Large linear plasmids of *Borrelia* species that cause relapsing fever

Running title: Large linear plasmids of RF *Borrelia* species

Shelley Campeau Miller¹,†, Stephen F. Porcella², Sandra J. Raffel³, Tom G. Schwan³, and Alan G. Barbour¹

¹ Departments of Microbiology & Molecular Genetics, Medicine, and Ecology & Evolutionary Biology, University of California Irvine, Irvine, CA 92697, USA

² Genomics Unit, Research Technologies Branch, Rocky Mountain Laboratories, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Hamilton, MT 59840, USA

³ Laboratory of Zoonotic Pathogens, Rocky Mountain Laboratories, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Hamilton, MT 59840, USA

† Corresponding author at current address: Department of Pathology and Laboratory Medicine, David Geffen School of Medicine at UCLA, Brentwood Annex, 10833 Le Conte Ave, Los Angeles, CA 90095-1732, USA
ABSTRACT

*Borrelia* species of relapsing fever (RF) and Lyme disease (LD) lineages have linear chromosomes and both linear and circular plasmids. Unique to RF species, and little characterized to date, are large linear plasmids of ~160 kilobases, or ~10% of the genome. By a combination of Sanger and next-generation methods, we determined the sequences of large linear plasmids of two New World species: *Borrelia hermsii* to completion of its 174 kb length, and *B. turicatae* partially to 114 kb of its 150 kb. These sequences were then compared to corresponding sequences of the Old World species *B. duttonii* and *B. recurrentis* and to plasmid sequences of LD *Borrelia* species. The large plasmids were largely collinear, except for their left ends, about 27 kb of which were inverted in New World species. Approximately 60% of the *B. hermsii* lp174 plasmid sequence was repetitive for 6 types of sequence, and half its open reading frames were hypothetical proteins, not discernibly similar to proteins in the database. The central ~25 kb of all 4 linear plasmids were syntenic for orthologous genes for plasmid maintenance or partitioning in *Borrelia* species. Of all the sequenced linear and circular plasmids in *Borrelia* species, the large plasmid’s putative partition/replication genes were most similar to those of the 54 kb linear plasmids of LD species. Further evidence for shared ancestry was the observation that two of the hypothetical proteins were predicted to be structurally similar to the LD species’ CspA proteins, which are encoded on the 54 kb plasmids.
INTRODUCTION

Plasmids are rich sources of variation for the adaptive evolution of bacteria (1). These extrachromosomal replicons are usually circular in structure, but linear plasmids have been noted in some bacteria taxa, including the genus *Borrelia*, which has linear chromosomes as well (2-4). Other bacterial genera with linear plasmids include *Streptomyces* and *Rhodococcus* species (5, 6). Some of the larger plasmids of these genera have been well characterized and shown to be largely occupied by genes involved in antibiotic production and metabolic processes (7, 8). Comparatively less is known about the linear plasmids of *Borrelia* species.

*Borrelias* are host-associated, arthropod-transmitted spirochetes. The genus includes the agents of Lyme disease (LD), such as *B. burgdorferi*, *B. afzelii*, and *B. garinii*, and of relapsing fever (RF), such as *B. hermsii* and *B. turicatae* of the New World and *B. duttonii* and *B. crocidurae* of the Old World. Although *Borrelia* species also have circular plasmids, most of their extrachromosomal content is in the form of linear replicons that range in length from 5 to ~200 kilobases (2-4, 9). The individual plasmids exist in approximately the same copy number as the chromosome in the cells and, as a set with a single chromosome, constitute one of the genomes in these polyploid organisms (3, 10, 11).

The telomeres of the *Borrelia* linear plasmids are hairpins that form dimer junctions during replication, which are recognized and resolved by a telomere resolvase, ResT (12, 13). Plasmid compatibility of the segmented *Borrelia* genome involves orthologous sets of proteins, one of which, designated Pfam32, is homologous to ParA, a plasmid partitioning protein found in other bacteria (14-16). Other *Borrelia* plasmid-encoded proteins of these sets, such as Pfam49, appear to be unique to this genus, are widely distributed on both linear and circular replicons, and presumably are also involved in replication and compatibility. One isolate of *B. hermsii* had
a large plasmid that existed as a circle (17), but more typically these replicons are stably-maintained in their full-length linear forms.

While LD and RF species share many characteristics, such as a dependence on arthropods for transmission and in most of their nutritional requirements, they differ in biology in other respects. LD *Borrelia* species in their vertebrate hosts cause a transient, low-grade bacteremia followed by a persistent infection in the skin and various organs (18). They are transmitted between a variety of different host species by hard ticks, and transovarial transmission is non-existent or rare (reviewed in (19)). In contrast, RF *Borrelia* species cause infections of comparatively shorter duration, but cell densities in the blood reach $10^7$ or higher per milliliter. With the exception of *B. recurrentis*, which uses the human body louse as a vector, RF *Borrelia* species are transmitted by soft tick vectors, and transovarial transmission is common (19).

Compared to LD *Borrelia* species, for which genome sequence is available for 20 or more strains among 5 species (20-23), less is known about the genomes of RF *Borrelia* species. Reportedly complete chromosome and plasmid sequences are available for only three Old World species: *B. duttonii*, *B. recurrentis*, and *B. crocidurae* (24, 25). The report on the genomes for *B. duttonii* and *B. recurrentis* focused on the chromosome and some of the smaller plasmids, with limited annotation of the features of the large linear plasmids of those species (24). To date, only a few genes have been mapped to the large RF linear plasmid, leaving much of the plasmid undefined (26-30).

Our goal was to provide for a fuller understanding of these large linear plasmids. To that end we determined the complete sequence of the 174 kb linear plasmid of *B. hermsii* and most of the sequence of the 150 kb plasmid of *B. turicatae*, another New World species. A
combination of Sanger sequencing and next-generation sequencing of genomic DNA and gap-
spanning PCR products was performed for this work. These sequences were compared with
the similarly sized linear plasmids of *B. duttonii*, *B. recurrentis*, and *B. crocidurae*. We found
that the large plasmids of New World and Old World RF species were syntenic over most of
their lengths and predominantly conserved in the central regions of their linear sequence.
Comparisons with a wide variety of *Borrelia* plasmids revealed large portions of these large
plasmids to be orthologous to genes of the ubiquitous lp54-type plasmids of LD *Borrelia*

**METHODS**

**Bacterial strains and growth conditions.** *B. hermsii* strain HS1 serotypes 7 and 33 (31) and
strain DAH (32) from eastern Washington, *B. turicatae* strain 91E135 (also known as “Oz1”)
from western Texas (33, 34), and *B. parkeri* strain HR1 from northern California were used.
Isolates HS1 and DAH are the same strain on the basis of restriction fragment length
polymorphism analysis and by multilocus sequence typing (35, 36). Their ~950 kb long
chromosomes are >99.99% identical in sequence (A.G.B., unpublished findings). *B. turicatae*
and *B. parkeri* are closely-related on the basis of sequences of several genes (36). Low-
passage isolates were cultured in BSK II medium with 6% or 12% rabbit serum and grown at
34°C, unless otherwise noted (37). Total DNA of *B. duttonii* strain Ly was kindly provided by
Sven Bergström, Umeå University, Sweden. Spirochetes were harvested by centrifugation at
9,500 x g for 20 min and washed twice with phosphate-buffered saline (PBS)-5 mM MgCl₂ at pH
7.4 (PBS-Mg).
Mouse infections. Female, 5- to 6- week-old CB17 Severe Combined Immunodeficiency (SCID) mice (Charles River Laboratories, Wilmington, MA) were inoculated intraperitoneally with 10-50 cells of serotype 7 of *B. hermsii* strain HS1. The infection was monitored by phase-contrast microscopy of a wet mount of tail vein blood. The mice were euthanized when there were \( \sim 10^8 \) spirochetes per ml of blood, and blood was collected by cardiac puncture with syringes coated with 3% sodium citrate solution.

Pulsed-field gel electrophoresis. Agarose plugs of genomic DNA were prepared as described previously with modifications (17, 31). In brief, low passage cells were harvested by centrifugation for 15 min at 8,000 x g at room temperature, washed with 150 mM NaCl–50 mM Tris–1 mM EDTA (TN buffer) with final resuspension in TN buffer. An equal volume of molten 2% low-melting-point SeaPlaque\textsuperscript{®} GTG\textsuperscript{®} agarose (Lonza, Rockland, ME) in TN buffer was added giving a final concentration of 10\(^9\) cells ml\(^{-1}\); 80 µl of the mixture was poured into each casting well (Bio-Rad, Hercules, CA). Agarose plugs were submerged in lysis solution containing 1 mg ml\(^{-1}\) proteinase K (Roche, Mannheim, Germany) in 50 mM Tris-50 mM EDTA-1% sodium dodecyl sulfate (SDS). Lysis was performed at 50°C for 24 h. Treated agarose blocks were washed twice in 10 mM Tris-1 mM EDTA (TE) buffer and one-third of the plug was loaded into wells of a 0.8% agarose gel. Pulsed-field gel electrophoresis was performed in circulating 0.5x TBE buffer (45 mM Tris-borate and 1 mM EDTA) at 14°C for 21 h on a CHEF Mapper\textsuperscript{®} XA apparatus with the following parameters: 6V cm\(^{-1}\), included angle of 120°, initial switch time of 50 s ramped to a final switch time of 90 s.

Restriction digestion and Southern blot analysis. For restriction enzyme digestion, total genomic DNA was isolated using a DNeasy Tissue Kit per manufacturer’s instructions; approximately 1 µg of DNA was used in each restriction digestion reaction. Digested fragments
were separated on a 0.7% TBE gel, and Southern blot analysis was carried out as described previously (27). Prehybridization and hybridization solutions contained 6X SSC, 5X Denhardt's, 0.5% SDS, 100 μg ml⁻¹ salmon sperm DNA and 50% formamide. A probe for the 5' end of the plasmid and corresponding to the fhbA gene (Table S1) was made using PCR amplicons as template DNA, purified using the Zymo DNA Clean and Concentrator kit (Zymo Research, Irvine, CA) and subsequently biotinylated by random labeling using the Phototope Labeling Kit (New England Biolabs, Ipswich, MA) per manufacturer's instructions. Detection of the hybridized probe was performed with the Phototope Detection Kit (New England Biolabs). A 60-mer oligo probe for the 3' end of the plasmid corresponding to part of bha167 was biotinylated at the 5' end (Sigma-Aldrich, St Louis, MO) (Table S1).

Polymerase chain reaction procedures. Genomic DNA, which was isolated using a DNeasy Tissue Kit (QIAGEN, Germantown, MD), was used as the template for amplification using Phusion DNA polymerase and HF buffer (Finnzymes, Woburn, MA). For all PCR reactions, the initial denaturation step at 98°C for 3 min was followed by 35-40 cycles of denaturation at 98°C for 10 sec, annealing for 30 sec and extension at 72°C for 2 min with a final 7 min extension at 72°C. Primer sets and annealing temperatures are listed in Supplementary Table S1. For reverse transcription and quantitative PCR, total RNA from (i) cultured B. hermsii cells was stabilized in RNAprotect Bacteria Reagent and then extracted using the QIAGEN RNeasy Mini Kit with automation by the QiaCube (QIAGEN), or (ii) RNA from cells in mouse blood was isolated using the Mouse RiboPure™-Blood RNA Isolation Kit (Ambion, Austin, TX). RNA was DNase treated with the DNA-free kit (Ambion) for 30 min at 37°C according to manufacturer’s instructions. The RNA was then reverse-transcribed to cDNA using random hexamers as primers and a Taqman® Gold RT-PCR kit (Ambion) with and without reverse transcriptase at
42°C for 30 min followed by incubation at 95°C. After cDNA synthesis, real-time PCR amplification was performed using Eurogentec's qPCR Master Mix Plus (AnaSpec, Fremont, CA) and a Rotor-Gene 3000 thermal cycler (QIAGEN). The real-time PCR reaction mixture contained cDNA, 0.3 μM of each primer, 0.25 μM of dual-labeled probe and 0.5 volumes qPCR Master Mix Plus. Samples and standards were run in duplicate. Reaction conditions were as follows: 50°C for 2 min, 95°C for 10 min, 40 cycles at 95°C for 15 s, and 60°C for 1 min. Primer and probe sets are listed in Supplementary Materials Table S1. Amplification curves were analyzed using Rotor Gene software version 6.0 (QIAGEN). Standard curves using 10-fold dilutions of circular plasmid DNA with known concentrations were generated to determine the relative amounts of cDNA. Statistical analysis, including a General Linear Model (GLM) procedure with default settings, was carried out with STATA version 10 (STATA Corp., College Station, TX).

**Genome sequencing of B. hermsii strain DAH and B. turicatae strain 91E135.** Sanger sequencing of shotgun libraries of 2 – 3 kb-sized fragments of genomic DNA was carried out as described previously (31, 38). For the present study, pyrosequencing using 454 FLX technology (Roche, Branford, CT) also was performed following the manufacturer’s recommended protocols. Fragment- and mate-paired libraries, each at 30X coverage, were obtained for both genomes. De novo assemblies were generated using Newbler (http://www.454.com).

The complete large plasmid sequence for *B. hermsii* strain HS1 was obtained by next-generation Illumina sequencing with a final coverage ≥ 10X (described in more detail below). As studies have previously shown, *B. hermsii* strains HS1 and DAH, both originating from the same location in eastern Washington, are very close genetically and thought to be perhaps isolates of the same strain. Initially, the *B. hermsii* DAH genome was sequenced by a combination of
Sanger and 454-based methods, resulting in sequences with \( \geq 30X \) coverage, and assembled

*de novo* at Rocky Mountain Labs (Hamilton, MT). The assembly of the large plasmid for *B. hermsii* HS1 was achieved by mapping against the preliminary assembly of the DAH large linear

plasmid. The latter sequence contained gaps for which primers were designed for use in PCR.

**Genome sequencing of *B. hermsii* strain HS1.** Genomic DNA was isolated using a DNeasy

Tissue Kit (QIAGEN). DNA libraries were sheared to an average size of between 150 and 200

bp, repaired to generate blunt ends, and adapters ligated to ends. Fragments with ligated

adaptor sequences were then amplified and validated by an Agilent Bioanalyzer (Santa Clara,

CA) and subsequently quantified by quantitative PCR (qPCR) using a KAPA Library

Quantification Kit (Kapa Biosystems, Woburn, MA). Based on qPCR concentrations, libraries

were normalized to 10 nM in 10 µl volume and from here were cluster generated for data

processing with an Illumina HiSeq 2000 NGS platform. The depth of coverage was \( \geq 10X \).

Sequences were then mapped to a reference assembly of the *B. hermsii* DAH large plasmid

(described below) using CLC Genomics Workbench version 5.1 (CLC bio, Aarhus, Denmark).

After the final assembly of the *B. hermsii* HS1 large linear plasmid sequence was done,

remaining gaps were manually closed by PCR followed by primer-directed Sanger sequencing

of the products at GENEWIZ (La Jolla, CA). Sequence was confirmed by resequencing of the

genome as follows: DNA extracted from culture harvest was first treated with RNase I

(Fermentas, Burlington, Ontario), purified as above, and then enzymatically sheared with an Ion

Express Plus Fragment Library Kit (Life Technologies, Grand Island, NY). Products were size-

selected by gel purification with the E-Gel system (Invitrogen). Templates were prepared by

emulsion PCR on a Ion Torrent OneTouch apparatus (Life Technologies), and these were

sequenced on an Ion Torrent Personal Genome Machine with Ion 314 chips (Life Technologies).
Single reads of ~100 bases were assembled into de novo contigs or mapped to reference sequences using CLC Genomics Workbench version 5.1.

**Annotation.** Open reading frames (ORFs) of ≥ 150 bp were identified with GLIMMER v. 3.02 and AMIGene and then manually curated for consensus start sites (39, 40). All putative ORFs were searched against the National Center for Biotechnology Information (NCBI) non-redundant protein database using blastp (Supplemental tables 2 and 3) (41). The BLAST search criteria for designating an ORF as homologous to a deduced protein of a LD *Borrelia* species or another organism were the following: *e*-values < 10^{-5}, pairwise amino acid identities >20%, and > 50% coverage of the smaller protein (42, 43). ORFs were analyzed using a collection of protein families represented by hidden Markov models (HMM) constructed from family member seed alignments (Protein family-A) or from un-annotated and automatically generated non-redundant clusters (Protein family-B) with Pfam version 24.0 using HMMER3 (http://pfam.sanger.ac.uk/) (44). Putative lipoproteins with signal peptidase II cleavage sites were identified using LipoP 1.0 (http://www.cbs.dtu.dk) (45). ORF names were assigned based on designations given to LD *Borrelia* plasmids: the first two letters representing the genus and species abbreviation, a third letter representing the plasmid compatibility group, and a gene number, starting from the arbitrarily designated left end of a linear plasmid.

**DNA sequence analysis.** Distance phylograms were constructed using maximum-likelihood protocol of PhyML, as implemented by the SeaView suite version 4.2.12 (46). Global alignments to identify rearrangements and co-linearity within the large plasmids were performed using Progressive MAUVE 2.1.0 aligner software according to the manufacturer’s instructions with a seed weight of 17 and local collinear block (LCB) weight of 500 (47). Total GC skews of large plasmids or portions of plasmids were calculated by determining (G-C)/(G+C) with settings
described in the text (http://www.genomicsplace.com). DNA matrix dot plots were constructed using the multiple alignment program MAFFT version 6 (http://mafft.cbrc.jp), which uses a fast Fourier transform (74). Default settings were a scoring matrix of 200 PAM/\kappa = 2, a gap opening penalty of 1.53, and a plot threshold score of 39 (E = 8.4e^{-11}). For a deduced amino acid sequence, virtual protein folding and then comparison of the predicted structure with the protein structure database was carried out on the UC Santa Cruz Bioinformatics server (http://compbio.soe.ucsc.edu) as described by Karplus et al. (48). Predicted tertiary structures were visualized with kiNG software (http://kinemage.biochem.duke.edu).

**Nucleotide sequence accession numbers.** The complete sequence for *B. hermsii* HS1 lp174 plasmid and the partial sequence for *B. turicatae* 91E135 lp150 plasmid were assigned GenBank numbers HM008709 and HM008710, respectively. Sequences for *B. hermsii* HS1 bha064 and bha065 and *B. turicatae* 91E135 bta037 and bta038 have accession numbers GQ141864 to GQ141867, respectively. Sequences used in the phylogram for Pfam32 genes were the following: *B. afzelii* ACA-1 lp54 (CP001247), cp32-3 (CP001237), cp26 (CP001250), and lp28-7 (CP001242); *B. afzelii* PKo chromosome (CP002933); *B. burgdorferi* B31 lp54 (AE000790), cp32-3 (AE001576), cp26 (AE000792), lp28-1 (AE000794), and chromosome (AE000783); *B. duttonii* Ly lp165 (CP000979), lp26 (CP000982), lp35 (CP000985), cp26 (CP000980), and chromosome (CP000976); *B. garinii* PBr lp54 (CP001308), cp32-10 (CP001306), cp26 (CP001305), and lp28-1 (CP001310); *B. hermsii* HS1 lp174 (HM008709), cp32-like (AF209440), lp53 (JN232111), lp28-2 (DQ172919), and chromosome (CP000048); *B. lonestari* LS-1 lp28-like (EF507519); *B. turicatae* 91E135 lp150 (HM008710) and chromosome (CP000049).
RESULTS

Sizes of the large linear plasmids. We estimated the sizes of the large linear plasmids of *B. hermsii, B. parkeri,* and *B. turicatae* by pulsed-field gel electrophoresis under conditions optimized for separation of linear duplex DNA in the range of 50 to 1000 kb. Size standards were bacteriophage lambda concatemers, *Saccharomyces cerevisiae* chromosomes, and the sized linear chromosome and plasmids of the *B. burgdorferi* type strain B31 (16, 43). Figure 1 shows a gel of the intact linear replicons of *B. hermsii, B. parkeri,* and *B. burgdorferi.* The sizes of the large plasmids of *B. hermsii* and *B. parkeri* were estimated to be 174 kb and 150 kb, respectively. In another gel the large linear plasmids of the closely-related species *B. parkeri* and *B. turicatae* had identical migrations (data not shown). Accordingly, the *B. hermsii (Bh)* and *B. turicatae (Bt)* large plasmids were designated lp174 and lp150, respectively.

Plasmid sequencing. We carried out whole genome sequencing of *B. hermsii* DAH and *B. turicatae* 91E135 by a combination of Sanger and 454 technologies, while *B. hermsii* HS1 was pursued with Illumina and Ion Torrent technologies. Direct Sanger sequencing of PCR products or their clones in plasmid vectors was used for gap closure, and the deduced physical map was confirmed by Southern blots of selected restriction fragments. While the sequencing of the *B. hermsii* large linear replicon began with strain DAH, the sequences of several other *B. hermsii* phylotyping genes suggested that HS1 and DAH were near identical or the same strain (31, 36). Our initial comparison of randomly-selected homologous sequences of the large plasmids of the two isolates supported this conclusion (unpublished findings). Accordingly, large contigs accounting for a total of 163,905 base pairs of lp174 of strain DAH generated from Sanger and 454 technologies, were used as a reference for the mapping of reads from the Illumina-based sequencing of HS1. Closure of gaps was carried out on HS1 DNA, and so that is the strain...
designation of record for this species' replicon described here. The completed assembly totaled 173,739 bp. The completed HS1 sequence and a draft DAH sequence were identical at 173,443 (99.8%) out of 173,739 sites compared in the alignment.

Recognizing that telomeric regions by nature of their structure may be difficult to sequence, we carried out Southern blot analysis with probes for the presumed left or right ends of lp174. The probe for the left end was based on an open reading frame that was identical to the \( fhbA \) gene of \( B. hermsii \) and subsequently numbered \( bha008 \). The probe was directed against the entire \( bha008 \) gene plus an additional 100 nucleotides flanking each side. The 60-mer oligo probe for the right end was based on an open reading frame subsequently designated as \( bha167 \). Table 1 gives the expected sizes of the hybridizing restriction fragments, based on the sequence assembly, and the observed sizes from the Southern blot analyses. Overall, the observed fragments with 6 different restriction enzymes were within 0.2 to 1.5% of what was predicted. Therefore, these results indicate that we either have or are very close to having the complete lp174 plasmid for HS1.

The base count for the final assembly of the \( B. turicatae \) lp150 plasmid was 114,195 bp. As detailed below, the estimated 35 kb of missing sequence was primarily at the right end of the plasmid. However, it is possible that sequence at the left end of the lp150 was also unaccounted for here. The lists of ORFs that were identified for the entire \( Bh \) lp174 and for most of \( Bt \) lp150 are given in Tables S2 and S3, respectively.

**General features.** The locations and extent of repetitive DNA are indicated in the dot plot matrix of the \( Bh \) lp174 sequence plotted directly against itself (Figure 2, panel A). Approximately 40% of the plasmid’s content is repetitive DNA, and this is mainly located in regions we designated as A, B, C, D, and E. Inverted repeats were infrequent and short.
Region D comprises tandemly-arrayed paralogous ORFs of various lengths and of unknown function. Database searches indicated that similar genes were found only in RF *Borrelia* species and then in varying numbers. Region E at the right end is a 5514 nt long open reading frame, which is homologous to a partial, unmapped ORF of *B. crocidurae* (accession number AFI32145) and it contains multiple copies of a ~400 nt long sequence with 70-80% identity between copies. Regions A-E of *Bh* lp174, which contain repetitive sequence, correspond to similar locational regions of repetitive sequence in the large plasmid of *B. turicatae* (Figure S1A), and to the large plasmid of *B. duttonii* (Figure S1B). Notably, repetitive regions A, B, C, and D appear locationally conserved in all three species, while region E appears variable relative to the species origin of the linear plasmid. Therefore, there appears to be a level of conservation of repetitive domains across the large linear plasmid from different species of RF *Borrelia*.

The large plasmids of *B. hermsii*, *B. turicatae*, and *B. duttonii* have an approximate G+C percentage of 30%, which is within the range of G+C percentage for both the chromosome and most plasmids of *Borrelia* species. The coding densities of the large plasmids are at ~75%, which is less than the coding densities of the chromosomes of these species (24, 25) (Accession No. CP000048 and CP000049). This lower level of coding density is however, similar to the coding densities of other linear plasmids of *Borrelia* species (22, 31, 43). Only 5-10% of the deduced ORFs of the 3 large plasmids defined here had identifiable homologs to proteins from organisms outside the *Borrelia* genus (Tables S2 and S3). These ORFs encode products involved with nucleotide biosynthesis (*thyX*, *nrdE*, *nrdF*, and *nrdI*) (38, 49) and putative bacterial plasmid replication and partitioning (Paralogous Family [PFam] 32, PFam49, PFam50, and PFam57) (16, 43). The remaining protein sequences of ORFs were found to be
homologous to proteins found in LD *Borrelia* species as well as in RF species. In addition, homologs were also found that were only present in RF *Borrelia* species and a few short ORFs were found that appeared unique to *B. hermsii* or *B. turicatae* (black ORFs Figures 2C and S2).

In order to analyze general plasmid synteny or look for the presence of large regions of conservation or regions of inversion, an alignment and Mauve Plot analysis was performed and these results are shown in Figure 2B. Three large collinear blocks (*I*, *II*, and *III*) were discovered between the *B. hermsii*, *B. turicatae*, and *B. duttonii* plasmids. These three collinear blocks constitute more than half the total length of each of these plasmids. Block *I* contains genes for two possible virulence factors of RF *Borrelia* species: *fhbA*, which encodes a Factor H-binding protein (26), and *bha007* and its orthologs, which encode proteins that bind complement pathway components or fibronectin (28, 50) (E.R.G. Lewis et al., submitted for publication). The longer length of *B. hermsii* block *I* relative to *B. turicatae* block *I* is attributable to a higher number of repeated “P35” (PFam54) homologs, a family of lipoproteins in LD *Borrelia* species (51). “P35”-like coding sequences were even fewer in the *B. duttonii*, as evidenced by the lack of homology (red lines) within block *I* (Figure 2B) or *B. recurrentis* large plasmids (data not shown), and were not discernible at all in the *B. crocidurae* large plasmid sequence (data not shown).

In the published sequence for the large plasmid of *B. duttonii* (NC011247) block *I* is inverted, involving approximately 27 kb of sequence, with respect to its orientation in the New World species (Figure 2B). We noted a similar inversion for the block *I* region in *B. crocidurae* (NC017778). To verify that this inversion was not the result of an assembly error, we carried out PCR on total DNA isolated from both *B. hermsii* and *B. duttonii*, using 6 pairs of primers that spanned the junction of blocks *I* and *II* for each of the two possible orientations. Each of the six
PCR reactions yielded appropriately sized fragments that support the current assembly (Table S4). Therefore, these results suggest that there was an inversion of this region when this location is compared between New World RF species *B. hermsii* and *B. turicatae*, relative to Old World RF species *B. duttonii* and *B. crocidurae*.

**Comparison with LD *Borrelia* plasmids.** Panel C of Figure 2 is a graphical representation of the 167 ORFs, including three pseudogenes, identified in the *Bh* lp174 plasmid. A comparable map of the 108 ORFs encoded on *Bt* lp150 is provided in Figure S2. These maps demonstrate ORFs that are unique to *B. hermsii* and at least one other RF *Borrelia* species (green) from those that appear to be at least genus-wide (orange, red or yellow), and those unique to a specific species (black) in their distribution. Among the genus-wide group, we further distinguish between ORFs whose closest orthologs are in the lp54 linear plasmid of *B. burgdorferi* and related species and ORFs in common with genes encoded on other *B. burgdorferi* plasmids.

Of the 76 ORFs described for the lp54 plasmid (AE000790), 28 of these ORFs were found to have orthologs on the RF large plasmids. These 28 orthologs appear predominantly in two extended regions of *Bh* lp174, namely, ~79 kb to 96 kb and 121 kb to 132 kb (Figure 2C), which also are largely syntenic with the lp54 linear plasmid of *B. burgdorferi*. In the region 79-96 kb, there are orthologs of the following ORFs of lp54: *bba38-43, bba45-49*, and *bba51-52*. In lp54, *bba50* resides between the genes *bba49* and *bba51*, however, in *B. hermsii* lp174 a 1,424 bp long ORF, designated *bha087*, that is unique to RF species resides in this location. In addition, for the large linear plasmids of *B. turicatae* and *B. duttonii*, genes homologous to *bha087* designated *bta062* and *bdu1086*, respectively, both of which are unique to RF *Borrelia* species, also reside in this location. Further to the right, and present in *B. hermsii* lp174, but not in the large plasmids of *B. turicatae, B. duttonii, B. recurrentis*, or *B. crocidurae*, are *bha099-102*. 

16
(yellow ORFs in Figure 2C), which are orthologs of 4 genes of *B. burgdorferi* circular plasmid 26 (cp26). These genes are three chitobiose transport genes (*chbABC*) and a gene for an integrin-binding protein (52-54). These results suggest that lp174 contains orthologs that reside on LD cp26 and LD lp54.

Continuing further downstream on the HS1 plasmid, we encountered the repetitive sequence region C, described earlier in Figure 2A that resides from ~100 kb to 112 kb and the genes *bha103* to *bha129* (Figure 2C). There were repeats of two alternating sequences, one of which was unique to RF species (*bha107, 109, 111, 113, 115, 118, 120*, and *126*), and the other of which is homologous to the PFam99 family of *B. burgdorferi* (*bha104, 106, 108, 110, 114, 117, 119, and 125*). In the midst of these repeats is the *bdrA* gene (*bha116*) of *B. hermsii* (55).

Paralogous genes were also arranged as tandem repeats in repetitive sequence region C of *B. turicatae* lp150 (Figure S2), but not in the large plasmids of the Old World species. This repetitive sequence region C in *B. hermsii* (112 kb to 121 kb) and *B. turicatae* (71 kb to 92 kb) was followed by a series of ORFs that were either unique to RF species, such as the *alp* gene (*bha128* or *bta096*) for the Arthropod-associated Lipoprotein (30), or common to both RF and LD groups, including 3 ORFs (*bha125-127* or *bta092-094*) with homology to ORFs restricted to the lp38 plasmid among *B. burgdorferi* plasmids.

From 121 kb to 132 kb of *Bh* lp174 there is sequence that is partially syntenic with lp54. Included here are four genes, *bha132-135* (encoding PFam32, PFam49, PFam50, and PFam57) constituting a plasmid maintenance locus (16, 43). All *Borrelia* plasmids characterized to date carry a subset of genes from PFams 32, 49, 50, and either 57 or 62, where PFam32 is homologous to the ParA protein of other bacteria (56). Skew-shifting allows identification of the origin of replication in an otherwise cryptic plasmid. Total GC-skew calculations and plots for *Bh*
lp174 and Bt lp150 show that the skew shifted at the site where this plasmid maintenance gene set is located, and more specifically, within the PFam57 (bha134) gene (Figure 3A), suggesting that this locus may be the origin of replication for the large plasmids (57).

Partition locus genes of lp54 and RF large linear plasmids share a common origin. The four gene products in this locus have 46-80% amino acid identity with their counterparts on the lp54 plasmid. Phylogenetic analysis of Pfam32 genes from LD linear and circular plasmids and the linear chromosomes of both RF and LD Borrelia species indicates that the orthologs of RF species large plasmid homologs are more closely related to Bb lp54 genes than they are to those of any other replicon, including other replicons in their own species (Figure 4). The same phylogenetic relationships were inferred from the sequences of other genes in the plasmid maintenance locus (data not shown). Therefore, plasmid maintenance genes show sequence bias towards the replicon structure that encodes them across species, rather than across replicons within a single species or isolate.

We noticed that in these regions of Bh lp174 (122 kb to 133 kb) and Bt lp150 (92 kb to 103 kb) that had partial synteny to lp54, two sets of lp54 ORFs, namely, bba22-29 and bba14-17, were absent in RF Borrelia species and without replacements by other ORFs. Were these ORFs lost from the lineage leading to B. hermsii and other RF species but not from the LD species lineage? Or, were they gained by the LD Borrelia lineage after the split from a common ancestor? Taking advantage of the plasmid’s sharp shift in GC skew in the midst of the plasmid maintenance locus of lp174 (Figure 3A), we compared the GC skew patterns of these syntenic regions in each replicon. We reasoned that if all or some of the ORFs unique to LD species were relatively recent acquisitions, this might be apparent in the GC skew pattern (58, 59).

Indeed, as analysis of 121 to 132 kb of Bh lp174 and the corresponding span of ORFs bba08-31
of *B. burdorferi* lp54 reveals (Figure 3B), the GC skew around the operons for Outer Surface
Proteins A and B (*bba15* and *bba16*) and for Decorin-Binding Proteins A and B (*bba24* and
*bba25*) is opposite to that of the surrounding sequence that is common to RF and LD species.
This is an indication that these operons, which are unique to LD species, were likely acquired
horizontally in the LD clade after the descent from an ancestral taxon, as appears to be the case
for other operons (49, 60).

**Regions unique to RF* Borrelia* species.** Part of local collinear block II, specifically from ~36
to 75 kb of *Bh* lp174 (Figure 2C) is just left or upstream of a functional pyrimidine biosynthesis
locus containing *thyX* (*bha078*), which encodes a thymidylate synthase (49), and the
ribonucleotide reductase subunit cluster *nrdFEI* (*bha075-077*) (38, 49). The majority of the
coding sequences in this region were on the minus strand (Figure 2C), and range in polypeptide
sizes of 200-300 residues. The proteins in this region appeared to be paralogs with considerable
divergence between each member in the same species and between orthologs in other collinear
block-containing species (data not shown). The exceptions were a pair of ORFs, *bha064* and
*bha065* in *B. hermsii* (Figure 2C), *bta037* and *bta038* in *B. turicatae* (Figure S3), and *bdu1066*
and *bdu1067* in *B. duttonii* (not shown), as well as their counterparts in *B. recurrentis* and *B. crocidurae* (not shown). BHA064 and BHA065 likely are paralogs duplicated prior to the
divergence of RF and LD lineages, with greater identity between corresponding loci of the three
species than between pair members of the same species (Figure S4). For example, BHA065 of
*B. hermsii* and BTA039 of *B. turicatae* were 78% identical in protein sequence, while the percent
identity between BHA065 and BHA064 was only 67%.

**Expression and characterization of *bha064***. There are near-consensus ribosomal binding
sequences (GGAGA) positioned 9 nt in front of the start codons for both *bha64* and *bha65* with
an intergenic distance of 174 bp. Corresponding distances in other RF species were 228 bp
between bta037 and bta038 of B. turicatae and 170 bp between bdu1066 and bdu1067 of B.
duttonii. Accordingly, we assessed the transcription of bha064 by quantitative reverse
transcriptase PCR with 5 replicates each for cells growing in culture medium under different
temperatures (23 vs. 37 °C) and pH (7.6 vs. 8.0) or in the blood of mice. Results of each
replicate were normalized by the values for copies of the constitutively and abundantly
expressed flaB gene for the same replicate. Expression for bha064 was significantly higher
under conditions that mimicked the tick environment, i.e. 23 °C and pH 8.0 (61, 62), compared
to culture medium at 37 °C at both pHs or in mouse blood (Figure 5). By linear regression
analysis the combination of temperature and pH accounted for the expression values in vitro
with a coefficient of 4.1 (95% confidence interval 3.3-4.9) and a coefficient of determination ($R^2$)
of 0.77 ($p < 0.0001$). Normalized expression of bha064 was also higher in vitro at 23 °C at pH
8.0 (5.40 [3.24-7.55]) than in mouse blood (2.02 [1.46-2.59]) (Mann-Whitney U $p = 0.009$).
These findings suggest that bha064 is potentially preferentially expressed in the tick vector
relative the mammalian host environment, but this remains to be determined.

Another clue about the possible function of BHA064 and BHA065 was provided by
prediction of their tertiary structures through virtual protein folding and threading of their amino
acid sequences against known protein structures in the database. Predictions for BHA064 and
BHA065 of B. hermsii, as well as their orthologs in B. turicatae, and B. duttonii, had the same
top alignment hit, PDB structure 1W33. The e-value scores for this call were $<10^{-2}$ with next
best scores $>1$. 1W33 is reference for the crystal structure of the B. burgdorferi CspA protein,
formerly known as the Complement Regulator-Acquiring Surface Protein or CRASP-1 (63)
(Figures S5 and S6). This largely α-helical protein is encoded by gene bba68 on B. burgdorferi
Ip54 and is a member of PFam54 gene family. The closest relative to BBA68 (CspA) is BBA69, a neighboring gene on Ip54, with 59% identity. Thus, LD species have tandem pairs of paralogs like BHA064/065 of B. hermsii, BTA037/038 of B. turicatae, and BDU1066/1067 of B. duttonii.

DISCUSSION

The large linear plasmids of five RF Borrelia species are fairly homologous, as demonstrated by (i) the extensive regions of collinearity between plasmids from different species, (ii) the possession of common biosynthesis genes, like those for ribonucleotide reductase, not present elsewhere in the genome, and (iii) the relatedness of their plasmid maintenance genes. Although there is a set of core genes shared among the large plasmids and an overall similarity in architecture, there is also considerable intermolecular variability, especially at the plasmids' ends, as was noted regarding the linear replicons of LD Borrelia species (16). Unique gene content was present in the large linear plasmid of one species but not in the large linear plasmids of other species. An example is on B. hermsii lp174, there is a cluster of large linear plasmid ORFs that are orthologous to genes associated with cp26 of B. burgdorferi (Figure 2C). But such instances were few. Differences in lengths of the plasmids were largely attributable to variation in the number of repeats within the extensive regions of repetitive DNA. In some cases, such as repetitive Region A with its tandemly-arrayed ORFs of “P35”-like sequences (Figure 2), the repeated DNA was homologous to sequences found in LD Borrelia species Ip54, Ip28-4 and Ip38 replicons (AE000790, AE000789, and AE000787). But in other regions, such as Region D, the repetitive ORFs appeared to be unique to RF species.

The shortest of the large linear plasmids was reported as 124 kb for B. recurrentis (24). But there is evidence that the deposited plasmid sequence for B. recurrentis was not complete.
Grosskinsky et al. demonstrated that *B. recurrentis* has a gene that is an ortholog to the *cihC* gene of *B. duttonii*, *bha007* of *B. hermsii*, and *bta001* of *B. turicatae*, all of which are near the left termini of their respective plasmids (28). This gene and adjoining genes, such as *fhbA* and the clusters of “P35”-like sequences, were not included in the sequence of “lp124” of *B. recurrentis*. In one study, a portion of the left end of the large plasmid, including loss of the *fhbA* gene, was spontaneously lost in a few strains of *B. hermsii* and in one strain of *B. turicatae* after 100 serial passes in vitro, yet remained viable in culture (53). It is possible that this phenomenon occurred in the *B. recurrentis* strain used for sequencing, however, pulsed-field gel electrophoresis showed that the large plasmid of *B. recurrentis* was about the same size as the *B. duttonii* plasmid at 160-170 kb (24, 28). Therefore, the *B. recurrentis* publically available linear plasmid is most likely missing 36 – 46kb of sequence.

A prominent feature of the large linear plasmids’ sequences was the ~60% proportion of repetitive DNA, much of which was collinear or locationally similar between replicons. Some of the repeated ORFs, such as the “P35”-like coding sequences, were homologous to sequences in LD *Borrelia* species, while others, like the repeats in Region D, appeared to be unique to RF species. It is possible that amplification and diversification of these proteins could have occurred as the spirochetes adapted to different hosts, arthropod or vertebrate, and their associated immune or environmental selective pressures. Gene duplications in other pathogens can result in an increase of functional genes used, for example, to adapt to or exploit novel environmental niches or function in a gene dosage effect (64, 65).

Our inference about a common ancestry for the large linear plasmids of RF *Borrelia* species was anticipated by prior findings of the *nrdEFI* genes on large plasmids across species (49). A less expected finding was the extent of shared heritage for the large linear plasmids of...
RF Borrelia species and in particular the lp54-type linear plasmids of LD species. Lescot et al observed syntenic regions between B. duttonii lp165 and B. burgdorferi lp54, but orthologous genes between the two plasmids were not described (24). Totaling ~22 kb, two regions containing lp54 orthologs were found in similar locations on each of the RF large plasmids and the more conserved interior portion of these plasmids. All LD Borrelia species and strains whose genome sequences are public have had an lp54-like plasmid of similar size.

Recognizing the phylogenetic evidence we present here regarding the PFam32 genes and the 3 other genes of the plasmid maintenance locus, which are borne by the large plasmids and those of lp54-type plasmids, we propose assigning the RF large plasmids to the type A compatibility group for Borrelia plasmids (66, 67), and, accordingly, use “a” as the third letter after “bh”, “bt”, etc., for coding sequence designations.

The lp54 plasmids bear genes whose expression is regulated during the infectious cycle and that in their respective presence or absence distinguish LD species from RF species. These include OspA (bba15) and OspB (bba16) (68, 69), and DbpA (bba24) and DbpB (bba25) (70). GC skew analysis indicates that the ospAB and dbpAB operons were acquired horizontally by the LD lineage after the putative split from the lineage leading to RF species.

Can we identify lp174 genes that distinguish RF species from LD species? Three examples are fhbA, bipA, and alp (26, 29, 30). Two other lp174 genes that first appeared to be unique to RF species on the basis of standard blast search criteria were bha064 and bha065. But a search for structural similarities revealed what may be homology of these proteins to BBA68, the CspA (formerly CRASP-1) protein and a PFam54 family member. A different “CRASP-1” protein was named in B. hermsii (71), but this protein is the same as FhbA, encoded by bha008.

Surrounding bha064 and bha065 are other ORFs that are more divergent in sequence from this
pair, but possibly are paralogs and, as such, additional candidates for virulence determinants.

To the right and downstream of bha064 and bha065 is the thyX gene, bha076 (Figure 2C). A thyX pseudogene occupies a similar position to the PFam54 genes on lp54 (49). If one also includes these in the list of shared genes, then an even larger proportion of lp174 has common ancestry with the lp54-type plasmids.

While the functions of BHA064 and BHA065 in B. hermsii and orthologous ORFs in other species remain to be determined, the higher expression of bha064 under in vitro conditions that mimic the environment in the arthropod suggests to us a vector-associated role in the RF life cycle. If BHA064, as well as its orthologs in four other Borrelia species, are homologous as well as structurally similar to CspA of LD species, how does the experimental expression data for bha064 compare with a proposed function of CspA, i.e. providing serum resistance (72)?

Somewhat at odds with the serum resistance claim is the report that CspA expression is increased in the tick during feeding but rapidly decreases once the spirochetes are transmitted to the mammal (73). Whatever activity BHA064 and the other ORFs turn out to actually have, the present study provides a sharper resolution view of regions of largely unknown function on these unique large linear plasmids and provides for insights about their evolution.

**Acknowledgements.** The work was supported by NIH grant AI24424 and the intramural program of the National Institute of Allergy and Infectious Diseases. We thank Sherwood Casjens for advice and Fong Hue and Renee Marcsisin for technical assistance.
Literature Cited


**FIGURE LEGENDS**

**Figure 1.** Pulsed field gel of total DNA of *B. hermsii*, *B. parkeri* and *B. burgdorferi*. Cells were embedded in agarose plugs and lysed *in situ* with proteinase K. Total DNA was separated by pulsed-field gel electrophoresis (0.8% agarose) with parameters for separating 50 – 1,000 kb DNA molecules per manufacturer’s instructions: 6V cm⁻¹, included angle of 120°, initial switch time of 50 s ramped to a final switch time of 90 s for 21 h at 14°C. The large plasmids of *B. hermsii* and *B. parkeri* are indicated by arrows and the 54 kb linear plasmid of *B. burgdorferi* by an arrowhead. Asterisks indicate the linear chromosomes of all species. The sizes (in kb) of selected *Saccharomyces cerevisiae* chromosomes and lambda DNA concatemers are shown on the left and right, respectively.

**Figure 2.** Organization of repetitive DNA, collinear sequence, and individual open reading frames of relapsing fever *Borrelia* species.

*A.* Dot matrix plot of *B. hermsii* lp174 versus itself. 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. Regions of repetitive DNA are labeled A-E above the plot.

*B.* Schematic picture of alignment of the large linear plasmid sequences of *B. hermsii*, *B. turicatae*, and *B. duttonii*. Regions of contiguously homologous sequence are represented by colored local collinear blocks (LCBs I – III). A LCB is defined as a conserved segment of sequence that appears to be internally free from genome rearrangements relative to the reference sequence. Grey lines bound the orthologous LCBs. Similarity profiles (vertical lines within the LCBs) whose heights correspond to average level of sequence similarity are shown in each LCB. LCBs below the centerline represent sequence in the opposite orientation relative to
the reference sequence, *B. hermsii* lp174. Nucleotide positions (in kb) are indicated above the alignment.

**C.** Physical map of *B. hermsii* lp174 (Accession No. HM008709) oriented left to right. Genes encoded on the positive strand are indicated by boxes above the centerline and those on the negative strand are below the centerline. Beige shaded boxes (A-E) correspond to regions of repetitive DNA shown in 2A. Grey shaded boxes (I-III) correspond to LCBs shown in 2B.

Selected genes previously mapped to the large plasmids are underlined: *fhbA* (*bha008*), *nrdF* (*bha075*), *nrdE* (*bha076*), *nrdI* (*bha077*), *thyX* (*bha078*), *bipA* (*bha123*), and *alp* (*bha0128*).

Newly identified genes discussed in the text are also indicated: *bha064*, *bha065*, and the plasmid replication and partition loci (*bha132-135*). Colored ORFs represent homologs to the LD species P35 antigen (region A) or to PFam99 (orange), *B. burgdorferi* lp54 (red), *B. burgdorferi* cp26 (yellow), ORFs unique and conserved to RF species (green), or ORFs unique to a single species (black). Locus tags or gene names are indicated above the corresponding ORF.

**Figure 3.** GC skew analysis of *Borrelia* plasmid sequences.

(A) GC skew diagrams of *B. hermsii* lp174 and *B. turicatae* partial lp150 sequences. Shaded grey boxes indicate the plasmid replication and partitioning loci. Red lines at the top indicate the window and the point at which the skew shifts. Total GC skew was calculated with a sliding window size of 500 bp and a step size of 50 bp.

(B) Circular G-C skew representation of orthologous regions from *B. hermsii* lp174 and *B. burgdorferi* lp54. Deviation from the average GC skew of the entire sequence is represented by green lines (above average) or purple lines (below average). Radiating lines originating from the
center circle indicate ORF boundaries. Orthologous ORFs are noted by the same font colors (red, cyan or blue) with thick diagonal lines enclosing regions of non-homologous sequence (ORFs indicated in black). GC skew was calculated with a 500 bp window and step size of 1 bp (75).

**Figure 4.** Phylogram of codon-aligned nucleotide sequences for selected PFam32 orthologs from different *Borrelia* replicons. Genes from the RF large plasmids (squares) and LD lp54 plasmids (circles) are indicated. The replicons and accession numbers used for each species and strain are provided in the methods section. Nodes with bootstrap values > 70% support by neighbor-joining distance criteria from 10,000 replicates are indicated. Scale bar represents nucleotide substitutions per site.

**Figure 5.** Gene expression analysis of *bha064*. Quantitative real-time PCR analysis of *bha064* cDNA copies normalized to copies of the constitutively expressed *flaB* for *B. hermsii* grown at 37°C (grey bars) in the blood of immunodeficient mice or in culture, or grown at 23°C (black bars) in culture. Ninety-five percent confidence intervals are shown for each condition. Asterisks above lines indicate significant differences as measured by t-test (*p* value < 0.001).
Table 1. Expected and observed sizes in bp of hybridizing restriction fragments

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>bha008 probe (left end)</th>
<th>bha167 probe (right end)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Expected</td>
<td>Observed</td>
</tr>
<tr>
<td>Bam HI</td>
<td>7,993</td>
<td>8,850</td>
</tr>
<tr>
<td>Eco RI</td>
<td>3,213</td>
<td>3,208</td>
</tr>
<tr>
<td>Eco RV</td>
<td>6,140</td>
<td>6,538</td>
</tr>
<tr>
<td>Kpn I</td>
<td>11,237</td>
<td>11,037</td>
</tr>
<tr>
<td>Pst I</td>
<td>10,698</td>
<td>10,218</td>
</tr>
<tr>
<td>Xba I</td>
<td>4,284</td>
<td>4,360</td>
</tr>
<tr>
<td>Total</td>
<td>43,565</td>
<td>44,211</td>
</tr>
</tbody>
</table>

a na, not applicable
B. hermsii lp174
B. turicatae lp150
B. duttonii lp165
B. burgdorferi lp54
B. garinii lp54
B. afzelii lp54
B. burgdorferi cp32-3
B. garinii cp32-like
B. afzelii cp32-3
B. hermsii cp32-like
B. duttonii lp26
B. duttonii lp35
B. burgdorferi cp26
B. garinii cp26
B. afzelii cp26
B. duttonii lp23
B. hermsii lp53
B. hermsii lp28-2
B. garinii /lp28-like
B. burgdorferi lp28-1
B. lonestari /lp28-like
B. speilmanii chromosome
B. burgdorferi chromosome
B. hermsii chromosome
B. turicatae chromosome
B. duttonii chromosome
Mean number of bha064 copies per 1000 flaB copies

Growth conditions:
- Mouse blood
- Medium pH 7.6
- Medium pH 8.0

37 37 37 23 23