with iron and copper binding properties was described (Wang et al., 2012). In most cases that have been examined manganese- or zinc-dependent enzymes are used instead of iron-dependent enzymes (Nguyen et al., 2007; Ouyang et al., 2009; Groshong and Blevins, 2014). Given this limited biosynthetic capacity, the spirochete must acquire nearly all carbohydrates, amino acids, nucleosides, lipids and cofactors from its environment. Numerous transporter and permease components are encoded on the chromosome (with several others encoded on plasmids); these include those for transport/uptake of carbohydrates (14 genes) and amino acids/peptides (15 genes) (Fraser et al., 1997; Corona and Schwartz, 2015; Groshong et al., 2017). (See also Radolf and Samuels, 2021, for additional discussion of metabolism.)

The initial annotation of the B31 chromosome suggested a surprisingly small number of genes identified as encoding transcriptional regulators. This was not expected as numerous *B. burgdorferi* genes are known to be regulated by environmental signals (Samuels, 2011; Radolf et al., 2012). Only two alternative sigma factors (RpoS and RpoN) and two two-component systems (Hk1-Rrp1 and Hk2-Rrp2) initially were identified by sequence homology. Since then several additional transcriptional regulators have been identified and characterized, including BosR, BadR, CsrA, Rel_{Bbu} and DksA (Samuels, 2011; Radolf et al., 2012; Groshong and Blevins, 2014; Arnold et al., 2018) (see Radolf and Samuels, 2021, for more details).

Chromosomes of Other Lyme Disease Borrelia species

The seminal publication of the genome of the *B. burgdorferi* type strain B31 (Fraser et al., 1997) revealed its complexity. It took several years before the full closure of all plasmids was achieved and the diversity of the plasmids, even within one species, was fully appreciated (Casjens et al., 2000; Casjens et al., 2012; Casjens et al., 2017; Casjens et al., 2018). The stability and co-linearity of chromosome and core plasmids (lp54, cp26) between *Borrelia* species was initially demonstrated by Glöckner et al. (2004; 2006). These authors sequenced the genomes of *B. bavariensis* PBi (at the time still designated *B. garinii*) and *B. afzelii* PKo and showed that the chromosomes of these two species were co-linear with that of *B. burgdorferi* B31. The chromosomes of PBi and PKo had only three insertions and six and nine deletions of >100 bp, respectively (Glockner et al., 2006). Two of these insertions were >1000 bp; one resided within the coding sequence of the basic membrane protein A and the other was near the origin or replication. A large indel of >600 bp in the genome of PBi was found in the coding region of LMP1 (BB_0210) which contains variable number of tandem repeats (Mongodin et al., 2013; Tyler et al., 2018).

Subsequently determined genome sequences for many of the 22 Lyme disease *Borrelia* species are now available (Table 1), either as completed genomes, draft genomes or in the form of raw reads in the NCBI sequence read archive (SRA) (https://www.ncbi.nlm.nih.gov/sra) (Casjens et al., 2011a; Casjens et al., 2011b; Schutzer et al., 2011; Schutzer et al., 2012; Mongodin et al., 2013; Ivanova et al., 2014; Jacquot et al., 2014; Gatzmann et al., 2015; Becker et al., 2016; Castillo-Ramirez et al., 2016; Kingry et al., 2016; Margos et al., 2017a; Margos et al., 2017b; Margos et al., 2017c; Margos et al., 2018; Margos et al., 2019a; Margos et al., 2020). The genome structure, i.e. its composition of a linear main chromosome and a number of circular and linear plasmids, is conserved in all species investigated (Schutzer et al., 2012; Mongodin et al., 2013; Ivanova et al., 2014; Becker et al., 2016; Kingry et al., 2016; Margos et al., 2019a; Margos et al., 2020). The chromosome of the different LB species varies in size, ranging from 900 kbp (900,694 in *B. chilensis*) to 922 kbp (922,901 in *B. burgdorferi* JD1) (Table 1).

Right end extensions (plasmid fusion) of chromosomes have been described in *B. burgdorferi* , *B. mayonii* and *B. valaisiana* (Casjens et al., 1997a; Huang et al., 2004; Banik et al., 2011; Casjens et al., 2012; Casjens et al., 2018) and has most recently has been found in the *B. turdi* isolate TPT2017, in which an lp28-2-like plasmid was found fused to the right end of its chromosome (Margos et al., 2019a).

The rRNA gene profiles of LB spirochetes were described as consisting of one 16S rRNA locus that is located approximately 2 kb upstream of tandemly duplicated 5S-23S rRNA loci separated by an intergenic spacer of approximately 200-350 bp (Schwartz et al., 1992; Gazumyan et al., 1994); this arrangement appeared to be similar in other LB species (Ojaimi et al., 1994). However, differences in the rRNA gene profiles have been identified in *B. bavariensis*, *B. afzelii*, *B. spielmanii, B. turdi* and *B. yangtzensis* which contain two 16S rRNA genes. Other species, such as *B. americana* (3), *B. carolinensis* (4), *B. japonica* (4) and *B. kurtenbachii* (4) appear to have three or four 5S-23S rRNA loci (Table 1).

In two isolates of *B. andersonii,* 19857 and 21038, and the *B. japonica* isolate lKA2, a single 5S rRNA locus and insertions in the second copy of the 23S rRNA were reported (Marconi et al., 1995). Curiously, two genome assemblies of *B. bavariensis,* isolates NMJW1 and SZ, (erroneously named *B. garinii* in GenBank), seem to possess only one 5S and one 23S rRNA locus but two 16S rRNA loci. However, whether these deviations in the arrangement of the rRNA loci are correct or may in some cases be due to mis-assembly of sequence data needs to be confirmed. In *B. chilensis,* 31 tRNAs and 1 non-coding (nc) RNA are present (Ivanova et al., 2014), whereas the number of tRNAs and ncRNAs in the genomes of the majority of LB species with completed chromosome sequences in GenBank were 32 and 3, respectively (Table 1).

In general, the synteny of the main chromosome appears to be well conserved in most investigated LB species. The length of the constant chromosomal regions was reported to be approximately 903 kbp with only a few indels or differences between the chromosomes of LB species (although in some it may be shorter (e.g. in *B. bissettiae*, *B. chilensis* or *B. maritima* chromosomes are only 900 - 902 kbp*;* Table 1) (Mongodin et al., 2013; Ivanova et al., 2014; Margos et al., 2020). Interspecies differences in the three repeat-containing genes (*bb0210*, *bb0546* and *bb0801*) were found. In addition, indels larger than 25 bp differentiate the chromosomes of *B. burgdorferi* and *B. afzelii* (Glockner et al., 2006; Mongodin et al., 2013). Furthermore, compared to strain B31, *B. bavariensis* isolate PBi and *B. afzelii* PKo had duplications in the *bmp* gene region, five indels of <330 bp in intergenic regions and indels of <150 bp in genes *bb0309*, *bb0704* and *bb0749* (Glockner et al., 2006; Mongodin et al., 2013). Additional comparative genome analyses of *B. maritima* and its closest relative *B. bissettiae* isolate DN127 showed that *B. maritima* lacked five chromosomal protein coding sequences, including three hypothetical proteins, the orthologue of *bb0049* and the gene encoding competence protein ComEC/Rec2, but included an additional three genes in its repertoire --- a gene encoding an apolipoprotein N-acetyl-transferase, a hypothetical protein and the arginine/ornithine antiporter ArcD (Margos et al., 2020).

Although different statistical methods were used for calculating genetic distances within and between species, it is apparent that the genetic divergence is much smaller within species than between species, confirming previous data using MLST/MLSA (Margos et al., 2009). Within *B. burgdorferi* only minor differences between chromosomes were noted, e.g. ZS7 and Bol26 differed by 0.084% and 94b and 29805 by 0.625% (Mongodin et al., 2013), whereas distances calculated between species ranged from 2.8% - 8 % (Mongodin et al., 2013; Jacquot et al., 2014). Consistent with this, the most recently described LB species, *B. mayonii* and *B. maritima* shared average nucleotide identities of 93.8% and 90.6%, respectively, with other LB species. Thus, phylogenetic analyses confirm the genetic distinctness of the different LB species, which is likely related to their ecology (see below).

Lyme Disease *Borrelia* plasmids

Members of the *Borrelia* genus carry more plasmids than any other bacteria. LB isolates that have been analyzed carry between 7 and 21 different plasmids (Figure 1). These plasmids are both linear and circular and range from 5 to 84 kbp in size (Casjens et al., 2000; Casjens et al., 2012; Kingry et al., 2016; Casjens et al., 2017; Margos et al., 2017b; Casjens et al., 2018; Jabbari et al., 2018; Margos et al., 2020). Most of the plasmids, except for cp26 (Byram et al., 2004), are not required for growth in culture (Sadziene et al., 1992; Casjens et al., 1997b). Table 2 lists the plasmid complement for the 32 LB isolates

Table 1. Chromosomal genetic features of the Lyme Disease *Borrelia* species complex

Species	Isolate name	No. of tRNA/ncRNA	No. of rRNAs rrs/rrf-rrl	Average G+C % *	Chromosome size°	No. of sequenced genomes (source)
B. afzelii	Pko		2/2	27.9	903690	11 (GB#)
	HLJ01	32/3	2/2	28.3	905471	12 (SRA##)
	K78	32/3	2/2	27.9	905949	
	Tom3107	32/3	2/2	28.1	905861	
	BO23	32/3	2/2	27.8	905349	
B. americana	SCW-41	nd	1/ 3	nd	nd	1 (SRA)
B. bavariensis	PBi	32/3	2/2	27.8	905817	1 (GB)
	BgVir	32/3	2/2	28.2	905534	25 (SRA)
	NMJW1	32/3	2/1	28.4	902789	
	SZ	32/3	2/1	28.4	902487	
B. bissettiae	DN-127	32/3	1/ 2	28.3	900755	2 (GB) 1 (SRA)
B. burgdorferi sensu stricto	B31	31/1?	1/ 2	28.1	910724	111 (GB, many draft)
	JD1	32/2	1/ 2	28.3	922801	16 (SRA)
	CA382	32/3	1/ 2	28.6	910736	
	PAbe	32/3	1/ 2	28.5	910728	
	PAli	32/3	1/ 2	28.3	909921	
	MM1	32/3	1/ 2	28.2	908512	
	ZS7	32/3	1/ 1	28.2	906701	
	B331	32/3	1/ 2	28.2	904573	
	N40	32/3	1/ 2	28.2	902191	
B. californiensis	CA446	32/?	1/ 2	28.1	nd	1 (SRA)
B. carolinensis	CA446	nd	1/ 4	nd	nd	1 (SRA)
B. chilensis	VA1	31/1	1/ 2	28.5	900694	1 (GB)
B. garinii	20047	32/3	1/ 2	28.1	906449	42 (GB, mostly draft)
	Far04	29/3	1/ 2 (1 partial 23S)	nd	nd	26 (SRA)
B. japonica	HO14	33/?	1/ 4	28.1	nd	1 (SRA)
B. kurtenbachii	25015	33/?	1/ 4	28.4	nd	1 (SRA)
B. lanei	CA28-91	Nd	Nd	Nd	Nd	
B. lusitaniae	PoTiB2		1/ 2			1 (SRA)
B. maritima	CA690	33/1	1/ 2	27	902176	1 (GB)
B. mayonii	M14-1420	32/3	1/ 2	27.8	904387	2 (GB)
	M14-1539	32/3	1/ 2	27.8	904387	
B. spielmanii	PMew	Nd	2/2	27.6	nd	1 (GB) 3 (SRA)
B. turdi	TPT2017	33/?	2/2	27.7	935973	3 (GB, draft)
	T1990A		2/2	27.7	906931	
	T2084	33?/	2/2	27.6	901717	
B. valaisiana	VS116	32/2	1/ 2	27.4	913294	2 (GB)
	Tom4006	32/3	1/ 2	28.0	912160	2 (SRA)
B. yangtzensis	Okinawa CW62	33/?	2/ 1	27.8	nd	1 (SRA)

*may depend on whether and which plasmids are included
°Telomere regions may not be included in all samples
#GB=GenBank
##SRA=Sequence Read Archive

for which complete or nearly complete genome sequences have been reported to date.

Plasmid types and nomenclature

B. burgdorferi plasmids were originally named in the type strain B31 with “lp” for linear plasmid and “cp” for circular plasmid and numerical designations of their approximate kbp sizes as determined by agarose gel electrophoresis and sequencing that did not always reach the linear plasmid telomeres (Fraser et al., 1997; Casjens et al., 2000). Naming the plasmids in all strains according to their size,

however, creates several difficulties: (i) most of the plasmids are in the 25 to 31 kbp size range, so the number of different names based only on size is limited, (ii) plasmids from different isolates with closely related segments can, nonetheless, harbor major differences in size and organization, and (iii) names based purely on size have no biological significance. Naming plasmids according to more biologically meaningful criteria would be more informative. It was noticed early on that all of the B31 plasmids larger than 15 kbp carry a set of four genes, usually in a contiguous cluster, that are predicted to be involved in plasmid replication and maintenance. These genes encode proteins that are members of the LB protein families (PFams) named PFam32, PFam49, PFam50 and PFam57/62 (originally defined as *paralogous* families within strain B31 in Casjens *et al.* (2000) and extended to other strains in Casjens *et al.* (2012) (Figure 1). Smaller plasmids lp5, lp21 and cp9 often lack one or more of these four gene types but always carry a PFam57/62 gene. The PFam32 proteins are homologues of the well-studied ParA proteins that play a central role in the partitioning of many bacterial plasmids (Shih and Rothfield, 2006); the other three families have no strong similarity to genes of known function. Experimental studies have strongly supported the notion that the PFam49, PFam50 and PFam57/62 protein types are required for plasmid maintenance (Stewart et al., 2001; Eggers et al., 2002). The PFam32 proteins of circular cp26 or linear lp17 are apparently not required for maintenance (Beaurepaire and Chaconas, 2005; Tilly et al., 2012); however, the mutational analysis of Lin et al. (2012) suggests that they are required for other plasmids. PFam32 proteins are quite diverse and can be robustly parsed into different sequence types; strikingly, no LB cell has been found that carries two plasmids that encode PFam32 proteins of the same sequence type. This PFam32-type mutual exclusivity currently holds true for the 669 sequenced plasmids in 50 completely or nearly completely sequenced isolates (423 plasmids in 29 isolates with completely sequenced and published genomes (Table 2) and 21 additional unpublished LB genomes that carry 246 plasmids; Casjens, unpublished results). This strongly suggests that two plasmids with the same PFam32 type gene cannot coexist in the same cell and so are incompatible in spite of the fact that this protein is not always essential for plasmid maintenance in the laboratory. Consequently, it has been proposed that, when possible, LB plasmids be named according to the PFam32 protein they encode (reviewed by Casjens *et al.*, 2006; 2010; 2018). Rather than inventing a completely new set of plasmid names (the B31 names had been in use for nearly a decade), each PFam32 type was given its already existing B31 name (*e.g.*, cp26, lp28-2, lp36, *etc.*). As new PFam types were discovered in other isolates, they were given new names of the same style (*e.g.,* lp28-5 discovered in *B. burgdorferi* strain N40; lp28-8 discovered in *B. burgdorferi* strain 94a). The various B31 lp28 plasmids as well as lp21, lp25, lp36, lp38 and lp56 plasmids are all very variable in size between strains, and all of these plasmids contain many overlapping paralogous gene types.

Thus, when a newly discovered plasmid type contains members of these same gene sets it is given the next number in the lp28 series which is currently up to lp28-13 (see below). Linear plasmids that carry a cp32 type PFam32 gene are given a name that corresponds to its cp32 type (*e.g,* a linear plasmid with a cp32-6 type PFam32 gene would be lp32-6) (Casjens et al., 2017; Casjens et al., 2018). Table 3 lists all the published plasmids and their size ranges.

This PFam32 type nomenclature scheme has the major advantage that plasmids of the same compatibility type have related names, but it also has disadvantages: (i) the PFam32 gene on a given plasmid must be sequenced in order to determine its final designation, and (ii) plasmids of the same PFam32 type often differ substantially in gene content (see below). PFam32 sequence types are not limited to single species as most types are found in multiple LB species (Casjens et al., 2018). Of course, any biological nomenclature system will have some complications, and LB isolates are no exception. First, several cases exist in which individual plasmids carry more than one apparently intact PFam32 gene (though adhering to the “only one copy of any PFam32 type gene per cell” rule). There are examples of circular plasmids integrated into linear plasmids (cp32-10 integrated into lp56 in *B. burgdorferi* strain B31 and cp32-11 into lp54 in *B. finlandensis* strain SV1), apparent end-to-end fusion of linear plasmids lp36 and lp28-4 in *B. burgdorferi* strain CA-11.2A, and fusion of two or more circular plasmids (*e.g.*, cp32-1 and cp32-5 in *B. burgdorferi* strain JD1; fusion of parts of four cp32s in cp32-quad in *B. bissettiae* strain DN127) (Casjens et al., 2012; Casjens et al., 2018). The names of fused plasmids have not always been consistent, but the recommendation is that they be named with both PFam32 types, for example “cp32-1+5”. A second more serious complication is that it is not known how

Table 2. Total plasmid numbers in 32 Lyme Disease *Borrelia* isolates

Species	Isolate (X[1])	cp9	cp26	cp32[2]	lp5	lp17	linears[3]	lp54	Total[4]	Reference
B. burgdorferi										
	64b (A)	-	1	9[8]	-	1	7	1	18	Casjens, 2017
	B31 (A)	1[5]	1	8[6]	1	1	10	1	21[2,5]	Casjens, 2000
	Bol26 (A)	-	1	6[7,8]	-	1	4	1	13	Casjens, 2017
	PAli (A[1])	-	1	7[6]	-	1	4	1	13	Margos, 2017
	PAbe (A[1])	-	1	7[6]	-	1	4[9]	1	11	Margos, 2017
	ZS7(A)	-	1	6[8]	-	1	6	1	14	Casjens, 2017
	29805 (B)	-	1	5[8]	-	1	7	1	15	Casjens, 2017
	N40 (B)	1	1	6[8]	-	1	6	1	16	Casjens, 2017
	WI91-23 (B)	2	1	7[8]	1	1	9	1	21	Casjens, 2017
	72a (C)	-	1	6[8]	-	1	4	1	13	Casjens, 2017
	94a (C)	1	1	5	-	1	5[7]	1	14	Casjens, 2017
	118a (C)	1	1	8[8]	-	1	8	1	19	Casjens, 2017
	CA11_2a (C)	-	1	5	-	1	4[10]	1	12	Casjens, 2017
	MM1 (C[1])	1	1	6	-	1	5	1	15	Jabbari, 2018
	156a (D)	-	1	8[8]	-	1	9	1	20	Casjens, 2017
	297 (D[11])	-	1	9[8]	-	1	7	1	19	Casjens, 2017
	JD1 (D)	-	1	9[8]	-	1	9	1	20	Casjens, 2017
B. "finlandensis"										
	SV1	-	1	4[8,12]	-	1	4	1	10	Casjens, 2018
B. bissettiae										
	DN127	1	1	9[8]	-	1	5	1	16	Casjens, 2018
B. mayonii										
	MN14-1420	1	1	5[8]	1	1	5	1	15	Kingry, 2016
	MN14-1539	1	1	5[8]	-	1	5	1	14	Kingry, 2016
B. bavariensis										
	PBi	-	1	4[13]	-	1	6[13]	1	11[13]	Glockner, 2006 Margos, 2018
B. garinii										
	20047	-	1	2[13]	-	1	5[13]	1	>6	Margos, 2018-[15]
	Far04	-	1	-	-	1	4	1	7	Casjens, 2018
	PBr	-	1	2[8]	-	1	6	1	11	Casjens, 2018
B. afzelii										
	ACA-1	-	1	4[8]	-	1	7	1	14	Casjens, 2018-[15]
	BO23	1	1	?[13]	-	1	5[13,14]	1	>9	
	K78	-	1	4	-	1	6	1	13	Schuler, 2015
	PKo	-	1	7[8]	-	1	7	1	17	Casjens, 2018
B. spielmanii										
	A14S	1[7]	1	4[7]	-	1	5[7]	1	13	Casjens, 2018
B. valaisiana										
	VS116	1	1	3[8]	-	1	4	1	11	Casjens, 2018
B. maritima										
	CA690	-	1	2	-	1	3	1	7	Margos, 2019 Casjens, unpub
B. turdi										
	T1990A	1	1	3[2]	-	1	3	1	10	Margos, 2019
	TPT2017	-	1	3[2]	-	1	4	1	10	Margos, 2019
	T2084	-	1	3[2]	-	1	3	1	9	Margos, 2019

[1]The sequences were analyzed in the following references (Kingry *et al.*, 2016; Casjens *et al.*, 2017; 2018; Margos *et al.*, 2017; 2019; Jabbari *et al.*, 2018). The Bbss chromosomal SNP type (Mongodin *et al.*, 2013) is indicated in parentheses; The SNP types were predicted for PAbe, PAli and MM1 chromosomes by our analysis of cp26, lp54 and chromosome right end (see text) (S. Casjens and W. Qiu, unpublished).

[2]Number of cp32 PFam32 compatibility types, not DNA molecules. Strain B31 may have had 9[th] cp32 (cp32-5) that was lost before genome sequencing (Zuckert and Meyer, 1996; Casjens *et al.*, 1997).

[3]This category includes all the linear plasmids except lp5, lp17 and lp54 (see text).

[4]The number of different plasmid DNA molecules; fused plasmids counted as one.

[5]B31 may have had a second cp9 that was lost before genome sequencing (Miller *et al.*, 2000).

[6]cp32-10 is integrated into the lp56 linear plasmid (Casjens *et al.*, 2000).

[7]Assembly of one or more of these plasmids sequences was incomplete.

[8]One or more cp32s either truncated or fused with other cp32s.

[9]Includes lp28-1 since Illumina sequencing methods found some lp28-1 sequences; it may have been lost in most cells in the culture.

[10]Lp36 and lp28-4 are fused end-to-end.

[11]The 297 chromosome has not been sequenced. Its chromosomal SNP type is inferred from the cp26 and lp54 plasmids (Mongodin *et al.*, 2013).

[12]Cp32-11 is integrated into lp54 (Casjens *et al.*, 2018).

[13]There may be additional plasmids in this group whose sequences were not completely assembled.

[14]Lp54 and lp38 are reported to be fused end-to-end in CA-11.2A.

[15]Sébastien Bontemps-Gallo, personal communication and unpublished; cp32s were not assembled; see relevant GenBank entries.

Table 3. Total Lyme Disease *Borrelia* plasmid numbers and sizes in 29 genomes from 9 species.

	burgdorferi	*finlandensis*	*bissettiae*	*mayonii*	*afzelii*	*garinii*	*spielmanii*	*valaisiana*	*maritima*	Size range
	17[1]	1	1	2	3	2	1	1	1	(kbp)[3]
Plasmid										
lp5	2	-	1	-	-	-	-	-	-	5
lp17	17	1	1	2	3	2	1	1	1	15-28
lp21	2	-	-	-	-	-	-	-	-	15-19
lp25	9[2]	-	1	2	-	2	-	1	-	22-33
lp28-1	8	-	-	-	-	-	-	-	-	24-30
lp28-2	5	1	-	-	3	-	-	-	1	21-36
lp28-3	14[2]	-	1	2	2	1	1	1	-	22-82
lp28-4	15	1	1	1	3	1	1	-	1	22-41[4]
lp28-5	6	-	-	-	-	-	-	-	-	24-28
lp28-6	6	-	-	-	-	-	-	-	-	27-31
lp28-7	1	-	1	-	2	1	-	-	-	27-34
lp28-8	1	-	-	2	2	1	1	1	-	18-23
lp28-9	1	-	-	-	2	2	-	-	-	27-39
lp32-3	2	-	-	-	-	-	-	-	-	17
lp32-6	-	1	-	-	-	-	-	-	-	48
lp32-10	-	-	-	-	2	1	-	-	-	30-42
lp32-12	-	1	-	-	-	-	-	-	-	53
lp36	17	-	-	2	-	2	2	2	1	17-46[4]
lp38	12	-	-	-	3	-	1	-	-	20-39
lp54	17	1	1	2	3	2	1	1	1	53-61[4]
lp56	6	-	1	-	-	-	-	-	-	21-30[4]
cp9	8[2]	-	1	2	-	-	1	1	-	8-11
cp26	17	1	1	2	3	2	1	1	1	26-27
cp32's	108[5]	3	7	10	15	2	4	3	1	27-32[6]

[1]Below the species name is the number of genomes analyzed in the following references (Kingry *et al.*, 2016; Casjens *et al.*, 2017; 2018; Margos *et al.*, 2017; 2019; Jabbari *et al.*, 2018); they are those in table 2 but not including strains PBi, 20047 and BO23 since their plasmids are likely incompletely described.

[2]Includes unsequenced 297 lp25 and N40 lp28-3 and second B31 cp9 type (see Casjens *et al.*, 2017).

[3]The upper size estimates can be slight underestimates since most sequences do not extend quite to the termini of the plasmids (see Casjens *et al.*, 2012).

[4]Sizes do not include neatly integrated ~30 kbp cp32s in B31 lp56 and SV1 lp54 or fused lp36 and lp28-4 in CA-11.2A

[5]Number of cp32 apparently intact PFam32 genes and not DNA molecules (see text).

[6]All apparently intact cp32s are between 27 and 32 kbp; however, a significant fraction of cp32s exhibit large deletions and these can be as small as 13.5 kbp.

different the PFam32 proteins must be to define different plasmid compatibilities. Although the PFam32 proteins form well-defined clusters, there is insufficient experimental evidence to determine whether such clusters indeed always have different compatibilities. The most closely related PFam32 types that must have different compatibilities, since they reside in the in the same cell, are cp32-8, cp32-10 and cp32-12, all of which are present in *B. burgdorferi* strain JD1and are about 25% different in amino acid sequence (see figure S1 in Casjens *et al.*, 2018). lp28-1 and lp28-9 (17-18% different PFam32s) have been arbitrarily split into different PFam32 types (Casjens et al., 2017), whereas it has been argued that lp28-10 (Kingry et al., 2016) and lp28-8, at 12-15% different, may be the same compatibility group (Casjens et al., 2018). However, it should be emphasized that in the absence of finding a cell that contains both types, experimental tests of compatibility in these cases will be required to ascertain whether these groups have the same or different compatibilities. Finally, plasmids smaller than ~15 kbp, lp5 and cp9, often do not encode a PFam32 protein and so cannot be categorized in this

fashion. For example, *B. burgdorferi* strain WI91-23 (Casjens et al., 2017) and perhaps the original *B. burgdorferi* B31 isolate (Miller et al., 2000) contain two different cp9 plasmids that lack PFam32 genes, indicating that there are multiple compatibility types for this plasmid. These may correlate with differences in their PFam57/62 proteins (which are encoded by *all* LB species plasmids), but this remains speculative.

Plasmid relationships and gene content

How many plasmid compatibility types exist in LB species? Thirty-two different PFam32 types have been reported (Casjens et al., 2018). These include two types that were previously noted as: (i) a second type of PFam32 gene on *B. burgdorferi* B31 lp28-1 (*bbf13*) and recently found as the only PFam32 gene on a *B. burgdorferi* strain B331 plasmid named lp28-11 (Schwartz et al., unpublished), and (ii) another type (unnamed because the plasmids they reside on have not been completely sequenced or described) present in contiguous sequences (contigs) in a *B. afzelii* PKo draft genome sequence (locus_tag BAPKO_2556) and a *B. japonica* HO14 draft genome sequence (locus_tag SAMN02983004_01117). In addition, two new types recently discovered in *B. turdi* have been named lp28-12 and lp28-13 (the former designated lp30 in (Margos et al., 2019a; S. Casjens and W. Qiu, unpublished). This brings the current total to 34 PFam32 types. Thus, with lp5 and *at least* two types of cp9, there appears to be a minimum of 37 LB plasmid compatibility types, suggesting that it would be theoretically possible that a single cell could harbor 37 different plasmids. The rate of discovery of new PFam32 types has greatly diminished as more genome sequences have been determined, so it is likely that nearly all extant types are now known.

For LB strains in which the plasmid copy numbers have been measured, the range is one to two copies per chromosome. The relative staining intensities/bp of all the plasmid and chromosome electrophoresis bands are typically quite similar, suggesting that plasmids are generally present in low copy number (Hinnebusch and Barbour, 1992; Casjens and Huang, 1993; Glockner et al., 2004; Tilly et al., 2012). The observation that sequence coverage was 10-fold higher for *B. burgdorferi* strain JD1 lp28-6 than for other plasmids suggests that there may be exceptions to this rule (Casjens et al., 2012). The LB plasmids also have a number of additional interesting and unusual features: (i) Only a small fraction of the plasmid genes encode proteins with recognizable similarity to proteins outside of the *Borrelia* genus; these include proteins related to previously known plasmid partitioning proteins (above), small molecule transporters, DNA restriction-modification systems, and nucleotide and DNA metabolism enzymes. (ii) The plasmids encode many N-terminally lipidated proteins most of which are targeted to the outer surface of the bacteria (*e.g.*, Casjens et al., 2000; Dowdell et al., 2017). These proteins are important mediators of interactions between spirochetes and their hosts and are potential vaccine and detection targets. (iii) A large number of paralogous gene families and paralogous intergenic sequences are present on the plasmids (Casjens et al., 2000). (iv) Like the main chromosome, the linear plasmids have covalently closed hairpin ends (Barbour and Garon, 1987; Hinnebusch et al., 1990; Hinnebusch and Barbour, 1991, 1992) (see Radolf and Samuels, 2021). (v) Most of the LB plasmid types appear to have undergone structural rearrangements at a significant frequency in nature, judging from comparisons of cognate PFam32 type plasmids present in different isolates. Each resulting structural variant or "organizational subtype" is typically restricted to a single species. The different plasmid types and their rearrangements are discussed in more detail below.

Cp9

Six of the seventeen *B. burgdorferi* complete genome sequences and five of the fifteen genomes from nine other LB species include a cp9. Although the cp9s from *B. burgdorferi* do not carry a PFam32 gene, those from the other LB species encode a unique PFam32 type (Casjens et al., 2018). The different species harbor different cp9 organizational subtypes, and three are known in *B. burgdorferi*, including two in strain WI91-23. The *B. burgdorferi* cp9s encode PFam95 *eppA* genes (Champion et al., 1994; Miller and Stevenson, 2003), and all cp9s encode PFam96 and PFam165 proteins, also called BppC and BppA, respectively (El-Hage and Stevenson, 2002), which also are encoded by all intact cp32 plasmids; the roles of these proteins are not known. Cp9s may be readily lost during in vitro passage (Grimm et al., 2003), and there is no current evidence that they are involved in virulence.

Cp26

All LB isolates analyzed to date carry very similar, syntenic cp26 plasmids (Tilly et al., 1997; Terekhova et al., 2006; Casjens et al., 2012; 2017; 2018) and it

has been shown to be essential for growth in culture in *B. burgdorferi* strain B31 (Byram et al., 2004; Jewett et al., 2007a). These plasmids encode proteins involved in GMP synthesis (Margolis et al., 1994), chitobiose import (Tilly et al., 2004), host integrin binding (Behera et al., 2008), oligopeptide import (Bono et al., 1998), and the telomere resolvase that creates the hairpin ends of the linear replicons (Ravin et al., 2000; Kobryn and Chaconas, 2002) (see Radolf and Samuels, 2021). It also encodes the important surface antigen lipoprotein, OspC, that is expressed early in infection of a mammalian host and is essential for establishment of infection (Wilske et al., 1986; Grimm et al., 2004; Tilly et al., 2006; Fingerle et al., 2007). *ospC*, one of the most variable single-copy genes in the LB genome (Mongodin et al., 2013), appears to have undergone a number of short horizontal exchange events (Livey et al., 1995; Qiu et al., 1997; Wang et al., 1999a; Wang et al., 1999c; Barbour and Travinsky, 2010; Brisson et al., 2010; Haven et al., 2011) Single nucleotide polymorphism (SNP) tree analysis shows that, in the strains whose whole genome sequences have been determined, whole cp26 plasmids have not exchanged between chromosomal lineages, either within or between LB species (Mongodin et al., 2013).

Cp32s

The cp32 plasmids are largely syntenic bacteriophage prophages (see below) that can be quite similar throughout their lengths - in some cases 99% identical over several kbp (which makes unambiguous sequence assembly difficult without very long run sequence technology). Most LB isolates carry one or more cp32 plasmids (Casjens et al., 1997b; Iyer et al., 2003; Miller and Stevenson, 2003; Casjens et al., 2017; 2018). The two *B. garinii* isolates and one *B. maritima* isolate whose genomes have been sequenced have the fewest with two or fewer (*B. garinii* Far04 is the only isolate that carried no cp32s when its genome was sequenced) (Casjens et al., 2018; Margos et al., 2020). Their putative phage virion assembly proteins (see below) are presumably only expressed when the prophages are induced to lytic growth (Zhang and Marconi, 2005), but the cp32s also encode a number of surface exposed lipoproteins that are expressed from the prophage. The latter include the *revA* genes whose surface lipoprotein products bind fibronectin (Brissette et al., 2009a; Byram et al., 2015), the *mlp* encoded surface lipoproteins (Theisen, 1996; Yang et al., 1999; Porcella et al., 2000), the *bdr* (*Borrelia* direct repeat) genes whose functions are unknown (Zuckert et al., 1999; Zuckert and Barbour, 2000), and the complex family of *erp* (also called *ospE, ospF* or *elp*) genes whose various members have been shown to encode surface lipoproteins that bind to plasminogen (Brissette et al., 2009b), laminin (Brissette et al., 2009c), and factor H complement regulatory factor (Alitalo et al., 2002; Metts et al., 2003; Kraiczy et al., 2004), as well as having possible cell adhesin function (Antonara et al., 2007).

Twelve different types of PFam32 proteins are encoded by the cp32 group of circular plasmids, and these apparently can exist in the same cell (the name cp32-2 is not currently in use as it has the same PFam32 type as cp32-7 (Fraser et al., 1997; Casjens et al., 2000). Relatively frequent homologous recombination appears to shuffle sequences among the cp32s as well as allowing fusions between them (Brisson et al., 2013; Casjens et al., 2017; Margos et al., 2017b). In addition, about 15% of the cp32s that have been sequenced have undergone some apparently nonhomologous organizational rearrangements such as truncation, fusion and inversion.

Lp5

Lp5 is a relatively rare small plasmid that has been found only in two isolates of *B. burgdorferi* and one of *B. mayonii* (Casjens et al., 2000; Kingry et al., 2016; Casjens et al., 2017). They appear to encode only a PFam57/62 plasmid maintenance protein, a PFam137 protein of unknown function and several pseudogenes. There is no current indication that lp5 is involved in tick transmission or mouse virulence.

Lp17

Lp17, along with cp26 and lp54, is one of the three plasmids found in all natural LB isolates. It encodes unique PFam32 and PFam57/62 plasmid maintenance/compatibility proteins; in addition, all known lp17s encode homologues of *B. burgdorferi* B31 proteins BBD9, 10, 11, 15 and 18. Except for BBD18, a global regulator that affects the function of the alternative sigma factor RpoS and thereby controls the transition from mammalian host to tick vector (Dulebohn et al., 2014; Hayes et al., 2014), none of the other genes have a known function. Curiously, among the 32 lp17s in completely sequenced genomes, there are currently 16 different organizational variants (Figure 2) and nearly all are restricted to one species. Nearly all such variants are restricted to one species, but a single species can

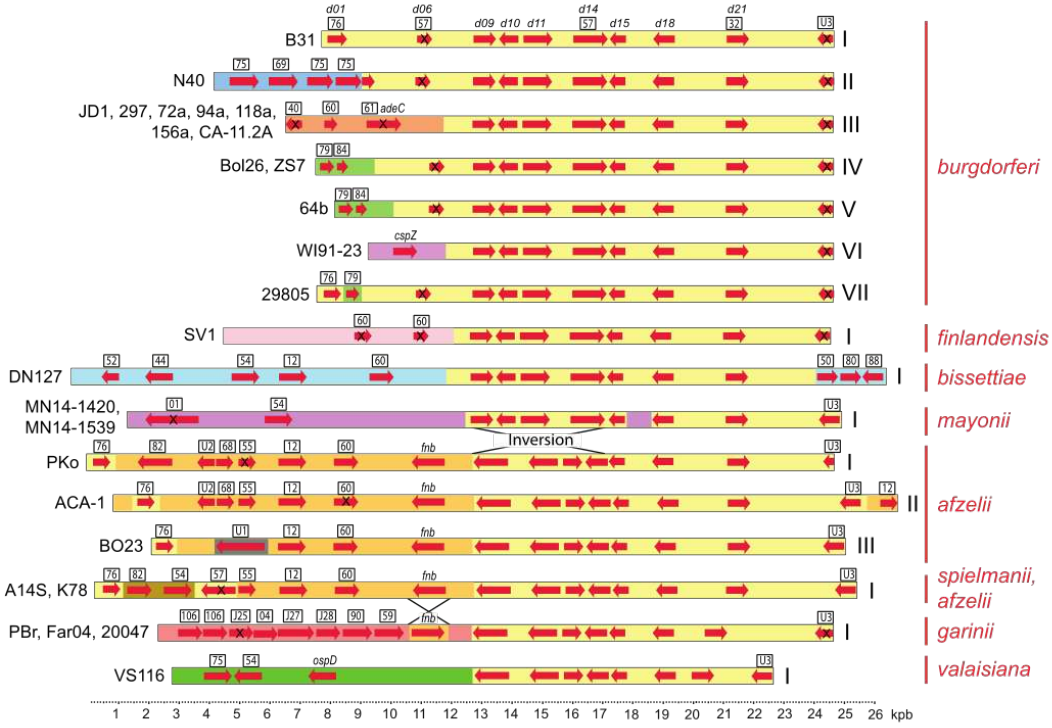

Figure 2. ORF maps of lp17 plasmids**.** The horizontal bars represent all the currently sequenced and annotated lp17 plasmids. identical background colors indicate regions of homologous DNA in the different plasmids. The bacterial species and organizational subtypes are indicated by Roman numerals on the right, and isolates that carry each subtype are indicated on the left. Selected genes are indicated by red arrows, and black "X"s mark pseudogenes. *B. burgdorferi* paralogous family (PFam) numbers (Casjens et al., 2012) are indicated in the boxes above; "U"s in boxes are proteins for which no intact gene is known in *B. burgdorferi*; "J's in boxes indicate strain B31 homologous proteins for which there is only one homologue in B31 (*i.e.*, no PFam number exists); *adeC, cspZ* and *fbn* refer to adenine deaminase, complement regulator-acquiring surface protein Z and fibronectin binding protein encoding genes, respectively. Selected strain B31 gene names are indicated above its map. This figure is reproduced with permission from Casjens *et al*. (2018).

have up to seven different variants (as in the case of *B. burgdorferi* where 17 isolates have been sequenced (Casjens et al., 2017; Margos et al., 2017b; Casjens et al., 2018; Jabbari et al., 2018; Margos et al., 2020). Interestingly, all the LB lp17 variants include alternate sequences (3-21 kbp long) at their left ends, and all of these alternate left end sequences encode proteins that are homologous to proteins encoded by other linear plasmids.

Lp54

Lp54 plasmids are present in all LB isolates and are relatively uniform in gene content and organization. Two exceptions are apparent fusions with lp38 in strain BO23 and with cp32-11 in SV1. One variation among canonical, unfused lp54s is the presence of different 2-3 kbp left end extensions in (i) *B. chilensis,* (ii) *B. japonica* and (iii) *B. afzelii, B. garinii* and *B. spielmanii* that encode different proteins (Samuels et al., 1993; Casjens et al., 2018). The lack of repeated sequences on the lp54s allows easier sequence assembly, so more lp54 plasmid sequences have been reported than are present in the complete sets of plasmids that are listed in Tables 2 and 3.

Lp54 carries a number of genes that are important in mouse infection and tick transmission. The expression of one-third of lp54 genes is regulated by RpoS; these include genes required during both the tick and mammalian phases of the spirochete life cycle (Caimano et al., 2007; Iyer et al., 2015; Caimano et al., 2019). OspA and OspB are major surface proteins that are expressed in ticks and are important for vector colonization (Pal et al., 2004; Yang et al., 2004; Tilly et al., 2012). Decorin binding proteins (DbpA and B) are important in mouse infection (Blevins et al., 2008; Shi et al., 2008; Weening et al., 2008; Salo et al., 2015). BBA34 (OppA5) is an oligopeptide binding protein that is required for mammalian infection (Bono et al., 1998; Groshong et al., 2017; Caimano et al., 2019) and BBA57 protein is critical for early infection success (Bernard et al., 2018). Bestor *et al*. (2010; 2012) and Promnares *et al*. (2009) used directed genetic analysis to show that the BBA03 and BBA62 (lp6.6) proteins, respectively, are important during tick transmission to mice. The PFam54 gene cluster on lp54, which encodes proteins BBA64-A73 in strain B31, includes genes that affect host complement function and/or bind host plasminogen (Wallich et al., 2005; Hallstrom et al., 2013; Koenigs et al., 2013; Hammerschmidt et al., 2016; Kraiczy, 2016). BBA64 is required for tick transmission (Gilmore et al., 2010). The central portion of this tandem gene array (*e.g.,* B31 genes *bba68-bba72*) varies in gene content and gene number both within and between species (Qiu et al., 1997; Wywial et al., 2009). Four conserved PFam54 genes make up the constant outside portions of this cluster, and there are 28 different full-length gene types in the variable regions of 28 LB genomes. The number of PFam54 genes in the central variable region ranges from two in *B. chilensis* to six or eight in *B. afzelii, B. bissettiae, B. garinii* and *B. valaisiana* isolates (Casjens et al., 2018). Essentially identical arrays are often present in different isolates of the same species, and any given array type is apparently restricted to a single species. This observation, along with single nucleotide polymorphism (SNP) tree analysis, shows that whole lp54 plasmids have not exchanged among the LB species analyzed (Mongodin et al., 2013).

Other linear plasmids

The remaining LB plasmids are linear and vary in size from 15 kbp to 82 kbp, although the vast majority are between 24 and 40 kbp. These include lp21, lp25, lp28-1, -2, -3, -4, -5, -6, -7, -8, -9, -11, lp32-3, -6, -10, -12, -13, lp36, lp38 and lp56 (lp28-10 is not used here because it was used by Kingery *et al.* (2016), and analyses of lp28-11, -12 and -13 mentioned above have not yet been published) (see Table 2). In addition, the left end of lp17 (above) and the right end of the chromosome (Banik et al., 2011; Casjens et al., 2012; 2018) have a number of different extensions that appear to be fragments derived from these "other" plasmids. These plasmids are discussed together since they have the following common properties: (i) They contain numerous similar sequence patches. (ii) Many of them have an exceptionally low (for bacteria) density of protein coding genes. (iii) They carry an unusually large (for bacteria) number of pseudogenes with reading frames that have been damaged by frame-disrupting mutations and/or large deletions and insertions relative to intact homologues. The ten *B. burgdorferi* B31 plasmids which contain most (87%) of this strain's pseudogenes have a total predicted protein coding fraction of 0.42, a much lower value than that (>0.90) typical of bacterial DNA (Casjens, 2000). (iv) Each PFam32 type includes plasmids from different isolates of quite different lengths (see Table 3) (Palmer et al., 2000; Casjens et al., 2012; 2017; 2018). (v) Finally, these plasmids have patchwork or mosaic relationships wherein patches of similar sequence appear to be more or less randomly scattered. When two "cognate" (*i.e.*, same PFam32 type) plasmids from this group are compared, they typically contain patches of high sequence similarity and yet may have immediately adjacent sequences that are completely different. The rearrangements that generated the mosaic relationship patterns appear to be largely non-homologous (Casjens et al., 2012; 2017; 2018). These observations suggest a tumultuous history of (often duplicative) random rearrangements followed by decay of broken and redundant genes. The reasons for this type of evolution remain mysterious.

Such apparently chaotic evolution has generated multiple organizational subtypes for all of this group of linear plasmid types. The number of such subtypes is not small (*e.g.,* see lp17 section above), but the fact that essentially identical cognate plasmids have been found in many strains suggests that nonetheless a limited number of subtypes exist for each plasmid type. The rearrangement process that generates these organizational subtypes is ongoing, since different paralogous segments have diverged to different degrees and individual subtypes are nearly always limited to one species or one lineage within a species (Casjens et al., 2017; 2018). On the

other hand, the process is not so fast that every isolate or lineage has a completely unique set of organizational subtypes. The fact that no such rearrangement has been observed during laboratory cultivation indicates that such events are quite rare. Nonetheless, a small number of exceptionally similar plasmids have been found in more than one species. In particular, nearly identical lp5s in two *B. burgdorferi* lineages and *B. mayonii* and extremely similar lp17s in a *B. afzelii* strain and a *B. spielmanii* strain suggest that rare whole linear plasmid transfer may occur (Casjens et al., 2017; 2018).

Plasmid roles in Lyme disease

Numerous directed gene knockout and global scanning strategies have been used to identify plasmids or encoded genes in *B. burgdorferi* strain B31 that are required for infection of mice or survival in and transmission by ticks (*e.g.,* (Purser and Norris, 2000; Labandeira-Rey and Skare, 2001; Grimm et al., 2005; Strother et al., 2005; Lin et al., 2012; Ellis et al., 2013; Krupna-Gaylord et al., 2014; Troy et al., 2016; Ramsey et al., 2017; Phelan et al., 2019) (see Radolf and Samuels, 2021). These studies found several plasmids, lp25, lp28-1, lp36, lp54 and cp26, to be essential for strain B31 pathogenesis in mice, while lp25, lp28-1 and lp28-4 are critical for tick transmission. Numerous individual genes on these plasmids have been identified as having essential or modulatory roles in these processes, including the following: lp28-1 carries the *vlsE* gene that encodes an outer surface protein and the cassettes with which *vlsE* exchanges sequences to produce antigenic variation (Norris, 2014; Verhey et al., 2019) and *arp* (Barthold et al., 2006). Lp25 contains the nicotinamidase-encoding *pncA* gene that is critical for mouse infection (Purser et al., 2003; Jewett et al., 2011) and *bptA* which is essential for tick transmission (Revel et al., 2005). Lp36 carries *bbk13*, which encodes a non-surface-exposed membrane protein important in early skin infection (Aranjuez et al., 2019); *bbk17 (adeC)*, which encodes adenine deaminase, important for mouse infectivity (Jewett et al., 2007b); *bbk32*, which encodes a surface exposed lipoprotein that binds fibronectin and certain glycosaminoglycans through its N-terminal domain and inhibits the classical complement pathway through its C-terminal domain (Probert and Johnson, 1998; Fischer et al., 2006; Seshu et al., 2006; Xie et al., 2019); and *bbk46*, which is critical for evasion of the humoral immune response in mice (Ellis et al., 2013). The roles of individual genes on lp54 and cp26 were discussed above. Numerous other candidates for plasmid genes that play roles during infection of mice or ticks have been identified by various global strategies, but have not yet been studied in detail (Lin et al., 2012; Dulebohn et al., 2013; Ellis et al., 2013; Phelan et al., 2019). Although they have not been shown to be directly important in pathogenesis, strain B31 lp25 and lp56 genes that encode restriction/modification proteins are technically important in that their inactivation is required for efficient transformation of *B. burgdorferi* strain B31 cells by exogenous DNA (Lawrenz et al., 2002; Kawabata et al., 2004; Rego et al., 2011). We note that, because of the past rearrangements among these plasmids, orthologs of these genes are not always carried on plasmids of the same PFam32 type. For example, the *vlsE* gene/*vls* cassette system is carried on lp28-1, -3, -8, -9, lp32-3, -6 or lp36 in different LB isolates, and *pncA* is found on lp25, lp28-2 or lp38 (Casjens et al., 2017; 2018).

Almost no genes with homology to previously known virulence genes in other bacterial families have been found in the LB. The one exception is a cluster of genes (so far, always on an lp28-8 plasmid) that encode, modify and probably export a short 25 amino acid peptide toxin (Molloy et al., 2015). This toxin, borreliolysin S, and its putative extensive modifications are very similar to streptolysin S, a cytolytic molecule produced by *Streptococcus pyogenes* strains, whose mechanism of action is not well understood. Members of this family of toxins are reported to target eukaryotic or prokaryotic cells (Quereda et al., 2017; Tsao et al., 2019). Interestingly, this set of toxin forming genes is present in some isolates of *B. afzelii, B. spielmanii, B. valaisiana, B. japonica* and *B. lusitaniae* (its presence in *B. garinii* is still uncertain), but they have not been found in *B. burgdorferi*, *B. finlandensis, B. mayonii*, or *B. bavariensis* isolates (Molloy et al., 2015; Casjens et al., 2018; S. Casjens, unpublished). Thus, the presence of the toxin genes does not correlate with the LB species that cause human Lyme disease and it is not a critical requirement for that disease, although it could be involved in acrodermatitis chronica atrophicans, a cell chronic inflammatory of Lyme disease that is most commonly correlated with *B. afzelii* infection (Maraspin et al., 2019).

Comparative Genomics

Typing methods

Typing of bacterial strains has great importance for clinical and epidemiological studies, for diagnostics, as well as for population genetic or evolutionary

research of LB strains. While many PCR-based approaches, especially for diagnostic purposes, target single genes (based on time and financial considerations), a technique that was termed multilocus sequence typing (MLST) was invented in the late 1990s and has been used with great effect for bacterial epidemiology (Enright and Spratt, 1999; Spratt, 1999; Urwin and Maiden, 2003) In addition to methods described in this section, several other molecular typing methods, including DNA-DNA reassociation analysis, RNA gene restriction analysis (ribotyping), DNA sequencing of 16S rRNA or other conserved genes (e.g., *fla*, *ospA*, *hbb*), species-specific PCR, PCR-based restriction fragment length polymorphism (RFLP) analysis, pulsed-field gel electrophoresis (PFGE), randomly amplified polymorphic DNA (RAPD) fingerprinting, and variable-number tandem repeat (VNTR), have been developed and used for the identification and classification of LB species (Wang et al., 1999b; Farlow et al., 2002; Liveris, 2013). Except for OspC, these methods rely largely on chromosomal loci for typing determinations.

Discriminatory power varies among different molecular typing methods. In general, whole genome-based approaches such as PFGE, RAPD, MLST and whole genome sequencing (WGS) have the highest resolution. *ospC* typing was frequently used to differentiate strains within the same species, in particular in *B. burgdorferi*, and for disease association studies (Seinost et al., 1999; Wang et al., 1999c; Qiu et al., 2002; Brisson and Dykhuizen, 2004; Hanincova et al., 2008; Qiu et al., 2008; Wormser et al., 2008), but MLST is better suited than *ospC* for differentiation of disease associated sequence types (Hanincova et al., 2013) (see below). Depending on the study purpose (i.e., required discriminatory power) and available resources, appropriate selection of one or more molecular typing methods is critical to yield species, subtype and strain-specific information. Multilocus sequence typing/multilocus sequence analysis (MLST/MLSA) and WGS have been widely employed for new species delineation, epidemiology, population genetics and evolutionary studies of different LB species (Vitorino et al., 2008; Hoen et al., 2009; Margos et al., 2009; Margos et al., 2010; Ogden et al., 2010; Ogden et al., 2011; Vollmer et al., 2011; Margos et al., 2012; Hanincova et al., 2013; Vollmer et al., 2013; Jacquot et al., 2014; Carpi et al., 2015; Gatzmann et al., 2015; Mechai et al., 2015; Castillo-Ramirez et al., 2016; Walter et al., 2016; Walter et al., 2017; Tyler et al., 2018; Margos et al., 2020).

In this section, the focus will be on the following methods: *rrs-rrlA* (16S-23S) intergenic spacer (IGS) typing, *rrf-rrl* (5S-23S) IGS typing, *ospC* typing, MLST/MLSA and WGS analysis. These methods can be employed alone or in combination for molecular characterization of LB cultured isolates and directly or indirectly on uncultivated organisms in ticks, mammalian host reservoirs and human clinical specimens. Most PCR based methods can be used for molecular typing of *B. burgdorferi* directly in a variety of field-collected and clinical specimens, but mixed infections which have been reported in many sample types (Qiu et al., 2002; Vitorino et al., 2008; Hoen et al., 2009; Ogden et al., 2010; Vollmer et al., 2011; Mechai et al., 2015; Walter et al., 2016), may not be disentangled (Margos et al., 2011). Some molecular typing methods (i.e., ribotyping, PFGE, RAPD) do require pure cultured spirochetes. Selection of particular genotypes or strains of *B. burgdorferi* during *in vitro* cultivation has been demonstrated (Norris et al., 1997; Liveris et al., 1999) and this potential “culture bias” needs to be considered in data interpretation when cultured isolates are used for molecular typing studies.

rrs-rrlA (16S-23S) intergenic spacer (IGS)

As described earlier, LB species possess a unique rRNA gene organization that is distinct from that of other prokaryotes (Schwartz et al., 1992; Ojaimi et al., 1994). In *B. burgdorferi,* the rRNA gene locus consists of a single 16S rRNA gene (*rrs*) followed by a large intergenic spacer that varies in size among different LB species. This is followed by a tandem repeat of 23S rRNA-5S rRNA genes in most LB species that is separated by a short spacer of 200-350 bp (Schwartz et al., 1992; Gazumyan et al., 1994; Ojaimi et al., 1994; Fraser et al., 1997). Two different rRNA-based PCR-RFLP typing methods have been developed for *B. burgdorferi*, targeting either the *rrs-rrlA* spacer or the *rrfA-rrlB* spacer (Postic et al., 1994; Liveris et al., 1995; Rijpkema et al., 1995).

PCR amplification of the proximal 941 bp of the *rrs-rrlA* spacer followed by RFLP analysis using restriction endonuclease digestion distinguishes *B. burgdorferi* strains into three ribosomal spacer types referred to as RST1, RST2 and RST3. Studies have shown that RST types correlate with pathogenic potential in humans and mice (Liveris et al., 1999;

Wormser et al., 1999; Wang et al., 2001; Wang et al., 2002; Wormser et al., 2008). Application of this typing scheme to uncultivated spirochetes from LD patients and field-collected ticks showed that both may be simultaneously infected with one or more distinct genotypes of *B. burgdorferi* (Liveris et al., 1999; Ranka et al., 2004). The *rrs-rrlA* spacer PCR product can be subjected to direct sequencing and sequence polymorphism within the first 250 nucleotides of this PCR product, resulting in delineation of 10 distinct Intragenic Sequence (IGS) types (Bunikis et al., 2004). Using this approach, 127 *B. burgdorferi* clinical isolates were classified into 16 IGS types and demonstrated linkage disequlibrium between the *rrs-rrlA* IGS and *ospC* loci, indicating a lack of random distribution of the plasmid that harbors *ospC* (Hanincova et al., 2008) (see section below on population genomics).

rrfA-rrlB (5S-23S) IGS

The 5S-23S IGS was used extensively for species determination via PCR-RFLP, sequencing or reverse line blot analysis (e.g., Postic et al., 1994; Rijpkema et al., 1995; Postic et al., 1998; Saint Girons et al., 1998; Kurtenbach et al., 2001). It is still used today for sample screening and LB species determination, as well as for resolving mixed LB species infection, (e.g., Jenkins et al., 2012; Coipan et al., 2016; Blazejak et al., 2018). While single locus approaches appear attractive and economical for sample screening or diagnostic purposes, the problems associated with single locus approaches for bacterial species determination or strain characterization are well known. Single loci provide a genealogy only for the targeted locus not for the organism, do not buffer against exchange of genetic material and PCR may show differential sensitivity (Urwin and Maiden, 2003; Lager et al., 2017). Phylogenetic analyses of IGS sequences may give deceptive results in species assignment, as exemplified by *B. afzelii, B. lanei* (previously named genomospecies 2), and *B. garinii/ B. bavariensis* (Postic et al., 2007; Margos et al., 2009; Coipan et al., 2016; Margos et al., 2017a; Sabitova et al., 2018).

ospC typing

As noted, OspC is essential for *B. burgdorferi* to establish a productive initial infection in mammals (Tilly et al., 2006). *ospC* was among the first single gene loci to be targeted for strain typing and has been used most widely for *B. burgdorferi* strain differentiation. *B. burgdorferi* samples are most commonly genotyped by amplifying a 617 bp region of *ospC* (Seinost et al., 1999; Wang et al., 1999c; Qiu et al., 2002). *ospC* typing has been employed to investigate *B. burgdorferi* genetic diversity in environmental samples (e.g. ticks) (Wang et al., 1999c; Qiu et al., 2002; Brisson and Dykhuizen, 2004; Barbour and Travinsky, 2010) and different allelic variants are associated with a differential capacity to cause disseminated infection in mammals (Seinost et al., 1999; Wormser et al., 1999; Wang et al., 2001; 2002; Wormser et al., 2008). *ospC* is one of the most highly polymorphic gene in LB species; by convention, *ospC* major groups are designated using the criteria that *ospC* alleles are <2% different within a group and >8% divergent between groups (Wang et al., 1999c). Approximately 25 *ospC* major groups have been identified in *B. burgdorferi* populations (Barbour and Travinsky, 2010; Tyler et al., 2018). Aspects of *ospC* in relation to population genomics are presented in later sections.

MLST and MLSA

MLST and MLSA (the application of MLST at the genus level) were developed from MultiLocus Enzyme Electrophosis (MLEE) and provide a sensitive method for bacterial strain characterization and species delimitation (Enright and Spratt, 1999; Spratt, 1999; Urwin and Maiden, 2003; Bishop et al., 2009). Typical MLST/MLSA schemes are based on evolutionary congruent (Loza Reyes, 2010) housekeeping loci and certain criteria should be considered for selecting such genes (Urwin and Maiden, 2003). These include that the chosen housekeeping genes: i) are scattered across the linear chromosome to avoid local bias, ii) should not be flanked by genes under strong positive selection pressure as this may influence the neighboring genes, and iii) all should have similar levels of genetic diversity so that each gene contributes to phylogenetic analyses and no single gene dominates a tree generated by use of the concatenated sequences.

Although several MLST/MLSA schemes have been introduced for LB, several of these depart from typical MLST in that they combine different categories of loci, such as genes encoding outer surface proteins, conserved housekeeping genes, or noncoding loci (Bunikis et al., 2004; Qiu et al., 2004; Richter et al., 2006; Postic et al., 2007). If loci that are not evolutionary congruent are combined in phylogenetic analyses, mixture models must be applied (Loza Reyes, 2010). The MLST scheme described here (Margos et al., 2008) adheres to the

principles of MLST systems as originally intended (Spratt, 1999; Urwin and Maiden, 2003). It has been integrated into the MLST database maintained at the University of Oxford at https://pubmlst.org/borrelia (Jolley et al., 2004; Jolley, 2009; Margos et al., 2015a).

The MLST method developed for the *B. burgdorferi* sensu lato species complex (Margos et al., 2011) also has been utilized in modified form for relapsing-fever species (Margos et al., 2015a; Boden et al., 2016; Fingerle et al., 2016; Kingry et al., 2018; Stete et al., 2018) (see Radolf and Samuels, 2021). It permits characterization of isolates at the different levels required for evolutionary, epidemiological and population/landscape genetic studies. The method consists of three steps: i) data acquisition, ii) data assembly and iii) analysis (reviewed by (Margos et al., 2011)). Data acquisition involves sample collection, DNA extraction, targeted PCR of housekeeping loci and sequencing, either by Sanger sequencing or via Illumina MiSeq sequencing (Kingry et al., 2018). Comparison of sequence fragments with each other or to isolates in the MLST database allows species identification, determination of DNA variation in strains and species, as well as identification of recombination. It reflects the evolutionary history of bacterial isolates more realistically than single locus approaches.

As the method considers single point mutation differences, amplicons need to be sequenced in forward and reverse directions for comparison when using Sanger sequencing. Good quality sequences of the same fragment length as available through the MLST website (https://pubmlst.org/borrelia) permit obtaining allele numbers. The allele number for each locus results in a chain of eight integers which corresponds to the allelic profile of the isolate and defines the sequence type (ST). Novel alleles or novel sequence types must be submitted to the MLST database (via the curator, see https://pubmlst.org/borrelia/submission of data) to obtain sequential numbers. For analysis of MLST/MLSA data either concatenated sequences for all loci or allelic profiles can be used in downstream methods such as phylogenetic analysis, eBURST/goeBURST (Francisco et al., 2012), or population genetics methods. Currently, the *Borrelia* MLST database (https://pubmlst.org/borrelia) contains data for >3,100 isolates and >900 STs (as of 05.05.2020).

The species threshold level, adjusted to the MLST loci employed in this system, was determined to be 0.017 (Margos et al., 2009) and is comparable to the species threshold suggested by DNA-DNA hybridization (Postic et al., 2007). As for DNA-DNA hybridization, slight variations in genetic differences around the threshold may be observed using MLST. This poses the question whether such isolates should be regarded as new species. For example, when *B. yangtzensis* was described, some isolates were below the species threshold compared to the type strain but not compared to other isolates of the species. It was concluded that all isolates should belong to the same species (Margos et al., 2015b), indicating that slight variations around the species threshold should be considered when proposing new species. Also, when delineating new LB species by MLSA, ecological data should be taken into account whenever possible as this has been an important consideration when describing new bacterial species in general (Cohan, 2002).

Whole Genome Sequencing

Next generation sequencing (NGS) has advanced bacterial genome sequencing at a very fast pace (reviewed by Niedringhaus et al., 2011; Kulski, 2016; Levy and Myers, 2016). However, sequencing and assembly of LB genomes using NGS methods has shown that in addition to accurate short-read methods, technologies that provide long-reads covering several thousand bases (such as Oxford Nanopore Technology (ONT) or Pacific Bioscience single molecule real-time (SMRT) technology) are required for complete genome assembly, including plasmids (Kingry et al., 2016; Margos et al., 2017b; Gofton et al., 2018; Kuleshov et al., 2020). The value of completed bacterial genomes vs. high-quality draft status has been emphasized (Casjens et al., 2000; 2012) (see plasmid section above) and analyses that are contingent on multi-copy genes important for bacterial niche adaptation can only be achieved using completed genomes (Margos and Becker, unpublished).

Currently, the most commonly used technologies for bacterial genome sequencing are Illumina, Pacific Bioscience, and Oxford Nanopore technologies. The principles of these sequencing technologies are well described in several reviews (Niedringhaus et al., 2011; Kulski, 2016; Levy and Myers, 2016). All three technologies have been used for sequencing LB genomes. The accuracy of sequences is best with Illumina. However, due to the shortness of reads (600

bp max) difficult regions in the genome (in particular plasmid sequences) may not be completely assembled using only Illumina reads (Margos et al., 2017b; 2017c; Tyler et al., 2018; Margos et al., 2019a; 2020). For example, repetitive sequences on the chromosome (e.g. *lmp1*) prevent complete assembly of these genomic elements (Tyler et al., 2018; Margos et al., 2019a). Good results have been obtained with the PacBio RSII system sequencing often in combination with Illumina technology to correct small insertions/deletions. However, it has been noted that plasmids may appear incorrectly fused in the assembly or a single plasmid may consist of two or more contigs even using SMRT technology (Margos et al., 2017b; Gofton et al., 2018; Kuleshov et al., 2020). Another issue is that long sequence "overhangs" can be found at the ends of the resulting contigs due to the adapters added during SMRT library production and circular sequencing; these overhangs must be removed. On the other hand, SMRT sequencing provides valuable information on telomere regions of linear replicons which may be difficult or impossible to obtain with the other sequencing methods (Margos and Casjens, unpublished). Although ONT provides the longest reads, the error rate in DNA sequencing is high (Margos et al., 2017b). This may have an effect on genome assembly. Assembly software available for hybrid assembly such as SPAdes (Bankevich et al., 2012) or Canu (Koren et al., 2016) will not always assemble all plasmids accurately; there may be several copies of the same plasmid or plasmids may be lost during assembly (in particular cp32s) (Margos and Casjens, unpublished).

There is great utility of WGS for species and strain typing as well as for phylogenetic investigations. Without doubt, WGS provides the highest level of resolution for isolate discrimination. MLST data and sequences for other genes are easily extracted from WGS data and this provides an opportunity to compare isolates between studies (e.g.Gatzmann et al., 2015; Becker et al., 2016; Castillo-Ramirez et al., 2016; Walter et al., 2017; Tyler et al., 2018). Tools for MLST and genomic analyses are available via several websites, among them https://pubmlst.org/borrelia/ or the website of the Center for Genomic Epidemiology (http://www.genomicepidemiology.org/). Several studies have shown that phylogenies based on a larger number of conserved single copy genes (i.e., more than eight MLST genes) or genomic SNPs provide better node support even for internal nodes (Becker et al., 2016; Tyler et al., 2018). Some examples and details are given in the following section.

Genome phylogeny and biogeography

The application of next-generation sequencing methods to cultured LB species (Jacquot et al., 2014; Gatzmann et al., 2015; Becker et al., 2016; Tyler et al., 2018) and *B. burgdorferi*-infected ticks using an oligo-capture approach (Carpi et al., 2015; Walter et al., 2017) permitted assembly of the chromosome and contiguous sequences (contigs) for several, more or less completed, plasmids. These data allowed the re-construction of robust phylogenies based on chromosomal or genomic SNPs or large numbers of orthologous single copy genes with high support for external and even internal nodes (Mongodin et al., 2013; Becker et al., 2016; Tyler et al., 2018). The utility of genomics for phylogenetic/biogeographic analyses is briefly described here as the biogeography of species and their populations are described in more detail in Radolf and Samuels (2021).

Investigations on Lyme disease risk in the United States revealed a pattern of distinct hotspots in the Northeast, the Midwest and California (Bacon et al., 2008; Schwartz et al., 2017). The pattern of Lyme disease occurrence is driven by the distribution of tick vectors (Diuk-Wasser et al., 2006) and wildlife hosts (Hanincova et al., 2006; Hamer et al., 2010; Ostfeld et al., 2018). Lyme disease risk also was affected by historical events impacting the distribution of *Ixodes* vectors and their reproductive hosts (Barbour and Fish, 1993; Spielman, 1994) followed by independent expansions of *Ixodes spp.* and *B. burgdorferi* in the Northeast and Midwest US (Hoen et al., 2009) (see also Radolf and Samuels, 2021). The pattern of sequence type distribution demonstrated that the focal *B. burgdorferi* populations east (NE and MW USA) and west of the Rocky Mountains were genetically distinct but phylogenetically related (Hoen et al., 2009; Margos et al., 2012; Hanincova et al., 2013). These data also suggested that *B. burgdorferi* had been resident in North America for several thousand, if not millions of years (Hoen et al., 2009), findings that were confirmed by phylogenomic analyses of *B. burgdorferi* from North America (Walter et al., 2017). The predicted (Ogden et al., 2006) and recently observed population expansion of *I. scapularis* and *B. burgdorferi* into Canada (Ogden et al., 2010; 2011) indicated an influx of *B. burgdorferi* genotypes from the different focal populations of the USA likely driven by conseq-

uences of climate change (Ogden et al., 2006; Mechai et al., 2015; Tyler et al., 2018). Unique genotypes found in Canada, both by MLST and genome analyses (Mechai et al., 2015; Tyler et al., 2018), clustered next to rare genotypes from the USA (Walter et al., 2017) which suggested that these are also likely to originate from the USA (Tyler et al., 2018) (Figure 3). All phylogenetic/phylogenomic investigations conducted to date showed that *B. burgdorferi* populations do not cluster according to geography; that is, Northeast, Midwest and California isolates are intermingled in phylogenies and even European isolates form clades nested between North American clades (Figure 3) (Hoen et al., 2009; Hanincova et al., 2013; Castillo-Ramirez et al., 2016; Walter et al., 2017; Tyler et al., 2018; but see Qiu et al., 2008). While this is suggestive of migration events between Europe and North America and perhaps numerous migration events between populations within North America, the exact nature or mechanisms of these exchanges (and whether they are historical or contemporary) is currently unknown (Qiu et al., 2008; Castillo-Ramirez et al., 2016; Walter et al., 2017). Some *B. burgdorferi* genotypes isolated

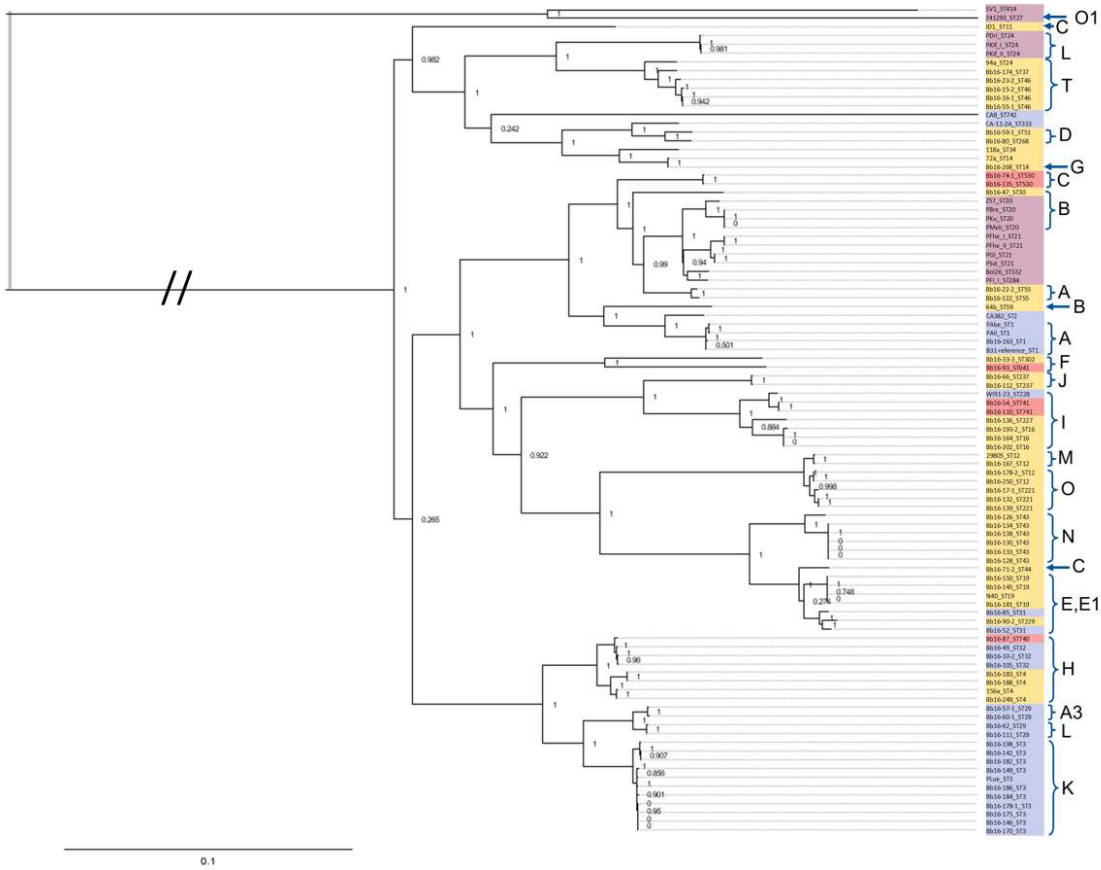

Figure 3. Population structure of *B. burgdorferi* inferred from genomic data. The figure shows a midpoint rooted phylogenetic tree of the *B. burgdorferi* chromosome generated via the SNVPhyl pipeline. All sequences are currently published full sequences from samples collected in North America and Europe. Color coding indicates the geographic occurrence of the samples: blue = USA only; red = Canada only, yellow = USA and Canada; violet = Europe. Letters to the right of the tree indicate the *ospC* major groups of the strains. Likelihood values for branches with a value less than 0.9 are shown. Reprinted with permission from (Tyler et al., 2018).

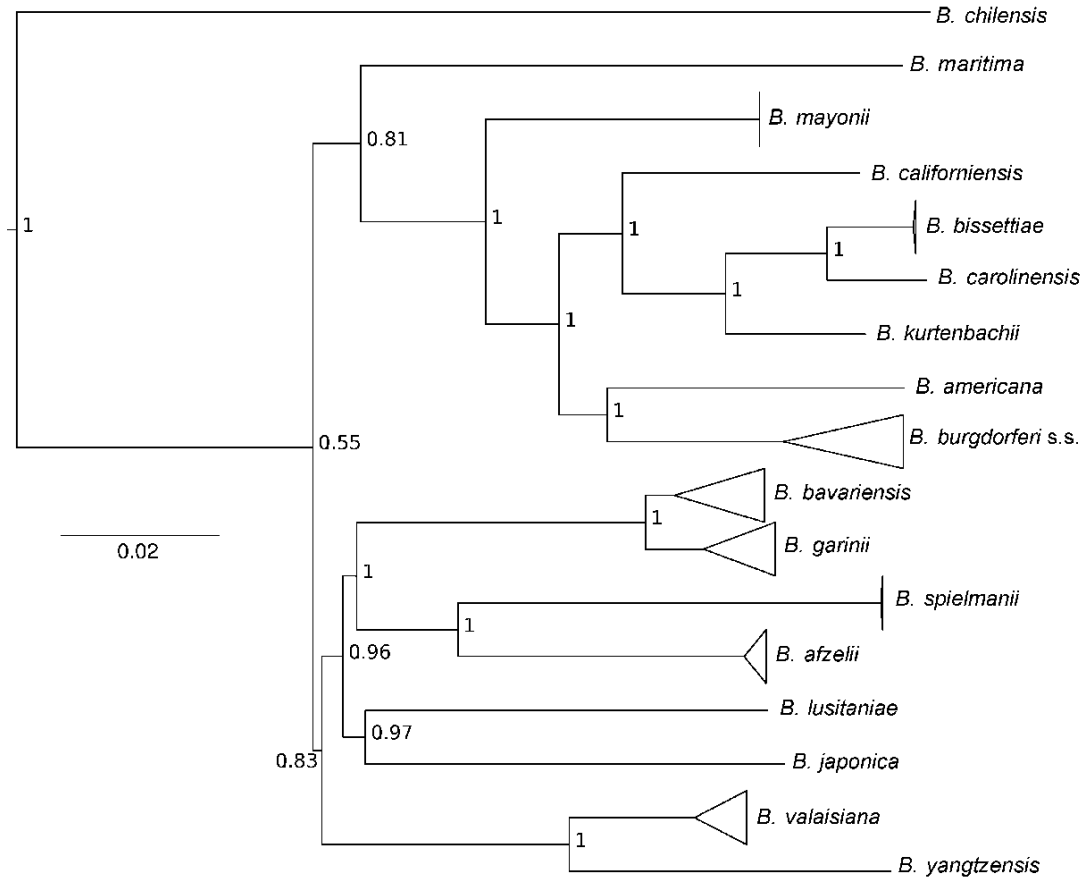

Figure 4. Phylogeny of 17 Lyme disease *Borrelia* species. Homologous sequences of chromosomal single copy genes were identified via BLASTn (v. 2.2.26) using *B. burgdorferi* B31 as reference (for details see Becker et al. 2016). DNA sequences of 113 genes were aligned in muscle (v. 3.8.31) and an unrooted maximum clade credibility tree was generated in BEAST (v.1.7.5). The tree contains 115 samples belonging to 17 of the currently known species. The nodes (branches) are very well supported; posterior probabilities are given next to the nodes. Scale bar corresponds to number of base substitutions per site. In the tree, *B. chinensis* takes a more basal position compared with other species. Two major lineages corresponding to a predominantly Eurasian clade and another predominantly North American clade are discernible. Figure modified from (Margos et al., 2020), with permission.

from patients in Europe were more closely related to North American genotypes than to tick-associated European genotypes; genome analyses suggested that these individuals may have acquired their infection in the USA (Castillo-Ramirez et al., 2016).

Approximately 90 genomes of various LB species were sequenced and used in combination with 18 previously published genomes for phylogenomic analyses (Becker et al., 2016). Phylogenetic reconstruction based on >100 conserved chromosomal single copy genes revealed that the genospecies formed clearly distinguishable clusters corresponding to (predicted) host associations. Populations of only two species (as investigated by MLST), namely *B. lusitaniae* and *B. bavariensis*, formed clusters that corresponded to geography (Vitorino et al., 2008; Margos et al., 2013). These data implied that LB host associations are an important driver of diversification and speciation but

that vector associations may also drive speciation (Gatzmann et al., 2015; Becker et al., 2016). There was strong node support throughout the tree and its topology, suggesting that host switching followed by speciation occurred several times in the evolutionary history of the LB species complex and is still ongoing. A phylogenetic tree for 17 LB species is presented in Figure 4.

A particularly interesting example relates to *B. garinii* and *B. bavariensis*. *B. garinii* is considered one of the most heterogeneous species within the LB species complex (Jacquot et al., 2014). Its wide distribution throughout Eurasia, in the Northern Hemisphere (including rarely the Atlantic coast of North America (Smith et al., 2006; Munro et al., 2019)) and Southern Hemisphere (Comstedt et al., 2011) is likely owing to its adaptation to terrestrial and sea birds as hosts and its ability to utilize several tick species as vectors (*Ixodes ricinus*, *Ixodes persulcatus*, *Ixodes. uriae* and *Ixodes pavlovsky*) (Gomez-Diaz et al., 2011; Vollmer et al., 2011; Mukhacheva and Kovalev, 2013; Vollmer et al., 2013; Norte et al., 2020). It has been speculated that the absence of *B. garinii* from terrestrial transmission cycles of North America may be due to vector incompetence of *I. scapularis* and *I. pacificus* (reviewed by Margos et al., 2019b). The population structure of *B. garinii* differs from that of *B. afzelii* (another widely distributed species throughout Eurasia) in that *B. garinii* MLST sequence types are more geographically mixed than those of *B. afzelii;* this has been attributed to migration patterns of their hosts (at least at a European scale) (Vollmer et al., 2011; Vollmer et al., 2013; Jacquot et al., 2014; Munro et al., 2019; Norte et al., 2020).

The closest related species to *B. garinii* is *B. bavariensis* with which it shares a most recent common ancestor (Margos et al., 2013; Gatzmann et al., 2015; Becker et al., 2016) (Figure 4). Speciation of *B. bavariensis* was likely due to a host switch from bird to rodent (Hu et al., 2001; Huegli et al., 2002). Phylogeographic analyses employing genetic data suggested that *B. bavariensis* has undergone a recent range expansion, invading Europe from Asia, its likely geographic origin. This range expansion was accompanied by a drastic genetic bottleneck, suggesting that the colonization of Europe was a single event, probably coinciding with a vector switch. Phylogenomics and biogeographic investigation confirmed the existence of two *B. bavariensis* populations, one in Eastern Europe/Asia that utilizes *I. persulcatus* as vector and a Western European population adapted to *I. ricinus*. Genome analyses of the chromosome and two plasmids (lp54 and cp26) confirmed a strong genetic bottleneck in, and supports the hypothesis for clonal expansion of, the European *B. bavariensis* population (reviewed in Margos et al., 2019b). Genes that were found under positive selection in genome analysis in the European population included nine chromosomal loci (BG_0142, BG_0275, BG_0298, BG_0327, BG_0359, BG_0368, BG_0706, BG_0748, BG_0827 although the latter showed a high value for recombination) and two loci on lp54 (BGA04, BGA37) (Gatzmann et al., 2015).

Genome analysis of *B. maritima*, a species described from California, revealed a small genome comprised of a chromosome and only seven plasmids, two circular and five linear (Margos et al., 2020). It remains to be investigated whether this is indicative of a narrow host and vector association. In the phylogenetic reconstruction from MLST data, *B. maritima* isolate CA690 was positioned close to the European species *B. afzelii,* whereas in the genomic-based phylogeny it formed a sister clade to *B. chilensis* and occupied a basal position compared to other LB species occurring in North America (Fedorova et al., 2014; Margos et al., 2020) (Figure 4). These data particularly highlight the utility of whole genome analyses for elucidation of phylogenetic relationships.

Lyme Disease *Borrelia* population genomics

Con-specific genomic groups in bacteria: selection or rare recombination?

Population genomics quantifies genome-scale genetic variations within natural populations of a single biological species and aims to understand underlying evolutionary processes including mutation, recombination, genetic drift and natural selection (Li et al., 2008; Ellegren, 2014). Population genomics can be studied through large-scale comparisons of DNA sequences of natural populations, and with the advent of high-throughput DNA sequencing technologies this term is replacing the more commonly used term “population genetics” (Luikart et al., 2003). For bacterial species, an absence of apparent genetic barriers to DNA exchange makes it challenging to analyze genetic diversification within populations as an evolutionary process distinct from genetic divergence between populations (Rocha, 2018). Like some other obligate and vector-borne parasites, however, LB bacteria show strong species-specific geographic distributions

such that genomic variations within and between populations can be productively analyzed (Kurtenbach et al., 2006; Margos et al., 2012; Jacquot et al., 2014; Qiu and Martin, 2014; Seifert et al., 2015; Walter et al., 2016).

Unlike eukaryotes, bacteria can often exchange DNA across large phylogenetic distances. Recombination among con-specific bacterial strains is not negligible, but occurs at a rate that is not as high as in sexual eukaryotes (Smith et al., 1993; Rocha, 2018). The promiscuous nature and intermediate rates of bacterial recombination make the study of genomic variations challenging in practice. One apparent difficulty is the decreased statistical confidence in classification of species and strains based on phylogenetic reconstruction, which assumes an absence of recombination. Special phylogenetic methods are often necessary to either identify and exclude recombination hotspots prior to phylogenetic reconstruction or to estimate recombination rates in conjunction with strain phylogeny (Marttinen et al., 2012; Ansari and Didelot, 2014).

Furthermore, quantifying bacterial recombination rates is critical for identification of genes under adaptive evolution in bacterial pathogens. For example, loci encoding serotype-determining surface antigens often exhibit the highest recombination rates as well as the highest sequence variability in the genomes of pathogenic bacteria, including *E. coli* and LB bacteria (Haven et al., 2011; Bobay et al., 2015). Indeed, recombination is a universal genetic mechanism for acceleration of species adaptation, without which beneficial mutations may not be maintained due to interference with each other ("Hill-Robertson effect") and deleterious mutations cannot be easily purged ("Muller's Ratchet") (Muller, 1964; Hill and Robertson, 1966).

The limited but non-negligible levels of recombination in bacteria result in a mosaic genomic structure, whereby genetic variations are strongly linked across the entire chromosome yet are unlinked at recombination hotspots (Milkman and Bridges, 1990; Smith et al., 1993). As such, a limited number of distinct sequence types - far fewer than one would expect from random recombination but more than one would expect from strict clonality - exist within bacterial populations and are known as "clonal frames" (Milkman and Bridges, 1990). Here we refer to these clonal frames or genome-sequence clusters, representing recognizable evolutionary lineages within a con-specific LB population, as genome-wide linkage groups or "genomic groups" for short. Contrary to the original and long-enduring belief that the presence of genomic groups within a con-specific bacterial population is primarily a reflection of low recombination rates, recombination typically occurs at a higher frequency than *de novo* mutations in bacterial pathogens, supporting the critical role natural selection plays in the maintenance of genomic groups (Smith et al., 1993). Indeed, in the absence of natural selection such as adaptive diversification, a recombination rate on par with the rate of mutation is sufficient to render genomic groups within a bacterial population unrecognizable (Fraser et al., 2007). In the subsequent sections, a review of studies identifying the genomic groups within LB populations based on phylogenetic reconstruction of DNA sequences from across the main chromosome is presented. The role of natural selection targeting two highly variable surface antigen loci (*ospC* and *vls*) in maintaining con-specific genomic groups in LB populations is then examined.

LB populations contain well defined genomic groups

Diverse genomic groups co-exist within LB populations, often infecting a single tick and presumably a single vertebrate host (Rauter and Hartung, 2005; Andersson et al., 2013; Di et al., 2018). Multilocus linkage disequilibrium (nonrandom distribution of genetic alleles) among LB strains was first noted at three genetic loci (*p93*, *fla*, and *ospA*) based on a number of isolates which are now recognized as distinct LB species (Dykhuizen et al., 1993). Subsequently, linkage disequilibrium was observed within natural populations within the same LB species as well, between alleles at two distinct loci (Qiu et al., 2002), at multiple loci (Bunikis et al., 2004; Hanincova et al., 2008; Margos et al., 2008; Qiu et al., 2008; Hanincova et al., 2013) and, most importantly in terms of defining genomic groups, at the whole genome level (Mongodin et al., 2013; Walter et al., 2017; Tyler et al., 2018). It should be noted that genome-wide linkage disequilibrium is particularly strong within LB populations that have experienced recent expansion, such as those in the northeastern USA and southern Canada (Qiu et al., 2002; Margos et al., 2012; Tyler et al., 2018). When sampling is performed over a wider geographic area with a longer history of indigenous populations, such as those in Western and Midwestern US and Europe, decay of linkage disequilibrium is noted especially between plasmid-borne loci (e.g, *ospC*) and

chromosomal loci (e.g., IGS) (Travinsky et al., 2010; Margos et al., 2012; Hanincova et al., 2013).

Con-specific genomic groups within LB populations displaying distinct evolutionary lineages are maintained to a greater extent by natural selection than by a low genetic recombination rate (Brisson and Dykhuizen, 2004; Haven et al., 2011). Indeed, rates of homologous recombination among coexisting *B. burgdorferi* strains in the Northeast US is about three times the rate of point mutations based on a comparison of three co-existing genomes (Qiu et al., 2004). A subsequent simulation-based study showed that diversifying selection at a single surface-antigen locus was sufficient to overcome sequence-homogenizing effects of recombination and maintain the coexistence of within-population genomic groups (Haven et al., 2011). Recombination occurs more frequently within than between LB species in areas where they coexist, perhaps due to vector and host specialization providing more opportunities for strain mixing (Jacquot et al., 2014; Gatzmann et al., 2015). Nevertheless, cross-species horizontal transfers of whole plasmids in nature have been documented among coexisting LB species, including lp5 exchange between *B. mayonii* and *B. burgdorferi* in the Midwest US, lp56 exchange between *B. bissettiae* and *B. burgdorferi* in Western US, lp17 exchange between *B. spielmanii* and *B. afzelii* in Europe, and lp28-9 exchange between *B. burgdorferi* and *B. garinii* or *B. afzelii* in Europe (Casjens et al., 2018).

The genomic diversity of LB species and populations within them has profound public-health and clinical consequences. First, LB species differ in ecological prevalence, human pathogenicity, and clinical manifestations. In the USA, *B. burgdorferi* is the predominant pathogenic LB species, present in over 50% of adult *Ixodes scapularis* ticks in the Midwest and Northeast (Greay et al., 2018; Chauhan et al., 2019; Tokarz et al., 2019). While humans are incidental hosts, *B. burgdorferi* is pathogenic to humans and has caused over 275,000 cases of Lyme disease from 2008-2015 (Schwartz et al., 2017). On the other hand, *B. mayonii,* a more recently discovered LB species in the Midwest USA, has a prevalence of ~2% in ticks and causes acute spirochaetaemia (Pritt et al., 2016; Cross et al., 2018). No documented human disease has been associated with other recognized North American LB species (with the possible exception of *B. bissettiae* DNA in human serum; see Girard et al., 2011), including *B. americana, B. andersonii, B. californiensis, B. carolinensis, B. kurtenbachii, B. lanei* and the most recently named *B. maritima* (Barbour, 2019; Margos et al., 2020). In Eurasia, human infections are caused more frequently by LB species associated with mammalian hosts (*B. afzelii, B. burgdorferi*, *B. bavariensis*, and *B. spielmanii*) than by those associated with bird and lizard hosts (*B. garinii*, *B. lusitaniae*, and *B. valaisiana*) (Stanek and Reiter, 2011; Coipan et al., 2016). Remarkably, *B. bavariensis* is infrequently found in the tick vectors in Europe, but was frequently isolated from human patients (Margos et al., 2013; Springer et al., 2020). For *B. bissettiae*, a single human isolate has been obtained in Germany from a patient without travel history, whereas two human isolates exist for *B. lusitaniae* (Collares-Pereira et al., 2004; de Carvalho et al., 2008; Margos et al., 2016). In Europe, the prevalence of *B. bissettiae* in ticks is extremely low (Coipan et al., 2016) and *B. lusitaniae* is moderately prevalent in countries around the Mediterranean Sea (Baptista et al., 2004). It has been proposed that *B. valaisiana* is non-pathogenic to humans because no single human isolate has been obtained to date (Margos et al., 2017d), although the species is found as frequently in *Ixodes ricinus* ticks (the vector of *Borrelia* species in Europe) as other known human-pathogenic *Borrelia* species (Rauter and Hartung, 2005). It remains an open question whether some *Borrelia* species do not cause human disease because they are rare in the environment or utilize tick vectors that do not bite humans at any frequency.

Second, genomic groups within *B. burgdorferi* vary in disease propensities. At least 18 *B. burgdorferi* genomic groups, corresponding to major sequence variations at *ospC* (allelic types A-O, T, and U) co-segregate with chromosome-based phylogeny in northeast USA populations (Wang et al., 1999c; Barbour and Travinsky, 2010; Di et al., 2018). Initially, four groups (*ospC* types A, B, I, and K) were found to be strongly associated with disseminated (i.e. extracutaneous) Lyme disease (Seinost et al., 1999). This was confirmed and expanded to include ospC type H (Wormser et al., 2008). A subsequent MLST-based study demonstrated that there are *B. burgdorferi* sequence types (ST) with different capacities to cause disseminated infection in humans (Hanincova et al., 2013). Considering that all *ospC* groups vary in human virulence, invasiveness of different *ospC* type clonal groups has been ranked by scaling the frequency of their presence in blood and cerebrospinal fluid cultures by their natural prevalence in ticks (Figure 5) (Dykhuizen et al.,

2008). Decreasing levels of *ospC* type diversity from tick to skin, blood, synovial fluid and cerebrospinal fluid as biological niches have been further confirmed by (Brisson et al., 2011).

Third, population genomics holds the promise of unbiased identification of molecular mechanisms underlying variations in clinical manifestations among LB strains (Seifert et al., 2015; Tufts et al., 2019). In the following sections, patterns of molecular polymorphisms at two highly variable surface antigen loci, *ospC* and *vlsE*, are summarized and the molecular and evolutionary mechanisms involved in sustaining LB prevalence in nature and pathogenicity in humans are discussed.

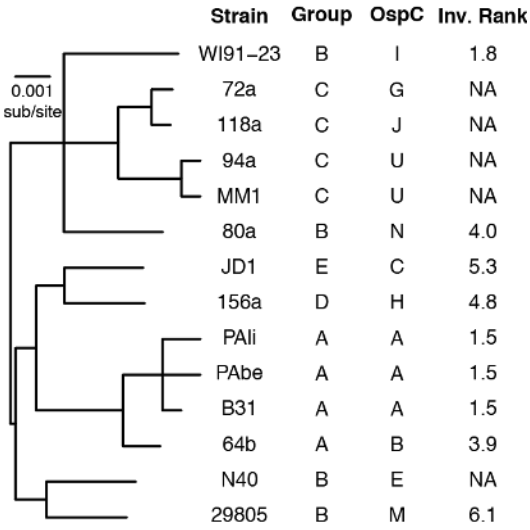

Figure 5. Genomic groups in *B. burgdorferi* from Northeast and Midwest United States. A phylogenetic tree of published complete *B. burgdorferi* genomes (n=14) (Mongodin et al., 2013) (left) was reconstructed based on nucleotide sequences at 20 loci, chosen randomly based on their approximate equal spacing on the main chromosome (*BB_0057, BB_0178, BB_0198, BB_0199, BB_0228, BB_0243, BB_0251, BB_0253, BB_0259, BB_0271, BB_0328, BB_0344, BB_0369, BB_0388, BB_0435, BB_0512, BB_0545, BB_0548, BB_0622, and BB_0809*). Sequences were translated, codon-aligned, and concatenated using MUSCLE (Edgar, 2004) and BpWrapper (Hernandez et al., 2018). An approximate maximum likelihood tree was inferred using FastTree (Price et al., 2010). All branches shown are supported by >0.9 bootstrap values. The tree was mid-point rooted and plotted using the APE package in R (Paradis et al., 2004). Phyogenetic groups (2nd column), major OspC allelic types (3rd column), and an invasiveness rank (4th column, "NA" for not available) were marked for each strain (Dykhuizen et al., 2008; Casjens et al., 2017). The nearly perfect linkage disequilibrium (and, in fact, the one-to-one correspondence) between major-group OspC alleles and distinct genomic groups suggests the critical role of OspC in maintaining within-species genomic diversity.

ospC variability: a serotype- and lineage-determinant of Lyme disease Borrelia

Two main selective forces have been proposed as evolutionary mechanisms that maintain the even and stable coexistence of diverse con-specific genomic groups within local LB populations despite the presence of recombination (Brisson et al., 2012; Andersson et al., 2013; Qiu and Martin, 2014; Seifert et al., 2015). First, the various LB genomic groups may prefer different vertebrate host species (Brisson and Dykhuizen, 2004; Mechai et al., 2016; Tufts et al., 2019). Second, diverse LB genomic groups may be maintained by negative frequency-dependent selection by which strains displaying rare antigenic variants have higher fitness than strains carrying more common antigen alleles (Haven et al., 2011; Durand et al., 2015; Durand et al., 2017a). The *ospC* locus encodes the major serotype determinant in LB (Wilske et al., 1993; Barbour and Travinsky, 2010). As noted earlier, it is also the most variable non-paralogous gene in the LB genome (Mongodin et al., 2013). The nearly one-to-one correspondence between the sequence allele types of OspC and *B. burgdorferi* genomic groups in the Northeast USA supports the mechanism of immune-mediated balancing selection for maintaining diverse coexisting genomic groups (Figures 3 and 5) (Attie et al., 2007; Hanincova et al., 2008); *ospC* alleles are highly divergent from one another (an average of ~15% pairwise differences in amino acid sequences of different OspC proteins) (Wang et al., 1999c). This pattern of genetic variation at immune-dominant loci is expected from, and strongly supports, the predominance of negative frequency-dependent selection in shaping microbial populations in nature, including those of LB (Haven et al., 2011; Andersson et al., 2013; Jacquot et al., 2014; Durand et al., 2015; Strandh and Raberg, 2015).

Critical ecological and evolutionary roles of *ospC* in shaping population structure of natural *B. burgdorferi* populations from the Northeast US are manifested, first, by a nearly complete linkage disequilibrium between sequence variations at *ospC* and genetic variations elsewhere in the genome (Figures 3 and 5) (Wang et al., 1999c; Hanincova et al., 2008; Wormser et al., 2008; Haven et al., 2011). The linkage disequilibrium is high, but less than complete at a larger geographic scale in the Western USA

(Barbour and Travinsky, 2010; Travinsky et al., 2010; Tyler et al., 2018). On the other hand, population genomic studies on *B. afzelii, B. garinii* and *B. burgdorferi* in Europe showed lower level of linkage disequilibrium between *ospC* and chromosomal loci, likely caused by linkage decay in older populations (Jacquot et al., 2014). Second, as noted above, *ospC* is the most polymorphic non-paralogous locus in the genome, driven by a high rate of recombination that spills over to its genomic neighbors (Figure 6). Third, *B. burgdorferi* strains with major sequence variations at *ospC* are rather evenly represented in local populations among infected *Ixodes* ticks, consistent with expectations of negative frequency-dependent selection (Rannala et al., 2000; Qiu et al., 2002; Di et al., 2018). However, this was not found in a population of *B. afzelii* sampled at one habitat in Switzerland over a period of 11 years where one OspC major group dominated (Durand et al., 2017b).

A closer look at *ospC* sequence variability among 10 LB species reveals non-uniform selective pressures among gene regions associated with functional and structural domains (Figure 6). The first 18 codons (nucleotide positions 1-54) encode a signal peptide and are conserved not only for amino acid sequences (low evolution rates at 1st and 2nd codon positions), but also for synonymous nucleotides (low rates at the 3rd codon positions) (Wang et al., 1999c). Following the signal peptide, the N-terminal membrane-tethering motif (codons 19-30, or nucleotide positions 55-123) lacks sequence conservation (Kumru et al., 2011). The OspC molecules form dimeric α-helical bundles that are covalently anchored on the outer membrane (Kumaran et al., 2001; Eicken et al., 2002; Kumru et al., 2011). The α-helices (e.g., α1, α2, and α3) tend to be conserved at 1st and 2nd codon positions but not at the 3rd codon position, indicating influence of purifying selection. The surface-exposed regions (e.g, β1 and β2) tend to evolve at nearly the same rates among the three codon positions, indicating influence of positive selection driven by antagonistic co-evolution with host immune defense (Figure 6). LB strains differing in *ospC* alleles are immunologically distinct and wildlife hosts are frequently infected by multiple LB strains carrying distinct *ospC* alleles (Andersson et al., 2013; States et al., 2014; Bhatia et al., 2018). Serotype-determining antibody-binding epitopes have been mapped to hypervariable regions, making it challenging to design broadly effective OspC-based diagnostics or vaccines (Buckles et al., 2006; Ivanova et al., 2009; Baum et al., 2013; Izac et al., 2019). An OspC-based vaccine designed with a linear concatenation of eight serotype-specific epitopes has been approved for veterinary use against Lyme disease in dogs (Earnhart and Marconi, 2007; Izac and Marconi, 2019). The potential of conserved regions of *ospC* (Figure 6) as immune-protective vaccinogens against a broad array of coexisting LB strains within local endemic regions of Lyme disease remains to be explored (Ivanova et al., 2009; Baum et al., 2013).

vls variability: fast sequence and duplicative evolution

Lyme disease *Borrelia* express numerous lipoproteins in a highly regulated fashion during its enzootic life cycle (Radolf et al., 2012) (see Radolf and Samuels, 2021). For example, OspA is highly expressed during tick infection and OspC is expressed during initial mammalian host invasion (Tilly et al., 2006; Mulay et al., 2009; Caimano et al., 2019). In contrast, the vls (*vmp* [variable major protein]-like sequences) family of lipoproteins is expressed during persistent infection of vertebrate hosts (Tilly et al., 2013; Norris, 2014; Chaconas et al., 2020), a key mechanism of antigenic variation in LB underlining its evasion host adaptive immunity. Comparative sequence analysis revealed distinct patterns of genomic variation, perhaps reflective their distinct biological functions. For example, OspA sequences vary greatly between LB species, but show minimal variation within species. In contrast, OspC sequences are substantially different both between and within LB species. The *vls* locus varies in both gene sequence, as well as in copy number, even within the same LB species (Mongodin et al., 2013; Tufts et al., 2019).

The *vls* system of antigenic variation in LB was discovered by sequence homology to the *vlp*/*vsp* (variable large/small protein) system in relapsing fever (RF) *Borrelia* (Zhang et al., 1997; Norris, 2014; Barbour, 2016) (See Radolf and Samuels, 2021). This system consists of a plasmid-borne expression locus, *vlsE*, for the variable surface antigen protein and an adjacent set of typically 10-20 tandem unexpressed "*vls"* cassettes containing variants of the *vlsE* sequence (Figure 7). Genetic variation at the *vlsE* expression locus is mediated by frequent non-homologous recombination between the silent *vls* cassettes and the expression locus (Coutte et al., 2009; Lin et al., 2009; Norris, 2014; Chaconas et al., 2020). Similar to the *vmp* system in RF *Borrelia*, capsule variability in *Neisseria* and the Variant Surface Glycoprotein (VSG) system in *Trypanosoma*,

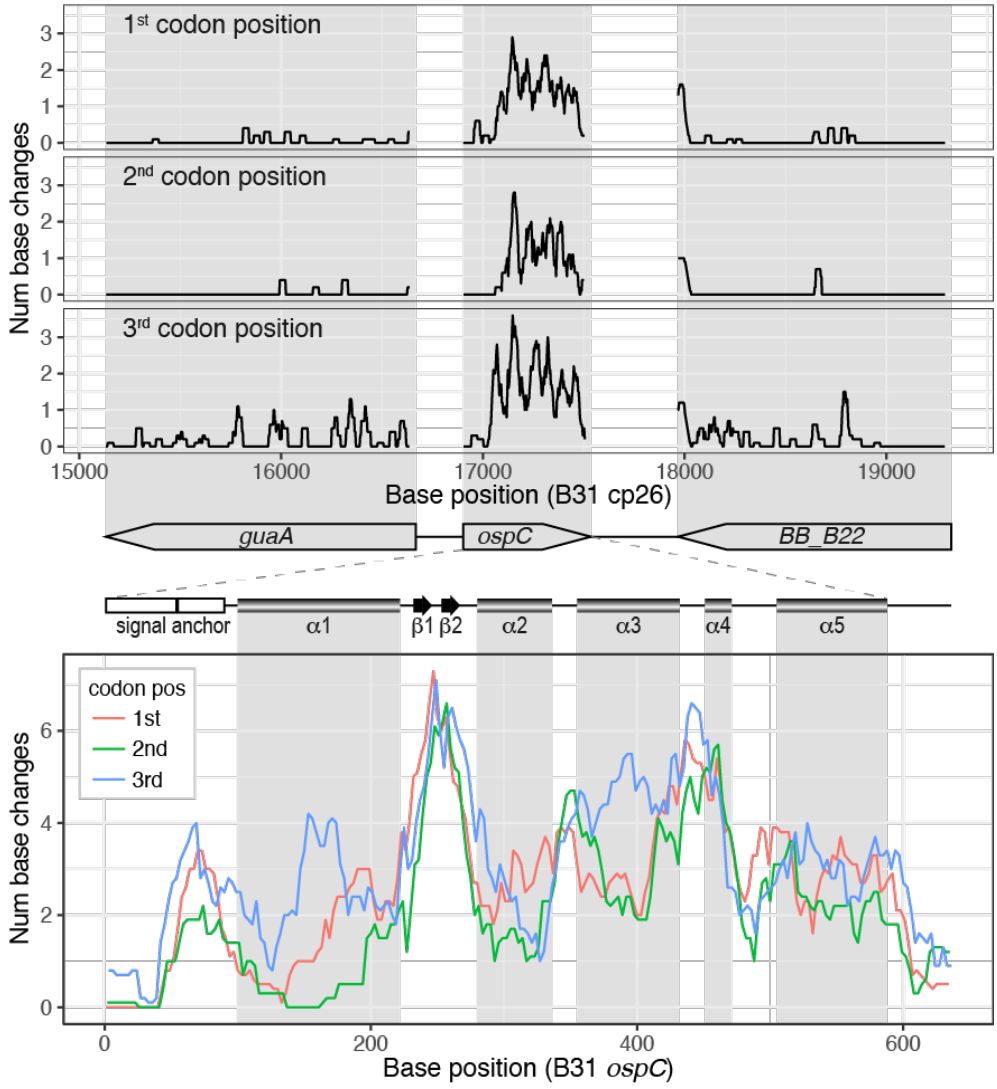

Figure 6. Association of *ospC* sequence variability with its molecular function and structure. (Top 3 Panels) Within-population evolutionary rates at *ospC* and its neighboring loci (*BB_B18* [*guaA*] and *BB_B22*). Genomic sequences at the three loci from 14 *B. burgdorferi* strains from Northeast USA (same strains as in Figure 5) were aligned by preserving codon positions using BIOALN of the BpWrapper sequence utilities (Hernandez et al., 2018). X-axis shows genomic coordinates of the B31 cp26 plasmid and a genome map of the region. Y-axis shows the number of nucleotide substitutions at the three nucleotide positions within each codon, calculated with the program DNACOMP of the PHYLIP package (Felsenstein, 1989). (Bottom panel) Evolutionary rates at the *ospC* locus among 32 *LB* strains of 10 LB species. The species include *B. burgdorferi* (16 strains; 118a, 156a, 29805, 64b, 72a, 94a, B31, BOL26, JD1, MM1, N40, PAbe, PAli, ZS7, CA-11-2A, and WI91-23), *B. spielmanii* (A14S), B. afzelii (ACA-1, K78, Tom3107, and PKo), *B.bavariensis* (BgVir and PBi), *B. bissettiae* (DN127), *B. garinii* (Far04 and PBr), *B. mayonii* (MN14-1420 and MN14-1539), *B. finlandensis* (SV1), *B. valaisiana* (Tom4006 and VS116), and *B. chinensis* (VA1). The positions of two functional domains (the signal peptide and the outer-membrane anchoring domain) and the secondary structural elements of OspC (Eicken et al., 2002; Kumru et al., 2011) are indicated above the panel.

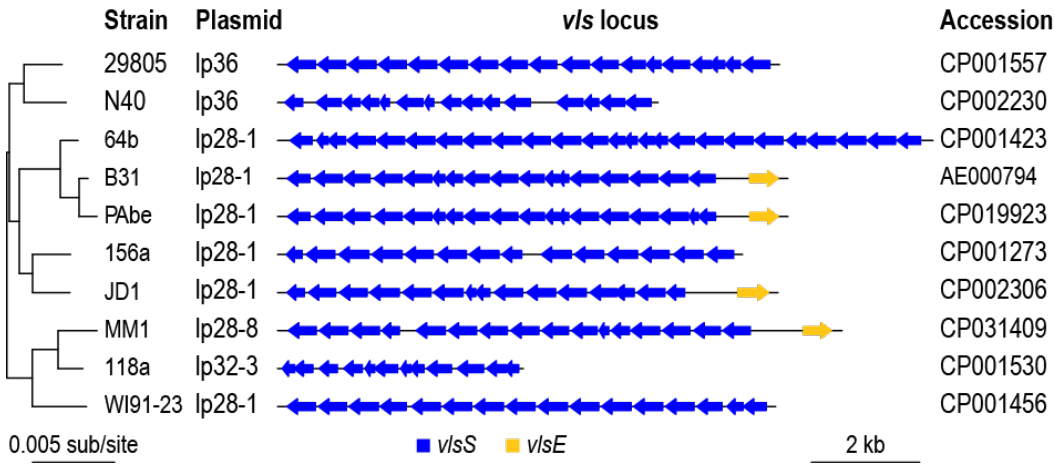

Figure 7. Copy-number and sequence variations at the *vls* locus. Genomic map of the *vls* locus in 10 *B. burgdorferi* strains. MLST phylogeny (left) was derived from Figure 5. Individual cassette sequences were originally identified manually (Casjens et al., 2017) and re-derived here by using HMMER (Potter et al., 2018) with B31 *vls* sequences (Zhang et al., 1997) as queries. *vlsE* expression locus is indicated in yellow. Note that the absence of *vlsE*, located in the telomeric regions of linear plasmids, in most strains is mostly likely due to incomplete genome assemblies.

generation of random and abundant genetic variants mediated by intra-genomic recombination serve as a population-level adaptive mechanism (Vink et al., 2012; Graves et al., 2013; Chaconas et al., 2020).

The *vls* system in LB bacteria varies in both paralogous copy numbers and in cassette sequences (Graves et al., 2013) (Figure 8). Greater sequence similarity of *vls* cassettes within than between the closely related strains (Figure 8) could be due to either concerted evolution or rapid duplication and loss of gene copies (Nei and Rooney, 2005). Concerted evolution is the mechanism by which paralogous gene copies (e.g., ribosomal RNA genes) in a species evolve non-independently from each other and converge in DNA sequences by intra-genomic recombination (Elder and Turner, 1995). The gene birth-death hypothesis, on the other hand, proposes a mechanism by which sequences converge by independent gene duplications and losses while mutations accumulate, also independently, among paralogous gene copies (Nei and Rooney, 2005). The birth-death mechanism appears to be a more parsimonious model that does not require the assumption of the additional process of intra-genomic recombination. Gene tree analysis of *vls* cassettes from sequenced *B. burgdorferi* genomes supports the birth-death model of *vls* cassette evolution. First, the cassette copy number varies among the closely related *B. burgdorferi* strains, suggesting frequent gene duplications and losses (Figure 7).

Second, cassette sequences have diverged greatly within the same genome, suggesting that mutations accumulate independently among cassette copies, with a lack of (or infrequent) intra-genomic recombination (Figure 8). Third, *vls* cassettes among the most recently diverged sister-group *B. burgdorferi* strains do not sort into strain-specific sequence groups, as one would expect from concerted evolution. Examples of recently diverged sister groups include B31/64b/PAbe, 72a/118a, and 94a/MM1 (Figure 8). Fourth, although distantly related genomic groups tend to show strain-specific *vls* cassette sequences, exceptions exist. For example, SNP Groups D (156a) and E (JD1) are a pair of distantly related sister groups (Figure 5), but their vis sequences fail to evolve into distinct strain-specific lineages as one would expect from concerted

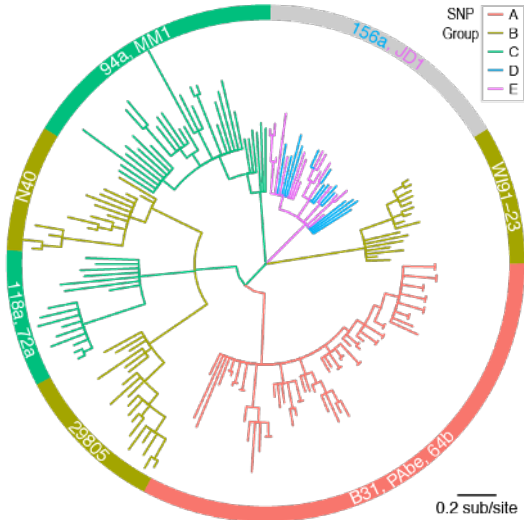

Figure 8. An unrooted maximum likelihood tree of *vls* cassette protein sequences. Genomes analyzed are those in Figure 3 plus 72a and 94a (Figure 5). Tree shows fast divergence of *vls* cassette sequences within and among con-specific genomic groups. Each tip represents a single cassette sequence. All branches are supported by a bootstrap value of 0.9 or greater, as determined by FastTree (Price et al., 2010). Clades (i.e., monophyletic groups) are colored by "SNP Group", referring to phylogenetic groups defined by chromosomal SNPs (Mongodin et al., 2013). This tree is an update of a previously published study using a similar set of genomes (Graves et al., 2013).

evolution within each strain (Figure 8). It is possible that natural selection has kept the *vls* cassettes in these two strains from diverging from each other. Further comparative analysis of *vls* sequences among closely related genomes would enable quantification of precise rates of cassette duplication, gene loss, mutation, intra-genomic recombination, and natural selection in this key antigenic variation system responsible for LB persistence in hosts.

Deep homology between diverse mechanisms of antigenic variation in Borrelia

The antigenic variation systems in LB and the RF *Borrelia* apparently descended from a common ancestral molecular system (Norris, 2006). While the two systems share characteristics such as sequence homology, structural resemblance, plasmid-borne locations, and antigenic variation mediated by recombination between archival and expressed copies, they have diverged in genomic organization and molecular mechanisms of recombination. For example, the archival copies in RF *Borrelia* are dispersed among plasmids, whereas the *vls* sequences are arranged in tandem arrays of cassettes adjacent to *vlsE* on a single linear plasmid (Norris, 2006; Barbour, 2016; Casjens et al., 2017; Casjens et al., 2018) (Figure 7).

Furthermore, sequence and structural analyses revealed a deep molecular homology between *vls* and *ospC* systems of antigenic variation. VlsE, OspC, and Vsp share a similar, predominantly α-helical molecular structure (Kumaran et al., 2001; Eicken et al., 2002; Kumru et al., 2011). Genetic complementation experiments showed interchangeable roles for OspC and VlsE in maintaining infection in immune-deficient mice, but they have divergent molecular functions in early and persistent infections, respectively, in natural hosts (Tilly et al., 2013). Primers targeting the upstream and downstream conserved regions of the *ospC* locus could also be successfully used to amplify a full-length *vsp* gene in *B. miyamotoi* (an RF *Borrelia*), hinting at shared gene regulatory mechanisms (Di et al., 2018). In sum, the three main molecular systems of antigenic variations in *Borrelia*, including the *ospC* and *vls* systems in LB and the *vsp/vlp* system in RF *Borrelia,* likely share a common evolutionary origin. Comparative analyses among closely related LB genomes would further illuminate the origin, diversification, and functions of these highly adaptable molecular systems that are keys to LB survival as an obligate tick-borne parasite of vertebrates.

Population genomics is indispensable for reconstructing biogeographic histories of natural LB populations (Margos et al., 2012; Gatzmann et al., 2015; Walter et al., 2017; Tyler et al., 2018). In addition, population analyses of con-specific genomic groups are essential for uncovering mechanisms of LB evolution in nature including recombination, genome divergence driven by adaptation to host and vector, and genome diversification driven by host immunity (Kurtenbach et al., 2006; Qiu and Martin, 2014; Seifert et al., 2015; Becker et al., 2016). Furthermore, population genomics informs the investigation of key genes and genetic mechanisms (e.g., *ospC* and *vls* cassettes) associated with invasion of, and persistence in, hosts including humans (Mongodin et al., 2013; Jacquot et al., 2014; Seifert et al., 2015).

Bacteriophages of the Lyme disease *Borrelia*

The ϕBB-1 bacteriophage

A number of studies in which supernatants of *B. burgdorferi* cultures were examined by electron

microscopy have reported the presence of tailed bacteriophage-like particles (Hayes et al., 1983; Barbour and Hayes, 1986; Neubert et al., 1993; Schaller and Neubert, 1994; Eggers and Samuels, 1999; Eggers et al., 2000). Sequence analysis also recognized that several *B. burgdorferi* plasmids could be prophages (see below). These observations have been borne out by more directed studies in strain CA-11.2A, where the supernatant yielded phage-like particles with symmetrical heads and apparently contractile tails (Eggers and Samuels, 1999; Eggers et al., 2000; 2001a). These particles were purified and shown to contain the cell's complement of cp32 DNA molecules in a linearized form (Eggers and Samuels, 1999) (Figure 9). This phage has been designated φBB-1.

The presence of cp32 DNA in the particles strongly suggests that these plasmids are prophage. With the exception of *B. garinii* strain Far04, cp32-like plasmids have been found in all other LB species that have been analyzed (see plasmid section above); however ϕBB-1 or similar phage particles have been detected in the supernatants of only a small number of *B. burgdorferi* strains and in one *B. bissettiae* strain following spontaneous release or after treatment with sub-lethal concentrations of 1-methyl-3-nitroso-nitroguanidine (MNNG), mitomycin C, or ethanol (Eggers and Samuels, 1999; Eggers et al., 2000; 2016). φBB-1 has not yet been propagated by lytic growth in the laboratory, and the reason(s) for this failure are not yet understood. Possibilities include: i) a high propensity to form a stable lysogen; ii) possible genetic defects in the machinery required for prophage induction; iii) a technical inability to generate a 'lawn' of *B. burgdorferi* cells in which to detect viral plaques; iv) a lack of sensitivity in the methods needed to detect phage from a small percentage of cells that are undergoing lysis; and/or (v) the absence *in vitro* of unique signals that induce fully lytic growth, such as might be found in either the tick vector or mammalian host.

Horizontal gene transfer (HGT) in *B. burgdorferi* appears to primarily involve the cp32s and other plasmid-borne loci consisting of fragments <2kb (Stevenson and Miller, 2003; Qiu et al., 2004; Schulte-Spechtel et al., 2006; Barbour and Travinsky, 2010; Haven et al., 2011; Brisson et al., 2012; 2013; Hanincova et al., 2013; Mongodin et al., 2013). The demonstration that ϕBB-1 can transduce a cp32 between two cells of the same *B. burgdorferi* strain and between different *B. burgdorferi* strains, suggests the possibility that φBB-1 is capable of enhancing the genetic diversity observed in the complement of cp32 plasmids found in different *B. burgdorferi* isolates (Casjens et al., 1997b; Caimano et al., 2000; Iyer et al., 2003; Stevenson and Miller, 2003; Casjens et al., 2018). Furthermore, the observation that ϕBB-1 can package small heterologous *E. coli/B. burgdorferi* shuttle vectors and transduce them between different *B. burgdorferi* strains (Eggers et al., 2016), suggests that this phage may act in generalized transduction, as well. With many bacteriophage, a small percentage (<1%) of the capsids are filled with bacterial host DNA rather than phage genomic DNA (Stanton, 2007; Thierauf et al., 2009; Muniesa et al., 2011; Penades et al., 2015). This mispackaging is thought to likely occur, at least in part, when the phage packaging mechanism recognizes packaging sequence homologs throughout the bacterial genome (Thierauf et al., 2009). Although the role of generalized transduction in HGT in naturally-occurring spirochetes is not yet well understood, "directed mispackaging" could be one factor as to why some non-phage genes, such as *ospC*, appear to be more likely than others to be subject to HGT and selective retainment in *B. burgdorferi* (Dykhuizen and Baranton, 2001; Brisson and Dykhuizen, 2004; Qiu et al., 2004; Attie et al., 2007).

cp32 as prophage

The cp32s share organizational properties with the genomes of other tailed bacteriophages. Comparative genomic studies of double-stranded DNA bacteriophage groups, such as the λ-like family of phages, have shown that their genomes are genetic mosaics in which each has an overall conserved gene order, with occasional non-homologous genes inserted into the genomes of different phages (Casjens et al., 1992; Hendrix, 2002; Casjens, 2005; Hendrix and Casjens, 2006; Casjens, 2008); this pattern is similar to that observed among the cp32s (Casjens et al., 1997b; Caimano et al., 2000; Casjens et al., 2000; Stevenson et al., 2000b; Stevenson and Miller, 2003). The 32-kb circular plasmids can have long stretches of nearly identical DNA, broken by three regions of significant variability; these regions correspond to the partitioning/compatibility regions that allow several cp32 isoforms to be stably maintained in a single isolate (see plasmid section above), and two regions that encode multiple families of differentially-expressed OspE/OspF/Elp and Mlp lipoproteins (Marconi et al., 1996; Porcella et al., 1996;

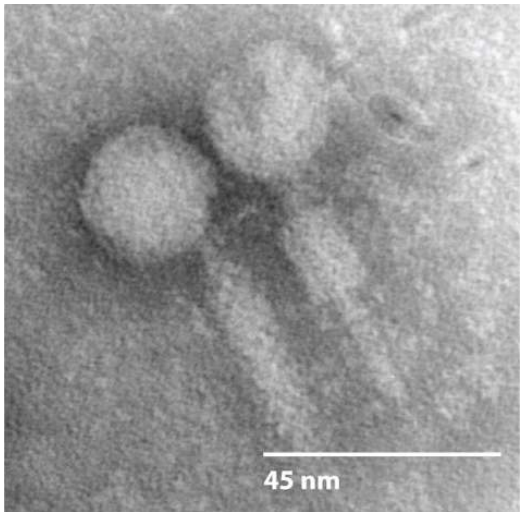

Figure 9. *B. burgdorferi* bacteriophage φBB-1 virions. Phage particles from the supernatant of an MNNG-treated culture of *B. burgdorferi* strain CA-11.2A were negatively stained for transmission electron microscopy with phosphotungstic acid; bar is 45 nm. Reproduced from Eggers et. al (Eggers et al., 2001b) with permission of the publisher.

Stevenson et al., 1996; Zuckert and Meyer, 1996; Casjens et al., 1997b; Akins et al., 1999; Yang et al., 1999; Caimano et al., 2000; Casjens et al., 2000; Porcella et al., 2000; Yang et al., 2003; Casjens et al., 2012).

cp32 phage-like genes

Two genes that encode proteins with homology to phage virion head assembly proteins are present on the cp32s. Genes L43 (and paralogues on other cp32s; L43 is on cp32-8 in *B. burgdorferi* strain B31) and L01 (and paralogues) encode a putative phage large subunit terminase and a portal protein, respectively (Casjens et al., 2000; Eggers et al., 2000; 2001a). Terminases form the ATP-dependent motors that translocate phage DNA into virions, and the portal proteins form the hole through which the DNA is packaged into the preformed phage heads (Catalano, 2005; Prevelige and Cortines, 2018). These two proteins are the most highly conserved tailed-phage proteins, and no homologues have been found that have different functions (Casjens, 2003). The presence of these genes strongly supports the notion that cp32 plasmids are in fact prophages. The positions of genes L43 and L01 (and cp32 paralogs) are consistent with the typical arrangement of virion head assembly genes in other dsDNA temperate bacteriophages, where the large terminase and portal are often the second and third genes, respectively, in the phage's late operon (see below) (Casjens, 2003; 2005; 2008). In contrast to L43 and L01, a majority of the predicted proteins encoded on the cp32s have no homologues in current databases outside of *B. burgdorferi*, while the structural proteins for φBB-1 have not yet been identified. This lack of homology may not be surprising, as virion structural proteins are known to be extremely variable, and many phage structural proteins have no recognizable homologues, even among proteins with identical functions in other phages (Casjens et al., 1992; Hendrix et al., 1999; 2000; Casjens, 2003; 2005). The known dsDNA tailed phages with genomes in the 20-60 kbp range typically have a single large operon that is expressed late in lytic infection and contains the 20-30 genes required for virion assembly and lysis (Casjens, 2005; 2008). The putative terminase and portal protein genes on the cp32s are followed by 25-30 contiguous genes, which are transcribed in the same direction, have no homologues outside the cp32s, and remain uncharacterized (Casjens et al., 2000). We previously have proposed that this region is likely the "late operon" of φBB-1 (Eggers et al., 2000; 2001a).

As part of their survival strategy, bacteriophages escape from their host cell by degrading the cell wall and lysing the cell after a cycle of lytic growth. To do this, all tailed bacteriophages encode an endolysin that degrades the cell wall and a holin whose permeabilization of the membrane allows the endolysin access to the cell wall thus controlling the timing of cell lysis (Young and Blasi, 1995). The genes that encode these proteins are sufficiently variable that they are often not recognizable in nucleotide sequence (Wang et al., 2000; Young, 2002). The function of these lysis genes necessitates that their expression be tightly controlled (holins in particular are lethal to bacteria (Young and Blasi, 1995), and they most often are clustered at the beginning or the end of the late operon (Casjens et al., 1992; Ford et al., 1998). On the cp32s, *blyA* (L23 and its paralogs) lies near the distal end of the putative phage late operon; the *blyA* protein product has membrane-disruptive activity and has been shown to mediate the release of both the phage lambda encoded endolysin (Damman et al., 2000) and an *E. coli* host-encoded cytolysin (Ludwig et al., 2007) in surrogate systems. These results suggest

that *blyA* may encode a holin, although no cp32-encoded endolysin has yet been identified.

Late operons are not expressed from the uninduced prophage DNA and are only expressed during lytic growth of the phage; thus, the above model predicts that these genes will only be expressed during "inducing" conditions. A second, slightly weaker, prediction of the model is that these genes will in fact be co-transcribed as an operon. Zhang and Marconi (2005) used RT-PCR across the gene boundaries to demonstrate that the thirty genes from BBL42 to BBL28 (using cp32-8 as a typical cp32) are in fact co-transcribed after *B. burgdorferi* cells are exposed to the φBB-1-inducing agent MNNG, and that these genes are apparently not expressed in uninduced cells. Other DNA array transcription studies are largely in agreement that most of these genes are expressed minimally, if at all, in the absence of induction (Ojaimi et al., 2003; Anderton et al., 2004; Tokarz et al., 2004). The expression of these genes is influenced by the *B. burgdorferi* stringent response regulator, Rel_{Bbu}; a *B. burgdorferi* strain in which this protein was inactivated exhibited a marked increase in their expression (Drecktrah et al., 2015). These data are completely consistent with the proposal that this block of genes comprise a late phage operon. Curiously, the expression pattern of *blyA* is an exception to this rule and not what is expected for a lethal holin. An increase of the *blyA* transcript, in the absence of a concomitant increase in the expression of other genes within the late phage operon, has been observed with an increase in temperature (Ojaimi et al., 2003), after treatment with a bactericidal antibody (Anderton et al., 2004) and in the mammalian host (Zhang and Marconi, 2005). Thus, the *blyA* (and the adjacent *blyB*) gene appears to be under different regulatory control from the rest of the putative cp32 late operon genes. Consequently, a final determination as to whether BlyA is in fact a holin awaits further experimentation. Similarly, the cp32 *bdr* genes (R27 in cp32-4 and paralogs), the most promoter distal member of this gene cluster, appear to be expressed under non-inducing conditions and so may not be part of a late operon (Zuckert et al., 1999; Roberts et al., 2000).

There also are cp32 genes outside of the putative late operon that have homology to phage-related genes; these include those encoding the PFam161, PFam165, and PFam96 proteins. The *pfam96* gene is predicted to be a recombinase. The genes for PFam161 and PFam165 encode a single-stranded DNA binding protein and a DNA nuclease, respectively, and were renamed *ssbP* and *nucP* (Chenail et al., 2012). The products of these genes could play a role in the circularization of the linear φBB-1 genome into cp32 or in phage-mediated recombination. Tokarz et al. (2004) observed that members of all three of these PFams are upregulated in response to blood, a stimulus that *B. burgdorferi* encounters during the blood meal within the tick vector or during infection of the mammalian host. Chenail, et al. (2012) demonstrated that *ssbP* and *nucP* are up-regulated by the transcription factor, BpaB, which also represses gene expression from the cp32 *ospE/ospF/elp* loci (see below) and is itself encoded on each cp32 within the region responsible for plasmid replication and maintenance. Chromatin immunoprecipitation using an antibody specific for BpaB revealed other cp32 binding sites for this transcription factor, including several within the putative late phage operon near the genes encoding the putative terminase and portal protein genes and one near the *blyA* (holin) gene. The coordinated regulation of the "early" and "late" genes of many better studied plasmid prophages remains rather poorly understood (Ravin et al., 2000; Lobocka et al., 2004; Lehnherr, 2006), so an understanding of these findings in a larger context is difficult at this time.

Each cp32 also encodes a set of genes that are involved in their replication and stable maintenance, including the PFam57, PFam32, and PFam49 genes (see plasmid section above). PFam57 and the untranslated region directly upstream were found to be essential and sufficient for replication of the prophage plasmid (Eggers et al., 2002). The PFam32 paralogs have significant homology to the partitioning gene, *parA* (Zuckert and Meyer, 1996; Casjens et al., 1997b, 2000; Stevenson et al., 2000a). Homologues of *parA* are found on a number of well-studied low copy number plasmids, including two *E. coli* prophages, P1 and N15, which exist as autonomously replicating plasmids (Austin and Abeles, 1985; Grigoriev and Lobocka, 2001; Surtees and Funnell, 2003; Lobocka et al., 2004). Finally, the PFam49 genes found downstream of the PFam32 paralogs encode the transcription factor and predicted ParB homolog, BpaB (Stevenson et al., 1996; Stevenson et al., 2000b; Eggers et al., 2002; Burns et al., 2010; Jutras et al., 2012).

Lysogenic conversion

Most prophages alter their hosts, often by modifying the bacterial surface, through expression of their

"host conversion" genes (Cheetham et al., 1995; Casjens and Hendrix, 2002; Canchaya et al., 2003; Casjens, 2003; Hendrix and Casjens, 2006; Dion et al., 2020), and the cp32s appear to be no exception. The cp32s express several families of proteins that have been well-studied. These include the Mlp, Bdr, Rev, BapA and OspE/OspF/Elp proteins (Gilmore and Mbow, 1998; Skare et al., 1999; Yang et al., 1999; Zuckert et al., 1999; Porcella et al., 2000; Bauer et al., 2001; Miller and Stevenson, 2003; Zhang et al., 2005; Stevenson et al., 2006; Brissette et al., 2008; Brissette et al., 2009b; Kenedy and Akins, 2011; Lin et al., 2015). All these are membrane proteins, many demonstrate increased expression during the bloodmeal, and most have been shown to be on the cell surface. While the functional roles for many of these gene products are not known, an exception is OspE and related proteins that are able to bind complement factor H and so appear to play a role in the ability of *B. burgdorferi* to evade the host complement system (Hellwage et al., 2001; Alitalo et al., 2002; Stevenson et al., 2002; Metts et al., 2003; McDowell et al., 2004; Miller and Stevenson, 2004; Hovis et al., 2006; Kenedy and Akins, 2011). Additionally, OspF and related proteins may act as adhesins during *B. burgdorferi* infection and have been shown to bind to host cell heparin sulfate glycosaminoglycans on glial epithelial cells (Antonara et al., 2007; Antonara et al., 2011; Lin et al., 2015). Thus, these genes appear to be involved in the interactions between the pathogenic bacterium and its host, and so can be considered lysogenic conversion genes.

A multiplicity of prophages

In many bacterial systems, prophages have evolved ways to enhance the resistance of their bacterial hosts to superinfection by similar phages that might represent a threat (i.e., cause cell death) (Bondy-Denomy et al., 2016). φBB-1 represents an unusual group of bacteriophages, in that there can be as many as nine quite similar prophages resident in a single cell. Although the PFam32 proteins play a role in determining the maintenance and compatibility of prophage plasmids (see plasmid section above), it is not clear what role, if any, these proteins have in the mechanisms governing superinfection. Furthermore, Eggers, *et al*. (2016) reported differences in the efficiency with which ϕBB-1 transduces DNA packaged from one *B. burgdorferi* strain into either the same strain or into other *B. burgdorferi* strains. These data suggest that differences in surface receptors, restriction modification systems (Lawrenz et al., 2002; Kawabata et al., 2004; Rego et al., 2011), resident plasmid complements, or other as-yet-unidentified factors may influence the initial establishment of a new prophage. The entire complement of cp32s within a given isolate is represented in the extracellular supernatant after prophage induction and are all presumably packaged individually within φBB-1 capsids (Eggers et al., 2001b). Additionally, the putative structural and functional phage genes on different cp32s are highly conserved; thus, we anticipate that all cp32s might well be packaged within virtually identical capsids, whether or not all individual cp32s are still fully functional prophages. For example, two B31 plasmids, cp32-4 and cp32-9, have frameshift mutations in one and seven putative late phage operon genes, respectively (Casjens et al., 2000; Eggers et al., 2000); however we have observed, that cp32-4 and/or cp32-9 (the two plasmids were indistinguishable by the methods used) are packaged in abundance within φBB-1 capsids released from *B. burgdorferi* strain B31 (C. H. Eggers and D. S. Samuels, unpublished).

Several other *B. burgdorferi* plasmids (for example, cp9, cp18, lp54, lp56 and lp28-2) contain DNA that is homologous to virion structural protein genes and so may be phage-related. Cp9 plasmids in *B. burgdorferi* and other LB species appear to have evolved from cp32 plasmids by a series of inversion and deletion mutations (Champion et al., 1994; Dunn et al., 1994; Stevenson et al., 2000b). Similarly, cp18 plasmids present in strains *B. burgdorferi* 297 and N40 appear to be defective prophages that arose from deletions of the putative late phage operons on cp32s (Stevenson et al., 1997; Caimano et al., 2000). Both of these plasmid types may represent satellite phages that lack structural and functional phage genes but rely on the machinery encoded on cp32s to spread through a population. *B. burgdorferi* strain B31 lp56 contains an integrated cp32 (Casjens et al., 2000). This integration, which appears to be a rather recent event, occurred in the middle of a putative late operon gene, and two of its other late operon genes have frameshift mutations (Casjens et al., 2000), suggesting that the late operon of this cp32 is beginning to decay; such decay is characteristic of prophage genes that are no longer under selection (Canchaya et al., 2003; Casjens, 2003). lp54 contains twenty noncontiguous genes that are similar (but diverged) in sequence to the putative cp32 late operon genes (including the putative terminase and portal genes). However, in this region of lp54 (from

approximately BB_A31 to BB_A55), there are two large insertions, one deletion and one substitution relative to the cp32 sequences (Casjens et al., 2000). Thus, whether lp54 is a productive prophage is unclear; neither this plasmid nor lp56 has been observed in preparations of phage from induced *B. burgdorferi* cultures. One other linear plasmid, lp28-2 (and its relatives lp28-6, lp28-7 and lp28-9), is only very distantly related to the cp32s, but also carries a cluster of genes that may be phage-related (Eggers et al., 2001a; Casjens et al., 2018). These genes should encode a potential large subunit terminase (G21), a portal (G20), a tape-measure protein (G08) that is responsible for determining tail length in other bacteriophages (Hendrix, 1988), and several proteins (G10, G22, G23 and G27) with very distant homology to other phage proteins. Linear plasmids with these genes are found in other LB species as well, but an lp28-2-like genome has not yet been demonstrated to be packaged into phage virions. Given that a number of different bacteriophages have been observed in association with LB, however, one of these could potentially be derived from a more intact lp28-2 or related plasmid.

Future prospects

Determination of the complete genome sequence of the *B. burgdorferi* type strain B31, including the chromosome and all plasmids, in the year 2000 was a major step forward in the study of the Lyme disease bacteria as it opened the door to many new experimental avenues. Complete genome sequences of other *B. burgdorferi* isolates and a number of other LB species in the subsequent two decades has provided a much more detailed and refined picture of genetic variation and relationships within and among these LB lineages. This information also formed the basis for more accurate strain typing and lineage tracking, as well as disease diagnosis and detection of LB spirochetes. Since NextGen sequencing is much cheaper and faster, elucidation of many more complete LB genome sequences is expected in the near future. This information should make analyses of LB phylogeny, population biology, as well as host species and phylogeographic correlations, much more robust and informative. The genome sequences of LB species that do not cause human disease may point the way towards identification of genes important in human Lyme disease.

Although significant and ongoing progress has been made in understanding the roles of particular LB genes in mouse infection, much remains to be learned. Understanding of the molecular pathogenesis of Lyme disease, especially in humans, remains quite incomplete. Further, the possible loss of plasmids during the initial isolation of LB from their natural hosts still causes uncertainty in such studies. Genetic manipulation of LB and the use of mutant LB spirochetes in animal infection studies has been largely limited to the B31 strain of *B. burgdorferi* in laboratory mice. This is due to technical considerations and the current inability to modify DNA of most other isolates. Overcoming these difficulties in future studies of other *B. burgdorferi* lineages and LB species should lead to new insights in the molecular pathogenesis of Lyme disease.

Acknowledgements

The authors graciously thank current and former members of their research groups and their colleagues for their contributions to LB genomics. The authors research was supported by grants AR041511 and AI045801 (IS), AI049003 (SRC) and AI139782 (WQ) from the National Institutes of Health and a Quinnipiac University School of Health Sciences Faculty Research Award (CHE).

References

Akins, D.R., Caimano, M.J., Yang, X., Cerna, F., Norgard, M.V., and Radolf, J.D. (1999). Molecular and evolutionary analysis of *Borrelia burgdorferi* 297 circular plasmid-encoded lipoproteins with OspE- and OspF-like leader peptides. Infect Immun *67*, 1526-1532.

Alitalo, A., Meri, T., Lankinen, H., Seppala, I., Lahdenne, P., Hefty, P.S., Akins, D., and Meri, S. (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.

Andersson, M., Scherman, K., and Raberg, L. (2013). Multiple-strain infections of *Borrelia afzelii*: a role for within-host interactions in the maintenance of antigenic diversity? Am Nat *181*, 545-554. https://doi.org/10.1086/669905

Anderton, J.M., Tokarz, R., Thill, C.D., Kuhlow, C.J., Brooks, C.S., Akins, D.R., Katona, L.I., and Benach, J.L. (2004). Whole-genome DNA array analysis of the response of *Borrelia burgdorferi* to a bactericidal monoclonal antibody. Infect Immun *72*, 2035-2044.

Ansari, M.A., and Didelot, X. (2014). Inference of the properties of the recombination process from whole bacterial genomes. Genetics *196*, 253-265. https://doi.org/10.1534/genetics.113.157172

Antonara, S., Chafel, R.M., LaFrance, M., and Coburn, J. (2007). *Borrelia burgdorferi* adhesins identified using in vivo phage display. Mol Microbiol *66*, 262-276.

Antonara, S., Ristow, L., and Coburn, J. (2011). Adhesion mechanisms of *Borrelia burgdorferi*. Adv Exp Med Biol *715*, 35-49. https://doi.org/10.1007/978-94-007-0940-9_3

Aranjuez, G.F., Kuhn, H.W., Adams, P.P., and Jewett, M.W. (2019). *Borrelia burgdorferi bbk13* Is Critical for Spirochete Population Expansion in the Skin during Early Infection. Infect Immun *87*. https://doi.org/10.1128/iai.00887-18

Arnold, W.K., Savage, C.R., Lethbridge, K.G., Smith, T.C., 2nd, Brissette, C.A., Seshu, J., and Stevenson, B. (2018). Transcriptomic insights on the virulence-controlling CsrA, BadR, RpoN, and RpoS

regulatory networks in the Lyme disease spirochete. PLoS One *13*, e0203286. https://doi.org/10.1371/journal.pone.0203286

Attie, O., Bruno, J.F., Xu, Y., Qiu, D., Luft, B.J., and Qiu, W.G. (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. https://doi.org/10.1016/j.meegid.2006.02.008

Austin, S., and Abeles, A. (1985). The partition functions of P1, P7, and F miniplasmids. Basic Life Sci *30*, 215-226.

Bacon, R.M., Kugeler, K.J., and Mead, P.S. (2008). Surveillance for Lyme disease--United States, 1992-2006. MMWR Surveill Summ *57*, 1-9.

Banik, S., Terekhova, D., Iyer, R., Pappas, C.J., Caimano, M.J., Radolf, J.D., and Schwartz, I. (2011). BB0844, an RpoS-regulated protein, is dispensable for *Borrelia burgdorferi* infectivity and maintenance in the mouse-tick infectious cycle. Infect Immun *79*, 1208-1217. https://doi.org/10.1128/IAI.01156-10

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., *et al.* (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol *19*, 455-477. https://doi.org/10.1089/cmb.2012.0021

Baptista, S., Quaresma, A., Aires, T., Kurtenbach, K., Santos-Reis, M., Nicholson, M., and Collares-Pereira, M. (2004). Lyme borreliosis spirochetes in questing ticks from mainland Portugal. Int J Med Microbiol *293 Suppl 37*, 109-116. https://doi.org/10.1016/s1433-1128(04)80016-0

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

Barbour, A.G. (2016). Multiple and Diverse *vsp* and *vlp* Sequences in *Borrelia miyamotoi*, a Hard Tick-Borne Zoonotic Pathogen. PLoS One *11*, e0146283. https://doi.org/10.1371/journal.pone.0146283

Barbour, A.G., and Fish, D. (1993). The biological and social phenomenon of Lyme disease. Science *260*, 1610-1616.

Barbour, A.G., and Garon, C.F. (1987). Linear plasmids of the bacterium *Borrelia burgdorferi* have covalently closed ends. Science *237*, 409-411.

Barbour, A.G., and Hayes, S.F. (1986). Biology of *Borrelia* species. Microbiol Rev *50*, 381-400.

Barbour, A.G., and Travinsky, B. (2010). Evolution and distribution of the *ospC* Gene, a transferable serotype determinant of *Borrelia burgdorferi*. MBio *1*, e00153-00110. https://doi.org/10.1128/mBio.00153-10

Barbour, A.Q., W. (2019). Borreliella. In Bergey's Manual of Systematics of Archaea and Bacteria (John Wiley & Sons, Inc. in association with Bergey's Manual Trust), pp. 1-22. https://doi.org/https://doi.org/10.1002/9781118960608.gbm01525

Barthold, S.W., Hodzic, E., Tunev, S., and Feng, S. (2006). Antibody-mediated disease remission in the mouse model of lyme borreliosis. Infect Immun *74*, 4817-4825. https://doi.org/74/8/4817 [pii] 10.1128/IAI.00469-06

Bauer, Y., Hofmann, H., Jahraus, O., Mytilineos, J., Simon, M.M., and Wallich, R. (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.

Baum, E., Randall, A.Z., Zeller, M., and Barbour, A.G. (2013). Inferring epitopes of a polymorphic antigen amidst broadly cross-reactive antibodies using protein microarrays: a study of OspC proteins of *Borrelia burgdorferi*. PLoS One *8*, e67445. https://doi.org/10.1371/journal.pone.0067445

Beaurepaire, C., and Chaconas, G. (2005). Mapping of essential replication functions of the linear plasmid lp17 of *B. burgdorferi* by targeted deletion walking. Mol Microbiol *57*, 132-142. https://doi.org/MMI4688 [pii] 10.1111/j.1365-2958.2005.04688.x

Becker, N.S., Margos, G., Blum, H., Krebs, S., Graf, A., Lane, R.S., Castillo-Ramirez, S., Sing, A., and Fingerle, V. (2016). Recurrent evolution of host and vector association in bacteria of the *Borrelia burgdorferi* sensu lato species complex. BMC Genomics *17*, 734. https://doi.org/10.1186/s12864-016-3016-4

Behera, A.K., Durand, E., Cugini, C., Antonara, S., Bourassa, L., Hildebrand, E., Hu, L.T., and Coburn, J. (2008). *Borrelia burgdorferi* BBB07 interaction with integrin alpha3beta1 stimulates production of pro-inflammatory mediators in primary human chondrocytes. Cell Microbiol *10*, 320-331. https://doi.org/CMI1043 [pii] 10.1111/j.1462-5822.2007.01043.x

Bernard, Q., Smith, A.A., Yang, X., Koci, J., Foor, S.D., Cramer, S.D., Zhuang, X., Dwyer, J.E., Lin, Y.P., Mongodin, E.F., *et al.* (2018). Plasticity in early immune evasion strategies of a bacterial pathogen. Proc Natl Acad Sci U S A *115*, E3788-e3797. https://doi.org/10.1073/pnas.1718595115

Bestor, A., Rego, R.O., Tilly, K., and Rosa, P.A. (2012). Competitive advantage of *Borrelia burgdorferi* with outer surface protein BBA03 during tick-mediated infection of the mammalian host. Infect Immun *80*, 3501-3511. https://doi.org/10.1128/iai.00521-12

Bestor, A., Stewart, P.E., Jewett, M.W., Sarkar, A., Tilly, K., and Rosa, P.A. (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. https://doi.org/IAI.01059-09 [pii] 10.1128/IAI.01059-09

Bhatia, B., Hillman, C., Carracoi, V., Cheff, B.N., Tilly, K., and Rosa, P.A. (2018). Infection history of the blood-meal host dictates pathogenic potential of the Lyme disease spirochete within the feeding tick vector. PLoS Pathogens *14*, e1006959. https://doi.org/10.1371/journal.ppat.1006959

Bishop, C.J., Aanensen, D.M., Jordan, G.E., Kilian, M., Hanage, W.P., and Spratt, B.G. (2009). Assigning strains to bacterial species via the internet. BMC Biol *7*, 3. https://doi.org/10.1186/1741-7007-7-3

Blazejak, K., Raulf, M.K., Janecek, E., Jordan, D., Fingerle, V., and Strube, C. (2018). Shifts in *Borrelia burgdorferi* (s.l.) geno-species infections in *Ixodes ricinus* over a 10-year surveillance period in the city of Hanover (Germany) and *Borrelia miyamotoi*-specific Reverse Line Blot detection. Parasit Vectors *11*, 304. https://doi.org/10.1186/s13071-018-2882-9

Blevins, J.S., Hagman, K.E., and Norgard, M.V. (2008). Assessment of decorin-binding protein A to the infectivity of *Borrelia burgdorferi* in the murine models of needle and tick infection. BMC Microbiol *8*, 82. https://doi.org/1471-2180-8-82 [pii] 10.1186/1471-2180-8-82

Bobay, L.M., Traverse, C.C., and Ochman, H. (2015). Impermanence of bacterial clones. Proc Natl Acad Sci U S A *112*, 8893-8900. https://doi.org/10.1073/pnas.1501724112

Boden, K., Lobenstein, S., Hermann, B., Margos, G., and Fingerle, V. (2016). *Borrelia miyamotoi*-Associated Neuroborreliosis in Immunocompromised Person. Emerg Infect Dis *22*, 1617-1620. https://doi.org/10.3201/eid2209.152034

Bondy-Denomy, J., Qian, J., Westra, E.R., Buckling, A., Guttman, D.S., Davidson, A.R., and Maxwell, K.L. (2016). Prophages mediate defense against phage infection through diverse mechanisms. ISME J *10*, 2854-2866. https://doi.org/10.1038/ismej.2016.79

Bono, J.L., Tilly, K., Stevenson, B., Hogan, D., and Rosa, P. (1998). Oligopeptide permease in *Borrelia burgdorferi*: putative peptide-binding components encoded by both chromosomal and plasmid loci. Microbiology *144*, 1033-1044.

Brissette, C.A., Bykowski, T., Cooley, A.E., Bowman, A., and Stevenson, B. (2009a). *Borrelia burgdorferi* RevA antigen binds host fibronectin. Infect Immun *77*, 2802-2812. https://doi.org/IAI.00227-09 [pii] 10.1128/IAI.00227-09

Brissette, C.A., Cooley, A.E., Burns, L.H., Riley, S.P., Verma, A., Woodman, M.E., Bykowski, T., and Stevenson, B. (2008). Lyme borreliosis spirochete Erp proteins, their known host ligands, and

potential roles in mammalian infection. Int J Med Microbiol *298*, 257-267. https://doi.org/10.1016/j.ijmm.2007.09.004

Brissette, C.A., Haupt, K., Barthel, D., Cooley, A.E., Bowman, A., Skerka, C., Wallich, R., Zipfel, P.F., Kraiczy, P., and Stevenson, B. (2009b). *Borrelia burgdorferi* infection-associated surface proteins ErpP, ErpA, and ErpC bind human plasminogen. Infect Immun *77*, 300-306. https://doi.org/IAI.01133-08 [pii] 10.1128/IAI.01133-08

Brissette, C.A., Verma, A., Bowman, A., Cooley, A.E., and Stevenson, B. (2009c). The *Borrelia burgdorferi* outer-surface protein ErpX binds mammalian laminin. Microbiology *155*, 863-872. https://doi.org/155/3/863 [pii] 10.1099/mic.0.024604-0

Brisson, D., Baxamusa, N., Schwartz, I., and Wormser, G.P. (2011). Biodiversity of *Borrelia burgdorferi* strains in tissues of Lyme disease patients. PLoS One *6*, e22926. https://doi.org/10.1371/journal.pone.0022926

Brisson, D., Drecktrah, D., Eggers, C.H., and Samuels, D.S. (2012). Genetics of *Borrelia burgdorferi*. Annu Rev Genet *46*, 515-536. https://doi.org/10.1146/annurev-genet-011112-112140

Brisson, D., and Dykhuizen, D.E. (2004). *ospC* diversity in *Borrelia burgdorferi*: different hosts are different niches. Genetics *168*, 713-722.

Brisson, D., Vandermause, M.F., Meece, J.K., Reed, K.D., and Dykhuizen, D.E. (2010). Evolution of northeastern and midwestern *Borrelia burgdorferi*, United States. Emerg Infect Dis *16*, 911-917.

Brisson, D., Zhou, W., Jutras, B.L., Casjens, S., and Stevenson, B. (2013). Distribution of cp32 prophages among Lyme disease-causing spirochetes and natural diversity of their lipoprotein-encoding *erp* loci. Appl Environ Microbiol *79*, 4115-4128. https://doi.org/AEM.00817-13 [pii] 10.1128/AEM.00817-13

Buckles, E.L., Earnhart, C.G., and Marconi, R.T. (2006). Analysis of antibody response in humans to the type A OspC loop 5 domain and assessment of the potential utility of the loop 5 epitope in Lyme disease vaccine development. Clin Vaccine Immunol *13*, 1162-1165. https://doi.org/10.1128/CVI.00099-06

Bunikis, J., Garpmo, U., Tsao, J., Berglund, J., Fish, D., and Barbour, A.G. (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.

Burns, L.H., Adams, C.A., Riley, S.P., Jutras, B.L., Bowman, A., Chenail, A.M., Cooley, A.E., Haselhorst, L.A., Moore, A.M., Babb, K., *et al.* (2010). BpaB, a novel protein encoded by the Lyme disease spirochete's cp32 prophages, binds to erp Operator 2 DNA. Nucleic Acids Res *38*, 5443-5455. https://doi.org/10.1093/nar/gkq284

Byram, R., Gaultney, R.A., Floden, A.M., Hellekson, C., Stone, B.L., Bowman, A., Stevenson, B., Johnson, B.J., and Brissette, C.A. (2015). *Borrelia burgdorferi* RevA Significantly Affects Pathogenicity and Host Response in the Mouse Model of Lyme Disease. Infect Immun *83*, 3675-3683. https://doi.org/10.1128/iai.00530-15

Byram, R., Stewart, P.E., and Rosa, P. (2004). The essential nature of the ubiquitous 26-kilobase circular replicon of *Borrelia burgdorferi*. J Bacteriol *186*, 3561-3569. https://doi.org/10.1128/JB.186.11.3561-3569.2004 186/11/3561

Caimano, M.J., Groshong, A.M., Belperron, A., Mao, J., Hawley, K.L., Luthra, A., Graham, D.E., Earnhart, C.G., Marconi, R.T., Bockenstedt, L.K., *et al.* (2019). The RpoS Gatekeeper in *Borrelia burgdorferi*: An Invariant Regulatory Scheme That Promotes Spirochete Persistence in Reservoir Hosts and Niche Diversity. Front Microbiol *10*, 1923. https://doi.org/10.3389/fmicb.2019.01923

Caimano, M.J., Iyer, R., Eggers, C.H., Gonzalez, C., Morton, E.A., Gilbert, M.A., Schwartz, I., and Radolf, J.D. (2007). Analysis of the RpoS regulon in *Borrelia burgdorferi* in response to mammalian host signals provides insight into RpoS function during the enzootic cycle. Mol Microbiol *65*, 1193-1217. https://doi.org/MMI5860 [pii] 10.1111/j.1365-2958.2007.05860.x

Caimano, M.J., Yang, X., Popova, T.G., Clawson, M.L., Akins, D.R., Norgard, M.V., and Radolf, J.D. (2000). Molecular and evolutionary characterization of the cp32/18 family of supercoiled plasmids in *Borrelia burgdorferi* 297. Infect Immun *68*, 1574-1586.

Canchaya, C., Proux, C., Fournous, G., Bruttin, A., and Brussow, H. (2003). Prophage genomics. Microbiol Mol Biol Rev *67*, 238-276.

Carpi, G., Walter, K.S., Bent, S.J., Hoen, A.G., Diuk-Wasser, M., and Caccone, A. (2015). Whole genome capture of vector-borne pathogens from mixed DNA samples: a case study of *Borrelia burgdorferi*. BMC Genomics *16*, 434. https://doi.org/10.1186/s12864-015-1634-x

Casjens, S. (1998). The diverse and dynamic structure of bacterial genomes. Annu Rev Genet *32*, 339-377. https://doi.org/10.1146/annurev.genet.32.1.339

Casjens, S. (2000). Borrelia genomes in the year 2000. J Mol Microbiol Biotechnol *2*, 401-410.

Casjens, S. (2003). Prophages in bacterial genomics: What have we learned so far? Molec Microbiol *249*, 277-300.

Casjens, S., Delange, M., Ley, H.L., 3rd, Rosa, P., and Huang, W.M. (1995). Linear chromosomes of Lyme disease agent spirochetes: genetic diversity and conservation of gene order. J Bacteriol *177*, 2769-2780.

Casjens, S., Eggers, C.H., and Schwartz, I. (2010). *Borrelia* genomics: chromsosome, plasmids, bacteriophges and genetic variation. In *Borrelia* molecular biology, host interaction and pathogenicity, D.S. Samuels, and J. Radolf, eds. (Norfolk: Caister Academic Press), pp. 27-53.

Casjens, S., Hatfull, G., and Hendrix, R. (1992). Evolution of dsDNA tailed-bacteriophage genomes. Sem Virol *3*, 383-397.

Casjens, S., and Hendrix, R. (2002). Bacteriophage roles in bacterial chromosome evolution. In The Bacterial Chromosome, P. Higgins, ed. (Washington, D.C.: ASM Press), p. 39-52.

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

Casjens, S., Murphy, M., DeLange, M., Sampson, L., van Vugt, R., and Huang, W.M. (1997a). Telomeres of the linear chromosomes of Lyme disease spirochaetes: nucleotide sequence and possible exchange with linear plasmid telomeres. Mol Microbiol *26*, 581-596. https://doi.org/10.1046/j.1365-2958.1997.6051963.x

Casjens, S., Palmer, N., van, V.R., Huang, W.M., Stevenson, B., Rosa, P., Lathigra, R., Sutton, G., Peterson, J., Dodson, R.J., *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.

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

Casjens, S.R. (2005). Comparative genomics and evolution of the tailed-bacteriophages. Curr Opin Microbiol *8*, 451-458.

Casjens, S.R. (2008). Diversity among the tailed-bacteriophages that infect the *Enterobacteriaceae*. Res Microbiol *159*, 340-348. https://doi.org/10.1016/j.resmic.2008.04.005

Casjens, S.R., Di, L., Akther, S., Mongodin, E.F., Luft, B.J., Schutzer, S.E., Fraser, C.M., and Qiu, W.G. (2018). Primordial origin and diversification of plasmids in Lyme disease agent bacteria. BMC Genomics *19*, 218. https://doi.org/10.1186/s12864-018-4597-x

Casjens, S.R., Fraser-Liggett, C.M., Mongodin, E.F., Qiu, W.G., Dunn, J.J., Luft, B.J., and Schutzer, S.E. (2011a). Whole genome sequence of an unusual *Borrelia burgdorferi* sensu lato isolate. J Bacteriol *193*, 1489-1490. https://doi.org/10.1128/JB.01521-10

Casjens, S.R., Gilcrease, E.B., Vujadinovic, M., Mongodin, E.F., Luft, B.J., Schutzer, S.E., Fraser, C.M., and Qiu, W.G. (2017). Plasmid diversity and phylogenetic consistency in the Lyme disease agent

Borrelia burgdorferi. BMC Genomics *18*, 165. https://doi.org/10.1186/s12864-017-3553-5

Casjens, S.R., Huang, W.M., Gilcrease, E.B., Qiu, W.G., McCaig, W.D., Luft, B.J., and Schutzer, S.E. (2006). Comparative genomics of *Borrelia burgdorferi*. In Molecular Biology of Spirochetes, F.C. Cabello, D. Hulinska, and H.P. Godfrey, eds. (Amsterdam: IOS Press), pp. 79-95.

Casjens, S.R., Mongodin, E.F., Qiu, W.G., Dunn, J.J., Luft, B.J., Fraser-Liggett, C.M., and Schutzer, S.E. (2011b). Whole-genome sequences of two *Borrelia afzelii* and two *Borrelia garinii* Lyme disease agent isolates. J Bacteriol *193*, 6995-6996. https://doi.org/10.1128/JB.05951-11

Casjens, S.R., Mongodin, E.F., Qiu, W.G., Luft, B.J., Schutzer, S.E., Gilcrease, E.B., Huang, W.M., Vujadinovic, M., Aron, J.K., Vargas, L.C., *et al.* (2012). Genome stability of Lyme disease spirochetes: comparative genomics of *Borrelia burgdorferi* plasmids. PLoS One *7*, e33280.

Castillo-Ramirez, S., Fingerle, V., Jungnick, S., Straubinger, R.K., Krebs, S., Blum, H., Meinel, D.M., Hofmann, H., Guertler, P., Sing, A., *et al.* (2016). Trans-Atlantic exchanges have shaped the population structure of the Lyme disease agent *Borrelia burgdorferi* sensu stricto. Sci Rep *6*, 22794. https://doi.org/10.1038/srep22794

Catalano, C.E. (2005). Viral genome packaging machines: genetics, structure, and mechanism (Georgetown, TX: Landes Bioscience).

Chaconas, G., Castellanos, M., and Verhey, T.B. (2020). Changing of the guard: How the Lyme disease spirochete subverts the host immune response. J Biol Chem *295*, 301-313. https://doi.org/10.1074/jbc.REV119.008583

Champion, C.I., Blanco, D.R., Skare, J.T., Haake, D.A., Giladi, M., Foley, D., Miller, J.N., and Lovett, M.A. (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.

Chauhan, G., McClure, J., Hekman, J., Marsh, P.W., Bailey, J.A., Daniels, R.F., Genereux, D.P., and Karlsson, E.K. (2019). Combining Citizen Science and Genomics to Investigate Tick, Pathogen, and Commensal Microbiome at Single-Tick Resolution. Front Genet *10*, 1322. https://doi.org/10.3389/fgene.2019.01322

Cheetham, B.F., Tattersall, D.B., Bloomfield, G.A., Rood, J.I., and Katz, M.E. (1995). Identification of a gene encoding a bacteriophage-related integrase in a vap region of the Dichelobacter nodosus genome. Gene *162*, 53-58.

Chenail, A.M., Jutras, B.L., Adams, C.A., Burns, L.H., Bowman, A., Verma, A., and Stevenson, B. (2012). *Borrelia burgdorferi* cp32 BpaB modulates expression of the prophage NucP nuclease and SsbP single-stranded DNA-binding protein. J Bacteriol *194*, 4570-4578. https://doi.org/10.1128/JB.00661-12

Cohan, F.M. (2002). What are bacterial species? Annu Rev Microbiol *56*, 457-487.

Coipan, E.C., Jahfari, S., Fonville, M., Oei, G.A., Spanjaard, L., Takumi, K., Hovius, J.W., and Sprong, H. (2016). Imbalanced presence of *Borrelia burgdorferi* s.l. multilocus sequence types in clinical manifestations of Lyme borreliosis. Infect Genet Evol *42*, 66-76. https://doi.org/10.1016/j.meegid.2016.04.019

Collares-Pereira, M., Couceiro, S., Franca, I., Kurtenbach, K., Schafer, S.M., Vitorino, L., Goncalves, L., Baptista, S., Vieira, M.L., and Cunha, C. (2004). First isolation of *Borrelia lusitaniae* from a human patient. J Clin Microbiol *42*, 1316-1318. https://doi.org/10.1128/jcm.42.3.1316-1318.2004

Comstedt, P., Jakobsson, T., and Bergstrom, S. (2011). Global ecology and epidemiology of *Borrelia garinii* spirochetes. Infect Ecol Epidemiol *1*. https://doi.org/10.3402/iee.v1i0.9545

Corona, A., and Schwartz, I. (2015). *Borrelia burgdorferi*: Carbon Metabolism and the Tick-Mammal Enzootic Cycle. Microbiol Spectr *3*. https://doi.org/10.1128/microbiolspec.MBP-0011-2014

Coutte, L., Botkin, D.J., Gao, L., and Norris, S.J. (2009). Detailed analysis of sequence changes occurring during *vlsE* antigenic variation in the mouse model of *Borrelia burgdorferi* infection. PLoS Pathogens *5*, e1000293. https://doi.org/10.1371/journal.ppat.1000293

Cross, S.T., Kapuscinski, M.L., Perino, J., Maertens, B.L., Weger-Lucarelli, J., Ebel, G.D., and Stenglein, M.D. (2018). Co-Infection Patterns in Individual *Ixodes scapularis* Ticks Reveal Associations between Viral, Eukaryotic and Bacterial Microorganisms. Viruses *10*. https://doi.org/10.3390/v10070388

Damman, C.J., Eggers, C.H., Samuels, D.S., and Oliver, D.B. (2000). Characterization of *Borrelia burgdorferi* BlyA and BlyB proteins: a prophage-encoded holin-like system. J Bacteriol *182*, 6791-6797.

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

de Carvalho, I.L., Fonseca, J.E., Marques, J.G., Ullmann, A., Hojgaard, A., Zeidner, N., and Nuncio, M.S. (2008). Vasculitis-like syndrome associated with Borrelia lusitaniae infection. Clin Rheumatol *27*, 1587-1591. https://doi.org/10.1007/s10067-008-1012-z

Di, L., Wan, Z., Akther, S., Ying, C., Larracuente, A., Li, L., Di, C., Nunez, R., Cucura, D.M., Goddard, N.L., *et al.* (2018). Genotyping and Quantifying Lyme Pathogen Strains by Deep Sequencing of the Outer Surface Protein C (*ospC*) Locus. J Clin Microbiol *56*. https://doi.org/10.1128/JCM.00940-18

Dion, M.B., Oechslin, F., and Moineau, S. (2020). Phage diversity, genomics and phylogeny. Nat Rev Microbiol *18*, 125-138. https://doi.org/10.1038/s41579-019-0311-5

Diuk-Wasser, M.A., Gatewood, A.G., Cortinas, M.R., Yaremych-Hamer, S., Tsao, J., Kitron, U., Hickling, G., Brownstein, J.S., Walker, E., Piesman, J., *et al.* (2006). Spatiotemporal patterns of host-seeking *Ixodes scapularis* nymphs (Acari: Ixodidae) in the United States. J Med Entomol *43*, 166-176.

Dowdell, A.S., Murphy, M.D., Azodi, C., Swanson, S.K., Florens, L., Chen, S., and Zuckert, W.R. (2017). Comprehensive Spatial Analysis of the *Borrelia burgdorferi* Lipoproteome Reveals a Compartmentalization Bias toward the Bacterial Surface. J Bacteriol *199*. https://doi.org/10.1128/JB.00658-16

Drecktrah, D., Lybecker, M., Popitsch, N., Rescheneder, P., Hall, L.S., and Samuels, D.S. (2015). The *Borrelia burgdorferi* RelA/SpoT homolog and stringent response regulate survival in the tick vector and global gene expression during starvation. PLoS Pathogens *11*, e1005160. https://doi.org/10.1371/journal.ppat.1005160

Dulebohn, D.P., Bestor, A., and Rosa, P.A. (2013). *Borrelia burgdorferi* linear plasmid 28-3 confers a selective advantage in an experimental mouse-tick infection model. Infect Immun *81*, 2986-2996. https://doi.org/10.1128/IAI.00219-13

Dulebohn, D.P., Hayes, B.M., and Rosa, P.A. (2014). Global repression of host-associated genes of the Lyme disease spirochete through post-transcriptional modulation of the alternative sigma factor RpoS. PLoS One *9*, e93141. https://doi.org/10.1371/journal.pone.0093141

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

Durand, J., Herrmann, C., Genne, D., Sarr, A., Gern, L., and Voordouw, M.J. (2017a). Multistrain Infections with Lyme Borreliosis Pathogens in the Tick Vector. Appl Environ Microbiol *83*. https://doi.org/10.1128/AEM.02552-16

Durand, J., Jacquet, M., Paillard, L., Rais, O., Gern, L., and Voordouw, M.J. (2015). Cross-Immunity and Community Structure of a Multiple-Strain Pathogen in the Tick Vector. Appl Environ Microbiol *81*, 7740-7752. https://doi.org/10.1128/AEM.02296-15

Durand, J., Jacquet, M., Rais, O., Gern, L., and Voordouw, M.J. (2017b). Fitness estimates from experimental infections predict the long-term

strain structure of a vector-borne pathogen in the field. Sci Rep *7*, 1851. https://doi.org/10.1038/s41598-017-01821-1

Dykhuizen, D.E., and Baranton, G. (2001). The implications of a low rate of horizontal transfer in *Borrelia*. Trends Microbiol *9*, 344-350.

Dykhuizen, D.E., Brisson, D., Sandigursky, S., Wormser, G.P., Nowakowski, J., Nadelman, R.B., and Schwartz, I. (2008). The propensity of different *Borrelia burgdorferi* sensu stricto genotypes to cause disseminated infections in humans. Am J Trop Med Hyg *78*, 806-810.

Dykhuizen, D.E., Polin, D.S., Dunn, J.J., Wilske, B., Preac-Mursic, V., Dattwyler, R.J., and Luft, B.J. (1993). *Borrelia burgdorferi* is clonal: implications for taxonomy and vaccine development. Proc Natl Acad Sci U S A *90*, 10163-10167. https://doi.org/10.1073/pnas.90.21.10163

Earnhart, C.G., and Marconi, R.T. (2007). Construction and analysis of variants of a polyvalent Lyme disease vaccine: approaches for improving the immune response to chimeric vaccinogens. Vaccine *25*, 3419-3427. https://doi.org/10.1016/j.vaccine.2006.12.051

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res *32*, 1792-1797. https://doi.org/10.1093/nar/gkh340

Eggers, C.H., Caimano, M.J., Clawson, M.L., Miller, W.G., Samuels, D.S., and Radolf, J.D. (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 spirochaete. Mol Microbiol *43*, 281-295.

Eggers, C.H., Casjens, S., Hayes, S.F., Garon, C.F., Damman, C.J., Oliver, D.B., and Samuels, D.S. (2000). Bacteriophages of spirochetes. J Mol Microbiol Biotechnol *2*, 365-373.

Eggers, C.H., Casjens, S., and Samuels, D.S. (2001a). Bacteriophges of *Borrelia burgdorferi* and other spirochetes. In The spirochetes Molecilar and celullar biology, M. Saier, and J. Garcia-Lara, eds. (Wiltshire: Hozison Scientific Press), pp. 35-44.

Eggers, C.H., Gray, C.M., Preisig, A.M., Glenn, D.M., Pereira, J., Ayers, R.W., Alshahrani, M., Acabbo, C., Becker, M.R., Bruenn, K.N., *et al.* (2016). Phage-mediated horizontal gene transfer of both prophage and heterologous DNA by fBB-1, a bacteriophage of *Borrelia burgdorferi*. Pathog Dis *74*. https://doi.org/10.1093/femspd/ftw107

Eggers, C.H., Kimmel, B.J., Bono, J.L., Elias, A.F., Rosa, P., and Samuels, D.S. (2001b). Transduction by fBB-1, a bacteriophage of *Borrelia burgdorferi*. J Bacteriol *183*, 4771-4778.

Eggers, C.H., and Samuels, D.S. (1999). Molecular evidence for a new bacteriophage of *Borrelia burgdorferi*. J Bacteriol *181*, 7308-7313.

Eicken, C., Sharma, V., Klabunde, T., Lawrenz, M.B., Hardham, J.M., Norris, S.J., and Sacchettini, J.C. (2002). Crystal structure of Lyme disease variable surface antigen VlsE of *Borrelia burgdorferi*. J Biol Chem *277*, 21691-21696. https://doi.org/10.1074/jbc.M201547200

El-Hage, N., and Stevenson, B. (2002). Simultaneous coexpression of *Borrelia burgdorferi* Erp proteins occurs through a specific, *erp* locus-directed regulatory mechanism. J Bacteriol *184*, 4536-4543. https://doi.org/10.1128/jb.184.16.4536-4543.2002

Elder, J.F., Jr., and Turner, B.J. (1995). Concerted evolution of repetitive DNA sequences in eukaryotes. Q Rev Biol *70*, 297-320. https://doi.org/10.1086/419073

Elias, A.F., Stewart, P.E., Grimm, D., Caimano, M.J., Eggers, C.H., Tilly, K., Bono, J.L., Akins, D.R., Radolf, J.D., Schwan, T.G., *et al.* (2002). Clonal polymorphism of *Borrelia burgdorferi* strain B31 MI: implications for mutagenesis in an infectious strain background. Infect Immun *70*, 2139-2150.

Ellegren, H. (2014). Genome sequencing and population genomics in non-model organisms. Trends Ecol Evol *29*, 51-63. https://doi.org/10.1016/j.tree.2013.09.008

Ellis, T.C., Jain, S., Linowski, A.K., Rike, K., Bestor, A., Rosa, P.A., Halpern, M., Kurhanewicz, S., and Jewett, M.W. (2013). In vivo expression technology identifies a novel virulence factor critical for *Borrelia burgdorferi* persistence in mice. PLoS Pathogens *9*, e1003567. https://doi.org/10.1371/journal.ppat.1003567

Enright, M.C., and Spratt, B.G. (1999). Multilocus sequence typing. Trends Microbiol *7*, 482-487.

Farlow, J., Postic, D., Smith, K.L., Jay, Z., Baranton, G., and Keim, P. (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.

Fedorova, N., Kleinjan, J.E., James, D., Hui, L.T., Peeters, H., and Lane, R.S. (2014). Remarkable diversity of tick or mammalian-associated Borreliae in the metropolitan San Francisco Bay Area, California. Ticks Tick Borne Dis *5*, 951-961. https://doi.org/10.1016/j.ttbdis.2014.07.015

Felsenstein, J. (1989). PHYLIP - Phylogeny Inference Package. Cladistics *5*, 164-166.

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

Fingerle, V., Goettner, G., Gern, L., Wilske, B., and Schulte-Spechtel, U. (2007). Complementation of a *Borrelia afzelii* OspC mutant highlights the crucial role of OspC for dissemination of *Borrelia afzelii* in *Ixodes ricinus*. Int J Med Microbiol *297*, 97-107. https://doi.org/10.1016/j.ijmm.2006.11.003

Fingerle, V., Pritsch, M., Wachtler, M., Margos, G., Ruske, S., Jung, J., Loscher, T., Wendtner, C., and Wieser, A. (2016). "*Candidatus Borrelia kalaharica*" Detected from a Febrile Traveller Returning to Germany from Vacation in Southern Africa. PLoS Neg Trop Dis *10*, e0004559. https://doi.org/10.1371/journal.pntd.0004559

Fischer, J.R., LeBlanc, K.T., and Leong, J.M. (2006). Fibronectin binding protein BBK32 of the Lyme disease spirochete promotes bacterial attachment to glycosaminoglycans. Infect Immun *74*, 435-441. https://doi.org/74/1/435 [pii] 10.1128/IAI.74.1.435-441.2006

Ford, M.E., Sarkis, G.J., Belanger, A.E., Hendrix, R.W., and Hatfull, G.F. (1998). Genome structure of mycobacteriophage D29: implications for phage evolution. J Mol Biol *279*, 143-164.

Francisco, A.P., Vaz, C., Monteiro, P.T., Melo-Cristino, J., Ramirez, M., and Carrico, J.A. (2012). PHYLOViZ: phylogenetic inference and data visualization for sequence based typing methods. BMC Bioinformatics *13*, 87. https://doi.org/10.1186/1471-2105-13-87

Fraser, C., Hanage, W.P., and Spratt, B.G. (2007). Recombination and the nature of bacterial speciation. Science *315*, 476-480. https://doi.org/10.1126/science.1127573

Fraser, C.M., Casjens, S., Huang, W.M., Sutton, G.G., Clayton, R., Lathigra, R., White, O., Ketchum, K.A., Dodson, R., Hickey, E.K., *et al.* (1997). Genomic sequence of a Lyme disease spirochaete, *Borrelia burgdorferi*. Nature *390*, 580-586.

Gatzmann, F., Metzler, D., Krebs, S., Blum, H., Sing, A., Takano, A., Kawabata, H., Fingerle, V., Margos, G., and Becker, N.S. (2015). NGS population genetics analyses reveal divergent evolution of a Lyme Borreliosis agent in Europe and Asia. Ticks Tick Borne Dis *6*, 344-351. https://doi.org/10.1016/j.ttbdis.2015.02.008

Gazumyan, A., Schwartz, J.J., Liveris, D., and Schwartz, I. (1994). Sequence analysis of the ribosomal RNA operon of the Lyme disease spirochete, *Borrelia burgdorferi*. Gene *146*, 57-65.

Gilmore, R.D., Jr., Howison, R.R., Dietrich, G., Patton, T.G., Clifton, D.R., and Carroll, J.A. (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. https://doi.org/10.1073/pnas.1000268107

Gilmore, R.D., Jr., and Mbow, M.L. (1998). A monoclonal antibody generated by antigen inoculation via tick bite is reactive to the *Borrelia burgdorferi* Rev protein, a member of the 2.9 gene family locus. Infect Immun *66*, 980-986.

Girard, Y.A., Fedorova, N., and Lane, R.S. (2011). Genetic diversity of *Borrelia burgdorferi* and detection of B. *bissettii*-like DNA in serum of

north-coastal California residents. J Clin Microbiol *49*, 945-954. https://doi.org/10.1128/JCM.01689-10

Glockner, G., Lehmann, R., Romualdi, A., Pradella, S., Schulte-Spechtel, U., Schilhabel, M., Wilske, B., Suhnel, J., and Platzer, M. (2004). Comparative analysis of the *Borrelia garinii* genome. Nucleic Acids Res *32*, 6038-6046. https://doi.org/32/20/6038 [pii] 10.1093/nar/gkh953

Glockner, G., Schulte-Spechtel, U., Schilhabel, M., Felder, M., Suhnel, J., Wilske, B., and Platzer, M. (2006). Comparative genome analysis: selection pressure on the *Borrelia vls* cassettes is essential for infectivity. BMC Genomics *7*, 211. https://doi.org/10.1186/1471-2164-7-211

Gofton, A.W., Margos, G., Fingerle, V., Hepner, S., Loh, S.M., Ryan, U., Irwin, P., and Oskam, C.L. (2018). Genome-wide analysis of *Borrelia turcica* and '*Candidatus Borrelia tachyglossi*' shows relapsing fever-like genomes with unique genomic links to Lyme disease *Borrelia*. Infect Genet Evol *66*, 72-81. https://doi.org/10.1016/j.meegid.2018.09.013

Gomez-Diaz, E., Boulinier, T., Sertour, N., Cornet, M., Ferquel, E., and McCoy, K.D. (2011). Genetic structure of marine *Borrelia garinii* and population admixture with the terrestrial cycle of Lyme borreliosis. Environ Microbiol *13*, 2453-2467. https://doi.org/10.1111/j.1462-2920.2011.02515.x

Graves, C.J., Ros, V.I., Stevenson, B., Sniegowski, P.D., and Brisson, D. (2013). Natural selection promotes antigenic evolvability. PLoS Pathogens *9*, e1003766. https://doi.org/10.1371/journal.ppat.1003766

Greay, T.L., Gofton, A.W., Paparini, A., Ryan, U.M., Oskam, C.L., and Irwin, P.J. (2018). Recent insights into the tick microbiome gained through next-generation sequencing. Parasit Vectors *11*, 12. https://doi.org/10.1186/s13071-017-2550-5

Grigoriev, P.S., and Lobocka, M.B. (2001). Determinants of segregational stability of the linear plasmid-prophage N15 of *Escherichia coli*. Molec Microbiol *42*, 355-368.

Grimm, D., Elias, A.F., Tilly, K., and Rosa, P.A. (2003). Plasmid stability during in vitro propagation of *Borrelia burgdorferi* assessed at a clonal level. Infect Immun *71*, 3138-3145.

Grimm, D., Tilly, K., Bueschel, D.M., Fisher, M.A., Policastro, P.F., Gherardini, F.C., Schwan, T.G., and Rosa, P.A. (2005). Defining plasmids required by *Borrelia burgdorferi* for colonization of tick vector *Ixodes scapularis* (Acari: Ixodidae). J Med Entomol *42*, 676-684.

Grimm, D., Tilly, K., Byram, R., Stewart, P.E., Krum, J.G., Bueschel, D.M., Schwan, T.G., Policastro, P.F., Elias, A.F., and Rosa, P.A. (2004). Outer-surface protein C of the Lyme disease spirochete: a protein induced in ticks for infection of mammals. Proc Natl Acad Sci U S A *101*, 3142-3147.

Groshong, A.M., and Blevins, J.S. (2014). Insights into the biology of *Borrelia burgdorferi* gained through the application of molecular genetics. Advances in applied microbiology *86*, 41-143. https://doi.org/10.1016/B978-0-12-800262-9.00002-0

Groshong, A.M., Dey, A., Bezsonova, I., Caimano, M.J., and Radolf, J.D. (2017). Peptide Uptake Is Essential for *Borrelia burgdorferi* Viability and Involves Structural and Regulatory Complexity of its Oligopeptide Transporter. mBio *8*. https://doi.org/10.1128/mBio.02047-17

Hallstrom, T., Siegel, C., Morgelin, M., Kraiczy, P., Skerka, C., and Zipfel, P.F. (2013). CspA from *Borrelia burgdorferi* inhibits the terminal complement pathway. MBio *4*. https://doi.org/10.1128/mBio.00481-13

Hamer, S.A., Tsao, J.I., Walker, E.D., and Hickling, G.J. (2010). Invasion of the Lyme Disease Vector *Ixodes scapularis*: Implications for *Borrelia burgdorferi* Endemicity. Ecohealth *DOI: 10.1007/s10393-010-0287-0*.

Hammerschmidt, C., Klevenhaus, Y., Koenigs, A., Hallstrom, T., Fingerle, V., Skerka, C., Pos, K.M., Zipfel, P.F., Wallich, R., and Kraiczy, P. (2016). BGA66 and BGA71 facilitate complement resistance of *Borrelia bavariensis* by inhibiting assembly of the membrane attack complex. Mol Microbiol *99*, 407-424. https://doi.org/10.1111/mmi.13239

Hanincova, K., Kurtenbach, K., Diuk-Wasser, M., Brei, B., and Fish, D. (2006). Epidemic spread of Lyme borreliosis, northeastern United States. Emerg Infect Dis *12*, 604-611.

Hanincova, K., Liveris, D., Sandigursky, S., Wormser, G.P., and Schwartz, I. (2008). *Borrelia burgdorferi* sensu stricto is clonal in patients with early Lyme borreliosis. Appl Environ Microbiol *74*, 5008-5014.

Hanincova, K., Mukherjee, P., Ogden, N.H., Margos, G., Wormser, G.P., Reed, K.D., Meece, J.K., Vandermause, M.F., and Schwartz, I. (2013). Multilocus sequence typing of *Borrelia burgdorferi* suggests existence of lineages with differential pathogenic properties in humans. PLoS One *8*, e73066.

Haven, J., Vargas, L.C., Mongodin, E.F., Xue, V., Hernandez, Y., Pagan, P., Fraser-Liggett, C.M., Schutzer, S.E., Luft, B.J., Casjens, S.R., *et al.* (2011). Pervasive Recombination and Sympatric Genome Diversification Driven by Frequency-Dependent Selection in *Borrelia burgdorferi*, the Lyme Disease Bacterium. Genetics *189*, 951-966. https://doi.org/genetics.111.130773 [pii] 10.1534/genetics.111.130773

Hayes, B.M., Dulebohn, D.P., Sarkar, A., Tilly, K., Bestor, A., Ambroggio, X., and Rosa, P.A. (2014). Regulatory protein BBD18 of the lyme disease spirochete: essential role during tick acquisition? MBio *5*, e01017-01014. https://doi.org/10.1128/mBio.01017-14

Hayes, S.F., Burgdorfer, W., and Barbour, A.G. (1983). Bacteriophage in the *Ixodes dammini* spirochete, etiological agent of Lyme disease. J Bacteriol *154*, 1436-1439.

Hellwage, J., Meri, T., Heikkila, T., Alitalo, A., Panelius, J., Lahdenne, P., Seppala, I.J., and Meri, S. (2001). The complement regulator factor H binds to the surface protein OspE of *Borrelia burgdorferi*. J Biol Chem *276*, 8427-8435.

Hendrix, R., and Casjens, S. (2006). Bacteriophage l and its genetic neighborhood. In The Bacteriophages, 2nd Edition, R. Calendar, ed. (New York City, N.Y.: Oxford Press), pp. 409-447.

Hendrix, R.W. (1988). Tail length determination in double-stranded DNA bacteriophages. Curr Top Microbiol Immunol *136*, 21-29.

Hendrix, R.W. (2002). Bacteriophages: evolution of the majority. Theor Popul Biol *61*, 471-480.

Hendrix, R.W., Lawrence, J.G., Hatfull, G.F., and Casjens, S. (2000). The origins and ongoing evolution of viruses. Trends Microbiol *8*, 504-508.

Hendrix, R.W., Smith, M.C., Burns, R.N., Ford, M.E., and Hatfull, G.F. (1999). Evolutionary relationships among diverse bacteriophages and prophages: all the world's a phage. Proc Natl Acad Sci USA *96*, 2192-2197.

Hernandez, Y., Bernstein, R., Pagan, P., Vargas, L., McCaig, W., Ramrattan, G., Akther, S., Larracuente, A., Di, L., Vieira, F.G., *et al.* (2018). BpWrapper: BioPerl-based sequence and tree utilities for rapid prototyping of bioinformatics pipelines. BMC bioinformatics *19*, 76. https://doi.org/10.1186/s12859-018-2074-9

Hill, W.G., and Robertson, A. (1966). The effect of linkage on limits to artificial selection. Genet Res *8*, 269-294.

Hinnebusch, J., and Barbour, A.G. (1991). Linear plasmids of *Borrelia burgdorferi* have a telomeric structure and sequence similar to those of a eukaryotic virus. J Bacteriol *173*, 7233-7239.

Hinnebusch, J., and Barbour, A.G. (1992). Linear- and circular-plasmid copy numbers in *Borrelia burgdorferi*. J Bacteriol *174*, 5251-5257.

Hinnebusch, J., Bergstrom, S., and Barbour, A.G. (1990). Cloning and sequence analysis of linear plasmid telomeres of the bacterium *Borrelia burgdorferi*. Mol Microbiol *4*, 811-820.

Hoen, A.G., Margos, G., Bent, S.J., Kurtenbach, K., and Fish, D. (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.

Hovis, K.M., Tran, E., Sundy, C.M., Buckles, E., McDowell, J.V., and Marconi, R.T. (2006). Selective binding of *Borrelia burgdorferi* OspE paralogs to factor H and serum proteins from diverse animals: possible expansion of the role of OspE in Lyme disease pathogenesis. Infect Immun *74*, 1967-1972.

Hu, C.M., Wilske, B., Fingerle, V., Lobet, Y., and Gern, L. (2001). Transmission of *Borrelia garinii* OspA serotype 4 to BALB/c mice by *Ixodes ricinus* ticks collected in the field. J Clin Microbiol *39*, 1169-1171.

Huang, W.M., Robertson, M., Aron, J., and Casjens, S. (2004). Telomere exchange between linear replicons of *Borrelia burgdorferi*. J Bacteriol *186*, 4134-4141. https://doi.org/10.1128/JB.186.13.4134-4141.2004

Huegli, D., Hu, C.M., Humair, P.F., Wilske, B., and Gern, L. (2002). *Apodemus* species mice are reservoir hosts of *Borrelia garinii* OspA serotype 4 in Switzerland. J Clin Microbiol *40*, 4735-4737.

Ivanova, L., Christova, I., Neves, V., Aroso, M., Meirelles, L., Brisson, D., and Gomes-Solecki, M. (2009). Comprehensive seroprofiling of sixteen *B. burgdorferi* OspC: implications for Lyme disease diagnostics design. Clin Immunol *132*, 393-400. https://doi.org/10.1016/j.clim.2009.05.017

Ivanova, L.B., Tomova, A., Gonzalez-Acuna, D., Murua, R., Moreno, C.X., Hernandez, C., Cabello, J., Cabello, C., Daniels, T.J., Godfrey, H.P*., et al*. (2014). *Borrelia chilensis*, a new member of the *Borrelia burgdorferi* sensu lato complex that extends the range of this genospecies in the Southern Hemisphere. Environ Microbiol *16*, 1069-1080. https://doi.org/10.1111/1462-2920.12310

Iyer, R., Caimano, M.J., Luthra, A., Axline, D., Jr., Corona, A., Iacobas, D.A., Radolf, J.D., and Schwartz, I. (2015). Stage-specific global alterations in the transcriptomes of Lyme disease spirochetes during tick feeding and following mammalian host adaptation. Mol Microbiol *95*, 509-538. https://doi.org/10.1111/mmi.12882

Iyer, R., Kalu, O., Purser, J., Norris, S., Stevenson, B., and Schwartz, I. (2003). Linear and circular plasmid content in *Borrelia burgdorferi* clinical isolates. Infect Immun *71*, 3699-3706.

Izac, J.R., Camire, A.C., Earnhart, C.G., Embers, M.E., Funk, R.A., Breitschwerdt, E.B., and Marconi, R.T. (2019). Analysis of the antigenic determinants of the OspC protein of the Lyme disease spirochetes: Evidence that the C10 motif is not immunodominant or required to elicit bactericidal antibody responses. Vaccine *37*, 2401-2407. https://doi.org/10.1016/j.vaccine.2019.02.007

Izac, J.R., and Marconi, R.T. (2019). Diversity of the Lyme Disease Spirochetes and its Influence on Immune Responses to Infection and Vaccination. Vet Clin North Am Small Anim Pract *49*, 671-686. https://doi.org/10.1016/j.cvsm.2019.02.007

Jabbari, N., Glusman, G., Joesch-Cohen, L.M., Reddy, P.J., Moritz, R.L., Hood, L., and Lausted, C.G. (2018). Whole genome sequence and comparative analysis of *Borrelia burgdorferi* MM1. PLoS One *13*, e0198135. https://doi.org/10.1371/journal.pone.0198135

Jacquot, M., Gonnet, M., Ferquel, E., Abrial, D., Claude, A., Gasqui, P., Choumet, V., Charras-Garrido, M., Garnier, M., Faure, B*., et al.* (2014). Comparative population genomics of the *Borrelia burgdorferi* species complex reveals high degree of genetic isolation among species and underscores benefits and constraints to studying intra-specific epidemiological processes. PLoS One *9*, e94384. https://doi.org/10.1371/journal.pone.0094384

Jenkins, A., Hvidsten, D., Matussek, A., Lindgren, P.E., Stuen, S., and Kristiansen, B.E. (2012). *Borrelia burgdorferi* sensu lato in *Ixodes ricinus* ticks from Norway: evaluation of a PCR test targeting the chromosomal *flaB* gene. Exp Appl Acarol *58*, 431-439. https://doi.org/10.1007/s10493-012-9585-2

Jewett, M.W., Byram, R., Bestor, A., Tilly, K., Lawrence, K., Burtnick, M.N., Gherardini, F., and Rosa, P.A. (2007a). Genetic basis for retention of a critical virulence plasmid of *Borrelia burgdorferi*. Mol Microbiol *66*, 975-990. https://doi.org/MMI5969 [pii] 10.1111/j.1365-2958.2007.05969.x

Jewett, M.W., Jain, S., Linowski, A.K., Sarkar, A., and Rosa, P.A. (2011). Molecular characterization of the *Borrelia burgdorferi* in vivo-essential protein PncA. Microbiology *157*, 2831-2840. https://doi.org/mic.0.051706-0 [pii] 10.1099/mic.0.051706-0

Jewett, M.W., Lawrence, K., Bestor, A.C., Tilly, K., Grimm, D., Shaw, P., VanRaden, M., Gherardini, F., and Rosa, P.A. (2007b). The critical role of the linear plasmid lp36 in the infectious cycle of *Borrelia burgdorferi*. Mol Microbiol *64*, 1358-1374.

Jolley, K.A. (2009). Internet-based sequence-typing databases for bacterial molecular epidemiology. Methods Mol Biol *551*, 305-312. https://doi.org/10.1007/978-1-60327-999-4_21

Jolley, K.A., Chan, M.S., and Maiden, M.C. (2004). mlstdbNet - distributed multi-locus sequence typing (MLST) databases. BMC bioinformatics *5*, 86. https://doi.org/10.1186/1471-2105-5-86

Jutras, B.L., Verma, A., Adams, C.A., Brissette, C.A., Burns, L.H., Whetstine, C.R., Bowman, A., Chenail, A.M., Zuckert, W.R., and Stevenson, B. (2012). BpaB and EbfC DNA-binding proteins regulate production of the Lyme disease spirochete's infection-associated Erp surface proteins. J Bacteriol *194*, 778-786. https://doi.org/10.1128/JB.06394-11

Kawabata, H., Norris, S.J., and Watanabe, H. (2004). BBE02 disruption mutants of *Borrelia burgdorferi* B31 have a highly transformable, infectious phenotype. Infect Immun *72*, 7147-7154. https://doi.org/72/12/7147 [pii] 10.1128/IAI.72.12.7147-7154.2004

Kenedy, M.R., and Akins, D.R. (2011). The OspE-related proteins inhibit complement deposition and enhance serum resistance of *Borrelia burgdorferi*, the Lyme disease spirochete. Infect Immun *79*, 1451-1457. https://doi.org/10.1128/IAI.01274-10

Kingry, L.C., Anacker, M., Pritt, B., Bjork, J., Respicio-Kingry, L., Liu, G., Sheldon, S., Boxrud, D., Strain, A., Oatman, S*., et al.* (2018). Surveillance for and Discovery of *Borrelia* Species in US Patients Suspected of Tickborne Illness. Clin Infect Dis *66*, 1864-1871. https://doi.org/10.1093/cid/cix1107

Kingry, L.C., Batra, D., Replogle, A., Rowe, L.A., Pritt, B.S., and Petersen, J.M. (2016). Whole Genome Sequence and Comparative Genomics of the Novel Lyme Borreliosis Causing Pathogen, *Borrelia mayonii*. PLoS One *11*, e0168994. https://doi.org/10.1371/journal.pone.0168994

Kitten, T., and Barbour, A.G. (1990). Juxtaposition of expressed variable antigen genes with a conserved telomere in the bacterium *Borrelia hermsii*. Proc Natl Acad Sci U S A *87*, 6077-6081.

Kobryn, K., and Chaconas, G. (2002). ResT, a telomere resolvase encoded by the Lyme disease spirochete. Mol Cell *9*, 195-201. https://doi.org/S1097276501004336

Koenigs, A., Hammerschmidt, C., Jutras, B.L., Pogoryelov, D., Barthel, D., Skerka, C., Kugelstadt, D., Wallich, R., Stevenson, B., Zipfel, P.F*., et al.* (2013). BBA70 of *Borrelia burgdorferi* is a novel plasminogen-binding protein. J Biol Chem *288*, 25229-25243. https://doi.org/10.1074/jbc.M112.413872

Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., and Phillippy, A.M. (2016). Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. bioRxiv.

Kraiczy, P. (2016). Hide and Seek: How Lyme Disease Spirochetes Overcome Complement Attack. Frontiers in immunology *7*, 385. https://doi.org/10.3389/fimmu.2016.00385

Kraiczy, P., Hartmann, K., Hellwage, J., Skerka, C., Kirschfink, M., Brade, V., Zipfel, P.F., Wallich, R., and Stevenson, B. (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.

Krupna-Gaylord, M.A., Liveris, D., Love, A.C., Wormser, G.P., Schwartz, I., and Petzke, M.M. (2014). Induction of type I and type III interferons by *Borrelia burgdorferi* correlates with pathogenesis and requires linear plasmid 36. PLoS One *9*, e100174. https://doi.org/10.1371/journal.pone.0100174

Kuleshov, K.V., Margos, G., Fingerle, V., Koetsveld, J., Goptar, I.A., Markelov, M.L., Kolyasnikova, N.M., Sarksyan, D.S., Kirdyashkina, N.P., Shipulin, G.A., *et al.* (2020). Whole genome sequencing of *Borrelia miyamotoi* isolate Izh-4: reference for a complex bacterial genome. BMC Genomics *21*, 16. https://doi.org/10.1186/s12864-019-6388-4

Kulski, J.K. (2016). Next-Generation Sequencing — An Overview of the History, Tools, and "Omic" Applications. In Next Generation Sequencing - Advances, Applications and Challenges, J.K. Kulski, ed. (London UK: IntechOpen). https://doi.org/DOI: 10.5772/61964

Kumaran, D., Eswaramoorthy, S., Luft, B.J., Koide, S., Dunn, J.J., Lawson, C.L., and Swaminathan, S. (2001). Crystal structure of outer surface protein C (OspC) from the Lyme disease spirochete, *Borrelia burgdorferi*. EMBO J *20*, 971-978. https://doi.org/10.1093/emboj/20.5.971

Kumru, O.S., Schulze, R.J., Rodnin, M.V., Ladokhin, A.S., and Zuckert, W.R. (2011). Surface localization determinants of Borrelia OspC/Vsp family lipoproteins. J Bacteriol *193*, 2814-2825. https://doi.org/10.1128/JB.00015-11

Kurtenbach, K., De Michelis, S., Sewell, H.S., Etti, S., Schafer, S.M., Hails, R., Collares-Pereira, M., Santos-Reis, M., Hanincova, K., Labuda, M., *et al.* (2001). Distinct combinations of *Borrelia burgdorferi* sensu lato genospecies found in individual questing ticks from Europe. Appl Environ Microbiol *67*, 4926-4929.

Kurtenbach, K., Hanincova, K., Tsao, J.I., Margos, G., Fish, D., and Ogden, N.H. (2006). Fundamental processes in the evolutionary ecology of Lyme borreliosis. Nat Rev Microbiol *4*, 660-669. https://doi.org/10.1038/nrmicro1475

Labandeira-Rey, M., and Skare, J.T. (2001). Decreased infectivity in *Borrelia burgdorferi* strain B31 is associated with loss of linear plasmid 25 or 28-1. Infect Immun *69*, 446-455. https://doi.org/10.1128/IAI.69.1.446-455.2001

Lager, M., Faller, M., Wilhelmsson, P., Kjelland, V., Andreassen, A., Dargis, R., Quarsten, H., Dessau, R., Fingerle, V., Margos, G., *et al.* (2017). Molecular detection of *Borrelia burgdorferi* sensu lato - An analytical comparison of real-time PCR protocols from five different Scandinavian laboratories. PloS ONE *12*, e0185434. https://doi.org/10.1371/journal.pone.0185434

Lawrenz, M.B., Kawabata, H., Purser, J.E., and Norris, S.J. (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.

Lehnherr, H. (2006). Bacteriophage P1. In The Bacteriophages 2nd edition, S.T. Calendar, ed. (New York: Oxford University Press), pp. 350-364.

Levy, S.E., and Myers, R.M. (2016). Advancements in Next-Generation Sequencing. Annu Rev Genom Hum Gen *17*, 95-115. https://doi.org/10.1146/annurev-genom-083115-022413

Li, Y.F., Costello, J.C., Holloway, A.K., and Hahn, M.W. (2008). "Reverse ecology" and the power of population genomics. Evolution *62*, 2984-2994. https://doi.org/10.1111/j.1558-5646.2008.00486.x

Lin, T., Gao, L., Edmondson, D.G., Jacobs, M.B., Philipp, M.T., and Norris, S.J. (2009). Central role of the Holliday junction helicase RuvAB in vlsE recombination and infectivity of *Borrelia burgdorferi*. PLoS Pathogens *5*, e1000679. https://doi.org/10.1371/journal.ppat.1000679

Lin, T., Gao, L., Zhang, C., Odeh, E., Jacobs, M.B., Coutte, L., Chaconas, G., Philipp, M.T., and Norris, S.J. (2012). Analysis of an ordered, comprehensive STM mutant library in infectious *Borrelia burgdorferi*: insights into the genes required for mouse infectivity. PLoS One *7*, e47532. https://doi.org/10.1371/journal.pone.0047532

Lin, Y.P., Bhowmick, R., Coburn, J., and Leong, J.M. (2015). Host cell heparan sulfate glycosaminoglycans are ligands for OspF-related proteins of the Lyme disease spirochete. Cell Microbiol *17*, 1464-1476. https://doi.org/10.1111/cmi.12448

Liveris, D., Gazumyan, A., and Schwartz, I. (1995). Molecular typing of *Borrelia burgdorferi* sensu lato by PCR-restriction fragment length polymorphism analysis. J Clin Microbiol *33*, 589-595.

Liveris, D., Varde, S., Iyer, R., Koenig, S., Bittker, S., Cooper, D., McKenna, D., Nowakowski, J., Nadelman, R.B., Wormser, G.P., *et al.* (1999). Genetic diversity of *Borrelia burgdorferi* in Lyme disease patients as determined by culture versus direct PCR with clinical specimens. J Clin Microbiol *37*, 565-569.

Liveris, D., Hanincova, K. and Schwartz, I. (2013). Borreliae. In Molecular Typing in Bacterial Infections, I.M. de Fillipis, M.L. McKee, ed. (New York: Humana Press), pp. 353-369. https://doi.org/10.1007/978-1-62703-185-1

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

Lobocka, M.B., Rose, D.J., Plunkett, G., 3rd, Rusin, M., Samojedny, A., Lehnherr, H., Yarmolinsky, M.B., and Blattner, F.R. (2004). Genome of bacteriophage P1. J Bacteriol *186*, 7032-7068. https://doi.org/10.1128/JB.186.21.7032-7068.2004

Loza Reyes, E. (2010). Classification of phylogenetic data via Bayesian mixture modelling. PhD thesis University of Bath, UK.

Ludwig, A., von Rhein, C., Mischke, A., and Brade, V. (2007). Release of latent ClyA cytolysin from Escherichia coli mediated by a bacteriophage-associated putative holin (BlyA) from *Borrelia burgdorferi*. Int J Med Microbiol 298, 473-481. DOI: 10.1016/j.ijmm.2007.07.014

Luikart, G., England, P.R., Tallmon, D., Jordan, S., and Taberlet, P. (2003). The power and promise of population genomics: from genotyping to genome typing. Nat Rev Genet *4*, 981-994. https://doi.org/10.1038/nrg1226

Lynch, M. (2006). Streamlining and simplification of microbial genome architecture. Annu Rev Microbiol *60*, 327-349. https://doi.org/10.1146/annurev.micro.60.080805.142300

Maraspin, V., Mrvic, T., Ruzic-Sabljic, E., Jurcic, V., and Strle, F. (2019). Acrodermatitis chronica atrophicans in children: Report on two cases and review of the literature. Ticks Tick Borne Dis *10*, 180-185. https://doi.org/10.1016/j.ttbdis.2018.10.009

Marconi, R.T., Liveris, D., and Schwartz, I. (1995). Identification of novel insertion elements, restriction fragment length polymorphism patterns, and discontinuous 23S rRNA in Lyme disease spirochetes: phylogenetic analyses of rRNA genes and their intergenic spacers in *Borrelia japonica* sp. nov. and genomic group 21038 (*Borrelia andersonii* sp. nov.) isolates. J Clin Microbiol *33*, 2427-2434.

Marconi, R.T., Sung, S.Y., Hughes, C.A., and Carlyon, J.A. (1996). Molecular and evolutionary analyses of a variable series of genes in *Borrelia burgdorferi* that are related to *ospE* and *ospF*, constitute a gene family, and share a common upstream homology box. J Bacteriol *178*, 5615-5626.

Margolis, N., Hogan, D., Tilly, K., and Rosa, P.A. (1994). Plasmid location of *Borrelia* purine biosynthesis gene homologs. J Bacteriol *176*, 6427-6432.

Margos, G., Becker, N.S., Fingerle, V., Sing, A., Ramos, J.A., Carvalho, I.L., and Norte, A.C. (2019a). Core genome phylogenetic analysis of the avian associated Borrelia turdi indicates a close relationship to *Borrelia garinii*. Mol Phylogenet Evol *131*, 93-98. https://doi.org/10.1016/j.ympev.2018.10.044

Margos, G., Binder, K., Dzaferovic, E., Hizo-Teufel, C., Sing, A., Wildner, M., Fingerle, V., and Jolley, K.A. (2015a). PubMLST.org - The new home for the Borrelia MLSA database. Ticks Tick Borne Dis *6*, 869-871. https://doi.org/10.1016/j.ttbdis.2015.06.007

Margos, G., Chu, C.Y., Takano, A., Jiang, B.G., Liu, W., Kurtenbach, K., Masuzawa, T., Fingerle, V., Cao, W.C., and Kawabata, H. (2015b). *Borrelia yangtzensis* sp. nov. a rodent associated species in Asia is related to *B. valaisiana*. Int J Syst Evol Microbiol. https://doi.org/10.1099/ijsem.0.000491

Margos, G., Fedorova, N., Becker, N.S., Kleinjan, J.E., Marosevic, D., Krebs, S., Hui, L., Fingerle, V., and Lane, R.S. (2020). *Borrelia maritima* sp. nov., a novel species of the *Borrelia burgdorferi* sensu lato complex, occupying a basal position to North American species. Int J Syst Evol Microbiol *70*, 849-856. https://doi.org/10.1099/ijsem.0.003833

Margos, G., Fedorova, N., Kleinjan, J.E., Hartberger, C., Schwan, T.G., Sing, A., and Fingerle, V. (2017a). *Borrelia lanei* sp. nov. extends the diversity of *Borrelia* species in California. Int J Syst Evol Microbiol *67*, 3872-3876. https://doi.org/10.1099/ijsem.0.002214

Margos, G., Fingerle, V., and Reynolds, S. (2019b). *Borrelia bavariensis*: Vector Switch, Niche Invasion, and Geographical Spread of a Tick-Borne Bacterial Parasite. Frontiers in Ecology and Evolution *7*. https://doi.org/10.3389/fevo.2019.00401

Margos, G., Gatewood, A.G., Aanensen, D.M., Hanincova, K., Terekhova, D., Vollmer, S.A., Cornet, M., Piesman, J., Donaghy, M., Bormane, A., *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. https://doi.org/10.1073/pnas.0800323105

Margos, G., Gofton, A., Wibberg, D., Dangel, A., Marosevic, D., Loh, S.M., Oskam, C., and Fingerle, V. (2018). The genus *Borrelia* reloaded. PLoS One *13*, e0208432. https://doi.org/10.1371/journal.pone.0208432

Margos, G., Hepner, S., Mang, C., Marosevic, D., Reynolds, S.E., Krebs, S., Sing, A., Derdakova, M., Reiter, M.A., and Fingerle, V. (2017b). Lost in plasmids: next generation sequencing and the complex genome of the tick-borne pathogen *Borrelia burgdorferi*. BMC Genomics *18*, 422. https://doi.org/10.1186/s12864-017-3804-5

Margos, G., Hepner, S., Mang, C., Sing, A., Liebl, B., and Fingerle, V. (2017c). Completed Genome Sequences of *Borrelia burgdorferi* Sensu Stricto B31(NRZ) and Closely Related Patient Isolates from Europe. Genome Announc *5*. https://doi.org/10.1128/genomeA.00637-17

Margos, G., Hojgaard, A., Lane, R.S., Cornet, M., Fingerle, V., Rudenko, N., Ogden, N., Aanensen, D.M., Fish, D., and Piesman, J. (2010). Multilocus sequence analysis of *Borrelia bissettii* strains from North America reveals a new *Borrelia* species, *Borrelia kurtenbachii*. Ticks Tick Borne Dis *1*, 151-158.

Margos, G., Lane, R.S., Fedorova, N., Koloczek, J., Piesman, J., Hojgaard, A., Sing, A., and Fingerle, V. (2016). *Borrelia bissettiae* sp. nov. and *Borrelia californiensis* sp. nov. prevail in diverse enzootic transmission cycles. Int J Syst Evol Microbiol *66*, 1447-1452. https://doi.org/10.1099/ijsem.0.000897

Margos, G., Sing, A., and Fingerle, V. (2017d). Published data do not support the notion that *Borrelia valaisiana* is human pathogenic. Infection *45*, 567-569. https://doi.org/10.1007/s15010-017-1032-1

Margos, G., Tsao, J.I., Castillo-Ramirez, S., Girard, Y.A., Hamer, S.A., Hoen, A.G., Lane, R.S., Raper, S.L., and Ogden, N.H. (2012). Two boundaries separate *Borrelia burgdorferi* populations in North America. Appl Environ Microbiol *78*, 6059-6067. https://doi.org/10.1128/AEM.00231-12

Margos, G., Vollmer, S.A., Cornet, M., Garnier, M., Fingerle, V., Wilske, B., Bormane, A., Vitorino, L., Collares-Pereira, M., Drancourt, M., *et al.* (2009). A new *Borrelia* species defined by Multilocus Sequence Analysis of Housekeeping Genes. Appl Environ Microbiol *75*, 5410-5416.

Margos, G., Vollmer, S.A., Ogden, N.H., and Fish, D. (2011). Population genetics, taxonomy, phylogeny and evolution of *Borrelia burgdorferi* sensu lato. Infect Genet Evol *11*, 1545-1563.

Margos, G., Wilske, B., Sing, A., Hizo-Teufel, C., Cao, W.C., Chu, C., Scholz, H., Straubinger, R.K., and Fingerle, V. (2013). *Borrelia bavariensis* sp. nov. is widely distributed in Europe and Asia. Int J Syst Evol Microbiol *63*, 4284-4288.

Marttinen, P., Hanage, W.P., Croucher, N.J., Connor, T.R., Harris, S.R., Bentley, S.D., and Corander, J. (2012). Detection of recombination events in bacterial genomes from large population samples. Nucleic Acids Res *40*, e6. https://doi.org/10.1093/nar/gkr928

McDowell, J.V., Wolfgang, J., Senty, L., Sundy, C.M., Noto, M.J., and Marconi, R.T. (2004). Demonstration of the involvement of outer surface protein E coiled coil structural domains and higher order structural elements in the binding of infection-induced antibody and the complement-regulatory protein, factor H. J Immunol *173*, 7471-7480.

Mechai, S., Margos, G., Feil, E.J., Barairo, N., Lindsay, L.R., Michel, P., and Ogden, N.H. (2016). Evidence for Host-Genotype Associations of *Borrelia burgdorferi* Sensu Stricto. PLoS One *11*, e0149345. https://doi.org/10.1371/journal.pone.0149345

Mechai, S., Margos, G., Feil, E.J., Lindsay, L.R., and Ogden, N.H. (2015). Complex population structure of *Borrelia burgdorferi* in southeastern and south central Canada as revealed by phylogeographic analysis. Appl Environ Microbiol *81*, 1309-1318. https://doi.org/10.1128/AEM.03730-14

Metts, M.S., McDowell, J.V., Theisen, M., Hansen, P.R., and Marconi, R.T. (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.

Milkman, R., and Bridges, M.M. (1990). Molecular evolution of the *Escherichia coli* chromosome. III. Clonal frames. Genetics *126*, 505-517.

Miller, J., Bono, J., Babb, K., El-Hage, N., Casjens, S., and Stevenson, B. (2000). A second allele of *eppA* in *Borrelia burgdorferi* strain B31 is located on the previously undetected circular plasmid cp9-2. J Bacteriol *182*, 6254-6258.

Miller, J.C., and Stevenson, B. (2003). Immunological and genetic characterization of *Borrelia burgdorferi* BapA and EppA proteins. Microbiology *149*, 1113-1125.

Miller, J.C., and Stevenson, B. (2004). Increased expression of *Borrelia burgdorferi* factor H-binding surface proteins during transmission from ticks to mice. Int J Med Microbiol *293 Suppl 37*, 120-125.

Molloy, E.M., Casjens, S.R., Cox, C.L., Maxson, T., Ethridge, N.A., Margos, G., Fingerle, V., and Mitchell, D.A. (2015). Identification of the minimal cytolytic unit for streptolysin S and an expansion of the toxin family. BMC Microbiol *15*, 141. https://doi.org/10.1186/s12866-015-0464-y

Mongodin, E.F., Casjens, S.R., Bruno, J.F., Xu, Y., Drabek, E.F., Riley, D.R., Cantarel, B.L., Pagan, P.E., Hernandez, Y.A., Vargas, L.C., *et al.* (2013). Inter- and intra-specific pan-genomes of *Borrelia burgdorferi* sensu lato: genome stability and adaptive radiation. BMC Genomics *14*, 693. https://doi.org/10.1186/1471-2164-14-693

Mukhacheva, T.A., and Kovalev, S.Y. (2013). Multilocus sequence analysis of *Borrelia burgdorferi* s.l. in Russia. Ticks Tick Borne Dis *4*, 275-279.

Mulay, V.B., Caimano, M.J., Iyer, R., Dunham-Ems, S., Liveris, D., Petzke, M.M., Schwartz, I., and Radolf, J.D. (2009). *Borrelia burgdorferi bba74* is expressed exclusively during tick feeding and is regulated by both arthropod- and mammalian host-specific signals. J Bacteriol *191*, 2783-2794. https://doi.org/10.1128/JB.01802-08

Muller, H.J. (1964). The Relation of Recombination to Mutational Advance. Mutat Res *106*, 2-9. https://doi.org/10.1016/0027-5107(64)90047-8

Muniesa, M., Imamovic, L., and Jofre, J. (2011). Bacteriophages and genetic mobilization in sewage and faecally polluted environments. Microb Biotechnol *4*, 725-734. https://doi.org/10.1111/j.1751-7915.2011.00264.x

Munro, H.J., Ogden, N.H., Mechai, S., Lindsay, L.R., Robertson, G.J., Whitney, H., and Lang, A.S. (2019). Genetic diversity of *Borrelia garinii* from *Ixodes uriae* collected in seabird colonies of the

northwestern Atlantic Ocean. Ticks Tick Borne Dis *10*, 101255. https://doi.org/10.1016/j.ttbdis.2019.06.014

Nei, M., and Rooney, A.P. (2005). Concerted and birth-and-death evolution of multigene families. Annu Rev Genet *39*, 121-152. https://doi.org/10.1146/annurev.genet.39.073003.112240

Neubert, U., Schaller, M., Januschke, E., Stolz, W., and Schmieger, H. (1993). Bacteriophages induced by ciprofloxacin in a *Borrelia burgdorferi* skin isolate. Zentralbl Bakteriol *279*, 307-315.

Nguyen, K.T., Wu, J.C., Boylan, J.A., Gherardini, F.C., and Pei, D. (2007). Zinc is the metal cofactor of *Borrelia burgdorferi* peptide deformylase. Arch Biochem Biophys *468*, 217-225. https://doi.org/S0003-9861(07)00483-3 [pii] 10.1016/j.abb.2007.09.023

Niedringhaus, T.P., Milanova, D., Kerby, M.B., Snyder, M.P., and Barron, A.E. (2011). Landscape of next-generation sequencing technologies. Anal Chem *83*, 4327-4341. https://doi.org/10.1021/ac2010857

Norris, D.E., Johnson, B.J., Piesman, J., Maupin, G.O., Clark, J.L., and Black, W.C.t. (1997). Culturing selects for specific genotypes of *Borrelia burgdorferi* in an enzootic cycle in Colorado. J Clin Microbiol *35*, 2359-2364.

Norris, S.J. (2006). Antigenic variation with a twist--the *Borrelia* story. Mol Microbiol *60*, 1319-1322. https://doi.org/10.1111/j.1365-2958.2006.05204.x

Norris, S.J. (2014). *vls* Antigenic Variation Systems of Lyme Disease *Borrelia*: Eluding Host Immunity through both Random, Segmental Gene Conversion and Framework Heterogeneity. Microbiol Spectr *2*, 1. https://doi.org/10.1128/microbiolspec.MDNA3-0038-2014

Norte, A.C., Margos, G., Becker, N.S., Albino Ramos, J., Nuncio, M.S., Fingerle, V., Araujo, P.M., Adamik, P., Alivizatos, H., Barba, E., *et al.* (2020). Host dispersal shapes the population structure of a tick-borne bacterial pathogen. Mol Ecol *29*, 485-501. https://doi.org/10.1111/mec.15336

Ogden, N.H., Bouchard, C., Kurtenbach, K., Margos, G., Lindsay, L.R., Trudel, L., Nguon, S., and Milord, F. (2010). Active and passive surveillance and phylogenetic analysis of *Borrelia burgdorferi* elucidate the process of Lyme disease risk emergence in Canada. Environ Health Perspect *118*, 909-914.

Ogden, N.H., Maarouf, A., Barker, I.K., Bigras-Poulin, M., Lindsay, L.R., Morshed, M.G., O'Callaghan C, J., Ramay, F., Waltner-Toews, D., and Charron, D.F. (2006). Climate change and the potential for range expansion of the Lyme disease vector *Ixodes scapularis* in Canada. Int J Parasitol *36*, 63-70.

Ogden, N.H., Margos, G., Aanensen, D.M., Drebot, M.A., Feil, E.J., Hanincova, K., Schwartz, I., Tyler, S., and Lindsay, L.R. (2011). Investigation of genotypes of *Borrelia burgdorferi* in *Ixodes scapularis* ticks collected during surveillance in Canada. Appl Environ Microbiol *77*, 3244-3254. https://doi.org/10.1128/AEM.02636-10

Ojaimi, C., Brooks, C., Casjens, S., Rosa, P., Elias, A., Barbour, A., Jasinskas, A., Benach, J., Katona, L., Radolf, J., *et al.* (2003). Profiling of temperature-induced changes in *Borrelia burgdorferi* gene expression by using whole genome arrays. Infect Immun *71*, 1689-1705.

Ojaimi, C., Davidson, B.E., Saint Girons, I., and Old, I.G. (1994). Conservation of gene arrangement and an unusual organization of rRNA genes in the linear chromosomes of the Lyme disease spirochaetes *Borrelia burgdorferi*, B. garinii and B. afzelii. Microbiology *140*, 2931-2940. https://doi.org/10.1099/13500872-140-11-2931

Ostfeld, R.S., Levi, T., Keesing, F., Oggenfuss, K., and Canham, C.D. (2018). Tick-borne disease risk in a forest food web. Ecology *99*, 1562-1573. https://doi.org/10.1002/ecy.2386

Ouyang, Z., He, M., Oman, T., Yang, X.F., and Norgard, M.V. (2009). A manganese transporter, BB0219 (BmtA), is required for virulence by the Lyme disease spirochete, *Borrelia burgdorferi*. Proc Natl Acad Sci U S A *106*, 3449-3454. https://doi.org/10.1073/pnas.0812999106

Pal, U., Li, X., Wang, T., Montgomery, R.R., Ramamoorthi, N., Desilva, A.M., Bao, F., Yang, X., Pypaert, M., Pradhan, D., *et al.* (2004). TROSPA, an *Ixodes scapularis* receptor for *Borrelia burgdorferi*. Cell *119*, 457-468. https://doi.org/10.1016/j.cell.2004.10.027

Palmer, N., Fraser, C., and Casjens, S. (2000). Distribution of twelve linear extrachromosomal DNAs in natural isolates of Lyme disease spirochetes. J Bacteriol *182*, 2476-2480. https://doi.org/10.1128/jb.182.9.2476-2480.2000

Paradis, E., Claude, J., and Strimmer, K. (2004). APE: Analyses of Phylogenetics and Evolution in R language. Bioinformatics *20*, 289-290. https://doi.org/10.1093/bioinformatics/btg412

Paster, B.J., Dewhirst, F.E., Weisburg, W.G., Tordoff, L.A., Fraser, G.J., Hespell, R.B., Stanton, T.B., Zablen, L., Mandelco, L., and Woese, C.R. (1991). Phylogenetic analysis of the spirochetes. J Bacteriol *173*, 6101-6109.

Penades, J.R., Chen, J., Quiles-Puchalt, N., Carpena, N., and Novick, R.P. (2015). Bacteriophage-mediated spread of bacterial virulence genes. Curr Opin Microbiol *23*, 171-178. https://doi.org/10.1016/j.mib.2014.11.019

Phelan, J.P., Kern, A., Ramsey, M.E., Lundt, M.E., Sharma, B., Lin, T., Gao, L., Norris, S.J., Hyde, J.A., Skare, J.T., *et al.* (2019). Genome-wide screen identifies novel genes required for *Borrelia burgdorferi* survival in its *Ixodes* tick vector. PLoS Pathogens *15*, e1007644. https://doi.org/10.1371/journal.ppat.1007644

Picardeau, M., Lobry, J.R., and Hinnebusch, B.J. (1999). Physical mapping of an origin of bidirectional replication at the centre of the *Borrelia burgdorferi* linear chromosome. Mol Microbiol *32*, 437-445. https://doi.org/10.1046/j.1365-2958.1999.01368.x

Plasterk, R.H., Simon, M.I., and Barbour, A.G. (1985). Transposition of structural genes to an expression sequence on a linear plasmid causes antigenic variation in the bacterium *Borrelia hermsii*. Nature *318*, 257-263.

Porcella, S.F., Fitzpatrick, C.A., and Bono, J.L. (2000). Expression and immunological analysis of the plasmid-borne *mlp* genes of *Borrelia burgdorferi* strain B31. Infect Immun *68*, 4992-5001.

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

Posey, J.E., and Gherardini, F.C. (2000). Lack of a role for iron in the Lyme disease pathogen. Science *288*, 1651-1653.

Postic, D., Assous, M.V., Grimont, P.A., and Baranton, G. (1994). Diversity of *Borrelia burgdorferi* sensu lato evidenced by restriction fragment length polymorphism of *rrf* (5S)-*rrl* (23S) intergenic spacer amplicons. Int J Syst Bacteriol *44*, 743-752.

Postic, D., Garnier, M., and Baranton, G. (2007). Multilocus sequence analysis of atypical *Borrelia burgdorferi* sensu lato isolates - description of *Borrelia californiensis* sp. nov., and genomospecies 1 and 2. Int J Med Microbiol *297*, 263-271.

Postic, D., Ras, N.M., Lane, R.S., Hendson, M., and Baranton, G. (1998). Expanded diversity among Californian *Borrelia* isolates and description of *Borrelia bissettii* sp. nov. (formerly *Borrelia* group DN127). J Clin Microbiol *36*, 3497-3504.

Potter, S.C., Luciani, A., Eddy, S.R., Park, Y., Lopez, R., and Finn, R.D. (2018). HMMER web server: 2018 update. Nucleic Acids Res *46*, W200-W204. https://doi.org/10.1093/nar/gky448

Prevelige, P.E., Jr., and Cortines, J.R. (2018). Phage assembly and the special role of the portal protein. Curr Opin Virol *31*, 66-73. https://doi.org/10.1016/j.coviro.2018.09.004

Price, M.N., Dehal, P.S., and Arkin, A.P. (2010). FastTree 2--approximately maximum-likelihood trees for large alignments. PLoS One *5*, e9490. https://doi.org/10.1371/journal.pone.0009490

Pritt, B.S., Mead, P.S., Johnson, D.K.H., Neitzel, D.F., Respicio-Kingry, L.B., Davis, J.P., Schiffman, E., Sloan, L.M., Schriefer, M.E., Replogle, A.J., *et al.* (2016). Identification of a novel pathogenic *Borrelia* species causing Lyme borreliosis with unusually high

spirochaetaemia: a descriptive study. Lancet Infect Dis *16*, 556-564. https://doi.org/10.1016/S1473-3099(15)00464-8

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

Promnares, K., Kumar, M., Shroder, D.Y., Zhang, X., Anderson, J.F., and Pal, U. (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. https://doi.org/10.1111/j.1365-2958.2009.06853.x

Purser, J.E., Lawrenz, M.B., Caimano, M.J., Howell, J.K., Radolf, J.D., and Norris, S.J. (2003). A plasmid-encoded nicotinamidase (PncA) is essential for infectivity of *Borrelia burgdorferi* in a mammalian host. Mol Microbiol *48*, 753-764. https://doi.org/3452

Purser, J.E., and Norris, S.J. (2000). Correlation between plasmid content and infectivity in *Borrelia burgdorferi*. Proc Natl Acad Sci USA *97*, 13865-13870. https://doi.org/Doi 10.1073/Pnas.97.25.13865

Qiu, W.G., Bosler, E.M., Campbell, J.R., Ugine, G.D., Wang, I.N., Luft, B.J., and Dykhuizen, D.E. (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. Hereditas *127*, 203-216.

Qiu, W.G., Bruno, J.F., McCaig, W.D., Xu, Y., Livey, I., Schriefer, M.E., and Luft, B.J. (2008). Wide distribution of a high-virulence *Borrelia burgdorferi* clone in Europe and North America. Emerg Infect Dis *14*, 1097-1104.

Qiu, W.G., Dykhuizen, D.E., Acosta, M.S., and Luft, B.J. (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.

Qiu, W.G., and Martin, C.L. (2014). Evolutionary genomics of *Borrelia burgdorferi* sensu lato: findings, hypotheses, and the rise of hybrids. Infect Genet Evol *27*, 576-593. https://doi.org/10.1016/j.meegid.2014.03.025

Qiu, W.G., Schutzer, S.E., Bruno, J.F., Attie, O., Xu, Y., Dunn, J.J., Fraser, C.M., Casjens, S.R., and Luft, B.J. (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.

Quereda, J.J., Nahori, M.A., Meza-Torres, J., Sachse, M., Titos-Jimenez, P., Gomez-Laguna, J., Dussurget, O., Cossart, P., and Pizarro-Cerda, J. (2017). Listeriolysin S Is a Streptolysin S-Like Virulence Factor That Targets Exclusively Prokaryotic Cells In Vivo. MBio *8*. https://doi.org/10.1128/mBio.00259-17

Radolf, J.D., Caimano, M.J., Stevenson, B., and Hu, L.T. (2012). Of ticks, mice and men: understanding the dual-host lifestyle of Lyme disease spirochaetes. NatRevMicrobiol *10*, 87-99.

Radolf J.D. and Samuels, D.S. (2021). Lyme Disease and Relapsing Fever Spirochetes: Genomics, Molecular Biology, Host Interactions and Disease Pathogenesis (Norfolk, UK: Caister Academic Press). https://doi.org/10.21775/9781913652616

Ramsey, M.E., Hyde, J.A., Medina-Perez, D.N., Lin, T., Gao, L., Lundt, M.E., Li, X., Norris, S.J., Skare, J.T., and Hu, L.T. (2017). A high-throughput genetic screen identifies previously uncharacterized *Borrelia burgdorferi* genes important for resistance against reactive oxygen and nitrogen species. PLoS Pathogens *13*, e1006225. https://doi.org/10.1371/journal.ppat.1006225

Ranka, R., Bormane, A., Salmina, K., and Baumanis, V. (2004). Identification of three clinically relevant *Borrelia burgdorferi* sensu lato genospecies by PCR-restriction fragment length polymorphism analysis of 16S-23S ribosomal DNA spacer amplicons. J Clin Microbiol *42*, 1444-1449. https://doi.org/10.1128/jcm.42.4.1444-1449.2004

Rannala, B., Qiu, W.G., and Dykhuizen, D.E. (2000). Methods for estimating gene frequencies and detecting selection in bacterial populations. Genetics *155*, 499-508.

Ras, N.M., Lascola, B., Postic, D., Cutler, S.J., Rodhain, F., Baranton, G., and Raoult, D. (1996). Phylogenesis of relapsing fever *Borrelia* spp. Int J Syst Bacteriol *46*, 859-865. https://doi.org/10.1099/00207713-46-4-859

Rauter, C., and Hartung, T. (2005). Prevalence of *Borrelia burgdorferi* sensu lato genospecies in *Ixodes* ricinus ticks in Europe: a metaanalysis. Appl Environ Microbiol *71*, 7203-7216. https://doi.org/10.1128/AEM.71.11.7203-7216.2005

Ravin, V., Ravin, N., Casjens, S., Ford, M.E., Hatfull, G.F., and Hendrix, R.W. (2000). Genomic sequence and analysis of the atypical temperate bacteriophage N15. J Mol Biol *299*, 53-73.

Rego, R.O., Bestor, A., and Rosa, P.A. (2011). Defining the plasmid-borne restriction-modification systems of the Lyme disease spirochete *Borrelia burgdorferi*. J Bacteriol *193*, 1161-1171. https://doi.org/JB.01176-10 [pii] 10.1128/JB.01176-10

Revel, A.T., Blevins, J.S., Almazan, C., Neil, L., Kocan, K.M., de la Fuente, J., Hagman, K.E., and Norgard, M.V. (2005). *bptA* (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. https://doi.org/0502565102 [pii] 10.1073/pnas.0502565102

Richter, D., Postic, D., Sertour, N., Livey, I., Matuschka, F.R., and Baranton, G. (2006). Delineation of *Borrelia burgdorferi* sensu lato species by multilocus sequence analysis and confirmation of the delineation of *Borrelia spielmanii* sp. nov. Int J Syst Evol Microbiol *56*, 873-881.

Rijpkema, S.G., Molkenboer, M.J., Schouls, L.M., Jongejan, F., and Schellekens, J.F. (1995). Simultaneous detection and genotyping of three genomic groups of *Borrelia burgdorferi* sensu lato in Dutch *Ixodes ricinus* ticks by characterization of the amplified intergenic spacer region between 5S and 23S rRNA genes. J Clin Microbiol *33*, 3091-3095.

Roberts, D.M., Theisen, M., and Marconi, R.T. (2000). Analysis of the cellular localization of Bdr paralogs in *Borrelia burgdorferi*, a causative agent of lyme disease: evidence for functional diversity. J Bacteriol *182*, 4222-4226.

Rocha, E.P.C. (2018). Neutral Theory, Microbial Practice: Challenges in Bacterial Population Genetics. Mol Biol Evol *35*, 1338-1347. https://doi.org/10.1093/molbev/msy078

Sabitova, Y., Fomenko, N., Tikunov, A., Stronin, O., Khasnatinov, M., Abmed, D., Danchinova, G., Golovljova, I., and Tikunova, N. (2018). Multilocus sequence analysis of *Borrelia burgdorferi* sensu lato isolates from Western Siberia, Russia and Northern Mongolia. Infect Genet Evol *62*, 160-169. https://doi.org/10.1016/j.meegid.2018.04.015

Sadziene, A., Rosa, P., Thompson, P., Hogan, D., and Barbour, A. (1992). Antibody-resistant mutants of *Borrelia burgdorferi*: in vitro selection and characterization. J Exp Med *176*, 799-809.

Sadziene, A., Wilske, B., Ferdows, M.S., and Barbour, A.G. (1993). The cryptic ospC gene of *Borrelia burgdorferi* B31 is located on a circular plasmid. Infect Immun *61*, 2192-2195.

Saint Girons, I., Gern, L., Gray, J.S., Guy, E.C., Korenberg, E., Nuttall, P.A., Rijpkema, S.G., Schonberg, A., Stanek, G., and Postic, D. (1998). Identification of *Borrelia burgdorferi* sensu lato species in Europe. Zentralbl Bakteriol *287*, 190-195.

Salo, J., Jaatinen, A., Soderstrom, M., Viljanen, M.K., and Hytonen, J. (2015). Decorin binding proteins of *Borrelia burgdorferi* promote arthritis development and joint specific post-treatment DNA persistence in mice. PLoS One *10*, e0121512. https://doi.org/10.1371/journal.pone.0121512

Samuels, D., Marconi, R., and 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.

Samuels, D.S. (2011). Gene regulation in *Borrelia burgdorferi*. Annu Rev Microbiol *65*, 479-499. https://doi.org/10.1146/annurev.micro.112408.134040

Schaller, M., and Neubert, U. (1994). Bacteriophages and ultrastructural alterations of *Borrelia burgdorferi* induced by ciprofloxacin. J Spirochetal Tick-borne Dis *1*, 37-40.

Schulte-Spechtel, U., Fingerle, V., Goettner, G., Rogge, S., and 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.

Schutzer, S.E., Fraser-Liggett, C.M., Casjens, S.R., Qiu, W.G., Dunn, J.J., Mongodin, E.F., and Luft, B.J. (2011). Whole-genome sequences of thirteen isolates of *Borrelia burgdorferi*. J Bacteriol *193*, 1018-1020. https://doi.org/JB.01158-10 [pii] 10.1128/JB.01158-10

Schutzer, S.E., Fraser-Liggett, C.M., Qiu, W.G., Kraiczy, P., Mongodin, E.F., Dunn, J.J., Luft, B.J., and Casjens, S.R. (2012). Whole-genome sequences of *Borrelia bissettii*, *Borrelia valaisiana*, and *Borrelia spielmanii*. J Bacteriol *194*, 545-546. https://doi.org/10.1128/JB.06263-11

Schwan, T.G., Burgdorfer, W., and Garon, C.F. (1988). Changes in infectivity and plasmid profile of the Lyme disease spirochete, *Borrelia burgdorferi*, as a result of in vitro cultivation. Infect Immun *56*, 1831-1836.

Schwan, T.G., Raffel, S.J., Schrumpf, M.E., and Porcella, S.F. (2007). Diversity and distribution of *Borrelia hermsii*. Emerg Infect Dis *13*, 436-442.

Schwartz, A.M., Hinckley, A.F., Mead, P.S., Hook, S.A., and Kugeler, K.J. (2017). Surveillance for Lyme Disease - United States, 2008-2015. MMWR Surveill Summ *66*, 1-12. https://doi.org/10.15585/mmwr.ss6622a1

Schwartz, J.J., Gazumyan, A., and Schwartz, I. (1992). rRNA gene organization in the Lyme disease spirochete, *Borrelia burgdorferi*. J Bacteriol *174*, 3757-3765.

Seifert, S.N., Khatchikian, C.E., Zhou, W., and Brisson, D. (2015). Evolution and population genomics of the Lyme borreliosis pathogen, *Borrelia burgdorferi*. Trends Genet *31*, 201-207. https://doi.org/10.1016/j.tig.2015.02.006

Seinost, G., Dykhuizen, D.E., Dattwyler, R.J., Golde, W.T., Dunn, J.J., Wang, I.N., Wormser, G.P., Schriefer, M.E., and Luft, B.J. (1999). Four clones of *Borrelia burgdorferi* sensu stricto cause invasive infection in humans. Infect Immun *67*, 3518-3524.

Seshu, J., Esteve-Gassent, M.D., Labandeira-Rey, M., Kim, J.H., Trzeciakowski, J.P., Hook, M., and Skare, J.T. (2006). Inactivation of the fibronectin-binding adhesin gene *bbk32* significantly attenuates the infectivity potential of *Borrelia burgdorferi*. Mol Microbiol *59*, 1591-1601. https://doi.org/MMI5042 [pii] 10.1111/j.1365-2958.2005.05042.x

Shi, Y., Xu, Q., McShan, K., and Liang, F.T. (2008). Both decorin-binding proteins A and B are critical for the overall virulence of *Borrelia burgdorferi*. Infect Immun *76*, 1239-1246. https://doi.org/IAI.00897-07 [pii] 10.1128/IAI.00897-07

Shih, Y.L., and Rothfield, L. (2006). The bacterial cytoskeleton. Microbiol Mol Biol Rev *70*, 729-754.

Skare, J.T., Foley, D.M., Hernandez, S.R., Moore, D.C., Blanco, D.R., Miller, J.N., and Lovett, M.A. (1999). Cloning and molecular characterization of plasmid-encoded antigens of *Borrelia burgdorferi*. Infect Immun *67*, 4407-4417.

Smith, J.M., Smith, N.H., O'Rourke, M., and Spratt, B.G. (1993). How clonal are bacteria? Proc Natl Acad Sci U S A *90*, 4384-4388. https://doi.org/10.1073/pnas.90.10.4384

Smith, R.P., Jr., Muzaffar, S.B., Lavers, J., Lacombe, E.H., Cahill, B.K., Lubelczyk, C.B., Kinsler, A., Mathers, A.J., and Rand, P.W. (2006). *Borrelia garinii* in seabird ticks (*Ixodes uriae*), Atlantic Coast, North America. Emerg Infect Dis *12*, 1909-1912.

Spielman, A. (1994). The emergence of Lyme disease and human babesiosis in a changing environment. Ann N Y Acad Sci *740*, 146-156.

Spratt, B.G. (1999). Multilocus sequence typing: molecular typing of bacterial pathogens in an era of rapid DNA sequencing and the Internet. Curr Opin Microbiol *2*, 312-316. https://doi.org/http://biomednet.com/elecref/1369527400200312

Springer, A., Raulf, M.K., Fingerle, V., and Strube, C. (2020). Borrelia prevalence and species distribution in ticks removed from humans in Germany, 2013-2017. Ticks Tick Borne Dis *11*, 101363. https://doi.org/10.1016/j.ttbdis.2019.101363

Stanek, G., and Reiter, M. (2011). The expanding Lyme Borrelia complex--clinical significance of genomic species? Clin Microbiol Infect *17*, 487-493. https://doi.org/10.1111/j.1469-0691.2011.03492.x

Stanton, T.B. (2007). Prophage-like gene transfer agents-novel mechanisms of gene exchange for *Methanococcus*, *Desulfovibrio*, *Brachyspira*, and *Rhodobacter* species. Anaerobe *13*, 43-49. https://doi.org/10.1016/j.anaerobe.2007.03.004

States, S.L., Brinkerhoff, R.J., Carpi, G., Steeves, T.K., Folsom-O'Keefe, C., DeVeaux, M., and Diuk-Wasser, M.A. (2014). Lyme disease risk not amplified in a species-poor vertebrate community: similar *Borrelia burgdorferi* tick infection prevalence and OspC genotype frequencies. Infect Genet Evol *27*, 566-575. https://doi.org/10.1016/j.meegid.2014.04.014

Stete, K., Rieg, S., Margos, G., Hacker, G., Wagner, D., Kern, W.V., and Fingerle, V. (2018). Case Report and Genetic Sequence Analysis of *Candidatus Borrelia kalaharica*, Southern Africa. Emerg Infect Dis *24*, 1659-1664. https://doi.org/10.3201/eid2409.171381

Stevenson, B., Bykowski, T., Cooley, A.E., Babb, K., Miller, J.C., Woodman, M.E., von Lackum, K., and Riley, S.P. (2006). The Lyme disease spirochete Erp protein family: structure, function and regulation of expression. In Moleculr biology of spirochetes, F.C. Cabello, D. Hulinska, and H.P. Godfrey, eds. (Amsterdam: IOS Press), pp. 354-372.

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

Stevenson, B., El-Hage, N., Hines, M.A., Miller, J.C., and Babb, K. (2002). Differential binding of host complement inhibitor factor H by *Borrelia burgdorferi* Erp surface proteins: a possible mechanism underlying the expansive host range of Lyme disease spirochetes. Infect Immun *70*, 491-497.

Stevenson, B., and Miller, J.C. (2003). Intra- and interbacterial genetic exchange of Lyme disease spirochete *erp* genes generates sequence identity amidst diversity. J Mol Evol *57*, 309-324.

Stevenson, B., Porcella, S.F., Oie, K.L., Fitzpatrick, C.A., Raffel, S.J., Lubke, L., Schrumpf, M.E., and Schwan, T.G. (2000a). The relapsing fever spirochete *Borrelia hermsii* contains multiple, antigen-encoding circular plasmids that are homologous to the cp32 plasmids of Lyme disease spirochetes. Infect Immun *68*, 3900-3908.

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

Stevenson, B., Zückert, W., and Akins, D. (2000b). Repetition, conservation, and variation: the multiple cp32 plasmids of *Borrelia* species. J Molec Microbiol Biotech, in press.

Stewart, P.E., Byram, R., Grimm, D., Tilly, K., and Rosa, P.A. (2005). The plasmids of *Borrelia burgdorferi*: essential genetic elements of a pathogen. Plasmid *53*, 1-13. https://doi.org/S0147-619X(04)00122-2 [pii] 10.1016/j.plasmid.2004.10.006

Stewart, P.E., Thalken, R., Bono, J.L., and Rosa, P. (2001). Isolation of a circular plasmid region sufficient for autonomous replication and transformation of infectious *Borrelia burgdorferi*. Mol Microbiol *39*, 714-721. https://doi.org/10.1046/j.1365-2958.2001.02256.x

Strandh, M., and Raberg, L. (2015). Within-host competition between *Borrelia afzelii ospC* strains in wild hosts as revealed by massively parallel amplicon sequencing. Philos Trans R Soc Lond B Biol Sci *370*. https://doi.org/10.1098/rstb.2014.0293

Strother, K.O., Broadwater, A., and De Silva, A. (2005). Plasmid requirements for infection of ticks by *Borrelia burgdorferi*. Vector Borne Zoonotic Dis *5*, 237-245. https://doi.org/10.1089/vbz.2005.5.237

Strother, K.O., and de Silva, A. (2005). Role of *Borrelia burgdorferi* linear plasmid 25 in infection of *Ixodes scapularis* ticks. J Bacteriol *187*, 5776-5781. https://doi.org/10.1128/JB.187.16.5776-5781.2005

Surtees, J.A., and Funnell, B.E. (2003). Plasmid and chromosome traffic control: how ParA and ParB drive partition. Curr Top Dev Biol *56*, 145-180.

Terekhova, D., Iyer, R., Wormser, G.P., and 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. https://doi.org/188/17/6124 [pii] 10.1128/JB.00459-06

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.

Thierauf, A., Perez, G., and Maloy, A.S. (2009). Generalized transduction. Methods Mol Biol *501*, 267-286. https://doi.org/10.1007/978-1-60327-164-6_23

Tilly, K., Bestor, A., and Rosa, P.A. (2013). Lipoprotein succession in *Borrelia burgdorferi*: similar but distinct roles for OspC and VlsE at different stages of mammalian infection. Mol Microbiol *89*, 216-227. https://doi.org/10.1111/mmi.12271

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

Tilly, K., Checroun, C., and Rosa, P.A. (2012). Requirements for *Borrelia burgdorferi* plasmid maintenance. Plasmid. https://doi.org/S0147-619X(12)00013-3 [pii] 10.1016/j.plasmid.2012.01.009

Tilly, K., Grimm, D., Bueschel, D.M., Krum, J.G., and 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.

Tilly, K., Krum, J.G., Bestor, A., Jewett, M.W., Grimm, D., Bueschel, D., Byram, R., Dorward, D., Vanraden, M.J., Stewart, P.*, et al.* (2006). *Borrelia burgdorferi* OspC protein required exclusively in a crucial early stage of mammalian infection. Infect Immun *74*, 3554-3564.

Tokarz, R., Anderton, J.M., Katona, L.I., and Benach, J.L. (2004). Combined effects of blood and temperature shift on *Borrelia burgdorferi* gene expression as determined by whole genome DNA array. Infect Immun *72*, 5419-5432.

Tokarz, R., Tagliafierro, T., Sameroff, S., Cucura, D.M., Oleynik, A., Che, X., Jain, K., and Lipkin, W.I. (2019). Microbiome analysis of *Ixodes scapularis* ticks from New York and Connecticut. Ticks Tick Borne Dis *10*, 894-900. https://doi.org/10.1016/j.ttbdis.2019.04.011

Travinsky, B., Bunikis, J., and Barbour, A.G. (2010). Geographic differences in genetic locus linkages for *Borrelia burgdorferi*. Emerg Infect Dis *16*, 1147-1150. https://doi.org/10.3201/eid1607.091452

Troy, E.B., Lin, T., Gao, L., Lazinski, D.W., Lundt, M., Camilli, A., Norris, S.J., and Hu, L.T. (2016). Global Tn-seq analysis of carbohydrate utilization and vertebrate infectivity of *Borrelia burgdorferi*. Mol Microbiol *101*, 1003-1023. https://doi.org/10.1111/mmi.13437

Tsao, N., Kuo, C.F., Cheng, M.H., Lin, W.C., Lin, C.F., and Lin, Y.S. (2019). Streptolysin S induces mitochondrial damage and macrophage death through inhibiting degradation of glycogen synthase kinase-3beta in *Streptococcus pyogenes* infection. Sci Rep *9*, 5371. https://doi.org/10.1038/s41598-019-41853-3

Tufts, D.M., Hart, T.M., Chen, G.F., Kolokotronis, S.O., Diuk-Wasser, M.A., and Lin, Y.P. (2019). Outer surface protein polymorphisms linked to host-spirochete association in Lyme *Borreliae*. Mol Microbiol *111*, 868-882. https://doi.org/10.1111/mmi.14209

Tyler, S., Tyson, S., Dibernardo, A., Drebot, M., Feil, E.J., Graham, M., Knox, N.C., Lindsay, L.R., Margos, G., Mechai, S.*, et al.* (2018). Whole genome sequencing and phylogenetic analysis of strains of the agent of Lyme disease *Borrelia burgdorferi* from Canadian emergence zones. Sci Rep *8*, 10552. https://doi.org/10.1038/s41598-018-28908-7

Urwin, R., and Maiden, M.C. (2003). Multi-locus sequence typing: a tool for global epidemiology. Trends Microbiol *11*, 479-487.

Verhey, T.B., Castellanos, M., and Chaconas, G. (2019). Antigenic variation in the Lyme spirochete: detailed functional assessment of recombinational switching at *vlsE* in the JD1 strain of *Borrelia burgdorferi*. Mol Microbiol *111*, 750-763. https://doi.org/10.1111/mmi.14189

Vink, C., Rudenko, G., and Seifert, H.S. (2012). Microbial antigenic variation mediated by homologous DNA recombination. FEMS Microbiol Rev *36*, 917-948. https://doi.org/10.1111/j.1574-6976.2011.00321.x

Vitorino, L., Margos, G., Ze-Ze, L., Kurtenbach, K., and Collares-Pereira, M. (2010). Plasmid profile analysis of Portuguese *Borrelia lusitaniae* strains. Ticks Tick Borne Dis *1*, 125-128. https://doi.org/10.1016/j.ttbdis.2010.07.001

Vitorino, L.R., Margos, G., Feil, E.J., Collares-Pereira, M., Ze-Ze, L., and Kurtenbach, K. (2008). Fine-scale Phylogeographic Structure of *Borrelia lusitaniae* Revealed by Multilocus Sequence Typing. PloS ONE *3*, e4002.

Vollmer, S.A., Bormane, A., Dinnis, R.E., Seelig, F., Dobson, A.D., Aanensen, D.M., James, M.C., Donaghy, M., Randolph, S.E., Feil, E.J.*, et al.* (2011). Host migration impacts on the phylogeography of Lyme Borreliosis spirochaete species in Europe. Environ Microbiol *13*, 184-192. https://doi.org/10.1111/j.1462-2920.2010.02319.x

Vollmer, S.A., Feil, E.J., Chu, C.Y., Raper, S.L., Cao, W.C., Kurtenbach, K., and Margos, G. (2013). Spatial spread and demographic expansion of Lyme borreliosis spirochaetes in Eurasia. Infect Genet Evol *14C*, 147-155. https://doi.org/10.1016/j.meegid.2012.11.014

Wallich, R., Pattathu, J., Kitiratschky, V., Brenner, C., Zipfel, P.F., Brade, V., Simon, M.M., and Kraiczy, P. (2005). Identification and functional characterization of complement regulator-acquiring surface protein 1 of the Lyme disease spirochetes *Borrelia afzelii* and *Borrelia garinii*. Infect Immun *73*, 2351-2359. https://doi.org/10.1128/IAI.73.4.2351-2359.2005

Walter, K.S., Carpi, G., Caccone, A., and Diuk-Wasser, M.A. (2017). Genomic insights into the ancient spread of Lyme disease across North America. Nat Ecol Evol *1*, 1569-1576. https://doi.org/10.1038/s41559-017-0282-8

Walter, K.S., Carpi, G., Evans, B.R., Caccone, A., and Diuk-Wasser, M.A. (2016). Vectors as Epidemiological Sentinels: Patterns of Within-Tick *Borrelia burgdorferi* Diversity. PLoS Pathogens *12*, e1005759. https://doi.org/10.1371/journal.ppat.1005759

Wang, G., Ojaimi, C., Iyer, R., Saksenberg, V., McClain, S.A., Wormser, G.P., and Schwartz, I. (2001). Impact of genotypic variation of *Borrelia burgdorferi* sensu stricto on kinetics of dissemination and severity of disease in C3H/HeJ mice. Infect Immun *69*, 4303-4312. https://doi.org/10.1128/IAI.69.7.4303-4312.2001

Wang, G., Ojaimi, C., Wu, H., Saksenberg, V., Iyer, R., Liveris, D., McClain, S.A., Wormser, G.P., and Schwartz, I. (2002). Disease severity in a murine model of lyme borreliosis is associated with the genotype of the infecting *Borrelia burgdorferi* sensu stricto strain. J Infect Dis *186*, 782-791. https://doi.org/10.1086/343043

Wang, G., van Dam, A.P., and Dankert, J. (1999a). Evidence for frequent OspC gene transfer between *Borrelia valaisiana* sp. nov. and

other Lyme disease spirochetes. FEMS Microbiol Lett *177*, 289-296. https://doi.org/10.1111/j.1574-6968.1999.tb13745.x

Wang, G., van Dam, A.P., Schwartz, I., and Dankert, J. (1999b). Molecular typing of *Borrelia burgdorferi* sensu lato: taxonomic, epidemiological, and clinical implications. Clin Microbiol Rev *12*, 633-653.

Wang, I.N., Dykhuizen, D.E., Qiu, W., Dunn, J.J., Bosler, E.M., and Luft, B.J. (1999c). Genetic diversity of *ospC* in a local population of *Borrelia burgdorferi* sensu stricto. Genetics *151*, 15-30.

Wang, I.N., Smith, D.L., and Young, R. (2000). Holins: the protein clocks of bacteriophage infections. Annu Rev Microbiol *54*, 799-825.

Wang, P., Lutton, A., Olesik, J., Vali, H., and Li, X. (2012). A novel iron- and copper-binding protein in the Lyme disease spirochaete. Mol Microbiol *86*, 1441-1451. https://doi.org/10.1111/mmi.12068

Weening, E.H., Parveen, N., Trzeciakowski, J.P., Leong, J.M., Hook, M., and Skare, J.T. (2008). *Borrelia burgdorferi* lacking DbpBA exhibits an early survival defect during experimental infection. Infect Immun *76*, 5694-5705. https://doi.org/IAI.00690-08 [pii] 10.1128/IAI.00690-08

Wilske, B., Preac-Mursic, V., Jauris, S., Hofmann, A., Pradel, I., Soutschek, E., Schwab, E., Will, G., and Wanner, G. (1993). Immunological and molecular polymorphisms of OspC, an immunodominant major outer surface protein of *Borrelia burgdorferi*. Infect Immun *61*, 2182-2191.

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

Wormser, G.P., Brisson, D., Liveris, D., Hanincova, K., Sandigursky, S., Nowakowski, J., Nadelman, R.B., Ludin, S., and Schwartz, I. (2008). *Borrelia burgdorferi* genotype predicts the capacity for hematogenous dissemination during early Lyme disease. J Infect Dis *198*, 1358-1364.

Wormser, G.P., Liveris, D., Nowakowski, J., Nadelman, R.B., Cavaliere, L.F., McKenna, D., Holmgren, D., and Schwartz, I. (1999). Association of specific subtypes of *Borrelia burgdorferi* with hematogenous dissemination in early Lyme disease. J Infect Dis *180*, 720-725. https://doi.org/JID981202 [pii]10.1086/314922

Wywial, E., Haven, J., Casjens, S.R., Hernandez, Y.A., Singh, S., Mongodin, E.F., Fraser-Liggett, C.M., Luft, B.J., Schutzer, S.E., and Qiu, W.G. (2009). Fast, adaptive evolution at a bacterial host-resistance locus: The PFam54 gene array in *Borrelia burgdorferi*. Gene *445*, 26-37. https://doi.org/S0378-1119(09)00325-4 [pii] 10.1016/j.gene.2009.05.017

Xie, J., Zhi, H., Garrigues, R.J., Keightley, A., Garcia, B.L., and Skare, J.T. (2019). Structural determination of the complement inhibitory domain of *Borrelia burgdorferi* BBK32 provides insight into classical pathway complement evasion by Lyme disease spirochetes. PLoS Pathogens *15*, e1007659. https://doi.org/10.1371/journal.ppat.1007659

Xu, Y., Kodner, C., Coleman, L., and Johnson, R.C. (1996). Correlation of plasmids with infectivity of *Borrelia burgdorferi* sensu stricto type strain B31. Infect Immun *64*, 3870-3876.

Yang, X., Popova, T.G., Hagman, K.E., Wikel, S.K., Schoeler, G.B., Caimano, M.J., Radolf, J.D., and Norgard, M.V. (1999). Identification, characterization, and expression of three new members of the *Borrelia burgdorferi* Mlp (2.9) lipoprotein gene family. Infect Immun *67*, 6008-6018.

Yang, X.F., Hubner, A., Popova, T.G., Hagman, K.E., and Norgard, M.V. (2003). Regulation of expression of the paralogous Mlp family in *Borrelia burgdorferi*. Infect Immun *71*, 5012-5020.

Yang, X.F., Pal, U., Alani, S.M., Fikrig, E., and Norgard, M.V. (2004). Essential role for OspA/B in the life cycle of the Lyme disease spirochete. J Exp Med *199*, 641-648.

Young, R. (2002). Bacteriophage holins: deadly diversity. J Molec Microbiol Biotechnol *4*, 21-36.

Young, R., and Blasi, U. (1995). Holins: form and function in bacteriophage lysis. FEMS Microbiol Rev *17*, 191-205.

Zhang, H., and Marconi, R.T. (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.

Zhang, H., Raji, A., Theisen, M., Hansen, P.R., and Marconi, R.T. (2005). bdrF2 of Lyme disease spirochetes is coexpressed with a series of cytoplasmic proteins and is produced specifically during early infection. J Bacteriol *187*, 175-184.

Zhang, J.R., Hardham, J.M., Barbour, A.G., and Norris, S.J. (1997). Antigenic variation in Lyme disease borreliae by promiscuous recombination of VMP-like sequence cassettes. Cell *89*, 275-285. https://doi.org/10.1016/s0092-8674(00)80206-8

Zuckert, W.R., and Barbour, A.G. (2000). Stability of *Borrelia burgdorferi bdr* loci *in vitro* and *in vivo*. Infect Immun *68*, 1727-1730.

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

Zuckert, W.R., Meyer, J., and Barbour, A.G. (1999). Comparative analysis and immunological characterization of the *Borrelia* Bdr protein family. Infect Immun *67*, 3257-3266.