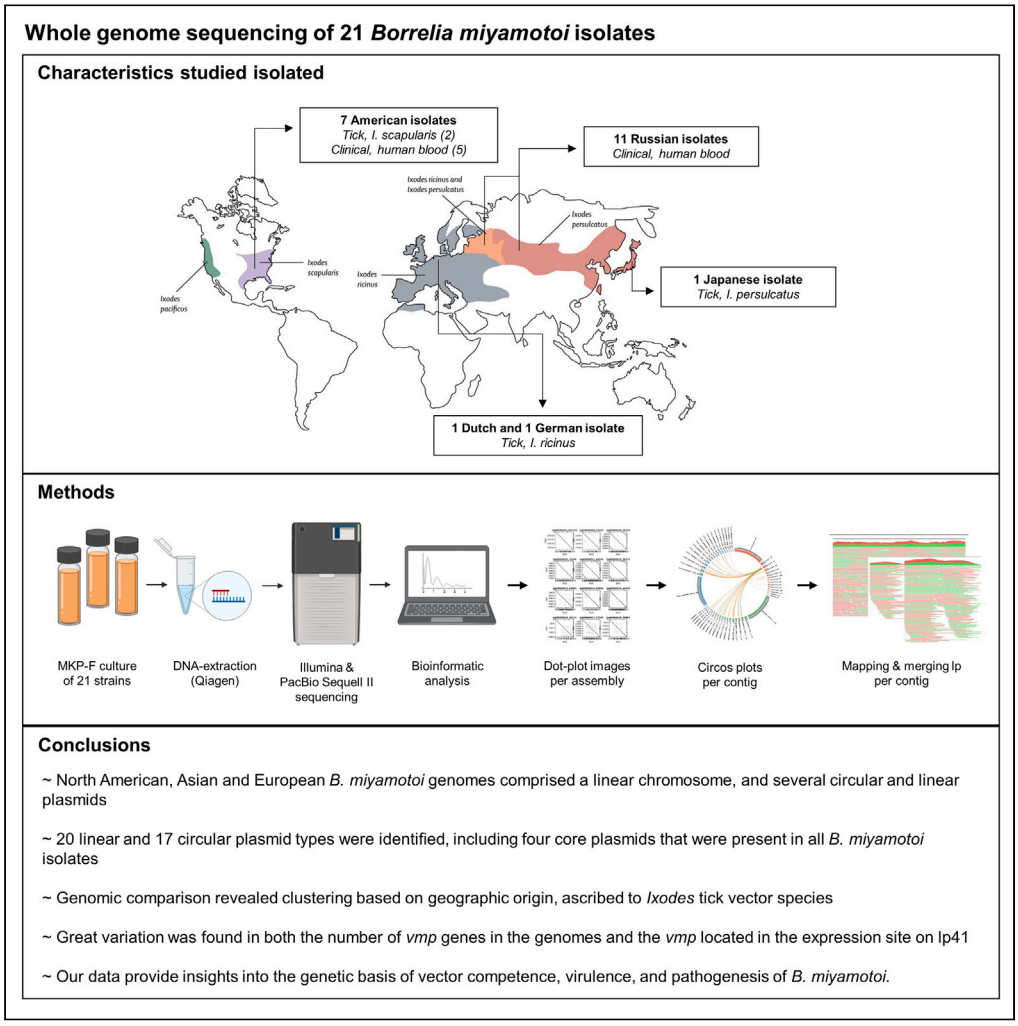
iScience

CellPress
OPEN ACCESS

Article

Combining short- and long-read sequencing unveils geographically structured diversity in *Borrelia miyamotoi*



Dieuwertje Hoornstra, Konstantin V. Kuleshov, Volker Fingerle, ..., Alexander E. Platonov, Gabriele Margos, Joppe W. Hovius

pandora@amsterdamumc.nl

Highlights

American, Asian, and European *B. miyamotoi* genomes comprised one linear chromosome

The isolates contained 20 linear and 17 circular plasmid types, with 4 core plasmids

Genomes clustered based on geographic origin ascribed to *Ixodes* vector species

The number of *vmps* per genome and the *vmp* in the lp41 expression site varied greatly

Hoornstra et al., iScience 27, 110616
September 20, 2024
https://doi.org/10.1016/j.isci.2024.110616

iScience

CellPress
OPEN ACCESS

Article

Combining short- and long-read sequencing unveils geographically structured diversity in *Borrelia miyamotoi*

Dieuwertje Hoornstra,[1,11] Konstantin V. Kuleshov,[2,11] Volker Fingerle,[3,4] Sabrina Hepner,[3,4] Alex Wagemakers,[1] Christina Strube,[5] Santiago Castillo-Ramírez,[6] Linda K. Bockenstedt,[7] Sam R. Telford 3rd,[8] Hein Sprong,[9] Alexander E. Platonov,[2] Gabriele Margos,[3,4] and Joppe W. Hovius[1,10,12,*]

SUMMARY

***Borrelia miyamotoi* is an emerging *Ixodes* tick-borne human pathogen in the Northern hemisphere. The aim of the current study was to compare whole genome sequences of *B. miyamotoi* isolates from different continents. Using a combination of Illumina and PacBio platforms and a novel genome assembly and plasmid typing pipeline, we reveal that the 21 sequenced *B. miyamotoi* isolates and publically available *B. miyamotoi* genomes from North America, Asia, and Europe form genetically distinct populations and cluster according to their geographical origin, where distinct *Ixodes* species are endemic. We identified 20 linear and 17 circular plasmid types and the presence of specific plasmids for isolates originating from different continents. Linear plasmids lp12, lp23, lp41, and lp72 were core plasmids found in all isolates, with lp41 consistently containing the *vmp* expression site. Our data provide insights into the genetic basis of vector competence, virulence, and pathogenesis of *B. miyamotoi*.**

INTRODUCTION

The genus *Borrelia* comprises multiple species, which can be divided into the two groups *Borrelia burgdorferi* sensu lato (s.l.) and relapsing fever *Borrelia* (RFB).[1,2] The *B. burgdorferi* s.l. group includes over 20 species, all of which are transmitted by hard-bodied *Ixodes* ticks. The RFB group includes over a dozen well-described species, most of which are transmitted by soft ticks, aside from *Borrelia recurrentis* (human body lice) and *Borrelia theileri, Borrelia lonestari* and *Borrelia miyamotoi* (hard-bodied ticks).[3] *Borrelia miyamotoi* clusters with other RFB species based on 16S ribosomal RNA and *flaB* gene sequences.

Similar to *B. burgdorferi* s.l., *B. miyamotoi* is transmitted by *Ixodes* ticks. In North America, *B. miyamotoi* is transmitted by *Ixodes scapularis* and *Ixodes pacificus*, in Asia by *Ixodes persulcatus*, and in Europe by *Ixodes ricinus*.[4–7] Genetic differences within Russian and American genotypes appear to be limited.[8,9] However, based on multilocus sequence typing, the Japanese isolates from *Ixodes ovatus* could not be classified as Asian, European or North American, potentially forming a new group.[10,11] Additionally, two North American strains from *I. scapularis* and *I. pacificus* differed in their 16S rRNA sequences.[12] Therefore, the geographical clustering appears to be vector-related.[13]

As in all other *Borrelia* genomes, the *B. miyamotoi* genome consists of a single large linear chromosome and circular and linear plasmids.[14] The North American *B. miyamotoi* tick isolate LB-2001 was found to contain RFB-specific genes, including *glpQ* on the linear chromosome.[7,14] Currently, draft genome sequences reported five North American, 12 Russian, one Asian and three European clinical or tick isolates of *B. miyamotoi*.[5,14–17] Due to high frequency of repetitive and paralogous sequences, so far only one Asian strain, Izh-4,[18] and one North American strain, LB-2001,[19] have been assembled as a whole genome. Therefore, knowledge on plasmid diversity of other *B. miyamotoi* isolates is still incomplete, although plasmid encoded genes are essential for pathogenicity and completion of the transmission cycle.[16–18] The Izh-4 isolate contained 850 open-reading frames on the conserved chromosome (906 kb), 428 on 12 linear plasmids (6-72 kb), and 82 on the two circular plasmids (30 kb).

[1]Center for Experimental and Molecular Medicine, Amsterdam University Medical Centers, Location Academic Medical Center, Amsterdam, the Netherlands
[2]Central Research Institute of Epidemiology, Moscow, Russia
[3]Bavarian Health and Food Safety Authority, Oberschleissheim, Germany
[4]German National Reference Centre for Borrelia, Oberschleissheim, Germany
[5]University of Veterinary Medicine, Hannover, Germany
[6]Center for Genome Sciences, National University of México, Cuernavaca, México
[7]Yale School of Medicine, New Haven, CT, USA
[8]Tufts Cummings School of Veterinary Medicine, Grafton, MA, USA
[9]National Institute for Public Health and the Environment, Bilthoven, the Netherlands
[10]Amsterdam Institute for Immunology & Infectious Diseases, Amsterdam University Medical Centers, location Academic Medical Center, Amsterdam, the Netherlands
[11]These authors contributed equally
[12]Lead contact
*Correspondence: pandora@amsterdamumc.nl
https://doi.org/10.1016/j.isci.2024.110616

Table 1. Characteristics of *Borrelia miyamotoi* strains

Name of stain	Geographical origin	Isolation source	Year	Isolation method	Passage	GenBank
LB-2001	USA, Connecticut	*I. scapularis*	2001	SCID acquisition and BSK isolation	X3	SAMN32260477
410	USA, Rhode Island	*H. sapiens*	2016	SCID inoculation and MKP-F isolation	1	SAMN32260473
483	USA, Massachusetts	*I. scapularis*	2004	SCID acquisition and MKP-F isolation	1	SAMN32260472
1938	USA, Rhode Island	*H. sapiens*	2016	SCID inoculation and MKP-F isolation	2	SAMN32260474
1995	USA, Rhode Island	*H. sapiens*	2017	SCID inoculation and MKP-F isolation	2	SAMN32260475
1999	USA, Rhode Island	*H. sapiens*	2016	SCID inoculation and MKP-F isolation	2	SAMN32260476
CT14D4	USA, Connecticut	*H. sapiens*	2014	Rag1$^{-/-}$ inoculation and BSK isolation	5	SAMN32260480
HT-31	Japan	*I. persulcatus*	1992	SCID acquisition and BSK isolation	X3	SAMN32260478
Izh-4	Russia, Izhevsk	*H. sapiens*	2016	Heparinized plasma MKP-F isolation	5	SAMN07572561
Izh-5	Russia, Izhevsk	*H. sapiens*	2016	Heparinized plasma MKP-F isolation	5	SAMN07572562
Izh-14	Russia, Izhevsk	*H. sapiens*	2016	Heparinized plasma MKP-F isolation	5	SAMN07572563
Izh-16	Russia, Izhevsk	*H. sapiens*	2016	Heparinized plasma MKP-F isolation	5	SAMN07572564
Yekat-1	Russia, Yekaterinburg	*H. sapiens*	2016	Heparinized plasma MKP-F isolation	4	SAMN07572565
Yekat-6	Russia, Yekaterinburg	*H. sapiens*	2016	Heparinized plasma MKP-F isolation	5	SAMN07572566
Yekat-17	Russia, Yekaterinburg	*H. sapiens*	2017	Heparinized plasma MKP-F isolation	1	SAMN10979508
Yekat-18	Russia, Yekaterinburg	*H. sapiens*	2018	Heparinized plasma MKP-F isolation	1	SAMN10979507
Yekat-19	Russia, Yekaterinburg	*H. sapiens*	2018	Heparinized plasma MKP-F isolation	1	SAMN10979509
Yekat-21	Russia, Yekaterinburg	*H. sapiens*	2018	Heparinized plasma MKP-F isolation	1	SAMN10979510
Yekat-76	Russia, Yekaterinburg	*H. sapiens*	2018	Heparinized plasma MKP-F isolation	1	SAMN10979512
ZStrull14-3	Germany	*I. ricinus*	2019	Larvae MKP-F isolation	3	SAMN32260479
NL-IR-1	Netherlands	*I. ricinus*	2018	Egg mass MKP-F isolation	4	SAMN12826994

Characteristics of *Borrelia miyamotoi* isolates used in this study, including the description of the strains by name, the origin of isolation (geographical location, source, year), the method of isolation and the GenBank BioSample number. The geographical location of strain isolation is provided for the USA per state, for Europe per country, and for Asian Russia per city province.

BSK, Barbour-Stoenner-Kelly; *H., Homo*; *I., Ixodes*; MKP-F, modified Kelly-Pettenkofer medium supplemented with heat-inactivated serum; Rag, rag1 deficient mice; SCID, Severe combined immunodeficient mice; USA, United States of America; X, unknown number of passages prior to obtaining the isolate from collaborator.

Linear plasmid 41 (lp41), the main virulence plasmid, has been shown to contain the expression site for a variable major protein (Vmp). Vmps comprise variable small proteins (Vsps) and variable large proteins (Vlps), grouped into α-, β-, γ-, and δ-subgroup families. Switching *vmp* gene expression is an established immune evasion mechanism of RFB responsible for the recurrent episodes of spirochetemia and the pattern of relapsing fever in infected humans.[18] *Borrelia miyamotoi* contains a Vmp-system similar to *Borrelia hermsii*.[20,21] Approximately 60 distinctive *vmp* gene cassettes are located throughout the genome, and a cassette is only expressed when it is translocated into the active expression locus.[22] The *vsp* and *vlp* genes appear to be located on linear plasmids.[18,20,21] Comparison of lp41 between two Asian and two North American *B. miyamotoi* isolates revealed that *vmp* expression sites were highly comparable. However, the number and order of *vmp*s and the particular gene in the expression site differed among strains.[18,21,23]

In this study, we used a novel genome assembly and plasmid typing pipeline to describe and compare the assembled genomes of 21 *B miyamotoi* isolates from ticks and patients from different continents. In addition, we included 21 publicly available genomes to create, to our knowledge, the most extensive dataset. This set has 42 isolates from seven different countries and covers a time span of 1991–2023. Our genomic data provide in-depth insights into the genetic basis of vector competence, virulence, and pathogenesis and may reveal genetic traits in *B. miyamotoi* isolates explaining the differences in disease prevalence throughout the Northern hemisphere.[24]

RESULTS

Genome sequences and assembly metrics

Twenty-one *B. miyamotoi* strains were included and categorized by geographical origin: seven strains from the USA,[14,25] 12 from Asia (one from Japan and 11 from Russia),[20,26,27] and two from Europe (Germany and The Netherlands)[5,28] (Table 1). Genome assembly of these isolates, using a combination of Illumina and PacBio platforms, resulted in 493 contigs. Validation of the assemblies and subsequent analyses of the contigs assembled from HiFi reads designated 344 contigs as borrelial chromosome or plasmids. Of these contigs, 21 met the criteria of chromosome, 228 of linear and 95 of circular plasmids. The remaining 149 linear contigs were excluded, as they were classified as repeats and bubbles from longer sequences.

The average sequencing depth per genome varied from 192-1600x in Illumina and 58-334x in PacBio sequencing (Table S1). Median assembly completeness was 99.2% (95%CI 99.1–99.4%), and the median accuracy was 58.4 (95%CI 55.0–61.7), indicating a high-quality of nucleotide sequence assembly. To ensure the 21 sequences belong to the *B. miyamotoi* genus an average nucleotide identity (ANI) analysis was performed (Table S2). Nevertheless, it was not always possible to assemble the 5′- and 3′-ends of the linear plasmids with a high end-homology, typically representing clusters of *vmp* genes which were nearly identical between different linear plasmids within the same genome.[21]

In general, the genomes showed a similar structure. Nevertheless, differences were found in total assembly length, due to the variable number of linear and circular plasmids and of protein-coding sequences (CDSs) in the genomes from different origins. North American, Asian, and European strains had a mean total assembly length of 1.35, 1.47, and 1.52MB, respectively, with North American strains being significantly smaller ($p = 0.00012$) than their Asian counterparts. There were fewer linear and circular plasmids in North American ($n = 12–14$) than in Asian ($n = 15–20$, $p = 0.0011$) or European strains ($n = 16–17$, $p = 0.056$). In addition, the North American strains were characterized by fewer CDSs, with an average of 1,329 CDSs, compared to 1,463 CDSs in Asian ($p = 0.0013$) and 1,525 CDSs in European strains.

Linear chromosome comparison

Throughout the comparison of linear chromosomes, no long insertions, deletions, duplications, or translocations were detected indicating a highly syntenic structural organization. However, minor differences were found in variable number of tandem repeats (VNTR), single nucleotide polymorphisms (SNPs), and small indels. The majority of these polymorphisms were base substitutions. Comparison within the geographical regions resulted in small average genetic differences of 15 (95%CI 11.0–20.0) in North American, 22 (95%CI 16.0–26.0) in Asian, and 52 in European strains. Whereas, comparison between the geographical regions resulted in large average genetic differences of 17,667 (95%CI 17,666-17,669) between Asian and North American, 22,052 (95%CI 22,048-22,056) between Asian and European, and 24,699 (95%CI 24,696-24,703) between North American and European strains.

The ANI of the linear chromosomes displayed an overall of 99.99%, and a group specific ANI of 99.99% (range 99.00–100.00%), 99.99% (range 99.98–100.00%), and 99.99% for North American, Asian, and European strains, respectively (Table S2). Group comparison showed the largest difference between the North American and European strains with an ANI value of 96.89% (range 96.87–96.90%). Interspecies comparison showed an average ANI of 84.15% (range 80.78–86.76%) comparing the sequenced *B. miyamotoi* strains to RFB (15 genospecies), while comparison to *B. burgdorferi* sensu lato (ten genospecies) showed an average ANI of 75.38% (range 75.18–75.57%).

Inter- and intraspecies phylogeny by core genes

For an interspecies phylogenetic analysis, we aligned 209 chromosomal core genes from 83 *Borrelia* genomes (Table S3). The phylogenetic tree demonstrated evident clustering of our 21 *B miyamotoi* strains within the RFB clade (Figure S1). The *B. miyamotoi* clade was divided distinctly into three major clades directly corresponding to geographic origin. The European clade formed a sister-group relationship with a clade containing the North American and the Asian clades. Hence, North American and Asian clades are more closely related to each other than to the European clade.

Subsequently, we performed an intraspecies phylogenetic analysis with a total of 42 high quality *B. miyamotoi* genomes and aligned 667 chromosomal core genes (Table S4). The intraspecies *B. miyamotoi* phylogeny corresponded with the interspecies *Borrelia* tree (Figure 1). Furthermore, it provided a more detailed subdivision within the geographical clades. Notably, this analysis demonstrated an evident clustering according to vector species. The North American clade contained one genetically distinct isolate from *I. pacificus* and eight more closely related strains, three from *I. scapularis* and five from patients from an *I. scapularis*-endemic area. The Asian clade contained one isolate from *I. pavlovskyi*, and 26 related strains, of which 14 were isolated from *I. persulcatus* and 12 from mammals (11 patients and 1 mouse) from an *I. persulcatus*-endemic area. The European clade comprised five strains from *I. ricinus* ticks.

Plasmid type identification and comparison

Plasmids of the 21 assembled sequenced genomes were typed by identifying the partition protein genes based on paralogous gene family (PF) classification, including PF32 and PF57/62. These gene families are commonly used for classification of plasmids, as they are generally present on plasmids.[29] Due to similarities to proteins of known function in other bacteria, they have been suggested to encode proteins involved in plasmid DNA partitioning and replication.[29,30] Reconstruction of the phylogenetic tree based on CDSs related to PF32 (Figure S2) and PF57/62 (Figure S3) identified a total of 36 plasmid types, of which 27 could be classified by PF32 genes. Because PF57/62 is present in all plasmids, and phylogenetic analyses of PF32 and PF57/62 or PF57/62 alone provided similar clustering, we suggest *B. miyamotoi* plasmid typing by PF57/62 only (Table S5). Seventeen plasmid types were identified among the circular plasmids and 20 among the linear plasmids (Table 2). Most types were found in at least two separate genomes, while cp30-9 was found only in German ZStruIII14-9.

Copy number analysis of the PF32 and PF57/62 genes on the different plasmid types provided insight into plasmid origin and evolution (Figure 2; Table S6 and Data S1). To illustrate, two phylogenetically distant PF32 copies were found on lp66, lp70, and lp97, suggesting that these linear plasmids were likely the product of fusion of two different circular plasmids. This hypothesis was supported by phylogenetic grouping of these PF32 copies within the circular plasmid group (Figure S2). All circular plasmids were characterized by the presence of one copy of PF32, with the exception of Yekat-18, Yekat-76, and ZStruIII14-9. These contained two groups of genes responsible for replication (cp30-4+cp19, cp30-6+cp30-11, and cp30-5+cp24-1, respectively), suggesting another potential evolutionary plasmid fusion (Figure S2). Overall, 3 out of 344 plasmids were fused (0.9%, 95%CI 0.2–2.5%). The reverse situation, plasmid size reduction, is defined as a plasmid shorter than 30 kb in length. This situation was observed in the cp30 plasmid group with an average length of 30 kb. The 19 kb plasmid from

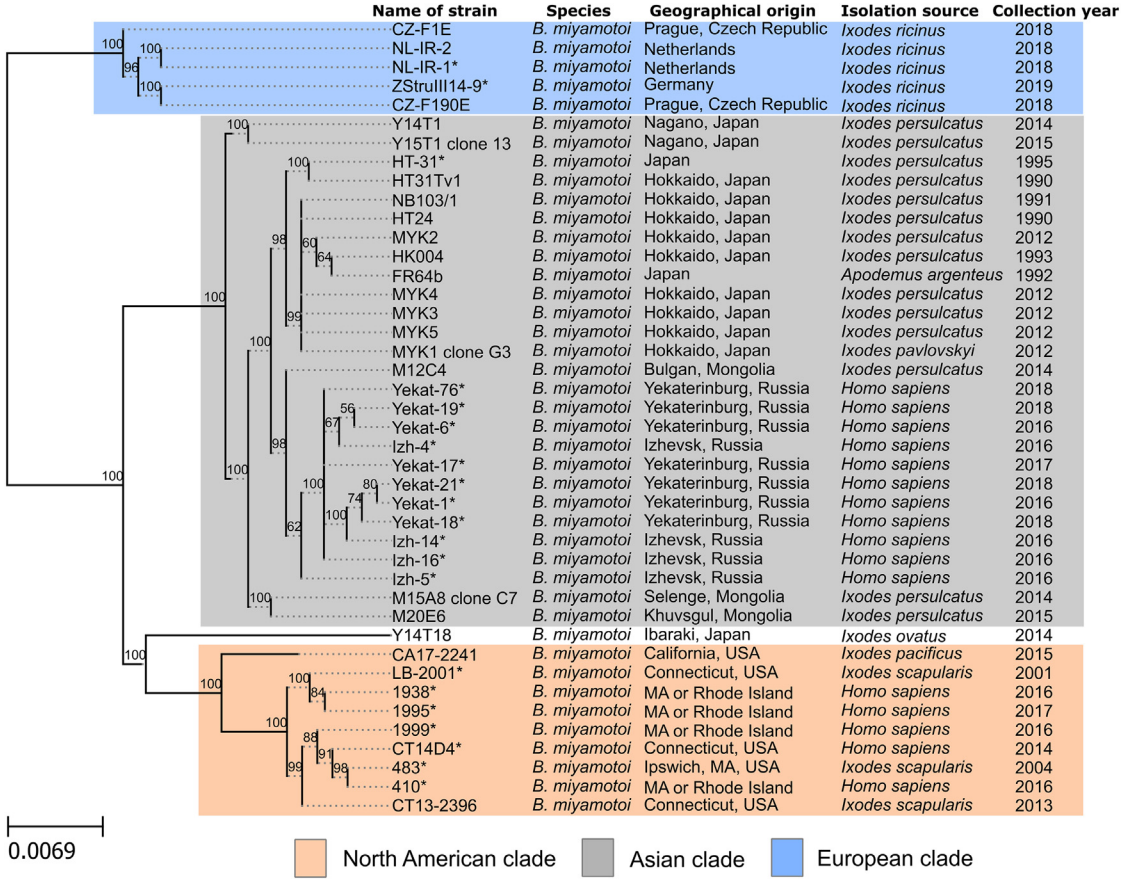


Figure 1. Phylogenetic tree of *Borrelia miyamotoi* species by core gene alignment

A phylogenetic tree of *Borrelia miyamotoi* strains based on the concatenated alignment of nucleotide sequences from 667 core genes located on the chromosome using Roary. A maximum likelihood tree was constructed by RAxML software using a nucleotide substitution model with a gamma distribution of variable positions (GTR + Γ). The resulting tree was rooted at the branch connecting the European clade to the rest of the tree. Scale bar indicates substitution rates. The figure includes the name of the strain, its origin concerning *Borrelia* species, geographical location, isolation source and year of collection. * Indicates the 21 *Borrelia miyamotoi* strains analyzed in this paper.

ZStruIII14-9 actually belonged to cp30-1, and the 15 kb plasmid from Yekat-76 belonged to cp30-4. Overall, 2 out of 344 plasmids were reduced (0.6%, 95%CI 0.1–2.1%).

Examining the number of plasmid types per strain revealed that each strain usually included only one copy of each plasmid type per genome. In contrast, lp20-1, lp24 and lp29-2 were found more than once in specific isolates (Table 2). Annotation of these repetitive linear plasmids revealed that their content varied due to variable gene clusters encoding Vmps. Therefore, these plasmids are thought to play a role in the storage and active transfer of *vmp* genes within the bacterial cell (Figure 2, Data S1). Thirty-one of 344 plasmids (9.0%, 95%CI 6.2–12.6%) were considered repetitive plasmids, which were observed in all Asian and European strains, and in one North American strain.

Core plasmids, identified in all strains, were lp12, lp23, lp41 and lp72. Comparative analysis of the entire plasmid set between strains showed that there are also group-specific plasmids for the North American, Asian and European strains (Table 2). Further analysis showed that the plasmid composition of North American strains was most stable. However, some variation was found due to the presence of lp20-2 in the 1999 strain and lp20-1-1, lp20-1-2 and lp20-2 in the CT14D4 strain. In European strains, linear plasmid composition was stable, with some variations in the cp30 group. Although the Asian strains had a conserved composition of linear plasmids, they were most diverse, as lp13, lp24-1, lp26 and lp29-1 were present in a subset of strains only. There was high variability and diversity in circular plasmids among Asian strains. The exception was cp19, which was similar in all Russian strains. Interestingly, the European and Asian strains carried common circular plasmids types (cp30-1, cp30-5, cp30-6, cp30-7, cp30-8, cp30-9, cp30-11, cp29-1), while the North American strains were characterized by a distinctive set.

Plasmid conservation

Thirty-seven plasmid types of the 344 plasmids were used to conduct a comparative analysis by pairwise alignment using the Pyani tool and calculating the Hadamard coefficient of similarity (Table 2). We distinguished groups of plasmids by Hadamard coefficient, which is inversely

Table 2. Spectrum of the chromosome and plasmids in *Borrelia miyamotoi* genomes determined by PF32 and PF57/62 typing

				North American							Asian												European		
PF typing	Plasmid	Hadamard coefficient (av.)	Hadamard coefficient (SD)	LB-2001	410	483	1938	1995	1999	CT14D4	HT-31	Izh-4	Izh-5	Izh-14	Izh-16	Yekat-1	Yekat- 6	Yekat-17	Yekat-18	Yekat-19	Yekat-21	Yekat-76	ZStrull 14-3	NL-IR-1	Total copies in dataset
PF32	chr	0.9920	0.0220	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	19
PF32	lp97	0.9998	0.0001	1	1	1	1	1	1	1															7
PF32	lp92	0.9902	0.0011																				1	1	0
PF32	lp72	0.9823	0.0799	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	19
PF32	lp70	0.9972	0.0053								1	1	1	1	1	1	1	1	1	1	1	1			12
PF32	lp66	0.9988	0.0000																				1	1	2
PF32	lp64	0.9835	0.0356								1	1	1	1	1	1	1	1	1	1	1	1			12
PF32	lp41	0.7741	0.1035	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	21
PF57/62	lp31	0.9413	0.0999	1	1	1	1	1	1	1															7
PF57/62	lp29-1	0.9251	0.1291											1		1			1		1				4
PF57/62	lp29-2	0.8118	0.2320								2	2	2	2	2	2	2	3	1	2	2	2	2	2	28
PF32	lp26	0.7444	0.2260	1	1	1	1	1	1	1	1			1									1	1	11
PF57/62	lp24-1	0.8622	0.1706								1	1	1	1	1		1	1	1	2	1	1			12
PF57/62	lp24-2	0.9996	0.0001																				1	1	2
PF32	lp23	0.6787	0.2004	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	21
PF57/62	lp20-1	1.1937	0.3279	1	1	1	1	1	1	2															8
PF57/62	lp20-2	0.9999	0.0000						1	1															2
PF57/62	lp18-1	0.9996	0.0002								1	1	1	1	1	1	1	1	1	1	1	1			12
PF57/62	lp18-2	0.9996	0.0004								1	1	1	1	1	1	1	1	1	1	1	1			12
PF32	lp12	0.9663	0.0287	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	21
PF57/62	lp13	1.0002	0.0013								1	1	1	1	1	1	1	1	1		1	1			11
PF32	cp30-1	0.7136	0.1598															1					19kb		2
PF32	cp30-2	0.9985	0.0008	1	1	1	1	1	1	1															7
PF32	cp30-3	0.9980	0.0008									1	1		1		1			1					5
PF32	cp30-4	0.8520	0.1606								1		1			1	1	1	cp30-4 + cp19	1	1	15kb			9
PF32	cp30-5	0.9790	0.0142								1	1	1	1		1	1		1		1		cp30-5 + cp24-1		9
PF32	cp30-6	0.8936	0.0806											1	1	1	1	1			1	cp30-6 + cp30-11		1	8
PF32	cp30-7	0.9194	0.0528								1							1					1	1	4
PF32	cp30-8	0.9208	0.0898										1	1	1	1		1	1				1		7
PF32	cp30-9	NA	NA																				1		1

(*Continued on next page*)

Table 2. *Continued*

				North American							Asian													European		
PF typing	Plasmid	Hadamard coefficient (av.)	Hadamard coefficient (SD)	LB-2001	410	483	1938	1995	1999	CT14D4	HT-31	Izh-4	Izh-5	Izh-14	Izh-16	Yekat-1	Yekat- 6	Yekat-17	Yekat-18	Yekat-19	Yekat-21	Yekat-76	ZStrull 14-3	NL-IR-1	Total copies in dataset
PF32	cp30-10	0.9993	0.0003	1	1	1	1	1	1	1															7
PF32	cp30-11	0.8530	0.0870									1	1		1		1	1				cp30-6 + cp30-11		1	7
PF32	cp29-1	0.9985	0.0012																1				1	1	3
PF32	cp29-2	0.8551	0.0517	1	1	1	1	1	1	1															7
PF32	cp24-1	NA	NA																				cp30-5 + cp24-1	1	2
PF32	cp24-2	0.9923	0.0232	1	1	1	1	1	1	1															7
PF32	cp19	0.9983	0.0019									1	1	1	1	1	1	1	cp30-4 + cp19	1	1	1			11
PF32	cp15	0.9959	0.0021																				1	1	2

Comparison of the spectrum of plasmids in *Borrelia miyamotoi* genomes determined by PF32 and PF57/62 typing. Per strain the number of replicons per chromosome or plasmid type are given. Of note, the some cells depict fusion plasmids and other cells size reduced plasmids in kb.

av., average; chr, chromosome; cp, circular plasmid; lp, linear plasmid; PF, partition protein genes based on paralogous gene family; SD, standard deviation.

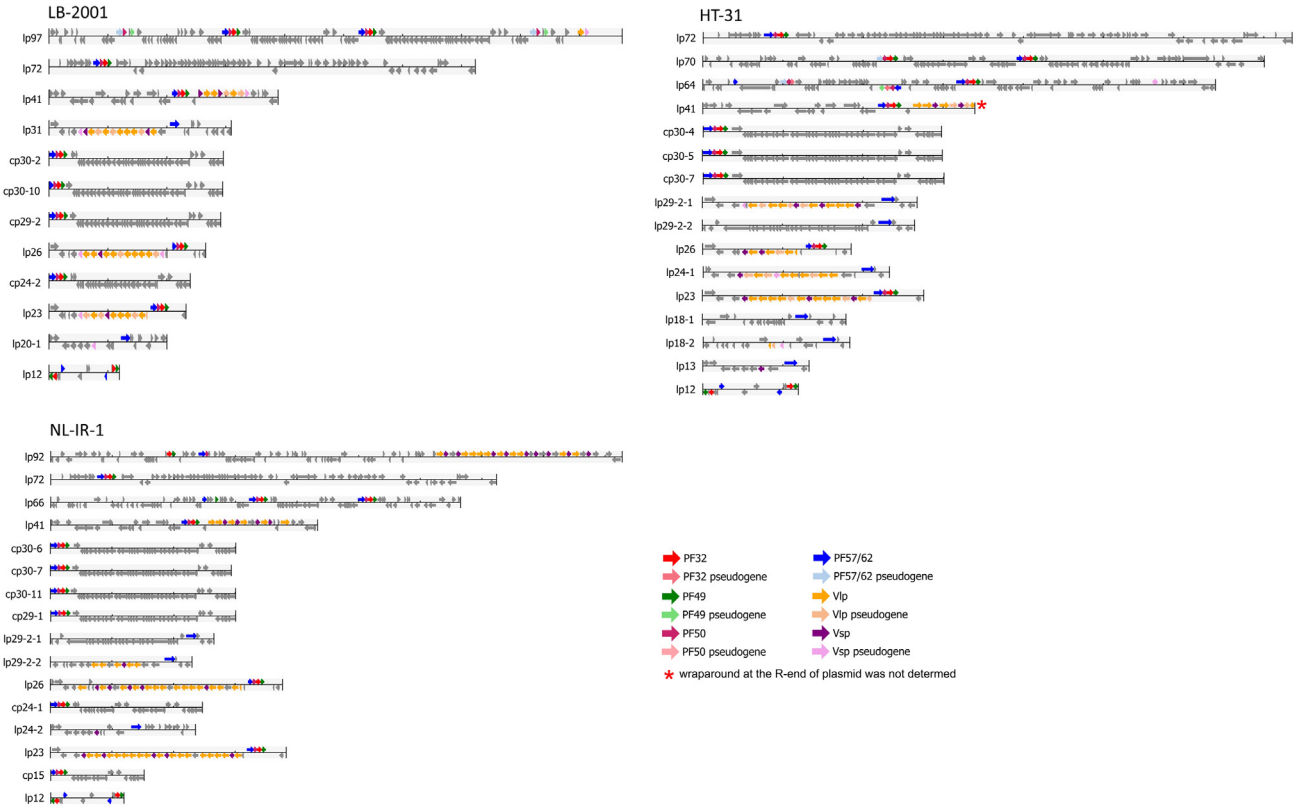


Figure 2. Plasmid diversity of three reference *Borrelia miyamotoi* genomes

A schematic representation of the plasmid diversity of the three reference *Borrelia miyamotoi* genomes. Identified are the genes and pseudogenes of PF32, 49, 50, 57/62, *vlp* and *vsp*. They are displayed in the order, amount and relative position on the linear and circular plasmids.

related to the level of plasmid type variation. The chromosome and most of the 20 linear plasmid types displayed a coefficient of >0.9, indicating a low level of variation and high levels of nucleotide sequence similarity and overlap. Exceptions included some linear plasmids due to the presence of variable clusters of *vmp* genes encoding Vmp proteins. Overall, the 17 circular plasmids showed lower coefficients than linear plasmid types, which were more variable due to differences in plasmid length.

Telomere identification

Identification of telomere sequences at both ends of a plasmid, corresponding to a hairpin structure, is the main validation of completeness in linear plasmids.[31,32] These regions were searched for manually in putative linear plasmids and compared among strains. The presence of a specific pattern (box 3), was observed in the chromosomes and in most linear plasmids with the consensus motif 'TWGTATA' on the 3′-end and the reverse sequence 'TATACWA' on the 5′-end, starting approximately from position 14 from both ends.[31,32] The linear plasmid types containing box 3 termini include lp13, lp18-1, lp18-2, lp20-1, lp20-2, lp23, lp24-1, lp29-1, lp29-2, lp31, lp64, lp66, lp70, lp72, lp92, and lp97, which were present in both Asian and European strains (Table S7). However, variation was found in the European lp41 3′ end, European lp24-2 both 3′- and 5′-ends, and the North American lp23, lp26 and lp41 5′-end, in which the putative site 'TTGTACA' was present. Comparative analysis of these terminal sequences on both chromosome and plasmid types showed that, as a rule, each of the geographical groups was characterized by a specific nucleotide sequence that included the conserved box 3.

It was not always possible to assemble the 5′- and 3′-ends of the linear plasmids. In 11 of the 498 terminal sequences criteria for the completeness of linear plasmids were not met, as the wrap arounds at the L- or E-end within the hairpin structures remained unidentified. This was observed at the right end of lp41 (n = 4) and the left ends of lp23 (n = 5), lp29 (n = 2), and lp24 (n = 1), in seven Asian and one North American strain, on plasmids characterized by clusters of *vmp* genes located on the ends (Table S7). The hairpin structures were confirmed on all other linear plasmids of these eight strains (ranging from 7 to 13 linear plasmids per strain), as well as in the remaining 13 genomes. Interestingly, when the hairpin structure was not determined, the number and order of ORFs at the end of the plasmid was largely similar to analogue plasmids of other strains (1938 lp41 right-end to Izh-4 lp23 left-end and lp24 -1 left-end; Izh-16 lp23 left-end to Yekat-6 lp23 left-end, Yekat-17 lp29-2-1 left-end and lp23 left-end), which suggests that these plasmids are close to completion.

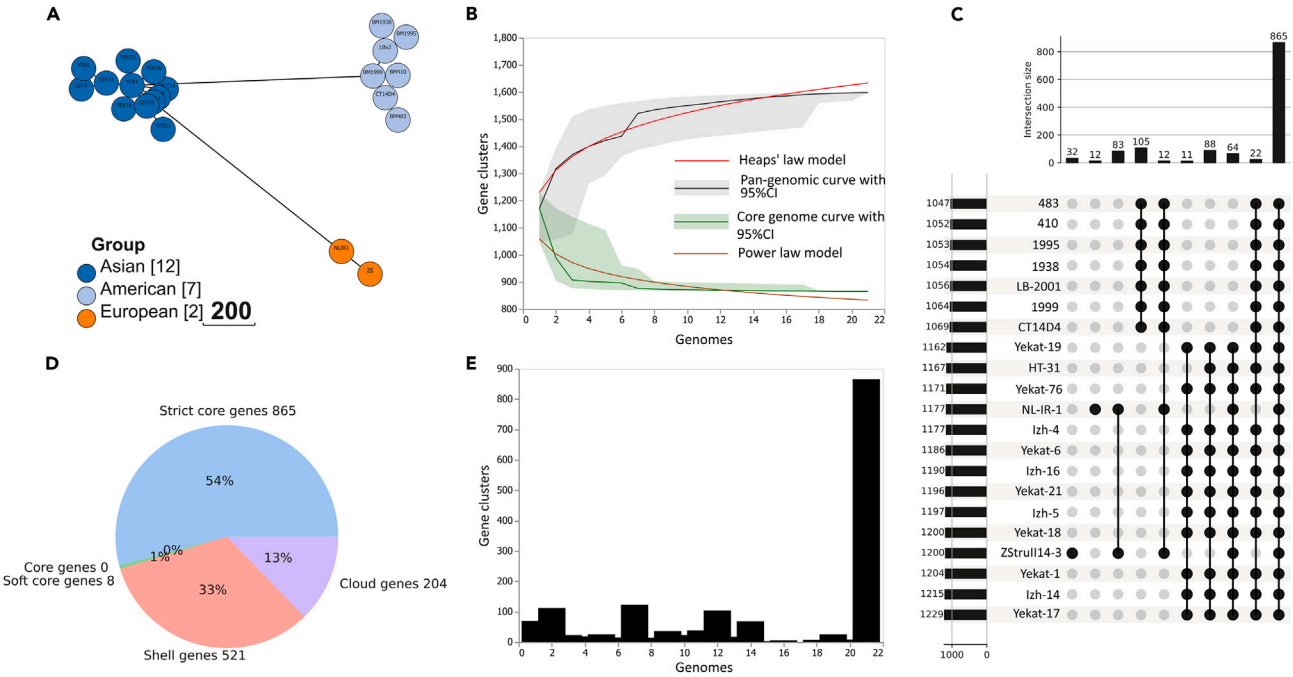


Figure 3. Pan-genome analysis of *Borrelia miyamotoi* performed by PEPPAN pipeline

(A) Whole-genome MLST tree based on PEPPAN output data. Scale indicates the number of genes different by alleles.

(B) Core-pan plots of studied *Borrelia miyamotoi* genomes. On the pan-genomic curve (black line with gray confident interval) Heaps' law model was calculated ($\gamma = 0.096 \pm 0.010$, $\kappa = 1{,}212{,}022 \pm 29{,}199$), resulting in approximately 7.292 new genes per new genome. On the core genome curve (green line with green confident interval) the power law model was applied ($\alpha = 63{,}115\text{-}\infty$, $\kappa = 1{,}142{,}212\text{-}\infty$), resulting in no new core genes by adding new genomes.

(C) Upset plot depicting the intersections of a sets of common or different gene clusters (≥ 10) among the 21 *B miyamotoi* genomes.

(D) Classification of the clusters of the genes based on frequency among 21 *B. miyamotoi* genomes.

(E) Histogram representing the gene cluster frequency spectrum G(k), of the cumulative genomes.

Pan-genome analysis

The 21 assembled genomes were included in a pan-genome analysis to investigate genetic diversity, specific features per strain and group, and similarity of gene content among linear and circular plasmids. Previously published genomes were excluded from this analysis due to lack of, or partially assembled plasmids only. Grouped genomes displayed a different allele profile of core genes (Figure 3A), resulting in 835 genes with different alleles between the Asian and European groups, and 837 between the Asia and North American groups. According to Heaps' Law and Power Law Model, this particular pan-genome is open with a γ-value of 0.096 (Figure 3B), indicating that including new genomes in the analysis will enable discovery of novel genes.[33]

A total of 1,598 gene clusters of CDS were identified, with 865 gene clusters (54.1%, 95%CI 51.7–56.6%) representing the strict core genes (Figures 3C and 3D). The core genes contained 756 chromosomal genes encoded by 99.0–100.0% of the strains. The core genome curve displayed a closed configuration, indicating no new core genes will be found by adding new genomes to the analysis. The softcore contained eight gene clusters (0.5%, 95%CI 0.2–1.0%) encoded by 95.0–99.0% of the genomes. The shell genes included 521 gene clusters (32.6%, 95%CI 30.3–35.0%) encoded by 15.0–95.0% of the genomes. The cloud genes featured 204 gene clusters (12.8%, 95%CI 11.2–14.5%) encoded by 0–15% of the genomes (Figure 3D). Noteworthy, the morphology of the gene frequency spectrum indicates that this set of genomes is not homogeneous by gene content (Figure 3E). Further analysis of gene cluster distribution concluded that 802 of 1,598 clusters (50.2%, 95%CI 47.7–52.7%) corresponded to the chromosome, while 796 clusters (49.8%, 95%CI 47.3–52.3%) were shared by plasmids, showing a conserved gene content of the chromosome.

The amount of pseudogenes on linear plasmids was 9.2% (range 0.0–29.2%, 0–19 genes/plasmid), and on circular plasmids was 2.8% (range 0.0–16.1%, 0–17 genes/plasmid) (Table S6), similar to the distribution found in *B. burgdorferi* s.l.[34] Simultaneously, a high saturation with paralogous genes of circular and linear plasmids was found. Of 796 plasmid clusters, 377 (47.4%, 95%CI 43.9–50.9%) were plasmid-type-specific and 419 (52.6%, 95%CI 49.1–56.2%) occurred in more than one plasmid type. The linear plasmid types lp72 (71 clusters), lp97 (29), lp64 (28), lp41 (26), lp70 and lp66 (21), and lp23 (17) had the highest number of unique genes among plasmid types. In general, gene content was similar among different circular plasmid types. Only cp30-5+cp24-1 and cp29-2 displayed 11 unique gene clusters.

By comparing gene clusters between different groups, we discovered 105 unique gene clusters for North American, 88 for Asian, and 83 for European strains (Figure 3C). Almost all identified unique clusters were located on circular and/or large linear plasmids, with only six North American, 12 Asian, and three European gene clusters found on the chromosome (Table S9).

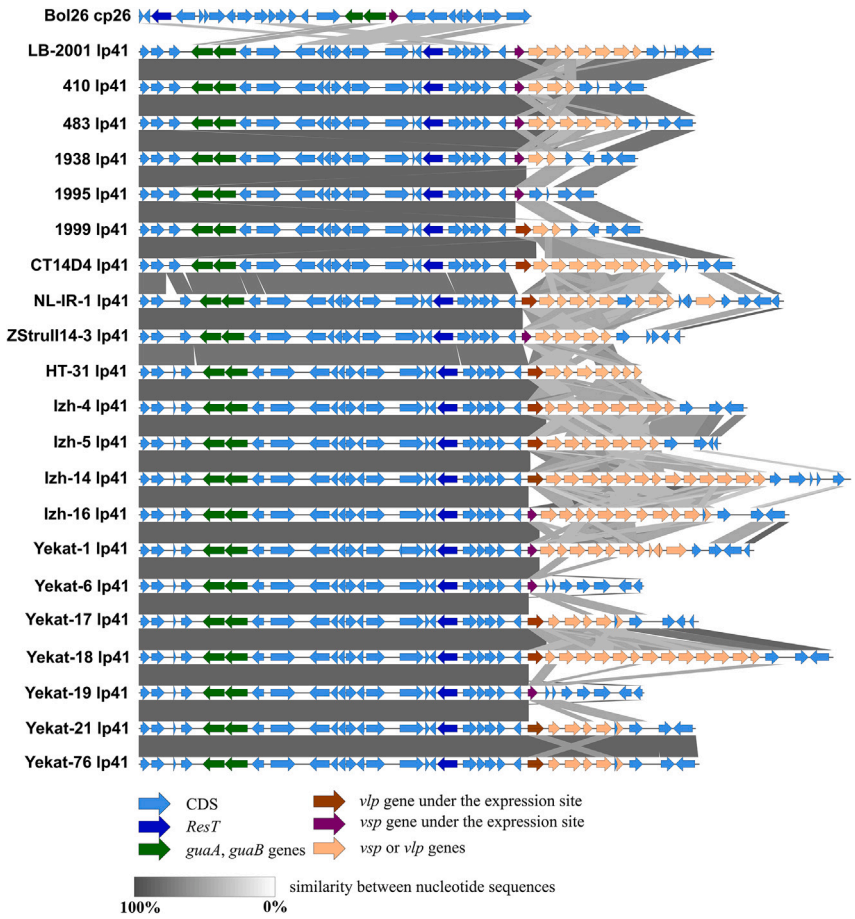


Figure 4. Nucleotide sequences comparison of virulence plasmid lp41 of *Borrelia miyamotoi*

Comparison of the nucleotide sequences of the virulence plasmid cp26 of the *B. burgdorferi* s.s. strain Bol26 (NC_012512) and the virulence plasmid lp41 of the 21 *B miyamotoi* isolates originating from America (LB-2001, 410, 483, 1938, 1995, 1999, and CT14D4), Asia (HT31, Izh-4, Izh-5, Izh-14, Izh-16, Yekat-1, Yekat-6, Yekat-17, Yekat-18, Yekat-19, Yekat-21, and Yekat-76), and Europe (ZStrulI14-9 and NL-IR-1). Plain gray blocks indicate similar areas (range of identity 70–100%) between plasmids. The arrows indicate genes and direction of open reading frames. The *vmp* block of genes are represented by the genes after the expression site, i.e., the expressed *vlp* gene in brown, the expressed *vsp* gene in purple, and the non-expressed *vmp* genes in orange. Other open reading frames are shown in arrows in light blue, *guaA* and *guaB* genes in green, and *resT* genes in dark blue.

Structural variation of the lp41 vmp expression site

Lp41 varied in size, with the plasmids from the North American strains ranging from 31 to 40 kb, the Asian from 34 to 48 kb, and the European from 37 to 44 kb (Figure 4). This discrepancy is attributed to a variable number of *vmp* genes and pseudogenes found downstream on the plasmid, ranging from one to 14, with an average of seven genes. Arrangement of the *vmp* expression sites was nearly identical, with a 100% similarity among all strains of the nucleotide composition of the ribosome binding signals within the '-10' and '-35' signals sequences. However, focusing on the individual strains showed that their genomes contained specific single nucleotide variations (Figure 5). Strain-specific variations were found based on the *vmp* gene located downstream of the expression site (Table S8). North American strains mostly displayed *vsp* genes, while two-thirds of the Asian strains showed a *vlp* gene in the expression site.

In 19 of 21 strains, the expressed Vmp also was determined by mass spectrometry, and the Vmps corresponding to the amino acid sequence results were compared to the *vmp* genes found in the expression sites by genomic sequencing (Table S8). For six strains, the amino acid sequence was 100% identical to the *vmp*. For nine isolates, this identity ranged from 99.7–98.9%. Interestingly, the identified amino acid sequence for 1995, Yekat-17, Yekat-21, and Yekat-76 did not match the *vmp* gene located at the expression site, but were 99.7–98.2% identical to *vmp* genes on another linear plasmid.

Number, location, and diversity of vmp genes

In the 21 sequenced isolates, 1,241 genes and pseudogenes encoding *vmp* genes were identified. Of these genes, 331 were found in North American strains, 737 in Asian, and 193 in European strains. However, the two European strains had significantly more *vsp* genes per isolate compared to American (p = 0.050) and Asian strains (p = 0.048), while North American isolates had the fewest *vmp* genes per isolate

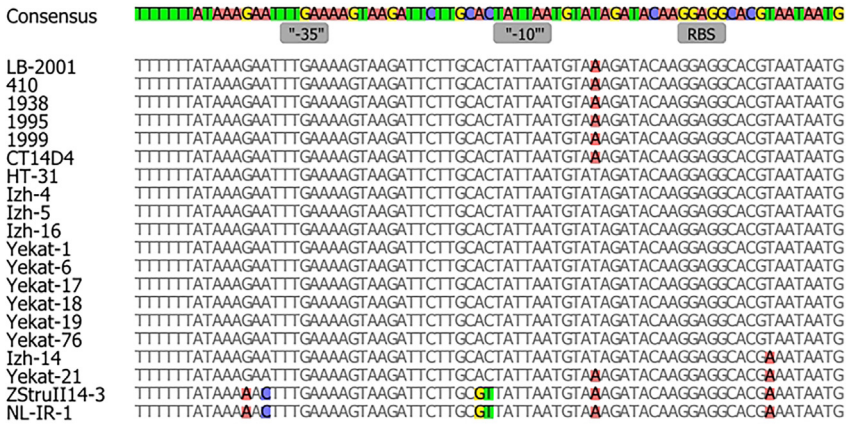


Figure 5. Alignment of the lp41 intergenic region upstream of the *vmp* expression site

Alignment of the intergenic region (73 bp) located upstream of the expressed *vmp* gene on the virulence plasmid lp41 expression site. The areas of a ribosome binding site (RBS), and the "-35" and "-10" sites are indicated in the gray boxes.

compared to Asian isolates ($p = 0.0011$). A pairwise comparison of strains from different geographical origin found significant differences ($p \leq 0.050$) in number of *vsp* genes, and a trend ($p \leq 0.056$) in the number of *vlp* genes (Figures S4A and S4B). All *vmp* genes were located on linear plasmids and organized in clusters. However, the distribution of *vmp* genes over different plasmid types differed geographically (Figure S4C). We found seven plasmid types carrying *vmp* gene clusters in North American strains, 11 in Asian strains, and six in European strains. The most saturated plasmid types were common to all strains and included lp23, lp26, and lp41. Interestingly, on long plasmid types, such as lp97, lp92, lp72, lp70, lp66, and lp64, we detected either large clusters or single *vmp* genes. For example, in the European-specific plasmid type lp92, an extended cluster was found located on the right side of the sequence (Figure 2, Data S1). The number and composition of *vmp* genes in this specific cluster were identical in NL-IR-1 and ZStruIII14-9, except for minor indels and nucleotide substitutions. In all North American strains, the right region of lp97 was characterized by a conserved cluster containing a *vlp* gene and *vsp* pseudogene. Finally, in all Asian strains, the right region of plasmid type lp64 contained a single *vsp* gene. Of note, similar regions on lp41, containing the *vmp* genes, were found within the organization of *vmp* clusters on various other linear plasmids within a specific genome (Table S8). The size of this region varied per strain, and included one *vmp*, several *vmp* genes, or the plasmid terminus. Such regions were most often found between lp41 and lp23, lp26, or lp29-2-1, and only incidentally seen between lp41 and lp24-1 or lp31 (Figure S5). Similar regions also were observed in long plasmids, e.g., lp64 from strains Yekat-6 and Yekat-19, and in lp97 from strain 1999. Interestingly, similar 3′-terminal regions were seen in Asian lp41 and lp64, including one *vsp* and seven other genes. Collectively, these findings indicate that all linear plasmids are likely involved in the mechanism of genetic recombination, also known as antigenic switching. In particular, the main method of recombination appears to be through the exchange of the terminal region of lp41 with another *vmp* gene(s)-carrying linear plasmid.

Phylogenetic analysis of vlp and vsp genes

Identified *vmp* genes and pseudogenes were compared to Interproscan databases for the domain presence of GenBank IPR000680 (*vlp* genes) or IPR001800 (*vsp* genes) to search for related genes. As a result, 913 *vlp* genes (644 coding, 269 pseudogenes), and 328 *vsp* genes (214 coding, 114 pseudogenes) were identified. The group of *vlp*-coding genes displayed an average length of 340aa (73-450aa, SD 47aa). For phylogenetic analysis, 21 *vlp* sequences were excluded based on length (≤299aa), as they could be the result of sequencing or annotation errors. The remaining 623 *vlp* genes were characterized by high diversity and grouped with a high degree of support into four distinct subfamilies, *vlp*-α, *vlp*-γ, *vlp*-δ, and *vlp*-β (Figures 6A and 6B). By phylogenetic analysis we showed that none of the *vlp* genes belonging to a particular group of strains formed group-specific clades. The group of *vsp* coding genes, represented in amino acid sequences, displayed an average length of 213aa (70-255aa, SD 14aa). For phylogenetic analysis, 3 *vsp* sequences were excluded by length (≤200aa), resulting in 211 included *vsp* coding genes. The topology of the phylogenetic tree indicated there is a subdivision into three main lineages, *vsp*-1, *vsp*-2 and *vsp*-3, with a bootstrap of 79–87% (Figures 6C and 6D). Interestingly, the genes within the *vsp*-2 group were found in only Asian and European *B. miyamotoi* strains, while the *vsp*-1 and *vsp*-3 groups contained genes from North American, Asian and European strains.

DISCUSSION

Genetically, *B. miyamotoi* is divided into distinct populations.[35,36] To date, several individual *B. miyamotoi* draft genomes have been published.[5,14–18] Rapid developments in sequencing techniques and higher data quality have enabled improvement of genomic assembly.[37] Furthermore, to investigate genetic differences in *B. miyamotoi* populations, comparative genomics on a larger number of strains isolated from North America, Asia and Europe, rather than on single isolates is required.[38,39]

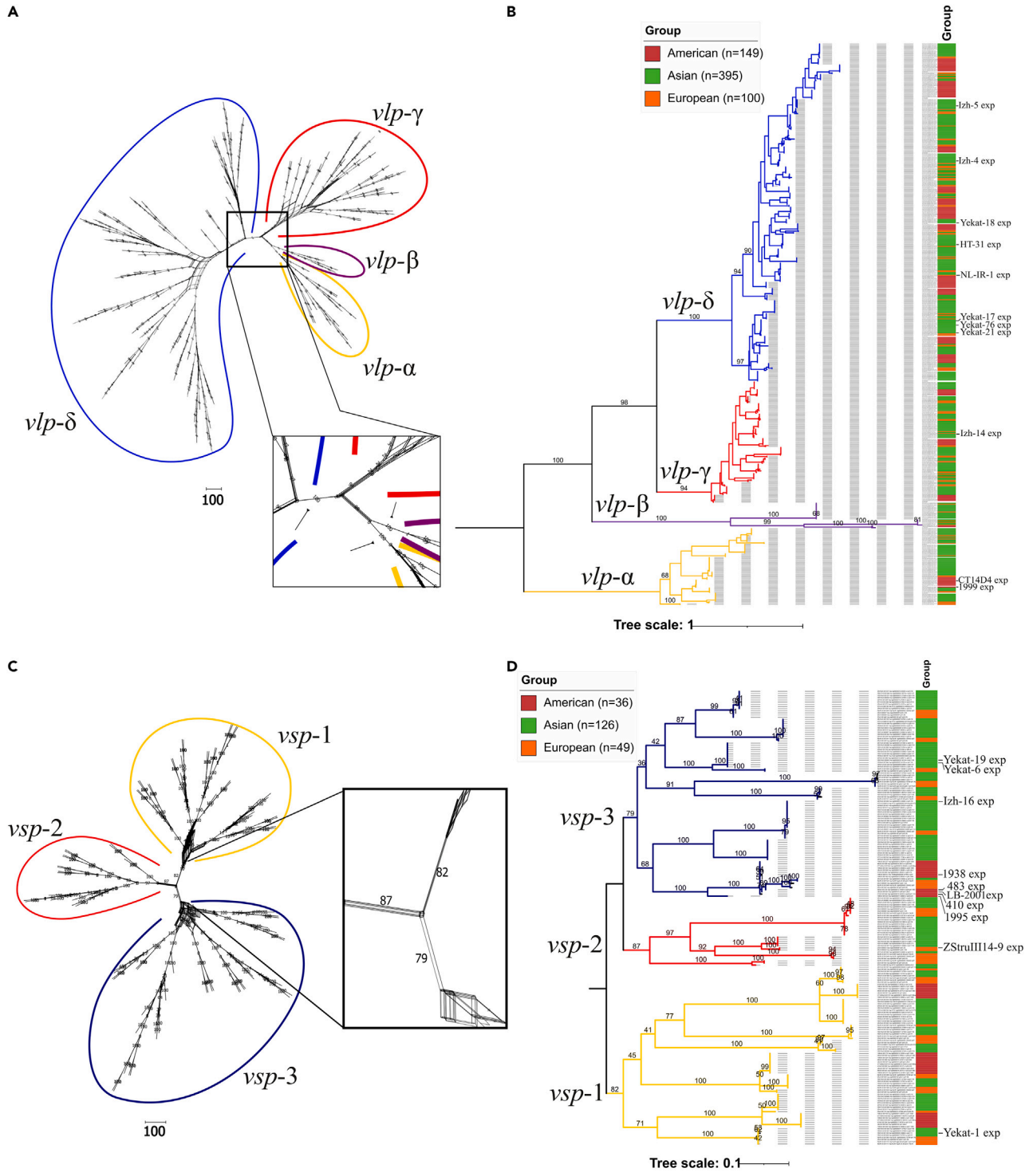


Figure 6. Phylogenetic diversity of *vlp* and *vsp* genes in *B. miyamotoi* genomes

Phylogenetic diversity of 623 *vlp* genes (A and B), and of 211 *vsp* genes (C, D) from 21 *B miyamotoi* genomes. The maximum likelihood tree was constructed using IQ-TREE v2.0.3. The trees were constructed based on amino acid sequences which contain domains corresponding to the lipoprotein 2 family (IPR000680) and OspC (IPR001800) for the *vlp* and *vsp* genes, respectively. According to BIC the WAG+F+G4 model was chosen for *vlp* genes, and the JTTDCMut+F+I + G4 model for *vsp* genes. The different subfamilies clades are indicated in panels (A and B) with *vlp*-α in yellow, *vlp*-γ in red, *vlp*-δ in blue, and *vlp*-β in purple; and in panels (C and D) with *vsp-1* in yellow, *vsp-2* in red, and *vsp-3* in blue.

Figure 6. *Continued*

(A and C) Unrooted maximum likelihood tree from the boot.splits.nex output visualized by the SplitsTree software v4. The tree shows support values for all bifurcations, computed as the occurrence frequencies in percentage in the bootstrap trees. The inlet of the tree root is enlarged to allow visualization of the branches, with 100% support values for the *vlp* genes (black arrows), and for the *vsp* genes 79–87% support values.

(B and D) Midpoint rooted maximum likelihood consensus tree from the boot.contree output. Nodes marked as 'exp' are genes located at the expression site on the lp41 plasmid. Colored bars indicate the *vmp* genes divided in the geographical groups of *B. miyamotoi* strains.

(B) The 149 American *vlp* genes in red, the 395 Asian *vlp* genes in green, and the 100 European *vlp* genes in orange.

(D) The 36 American *vsp* genes in red, the 126 Asian *vsp* genes in green, and the 49 European *vsp* genes in orange.

In the current study, by combining Illumina and PacBio sequencing, we were able to generate a high-quality assembly, allowing us to correct previously assembled genomes[5,15,18] and to validate the presence, typing, and completeness of plasmids.[40,41] Interestingly, more genomic structures such as circular plasmids were identified in the *B. miyamotoi* strain Izh-4 than previously published, using PacBio and MinIon platforms.[18] Another example is the detection of the inverted duplication of lp12, previously inaccurately assigned as lp6.[5,15,18] The high-quality data obtained in our study enabled a thorough assessment of genomic diversity among groups of *B. miyamotoi* strains and established a basis for universal nomenclature within this species.

We found that *B. miyamotoi* clustered genomically, based on not only the continent from where it was isolated, but also the *Ixodes* tick vector species.[4,13,42] However, in many cases, isolates were cultured from patients, and the tick vector only can be inferred from regional vector endemicity. Therefore, we grouped the isolates by continent of origin (North America, Europe, or Asia), rather than by vector species. Although we show that the genomic variation of *B. miyamotoi* is strongly associated with geographic spread, we speculate that this is actually based on the associated vectors, which are strongly associated with geography. In contrast to *B. burgdorferi* s.l., - which is generally transmitted horizontally - *B. miyamotoi* is predominantly transmitted vertically,[2,43,44] requiring specialized interaction between the pathogen and its vector. Therefore, *B. miyamotoi* largely relies on its vector for long-term survival. In contrast, the genetic variation of *B. burgdorferi* s.l. is strongly associated with its vertebrate hosts, with the amplifying hosts of *B. afzelii* and *B. garinii* being small mammals and birds, respectively.

Most plasmids contained clearly identifiable hairpin ends, telomeres, indicating completeness and linear structure. The sequence described as box 3 is thought to be involved in the interaction with the telomere resolvase ResT.[45] The terminal sequences of various plasmids identified in our study may be helpful in further research for rapid identification of telomeric sequences in newly sequenced genomes. Similar conserved motifs were previously described for *B. burgdorferi* s.l.[46]

Establishing a plasmid typing system based on PF32 and PF57/62 loci laid the foundation for universal plasmid nomenclature, enhancing typing applicability. As such, throughout the assembled genomes, we identified 17 circular and 20 linear plasmid types that differed between geographic regions, while core plasmids lp12, lp23, lp41, and lp72 were found in all isolates. Within RFB species, lp12 and lp41 are conserved.[40] The geographically based, group-specific plasmids most likely harbor genes essential for adaptation to a particular vector. The presence of core plasmids, on the other hand, indicates a vital role for these plasmids in the bacterium, as they contain the *vmp* expression site and may contain housekeeping genes that are essential for fundamental metabolic processes.

Previous studies established that lp41 in *B. miyamotoi* harbors the *vmp* gene expression site.[18,21,47] In the current study, lp41 showed a strong correlation with the previously studied cp26 of *B. burgdorferi* s.l. based on phylogeny of PF32 genes as well as based on similar gene sets including ResT (the telomere resolvase), GuaA (inosine 5-monophosphate dehydrogenase), GuaB (GMP synthase) genes, and plasmid maintenance genes (PF32, PF49, PF50, and PF57/62).[21,45,48] We identified a highly conserved expression site on lp41 in all isolates. However, specific nucleotide variations in promotor region near the ribosome binding site were found per geographic region, which suggests that nucleotide substitution impacts the level of expression. In an individual spirochete, the *vmp* gene that is expressed is located downstream of the expression site. Further analysis of this site revealed the expression of various *vmp* genes across the isolates. In addition, multiple *vmp* gene clusters or cassettes were found on various linear plasmids and not on circular plasmids, enabling recombinant translocation. Supported by the phylogeny, this linearization of DNA carrying the *vmp* virulence genes appears to highlight an evolutionary trend of the *Borrelia* genus. In particular, the *Borrelia burgdorferi* s.l. cp26 was linearized into lp41.

Vmp proteins are crucial for immune evasion by relapsing fever *Borrelia* and cause the characteristic relapsing fever pattern observed in infected humans, enabling evasion of the host organism's immune system by multiphasic antigenic variation.[49,50] Most of the *vmp* genes are non-functional, or silent, due to frameshifts and/or lack of promoter and ribosome binding signals. A silent *vmp* gene can be activated by translocation to a promotor region by recombination between extragenic repetitive elements.[51,52] As *vmp* gene cassettes were found on exclusively linear plasmids, we can assume their involvement in antigenic switching.[18,21,48,53–55] Moreover, the genomic analysis of all studied *B. miyamotoi* isolates revealed the potential for antigenic switching similar to other RFB spirochetes,[20,21] although clinically apparent relapsing fever is described in only a minority of patients.[24,56,57] Early antibiotic treatment, preventing relapses, may explain this divergent genomic and clinical observation. Notably, we found European isolates had the highest number of *vmp* genes, while North American strains had the lowest. In theory, this difference suggests a difference in virulence potential, as European isolates would have a broader set of *vmps* to switch between in order to avoid the humoral immune response elicited by the dominantly expressed Vmp. More European *B. miyamotoi* strains should be isolated from various sources – vectors, patients and other hosts – to study this group in more detail. Culturing egg masses originating from repleted adult female *I. ricinus* ticks may be especially practical, as this would skew toward cultivation of *B. miyamotoi* over *B. burgdorferi* s.l. isolates, which in contrast to *B. miyamotoi* are generally not transmitted transovarially.[24] In addition, a high level of suspicion of *B. miyamotoi* infection when confronted with European patients with an acute febrile illness after tick-bite, and in such cases empirically attempting to isolate *B. miyamotoi* from blood,[27] could result in much warranted European clinical isolates.

In general, the Vmp expression at the protein level confirmed the identity of the *vmp* genes found at the expression site of lp41 by genomic analysis. This finding affirms the previously described role of lp41 in bacterial virulence.[21] Both whole genome and protein sequencing were carried out from the same culture, albeit using different passages with a restriction of no more than five passages, which might explain discrepancies found in four of 19 cases.[58–60] Moreover, small differences could be explained because the mass spectrometry results were performed using previously available *B. miyamotoi* sequences, before sequences from the current study were annotated. Additionally, a recent study revealed by nanopore sequencing that LB-2001 already has minority populations with *vlps* in their expression site,[61] supporting the notion that *B. miyamotoi* continuously switches minority populations when cultured *in vitro* in the absence of antibody pressure.[62] This finding underscores the additional value of our current effort to expand genomic knowledge on this pathogen. However, more remains to be uncovered regarding the *vmp*-switching mechanism in *B. miyamotoi*.

Focusing on the specific *vmp* genes, the *vlp* genes were characterized by high diversity, and grouped with a high degree of support into the four distinct subfamilies *vlp*-α, *vlp*-γ, *vlp*-δ, and *vlp*-β, in accordance with previous RFB findings.[18,40] The topology of the phylogenetic *vsp* gene tree also suggests a subdivision into three main lineages, *vsp*-1, *vsp*-2 and *vsp*-3. This subdivision was not seen when analyzing the *vsp* genes from single genes only.[18,40]

To conclude, in this study we performed in-depth whole-genome sequencing on 21 *B miyamotoi* isolates from humans and ticks from the three endemic continents. The genomes of North American, Asian, and European strains consisted of one large linear chromosome and both circular and linear plasmids. Comparative analyses revealed clustering based on geographic region and *Ixodes* tick vector species. In total, we identified 20 linear and 17 circular plasmid types, of which linear plasmids lp12, lp23, lp41, and lp72 were shared between all isolates representing core plasmids most likely encoding genes vital for metabolic processes and virulence. In contrast, other plasmids were specific for isolates from certain geographic locations, presumably encoding genes involved in adaptation to a specific *Ixodes* tick species vector. Moreover, we discovered that *vmps* were located exclusively on linear plasmids and corroborated that the expression site was located universally on core plasmid lp41. However, we revealed great variation in both the number of *vmps* in the genomes of the *B. miyamotoi* isolates and the *vmp* located in the expression site. Collectively, our data provide novel insights into the genetic basis of vector competence, virulence, and pathogenesis of *B. miyamotoi*.

Limitations of the study

A limitation of the study is that we were bound to the number and type of *B. miyamotoi* strains we had available to perform in-depth analysis. However, this is the largest study of its kind, including seven North American, 12 Asian, and two European isolates only. As, to our knowledge, these are the only European *B. miyamotoi* strains thus far isolated. Additionally, from the human-derived isolates the genospecies of the tick that bit them is unknown, which can only be deduced by tick endemicity. Consequently, the geographical clustering found is assumed to be vector-dependent, but not ascertained. Also, pan-genome analysis shows that the pangenome is only just open; in other words, that adding more genomes would lead to finding new genes. Therefore, future studies should include more isolated, originated from a broader geographical range in the Northern hemisphere from both ticks and human patients. Moreover, it should be noted that in a small number of instances, we were unable to reliably detect hairpin structures at a single end of the identified plasmids, albeit in only 11 out of 498 terminal sequences. Modern sequencing platforms do not always allow for terminal wraparound identification due to the approaches used in which it is necessary to fragment DNA to prepare DNA libraries. Nevertheless, by using PacBio sequencing in the majority of cases we were able to find the wraparounds at the ends of the linear plasmids, showing that the hairpin structure, and thus the telomer region, is present. In Illumina sequencing and – depending on the library preparation kit used – in Nanopore sequencing, these hairpin structures are often missing. Finally, for four out of 19 isolates, discrepancies were found in the expressed Vmp depending on whether this was determined by mass spectrometry or by whole genome sequencing. Prolonged *in vitro* cultivation and passaging may lead to loss of plasmids and genetic rearrangements.[58–60] However, in this study *in vitro* cultivation was restricted to be no more than five passages and thus this seems unlikely. *In vitro* studies have previously shown that host humoral immune responses are able to induce antigenic *vmp*-switching in the *B. miyamotoi* isolate LB-2001.[23] Interestingly, a recent study revealed by nanopore sequencing that the predominantly Vsp1-expressing LB-2001 actually already has minority populations with *vlp*s being present in their expression site when cultured *in vitro*,[61] which may explain the differences in Vmp expression we observed for some isolates depending on the applied technique.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
  - Whole genome sequencing and data analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2024.110616.

ACKNOWLEDGMENTS

We would like to thank Durland Fish from Yale for the use of the American strain LB-2001, Barbara Johnson from the CDC for the use of the Japanese Ht31 strain, and Alexis Burnham for critically reviewing the manuscript.

AUTHOR CONTRIBUTIONS

J.W.H. and A.E.P. performed funding acquisition, provided formal supervisors as experts in the field, and were responsible for the acquisition of the various *Borrelia miyamotoi* isolates. D.H. and K.V.K. functioned as the project administrators and contributed equally to this manuscript. C.S. and V.F. obtained the German ZStruIII14-9 strain. S.T. obtained the five American strains, 483, 410, 1938, 1995, and 1999. L.K.B. obtained the CT14D4 strain. D.H. performed the *Borrelia miyamotoi* isolation, propagation, and DNA extraction. K.V.K. performed the whole genome sequence assembly. D.H. and K.V.K. executed data curation, methodology, formal analysis, validation, interpretation, visualization, and writing of the original draft. A.E.P. and J.W.H., as well as V.F., S.H., A.W., G.M., and H.S., assisted in conceptualization and interpretation of data. All authors had full access to the data and had final responsibility for the decision to submit for publication. All contributed to the data interpretation and writing by reviewing and editing the manuscript.

DECLARATION OF INTERESTS

All authors declare no competing interests.

Received: April 3, 2024
Revised: June 10, 2024
Accepted: July 26, 2024
Published: July 30, 2024

REFERENCES

1. Gupta, R.S., Mahmood, S., and Adeolu, M. (2013). A phylogenomic and molecular signature based approach for characterization of the phylum Spirochaetes and its major clades: proposal for a taxonomic revision of the phylum. Front. Microbiol. *4*, 217. https://doi.org/10.3389/fmicb.2013.00217.
2. Scoles, G.A., Papero, M., Beati, L., and Fish, D. (2001). A relapsing fever group spirochete transmitted by Ixodes scapularis ticks. Vector Borne Zoonotic Dis. *1*, 21–34. https://doi.org/10.1089/153036601750137624.
3. Margos, G., Fingerle, V., Cutler, S., Gofton, A., Stevenson, B., and Estrada-Peña, A. (2020). Controversies in bacterial taxonomy: The example of the genus Borrelia. Ticks Tick. Borne. Dis. *11*, 101335. https://doi.org/10.1016/j.ttbdis.2019.101335.
4. Takano, A., Toyomane, K., Konnai, S., Ohashi, K., Nakao, M., Ito, T., Andoh, M., Maeda, K., Watarai, M., Sato, K., and Kawabata, H. (2014). Tick surveillance for relapsing fever spirochete Borrelia miyamotoi in Hokkaido, Japan. PLoS One *9*, e104532. https://doi.org/10.1371/journal.pone.0104532.
5. Kuleshov, K.V., Hoornstra, D., Sprong, H., Platonov, A.E., and Hovius, J.W. (2019). Draft Whole-Genome Sequences of Two Western European Borrelia miyamotoi Isolates. Microbiol. Resour. Announc. *8*, e01314-19. https://doi.org/10.1128/MRA.01314-19.
6. Krause, P.J., Fish, D., Narasimhan, S., and Barbour, A.G. (2015). Borrelia miyamotoi infection in nature and in humans. Clin. Microbiol. Infect. *21*, 631–639. https://doi.org/10.1016/j.cmi.2015.02.006.
7. Barbour, A.G. (2014). Phylogeny of a relapsing fever Borrelia species transmitted by the hard tick Ixodes scapularis. Infect. Genet. Evol. *27*, 551–558. https://doi.org/10.1016/j.meegid.2014.04.022.
8. Mukhacheva, T.A., Salikhova, I.I., and Kovalev, S.Y. (2015). Multilocus spacer analysis revealed highly homogeneous genetic background of Asian type of Borrelia miyamotoi. Infect. Genet. Evol. *31*, 257–262. https://doi.org/10.1016/j.meegid.2015.02.009.
9. Rar, V., Livanova, N., Tkachev, S., Kaverina, G., Tikunov, A., Sabitova, Y., Igolkina, Y., Panov, V., Livanov, S., Fomenko, N., et al. (2017). Detection and genetic characterization of a wide range of infectious agents in Ixodes pavlovskyi ticks in Western Siberia, Russia. Parasit. Vectors *10*, 258. https://doi.org/10.1186/s13071-017-2186-5.
10. Iwabu-Itoh, Y., Bazartseren, B., Naranbaatar, O., Yondonjamts, E., Furuno, K., Lee, K., Sato, K., Kawabata, H., Takada, N., Andoh, M., et al. (2017). Tick surveillance for Borrelia miyamotoi and phylogenetic analysis of isolates in Mongolia and Japan. Ticks Tick. Borne. Dis. *8*, 850–857. https://doi.org/10.1016/j.ttbdis.2017.06.011.
11. Bunikis, J., Tsao, J., Garpmo, U., Berglund, J., Fish, D., and Barbour, A.G. (2004). Typing of Borrelia relapsing fever group strains. Emerg. Infect. Dis. *10*, 1661–1664. https://doi.org/10.3201/eid1009.040236.
12. Cook, V.J., Fedorova, N., Macdonald, W.P., Lane, R.S., and Barbour, A.G. (2016). Unique Strain of Borrelia miyamotoi in Ixodes pacificus Ticks, California, USA. Emerg. Infect. Dis. *22*, 2205–2207. https://doi.org/10.3201/eid2212.152046.
13. Nakao, R., Kasama, K., Boldbaatar, B., Ogura, Y., Kawabata, H., Toyoda, A., Hayashi, T., Takano, A., and Maeda, K. (2021). The evolution of hard tick-borne relapsing fever borreliae is correlated with vector species rather than geographical distance. BMC Ecol. Evol. *21*, 105. https://doi.org/10.1186/s12862-021-01838-1.
14. Hue, F., Ghalyanchi Langeroudi, A., and Barbour, A.G. (2013). Chromosome Sequence of Borrelia miyamotoi, an Uncultivable Tick-Borne Agent of Human Infection. Genome Announc. *1*, e00713-13. https://doi.org/10.1128/genomeA.00713-13.
15. Kuleshov, K.V., Koetsveld, J., Goptar, I.A., Markelov, M.L., Kolyasnikova, N.M., Sarksyan, D.S., Toporkova, M.G., Kirdyashkina, N.P., Shipulin, G.A., Hovius, J.W., and Platonov, A.E. (2018). Whole-Genome Sequencing of Six *Borrelia miyamotoi* Clinical Strains Isolated in Russia. Genome Announc. *6*, e01424-17. https://doi.org/10.1128/genomeA.01424-17.
16. Kingry, L.C., Replogle, A., Batra, D., Rowe, L.A., Sexton, C., Dolan, M., Connally, N., Petersen, J.M., and Schriefer, M.E. (2017). Toward a Complete North American Borrelia miyamotoi Genome. Genome Announc. *5*, e01557-16. https://doi.org/10.1128/genomeA.01557-16.
17. Kingry, L.C., Replogle, A., Dolan, M., Sexton, C., Padgett, K.A., and Schriefer, M.E. (2017). Chromosome and Large Linear Plasmid Sequences of a *Borrelia miyamotoi* Strain Isolated from Ixodes pacificus Ticks from California. Genome Announc. *5*, e00960-17. https://doi.org/10.1128/genomeA.00960-17.
18. Kuleshov, K.V., Margos, G., Fingerle, V., Koetsveld, J., Goptar, I.A., Markelov, M.L., Kolyasnikova, N.M., Sarksyan, D.S., Kirdyashkina, N.P., Shipulin, G.A., et al. (2020). Whole genome sequencing of Borrelia miyamotoi isolate Izh-4: reference for a complex bacterial genome. BMC Genom.

21, 16. https://doi.org/10.1186/s12864-019-6388-4.

19. Gilmore, R.D., Kneubehl, A.R., Lopez, J.E., Armstrong, B.A., Brandt, K.S., and Van Gundy, T.J. (2022). Modification of the multiplex plasmid PCR assay for Borrelia miyamotoi strain LB-2001 based on the complete genome sequence reflecting genomic rearrangements differing from strain CT13-2396. Ticks Tick. Borne. Dis. *13*, 101843. https://doi.org/10.1016/j.ttbdis.2021.101843.
20. Hamase, A., Takahashi, Y., Nohgi, K., and Fukunaga, M. (1996). Homology of variable major protein genes between Borrelia hermsii and Borrelia miyamotoi. FEMS Microbiol. Lett. *140*, 131–137. https://doi.org/10.1016/0378-1097(96)00168-1.
21. Barbour, A.G. (2016). Multiple and Diverse vsp and vlp Sequences in Borrelia miyamotoi, a Hard Tick-Borne Zoonotic Pathogen. PLoS One *11*, e0146283. https://doi.org/10.1371/journal.pone.0146283.
22. Stone, B.L., and Brissette, C.A. (2017). Host Immune Evasion by Lyme and Relapsing Fever Borreliae: Findings to Lead Future Studies for *Borrelia miyamotoi*. Front. Immunol. *8*, 12. https://doi.org/10.3389/fimmu.2017.00012.
23. Wagemakers, A., Koetsveld, J., Narasimhan, S., Wickel, M., Deponte, K., Bleijlevens, B., Jahfari, S., Sprong, H., Karan, L.S., Sarksyan, D.S., et al. (2016). Variable Major Proteins as Targets for Specific Antibodies against Borrelia miyamotoi. J. Immunol. *196*, 4185–4195. https://doi.org/10.4049/jimmunol.1600014.
24. Hoornstra, D., Azagi, T., van Eck, J.A., Wagemakers, A., Koetsveld, J., Spijker, R., Platonov, A.E., Sprong, H., and Hovius, J.W. (2022). Prevalence and clinical manifestation of Borrelia miyamotoi in Ixodes ticks and humans in the northern hemisphere: a systematic review and meta-analysis. Lancet. Microbe *3*, e772–e786. https://doi.org/10.1016/S2666-5247(22)00157-4.
25. Telford, S.R., Goethert, H.K., Molloy, P.J., and Berardi, V. (2019). Blood Smears Have Poor Sensitivity for Confirming Borrelia miyamotoi Disease. J. Clin. Microbiol. *57*, e01468-18. https://doi.org/10.1128/JCM.01468-18.
26. Fukunaga, M., Takahashi, Y., Tsuruta, Y., Matsushita, O., Ralph, D., McClelland, M., and Nakao, M. (1995). Genetic and phenotypic analysis of Borrelia miyamotoi sp. nov., isolated from the ixodid tick Ixodes persulcatus, the vector for Lyme disease in Japan. Int. J. Syst. Bacteriol. *45*, 804–810. https://doi.org/10.1099/00207713-45-4-804.
27. Koetsveld, J., Kolyasnikova, N.M., Wagemakers, A., Toporkova, M.G., Sarksyan, D.S., Oei, A., Platonov, A.E., and Hovius, J.W. (2017). Development and optimization of an in vitro cultivation protocol allows for isolation of Borrelia miyamotoi from patients with hard tick-borne relapsing fever. Clin. Microbiol. Infect. *23*, 480–484. https://doi.org/10.1016/j.cmi.2017.01.009.
28. Margos, G., Stockmeier, S., Hizo-Teufel, C., Hepner, S., Fish, D., Dautel, H., Sing, A., Dzaferovic, E., Rieger, M., Jungnick, S., et al. (2015). Long-term in vitro cultivation of Borrelia miyamotoi. Ticks Tick. Borne. Dis. *6*, 181–184. https://doi.org/10.1016/j.ttbdis.2014.12.001.
29. Casjens, S., Palmer, N., van Vugt, R., Huang, W.M., Stevenson, B., Rosa, P., Lathigra, R., Sutton, G., Peterson, J., Dodson, R.J., et al. (2000). A bacterial genome in flux: the twelve linear and nine circular extrachromosomal DNAs in an infectious isolate of the Lyme disease spirochete Borrelia burgdorferi. Mol. Microbiol. *35*, 490–516. https://doi.org/10.1046/j.1365-2958.2000.01698.x.
30. Zückert, W.R., and Meyer, J. (1996). Circular and linear plasmids of Lyme disease spirochetes have extensive homology: characterization of a repeated DNA element. J. Bacteriol. *178*, 2287–2298. https://doi.org/10.1128/jb.178.8.2287-2298.1996.
31. Casjens, S., Murphy, M., DeLange, M., Sampson, L., van Vugt, R., and Huang, W.M. (1997). Telomeres of the linear chromosomes of Lyme disease spirochaetes: nucleotide sequence and possible exchange with linear plasmid telomeres. Mol. Microbiol. *26*, 581–596. https://doi.org/10.1046/j.1365-2958.1997.6051963.x.
32. Schueler, W., Bunikis, I., Weber-Lehman, J., Comstedt, P., Kutschan-Bunikis, S., Stanek, G., Huber, J., Meinke, A., Bergstrom, S., and Lundberg, U. (2015). Complete genome sequence of Borrelia afzelii K78 and comparative genome analysis. PLoS One *10*, e0120548. https://doi.org/10.1371/journal.pone.0120548.
33. Zhou, Z., Charlesworth, J., and Achtman, M. (2020). Accurate reconstruction of bacterial pan- and core genomes with PEPPAN. Genome Res. *30*, 1667–1679. https://doi.org/10.1101/gr.260828.120.
34. Casjens, S.R., Gilcrease, E.B., Vujadinovic, M., Mongodin, E.F., Luft, B.J., Schutzer, S.E., Fraser, C.M., and Qiu, W.G. (2017). Plasmid diversity and phylogenetic consistency in the Lyme disease agent Borrelia burgdorferi. BMC Genom. *18*, 165. https://doi.org/10.1186/s12864-017-3553-5.
35. Cutler, S., Vayssier-Taussat, M., Estrada-Peña, A., Potkonjak, A., Mihalca, A.D., and Zeller, H. (2019). A new Borrelia on the block: Borrelia miyamotoi - a human health risk? Euro Surveill. *24*, 1800170. https://doi.org/10.2807/1560-7917.ES.2019.24.18.1800170.
36. Wagemakers, A., Staarink, P.J., Sprong, H., and Hovius, J.W.R. (2015). Borrelia miyamotoi: a widespread tick-borne relapsing fever spirochete. Trends Parasitol. *31*, 260–269. https://doi.org/10.1016/j.pt.2015.03.008.
37. Hepner, S., Kuleshov, K., Tooming-Kunderud, A., Alig, N., Gofton, A., Casjens, S., Rollins, R.E., Dangel, A., Mourkas, E., Sheppard, S.K., et al. (2023). A high fidelity approach to assembling the complex Borrelia genome. BMC Genom. *24*, 401. https://doi.org/10.1186/s12864-023-09500-4.
38. Margos, G., Hepner, S., Mang, C., Marosevic, D., Reynolds, S.E., Krebs, S., Sing, A., Derdakova, M., Reiter, M.A., and Fingerle, V. (2017). Lost in plasmids: next generation sequencing and the complex genome of the tick-borne pathogen Borrelia burgdorferi. BMC Genom. *18*, 422. https://doi.org/10.1186/s12864-017-3804-5.
39. Fraser, C.M., Eisen, J.A., Nelson, K.E., Paulsen, I.T., and Salzberg, S.L. (2002). The value of complete microbial genome sequencing (you get what you pay for). J. Bacteriol. *184*, 6403–6405. discusion 6405. https://doi.org/10.1128/JB.184.23.6403-6405.2002.
40. Kneubehl, A.R., and Lopez, J.E. (2023). Comparative genomics analysis of three conserved plasmid families in the Western Hemisphere soft tick-borne relapsing fever borreliae provides insight into variation in genome structure and antigenic variation systems. bioRxiv *11*, e00895-23. https://doi.org/10.1101/2023.03.06.531354.
41. Kneubehl, A.R., Krishnavajhala, A., Leal, S.M., Replogle, A.J., Kingry, L.C., Bermúdez, S.E., Labruna, M.B., and Lopez, J.E. (2022). Comparative genomics of the Western Hemisphere soft tick-borne relapsing fever borreliae highlights extensive plasmid diversity. BMC Genom. *23*, 410. https://doi.org/10.1186/s12864-022-08523-7.
42. Crowder, C.D., Carolan, H.E., Rounds, M.A., Honig, V., Mothes, B., Haag, H., Nolte, O., Luft, B.J., Grubhoffer, L., Ecker, D.J., et al. (2014). Prevalence of Borrelia miyamotoi in Ixodes ticks in Europe and the United States. Emerg. Infect. Dis. *20*, 1678–1682. https://doi.org/10.3201/eid2010.131583.
43. Breuner, N.E., Hojgaard, A., Replogle, A.J., Boegler, K.A., and Eisen, L. (2018). Transmission of the relapsing fever spirochete, Borrelia miyamotoi, by single transovarially-infected larval Ixodes scapularis ticks. Ticks Tick. Borne. Dis. *9*, 1464–1467. https://doi.org/10.1016/j.ttbdis.2018.07.006.
44. Han, S., Lubelczyk, C., Hickling, G.J., Belperron, A.A., Bockenstedt, L.K., and Tsao, J.I. (2019). Vertical transmission rates of Borrelia miyamotoi in Ixodes scapularis collected from white-tailed deer. Ticks Tick. Borne. Dis. *10*, 682–689. https://doi.org/10.1016/j.ttbdis.2019.02.014.
45. Byram, R., Stewart, P.E., and Rosa, P. (2004). The essential nature of the ubiquitous 26-kilobase circular replicon of Borrelia burgdorferi. J. Bacteriol. *186*, 3561–3569. https://doi.org/10.1128/JB.186.11.3561-3569.2004.
46. Bandy, N.J., Salman-Dilgimen, A., and Chaconas, G. (2014). Construction and characterization of a Borrelia burgdorferi strain with conditional expression of the essential telomere resolvase, ResT. J. Bacteriol. *196*, 2396–2404. https://doi.org/10.1128/JB.01435-13.
47. Barbour, A.G., Carter, C.J., Burman, N., Freitag, C.S., Garon, C.F., and Bergström, S. (1991). Tandem insertion sequence-like elements define the expression site for variable antigen genes of Borrelia hermsii. Infect. Immun. *59*, 390–397. https://doi.org/10.1128/iai.59.1.390-397.1991.
48. Barbour, A.G., and Garon, C.F. (1988). The genes encoding major surface proteins of Borrelia burgdorferi are located on a plasmid. Ann. N. Y. Acad. Sci. *539*, 144–153. https://doi.org/10.1111/j.1749-6632.1988.tb31847.x.
49. Barbour, A.G. (1988). Antigenic variation of surface proteins of Borrelia species. Rev. Infect. Dis. *10*, S399–S402. https://doi.org/10.1093/cid/10.supplement_2.s399.
50. Barbour, A.G. (1990). Antigenic variation of a relapsing fever Borrelia species. Annu. Rev. Microbiol. *44*, 155–171. https://doi.org/10.1146/annurev.mi.44.100190.001103.
51. Barbour, A.G., Burman, N., Carter, C.J., Kitten, T., and Bergström, S. (1991). Variable antigen genes of the relapsing fever agent Borrelia hermsii are activated by promoter addition. Mol. Microbiol. *5*, 489–493. https://doi.org/10.1111/j.1365-2958.1991.tb02132.x.
52. Dai, Q., Restrepo, B.I., Porcella, S.F., Raffel, S.J., Schwan, T.G., and Barbour, A.G. (2006). Antigenic variation by Borrelia hermsii occurs through recombination between extragenic repetitive elements on linear plasmids. Mol. Microbiol. *60*, 1329–1343. https://doi.org/10.1111/j.1365-2958.2006.05177.x.

53. Kitten, T., and Barbour, A.G. (1992). The relapsing fever agent Borrelia hermsii has multiple copies of its chromosome and linear plasmids. Genetics *132*, 311–324. https://doi.org/10.1093/genetics/132.2.311.
54. Tabuchi, N., Mitani, H., Seino, S., and Fukunaga, M. (2002). The 44-kb linear plasmid molecule in the relapsing fever agent Borrelia duttonii strain Ly serve as a preservation of vmp genes. Microbiol. Immunol. *46*, 159–165. https://doi.org/10.1111/j.1348-0421.2002.tb02681.x.
55. Barbour, A.G. (1993). Linear DNA of Borrelia species and antigenic variation. Trends Microbiol. *1*, 236–239. https://doi.org/10.1016/0966-842x(93)90139-i.
56. Platonov, A.E., Karan, L.S., Kolyasnikova, N.M., Makhneva, N.A., Toporkova, M.G., Maleev, V.V., Fish, D., and Krause, P.J. (2011). Humans infected with relapsing fever spirochete Borrelia miyamotoi, Russia. Emerg. Infect. Dis. *17*, 1816–1823. https://doi.org/10.3201/eid1710.101474.
57. Platonov, A.E., Toporkova, M.G., Kolyasnikova, N.M., Stukolova, O.A., Dolgova, A.S., Brodovikova, A.V., Makhneva, N.A., Karan, L.S., Koetsveld, J., Shipulin, G.A., and Maleev, V.V. (2017). [Clinical presentation of Ixodes tick-borne borreliosis caused by Borrelia miyamotoi in the context of an immune response to the pathogen]. Ter. Arkh. *89*, 35–43. https://doi.org/10.17116/terarkh2017891135-43.
58. Hovis, K.M., Jones, J.P., Sadlon, T., Raval, G., Gordon, D.L., and Marconi, R.T. (2006). Molecular analyses of the interaction of Borrelia hermsii FhbA with the complement regulatory proteins factor H and factor H-like protein 1. Infect. Immun. *74*, 2007–2014. https://doi.org/10.1128/IAI.74.4.2007-2014.2006.
59. Schwan, T.G., Burgdorfer, W., and Garon, C.F. (1988). Changes in infectivity and plasmid profile of the Lyme disease spirochete, Borrelia burgdorferi, as a result of in vitro cultivation. Infect. Immun. *56*, 1831–1836. https://doi.org/10.1128/iai.56.8.1831-1836.1988.
60. Biškup, U.G., Strle, F., and Ruzic-Sabljic, E. (2011). Loss of plasmids of Borrelia burgdorferi sensu lato during prolonged in vitro cultivation. Plasmid *66*, 1–6. https://doi.org/10.1016/j.plasmid.2011.02.006.
61. Gilmore, R.D., Armstrong, B.A., Brandt, K.S., Van Gundy, T.J., Hojgaard, A., Lopez, J.E., and Kneubehl, A.R. (2023). Analysis of variable major protein antigenic variation in the relapsing fever spirochete, Borrelia miyamotoi, in response to polyclonal antibody selection pressure. PLoS One *18*, e0281942. https://doi.org/10.1371/journal.pone.0281942.
62. Wagemakers, A., Koetsveld, J., Narasimhan, S., Wickel, M., Deponte, K., Bleijlevens, B., Jahfari, S., Sprong, H., Karan, L.S., Sarksyan, D.S., et al. (2016). Variable Major Proteins as Targets for Specific Antibodies against Borrelia miyamotoi. J. Immunol. *196*, 4185–4195. https://doi.org/10.4049/JIMMUNOL.1600014.
63. Wagemakers, A., Oei, A., Fikrig, M.M., Miellet, W.R., and Hovius, J.W. (2014). The relapsing fever spirochete Borrelia miyamotoi is cultivable in a modified Kelly-Pettenkofer medium, and is resistant to human complement. Parasit. Vectors *7*, 418. https://doi.org/10.1186/1756-3305-7-418.
64. Bushnell, B., Rood, J., and Singer, E. (2017). BBMerge - Accurate paired shotgun read merging via overlap. PLoS One *12*, e0185056. https://doi.org/10.1371/journal.pone.0185056.
65. Wood, D.E., and Salzberg, S.L. (2014). Kraken: ultrafast metagenomic sequence classification using exact alignments. Genome Biol. *15*, R46. https://doi.org/10.1186/gb-2014-15-3-r46.
66. Nurk, S., Walenz, B.P., Rhie, A., Vollger, M.R., Logsdon, G.A., Grothe, R., Miga, K.H., Eichler, E.E., Phillippy, A.M., and Koren, S. (2020). HiCanu: accurate assembly of segmental duplications, satellites, and allelic variants from high-fidelity long reads. Genome Res. *30*, 1291–1305. https://doi.org/10.1101/gr.263566.120.
67. Vaser, R., Sović, I., Nagarajan, N., and Šikić, M. (2017). Fast and accurate de novo genome assembly from long uncorrected reads. Genome Res. *27*, 737–746. https://doi.org/10.1101/gr.214270.116.
68. Seibt, K.M., Schmidt, T., and Heitkam, T. (2018). FlexiDot: highly customizable, ambiguity-aware dotplots for visual sequence analyses. Bioinformatics *34*, 3575–3577. https://doi.org/10.1093/bioinformatics/bty395.
69. Blum, M., Chang, H.Y., Chuguransky, S., Grego, T., Kandasaamy, S., Mitchell, A., Nuka, G., Paysan-Lafosse, T., Qureshi, M., Raj, S., et al. (2021). The InterPro protein families and domains database: 20 years on. Nucleic Acids Res. *49*, D344–D354. https://doi.org/10.1093/nar/gkaa977.
70. Kurtz, S., Phillippy, A., Delcher, A.L., Smoot, M., Shumway, M., Antonescu, C., and Salzberg, S.L. (2004). Versatile and open software for comparing large genomes. Genome Biol. *5*, R12. https://doi.org/10.1186/gb-2004-5-2-r12.
71. Rice, P., Longden, I., and Bleasby, A. (2000). EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet. *16*, 276–277. https://doi.org/10.1016/s0168-9525(00)02024-2.
72. Rhie, A., Walenz, B.P., Koren, S., and Phillippy, A.M. (2020). Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. Genome Biol. *21*, 245. https://doi.org/10.1186/s13059-020-02134-9.
73. Page, A.J., Cummins, C.A., Hunt, M., Wong, V.K., Reuter, S., Holden, M.T.G., Fookes, M., Falush, D., Keane, J.A., and Parkhill, J. (2015). Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics *31*, 3691–3693. https://doi.org/10.1093/bioinformatics/btv421.
74. Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics *30*, 1312–1313. https://doi.org/10.1093/bioinformatics/btu033.
75. Abascal, F., Zardoya, R., and Telford, M.J. (2010). TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. Nucleic Acids Res. *38*, W7–W13. https://doi.org/10.1093/nar/gkq291.
76. Tatusova, T., DiCuccio, M., Badretdin, A., Chetvernin, V., Nawrocki, E.P., Zaslavsky, L., Lomsadze, A., Pruitt, K.D., Borodovsky, M., and Ostell, J. (2016). NCBI prokaryotic genome annotation pipeline. Nucleic Acids Res. *44*, 6614–6624. https://doi.org/10.1093/nar/gkw569.
77. Pritchard, L., Glover, R.H., Humphris, S., Elphinstone, J.G., and Toth, I.K. (2016). Genomics and taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens. Anal. Methods *8*, 12–24. https://doi.org/10.1039/c5ay02550h.
78. Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. *32*, 1792–1797. https://doi.org/10.1093/nar/gkh340.
79. Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A., and Lanfear, R. (2020). IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Mol. Biol. Evol. *37*, 1530–1534. https://doi.org/10.1093/molbev/msaa015.
80. Replogle, A.J., Sexton, C., Young, J., Kingry, L.C., Schriefer, M.E., Dolan, M., Johnson, T.L., Connally, N.P., Padgett, K.A., and Petersen, J.M. (2021). Isolation of Borrelia miyamotoi and other Borreliae using a modified BSK medium. Sci. Rep. *11*, 1926. https://doi.org/10.1038/s41598-021-81252-1.
81. Schmidt, F.L., Sürth, V., Berg, T.K., Lin, Y.P., Hovius, J.W., and Kraiczy, P. (2021). Interaction between Borrelia miyamotoi variable major proteins Vlp15/16 and Vlp18 with plasminogen and complement. Sci. Rep. *11*, 4964. https://doi.org/10.1038/s41598-021-84533-x.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and virus strains		
LB-2001	This study and Hue et al.[14]	N/A
410	This study	N/A
483	This study	N/A
1938	This study	N/A
1995	This study	N/A
1999	This study	N/A
CT14D4	This study	N/A
HT-31	This study and Hamase et al.[20], Fukunaga et al.[26]	N/A
Izh-4	This study and Kuleshov et al.[18]	N/A
Izh-5	This study and Kuleshov et al.[15]	N/A
Izh-14	This study and Kuleshov et al.[15]	N/A
Izh-16	This study and Kuleshov et al.[15]	N/A
Yekat-1	This study and Kuleshov et al.[15]	N/A
Yekat-6	This study and Kuleshov et al.[15]	N/A
Yekat-17	This study	N/A
Yekat-18	This study	N/A
Yekat-19	This study	N/A
Yekat-21	This study	N/A
Yekat-76	This study	N/A
ZStrull14-3	This study and Margos et al.[28]	N/A
NL-IR-1	This study and Kuleshov et al.[5]	N/A
Chemicals, peptides, and recombinant proteins		
Modified Kelly-Pettenkofer medium supplemented with heat-inactivated serum (MKP-F)	Wagemakers et al.[63]	Patented
mini-protean 4–20% SDS gel	Bio-rad, USA	#4561096
Critical commercial assays		
DNeasy Blood & Tissue Kit	Qiagen, Germany	178017818
Maxwell® 16 LED DNA kit	Promega, Germany	AS1030
NexteraXT DNA Library Kit	Illumina, USA	Sets A-B (20091654, 20091656)
DNA libraries were sequenced using a 500-cycle V2 reagent kit on a MiSeq or 150 bp paired-end reads using the Illumina NovaSeq 6000	Illumina, USA	Sets C-D (20091658, 20091660)
For long read sequencing, extracted isolate DNA was submitted to whole genome sequencing (WGS) using single-molecule real-time (SMRT) sequencingon 8M SMRT cells	Pacific Biosciences Technology platform at the Norwegian Sequencing Center	www.sequencing.uio.no
g-tube fragmentation	Covaris, USA	SKU 520079
Ampure Beads	Beckman Coulter, USA	20563300
Sequel-II Binding kit 2.0 and Sequencing chemistry v2.0	Sequel-II instrument	https://www.pacb.com/wp-content/uploads/Insert-Sequel-II-Binding-Kit-2.0.pdf

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Sonicator	Branson, USA	N/A
trypsin digestion followed by LC –MS/MS on an LTQ Orbitrap	Thermo Scientific, USA	Yale School of Medicine Keck proteomics laboratory
Deposited data		
LB-2001, see also Table S10	GenBank	SAMN32260477
410, see also Table S10	GenBank	SAMN32260473
483, see also Table S10	GenBank	SAMN32260472
1938, see also Table S10	GenBank	SAMN32260474
1995, see also Table S10	GenBank	SAMN32260475
1999, see also Table S10	GenBank	SAMN32260476
CT14D4, see also Table S10	GenBank	SAMN32260480
HT-31, see also Table S10	GenBank	SAMN32260478
Izh-4, see also Table S10	GenBank	SAMN07572561
Izh-5, see also Table S10	GenBank	SAMN07572562
Izh-14, see also Table S10	GenBank	SAMN07572563
Izh-16, see also Table S10	GenBank	SAMN07572564
Yekat-1, see also Table S10	GenBank	SAMN07572565
Yekat-6, see also Table S10	GenBank	SAMN07572566
Yekat-17, see also Table S10	GenBank	SAMN10979508
Yekat-18, see also Table S10	GenBank	SAMN10979507
Yekat-19, see also Table S10	GenBank	SAMN10979509
Yekat-21, see also Table S10	GenBank	SAMN10979510
Yekat-76, see also Table S10	GenBank	SAMN10979512
ZStrull14-3, see also Table S10	GenBank	SAMN32260479
NL-IR-1, see also Table S10	GenBank	SAMN12826994
Oligonucleotides		
PF32 chromosomal gene of *Leptospira interrogans serovar Copenhageni* strain FDAARGOS_203	A6J42_RS12750	N/A
Borrelia_lipo	IPR000680	N/A
Lipoprotein_OspC	IPR001800	N/A
Software and algorithms		
BBTools	v38.90[64]	https://jgi.doe.gov/data-and-tools/software-tools/bbtools/bb-tools-user-guide/
SMRTbell® ExpressTemplate Prep Kit	2.0 Pacific Biosciences protocol for Multiplexed Microbial Libraries	https://www.pacb.com/wp-content/uploads/Procedure-Checklist-%E2%80%93-Preparing-Multiplexed-Microbial-Libraries-Using-SMRTbell-Express-Template-Prep-Kit-2.0.pdf
SMRT Tools	SMRT Link v9.0.0.92188	https://www.pacb.com/wp-content/uploads/SMRT_Link_Installation_v90.pdf
Kraken tool	2.0.8-beta[65]	https://ccb.jhu.edu/software/kraken/
BLAST database	NCBI nucleotide database	https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&BLAST_SPEC=GeoBlast&PAGE_TYPE=BlastSearch
Canu assembler	v2.1.1[66]	https://canu.readthedocs.io/en/latest/
Racon	v1.4.20[67]	https://anaconda.org/bioconda/racon ?

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Flexidot	v1.06[68]	https://github.com/molbio-dresden/flexidot
Interproscan	v5.23–62.0[69]	https://github.com/ebi-pf-team/interproscan?tab=readme-ov-file
Mummer	v3.0[70]	https://mummer.sourceforge.net/
Einverted tool analysis	Emboss package[71]	https://emboss.sourceforge.net/
CLC Genomics Workbench software	v21.0.3	N/A
Merqury package	v2020-01-29[72]	https://github.com/marbl/merqury
Roary	v1.007002[73]	https://sanger-pathogens.github.io/Roary/
PhiPack software	v1.1	https://anaconda.org/bioconda/phipack
RAxML	v8.2.9[74]	http://sco.h-its.org/exelixis/web/software/raxml/
Ete3 (etetoolkit, ete3, ete_toolchain)	v3	https://anaconda.org/conda-forge/ete3
TranslatorX	Abascal et al.[75]	http://translatorx.co.uk/
NCBI Prokaryotic Genome Annotation Pipeline	v2022-12-13.build6494[76]	https://github.com/ncbi/pgap/tree/2022-12-13.build6494
PEPPAN tool	Zhou et al.[33]	https://github.com/zheminzhou/PEPPAN
Pyani tool	v0.2.12[77]	https://github.com/widdowquinn/pyani
Muscle	v5[78]	https://www.drive5.com/muscle/
IQ-TREE	v2.0.3[79]	http://www.iqtree.org/
MASCOT database	N/A	https://www.matrixscience.com/

RESOURCE AVAILABILITY

Lead contact

Datasets generated and/or analyzed during the current study are available upon reasonable request from the Lead Contact, J.W. Hovius (pandora@amsterdamumc.nl).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- *Data:* All raw and processed sequencing data generated in this study have been submitted to the NCBI Gene Expression Omnibus (GEO; https://www.ncbi.nlm.nih.gov/geo/). GenBank accession numbers can be found for the seven North American strains, including LB2001 (SAMN32260477), 410 (SAMN32260473), 483 (SAMN32260472), 1938 (SAMN32260474), 1995 (SAMN32260475), 1999 (SAMN32260476), CT14D4 (SAMN32260480); the 12 Asian strains, including HT-31 (SAMN32260478), Izh-4 (SAMN07572561), Izh-5 (SAMN07572562), Izh-14 (SAMN07572563), Izh-16 (SAMN07572564), Yekat-1 (SAMN07572565), Yekat-6 (SAMN07572566), Yekat-17 (SAMN10979508), Yekat-18 (SAMN10979507), Yekat-19 (SAMN10979509), Yekat-21 (SAMN10979510), and Yekat-76 (SAMN10979512); and the two Europeans strains, including ZStruIII14-9 (SAMN32260479), and NL-IR-1 (SAMN12826994). The specific chromosome and plasmid contig GenBank accession number references can be found in Table S10. Chemical and gene nomenclature are provided following IUPAC standards.
- *Code:* This paper does not report original code.
- *Other items:* This paper does not report any other items.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

In this study, 21 *B miyamotoi* strains were included and categorized by geographical origin: seven strains from the USA, 12 strains from Asia (one Japanese, 11 Russian), and two from Europe (one German, one Dutch) (Table 1). The strains were obtained by different isolation methods from *Ixodes* tick species (5 strains) or human patients clinically suspected of a tick-borne disease followed by molecular and/or culture detection of *B. miyamotoi* (16 strains). Cultures were performed using a medium-based technique with Barbour-Stoenner-Kelly (BSK) medium,[80] or modified Kelly-Pettenkofer medium supplemented with heat-inactivated serum (MKP-F).[63] The North American reference strain LB-2001 was isolated from *I. scapularis* by feeding on an immunodeficient mouse and propagation of the infected murine blood sample in BSK.[14] In a similar manner, the North American 483 strain was isolated from *I. scapularis* using MKP-F. The Japanese reference strain HT-31 was directly

isolated from *I. persulcatus* using BSKII.[20,26] The European strains, including the German ZStruIII14-9 and Dutch NL-IR-1 strain, were isolated using MKP-F from an *I. ricinus* nymph and egg batch of a fully fed female, respectively.[5,28] The five North American clinical strains (410, 1938, 1995, 1999 and CT14D4) were obtained by inoculating *B. miyamotoi* infected human blood into immunodeficient mice, followed by propagation of the infected murine blood in either MKP-F or BSK. Finally, the 11 Russian clinical strains (Izh-4, Izh-5, Izh-14, Izh-16, Yekat-1, Yekat-6, Yekat-17, Yekat-18, Yekat-19, Yekat-21, and Yekat-76) were obtained from PCR proven *B. miyamotoi* disease patient blood directly propagated in MKP-F.[27] As it was not certain which *Ixodes* species vectored the mammalian isolates, we grouped the *B. miyamotoi* isolates according to the geographical location where they were isolated. Low passage numbers of the acquired culture stocks were cultivated *in vitro* in MKP-F. After two to three passages, a total amount of $\geq 10^9$ spirochetes were washed in phosphate-buffered saline prior to sequencing.[18]

METHOD DETAILS

Whole genome sequencing and data analysis

DNA isolation and preparation

Total DNA was extracted from the *Borrelia* whole cell suspension using the DNeasy Blood & Tissue Kit (Qiagen, Germany). DNA of isolate ZStruIII14-9 was extracted using Maxwell 16 LED DNA kit in the corresponding Maxwell instrument (Promega, Germany). To optimize assembly, Illumina and PacBio sequencing was performed to obtain both short and long read sequences.

Illumina sequencing

For short read sequencing, libraries were prepared using the NexteraXT DNA Library Kit (Illumina, USA). DNA libraries were sequenced using a 500-cycle V2 reagent kit on a MiSeq or 150 bp paired-end reads using the Illumina NovaSeq 6000. Low quality reads and adapter sequences were removed by BBTools v38.90.[64]

PacBio sequencing

For long read sequencing, extracted isolate DNA was submitted to whole genome sequencing (WGS) using single-molecule real-time (SMRT) sequencing by the Pacific Biosciences Technology platform at the Norwegian Sequencing Center (www.sequencing.uio.no). Library preparation was performed using the SMRTbell ExpressTemplate Prep Kit 2.0, following the Pacific Biosciences protocol for Multiplexed Microbial Libraries. DNA fragmentation (10-16 kb) was executed through g-tubes (Covaris, USA), and fragments under 3 kb were removed by Ampure Beads. The size selected library was sequenced on an 8M SMRT cell by a Sequel-II instrument, using Sequel-II Binding kit 2.0 and Sequencing chemistry v2.0. Diffusion loading resulted in 30 h of movie and 2 h of pre-extension time. Obtained reads were demultiplexed through Demultiplex Barcodes pipeline using SMRT Tools (SMRT Link v9.0.0.92188). Library preparation, sequencing and demultiplexing was performed with slight alterations for the German strain ZStruIII14-9.[28] From these subreads High fidelity (HiFi) reads were generated using a circular consensus sequencing (CCS) pipeline (SMRT Link v9.0.0.92188) with default settings on ≥ 3 passes and ≥ 0.99 predicted accuracy. CCS reads were demultiplexed to assign reads to individual isolates. To remove contaminating reads belonging to another microorganism, the sequences were checked for taxonomic accessory using Kraken tool[65] and by a BLAST on the NCBI nucleotide database.

Genome assembly

To identify linear and circular replicons or plasmids, a hands-on-detection method was developed based on their specific characteristics. The linear plasmid features included (i) a uniform read depth along the full length, (ii) a covalent closed hairpin sequences (telomeres), and (iii) replications genes. The circular plasmid feature comprised overlapping ends detected by Canu assembler and dot-plot display.

Optimal assembly of a single genome was obtained from analysis of two independent assemblies performed with the Canu assembler v2.1.1.[66] HiFi reads were used in one, and the initially obtained subreads in the other assembly. The options for the HiFi assembly included 'genomeSize = 1.6m -pacbio-HiFi'. The options for subreads assembly comprised of the 'correct', 'trim' steps and following 'assemble' step with option 'genomeSize = 1.6m'. Each draft assembly was polished by Racon v1.4.20 with HiFi reads as input.[67] The contigs of independent assemblies were evaluated through Flexidot, resulting in similar regions.[68] Self-complementary regions at the ends designated a hairpin structure and indicated a linear plasmid. These regions were found at both ends, may contain the telomeres, and mark completeness of linear plasmids. In contrast, overlapping ends indicated a circular plasmid. Next, all contigs of HiFi and subreads assemblies were removed, resembling chromosomal parts.

To attribute each contig of the draft version of the assembly to a particular plasmid type, HiFi contigs were annotated by PGAP and Interproscan to identify and type the partition protein genes based on paralogous gene family, PF32 and PF57/62 classification. To generate a consensus set of contigs per genome, the HiFi and subread assembled contigs were compared by Mummer v3.0.[70] The HiFi contigs primarily were characterized by better metrics – length of assembly, and the presence of hairpin structures – than the subread contigs. To validate a contig representing a hypothetical linear plasmid, the hairpin structures were determined to confirm completeness. Accurate location of the hairpin structures were verified by extending the analyzed contig sequences by 5000 bp of poly-N sequences from both 5′ and 3′ ends, followed by mapping the HiFi or subreads to the modified contig. These poly-N sequences allowed CLC Genomics Workbench's mapper to map long reads at the end of the contigs, which provided insight into parts of the reads that extended beyond the boundaries of the contigs. The coordinates of the hairpin structures were determined manually by Einverted tool analysis of the Emboss package.[71]

In general, the circular contigs were identified through HiFi read assembly. Validation was executed by central contig ligation, mirror-positioning the resulting parts, followed by HiFi reads alignment and manual inspection of the mapping quality. In exceptional cases, cyclic contigs were not recognized due to the absence of overlapping end sequences, but could be identified by any CDS, pseudogene, or partition gene of circular plasmids of other sequenced strains. To verify, these contigs were extended as described above, revealing their overlapping ends.

For each genome, the manually selected contig set was polished in two steps using ht CLC Genomics Workbench software v21.0.3. First, HiFi reads were used (length fraction 0.90, similarity fraction 0.90) to produce a HiFi consensus. Second, Illumina reads were used (length fraction 0.95, similarity fraction 0.95, ignoring non-specific match handling) resulting in the final consensus sequences. Quality metrics of the assembly were assessed by the Merqury package,[72] which estimates completeness and the quality value (QV), by comparing the assembly to k-mers generated from Illumina short-reads, with an increasing QV indicating a rising consensus accuracy.

Finally, after completion of the previous assembly steps, the contig's association with a linear or circular plasmid was re-evaluated. Orientations of linear and circular plasmids were modified according to the plus direction of the partition protein encoding genes PF32 or PF57/62 family. The start positions of circular plasmids were assigned to the intergenic region located upstream of the gene cluster encoding the partition proteins.

Core genome phylogeny

Phylogenetic trees of *Borrelia* species were constructed based on concatenated alignment of core genes on the chromosome, facilitating the comparison between *B. burgdorferi* s.l. and RFB, and within RFB genomes. Both interspecies and intraspecies phylogenetic analyses were performed, incorporating different sets of *Borrelia* genomes. Identification of orthologous gene clusters and the core genome alignment of chromosomes or specific plasmids was performed using Roary v1.007002.[73] Orthologous gene clusters in which signs of recombination were detected using the PhiPack software were excluded from the concatenated alignment. For interspecies comparison, a minimum of 70% and 95% identity for BLAST searches was used for *Borrelia* chromosomes and *B. miyamotoi* chromosomes, respectively. A phylogenetic tree was inferred based on core genome alignments using RAxML v8.2.9 with a GTR+Γ nucleotide substitution model and 1000 bootstrap replicates. The phylogenetic tree was visualized using Python v2.7.11 and Ete3 Python module.

Plasmid typing

A more in depth determination of the contigs to a certain type of plasmid,[29] mainly based on the identification of PF32 or PF57/62 gene relation, was performed with several modifications to what was previously described.[18] First, clustering sequences were omitted using CD-HIT on a 90% level, after which all data obtained from the Interproscan analysis per gene or pseudogene was taken into account. Second, to determine the PF32 gene homologues, a selection of cd02042 domain only was applied, due to their phylogenetic relation to cd02038 with an distinctly distant clade from PF32. As an international nomenclature for linear and circular plasmids based on PF32 and PF57/62 typing in *Borrelia* spp. has yet to be developed, the clustering of PF32 and PF57/62 protein genes in distinct clades on the phylogenetic tree was followed. The phylogenetic tree was constructed by including all nucleotide sequences encoding PF proteins of plasmids. As an outgroup the sequence of the PF32 chromosomal gene (A6J42_RS12750) of *Leptospira interrogans serovar Copenhageni* strain FDAARGOS_203 was selected. The alignment of nucleotide sequences according to the amino acid sequence was carried out using TranslatorX.[75] The phylogenetic tree was constructed using the RAxML.[74] The assignment of circular plasmids cp24 and cp30 was given from the root of the tree.

Contig annotation

Annotation of contigs during assembly was performed using a local version of the NCBI Prokaryotic Genome Annotation Pipeline (version 2022-12-13.build6494).[76] Annotation of protein signatures of identified CDS after PGAP annotation was performed with Interproscan v5 with database version 5.23–62.0.[69]

Pan- and core-genome analysis

Pan- and core-genome analyses were performed using the PEPPAN tool (https://github.com/zheminzhou/PEPPAN) to provide insights regarding specific genomic features, diversity and evolution.[33] PEPPAN pipeline was used selecting '-o ml –clust_identity 0.8' to define orthologous groups. Core-pan plots of the studied genomes, as well as upset plot depicting the intersections of gene clusters, were made with output files of clustering CDS.

Plasmids comparison

To determine plasmid similarity within a particular PF type, comparative analysis was conducted using Pyani tool v0.2.12 (script average_nucleotide_identity.py, and option -m ANIm).[77] Nucleotide sequence alignment was performed by Mummer software v3.0.[70] The Hadamard score as a plasmid similarity assessment was applied, determined by the product of the identity percentages and overlap between the two compared sequences.

Vmp gene identification and phylogeny

All genes and pseudogenes related to *vlp* (IPR000680 Borrelia_lipo) and *vsp* (IPR001800 Lipoprotein_OspC) were identified by the Interproscan database. To construct the phylogenetic tree, coding genes – translated into amino acid sequences – were used. Sequence alignment

was performed using Muscle v5.[78] Phylogenetic reconstruction based on the resulting alignment was done through IQ-TREE v2.0.3 program.[79]

Amino acid sequencing of expressed vmps

Vmp expression was determined by mass spectrometry, following previous described method.[23] Part of the *Borrelia* whole cell suspensions were heat-inactivated for 20 min in a water bath at 56°C and sonicated six times for 15 s (Branson, USA). A total of 10.0 μg per strain was loaded in a mini-protean 4–20% SDS gel (Bio-rad, USA). Gel bands around the expected size of Vsps (±23 kDa) and Vlps (±37 kDa), discriminated from flagellin (±41 kDa) and GlpQ (±39 kDa) by comparing multiple gels,[23,81] were sent to the Yale School of Medicine Keck proteomics laboratory for trypsin digestion followed by LC –MS/MS on an LTQ Orbitrap (Thermo Scientific, USA) and BLAST on the MASCOT database.

QUANTIFICATION AND STATISTICAL ANALYSIS

Comparison of the numbers of genome characteristics between groups of *B. miyamotoi* strains is performed by the nonparametric Wilcoxon-Mann-Whitney test. The *p*-values were adjusted for false-discovery rate by Benjamini-Hochberg correction. A *p*-value ≤ 0.05 indicates a significant differences between groups.