Published in final edited form as:
Infect Genet Evol. 2014 October ; 27: 551–558. doi:10.1016/j.meegid.2014.04.022.

Phylogeny of a relapsing fever *Borrelia* species transmitted by the hard tick *Ixodes scapularis*

Alan G. Barbour
Departments of Microbiology & Molecular Genetics, Medicine, and Ecology & Evolutionary Biology, University of California Irvine, Irvine, CA, USA 92697 abarbour@uci.edu Phone: +1.949.824.5626

Abstract

The discovery of *Borrelia* species that were related to the agents of relapsing fever but were transmitted by hard ticks rather than soft ticks challenged previous taxonomies based largely on microbe-host specificities and geographic considerations. One of these newly-identified organisms is the *Borrelia miyamotoi* sensu lato strain LB-2001 from North America and transmitted by *Ixodes scapularis*. This or related strains have been identified as the cause of human disease, but comparatively little is known about their biology or genetics. Using recently acquired chromosome sequence of LB-2001 together with database sequences and additional sequences determined here, I carried out comparisons of the several species of *Borrelia*, including those in the two major clades: the relapsing fever group of species and the Lyme disease group of species. Phylogenetic inference at the species level was based on four data sets: whole chromosomes of ~1 Mb each, and concatenated sequences of 19 ribosomal protein genes, 3 conserved nucleic acid enzymes (*rpoC*, *recC*, and *dnaE*), and 4 contiguous genes for nucleotide salvage on a large plasmid. Analyses using neighbor-joining, maximum likelihood, and Bayesian methods were largely concordant for each of the trees. They showed that LB-2001 and related hard tick-associated organisms, like *B. lonestari*, are deeply positioned within the RF group of species and that these organisms did not, as some earlier estimations had suggested, constitute a paraphyletic group. The analyses also provided further evidence that major changes in host ranges and life cycles, such as hard to soft ticks or vice versa, may not correlate well with overall sequence differences. The genetic differences between LB-2001 and *B. miyamotoi* sensu stricto justify provisional use of the “sensu lato” designation for LB-2001.

Keywords

tick-borne disease; spirochetes; Lyme disease; genomics

1. INTRODUCTION

The genus *Borrelia* comprises several species that share these life-cycle features: parasitization, either as a specialist or a generalist, of one or more vertebrate species, and transmission between these hosts, with different degrees of specificity, by a hematophagous arthropod. For all species but one, the louse-borne *B. recurrentis*, the vector is a tick. For some *Borrelia* species the tick mainly serves as a vector. Infection may be transtadial, that is persisting through different stages. But, if the tick fails to feed on a competent reservoir

host, that bacterial lineage ceases with death of the tick. In other species, the tick may not only be a vector between vertebrate reservoirs, such as rodents, but also provide for cross-generational maintenance on its own through transovarial transmission (Rollend et al., 2013). Two other binary traits for classification of *Borrelia* species are based on their interactions with their host and can be expressed as questions: (i) Is the preferred vector a soft (argasid) tick, such as a member of the *Ornithodoros* genus, or a hard (ixodid) tick, such as members of the prostriate genus *Ixodes* or metastriate genus *Amblyomma*? (ii) Is the infection burden of the host at time of transmission to the vector greatest in the blood itself or in a tissue, particularly the skin (Barbour, 2005; Barbour and Hayes, 1986)?

The tick-borne species that were known to cause relapsing fever in humans or a similar disease in domestic animals (such as avian borreliosis caused by *B. anserina*) were vectored by argasid ticks, commonly were transovarially transmitted, and mainly populated the blood. Some of the best-studied tickborne RF agents were the Palearctic and Afrotropic ecozone ("Old World") species *B. duttonii* and *B. crocidurae* and Nearctic ("New World") species *B. hermsii*, *B. turicatae*, and *B. parkeri* (Barbour, 2005). Each of these was vectored by a soft tick species, such as *O. hermsi* for *B. hermsii*, one of the RF agents in western North America. There was the one exception of the hard tick-transmitted *B. theileri*, which caused a mild febrile disease of cattle, bovine borreliosis, and other large animals, but this was not known to cause human infection and was little studied (Smith et al., 1985).

The discovery that *B. burgdorferi* and related species, like *B. afzelii* and *B. garinii*, were the etiologies of Lyme disease (LD) in North America, Europe, and Asia, was followed by recognition that these organisms not only were genetically distinct from the RF group of species, but they also differed in several biological features in their vectors and reservoirs. Unlike the RF agents known at the time, LD-associated *Borrelia* species were transmitted by hard ticks of the instead of soft ticks, did not manifest transovarial transmission, and generally achieved higher burdens in the skin than in the blood.

But a blurring of lines between these two groups began with the observation of additional RF-like species besides *B. theileri* in hard ticks. These included *B. miyamotoi* in *Ixodes persulcatus* in Asia (Fukunaga et al., 1995), *B. lonestari* in *Amblyomma americanum* in North America (Barbour et al., 1996), another *Borrelia* sp. in *A. geoemydae* in Japan (Takano et al., 2012), and at least one species, *B. turcica*, that was transmitted among reptiles by other *Amblyomma* as well as *Hyalomma* species (Takano et al., 2010) Organisms similar to *B. miyamotoi* in their vector associations with *Ixodes* species ticks and in selected DNA sequences were subsequently reported from different regions of the United States (Barbour et al., 2009; Hamer et al., 2012; Mun et al., 2006; Scoles et al., 2001; Scott et al., 2010) and Europe (Fraenkel et al., 2002; Geller et al., 2012; Subramanian et al., 2012). (I refer to the original isolate, HT31 (Fukunaga et al., 1995) and closely-related strains in Asia as *B. miyamotoi* sensu stricto and to less similar organisms transmitted by *Ixodes* spp. ticks as *B. miyamotoi* sensu lato (Bunikis et al., 2004b), with the assumption, as discussed in section 4., that this terminology is provisional.)

A view of *B. miyamotoi* sensu stricto (Bmss) and *B. miyamotoi* sensu lato (Bmsl) as strictly enzootic, with little or no opportunity or capacity for causing disease in people, had to be

revised after reports of human infections first in Russia (Platonov et al., 2011) and then in the United States (Chowdri et al., 2013; Gugliotta et al., 2013; Krause et al., 2013) and western Europe (Hovius et al., 2013). Doubtless more will be revealed about and the pathogenetic mechanisms and impact on public health of Bmss and Bmsl from laboratory and epidemiologic investigations, but these efforts would be advanced if we knew more about the genomes of these organisms and the diversity of their strains.

Phylogenetic analyses, mainly based on 16S ribosomal RNA (rDNA) and/or flagellin gene sequences, were in agreement that *B. theileri*, *B. lonestari*, and *B. miyamotoi* clustered with the agents of relapsing fever, including *B. recurrentis*, rather the second major *Borrelia* clade comprising the LD agents (Barbour, 2001; Barbour et al., 1996; Fraenkel et al., 2002; Fukunaga et al., 1996; Ras et al., 1996; Rich et al., 2001; Scoles et al., 2001). But, with some exceptions, such as the identification of *B. turicatae* and *B. parkeri* as sister species, the relationships within what came to be called the RF group were not well-resolved by studies based on just one or two loci. Did the hard tick-associated species represent a paraphyletic clade? Are they justifiably placed basal to the soft tick-associated species when the outgroup is the LD clade?

To provide additional material for phylogenetic inference, my laboratory carried out sequencing of LB-2001, a Bmsl strain and original isolate in North America (Hue et al., 2013), as well as the chromosome and much of the plasmid content of *B. parkeri* (Barbour and Campeau Miller, 2014). For the present study, these were added to existing chromosome and plasmid sequences of other *Borrelia* spp. in the databases for a fuller representation across the genus. I also obtained sequence of extrachromosomal DNA from de novo assemblies of LB-2001, and carried out partial genome sequencing of the avian borreliosis agent *B. anserina*, a species that appeared to be positioned in the RF group but whose genetics had been little characterized.

2. MATERIALS and METHODS

2.1. Organisms and cultivation

B. anserina strain Es was originally provided by Russell Johnson (University of Minnesota) and was cloned by limiting dilution (Ferdows et al., 1996). The bacteria were cultivated in BSK II medium at 34°C (Barbour, 1984) and then harvested by centrifugation as described (Dai et al., 2006).

2.2. Genome sequencing

The procedures for the sequencing of genomes of Bmsl strain LB-2001, which was isolated from the blood of mice, and *B. parkeri* strain HR1 have been described (Barbour and Campeau Miller, 2014; Hue et al., 2013). DNA from a fresh culture harvest of *B. anserina* was extracted with Qiagen's DNeasy Blood/Tissue Kit (Valencia, CA) and then treated with RNase I. The library was produced with the Ion Xpress™ Plus Fragment Library Kit with size-selection with the E-Gel system (Invitrogen) before emulsion PCR on an Ion OneTouch apparatus (Life Technologies, Carlsbad, CA). Templates were sequenced on an Ion Torrent Personal Genome Machine with Ion 314 or Ion 316 chips (Life Technologies). Single reads

of ~100-200 bases were assembled into de novo contigs using the Assembly Cell program of Genomics Workbench version 6.1 (CLC bio, Aarhus, Denmark). Selected genes were identified by searches of assembled contigs with BLASTN or TBLASTN algorithms on a local standalone WWW Blast server (Altschul et al., 1997).

2.3. Sequence analysis

Alignment of DNA and protein sequences was carried out either with ClustalX v. 2 (Larkin et al., 2007) or, for longer sequences and for DNA matrix dot plots, with MAFFT v. 7.110 (http://mafft.cbrc.jp) with default settings. Alignments were then edited and codon-aligned manually with MacClade v. 4.10 (Sinauer Associates, Inc., Sunderland, MA) or Mesquite v. 2.75 (http://mesquiteproject.org). Nucleotide diversity (π), the ratio of the number of nonsynonymous substitutions per non-synonymous site (K_a) to the number of synonymous substitutions per synonymous site (K_s), and Tajima's *D* test (Tajima, 1989) were assessed with the DnaSP v. 5.10 suite (Librado and Rozas, 2009). For computation of a similarity matrix for ungapped, aligned nucleotide sequences, Mobyle@Pasteur's (http://mobyle.pasteur.fr) Distance Matrix program, version 3.67, which is an implementation of DNADIST of the PHYLIP package (http://evolution.genetics.washington.edu/phylip/doc/main.html) was used. Distance-based clustering algorithms for ungapped alignments were neighbor joining with the BioNJ protocol and the Jukes-Cantor model for substitution for codon-aligned proteins-encoding sequences and observed differences for whole chromosomes. Phylogenetic inference was carried out by maximum likelihood estimation, as implemented by PhyML v. 3.0 (Guindon et al., 2010) in the SeaView suite version 4.2.12 (Gouy et al., 2010) or at http://www.atgcmontpellier.fr/phyml. Evolutionary models for protein encoding regions and corresponding parameters for inference were evaluated by maximum likelihood with Los Alamos National Laboratory's FindModel (http://www.hiv.lanl.gov), which is an implementation of the Modeltest script of Posada and Crandall (Posada and Crandall, 1998) and uses PAML of Yang et al. to calculate likelihood (Yang, 1994). Bayesian analysis was performed using MrBayes v. 3.2 (Huelsenbeck and Ronquist, 2001). The maximum likelihood model for all analyses employed 6 substitution types (General Time Reversible [GTR]), with a substitution model for triplets (codons) of nucleotides and rate variation across sites modelled using a gamma distribution. The Markov chain Monte Carlo search was run with 4 chains for 1,000,000 generations, with trees sampled every 100 generations, and with discard of first 5000 trees sampled. The clade credibility values in percent are given, where “100” corresponds to a posterior probability for the node of 1.0.

2.4. Accession numbers

The Genbank accession numbers for the annotated complete chromosome of Bmsl strain LB-2001 and for chromosome and 114 kb of the lp150 linear plasmid of *B. parkeri* strain HR1 are CP006647 and CP007036, respectively. Unless otherwise specified, these were the sources for the sequences for these species used in the study. The sequences of the following ribosomal protein genes of *B. anserina* were determined for this study and were assigned accession numbers KU136495-KJ136513, respectively and consecutively: *rplA*, *rplB*, *rplC*, *rplD*, *rplE*, *rplF*, *rplI*, *rplJ*, *rplK*, *rplM*, *rpsA*, *rpsB*, *rpsC*, *rpsD*, *rpsE*, *rpsG*, *rpsH*, *rpsI*, and *rpsR*. Other *B. anserina* genes and sequences assigned accession numbers were *rpoC*

(KJ136514), *recC* (KJ136515), *dnaE* (KJ136516), and a 4 kb fragment containing full-length *thyX*, *nrdI*, *nrdE*, and *nrdF* (KJ136517). Other sequences (with accession numbers) that were determined for the present study were the following: *vsp*1 gene of LB-2001 (KF031441), *vsp1* gene of *B. anserina* (KJ136518), and a fragment of LB-2001's large linear plasmid containing the *thyX*, *nrdI*, *nrdE*, and *nrdF* genes(KJ141201).

The complete chromosomes and individual chromosomal genes of other species were obtained from these accession numbers: CP002933 for *B. afzelii* strain PKo, CP002746 for *B. bissettii* DN127, AE000783 for *B. burgdorferi* B31, CP002312 for *B. burgdorferi* JD1, CP002228 for *B. burgdorferi* N40, CP003426 for *B. crocidurae* Achema, CP000976 for *B. duttonii* Ly, CP003151 for *B. garinii* BgVir, CP000048 for *B. hermsii* DAH of genomic group I (Porcella et al., 2005), CP004146 for *B. hermsii* YOR of genomic group II, CP000993 for *B. recurrentis* A1, CP000049 for *B. turicatae* 91E135, and CP004217 for *B. miyamotoi* FR64b. The sources of *thyX*, *nrdI*, *nrdE*, and *nrdF* gene sequences of the large plasmid of other species were HM008710 for *B. turicatae*, CP007036 for *B. parkeri*, HM008709 for *B. hermsii*, CP003435 for *B. crocidurae*, NC_011247 for *B. duttonii*, and NC_011246 for *B. recurrentis*. The *B. lonestari* strain LS-1 sequences used were EF07522 for *rpoC*, EF07518 for *recC*, and EF507510 for *dnaE*. Some accession numbers are provided at the point in the text where the sequences are introduced.

3. RESULTS

3.1. Genome sequencing of *B. anserina* and LB-2001

For *B. anserina* a total of 2,043,539 single reads of average length 143 nt were obtained from two combined runs. From the de novo assembly, there were 68 non-overlapping contigs with a N50 of 45,615 nt, yielding a total of 1,040,570 bp with an average coverage of 274X . These contigs contained the complete coding sequences for all the 30S and 50S ribosomal protein genes of this study, *rpoC*, *recC*, *dnaE*, *thyX*, *nrdI*, *nrdE*, and *nrdF*, as well as a complete variable small protein (vsp) of this species. The output for the next-gen sequencing of the LB-2001 genome is described by Hue and Barbour (Hue et al., 2013). For the present study we identified in a contig of 37,477 nt of extrachromosomal DNA a contiguous sequence containing the *thyX*, *nrdI*, *nrdE*, and *nrdF* gene, and in an extrachromosomal contig of 23,781 nt a complete *vsp* at a putative expression site in this species.

3.2. Comparison of chromosomes

The LB-2001 genome had the following horizontally-acquired genes that distinguish the relapsing fever group of species from the LD group of species: hypoxanthine-guanine phosphoribosyltransferase) (*hpt*), adenylosuccinate synthetase (*purA*), adenylosuccinate lyase (*purB*) in 16S-23S intergenic spacer, as previously reported (Barbour et al., 2005; Pettersson et al., 2007), and the ribonucleotide reductase subunit alpha (*nrdE*), ribonucleotide reductase subunit beta (*nrdF*), accessory ribonucleoprotein (*nrdI*), and FAD-dependent thymidylate synthase (*thyX*) genes of the large plasmids of *Borrelia* species (Miller et al., 2013; Zhong et al., 2006), as determined here. LB-2001's genome did not have the genes for either the phosphotransferase system maltose-specific enzyme, IICB

component (*glvC*) or maltose-6'-phosphate glucosidase (*glvA*), whose presence in the 16S-23S intergenic spacer of the reptile-associated species *Borrelia turcica* and *Borrelia* sp. BF-16 distinguishes them from both RF group and LD group species (Takano et al., 2010).

The G+C content (%) of LB-2001's chromsome was 28.7, compared to 27.6 for *B. duttonii*, 27.5 for *B. recurrentis*, 27.6 for *B. crocidurae*, 28.9 for *B. parkeri*, 29.1 for *B. turicate*, 29.8 for *B. hermsii*, and 30.0 for *B. anserina*. The corresponding values for the LD group species *B. burgdorferi*, *B. bissettii*, *B. afzelii*, *B. garinii*, and *B. bavariensis* were 28.6, 28.7, 28.3, 28.4, and 28.3, respectively. While the G+C contents were similar across all species, the ratio of the chromosome content of CG dinucleotides to G+C content distinguished the LD group of species from the RF group of species (Figure 1). LB-2001's CG/G+C ratio plotted along the regression line for the RF group, which overall displayed greater variance in this ratio than did the LD group members.

As Takahashi et al. by physical mapping observed for Bmss (Takahashi and Fukunaga, 1996), the chromosome of LB-2001 was syntenic with the chromosomes not only of *B. hermsii* and other RF group species (not shown) but also with that of *B. burgdorferi* (Figure 2). Large inversions, duplications, or rearrangements were not noted. Given this synteny, we aligned the complete chromosomes sequences for LB-2001, *B. hermsii*, *B. turicatae*, *B. crocidurae*, and *B. burgdorferi*. The alignment comprised 968,954 sites, of which 855,554 were ungapped, and, of these, 626,287 were non-segregating. Table 1 gives the pairwise identities, and Figure 3 shows a phylogram based on observed differences for this mix of coding and non-coding sequence. By this analysis LB-2001 was not outside the cluster of RF species. Indeed, the Old World RF species was basal to it and the other RF species.

3.3. Ribosomal protein genes

To incorporate *B. anserina* sequences into the phylogenetic analyses we used a concatenated set of ribosomal protein genes. These are informational genes, according to Lake and colleagues' classification (Rivera et al., 1998), and thus more informative for phylogenetic inference than operational genes, such as an enzyme in carbohydrate metabolism pathway, which are more at risk of horizontal replacement or addition. Sets of ribosomal protein genes have been the basis of other validated phylogenetic procedures for both prokaryotic and eukaryotic organisms (Bachvaroff et al., 2014; Mende et al., 2013; Teeling and Gloeckner, 2006). Accordingly, DNA sequences for the following 19 ribosomal proteins, whose genes were distributed over different regions of the chromosome, were obtained from the database or, in the case of *B. anserina*, sequenced for this study: the 50S proteins L1, L2, L3, L4, L5, L6, L9, L10, L11, and L13, and 30S proteins S1, S2, S3, S4, S5, S7, S8, S9, and S18. Concatenated codon-aligned nucleotide sequences for these were compared among LB-2001, seven RF group *Borrelia* species (hermsii, turicatae, parkeri, crocidurae, duttonii, recurrentis, and *anserina*), and four LD group *Borrelia* species (*burgdorferi, bissettii, afzelii*, and *garinii*).

There were 11,568 ungapped sites, of which 3831 (33.1%) were segregating. The average transition (T_s) to transversion (T_v) ratio was 4.1. Overall nucleotide diversity (π) was 0.139 with values of 0.620 for synonymous sites and 0.072 for nonsynonymous sites. The K_a/K_s ratios for a total of 66 pairs of sequences, the means (95% confidence intervals) were 0.107

(0.102-0.113) for 28 pairs of 8 RF species, 0.074 (0.067-0.081) for 6 pairs of 4 LB species, and 0.123 (0.120-0.126) for 32 instances of a RF species paired with a LB species. This was evidence that these genes were not under positive selection not only within each of the two major clades but the genus as a whole. By Tajima's test (D = 0.21) the null hypothesis of neutrality was not rejected (p >0.10).

Figure 4 is a phylogram of these aligned sequences. There was strong support for the nodes by three different protocols: neighbor joining, maximum likelihood, and Bayesian analysis. For this alignment, which included four LD species, LB-2001 clusters with the North American species as well as *B. anserina*, while the three Old World species are relatively basal to LB-2001. As expected, *B. burgdorferi* and *B. bissettii* are sister species (Casjens et al., 2012; Schutzer et al., 2012), as are *B. turicatae* and *B. parkeri* (Schwan et al., 2005). The same topology was the inferred by Bayesian analysis with a smaller set of sequences that were approximately equally divergent: LB-2001, *B. hermsii*, *B. turicatae*, *B. crocidurae*, and d*B. anserina* with *B. burgdorferi* as the out-group. *B. crocidurae* was basal to LB-2001 with a posterior probability (p) of 0.99.

3.4. Relationships with *B. lonestari*

In the public database there was a limited set of randomly acquired sequences of *B. lonestari* strain LS-1. Among these were three complete or near-complete genes for enzymes involved in transcription and DNA replication and repair: 3445 nt of DNA-directed RNA polymerase, subunit beta (*rpoC*), 2183 nts of exodeoxyribonuclease V, subunit gamma (*recC*), and 3208 nt of DNA polymerase III, subunit alpha (*dnaE*). Corresponding sequences were available for *B. hermsii*, *B. crocidurae*, *B. parkeri*, *B. burgdorferi*, LB-2001, and, as determined for this study, for *B. anserina*. Concatenation of the sequences provided for an alignment of 8826 codon-aligned ungapped sites with 3078 (34.9%) segregating sites. The average T_s/T_v was 3.1. Overall π was 0.148 with 0.612 at synonymous sites and 0.085 at nonsynonymous sites. The results displayed in Figure 5A provides further evidence that Bmsl and *B. lonestari* are sister taxa and further confirmation of the clade relationships among the RF group species that were noted with whole chromosomes and ribosomal proteins.

3.5. Co-evolution of chromosomes and large linear plasmids

To assess the genetic linkage between the chromosomes and the large linear plasmids of species for which both chromosome and large plasmid sequences were available or determined here, I examined 4.2 kb of contiguous sequence that contained four genes, *thyX*, *nrdI*, *nrdE*, and *nrdF*, whose products catalyze the de novo synthesis of deoxyribonucleotides and which are not found in LD Borrelia species (Pettersson et al., 2007; Zhong et al., 2006). Concatenation of the sequences provided for an alignment of 4164 ungapped codon-aligned sites, out of a total of 4170, with 1214 (29.1%) segregating sites. The average T_s/T_v was 2.5. Overall π was 0.130 with 0.518 at synonymous sites and 0.072 at nonsynonymous sites. Tajima's D was -0.200. Although these genes plausibly were acquired by horizontal transmission by a common ancestor (Zhong et al., 2006), the findings preponderantly favor a model of purifying over balancing selection.

The tree topology of these plasmid-encoded sequences was concordant with that of chromsosomes, the ribosomal proteins and the second set of chromosomal genes (Figure 5B). The tree shown in the figure had a Bayesian posterior probability of 0.816. In the single alternative tree, which had a posterior probability (*p*) of 0.183, for a cumulative probability (*P*) of 0.999, LB-2001 and *B. anserina* resolved as sister species, while the other nodes, including the relatively basal position of *B. crocidurae*, were retained.

3.6. Variable membrane proteins

A distinguishing and consistent feature of the genus *Borrelia* are the carriage of one or more genes for a lipoprotein that has a characteristic structure designated as the Lipoprotein_6 (pfam01441) family (http://www.ncbi.nlm.nih.gov/Structure/cdd). In the LD group of species there is usually only one of these alleles per genome, while in the RF group there may be several, only one of which is expressed at a time (Barbour, 2005). The LD group proteins were named outer surface proteins C (OspC) and the RF group proteins were named variable small proteins (Vsp). Their corresponding genes are *ospC* and *vsp*, with a number added to the *vsp* to designate the allele in the genome. A *vsp* gene was identified in Bmss strain HT31 (Hamase et al., 1996). Using this sequence (accession number D78201), I searched extachromosomal contigs of LB-2001 and *B. anserina* for the present study and identified homologous sequences in both organisms (accession numbers KF031441 and KJ136518),. There is evidence of other *vsp* genes in these genomes (A.G.B., F. Hue, and Q. Dai, manuscript in preparation), so I cannot say whether these are paralogs rather than orthologs of the Bmss allele until more sequences of this organism are available. Nevertheless, the N-terminal regions of the deduced proteins can be compared, because they include a conserved signal peptide, which is generally shared among Vsp or OspC proteins in their full-length forms (Carter et al., 1994).

Figure 6 is an alignment of the N-terminal ends of the unprocessed proteins of the LB-2001 and *B. anserina* Vsp together with those of HT31, Vsp33 (also known as Vtp) of *B. hermsii* HS1 (AAA59359), Vsp1 of *B. turicatae* strain Oz1 (AAB65089), and for comparison, the OspC protein of *B. burdgdorferi* B31 (AAC66329). Bbss HT31 and LB-2001 proteins differed from the other proteins in the sequence at the predicted signal peptidase site with the addition of 6 residues. I found no evidence of homologous sequence for these peptides in searches of the deposited protein and nucleotide sequences for the chromosomes and plasmids of the “Old World” species *B. duttonii*, *B. recurrentis*, or *B. crocidurae*.

3.7. Strain structure

Our previous study of the strain structure of Bmsl in the northeastern US revealed identical 16S-23S intergenic spacer sequences in 22 tick extracts from Connecticut (Bunikis et al., 2004b). There was also an apparent lack of strain diversity noted in samples from 10 Bmsl-positive *I. ricinus* ticks in southern Sweden. We contrasted this limited strain diversity of Bmsl with richer population structure of *B. burgdorferi* and *B. afzelii* in the same respective collection areas (Bunikis et al., 2004a). Since then there have been additional relevant sequences added to the database. LB-2001 has identical partial 16S rDNA sequence over 516 nt to that of two uncultured Bmsl strains from Northern California (DQ025522 and DQ025521) (Mun et al., 2006). LB-2001 also has an identical 16s-23S intergenic spacer

over 461 nt to Bmsl from Tennessee (HQ658903). In contrast, over this length there were a combination of 46-47 substitutions and indels between the LB-2001 sequence and that of two strains of Bmss from Japan (AY363703 and AY363704).

This analysis led to further study of the differences between isolates falling under the “sensu lato” rubric. I asked how the genetic distances between LB-2001 and a representative of Bmss compared to distances between other pairs of *Borrelia* taxa that have been assigned their own species designations. Figure 7 shows two phylograms based on observed differences in ungapped aligned sequences of whole chromosome of selected sets of RF group species, on the left, and Lyme disease group species, on the right. The overall chromosome pairwise identity between strain LB2001 and the Bmss strain FR64b was 0.979. This was far less than between the 3 representative strains of *B. burgdorferi*, B31, JD1, and N40, which had pairwise identities of 0.994 to 0.996. The LB-2001 and FR64b identity was close to the 0.977 identity between the two different species of *B. turicatae* and *B. parkeri*, but greater than the 0.965 identity between strains DAH and YOR, representatives of the two genomic groups of *B. hermsii*. Thus, there was little concordance among these examples between species and strain distinctions and overall DNA sequence diversity.

4. DISCUSSION

This study further establishes the placement of LB-2001 deep in the major clade of *Borrelia* species containing the RF agents. Beside the evidence of the well-resolved and -supported trees provided here, LB-2001 has these other genetic characteristics of RF group species that unambiguously distinguish this clade from the LD clade: (i) *purA* and *purB* genes in the 16S-23S intergenic spacer (Barbour et al., 2005; Pettersson et al., 2007); (ii) contiguous *nrdE*, *nrdF*, *nrdI*, and *thyX* genes on an extrachromosomal replicon (Zhong et al., 2006); and (iii) a *glpQ* gene on the chromosome (Schwan et al., 2003). In contrast to taxonomic investigations in the pre-molecular biology era, we know more about the DNA sequences of Bmsl and Bmss organisms than we do about their biology at this point. Perhaps we are missing the proverbial forest for the trees by a too narrow focus on sequence. Microbe-vector associations may be important considerations that conceivably should trump sequence-based classification. One could compellingly argue that the otherwise closely-related louse-borne *B. recurrentis* and tick-borne *B. duttonii* are good case for this (Lescot et al., 2008). But in the instance of Bmsl, the biological features of higher densities achieved in the blood than in skin or other tissues (Barbour et al., 2009) and of their transovarial transmission in ticks (Rollend et al., 2013) are consistent with the traits displayed by other RF group members (Barbour and Hayes, 1986).

The findings here accord with the earlier phylogenetic estimations from a flagellin and 16S rDNA sequences that grouped the hard tick-transmitted Bmss, Bmsl, *B. lonestari*, and *B. theileri* with the soft tick-transmitted RF agents and not the hard tick-transmitted LD agents (Section 1.). But some of these earlier results could also be interpreted as placing Bmss and Bmsl either basal or paraphyletic to a large cluster of species that had long been recognized as agents of relapsing fever. The more extensive analysis here of whole chromosomes and three sets of genes does not provide insights about the evolution of genus *Borrelia* among

spirochetes more generally; the branches between *Borrelia* and other spirochete taxa we know about are too long for much confidence about those inferences (Paster and Dewhirst, 2000). But the study does provide for more confidence about the systematics within the RF group and further evidence that differences in vector-associations may poorly correlate with overall genetic distance. Although *B. duttonii* and *B. hermsii* are each vectored by a particular *Ornithodoros* soft tick species, *B. hermsii* was more distant in this analysis to *B. duttonii* than it was to either the hard tick-associated LB-2001 or *B. anserina*, which is associated with *Argas* spp. soft ticks and not *Ornithodoros*. In each of the analyses, Old World species were basal to the rest of the RF group species, which included LB-2001 and *B. anserina*. While the issue was not specifically addressed here, the findings also raise a question about the last ancestor for *B. recurrentis*. Is it *B. duttonii*, as originally proposed (Lescot et al., 2008), or, as Figures 4 and 5 suggest, either *B. crocidurae* or another, as yet uncharacterized Palearctic/Afrotropic species?

The present study focused on phylogeny at the species level, but it led to a consideration of strain differences under the "*Borrelia miyamotoi* sensu lato" category. The data presented in Figure 7 leads me to conclude that retention of a provisional "sensu lato" designation for LB-2001 and related North American strains is justified. As was seen with the LD group of species, the "sensu lato" designation can be a provisional until the phylogenetic relationships are more clear. After the original description of *B. burgdorferi* in North America and Europe, other organisms that were transmitted by *Ixodes* ticks but did not closely match the earliest examples of LD agents in their protein electrophoresis patterns or plasmid profiles were called "*B. burgdorferi* sensu lato" in the literature, as my own experience has shown (Barbour, 1988; Barbour et al., 1985). With time and further characterization at the biological, biochemical, and genetic levels, most of these "atypical" strains were categorized as one or another species distinct from *B. burgdorferi* sensu stricto.

One might argue now that LB-2001 merits a different species designation by citing the comparable genetic distance between two long-recognized species: *B. turicatae* and *B. parkeri* (Figure 7). On the other hand, the observed distance between LB-2001 and the Bmss representative was less than that between two strains of *B. hermsii*. The taxonomy of the hard tick-transmitted RF species of *Borrelia* may ultimately depend on future gains in the understanding of the biology of these organisms and identification of distinguishing traits, which may not be wholly concordant with the overall DNA sequence differences.

Acknowledgments

This work was supported by Public Health Service grants AI-24424 and AI-100236 from the National Institute of Allergy and Infectious Diseases. I thank Fong Hue for technical assistance.

References

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997; 25:3389–3402. [PubMed: 9254694]

Bachvaroff TR, Gornik SG, Concepcion GT, Waller RF, Mendez GS, Lippmeier JC, Delwiche CF. Dinoflagellate phylogeny revisited: using ribosomal proteins to resolve deep branching

dinoflagellate clades. Molecular phylogenetics and evolution. 2014; 70:314–322. [PubMed: 24135237]

Barbour AG. Isolation and cultivation of Lyme disease spirochetes. Yale J Biol Med. 1984; 57:521–525. [PubMed: 6393604]

Barbour AG. Plasmid analysis of Borrelia burgdorferi, the Lyme disease agent. J Clin Microbiol. 1988; 26:475–478. [PubMed: 3356787]

Barbour, AG. Borrelia: a diverse and ubiquitous genus of tick-borne pathogens.. In: Scheld, WM.; Craig, WA.; Hughes, JM., editors. Emerging Infections 5. ASM Press; Washington, D.C.: 2001. p. 153-174.

Barbour, AG. Relapsing fever.. In: Dennis, DT.; Goodman, JL.; Sonenshine, DE., editors. Tick-borne Diseases of Humans. ASM Press; Washington, D.C.: 2005. p. 220-236.

Barbour AG, Bunikis J, Travinsky B, Hoen AG, Diuk-Wasser MA, Fish D, Tsao JI. Niche partitioning of Borrelia burgdorferi and Borrelia miyamotoi in the same tick vector and mammalian reservoir species. Am J Trop Med Hyg. 2009; 81:1120–1131. [PubMed: 19996447]

Barbour AG, Campeau Miller S. Genome sequence of Borrelia parkeri, an agent of enzootic relapsing fever in western North America. Genome Announc 2. 2014 in press.

Barbour AG, Hayes SF. Biology of Borrelia species. Microbiol Rev. 1986; 50:381–400. [PubMed: 3540570]

Barbour AG, Heiland RA, Howe TR. Heterogeneity of major proteins in Lyme disease borreliae: a molecular analysis of North American and European isolates. J Infect Dis. 1985; 152:478–484. [PubMed: 2411827]

Barbour AG, Maupin GO, Teltow GJ, Carter CJ, Piesman J. Identification of an uncultivable Borrelia species in the hard tick Amblyomma americanum: possible agent of a Lyme disease-like illness. J Infect Dis. 1996; 173:403–409. [PubMed: 8568302]

Barbour AG, Putteet-Driver AD, Bunikis J. Horizontally acquired genes for purine salvage in Borrelia spp. causing relapsing fever. Infect Immun. 2005; 73:6165–6168. [PubMed: 16113341]

Bunikis J, Garpmo U, Tsao J, Berglund J, Fish D, Barbour AG. Sequence typing reveals extensive strain diversity of the Lyme borreliosis agents *Borrelia burgdorferi* in North America and *Borrelia afzelii* in Europe. Microbiology. 2004a; 150:1741–55. [PubMed: 15184561]

Bunikis J, Tsao J, Garpmo U, Berglund J, Fish D, Barbour AG. Typing of Borrelia relapsing fever group strains. Emerg Infect Dis. 2004b; 10:1661–1664. [PubMed: 15498172]

Carter CJ, Bergstrom S, Norris SJ, Barbour AG. A family of surface-exposed proteins of 20 kilodaltons in the genus Borrelia. Infect Immun. 1994; 62:2792–2799. [PubMed: 8005669]

Casjens SR, Mongodin EF, Qiu WG, Luft BJ, Schutzer SE, Gilcrease EB, Huang WM, Vujadinovic M, Aron JK, Vargas LC, et al. Genome stability of Lyme disease spirochetes: comparative genomics of Borrelia burgdorferi plasmids. PloS One. 2012; 7:e33280. [PubMed: 22432010]

Chowdri HR, Gugliotta JL, Berardi VP, Goethert HK, Molloy PJ, Sterling SL, Telford SR. Borrelia miyamotoi infection presenting as Human Granulocytic Anaplasmosis: a case report. Ann In Med. 2013; 159:21–27.

Dai Q, Restrepo BI, Porcella SF, Raffel SJ, Schwan TG, Barbour AG. Antigenic variation by Borrelia hermsii occurs through recombination between extragenic repetitive elements on linear plasmids. Mol Microbiol. 2006; 60:1329–1343. [PubMed: 16796672]

Ferdows MS, Serwer P, Griess GA, Norris SJ, Barbour AG. Conversion of a linear to a circular plasmid in the relapsing fever agent Borrelia hermsii. J Bacteriol. 1996; 178:793–800. [PubMed: 8550515]

Fraenkel CJ, Garpmo U, Berglund J. Determination of novel Borrelia genospecies in Swedish Ixodes ricinus ticks. J Clin Microbiol. 2002; 40:3308–3312. [PubMed: 12202571]

Fukunaga M, Okada K, Nakao M, Konishi T, Sato Y. Phylogenetic analysis of Borrelia species based on flagellin gene sequences and its application for molecular typing of Lyme disease borreliae. Int J Syst Bacteriol. 1996; 46:898–905. [PubMed: 8863416]

Fukunaga M, Takahashi Y, Tsuruta Y, Matsushita O, Ralph D, McClelland M, Nakao M. Genetic and phenotypic analysis of Borrelia miyamotoi sp. nov., isolated from the ixodid tick Ixodes persulcatus, the vector for Lyme disease in Japan. Int J Syst Bacteriol. 1995; 45:804–810. [PubMed: 7547303]

Geller J, Nazarova L, Katargina O, Jarvekulg L, Fomenko N, Golovljova I. Detection and genetic characterization of relapsing fever spirochete Borrelia miyamotoi in Estonian ticks. PLoS One. 2012; 7:e51914. [PubMed: 23251652]

Gouy M, Guindon S, Gascuel O. SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol Biol Evol. 2010; 27:221–224. [PubMed: 19854763]

Gugliotta JL, Goethert HK, Berardi VP, Telford SR 3rd. Meningoencephalitis from Borrelia miyamotoi in an immunocompromised patient. N Engl J Med. 2013; 368:240–245. [PubMed: 23323900]

Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol. 2010; 59:307–321. [PubMed: 20525638]

Hamase A, Takahashi Y, Nohgi K, Fukunaga M. Homology of variable major protein genes between Borrelia hermsii and Borrelia miyamotoi. FEMS Microbiol Lett. 1996; 140:131–137. [PubMed: 8764474]

Hamer SA, Hickling GJ, Keith R, Sidge JL, Walker ED, Tsao JI. Associations of passerine birds, rabbits, and ticks with Borrelia miyamotoi and Borrelia andersonii in Michigan, U.S.A. Parasites & Vectors. 2012; 5:231. [PubMed: 23057837]

Hovius JW, de Wever B, Sohne M, Brouwer MC, Coumou J, Wagemakers A, Oei A, Knol H, Narasimhan S, Hodiamont CJ, et al. A case of meningoencephalitis by the relapsing fever spirochaete Borrelia miyamotoi in Europe. Lancet. 2013; 382:658. [PubMed: 23953389]

Hue F, Ghalyanchi Langeroudi A, Barbour AG. Chromosome sequence of Borrelia miyamotoi, an uncultivable tick-borne agent of human infection. Genome Announc. 2013; 1(5):e00713–13. doi: 10.1128/genomeA.00713-13. [PubMed: 24029760]

Huelsenbeck JP, Ronquist F. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics. 2001; 17:754–755. [PubMed: 11524383]

Krause PJ, Narasimhan S, Wormser GP, Rollend L, Fikrig E, Lepore T, Barbour A, Fish D. Human Borrelia miyamotoi infection in the United States. N Engl J Med. 2013; 368:291–293. [PubMed: 23323920]

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, et al. Clustal W and Clustal X version 2.0. Bioinformatics. 2007; 23:2947–2948. [PubMed: 17846036]

Lescot M, Audic S, Robert C, Nguyen TT, Blanc G, Cutler SJ, Wincker P, Couloux A, Claverie JM, Raoult D, et al. The genome of Borrelia recurrentis, the agent of deadly louse-borne relapsing fever, is a degraded subset of tick-borne Borrelia duttonii. PLoS Genetics. 2008; 4:e1000185. [PubMed: 18787695]

Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics. 2009; 25:1451–1452. [PubMed: 19346325]

Mende DR, Sunagawa S, Zeller G, Bork P. Accurate and universal delineation of prokaryotic species. Nature Meth. 2013; 10:881–884.

Miller SC, Porcella SF, Raffel SJ, Schwan TG, Barbour AG. Large linear plasmids of Borrelia species that cause relapsing fever. J Bacteriol. 2013; 195:3629–3639. [PubMed: 23749977]

Mun J, Eisen RJ, Eisen L, Lane RS. Detection of a Borrelia miyamotoi sensu lato relapsing-fever group spirochete from Ixodes pacificus in California. J Med Entomol. 2006; 43:120–123. [PubMed: 16506458]

Paster BJ, Dewhirst FE. Phylogenetic foundation of spirochetes. J Mol Microbiol Biotechnol. 2000; 2:341–344. [PubMed: 11075904]

Pettersson J, Schrumpf ME, Raffel SJ, Porcella SF, Guyard C, Lawrence K, Gherardini FC, Schwan TG. Purine salvage pathways among Borrelia species. Infect Immun. 2007; 75:3877–3884. [PubMed: 17502392]

Platonov AE, Karan LS, Kolyasnikova NM, Makhneva NA, Toporkova MG, Maleev VV, Fish D, Krause PJ. Humans infected with relapsing fever spirochete Borrelia miyamotoi, Russia. Emerg Infect Dis. 2011; 17:1816–1823. [PubMed: 22000350]

Porcella SF, Raffel SJ, Anderson DE Jr. Gilk SD, Bono JL, Schrumpf ME, Schwan TG. Variable tick protein in two genomic groups of the relapsing fever spirochete Borrelia hermsii in western North America. Infect Immun. 2005; 73:6647–6658. [PubMed: 16177341]

Posada D, Crandall KA. MODELTEST: testing the model of DNA substitution. Bioinformatics. 1998; 14:817–818. [PubMed: 9918953]

Ras NM, Lascola B, Postic D, Cutler SJ, Rodhain F, Baranton G, Raoult D. Phylogenesis of relapsing fever Borrelia spp. Int J Syst Bacteriol. 1996; 46:859–865. [PubMed: 8863409]

Rich SM, Armstrong PM, Smith RD, Telford SR 3rd. Lone star tick-infecting borreliae are most closely related to the agent of bovine borreliosis. J Clin Microbiol. 2001; 39:494–497. [PubMed: 11158095]

Rivera MC, Jain R, Moore JE, Lake JA. Genomic evidence for two functionally distinct gene classes. Proc Natl Acad Sci U S A. 1998; 95:6239–6244. [PubMed: 9600949]

Rollend L, Fish D, Childs JE. Transovarial transmission of Borrelia spirochetes by Ixodes scapularis: a summary of the literature and recent observations. Ticks Tick Borne Dis. 2013; 4:46–51. [PubMed: 23238242]

Schutzer SE, Fraser-Liggett CM, Qiu WG, Kraiczy P, Mongodin EF, Dunn JJ, Luft BJ, Casjens SR. Whole-genome sequences of Borrelia bissettii, Borrelia valaisiana, and Borrelia spielmanii. J Bacteriol. 2012; 194:545–546. [PubMed: 22207749]

Schwan TG, Battisti JM, Porcella SF, Raffel SJ, Schrumpf ME, Fischer ER, Carroll JA, Stewart PE, Rosa P, Somerville GA. Glycerol-3-phosphate acquisition in spirochetes: distribution and biological activity of glycerophosphodiester phosphodiesterase (GlpQ) among Borrelia species. J Bacteriol. 2003; 185:1346–1356. [PubMed: 12562805]

Schwan TG, Raffel SJ, Schrumpf ME, Policastro PF, Rawlings JA, Lane RS, Breitschwerdt EB, Porcella SF. Phylogenetic analysis of the spirochetes Borrelia parkeri and Borrelia turicatae and the potential for tick-borne relapsing fever in Florida. J Clin Microbiol. 2005; 43:3851–3859. [PubMed: 16081922]

Scoles GA, Papero M, Beati L, Fish D. A relapsing fever group spirochete transmitted by Ixodes scapularis ticks. Vector Borne Zoonotic Dis. 2001; 1:21–34. [PubMed: 12653133]

Scott MC, Rosen ME, Hamer SA, Baker E, Edwards H, Crowder C, Tsao JI, Hickling GJ. High-prevalence Borrelia miyamotoi infection among wild turkeys (Meleagris gallopavo) in Tennessee. J Med Entomol. 2010; 47:1238–1242. [PubMed: 21175079]

Smith RD, Miranpuri GS, Adams JH, Ahrens EH. Borrelia theileri: isolation from ticks (Boophilus microplus) and tick- borne transmission between splenectomized calves. Am J Vet Res. 1985; 46:1396–1398. [PubMed: 4026019]

Subramanian G, Sekeyova Z, Raoult D, Mediannikov O. Multiple tick-associated bacteria in Ixodes ricinus from Slovakia. Ticks Tick Borne Dis. 2012; 3:406–410. [PubMed: 23182274]

Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics. 1989; 123:585–595. [PubMed: 2513255]

Takahashi Y, Fukunaga M. Physical mapping of the Borrelia miyamotoi HT31 chromosome in comparison with that of Borrelia turicatae, an etiological agent of tick-borne relapsing fever. Clin Diagn Lab Immunol. 1996; 3:533–540. [PubMed: 8877131]

Takano A, Goka K, Une Y, Shimada Y, Fujita H, Shiino T, Watanabe H, Kawabata H. Isolation and characterization of a novel Borrelia group of tick-borne borreliae from imported reptiles and their associated ticks. Environ Microbiol. 2010; 12:134–146. [PubMed: 19758349]

Takano A, Sugimori C, Fujita H, Kadosaka T, Taylor KR, Tsubota T, Konnai S, Tajima T, Sato K, Watanabe H, et al. A novel relapsing fever Borrelia sp. infects the salivary glands of the molted hard tick, Amblyomma geoemydae. Ticks Tick Borne Dis. 2012; 3:259–261. [PubMed: 22910061]

Teeling H, Gloeckner FO. RibAlign: a software tool and database for eubacterial phylogeny based on concatenated ribosomal protein subunits. BMC Bioinformatics. 2006; 7:66. [PubMed: 16476165]

Yang Z. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. J Mol Evol. 1994; 39:306–314. [PubMed: 7932792]

Zhong J, Skouloubris S, Dai Q, Myllykallio H, Barbour AG. Function and evolution of plasmid-borne genes for pyrimidine biosynthesis in Borrelia spp. J Bacteriol. 2006; 188:909–918. [PubMed: 16428394]

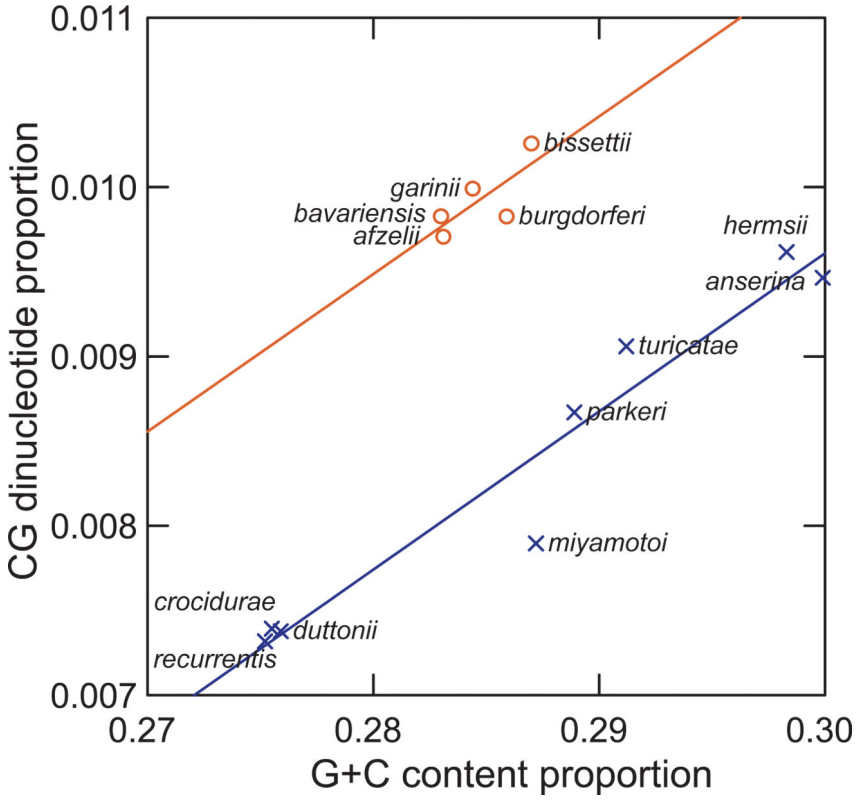

Figure 1.

Scatter plot with linear regression of the CG dinucleotide content on G+C nucleotide content of chromosomes of relapsing fever group *Borrelia* spp. (red) and Lyme disease group *Borrelia* spp. (blue).

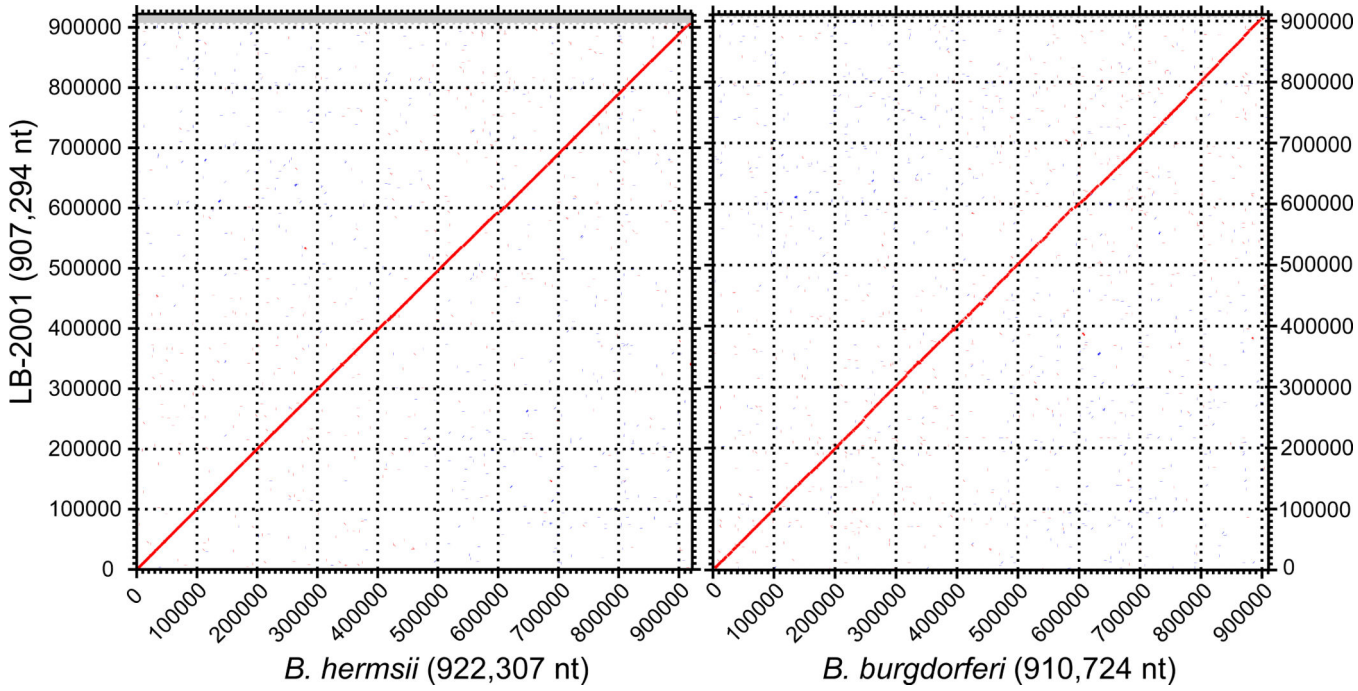

Figure 2.

Dot matrix plots of aligned LB-2001 chromosome sequence against *B. hermsii* strain DAH chromosome sequence (left panel) and *B. burgdorferi* strain B31 chromosome sequence. Lines shown are similarities on direct strands (red) or on opposite strands (blue). Sequences were aligned using the default settings of the MAFFT alignment program.

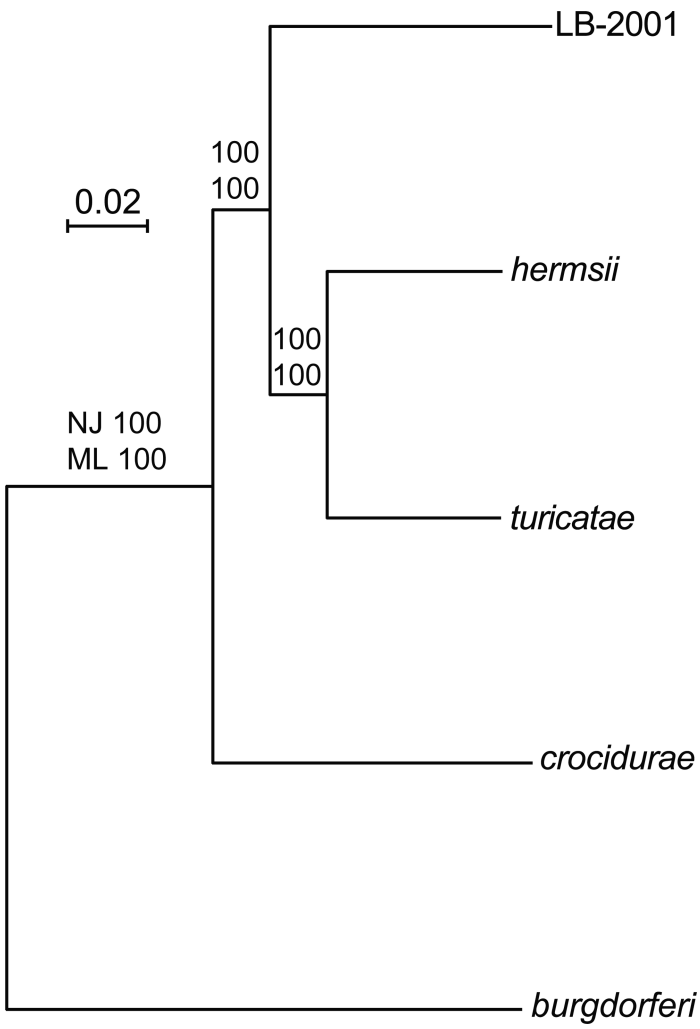

Figure 3.

Phylogram of ungapped aligned chromosome sequences of LB-2001 and four other *Borrelia* species by BioNJ protocol for observed differences (NJ) or by maximum likelihood (ML) with the General Time Reversible model. Nodes with bootstrap values of >70% support after 1000 replicates for NJ and 100 replicates for ML are shown. The scale bar represents nucleotide substitutions per site.

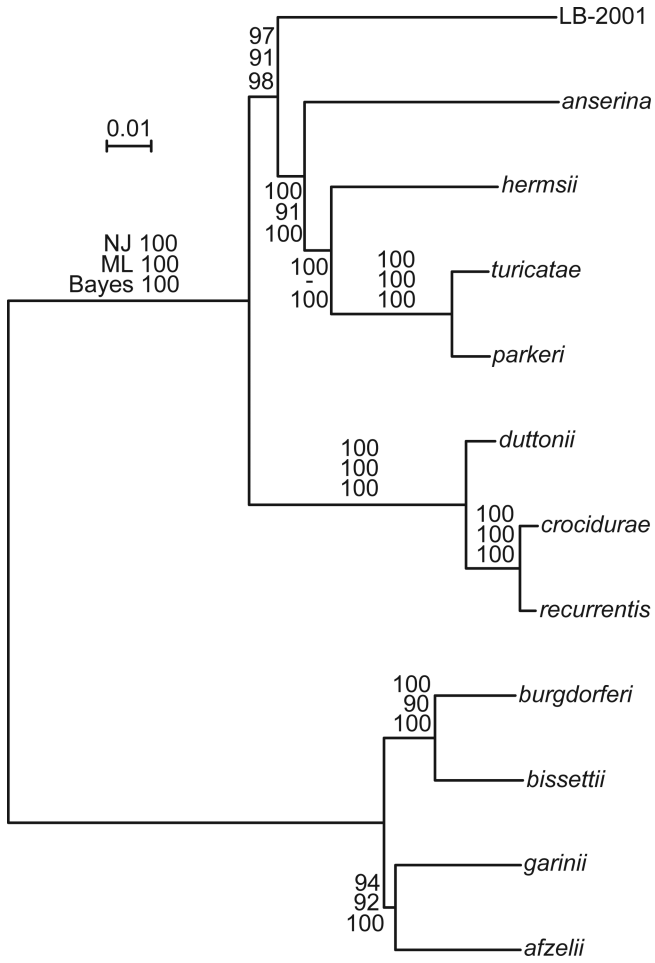

Figure 4.

Phylogram of codon-aligned ungapped concatenated nucleotide sequences of 19 ribosomal protein genes of LB-2001 and 11 other *Borrelia* species as inferred by neighbor-joining (NJ; 1000 replicates), maximum likelihood (ML; 100 replicates), and Bayesian analysis (see section 2.3). The evolutionary model for ML and Bayesian analysis was General Time Reversible (GTR) with a discrete gamma model of 4 categories and a shape parameter of 0.25; it began with a preliminary BioNJ tree. Node values are percent. The scale bar represents nucleotide substitutions per site.

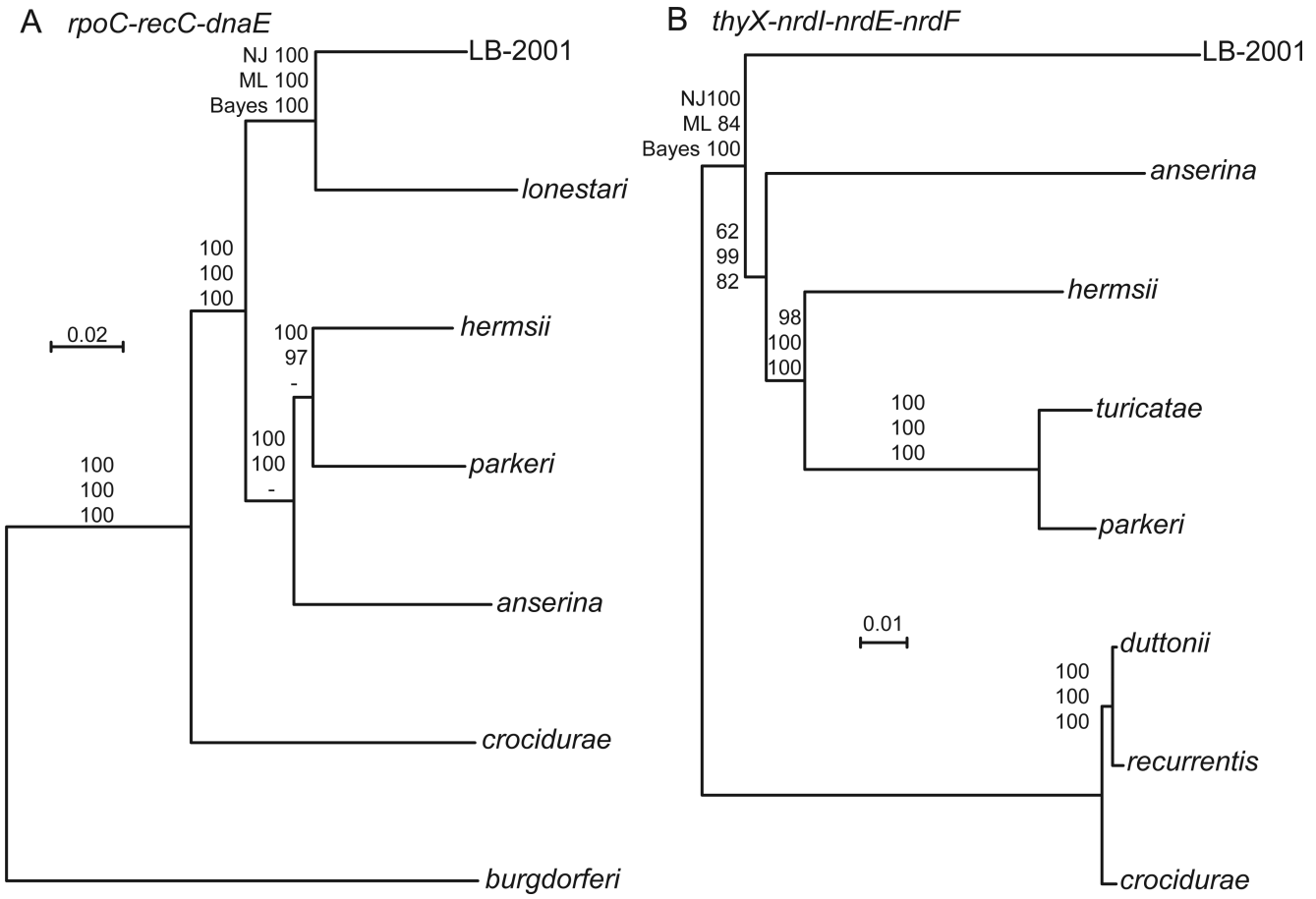

Figure 5.

Phylogram of codon-aligned ungapped concatenated nucleotide sequences for chromosome-encoded *rpoC*, *recC*, and *dnaE* (panel A) and plasmid-encoded *thyX*, *nrdI*, *nrdE*, and *nrdF* (panel B) of LB-2001 and other *Borrelia* species, as inferred by neighbor-joining (NJ; 1000 replicates), maximum likelihood (ML; 1000 replicates), and Bayesian analysis (see section 2.3). For models for both were GTR with discrete gamma model of 4 categories and a shape parameter of 0.29. The scale bar represents nucleotide substitutions per site.

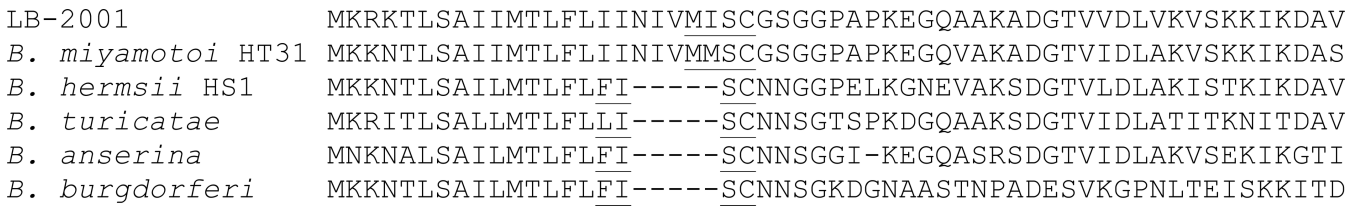

Figure 6.

Alignment of N-termini of selected unprocessed Vsp proteins of LB-2001, *B. miyamotoi* sensu stricto HT31, *B. hermsii* DAH, *B. turicatae* Oz1, and *B. anserina* Es, as well as the homologous OspC protein of *B. burgdorferi* B31. Underlined residues are the consensus signal peptidase 2 sites in each protein.

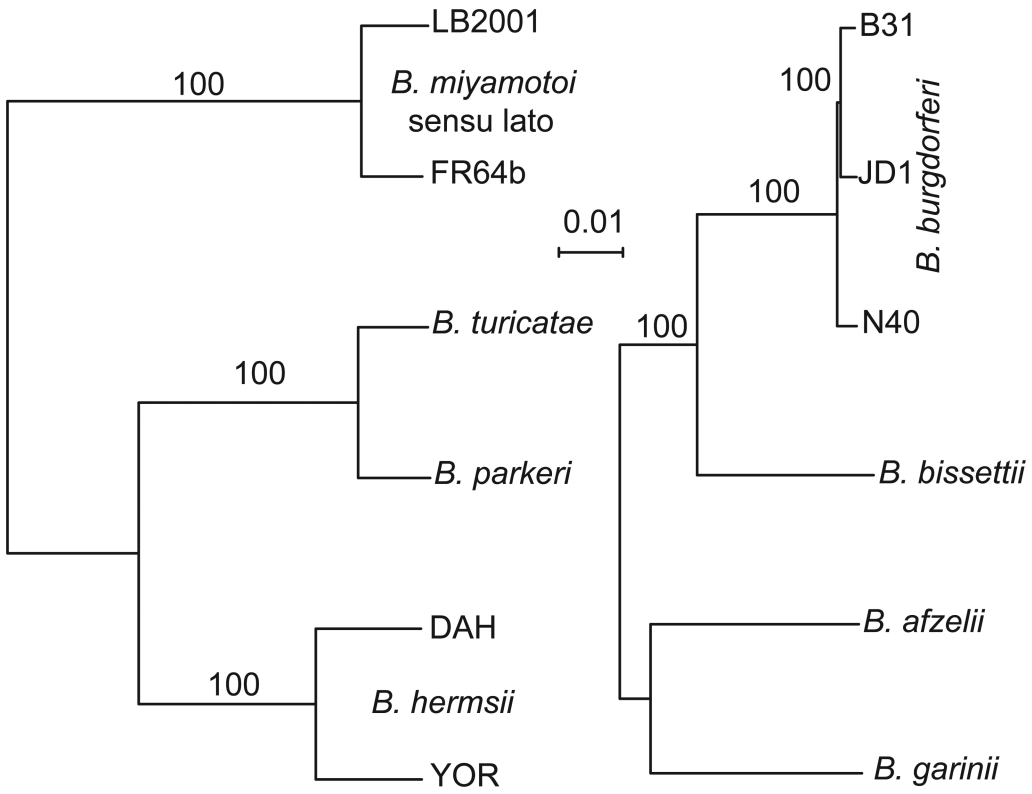

Figure 7.

Phylograms of ungapped aligned chromosome sequences of selected relapsing fever group *Borrelia* species and strains (left; 891,745 sites) and Lyme disease group species and strains (right; 893,252 sites) by BioNJ neighbor joining protocol for observed differences. Nodes with bootstrap values of >70% support after 1000 replicates. The phylograms are to the same scale. The bar represents nucleotide substitutions per site.

Table 1

Pairwise identity between *Borrelia* spp. chromosomes at 855,554 aligned ungapped sites

	LB-2001	***B. hermsii***	***B. turicatae***	***B. crocidurae***	***B. burgdorferi***
LB-2001	1.00	0.868	0.867	0.827	0.738
B. hermsii		1.00	0.909	0.847	0.739
B. turicatae			1.00	0.848	0.741
B. crocidurae				1.00	0.737
B. burgdorferi					1.00