AMERICAN SOCIETY FOR MICROBIOLOGY | mBio®

RESEARCH ARTICLE

Epigenomic Landscape of Lyme Disease Spirochetes Reveals Novel Motifs

Jenny Wachter,[a] Craig Martens,[b] Kent Barbian,[b] Ryan O. M. Rego,[c,d] Patricia Rosa[a]

[a]Laboratory of Bacteriology, Rocky Mountain Laboratories, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Hamilton, Montana, USA
[b]Research Technologies Branch, Rocky Mountain Laboratories, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Hamilton, Montana, USA
[c]Institute of Parasitology, Biology Centre, Czech Academy of Sciences, Ceske Budejovice, Czech Republic
[d]Faculty of Science, University of South Bohemia, Ceske Budejovice, Czech Republic

ABSTRACT *Borrelia burgdorferi*, the etiological agent of Lyme disease, persists in nature through an enzootic cycle consisting of a vertebrate host and an *Ixodes* tick vector. The sequence motifs modified by two well-characterized restriction/modification (R/M) loci of *B. burgdorferi* type strain B31 were recently described, but the methylation profiles of other Lyme disease *Borrelia* bacteria have not been characterized. Here, the methylomes of *B. burgdorferi* type strain B31 and 7 clonal derivatives, along with *B. burgdorferi* N40, *B. burgdorferi* 297, *B. burgdorferi* CA-11, *B. afzelii* PKo, *B. afzelii* BO23, and *B. garinii* PBr, were defined through PacBio single-molecule real-time (SMRT) sequencing. This analysis revealed 9 novel sequence motifs methylated by the plasmid-encoded restriction/modification enzymes of these *Borrelia* strains. Furthermore, while a previous analysis of *B. burgdorferi* B31 revealed an epigenetic impact of methylation on the global transcriptome, the current data contradict those findings; our analyses of wild-type *B. burgdorferi* B31 revealed no consistent differences in gene expression among isogenic derivatives lacking one or more restriction/modification enzymes.

IMPORTANCE The principal causative agent of Lyme disease in humans in the United States is *Borrelia burgdorferi*, while *B. burgdorferi*, *B. afzelii*, and *B. garinii*, collectively members of the *Borrelia burgdorferi sensu lato* species complex, cause Lyme disease in Europe and Asia. Two plasmid-encoded restriction/modification systems have been shown to limit the genetic transformation of *B. burgdorferi* type strain B31 with foreign DNA, but little is known about the restriction/modification systems of other Lyme disease *Borrelia* bacteria. This paper describes the methylation motifs present on genomic DNAs of multiple *B. burgdorferi*, *B. afzelii*, and *B. garinii* strains. Contrary to a previous report, we did not find evidence for an epigenetic impact on gene expression by methylation. Knowledge of the motifs recognized and methylated by the restriction/modification enzymes of Lyme disease *Borrelia* will facilitate molecular genetic investigations of these important human pathogens. Additionally, the similar motifs methylated by orthologous restriction/modification systems of Lyme disease *Borrelia* bacteria and the presence of these motifs within recombinogenic loci suggest a biological role for these ubiquitous restriction/modification systems in horizontal gene transfer.

KEYWORDS *Borrelia*, Lyme disease, methylation, motifs, epigenetic regulation

The spirochetal bacteria that cause Lyme disease exist in nature through an enzootic cycle consisting of a vertebrate host and an *Ixodes* tick vector (1–3). In the United States, *Borrelia burgdorferi sensu stricto* is the principal causative agent of Lyme disease in humans, while *B. burgdorferi*, *B. afzelii*, and *B. garinii*, members of the *Borrelia burgdorferi sensu lato* species complex, are known to cause human Lyme disease in Europe

Citation Wachter J, Martens C, Barbian K, Rego ROM, Rosa P. 2021. Epigenomic landscape of Lyme disease spirochetes reveals novel motifs. mBio 12:e01288-21. https://doi.org/10.1128/mBio.01288-21.

Editor Steven J. Norris, McGovern Medical School

Address correspondence to Jenny Wachter, jenny.wachter@nih.gov.

This article is a direct contribution from Patricia A. Rosa, a Fellow of the American Academy of Microbiology, who arranged for and secured reviews by Sherwood Casjens, University of Utah, and Weigang Qiu, Hunter College of City University of New York.

Received 3 May 2021
Accepted 7 May 2021
Published 22 June 2021

and Asia (4, 5). The segmented genome of *B. burgdorferi sensu lato* is unstable during *in vitro* cultivation, which can lead to the loss of native linear plasmids (lp) and circular plasmids (cp). Some of the plasmids that are readily lost during culture of *B. burgdorferi* type strain B31, such as lp25 and lp28-1, are required for infectivity in a mammalian host (6–8). However, the loss of other plasmids, such as lp56, has no impact on the experimental mouse-tick infectious cycle and can aid in the genetic manipulation of the spirochete (9, 10). The enhanced transformation of spirochetes lacking lp25 and lp56 led to the discovery of the type II restriction/modification (R/M) systems encoded by the *bbe02* and *bbq67* loci on these plasmids, respectively (9–13). An additional R/M gene, *bbh09* on lp28-3, is a homolog of *bbe02* but does not appear to affect transformation efficiency (9, 13).

DNA modifications, including methylation, have diverse functions, such as impeding or aiding in the cleavage of DNA by restriction endonucleases, mismatch repair, and/or regulation of gene expression (14–16). In bacteria and archaea, methyltransferase enzymes (MTases) are responsible for methylating DNA in the form of either 6-methyladenosine (m6A), 4-methylcytosine (m4C), or 5-methylcytosine (m5C), with m6A being the most common form of methylation in prokaryotes (17, 18). Within the majority of defined bacterial methylomes, the genomic R/M motifs are almost completely methylated in the presence of the responsible R/M genes (19, 20). In fact, R/M systems are responsible for the most abundant, widely distributed DNA methylation among prokaryotes (21). Despite the high prevalence of MTases within the >200,000 bacterial and archaeal complete or draft genomes sequenced, the DNA motifs that they modify and their biological roles remain largely undefined (22–25). However, sequencing technologies capable of detecting all three major forms of bacterial DNA methylation, such as single-molecule real-time (SMRT) sequencing, have been used to generate the majority of the >2,000 mapped bacterial and archaeal methylomes (19, 23, 26–32).

Strong evidence exists for horizontal gene transfer between different strains of *B. burgdorferi sensu lato* in nature (33–38), but it is unknown if endogenous R/M genes facilitate or impede such events, particularly if heterologous R/M systems recognize and methylate different motifs. The nucleotide sequence motifs modified by the m6A activity of BBE02 and BBQ67 in *B. burgdorferi* B31 have been determined by SMRT sequencing (39), but the R/M motifs of other *B. burgdorferi sensu lato* strains remain undefined. Additionally, while strain B31 derivatives lacking *bbe02* and *bbq67* loci were reported to exhibit striking differences in their transcriptomic profiles relative to the wild type (wt), these comparisons were made between nonisogenic variants that differed in total plasmid contents in addition to R/M loci (39). Here, we describe the methylomes of multiple *B. burgdorferi sensu stricto*, *B. garinii*, and *B. afzelii* strains and define the novel sequence motifs modified by their respective plasmid-encoded R/M systems. Additionally, engineered deletions of R/M loci in wt B31 permitted direct comparisons of the transcriptomes of isogenic variants lacking only *bbe02*, only *bbq67*, or both loci relative to the wt clone. Contrary to a previous report of an epigenetic impact of methylation on gene expression in *B. burgdorferi* strain B31 (39), our current RNA sequencing (RNA-seq) and quantitative real-time PCR (qRT-PCR) analyses of the B31 transcriptome contradict those findings. Importantly, the identification of the motifs recognized and methylated by the R/M enzymes of Lyme disease *Borrelia* bacteria will facilitate the genetic manipulation of these important human pathogens and provide data to define a biological role for these ubiquitous R/M systems.

RESULTS

Homology and REBASE searches for *Borrelia* restriction/modification genes. Searches for BBE02/BBH09 and BBQ67 homologs were performed with whole-genome BLAST protein and nucleotide searches of *Borrelia burgdorferi sensu lato* (see Table S1 in the supplemental material [R/M homologs]) (13). These searches identified 3 R/M genes in *B. burgdorferi* N40, 3 in *B. burgdorferi* CA-11.2A, 4 in *B. burgdorferi* 297, 4 in *B. afzelii* PKo, 2 in *B. afzelii* BO23, and 2 in *B. garinii* PBr. REBASE searches were also performed to identify any additional R/M

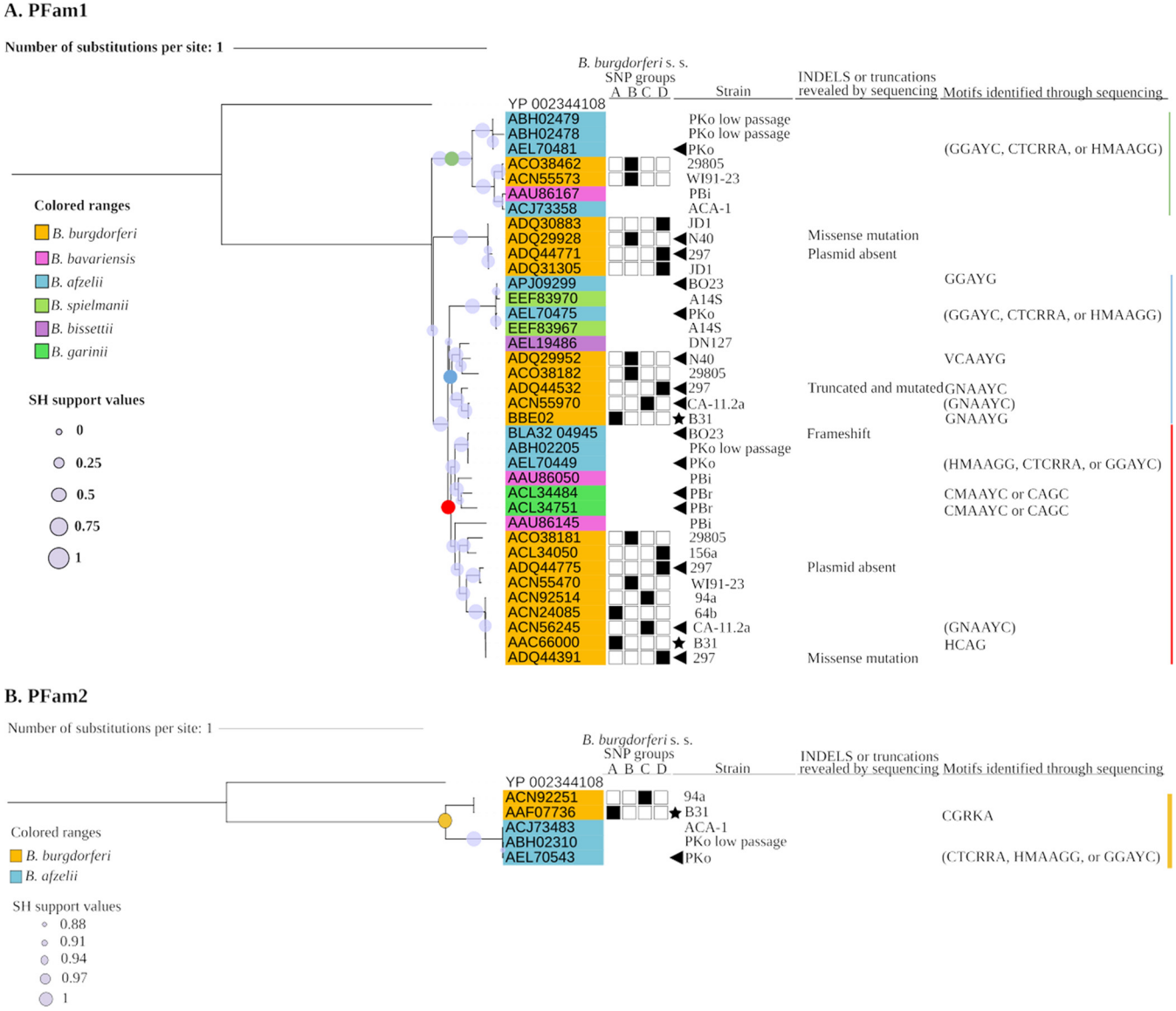

FIG 1 Phylogenetic trees depicting homologs of the BBE02 (Pfam1) (A) and BBQ67 (Pfam2) (B) type II R/M systems of *Borrelia burgdorferi* B31. The trees were rooted by the outgroup (GenBank accession number YP_002344108). The CAT approximations of the Shimodaira-Hasegawa (SH) test were used to determine the local support values (SH support values) with 1,000 resamplings. The *B. burgdorferi sensu stricto* (s. s.) single nucleotide polymorphism (SNP) groups are indicated by the boxes next to the R/M protein GeneIDs. Genes belonging to *B. burgdorferi* B31 are indicated by stars, while additional *Borrelia* strains that underwent SMRT sequencing are indicated with triangles. Any truncations identified in GenBank and any indels identified through sequencing are indicated next to the strain name. The identified motifs are indicated in the rightmost column next to the R/M enzyme responsible for its methylation; however, if it is not known which protein is responsible for motif recognition, all motifs identified within that strain are shown in parentheses. The colored circles and similarly colored lines indicate orthologs whose clade matches the expected strain phylogeny. The sequences used to create the tree are located in Table S1 in the supplemental material (R/M homologs).

genes; however, these generally revealed R/M enzymes that due to either size or frameshifts would not produce functional proteins (Table S1, REBASE). The phylogenetic trees were assembled from amino acid sequences identified through BLAST searches, with a greater number of BBE02/BBH09 homologs than BBQ67 homologs being identified (Fig. 1), in agreement with previous reports (13).

Single-molecule real-time sequencing of *Borrelia* reveals methylation of unique nonpalindromic m6A motifs by multiple strains. SMRT sequencing was performed to determine the motifs recognized by the R/M enzymes of 8 different *B. burgdorferi* B31 derivatives (13, 40–43), 3 different *B. burgdorferi* N40 derivatives (44), *B. burgdorferi* 297 (2), *B. burgdorferi* CA-11 (45) and clonal CA-11.2A (46), *B. garinii* PBr (47), and *B. afzelii* PKo (47) and BO23 (48), resulting in 75- to 776-fold coverage and >99.99% overall concordance with

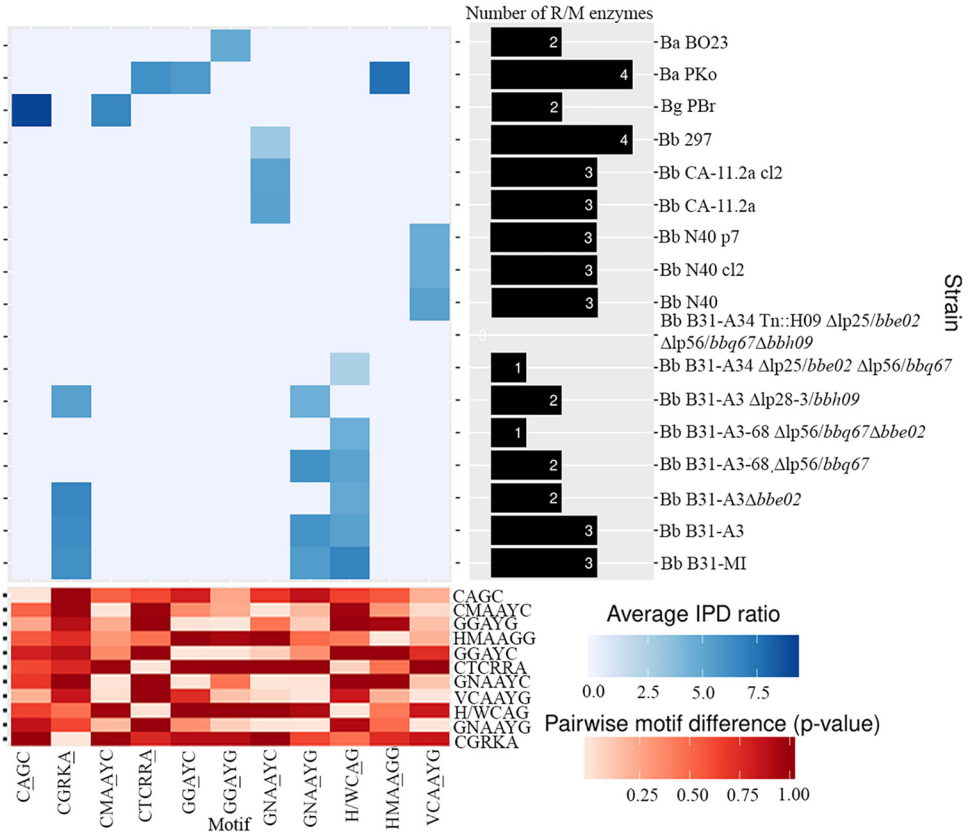

FIG 2 Heat map depicting the methylated motifs per genome and their interpulse duration (IPD) ratios. The horizontal heat map (red) indicates the detected motifs and the pairwise motif differences (based on the Pearson correlation coefficient), with the m6A-methylated bases underlined. The vertical bar plot demonstrates the number of R/M enzymes encoded within each *Borrelia* strain and derivative, identified to the right of the heat map. Ambiguity codes for nucleotides are as follows: H is A, C, or T; K is G or T; M is A or C; N is any base; R is A or G; V is A, C, or G; and Y is C or T. Ba, *B. afzelii*; Bg, *B. garinii*; Bb, *B. burgdorferi*.

the published genome sequences (Table S2, coverage stats). Due to the absence of complete chromosomal sequences, contigs were utilized for chromosomal alignments of *B. burgdorferi* CA-11 and *B. afzelii* BO23. Additionally, as *B. burgdorferi* 297 does not have a sequenced chromosome, a previously sequenced and assembled chromosomal sequence of a close relative, JD1, was used for alignment. Analysis of SMRT sequencing data revealed that, for the most part, genomic DNAs (gDNAs) from different *Borrelia* strains exhibited distinct methylation motifs, although they carry homologous R/M genes (Fig. 2). The variation of motifs within the same species is consistent with the rapid evolution of the R/M system in *Borrelia*. While there is a significant strain-specific distribution of motifs, through analysis of the pairwise motif differences and the phylogenetic distributions, orthologous R/M enzymes methylate motifs that are more similar than those of paralogous R/M enzymes (Fig. 1 and 2). All motifs were of the m6A type, with no confident m4C or m5C motifs being detected. Although the *Bbun40_y07* gene product of *B. burgdorferi* N40 is reported to methylate cytosine (49), cytosine methylation was not detected in either nonclonal or clonal N40 gDNA (Table 1). The number of unique methylated motifs detected for each *B. burgdorferi* B31 and N40, *B. garinii* PKo, and *B. afzelii* PBr strain was the same as the number of m6A R/M enzymes either known or predicted to be encoded in their genomes (Fig. 2). In contrast, *B. burgdorferi* N40, CA-11, and 297 and *B. afzelii* BO23 are predicted to harbor more R/M genes than the number of unique methylated motifs that were identified. While CA-11 harbored intact R/M genes, *Bbun40_e01* in N40 contains indels; *B. burgdorferi* 297 lacks plasmid lp28-5 carrying *Bbu297_y09* and *Bbu297_y05*, along with indels in *Bbu297_h03* and *Bbu297_j01*; and *B. afzelii* BO23 contains a frameshift in gene *BLA32_04945*. However, it is unknown if the indels in *BbuN40_e01*, *Bbu297_h03*, and

TABLE 1 Motifs and their prevalence within *Borrelia* genomes[a]

Strain or derivative	Motif	Type of methylation	No. of genomes with motif/total no. of genomes	% modified	Mean coverage[f]	Mean QV[b]
B. burgdorferi B31-MI[c]	CGRKA	m6A	2,058/2,237	92.00	132.77	130.20
	GNAAYG	m6A	2,562/2,980	85.97	122.18	106.97
	HCAG	m6A	10,311/16,677	61.83	148.18	104.20
B. burgdorferi B31-A3[d]	CGRKA	m6A	2,192/2,236	98.03	119.96	121.09
	GNAAYG	m6A	2,896/2,979	97.21	108.50	105.62
	HCAG	m6A	7,358/16,633	44.23	145.10	70.37
B. burgdorferi B31-A3-68 Δlp56/*bbq67*[d]	GNAAYG	m6A	2,885/2,907	99.24	119.72	118.81
	HCAG	m6A	8,123/16,113	50.41	147.75	70.06
B. burgdorferi B31-A3 Δ*bbe02*[d]	CGRKA	m6A	2,169/2,236	96.79	110.24	115.76
	WCAG	m6A	4,427/12,658	34.97	145.06	57.14
B. burgdorferi B31-A3-68 Δlp56/*bbq67* Δ*bbe02*[d]	H/WCAG	m6A	5,243/16,127	32.51	162.08	62.15
B. burgdorferi B31-A3 Δlp28-3/*bbh09*[d]	CGRKA	m6A	2,195/2,198	99.86	229.90	282.86
	GNAAYG	m6A	2,940/2,944	99.86	210.64	225.92
B. burgdorferi B31-A34 Tn::*bbh09* Δlp25/*bbe02* Δlp56/*bbq67* Δ*bbh09*[d]						
B. burgdorferi N40[c]	VCAAYG	m6A	2,182/2,277	95.83	137.73	134.17
B. burgdorferi N40 clone 2[d,e]	VCAAYG	m6A	2,165/2,227	97.22	234.64	257.67
B. burgdorferi N40 p7[d]	VCAAYG	m6A	2,180/2,277	95.74	218.29	238.61
B. burgdorferi CA11[c]	GNAAYC	m6A	2,499/2,535	98.58	234.85	306.45
B. burgdorferi CA-11.2A clone 2[d]	GNAAYC	m6A	2,504/2,535	98.78	233.08	303.94
B. burgdorferi 297[e]	GNAAYC	m6A	2,222/2,581	86.09	199.87	101.11
B. afzelii PKo[c]	CTCRRA	m6A	1,546/1,700	90.94	124.88	113.51
	GGAYC	m6A	1,347/1,363	98.83	118.81	159.64
	HMAAGG	m6A	3,583/5,485	65.32	133.64	95.51
B. afzelii BO23[d]	GGAYG	m6A	1,146/1,150	99.65	250.58	293.71
B. garinii PBr[c]	CMAAYC	m6A	1,402/2,169	64.64	124.61	115.72
	CAGC	m6A	3,346/4,111	79.14	119.94	116.96

[a]The methylated adenine is underlined.
[b]QV, quality value, a prediction of the error probability of a base call. Only base calls with a QV of 20 (99% accuracy) were used for analysis.
[c]Genomic DNA sequenced was from a nonclonal heterogeneous population.
[d]Genomic DNA sequenced was from a clonal population.
[e]*B. burgdorferi* N40 clone 2 is missing lp38, and *B. burgdorferi* 297 is missing lp28-1, lp28-4, and lp28-5.
[f]The mean of the total number of bases mapped to the methylated base; this includes reads from the same library molecules.

Bbu297_j01 would inactivate the resulting enzymes. All other *Borrelia* strains and derivatives did not contain any indels within their respective R/M genes (Table S2, R/M sequencing). Apart from the H/WCAG motifs (the methylated adenine is underlined), clonal derivatives, nonclonal *B. burgdorferi* N40 and CA-11, and *B. afzelii* PKo had efficient methylation (>95%) of m6A motifs encoded throughout the genome, while nonclonal *B. burgdorferi* B31 MI and 297 and *B. garinii* PBr did not methylate the majority of sites containing their respective motifs. The incomplete methylation of gDNA extracted from nonclonal populations is presumably due to the loss of plasmids encoding the R/M enzymes by a subset of the population (Table 1 and Fig. 2) (50). The two highly methylated motifs in *B. burgdorferi* B31, CGRKA modified by BBQ67 and GNAAYG modified by BBE02, are consistent with those identified

by Casselli et al. (39). However, the H/WCAG motif of *B. burgdorferi* B31 has not been previously described (22, 51). The methylated motifs identified within *B. burgdorferi* N40, CA-11, and 297; *B. garinii* PBr; and *B. afzelii* PKo and BO23 are novel (22, 51). Further REBASE searches identified additional motifs that contain some form of the *Borrelia* motifs identified within this study, except for GNAAYG, GNAAYC, and HMAAG (Table S2, REBASE motifs).

The number of motifs present in *Borrelia* DNA is not more than would be expected by chance. To determine if individual motifs confer some type of advantage and are overrepresented in strains carrying the cognate R/M enzyme, the numbers of motifs within each genome were analyzed. This analysis assessed the numbers of motifs present in each replicon versus randomly assembled, noncoding sequences of the same length to determine if there were significantly more motifs than expected by chance in any particular replicon (Table S2, motif analysis). A false-positive "replicon" occurs if a randomly assembled sequence contains more motifs than would be expected by chance. Significance is determined by scanning primary and control sequences for each motif and computing an odds score for each position in each sequence; these scores are combined through the average odds score, and the sum test is applied to determine if the primary sequences have significantly higher scores (52). Of all the genomes analyzed, there was no evidence of selective pressure on the prevalence of R/M motifs (Table S2, motif analysis). Additionally, all genes harbored within each sequenced genome were synonymously shuffled at the third position of each codon to produce 1,000 synonymously shuffled sequences (53). Comparisons revealed that the majority of genes encoded fewer motifs than the synonymously shuffled sequences (Fig. 3).

The number of R/M motifs in the *ospC* gene is higher than would be expected by chance. Although there is no evidence for selective pressure on the total number of R/M motifs per replicon, we analyzed the distribution of R/M motifs in plasmid-borne genes of *B. burgdorferi* B31 for which there is strong evidence of horizontal gene transfer and recombination (Fig. 4 and Table S2, genes of interest) (33–38, 54–57). For most of the R/M motifs in the 7 strains analyzed, *B. burgdorferi* B31 *ospC* contained a significantly higher number than would be expected by chance. Additionally, analysis of the *ospC* genes of the other 6 strains revealed a significantly higher occurrence of some motifs (Table S2, *ospC*). In contrast, *B. burgdorferi* B31 *ospD*, *dbpA*, *dbpB*, and the *erp* genes revealed higher numbers of some R/M motifs than would be expected by chance, but there was also a high number (≥25%) of false-positive sequences. However, there is no appreciable difference (as determined by the log fold change [logFC]) between the numbers of motifs encoded within these genes and their synonymously shuffled counterparts (Table S2, CodonShuffle). Together, these data indicate the clustering of R/M motifs in *ospC* but not in other plasmid-borne genes for which there is also evidence of horizontal gene transfer, indicating an *ospC*-specific selective pressure.

The number of motifs and the prevalence of methylated motifs are independent of the replicon and strain, with some striking exceptions. To determine if motif prevalence varies among replicons (Fig. S2 and S3), the numbers of encoded m6A sites were compared to the length of the replicon and found to correlate (Fig. 5). Therefore, the poor correlation between the number of methylated motifs and the total number of motif sites on a replicon can be explained by either variable methylase activity or the absence of the responsible R/M locus. Analysis of methylated motifs revealed that both the CGRKA and GNAAYG motifs of *B. burgdorferi* B31 and the GGAYG motif of *B. afzelii* BO23 significantly correlate with the number of motif sites present on the replicons, regardless of the derivative being analyzed, as would be expected with methylation of ~90% of encoded motifs (Table 1 and Fig. S4A). Similarly, methylation of CTCRRA and GGAYC of *B. afzelii* PKo, CAGC of *B. garinii* PBr, VCAAYG of *B. burgdorferi* N40, and GNAAYC of *B. burgdorferi* CA-11 and 297 significantly correlated with the number of motifs present on the replicons and methylated about 80% of their encoded motifs despite the fact that these were sequenced as nonclonal cultures (Table 1 and Fig. S4B and C). In striking contrast, however, the numbers of methylated H/WCAG motifs in *B. burgdorferi* B31, HMAAGG motifs in *B. afzelii* PKo, and CMAAYC motifs in *B. garinii* PBr

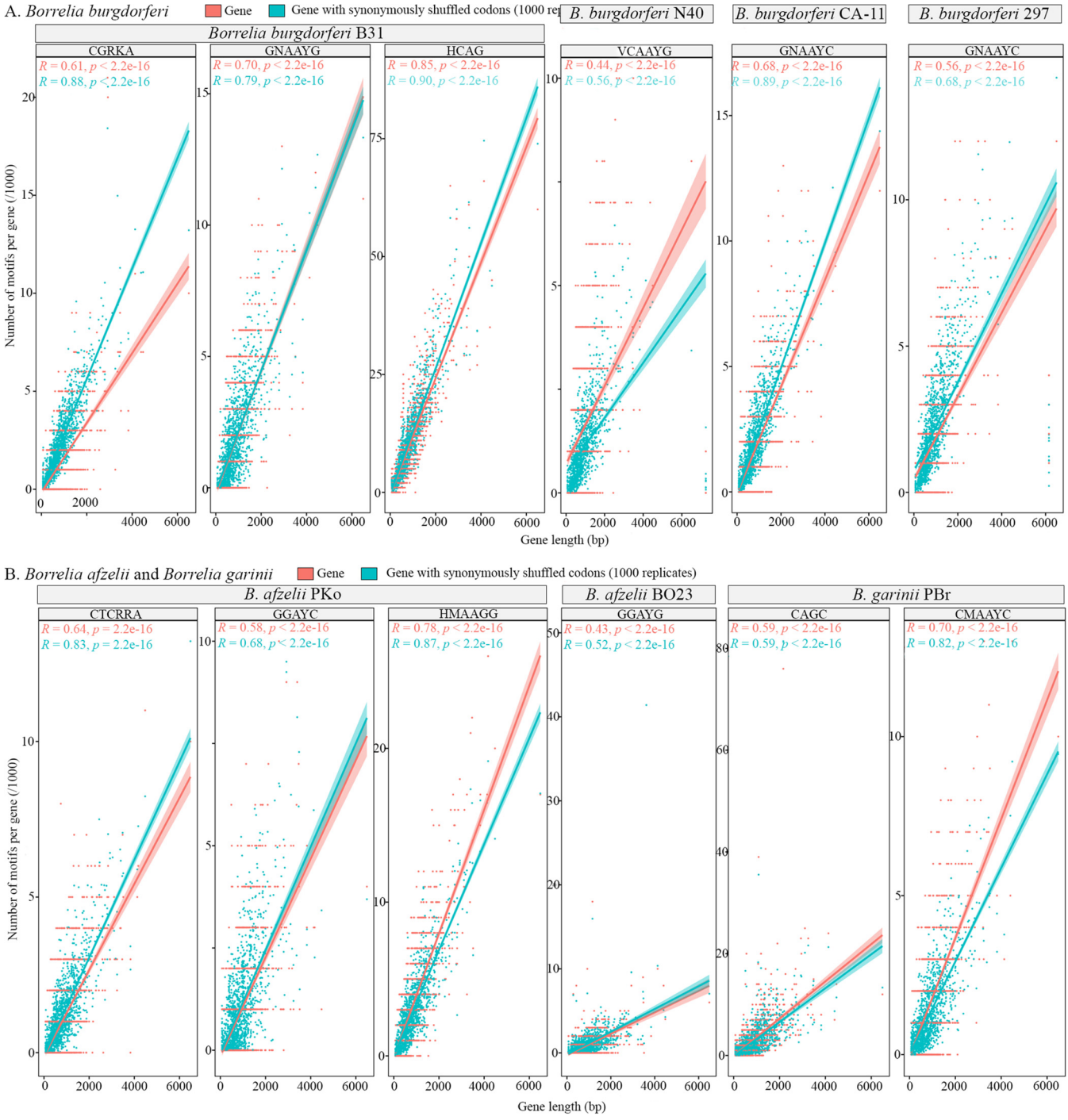

FIG 3 Pearson correlation of the number of motifs per gene (per 1,000 bp) and the average number of motifs per 1,000 synonymously shuffled genes. An *R* value close to 1 with a significant *P* value ($P < 0.01$) indicates a correlation between the number of motifs and gene length. The regression line and confidence interval along with the *R* and *P* values for each motif of *B. burgdorferi* derivatives (A) and *B. afzelii* derivatives and *B. garinii* (B) are shown. Other than the HMAAG motif of *B. afzelii* PKo and the CAGC and CMAAYC motifs of *B. garinii* PBr, the majority of genes contain fewer motifs than their synonymously shuffled counterparts.

do not correlate with the numbers of the respective motif sites present on the replicon and were found to methylate <65% of their respective sites (Table 1 and Fig. S4). The partial methylation of gDNA by BBH09 in *B. burgdorferi* B31 likely reflects the variable activity of the enzyme in clonal B31 derivatives (A3, A3 Δ*bbe02*, A3-68, A3-68 Δ*bbe02*, and A34). However, we cannot say if the incomplete methylation of gDNA from non-clonal cultures of *B. afzelii* PKo and *B. garinii* PBr reflects the activities of the enzymes

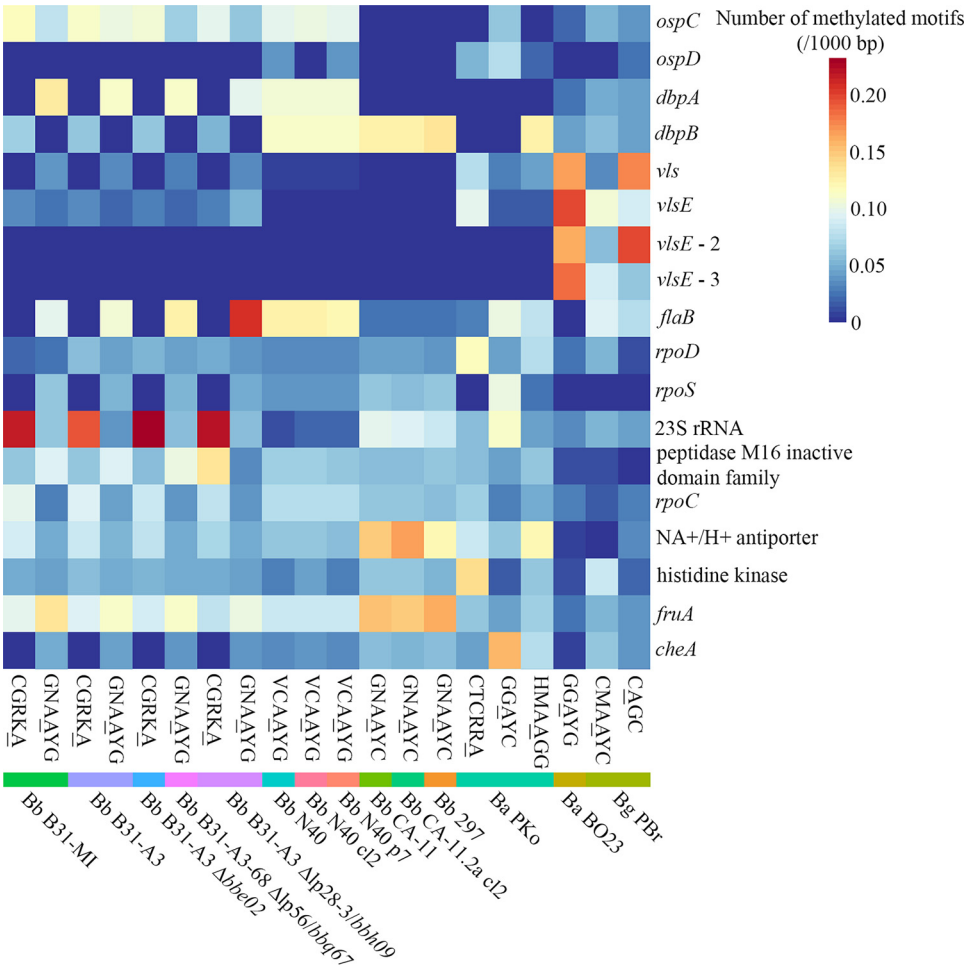

FIG 4 Heat map of methylated motifs per 1,000 bp within select genes of sequenced Lyme disease *Borrelia* spirochetes. The numbers of methylated motifs within each gene are displayed as the ratio of methylated motifs to the gene length per 1,000 bp so that each column (strain) totals 1. Each strain and derivative is displayed as a different color below the motif.

or the loss of plasmids harboring the respective R/M loci by some spirochetes in a heterologous population (Fig. 5 and Fig. S4).

***bbe02* and *bbq67* transcript levels are higher in culture than *in vivo*.** While the R/M enzymes of *B. burgdorferi* B31 are active *in vitro*, with BBE02 and BBQ67 methylating >95% of all genomic sites in clonal populations (Table 1), the transcription of these genes *in vivo* is unknown. Therefore, we measured *bbe02* and *bbq67* transcript levels in *B. burgdorferi* B31-A3 spirochetes grown *in vitro* and throughout the *in vivo* mouse-tick infectious cycle using qRT-PCR (Fig. 6). Transcripts for both genes were present at low levels *in vitro* relative to the constitutive *flaB* gene, and neither was detectable by qRT-PCR in total RNA isolated from infected mouse tissues. Analysis of *B. burgdorferi* B31-A3-infected ticks revealed low levels of *bbe02* and *bbq67* transcripts in larval ticks and unfed nymphs, with a transient increase in the levels of both transcripts in nymphs after feeding to repletion. Thus, *bbe02* and *bbq67* are expressed at very low levels throughout the mouse-tick infectious cycle, with a relative increase in transcript levels immediately following the nymphal blood meal.

Methylation by BBE02 and BBQ67 does not control the expression of genes in *B. burgdorferi* B31-A3. Previous analyses detailing the impact of *bbq67* on shuttle vector transformation efficiency and epigenetic regulation have relied on strains lacking lp56 (9–13, 39). Therefore, deletion of the *bbq67* locus was conducted to assess the effects of gene regulation on cells containing lp56. RNA-seq was performed on *B. burgdorferi* B31-A3 and B31-A3 derivatives containing isogenic deletions of *bbe02*, *bbq67*,

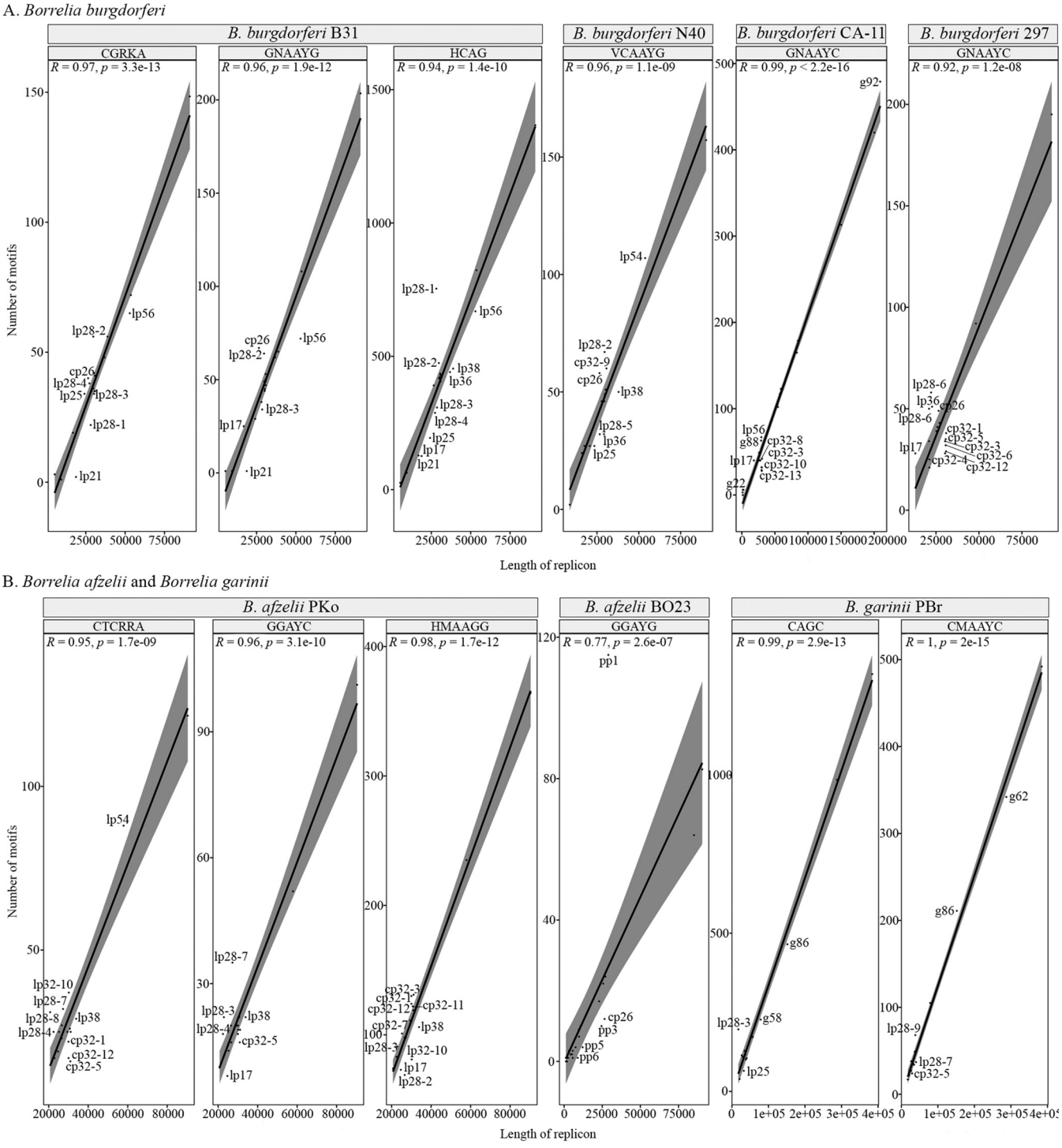

FIG 5 Pearson correlation of the number of motifs to the length of the replicon in *B. burgdorferi* strains B31, N40, CA-11, and 297 (A) and *B. afzelii* strains PKo and BO23 and *B. garinii* strain PBr (B). An *R* value close to 1 with a significant *P* value ($P < 0.01$) indicates a correlation between the number of motifs and the length of the replicon. The regression line and confidence interval along with the *R* and *P* values for each motif are shown. Only replicons that lie outside the confidence interval are labeled. The number of sites for each motif within their respective genomes significantly correlates with the length of the replicon.

and both *bbe02* and *bbq67*. Analysis of the RNA-seq data revealed that there was no obvious difference in gene expression regardless of the presence or absence of the R/M gene(s) when visualized with a principal-component analysis (PCA) plot to determine and visualize the maximum variation between samples (Fig. S5A). Additionally, both edgeR and DESeq2 analyses, which both normalize and analyze the data to determine differentially

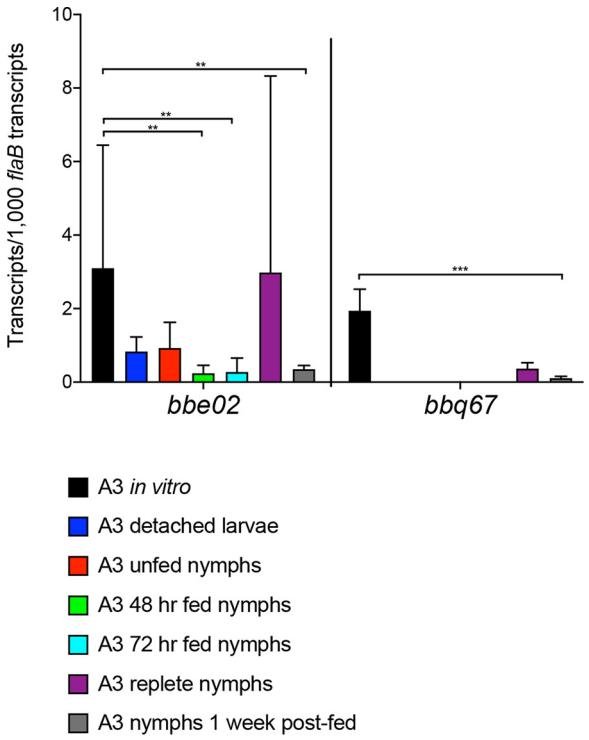

FIG 6 *bbe02* and *bbq67* transcript levels of *B. burgdorferi* B31-A3 grown *in vitro* or within *I. scapularis*. cDNA of total RNA from *B. burgdorferi* B31-A3 *in vitro* or within *I. scapularis* at different stages underwent qPCR. *bbe02* and *bbq67* were expressed at lower levels within *I. scapularis* than in *in vitro*-grown organisms, with *bbq67* being present at undetectable levels in larvae and feeding nymphs. The exception is the expression of *bbe02* at nymphal repletion. Significance was determined by Dunn's multiple comparison of the Kruskal-Wallis test (n = 3 biological and 3 technical replicates each). **, P value of <0.002; ***, P value of <0.0002.

expressed genes, were conducted, with similar results. A comparison of strains containing isogenic deletions of *bbe02*, *bbq67*, and both *bbe02* and *bbq67* revealed two genes that were differentially expressed. No other genes were determined to be significantly differentially expressed by edgeR and DESeq2 (Fig. 7).

As the lack of differential gene expression due to the absence of one or more R/M genes in B31-A3 is quite different than what was reported previously by Casselli et al. (39), additional RNA-seq analyses were performed with independently prepared cDNA libraries of B31-A3, B31-A3 Δ*bbe02*, B31-A3-68, B31-A3-68 Δ*bbe02*, B31-A3 lp28-3^{-}, B31-A34, and B31-A34 Tn::H09 (Table S1). However, similar to our previous data set, these seven strains did not reveal significant differences in gene expression that correlated with R/M gene content. The highly passaged B31-A34 and B31-A34 Tn::H09 strains lack 9 plasmids present in the B31-A3 derivatives, causing these strains to cluster separately in the resulting PCA plot. However, the lack of *bbh09* in the B31-A34 Tn::H09 strain does not result in differential gene expression compared to B31-A34, the isogenic strain from which it was derived (Fig. S5B). Additionally, edgeR and DESeq2 analyses of B31-A3 derivatives revealed that the majority of differentially expressed genes reflected the absence of lp56 in the B31-A3-68 derivatives (Fig. 8).

To confirm the RNA-seq data, qRT-PCR was performed to include differentially regulated genes with a log fold change of >1. Overall, qRT-PCR could not confirm all differentially expressed genes under a less stringent log fold change of >1 (Table 2, Fig. 7, and Table S2, increased and decreased expression).

In addition to the differential expression of numerous genes, Casselli et al. reported the differential expression of the alternative sigma factor *rpoS* and posttranscriptional regulators of *rpoS*, such as *bbd18*, in strains lacking *bbe02* and *bbq67* (39). As *rpoS* is a transcription factor that regulates a large number of genes, this would have a significant

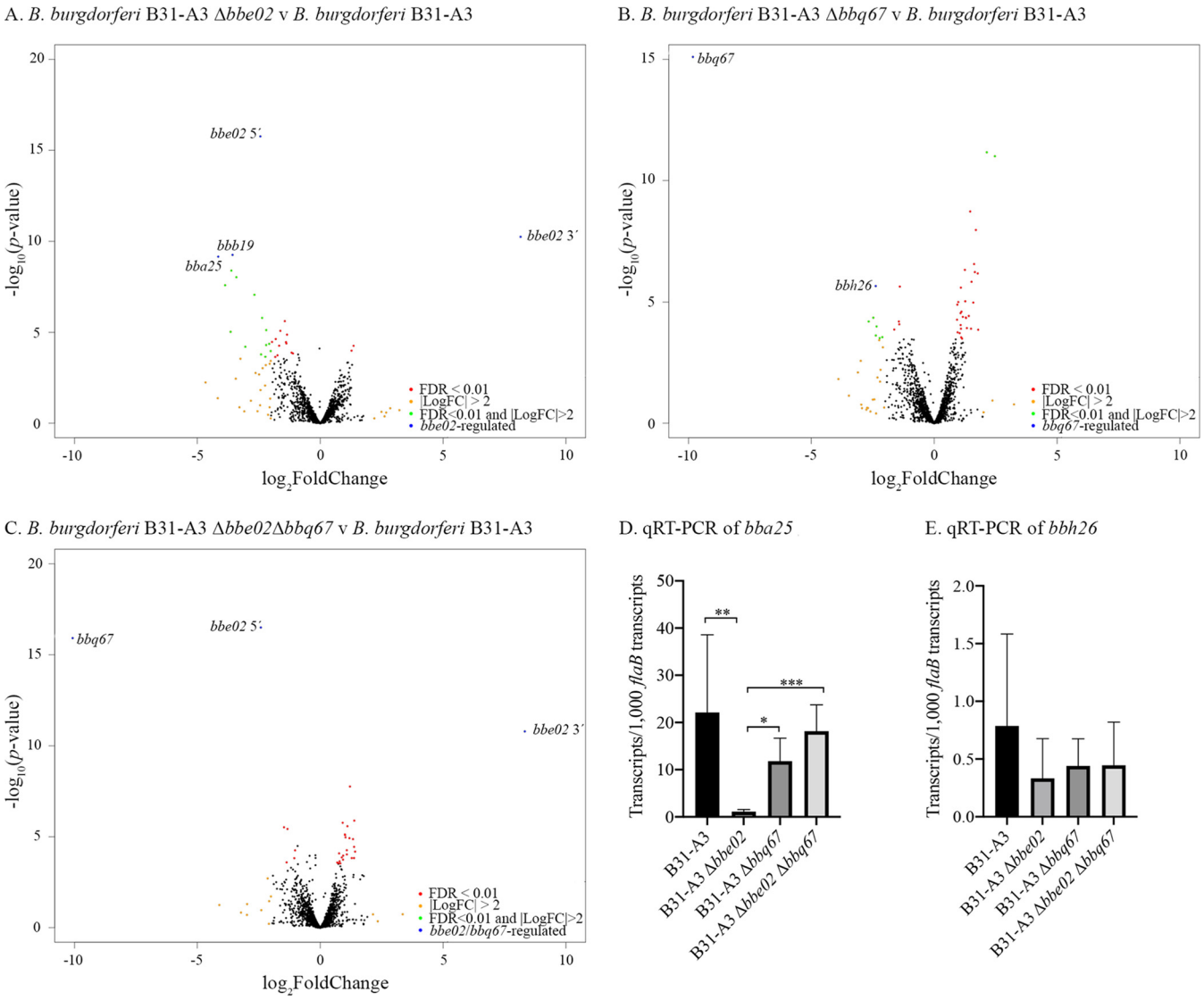

FIG 7 Volcano plots and qRT-PCR graphs of differentially expressed genes. (A to C) Volcano plots of *B. burgdorferi* B31-A3 Δ*bbe02* (A) and *B. burgdorferi* B31-A3 Δ*bbq67* (B) gene expression compared to B31-A3 gene expression and *B. burgdorferi* B31-A3 gene expression compared to B31-A3 Δ*bbe02* Δ*bbq67* gene expression (C). Differentially expressed genes with a false discovery rate (FDR) of <0.01 as determined by DESeq2, those with a log fold change (|logFC|) of >2 as determined by DESeq2, and genes that have an FDR of <0.01 and a log fold change of >2 as determined by DESeq2 are shown. Genes that have an FDR of <0.01 and a log fold change of >2 in both DESeq2 and edgeR are labeled and shown in blue. (D and E) qPCR of B31-A3, B31-A3 Δ*bbe02*, B31-A3 Δ*bbq67*, and B31-A3 Δ*bbe02* Δ*bbq67* to verify decreases in *bba25* gene expression within B31-A3 Δ*bbe02* (D) and *bbh26* gene expression within B31-A3Δ*bbq67* (E) compared to B31-A3 from RNA-seq data. Decreased *bba25* expression in B31-A3 Δ*bbe02* was confirmed by qPCR, but B31-A3 Δ*bbe02* Δ*bbq67* cells do not exhibit significantly lower expression levels of *bba25*. Decreased *bbh26* expression in B31-A3 Δ*bbq67* was not confirmed by qPCR. Significance was determined by Dunn's multiple comparison of the Kruskal-Wallis test (n = 3 biological and 3 technical replicates each). *, P value of <0.05; **, P value of <0.002; ***, P value of <0.0002.

impact on global gene expression (58–64). We therefore carefully examined the expression of sigma factors and response regulators in both RNA-seq data sets and further verified the relative levels of *rpoN*, *rpoS*, and *bbd18* transcripts by qRT-PCR. Consistent with our previous analyses and different from those of Casselli et al. (39), the expression levels of sigma factors and response regulators did not differ significantly between isogenic B31 strains containing and those lacking R/M loci by either analysis (Table S2).

DISCUSSION

DNA modification by R/M enzymes has been recognized as having diverse impacts on microorganisms (14–16). In *B. burgdorferi* type strain B31, two plasmid-encoded type II R/M enzymes, BBE02 and BBQ67, significantly impact the stable uptake and incorporation of exogenous DNA (9–13). Analysis of 8 *B. burgdorferi* B31 derivatives

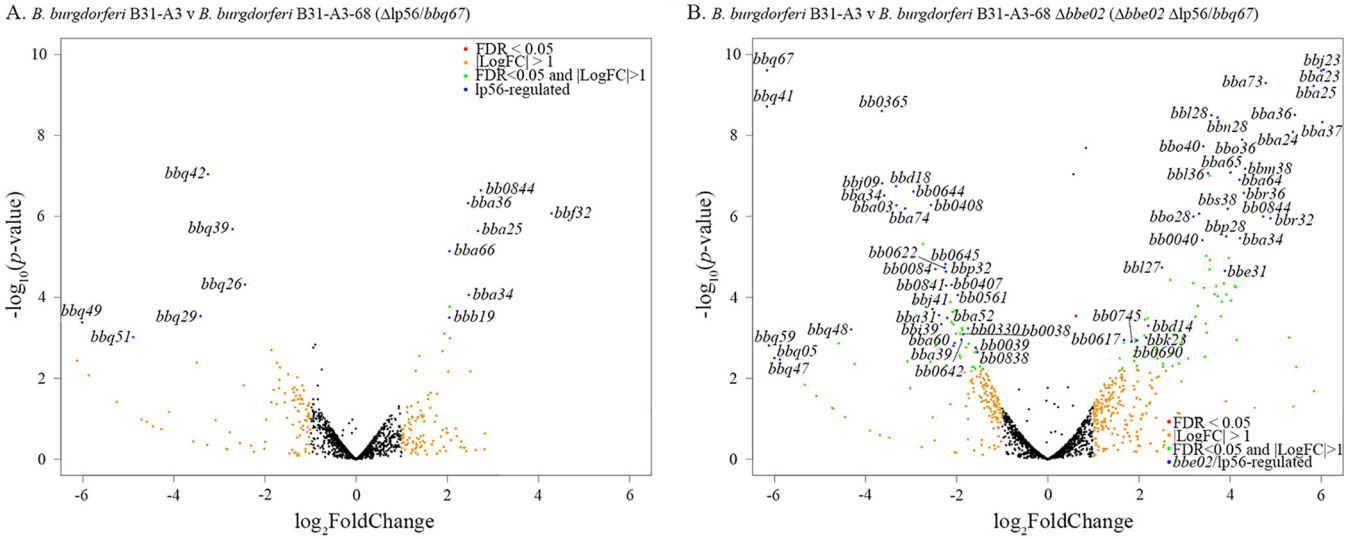

FIG 8 Volcano plots of *B. burgdorferi* B31-A3 gene expression compared to B31-A3-68 (lp56$^-$) gene expression (A) and *B. burgdorferi* B31-A3 gene expression compared to B31-A3-68 Δ*bbe02* (Δ*bbe02*/lp56$^-$) gene expression (B). Differentially expressed genes with an FDR of <0.01 as determined by DESeq2, those with a log fold change of >2 as determined by DESeq2, and genes that have an FDR of <0.01 and a log fold change of >2 in DESeq2 are shown. Genes that have an FDR of <0.01 and a log fold change of >2 by both DESeq2 and edgeR are labeled and shown in blue.

along with *B. burgdorferi* N40, CA-11, and 297; *B. afzelii* PKo and BO23; and *B. garinii* PBr revealed 9 methylation motifs. In addition to confirming the recently described GNAAYG (modified by BBE02) and CGRKA (modified by BBQ67) motifs, an H/WCAG motif modified by BBH09 was identified within *B. burgdorferi* type strain B31 (39). Despite the high sequence identity among BBE02 (~64 to 98%) and BBQ67 (~65%) orthologs analyzed in this study, the majority of methylated motifs are distinct and novel, suggesting a possible impact on genetic exchange between species and strains (Table 1 and Fig. 1 and 2). However, orthologous genes contain similar motif specificities, as observed in *B. burgdorferi* CA-11 and 297, which methylate the same motif, GNAAYC, which is similar to the *B. burgdorferi* B31 BBE02 motif (Fig. 1 and 2). The cause for motif variability, therefore, is more likely through the acquisition, loss, and pseudogenization of paralogous R/M genes than by sequence evolution among orthologs. While the majority of genomes analyzed carry R/M genes without indels, *B. burgdorferi* 297 lacked lp28-5, which carries the R/M genes *Bbu297_y09* and *Bbu297_y05* and

TABLE 2 Genes with significantly decreased expression in B31-A3 Δ*bbe02* and B31-A3 Δ*bbq67* compared to B31-A3, with an FDR of <0.05 and a $\log_2$ fold change of >1

Gene	$\log_2$ fold change	FDR[a]	No. of BBE02 sites	No. of BBQ67 sites	Confirmation by qPCR?
bba25[b]	edgeR, −3.75 DESeq2, −4.15	edgeR, 4.56×10^{-2} DESeq2, 2.35×10^{-7}	0	1	Yes, but expression does not appear to be significantly different in the absence of both *bbe02* and *bbq67*[b]
bbb19	edgeR, −3.15 DESeq2, −3.57	edgeR, 4.89×10^{-2} DESeq2, 2.35×10^{-7}	2	2	Did not determine as *ospC* expression is variable
bbe02	edgeR, −2.40 DESeq2, −2.43	edgeR, 9.11×10^{-3} DESeq2, 2.33×10^{-13}	4	4	Did not determine as *bbe02* contains a deletion/insertion
bbh26[b]	edgeR, −2.21 DESeq2, −2.38	edgeR, 2.04×10^{-2} DESeq2, 2.50×10^{-4}	0	0	No[b]
bbq67	edgeR, −10.45 DESeq2, −9.82	edgeR, 1.36×10^{-5} DESeq2, 1.03×10^{-12}	2	1	Did not determine as *bbq67* is deleted

[a]FDR, false discovery rate, the adjusted *P* value that gives the proportion of false positives expected among differentially expressed genes.
[b]See Fig. 7.

contains several silent and missense variations within its other R/M genes, *Bbu297_h03* and *Bbu297_j01*. Additionally, all *B. burgdorferi* N40 derivatives contain a single missense variation within the R/M gene *Bbun40_e01*, and *BLA32_04945* in *B. afzelii* BO23 contains a frameshift (see Table S2 in the supplemental material [R/M sequencing]). However, it is unknown if the identified indels would inactivate the resulting R/M enzyme. All previously and newly identified *Borrelia* motifs were of the m6A type, and the number of motifs identified within each genome corresponds to the number of encoded, intact m6A R/M enzymes, except for *B. burgdorferi* CA-11 (Table 1 and Fig. 2). Further work is required to assess and confirm the absence of cytosine methylation within all sequenced *B. burgdorferi* N40 derivatives that carry *Bbun40_y07*, a homolog of a *Haemophilus haemolyticus* cytosine methyltransferase (49).

Similar to other bacterial R/M systems, the majority of *Borrelia* R/M genes analyzed in this study methylated ≥95% of their respective motifs (20). However, the H/WCAG motif of *B. burgdorferi* B31, the GNAAYG and CGRKA motifs of *B. burgdorferi* B31-MI, the GNAAYC motif of *B. burgdorferi* 297, the HMAAGG motif of *B. afzelii* PKo, and the CMAAYC and CAGC motifs of *B. garinii* PBr were not efficiently methylated. Apart from *Bbu297_h03*, *Bbu297_j01*, *Bbun40_e01*, and *BLA32_04945*, it is not believed that mutations have resulted in this decrease in methylation activities as all other R/M genes lack indels (Table S2, R/M sequencing). The relatively modest level of methylation (~65 to 92%) in nonclonal *B. burgdorferi* B31-MI and 297, *B. afzelii* PKo, and *B. garinii* PBr can be explained by the absence of the plasmids carrying the R/M genes in subsets of their populations (Table 1) (41).

Successful transformation of *B. burgdorferi* strain B31 derivatives largely depends on the presence or absence of R/M genes and on the transforming DNA. As previously noted, *B. burgdorferi* B31 derivatives that lack one or more R/M genes, such as B31-A34 and B31-S9, are more readily transformed with shuttle vectors (9–13). Additionally, the transformation of *B. burgdorferi* N40 was shown to be enhanced after the deletion of *Bbun40_e01* (65), and shuttle vector transformation has been successful in *B. burgdorferi* CA-11.2A (34) and 297 (64) and *B. afzelii* BO23 (66). However, the impacts of the R/M genes in *B. afzelii* PKo and *B. garinii* PBr on the transformation efficiency are unknown. As the relative number of *B. burgdorferi* B31 CGRKA and GNAAYG motifs within the shuttle vectors pBSV2, pBSV2G, and pKFSS1 is consistent with previously reported transformation efficiencies (13), it is believed that the large number of *B. afzelii* PKo and *B. garinii* PBr motifs in these shuttle vectors would have a negative impact on the transformation efficiency.

In *B. burgdorferi* B31, it has been shown that *in vitro* methylation with the CpG methyltransferase M.SssI increased the rate of pBSV2 and pKFSS1 shuttle vector transformation in a BBQ67-dependent manner. Investigation of these shuttle vectors revealed that all CGRKA sites and ~75% of GNAAYG sites contain a cytosine that would be methylated by M.SssI (12). Furthermore, previously identified sites protected from RsaI cleavage by BBQ67 methylation were found to overlap CGRKA methylated motifs in pBSV2, pBSV2G, and pKFSS1 (13) (Fig. S6). As the presence of R/M genes primarily affects shuttle vector transformation and has relatively little influence on the rate of allelic exchange, the construction of a shuttle vector lacking the CGRKA and GNAAYG motifs should allow the more efficient transformation of *B. burgdorferi* B31 derivatives expressing BBQ67 and BBE02 (9–13, 67).

Our attempts to create a functional shuttle vector devoid of all CGRKA and GNAAYG motifs have been unsuccessful to date. The removal of these sites in the ColE1 origin of replication or within broad-host-range plasmids such as pEP2 resulted in plasmids that were unable to replicate within *Escherichia coli* (Table S2, ori). However, codon optimization and the removal of CGRKA and GNAAYG motifs within the *flg* promoter and the antibiotic cassettes *aadA1* (GenBank accession number MW473471), *aacC1* (accession number MW473470), and *aph* (accession number MW473472) conferred resistance to the respective antibiotics in *E. coli*. Furthermore, modified *flgp-aph* was shown to confer kanamycin resistance to *B. burgdorferi* B31-A3 following allelic exchange at *bbq67*.

Recent work has demonstrated that a putative methyltransferase gene of *Babesia bigemina*, a pathogen associated with tick vectors, is expressed only within the tick

environment (68). Analysis of the relative levels of *bbe02* and *bbq67* transcripts showed significantly higher expression levels *in vitro* than within *B. burgdorferi* B31-A3-infected larvae and nymphs, except immediately following the detachment of fully engorged nymphs from naive mice (Fig. 6). However, the role of R/M gene expression within the tick, where horizontal gene transfer among *Borrelia* bacteria likely occurs, remains unknown (67). Analyses of *B. burgdorferi* B31 genes that have strong evidence for gene transfer, such as *ospC* (54), *ospD* (55), the *erp* genes (37, 56), and the decorin binding protein (*dbp*) genes (57), revealed that there is no appreciable difference in the numbers of motifs when the codons are synonymously shuffled (as determined by the log fold change [Table S2, CodonShuffle]). This may indicate that motif presence is driven by codon bias. However, a comparison of these genes with randomly assembled noncoding sequences of the same length revealed that 6 out of 15 genes and their flanking regions contained significantly more GNAAYG or CGRKA motifs than would be expected by chance (Table S2, genes of interest). *ospC* was the only gene analyzed to contain significantly higher numbers of both GNAAYG and CGRKA motifs than would be expected by chance. Additionally, *ospC* carried by the other strains in this study showed a significantly higher rate of occurrence of at least one of the newly identified motifs. Further analysis into the presence of the other known *Borrelia* R/M motifs in the B31 *ospC* gene revealed that the majority of identified motifs were also present within the central portion of the coding region. A previous investigation into recombination via gene transfer at *ospC* demonstrated the highest recombination breakpoints and diversity in the middle of the coding sequence (54). This leads us to speculate that the presence of *Borrelia* R/M motifs within *ospC* could facilitate the transduction of cp26 DNA or recombination via gene transfer (34, 54, 69). Therefore, while R/M motifs are typically viewed as a means to degrade foreign DNA, in the case of *ospC*, and perhaps other horizontally transferred DNAs, it may be beneficial by providing breakpoints for packaging DNA into phage or for recombination into the native locus. The utilization of R/M motifs for this purpose would provide a higher probability of generating novel, chimeric, immunodominant serotype-defining *ospC* alleles (54).

In an attempt to understand the incomplete methylation of HCAG motifs within *B. burgdorferi* B31, the methylation patterns of each replicon were analyzed. This revealed that the number of CGRKA, GNAAYG, or HCAG motifs correlates with the length of the replicons, while the number of methylated HCAG motifs does not (Fig. 5 and Fig. S4). Neither the absence of *bbe02* and *bbq67* in *B. burgdorferi* B31-A3-68 Δ*bbe02* and B31-A34 nor the overexpression of *bbh09* improves the methylation of HCAG motifs (Table S2, A3-68-LS). While this could be due to the activity of the enzyme, it could also be due to a blockage of the HCAG motif by DNA binding proteins. Of the 16,677 HCAG motifs within the *B. burgdorferi* B31 genome, there are 8 BosR binding sites (70, 71), 2,203 EbfC binding sites (72), and 10,362 BpaB binding sites (73) within 100 bp surrounding an HCAG motif, while there are 0 and 4 BosR sites, 280 and 394 EbfC sites, and 1,506 and 1,861 BpaB sites within 100 bp of CGRKA and GNAAYG motifs, respectively. Within all B31 derivatives except A34, there are 2,617 HCAG motifs that are consistently methylated and 4,682 that are never methylated. Further analysis of 100 bp around the 4,682 unmethylated motifs revealed that 1 contains the BosR binding site, 540 contain the EbfC binding site, and 2,606 contain the BpaB binding site. In comparison, there are 179 unmethylated CGRKA motifs and 418 GNAAYG motifs within B31-MI, with 0 and 1 BosR sites, 15 and 70 EbfC sites, and 103 and 221 BpaB sites within 100 bp, respectively. While this does not explain all of the unmethylated HCAG motifs, DNA binding proteins may have a negative impact on DNA methylation by BBH09.

Although the majority of methylated motifs occur fairly consistently among the replicons of each *Borrelia* genome analyzed, the CAGC motif of *B. garinii* PBr is highly prevalent (174 sites within ~5,000 bp) and methylated at the left telomeric end of lp28-3 until it encounters the 3′ end of the *bgapbr_h0006* R/M enzyme (Fig. 4 and Fig. S3). Unfortunately, as *B. garinii* PBr encodes two R/M enzymes, it is unknown whether this motif is methylated by BGAPBR_H0006. These motifs lie within the *vlsE* expression locus and *vls* silent cassettes of *B. garinii* PBr. In fact,

a closer analysis of the *vls* loci within *B. burgdorferi* B31, 297, and N40 and *B. afzelii* PKo revealed a high concentration of CAGC motifs within these strains and derivatives as well. As the high number of *Borrelia* R/M motifs within the *ospC* gene is presumed to aid in the horizontal transfer and antigenic variation of immunodominant OspC, the high number of CAGC motifs within the *vls* genes may aid in antigenic variation through gene conversion of *vlsE* within spirochetes.

While a previous investigation of the *B. burgdorferi* B31 methylome concluded that the BBE02 and BBQ67 R/M enzymes had a global impact on gene regulation, our current study failed to find any significant evidence for the epigenetic regulation of the *Borrelia* transcriptome by either BBE02 or BBQ67 (Table 2 and Fig. 7). Despite lowering the stringency of RNA-seq analysis to identify differentially expressed genes, these could not be consistently validated by qRT-PCR (Table S2, increased and decreased expression). Furthermore, the current study analyzed *B. burgdorferi* B31-A3, B31-A3 Δ*bbe02*, B31-A3 Δ*bbq67*, and B31-A3 Δ*bbe02* Δ*bbq67* that retained all native plasmids, while the previous study analyzed RNA-seq data generated from *B. burgdorferi* B31-A3 and B31-A3 Δ*bbe02* that lacked lp38 and from 5A18-NP1 Δ*bbe02* that lacked lp28-4 in addition to lp56 (39). Additionally, the striking impact of R/M gene content on the expression of sigma factors and posttranscriptional regulators evidenced by Casselli et al. was not reproduced in two independently generated RNA-seq data sets during our current analysis. Variable culture conditions leading to differential sigma factor expression were most likely responsible for the putative epigenetic impact of R/M loci on gene expression in the previous study (39) and would explain this discrepancy.

MATERIALS AND METHODS

Phylogenetic tree construction. The BBE02 and BBQ67 nucleotide and amino acid sequences were used to identify similar genes within both Lyme disease and tick-borne relapsing fever spirochetes within the NCBI database as well as to search for *Borrelia* R/M genes within REBASE (see Table S1 in the supplemental material [R/M homologs and REBASE]). The identified proteins were aligned with Clustal Omega, and the percentages of identity and similarity between genes were determined with the Ident and Sim sequence analysis tool from the Sequence Manipulation Suite (74, 75). The phylogenetic tree was inferred by approximately maximum likelihood methods implemented through FastTree (76). This analysis utilized the JTT (Jones-Taylor-Thornton) model of amino acid evolution with the CAT approximation to account for the various rates of evolution across sites and estimated the local support values with the Shimodaira-Hasegawa (SH) test (77–79). The resulting tree was visualized with iTOL v 4.4.2 (80).

***B. burgdorferi* strains and growth conditions.** *B. burgdorferi* strains were cultured in Barbour-Stoenner-Kelly II (BSKII) medium supplemented with 6% rabbit serum (PelFreez Biologicals, Rogers, AZ) and the appropriate antibiotics (streptomycin at 50 μg/ml and kanamycin at 200 μg/ml) at 35°C under 2.5% CO_2 (81). *B. burgdorferi* B31 MI obtained from MedImmune (now AstraZeneca, Gaithersburg, MD) was subjected to PacBio sequencing as this was the gDNA used to generate the B31 genomic sequencing data for the type strain. *B. burgdorferi* strain B31-A3, an infectious clone derived from B31 MI lacking cp9, was used as the wild-type strain in this study (41). Cloning vectors were propagated using *E. coli* strain TOP10 (Invitrogen, Carlsbad, CA). All *B. burgdorferi* strains and derivatives along with plasmids utilized in this study are shown in Table S1 (bacterial strains).

Assembly of constructs and transformation of *B. burgdorferi*. *B. burgdorferi* was transformed by electroporation as previously described (82). Competent *B. burgdorferi* bacteria were freshly prepared from an exponential-phase culture and electroporated with 15 to 30 μg of plasmid DNA prepared from *E. coli*. Transformants were confirmed by PCR and sequencing, and the plasmid content was determined to ensure that no plasmids were lost during transformation (83).

All primers used in this study are listed in Table S1 (oligonucleotides). To generate the B31-A3 Δ*bbq67* and B31-A3 Δ*bbe02* Δ*bbq67* strains, 500 bp upstream and downstream of *bbq67* were PCR amplified and cloned flanking a kanamycin resistance cassette driven by the *flgB* promoter into the pCR-Blunt II-TOPO vector (Thermo Fisher Scientific). B31-A3 and B31-A3 Δ*bbe02* were electroporated with the resulting construct and plated on solid medium containing either kanamycin (B31-A3 Δ*bbq67*) or streptomycin and kanamycin (B31-A3 Δ*bbe02* Δ*bbq67*) to generate the B31-A3 Δ*bbq67* and B31-A3 Δ*bbe02* Δ*bbq67* strains (Fig. 9).

The antibiotic cassettes *aacC1* (gentamicin 3-*N*-acetyltransferase), *aadA1* (streptomycin 3-adenylyltransferase), and *aph* (aminoglycoside phosphotransferase) (kanamycin resistance) were optimized for electroporation into *B. burgdorferi* B31 through the removal of CGRKA and GNAAYG sites along with *Borrelia* codon optimization (GenScript, Piscataway, NJ). These optimized cassettes were driven by a *B. burgdorferi* R/M-minus *flg* promoter and transformed into *E. coli* TOP10 cells as part of a pUC57 vector. These optimized antibiotic cassettes were shown to confer the appropriate antibiotic resistance (gentamicin at 5 μg/ml, spectinomycin at 100 μg/ml, and kanamycin at 50 μg/ml) in transformed *E. coli*. Additionally, *flg*-driven *aph* was electroporated into *B. burgdorferi* B31-A3 and conferred kanamycin resistance to transformed cells through allelic exchange at *bbq67*.

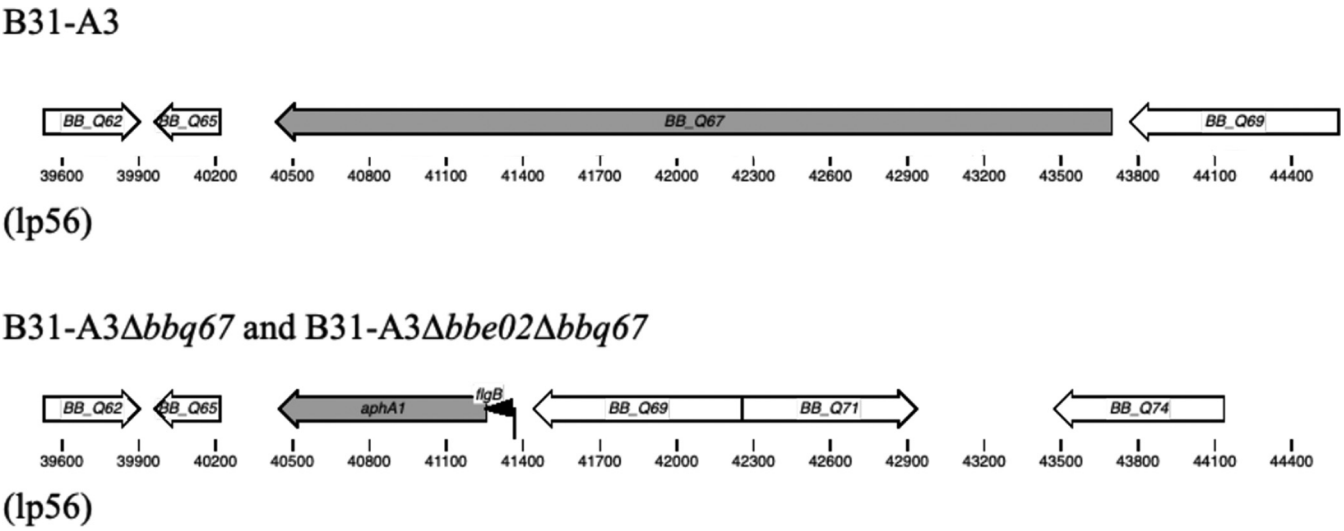

FIG 9 *bbq67* isogenic deletion. To create the B31-A3 Δ*bbq67* and B31-A3 Δ*bbe02* Δ*bbq67* strains, the 3,261-bp *bbq67* gene was replaced with the 919-bp *flgB*::*aphA1* antibiotic resistance cassette in the same orientation as *bbq67*.

Genome sequencing and motif and methylation analyses. Genomic DNA (gDNA) was isolated from *B. burgdorferi* MI, B31-A3, B31-A3 Δ*bbe02*, B31-A3-68, B31-A3-68 Δ*bbe02*, B31-A3 Δlp28-3, B31-A34, B31-A34 Tn::H09, N40, CA-11, CA-11.2a, and 297; *B. garinii* PBr and BO23; and *B. afzelii* PKo with Qiagen genomic DNA extraction according to the manufacturer's protocol (Qiagen, Hilden, Germany).

gDNA was used to generate 16 barcoded SMRTbell libraries (adapter kits 8A and 8B) and subjected to SMRT sequencing according to the manufacturer's instructions for multiplex microbial SMRTbell libraries v2 (Pacific BioSciences, Menlo Park, CA) with the following modifications. Two pools were independently generated for sequencing on two SMRT cells on the PacBio Sequel platform (Pacific BioSciences, Menlo Park, CA). The primer and polymerase were annealed to the first pool, which consisted of all 16 libraries, according to a non-size-selection protocol. The second pool was the same except for following the optional size selection protocol. Each SMRT cell underwent diffusion loading with a preextension time of 120 min and a 10-h movie time. The non-size-selected pool was loaded at 4.5 pM, while the size-selected pool was loaded at 7 pM. For each SMRT cell, SMRT Link RunQC showed a P1 value of >50% with N_{50} longest subreads of 9,250 bp and 12,750 bp, respectively. Reads were processed and mapped to the appropriate references using the pbsmrtpipe ds_modification_detection and sa3_ds_resequencing_fat pipelines (Pacific BioSciences). The references used for assembly were GenBank accession numbers AE000783 to AE000794 and AE001575 to AE001584 for B31; CP001651 and CP002227 to CP002242 for N40; ABJV02000001 to ABJV02000005 and CP001301 to CP001311 for PBr; CP002933 to CP002950 for PKo; CP001653, CP002312, and CP002253 to CP002270 for 297; ABJY02000001 to ABJY02000014 and CP001473 to CP001484 for CA-11; and CP018262 to CP018293 for BO23. Only base calls that had a quality value of 20 or higher were used for analysis.

To determine if the number of encoded motifs is higher than would be expected by chance, the Analysis of Motif Enrichment (AME) program of the MEME suite 5.0.5 package was utilized (52, 84). This analysis shuffles the FASTA sequences of each genome or gene to conserve 2-mer frequencies and create, at minimum, 1,000 randomized control sequences for motif comparison. AME then determines if each motif is enriched in the primary sequence using one-tailed Fisher's exact test and performs partition maximization over all possible position-weight matrix (PWM) thresholds, with false (control sequences) and true (input sequences) positives determined by their PWM scores (52). Codons within genes were synonymously shuffled by utilizing the N3 script of CodonShuffle in which the third position of each codon was shuffled to generate 1,000 shuffled sequences for each gene (53).

The number of motifs and the number of methylated motifs per 1,000 bp were determined and mapped using Circos version 0.69-7 (85) and analyzed using Pearson correlation. Graphs were generated using the ggpubr package of the ggplot2 version 3.2.0 tool of the tidyverse package in R (86).

RNA isolation, sequencing, and quantitative real-time PCR. Total RNA was isolated from three biological replicates of *B. burgdorferi* strains B31-A3, B31-A3 Δ*bbe02*, B31-A3 Δ*bbq67*, and B31-A3 Δ*bbe02* Δ*bbq67* and from two biological replicates of *B. burgdorferi* B31-A3, B31-A3 Δ*bbe02*, B31-A3-68, B31-A3-68 Δ*bbe02*, B31-A3 Δlp28-3, B31-A34, and B31-A34 Tn::H09 in mid-log phase (5×10^7 to 9×10^7 cells/ml) (Table S1). Cells were collected by centrifugation and treated with RNAprotect (Qiagen). RNA was isolated using TRIzol reagent (Life Technologies, Carlsbad, CA) according to the manufacturer's instructions and treated with 1 U of DNase I (Ambion, Foster City, CA) for 1 h at 37°C. RNA was quantified and subjected to Agilent Bioanalyzer 2200 Tape Station (Agilent, Santa Clara, CA) quality assessment. RNAs possessing RNA integrity number (RIN) values of ≥7.4 were used for downstream analysis.

cDNA libraries for RNA-seq and quantitative PCR (qPCR) were generated from these total RNAs. cDNA libraries used for RNA-seq were synthesized using Ribo-Zero and ScriptSeq complete bacterial kits with indexing primers for the synthesis of directional libraries (Illumina, San Diego, CA) according to the manufacturer's instructions. The

quality and quantity of the resulting cDNA libraries were assessed with an Agilent DNA 1000 assay on an Agilent 2100 Bioanalyzer (Agilent) and with a Kapa library quantification kit (Kapa Biosystems, Wilmington, MA) prior to RNA-seq. RNA-seq libraries were sequenced on an Illumina MiSeq platform using v3 chemistry with 2- by 75-bp reads.

RNA-seq reads were compiled, filtered to remove any reads with PHRED scores of less than 10, and aligned to the *B. burgdorferi* B31 genome (GenBank accession numbers AE000783 to AE000794 and AE001575 to AE001584) using bowtie2 (87). Reads for annotated genes were determined using featureCounts (88, 89). Differential expression analysis was conducted with edgeR (90) and DESeq2 (91).

For analysis of *bbe02* and *bbq67* expression within the tick vector, the total RNA and resulting cDNA from *B. burgdorferi* B31-A3-infected *Ixodes scapularis* ticks that were previously described were used for qPCRs (92, 93). To generate cDNA used for qPCR from *B. burgdorferi* cultured *in vitro*, 1 μg of RNA was reverse transcribed using the high-capacity cDNA reverse transcriptase kit (Life Technologies) according to the manufacturer's instructions. qPCRs were performed using IQ SYBR green supermix (Bio-Rad Life Sciences, Hercules, CA) with gene-specific primer sets (500 nM) (Table S1). Reactions were performed so that each experiment contained a biological and technical triplicate on a Viia7 real-time PCR system (Applied Biosystems, Foster City, CA) and analyzed with PRISM software. Negative-control reactions of primers lacking a template and on RNA samples that underwent cDNA reactions in the absence of reverse transcriptase were performed with each reaction to ensure that threshold cycle (C_T) values were not obtained from primer-primer interactions or from contaminating genomic DNA. Additionally, the melt curve was analyzed for each reaction. All primers used for qPCR are listed in Table S1 (oligonucleotides).

Data availability. Sequences of standard *Borrelia* selectable markers that were optimized for *B. burgdorferi* B31 transformation by the removal of CGRKA and GNAAYG motifs are located in GenBank under accession numbers MW473470 (*aacC1*), MW473471 (*aadA1*), and MW473472 (*aph*). The RNA-seq data are available under GEO accession number GSE169460.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

FIG S1, TIF file, 0.2 MB.

FIG S1A to C, TIF file, 0.2 MB.

FIG S1D to G, TIF file, 0.2 MB.

FIG S3, TIF file, 1.7 MB.

FIG S4, TIF file, 0.8 MB.

FIG S5, TIF file, 0.2 MB.

FIG S6, TIF file, 0.9 MB.

TABLE S1, XLSX file, 0.1 MB.

TABLE S2, XLSX file, 0.7 MB.

ACKNOWLEDGMENTS

We thank Gang Fang, John Beaulaurier, and Pedro Oliveira for initial PacBio SMRT sequencing, analysis, and helpful discussions on *Borrelia burgdorferi* R/M motifs. We also acknowledge Stacy Ricklefs and Dan Bruno for their help with RNA sequencing. We are grateful to Kendal Cooper, Adam Nock, Sherwood Casjens, and Weigang Qui for reviewing the manuscript and providing helpful comments.

This research was supported by the Intramural Research Program of the National Institute of Allergy and Infectious Diseases, National Institutes of Health.

REFERENCES

1. Burgdorfer W, Barbour A, Hayes S, Benach J, Grunwaldt E, Davis J. 1982. Lyme disease—a tick-borne spirochetosis? Science 216:1317–1319. https://doi.org/10.1126/science.7043737.
2. Steere AC, Grodzicki RL, Kornblatt AN, Craft JE, Barbour AG, Burgdorfer W, Schmid GP, Johnson E, Malawista SE. 1983. The spirochetal etiology of Lyme disease. N Engl J Med 308:733–740. https://doi.org/10.1056/NEJM198303313081301.
3. Lane RS, Piesman J, Burgdorfer W. 1991. Lyme borreliosis: relation of its causative agent to its vectors and hosts in North America and Europe. Annu Rev Entomol 36:587–609. https://doi.org/10.1146/annurev.en.36.010191.003103.
4. Baranton G, Assous M, Postic D. 1992. Three bacterial species associated with Lyme borreliosis. Clinical and diagnostic implications. Bull Acad Natl Med 176:1075–1085.
5. Steere AC, Strle F, Wormser GP, Hu LT, Branda JA, Hovius JWR, Li X, Mead PS. 2016. Lyme borreliosis. Nat Rev Dis Primers 2:16090. https://doi.org/10.1038/nrdp.2016.90.
6. Xu Y, Kodner C, Coleman L, Johnson RC. 1996. Correlation of plasmids with infectivity of *Borrelia burgdorferi* sensu stricto type strain B31. Infect Immun 64:3870–3876. https://doi.org/10.1128/IAI.64.9.3870-3876.1996.
7. Purser JE, Norris SJ. 2000. Correlation between plasmid content and infectivity in *Borrelia burgdorferi*. Proc Natl Acad Sci U S A 97:13865–13870. https://doi.org/10.1073/pnas.97.25.13865.
8. Labandeira-Rey M, Skare JT. 2001. Decreased infectivity in *Borrelia burgdorferi* strain B31 is associated with loss of linear plasmid 25 or 28-1. Infect Immun 69:446–455. https://doi.org/10.1128/IAI.69.1.446-455.2001.
9. Lawrenz MB, Kawabata H, Purser JE, Norris SJ. 2002. Decreased electroporation efficiency in *Borrelia burgdorferi* containing linear plasmids lp25 and lp56: impact on transformation of infectious *B. burgdorferi*. Infect Immun 70:4798–4804. https://doi.org/10.1128/iai.70.9.4798-4804.2002.
10. Jacobs MB, Norris SJ, Phillippi-Falkenstein KM, Philipp MT. 2006. Infectivity of the highly transformable BBE02$^-$ lp56$^-$ mutant of *Borrelia burgdorferi*, the Lyme disease spirochete, via ticks. Infect Immun 74:3678–3681. https://doi.org/10.1128/IAI.00043-06.

11. Kawabata H, Norris SJ, Watanabe H. 2004. BBE02 disruption mutants of *Borrelia burgdorferi* B31 have a highly transformable, infectious phenotype. Infect Immun 72:7147–7154. https://doi.org/10.1128/IAI.72.12.7147-7154.2004.
12. Chen Q, Fischer JR, Benoit VM, Dufour NP, Youderian P, Leong JM. 2008. *In vitro* CpG methylation increases the transformation efficiency of *Borrelia burgdorferi* strains harboring the endogenous linear plasmid lp56. J Bacteriol 190:7885–7891. https://doi.org/10.1128/JB.00324-08.
13. Rego ROM, Bestor A, Rosa PA. 2011. Defining the plasmid-borne restriction-modification systems of the Lyme disease spirochete *Borrelia burgdorferi*. J Bacteriol 193:1161–1171. https://doi.org/10.1128/JB.01176-10.
14. Doerfler W. 1983. DNA methylation and gene activity. Annu Rev Biochem 52:93–124. https://doi.org/10.1146/annurev.bi.52.070183.000521.
15. Modrich P, Lahue R. 1996. Mismatch repair in replication fidelity, genetic recombination, and cancer biology. Annu Rev Biochem 65:101–133. https://doi.org/10.1146/annurev.bi.65.070196.000533.
16. Casadesus J, Low D. 2006. Epigenetic gene regulation in the bacterial world. Microbiol Mol Biol Rev 70:830–856. https://doi.org/10.1128/MMBR.00016-06.
17. Korlach J, Turner SW. 2012. Going beyond five bases in DNA sequencing. Curr Opin Struct Biol 22:251–261. https://doi.org/10.1016/j.sbi.2012.04.002.
18. Beaulaurier J, Schadt EE, Fang G. 2019. Deciphering bacterial epigenomes using modern sequencing technologies. Nat Rev Genet 20:157–172. https://doi.org/10.1038/s41576-018-0081-3.
19. Murray IA, Clark TA, Morgan RD, Boitano M, Anton BP, Luong K, Fomenkov A, Turner SW, Korlach J, Roberts RJ. 2012. The methylomes of six bacteria. Nucleic Acids Res 40:11450–11462. https://doi.org/10.1093/nar/gks891.
20. Oliveira PH, Ribis JW, Garrett EM, Trzilova D, Kim A, Sekulovic O, Mead EA, Pak T, Zhu S, Deikus G, Touchon M, Lewis-Sandari M, Beckford C, Zeitouni NE, Altman DR, Webster E, Oussenko I, Bunyavanich S, Aggarwal AK, Bashir A, Patel G, Wallach F, Hamula C, Huprikar S, Schadt EE, Sebra R, van Bakel H, Kasarskis A, Tamayo R, Shen A, Fang G. 2020. Epigenomic characterization of Clostridioides difficile finds a conserved DNA methyltransferase that mediates sporulation and pathogenesis. Nat Microbiol 5:166–180. https://doi.org/10.1038/s41564-019-0613-4.
21. Noyer-Weidner M, Trautner TA. 1993. Methylation of DNA in prokaryotes, p 39–108. *In* Jost JP, Saluz HP (ed), DNA methylation: molecular biology and biological significance, vol 64. Birkhauser, Basel, Switzerland.
22. Roberts RJ, Vincze T, Posfai J, Macelis D. 2015. REBASE—a database for DNA restriction and modification: enzymes, genes and genomes. Nucleic Acids Res 43:D298–D299. https://doi.org/10.1093/nar/gku1046.
23. Blow MJ, Clark TA, Daum CG, Deutschbauer AM, Fomenkov A, Fries R, Froula J, Kang DD, Malmstrom RR, Morgan RD, Posfai J, Singh K, Visel A, Wetmore K, Zhao Z, Rubin EM, Korlach J, Pennacchio LA, Roberts RJ. 2016. The epigenomic landscape of prokaryotes. PLoS Genet 12:e1005854. https://doi.org/10.1371/journal.pgen.1005854.
24. Davis BM, Chao MC, Waldor MK. 2013. Entering the era of bacterial epigenomics with single molecule real time DNA sequencing. Curr Opin Microbiol 16:192–198. https://doi.org/10.1016/j.mib.2013.01.011.
25. Zhang Z, Wang J, Wang J, Wang J, Li Y. 2020. Estimate of the sequenced proportion of the global prokaryotic genome. Microbiome 8:134. https://doi.org/10.1186/s40168-020-00903-z.
26. Fang G, Munera D, Friedman DI, Mandlik A, Chao MC, Banerjee O, Feng Z, Losic B, Mahajan MC, Jabado OJ, Deikus G, Clark TA, Luong K, Murray IA, Davis BM, Keren-Paz A, Chess A, Roberts RJ, Korlach J, Turner SW, Kumar V, Waldor MK, Schadt EE. 2012. Genome-wide mapping of methylated adenine residues in pathogenic *Escherichia coli* using single-molecule real-time sequencing. Nat Biotechnol 30:1232–1239. https://doi.org/10.1038/nbt.2432.
27. Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, Peluso P, Rank D, Baybayan P, Bettman B, Bibillo A, Bjornson K, Chaudhuri B, Christians F, Cicero R, Clark S, Dalal R, Dewinter A, Dixon J, Foquet M, Gaertner A, Hardenbol P, Heiner C, Hester K, Holden D, Kearns G, Kong X, Kuse R, Lacroix Y, Lin S, Lundquist P, Ma C, Marks P, Maxham M, Murphy D, Park I, Pham T, Phillips M, Roy J, Sebra R, Shen G, Sorenson J, Tomaney A, Travers K, Trulson M, Vieceli J, Wegener J, Wu D, Yang A, Zaccarin D, et al. 2009. Real-time DNA sequencing from single polymerase molecules. Science 323:133–138. https://doi.org/10.1126/science.1162986.
28. Flusberg BA, Webster DR, Lee JH, Travers KJ, Olivares EC, Clark TA, Korlach J, Turner SW. 2010. Direct detection of DNA methylation during single-molecule, real-time sequencing. Nat Methods 7:461–465. https://doi.org/10.1038/nmeth.1459.
29. Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, Hall KP, Evers DJ, Barnes CL, Bignell HR, Boutell JM, Bryant J, Carter RJ, Keira Cheetham R, Cox AJ, Ellis DJ, Flatbush MR, Gormley NA, Humphray SJ, Irving LJ, Karbelashvili MS, Kirk SM, Li H, Liu X, Maisinger KS, Murray LJ, Obradovic B, Ost T, Parkinson ML, Pratt MR, Rasolonjatovo IMJ, Reed MT, Rigatti R, Rodighiero C, Ross MT, Sabot A, Sankar SV, Scally A, Schroth GP, Smith ME, Smith VP, Spiridou A, Torrance PE, Tzonev SS, Vermaas EH, Walter K, Wu X, Zhang L, Alam MD, Anastasi C, et al. 2008. Accurate whole human genome sequencing using reversible terminator chemistry. Nature 456:53–59. https://doi.org/10.1038/nature07517.
30. Krebes J, Morgan RD, Bunk B, Spröer C, Luong K, Parusel R, Anton BP, König C, Josenhans C, Overmann J, Roberts RJ, Korlach J, Suerbaum S. 2014. The complex methylome of the human gastric pathogen *Helicobacter pylori*. Nucleic Acids Res 42:2415–2432. https://doi.org/10.1093/nar/gkt1201.
31. Furuta Y, Namba-Fukuyo H, Shibata TF, Nishiyama T, Shigenobu S, Suzuki Y, Sugano S, Hasebe M, Kobayashi I. 2014. Methylome diversification through changes in DNA methyltransferase sequence specificity. PLoS Genet 10:e1004272. https://doi.org/10.1371/journal.pgen.1004272.
32. Lluch-Senar M, Luong K, Lloréns-Rico V, Delgado J, Fang G, Spittle K, Clark TA, Schadt E, Turner SW, Korlach J, Serrano L. 2013. Comprehensive methylome characterization of *Mycoplasma genitalium* and *Mycoplasma pneumoniae* at single-base resolution. PLoS Genet 9:e1003191. https://doi.org/10.1371/journal.pgen.1003191.
33. Qiu W-G, Schutzer SE, Bruno JF, Attie O, Xu Y, Dunn JJ, Fraser CM, Casjens SR, Luft BJ. 2004. Genetic exchange and plasmid transfers in *Borrelia burgdorferi* sensu stricto revealed by three-way genome comparisons and multilocus sequence typing. Proc Natl Acad Sci U S A 101:14150–14155. https://doi.org/10.1073/pnas.0402745101.
34. Eggers CH, Kimmel BJ, Bono JL, Elias AF, Rosa P, Samuels DS. 2001. Transduction by φBB-1, a bacteriophage of *Borrelia burgdorferi*. J Bacteriol 183:4771–4778. https://doi.org/10.1128/JB.183.16.4771-4778.2001.
35. Haven J, Vargas LC, Mongodin EF, Xue V, Hernandez Y, Pagan P, Fraser-Liggett CM, Schutzer SE, Luft BJ, Casjens SR, Qiu W-G. 2011. Pervasive recombination and sympatric genome diversification driven by frequency-dependent selection in Borrelia burgdorferi, the Lyme disease bacterium. Genetics 189:951–966. https://doi.org/10.1534/genetics.111.130773.
36. Rosa PA, Schwan T, Hogan D. 1992. Recombination between genes encoding major outer surface proteins A and B of Borrelia burgdorferi. Mol Microbiol 6:3031–3040. https://doi.org/10.1111/j.1365-2958.1992.tb01761.x.
37. Stevenson B, Casjens S, Rosa P. 1998. Evidence of past recombination events among the genes encoding the Erp antigens of *Borrelia burgdorferi*. Microbiology (Reading) 144:1869–1879. https://doi.org/10.1099/00221287-144-7-1869.
38. Livey I, Gibbs CP, Schuster R, Dorner F. 1995. Evidence for lateral transfer and recombination in OspC variation in Lyme disease Borrelia. Mol Microbiol 18:257–269. https://doi.org/10.1111/j.1365-2958.1995.mmi_18020257.x.
39. Casselli T, Tourand Y, Scheidegger A, Arnold WK, Proulx A, Stevenson B, Brissette CA. 2018. DNA methylation by restriction modification systems affects the global transcriptome profile in *Borrelia burgdorferi*. J Bacteriol 200:e00395-18. https://doi.org/10.1128/JB.00395-18.
40. Fraser CM, Casjens S, Huang WM, Sutton GG, Clayton R, Lathigra R, White O, Ketchum KA, Dodson R, Hickey EK, Gwinn M, Dougherty B, Tomb JF, Fleischmann RD, Richardson D, Peterson J, Kerlavage AR, Quackenbush J, Salzberg S, Hanson M, van Vugt R, Palmer N, Adams MD, Gocayne J, Weidman J, Utterback T, Watthey L, McDonald L, Artiach P, Bowman C, Garland S, Fuji C, Cotton MD, Horst K, Roberts K, Hatch B, Smith HO, Venter JC. 1997. Genomic sequence of a Lyme disease spirochaete, *Borrelia burgdorferi*. Nature 390:580–586. https://doi.org/10.1038/37551.
41. Elias AF, Stewart PE, Grimm D, Caimano MJ, Eggers CH, Tilly K. 2002. Clonal polymorphism of *Borrelia burgdorferi* strain B31 MI: implications for mutagenesis in an infectious strain background. Infect Immun 70:2139–2150. https://doi.org/10.1128/iai.70.4.2139-2150.2002.
42. Dulebohn DP, Bestor A, Rosa PA. 2013. *Borrelia burgdorferi* linear plasmid 28-3 confers a selective advantage in an experimental mouse-tick infection model. Infect Immun 81:2986–2996. https://doi.org/10.1128/IAI.00219-13.
43. Jewett MW, Byram R, Bestor A, Tilly K, Lawrence K, Burtnick MN, Gherardini F, Rosa PA. 2007. Genetic basis for retention of a critical virulence plasmid of *Borrelia burgdorferi*. Mol Microbiol 66:975–990. https://doi.org/10.1111/j.1365-2958.2007.05969.x.
44. Barthold SW, Moody KD, Terwilliger GA, Duray PH, Jacoby RO, Steere AC. 1988. Experimental Lyme arthritis in rats infected with *Borrelia burgdorferi*. J Infect Dis 157:842–846. https://doi.org/10.1093/infdis/157.4.842.

45. Schwan TG, Schrumpf ME, Karstens RH, Clover JR, Wong J, Daugherty M, Struthers M, Rosa PA. 1993. Distribution and molecular analysis of Lyme disease spirochetes, *Borrelia burgdorferi*, isolated from ticks throughout California. J Clin Microbiol 31:3096–3108. https://doi.org/10.1128/JCM.31.12.3096-3108.1993.
46. Margolis N, Rosa PA. 1993. Regulation of expression of major outer surface proteins in *Borrelia burgdorferi*. Infect Immun 61:2207–2210. https://doi.org/10.1128/IAI.61.5.2207-2210.1993.
47. Preac-Mursic V, Wilske B, Schierz G. 1986. European *Borrelia burgdorferi* isolated from humans and ticks culture conditions and antibiotic susceptibility. Zentralbl Bakteriol Mikrobiol Hyg A 263:112–118. https://doi.org/10.1016/S0176-6724(86)80110-9.
48. Canica MM, Nato F, du Merle L, Mazie JC, Baranton G, Postic D. 1993. Monoclonal antibodies for identification of *Borrelia afzelii* sp. nov. associated with late cutaneous manifestations of Lyme borreliosis. Scand J Infect Dis 25:441–448. https://doi.org/10.3109/00365549309008525.
49. Casjens SR, Mongodin EF, Qiu D, Luft BJ, Schutzer SE, Gilcrease EB, Huang WM, Vujadinovic M, Aron JK, Vargas LC, Freeman S, Radune D, Weidman JF, Dimitrov GI, Khouri HM, Sosa JE, Halpin RA, Dunn JJ, Fraser CM. 2012. Genome stability of Lyme disease spirochetes: comparative genomics of *Borrelia burgdorferi* plasmids. PLoS One 7:e33280. https://doi.org/10.1371/journal.pone.0033280.
50. Grimm D, Elias AF, Tilly K, Rosa PA. 2003. Plasmid stability during in vitro propagation of Borrelia burgdorferi assessed at a clonal level. Infect Immun 71:3138–3145. https://doi.org/10.1128/iai.71.6.3138-3145.2003.
51. Roberts RJ, Vincze T, Posfai J, Macelis D. 2010. REBASE—a database for DNA restriction and modification: enzymes, genes and genomes. Nucleic Acids Res 38:D234–D236. https://doi.org/10.1093/nar/gkp874.
52. McLeay RC, Bailey TL. 2010. Motif enrichment analysis: a unified framework and an evaluation on ChIP data. BMC Bioinformatics 11:165. https://doi.org/10.1186/1471-2105-11-165.
53. Jorge DM, Mills RE, Lauring AS. 2015. CodonShuffle: a tool for generating and analyzing synonymously mutated sequences. Virus Evol 1:vev012. https://doi.org/10.1093/ve/vev012.
54. Barbour AG, Travinsky B. 2010. Evolution and distribution of the ospC gene, a transferable serotype determinant of Borrelia burgdorferi. mBio 1:e00153-10. https://doi.org/10.1128/mBio.00153-10.
55. Marconi RT, Samuels DS, Landry RK, Garon CF. 1994. Analysis of the distribution and molecular heterogeneity of the ospD gene among the Lyme disease spirochetes: evidence for lateral gene exchange. J Bacteriol 176:4572–4582. https://doi.org/10.1128/jb.176.15.4572-4582.1994.
56. Stevenson B, Miller JC. 2003. Intra- and interbacterial genetic exchange of Lyme disease spirochete erp genes generates sequence identity amidst diversity. J Mol Evol 57:309–324. https://doi.org/10.1007/s00239-003-2482-x.
57. Schulte-Spechtel U, Fingerle V, Goettner G, Rogge S, Wilske B. 2006. Molecular analysis of decorin-binding protein A (DbpA) reveals five major groups among European Borrelia burgdorferi sensu lato strains with impact for the development of serological assays and indicates lateral gene transfer of the dbpA gene. Int J Med Microbiol 296(Suppl 40):250–266. https://doi.org/10.1016/j.ijmm.2006.01.006.
58. Yang XF, Alani SM, Norgard MV. 2003. The response regulator Rrp2 is essential for the expression of major membrane lipoproteins in *Borrelia burgdorferi*. Proc Natl Acad Sci U S A 100:11001–11006. https://doi.org/10.1073/pnas.1834315100.
59. Caimano MJ, Iyer R, Eggers CH, Gonzalez C, Morton EA, Gilbert MA, Schwartz I, Radolf JD. 2007. Analysis of the RpoS regulon in *Borrelia burgdorferi* in response to mammalian host signals provides insight into RpoS function during the enzootic cycle. Mol Microbiol 65:1193–1217. https://doi.org/10.1111/j.1365-2958.2007.05860.x.
60. Caimano MJ, Eggers CH, Hazlett KRO, Radolf JD. 2004. RpoS is not central to the general stress response in *Borrelia burgdorferi* but does control expression of one or more essential virulence determinants. Infect Immun 72:6433–6445. https://doi.org/10.1128/IAI.72.11.6433-6445.2004.
61. Caimano MJ, Eggers CH, Gonzalez CA, Radolf JD. 2005. Alternate sigma factor RpoS is required for the in vivo-specific repression of *Borrelia burgdorferi* plasmid lp54-borne *ospA* and *lp6.6* genes. J Bacteriol 187:7845–7852. https://doi.org/10.1128/JB.187.22.7845-7852.2005.
62. Fisher MA, Grimm D, Henion AK, Elias AF, Stewart PE, Rosa PA, Gherardini FC. 2005. *Borrelia burgdorferi* σ(54) is required for mammalian infection and vector transmission but not for tick colonization. Proc Natl Acad Sci U S A 102:5162–5167. https://doi.org/10.1073/pnas.0408536102.
63. Brooks CS, Hefty PS, Jolliff SE, Akins DR. 2003. Global analysis of *Borrelia burgdorferi* genes regulated by mammalian host-specific signals. Infect Immun 71:3371–3383. https://doi.org/10.1128/iai.71.6.3371-3383.2003.
64. Hübner A, Yang X, Nolen DM, Popova TG, Cabello FC, Norgard MV. 2001. Expression of *Borrelia burgdorferi* OspC and DpbA is controlled by a RpoN-RpoS regulatory pathway. Proc Natl Acad Sci U S A 98:12724–12729. https://doi.org/10.1073/pnas.231442498.
65. Chan K, Alter L, Barthold SW, Parveen N. 2015. Disruption of *bbe02* by insertion of a luciferase gene increases transformation efficiency of *Borrelia burgdorferi* and allows live imaging in Lyme disease susceptible C3H mice. PLoS One 10:e0129532. https://doi.org/10.1371/journal.pone.0129532.
66. Bontemps-Gallo S, Lawrence KA, Richards CL, Gherardini FC. 2018. Genomic and phenotypic characterization of *Borrelia afzelii* BO23 and *Borrelia garinii* CIP 103362. PLoS One 13:e0199641. https://doi.org/10.1371/journal.pone.0199641.
67. Stewart PE, Rosa PA. 2017. Physiologic and genetic factors influencing the zoonotic cycle of *Borrelia burgdorferi*, p 63–82. *In* Adler B (ed), Spirochete biology: the post genomic era, vol 415. Springer, Cham, Switzerland.
68. Bohaliga GAR, Johnson WC, Taus NS, Hussein HE, Bastos RG, Suarez CE, O'Connor R, Ueti MW. 2018. Identification of a putative methyltransferase gene of *Babesia bigemina* as a novel molecular biomarker uniquely expressed in parasite tick stages. Parasit Vectors 11:480. https://doi.org/10.1186/s13071-018-3052-9.
69. Eggers CH, Gray CM, Preisig AM, Glenn DM, Pereira J, Ayers RW, Alshahrani M, Acabbo C, Becker MR, Bruenn KN, Cheung T, Jendras TM, Shepley AB, Moeller JT. 2016. Phage-mediated horizontal gene transfer of both prophage and heterologous DNA by ϕBB-1, a bacteriophage of *Borrelia burgdorferi*. FEMS Pathog Dis 74:ftw107. https://doi.org/10.1093/femspd/ftw107.
70. Ouyang Z, Deka RK, Norgard MV. 2011. BosR (BB0647) controls the RpoN-RpoS regulatory pathway and virulence expression in *Borrelia burgdorferi* by a novel DNA-binding mechanism. PLoS Pathog 7:e1001272. https://doi.org/10.1371/journal.ppat.1001272.
71. Ouyang Z, Zhou J, Brautigam CA, Deka R, Norgard MV. 2014. Identification of a core sequence for the binding of BosR to the *rpoS* promoter region in *Borrelia burgdorferi*. Microbiology (Reading) 160:851–862. https://doi.org/10.1099/mic.0.075655-0.
72. Riley SP, Bykowski T, Cooley AE, Burns LH, Babb K, Brissette CA, Bowman A, Rotondi M, Miller MC, DeMoll E, Lim K, Fried MG, Stevenson B. 2009. *Borrelia burgdorferi* EbfC defines a newly-identified widespread family of bacterial DNA-binding proteins. Nucleic Acids Res 37:1973–1983. https://doi.org/10.1093/nar/gkp027.
73. Chenail AM, Jutras BL, Adams CA, Burns LH, Bowman A, Verma A, Stevenson B. 2012. Borrelia burgdorferi cp32 BpaB modulates expression of the prophage NucP nuclease and SsbP single-stranded DNA-binding protein. J Bacteriol 194:4570–4578. https://doi.org/10.1128/JB.00661-12.
74. Madeira F, Park YM, Lee J, Buso N, Gur T, Madhusoodanan N, Basutkar P, Tivey ARN, Potter SC, Finn RD, Lopez R. 2019. The EMBL-EBI search and sequence analysis tools APIs in 2019. Nucleic Acids Res 47:W636–W641. https://doi.org/10.1093/nar/gkz268.
75. Stothard P. 2000. The Sequence Manipulation Suite: JavaScript programs for analyzing and formatting protein and DNA sequences. Biotechniques 28:1102–1104. https://doi.org/10.2144/00286ir01.
76. Price MN, Dehal PS, Arkin AP. 2010. FastTree 2—approximately maximum-likelihood trees for large alignments. PLoS One 5:e9490. https://doi.org/10.1371/journal.pone.0009490.
77. Jones DT, Taylor WR, Thornton JM. 1992. The rapid generation of mutation data matrices from protein sequences. Comput Appl Biosci 8:275–282. https://doi.org/10.1093/bioinformatics/8.3.275.
78. Stamatakis A. 2006. Phylogenetic models of rate heterogeneity: a high performance computing perspective, p 8. *In* Proceedings of 20th IEEE International Parallel & Distributed Processing Symposium. IEEE, Piscataway, NJ.
79. Shimodaira H, Hasegawa M. 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol Biol Evol 16:1114–1116. https://doi.org/10.1093/oxfordjournals.molbev.a026201.
80. Letunic I, Bork P. 2019. Interactive Tree of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res 47:W256–W259. https://doi.org/10.1093/nar/gkz239.
81. Barbour AG. 1984. Isolation and cultivation of Lyme disease spirochetes. Yale J Biol Med 57:521–525.

82. Samuels DS, Mach KE, Garon CF. 1994. Genetic transformation of the Lyme disease agent *Borrelia burgdorferi* with coumarin-resistant *gyrB*. J Bacteriol 176:6045–6049. https://doi.org/10.1128/jb.176.19.6045-6049.1994.
83. Norris SJ, Howell JK, Odeh EA, Lin T, Gao L, Edmondson DG. 2011. High-throughput plasmid content analysis of *Borrelia burgdorferi* B31 by using Luminex multiplex technology. Appl Environ Microbiol 77:1483–1492. https://doi.org/10.1128/AEM.01877-10.
84. Bailey TL, Bodén M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS. 2009. MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res 37:W202–W208. https://doi.org/10.1093/nar/gkp335.
85. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. 2009. Circos: an information aesthetic for comparative genomics. Genome Res 19:1639–1645. https://doi.org/10.1101/gr.092759.109.
86. Wickham H. 2016. ggplot2: elegant graphics for data analysis, 1st ed, vol 8. Springer-Verlag, New York, NY.
87. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359. https://doi.org/10.1038/nmeth.1923.
88. Liao Y, Smyth GK, Shi W. 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30:923–930. https://doi.org/10.1093/bioinformatics/btt656.
89. Liao Y, Smyth GK, Shi W. 2019. The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. Nucleic Acids Res 47:e47. https://doi.org/10.1093/nar/gkz114.
90. McCarthy DJ, Chen Y, Smyth GK. 2012. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. Nucleic Acids Res 40:4288–4297. https://doi.org/10.1093/nar/gks042.
91. Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15:550. https://doi.org/10.1186/s13059-014-0550-8.
92. Savage CR, Jutras BL, Bestor A, Tilly K, Rosa P, Tourand Y, Stewart PE, Brissette CA, Stevenson B. 2018. *Borrelia burgdorferi* SpoVG DNA- and RNA-binding protein modulates the physiology of the Lyme disease spirochete. J Bacteriol 200:e00033-18. https://doi.org/10.1128/JB.00033-18.
93. Jutras BL, Savage CR, Arnold WK, Lethbridge KG, Carroll DW, Tilly K, Bestor A, Zhu H, Seshu J, Zückert WR, Stewart PE, Rosa PA, Brissette CA, Stevenson B. 2019. The Lyme disease spirochete's BpuR DNA/RNA-binding protein is differentially expressed during the mammal-tick infectious cycle, which affects translation of the SodA superoxide dismutase. Mol Microbiol 112:973–991. https://doi.org/10.1111/mmi.14336.