

Title: Targeting multicopy prophage genes for the differential diagnosis of Lyme disease

Jinyu Shan^{*,†,1}, Ying Jia^{†,1}, Louis Teulières², Faizal Patel¹, Martha R.J. Clokie^{*1}

¹Department of Genetics and Genome Biology, University of Leicester, Leicester, LE1 9HN, UK; ²PhelixRD Charity 230 Rue du Faubourg St Honoré, Paris, 75017, France.

[†]: These authors contributed equally to this work

* Correspondence address:

Jinyu Shan, Department of Genetics and Genome Biology, University of Leicester, Leicester, LE1 9HN, UK

Tel: 00 44 116 2711271

E-mail: js401@le.ac.uk

Martha R. J. Clokie, Department of Genetics and Genome Biology, University of Leicester, Leicester, LE1 9HN, UK

Tel: 00 44 116 252 2959

E-mail: mrjc1@le.ac.uk

Abstract: The successful treatment of Lyme disease (LD) requires an accurate diagnostic test; however, most tests are insensitive and unspecific. To overcome these challenges, we developed and validated an internally-controlled quantitative PCR (Ter-qPCR) that targets the multicopy terminase large subunit (*terL*) gene encoded by prophages that are only found in LD-causing bacteria. The *terL* protein helps phages pack their DNA. Strikingly, the detection limit of the Ter-qPCR was analytically estimated to be 22 copies and one bacterial cell in bacteria spiked blood. Furthermore, significant quantitative differences in terms of the amount of *terL* detected in healthy individuals and patients with either early or late disease. Together, the data suggests that the prophage-targeting PCR has significant power to provide a differential diagnosis for LD. Prophage encoded markers are prevalent in many other pathogenic bacteria rendering this approach highly applicable to bacterial identification in general, potentially revolutionising the detection of disease.

Introduction: Lyme disease (LD) is the most common tick-borne disease with more than 300, 000 cases in the US and 100, 000 cases in Europe reported annually^{1,2}. LD is caused by a group of bacteria classified together as the *Borrelia burgdorferi* sensu lato (s.l.) complex, that comprises a clade of more than 20 species including *B. burgdorferi* sensu stricto (s.s.) which dominates in US, and *B. garinii* and *B. afzelii* which are prevalent in Europe and Asia. The LD-causing bacteria are generally transmitted to humans after they are bitten by ticks of the *Ixodes* family infected with LD causing *Borrelia*. However, recent reports have raised concerns over *Borrelia* transmission through blood transfusion based on observations that *Borrelia* can survive and circulate in the human bloodstream³.

Currently, LD diagnosis is based on the overt clinical manifestation of disease in the form of erythema migrans (EM) skin lesions, commonly known as a ‘bull’s-eye’ rash⁴ and a history of tick exposure⁵. Although EM lesions occur in 70 to 80% of infected individuals^{5,6}, only a third of these patients develop the classic ‘bull’s-eye’ rash⁷, and many other types of skin lesions can occur which are easily confused with EM^{4,8,9}. In addition to the EM uncertainty, other

common symptoms of LD such as fatigue, muscle pain, headache, and perceived cognitive dysfunction largely overlap with an array of other diseases, including other tick-borne diseases. One such example is Relapsing Fever (RF), which is caused by close relatives of the LD-causing bacteria, such as *Borrelia miyamotoi*^{9,10}. The two *Borrelia* 'groups' responsible for LD and RF have caused great concern and clinical confusion, as they are morphologically similar and present with almost indistinguishable clinical symptoms¹¹. Despite this, they respond to different antibiotics and treatment regimens^{12,13}. Another example of confusion surrounding LD is the co-infection caused by *Bartonella* spp. This genus of bacteria is emerging as an increasingly common human infection¹⁴. Much of the controversy surrounding LD and co-infections with *Bartonella* and/or *B. miyamotoi* is due to the lack of a reliable and sensitive diagnostic method to detect and distinguish between the three groups of bacteria, the LD and RF causing *Borrelia* and *Bartonella*¹⁵. Therefore, laboratory tests to determine and distinguish between LD and co-infections play a vital role in the correct diagnosis and consequent treatment with different antibiotics.

Scientists have faced several challenges with LD detection including patients presenting with a delayed antibody response and a low number of *Borrelia* cells typically found in human clinical samples¹⁶⁻¹⁹. Serological tests for LD are also hindered by the cross-reactivity present between LD and RF^{16,18,20}. Although it is particularly difficult to diagnose LD early, it is critical, as it is far easier to treat the disease when it is detected at an early stage^{19,21,22 4,6,23,24}. Bacteria-targeting approaches, such as polymerase chain reaction (PCR) detecting the *Borrelia* chromosomal DNA, can potentially identify early LD but is relatively insensitive detecting only between 30-50% of positive cases, and is therefore deemed to have little clinical utility^{15,17,19}. The reasons behind the poor sensitivity of the current PCR methods in Lyme diagnosis are twofold; first, the current PCRs target *Borrelia* genomic DNA regions that have only one copy in each bacterium, such as the bacterial 16S rRNA gene, *RecA* gene, and the 5S-23S intergenic regions^{15,16,25-27}. Second, at least some *Borrelia* species are 'tissue-bound' and are only transiently found circulating in the blood²⁸.

In response to these diagnostic challenges, we adopted a novel approach, taking advantage of the fact that most pathogenic bacteria carry multiple complete or partial prophages (phages associated with bacteria)²⁹. These prophage sequences can form the bases of a template from which quantitative PCR (qPCR) primers and probes can be designed. It is known that *Borrelia* carry a large number of linear and circular plasmids (comprising between 33-40% of the *Borrelia* genome), among which the cp26 and cp32, and the lp54 linear plasmid, are evolutionarily stable³⁰. Of these paralogous plasmids, cp32 has been experimentally determined to be a *Borrelia burgdorferi* prophage thus it is highly likely that many of its homologues are also prophages³⁰⁻³⁴.

Each *Borrelia* species has a distinct amount of species specific variation in its prophage sequences; thus these prophages can be used as a proxy to identify the bacteria because of the tight correlation between them and the exact prophages found in each *Borrelia* host. As there are multiple prophages per *Borrelia* cell, the detectable signal is higher for prophages than bacteria. Furthermore, evidence suggests that *Borrelia* prophages can be released outside the *Borrelia* cells following encounters with stressors such as antibiotics or exposure to macrophages^{31-33,35,36}. In this study, we confirmed that *Borrelia* prophages can escape from the bacterial host cell in a spontaneous manner. Taking advantage of the multicopy

and free movement of *Borrelia* prophages, the approach to target prophages instead of bacteria will bypass the cryptic and tissue-bound feature that typifies human *Borrelia* infections²⁸. Thus, we have a greater chance of detecting the prophages in blood even when the bacteria may not be present or present in extremely low numbers. In this sense, prophages are somewhat analogous to *Borrelia* ‘footprints’.

Another challenge in detecting bacteria from blood samples is the successful extraction of DNA that forms the PCR target. It is well-known that DNA extraction from blood samples plays a pivotal role in the success of PCR diagnosis of bacterial infections^{37,38}. It is also true that PCR assays may fail to amplify if bacteria are sparse, or if an inappropriate DNA extraction method was adopted that favoured an overwhelming dominance of human DNA in the final DNA product³⁸⁻⁴⁰. To find the most suitable DNA extraction method, three blood DNA extraction methods were applied to blood samples spiked with serial dilutions of *Borrelia* cells and assessed by qPCR. We determined the best DNA extraction method by assessing each sample in order to determine the highest detectable phage copy number from the *Borrelia* spiked blood.

In this paper we have demonstrated for the first time in *Borrelia*-related diagnostics that it is possible to overcome the sensitivity challenges associated with LD detection. We highlight the enormous potential of our test to discriminate between healthy volunteers, early LD, and late LD patients. We present data from a systematic and comprehensive study that validates the use of the multicopy phage terminase large subunit (*terL*) gene as a molecular marker for the detection of *Borrelia* species. The analytical performance of the *terL*-targeting qPCR (referred to as Ter-qPCR) was thoroughly evaluated, and the test was shown to be able to detect one single *Borrelia* cell from blood samples. The diagnostic potential was evaluated using a set of blood and serum samples collected from healthy volunteers and individuals who were clinically diagnosed with LD.

In summary, we demonstrate that a quantitative phage-based PCR has the potential to revolutionise the differential diagnosis of LD from blood samples. This approach of looking for induced and specific phages may be useful for a plethora of blood-borne bacterial pathogens that cause sepsis, such as *Staphylococcus aureus*, *Escherichia coli*, and *Pseudomonas aeruginosa*⁴¹.

Results and Discussion

Multicopy, *terL* genes are widespread within *Borrelia* species

To determine which prophage gene to use as a marker the *B. burgdorferi* B31 genome was examined and shown to carry the multicopy *terL* gene (NC_000948.1). This gene encodes for the *terL* protein which is responsible for packing phage genomes and is essential for phage survival⁴². The gene was found to be present on seven of the circular plasmids from the cp32 series and on three linear plasmids within B31 genome (Supplementary information 1). Blastn analysis revealed that the *terL* homologs are widespread in LD and RF *Borrelia* spp., including *B. miyamotoi*. As summarised in Table 1, the *terL* homologs present in LD *Borrelia* species are mainly located on the cp32 plasmids with E values of zero and query cover of 100%; thus, primer design specific to species was carried out. There are 13 versions of *terL*-bearing cp32 plasmids (cp32-1 to cp32-13), all of which are present in the *B.*

burgdorferi spp. In contrast, *B. afzelii* and *B. garinii* encode eight and four such cp32 plasmids, respectively.

TerL homologs are also found in RF and *B. miyamotoi* species (E values ranging from 5e-164 to 0.03 and a query cover ranging from 90% to 5%). Thus, the phage *terL* gene appears to be a useful marker; indeed, it has previously been used as a marker to reveal the evolutionary relationships within prophages of the environmental *Paraburkholderia* species and lytic phages of *Edwardsiella ictaluri*, the causative agent of enteric septicaemia of catfish^{43,44}. In summary, the prevalence, amount of variability in the sequence, and multicopy nature of *terL* suggests that it could be a suitable marker to indicate *Borrelia* presence. Therefore, we predicted that a qPCR targeting *terL* homologs will detect *Borrelia* species to the strain level and distinguish LD *Borrelia* species from RF species and *B. miyamotoi*.

Phylogenetic analysis of the *terL* gene

To determine the potential of the *terL* gene as a marker for specific *Borrelia* species, phylogenetic analyses were carried out using neighbour-joining (NJ) (Fig. 1) and likelihood (ML) methods (Supplementary Fig. 1). Both trees were concordant with each other and demonstrated three well-defined clades of LD, RF and *B. miyamotoi*, indicating that the *terL* gene is evolutionary stable and offers resolution at the *Borrelia* species level. As shown in Fig. 1, the *terL* gene resolves *Borrelia* into genospecies, revealing an independent sub-group of LD *Borrelia* species that is well-separated from the other RF *Borrelia* group (including *B. miyamotoi*) with statistically a significant bootstrap value.

Encouragingly, and of significant importance to diagnosis and treatment, our analyses show a well-supported resolution within the LD and RF lineages. Moreover, it is important to note that the *terL* phylogenetic tree largely agrees with the 16S rRNA gene based *Borrelia* phylogeny. The only exception being that the 16S tree placed *B. miyamotoi* within the RF clade, while the *terL* tree places it outside both the LD and RF clades⁴⁵⁻⁴⁸. This variable phylogenetic position of *B. miyamotoi* indicates that it is distantly related to both LD and RF, but is more closely related to RF than to LD *Borrelia*. This detailed resolution offered by the *terL* gene is useful on two fronts: 1) it reflects the fact that the pathogenesis of *B. miyamotoi* is distinct to RF^{46,47}; and 2) as *B. miyamotoi* is the only RF *Borrelia* that can be co-transmitted with LD *Borrelia* species by hard-bodied ticks^{47,49,50}, it is really useful to have a molecular marker that can distinguish between LD and RF-causing *Borrelia* spp., and *B. miyamotoi*.

To summarise this section, the *terL* based phylogenetic analysis tightly correlates the *Borrelia* species. The phylogenetic power of the *terL* gene combined with the multi-copy nature suggests that it can be developed as a diagnostic marker for accurate identification of LD, and can be used to differentiate LD from related infections and co-infections such as RF and diseases caused by *B. miyamotoi*.

Ter-qPCR analytical specificity, sensitivity, and efficiency

To maximise the specificity and sensitivity of our test, we designed a set of primers and a probe to the most conserved regions of *terL* (to target nine out of 13 *terL* copies) (Fig. 2 and Supplementary information 2)⁵¹. The specificity of the primer/probe was confirmed by Blastn and *In silico* PCR (<http://insilico.ehu.es/PCR/>). Positive Ter-qPCR results were also obtained from all the LD *Borrelia* strains listed in Table 2. No positive Ter-qPCR was

observed from *B. spielmanii*, or RF-causing *Borrelia* strains as listed in Table 2, or other non-*Borrelia* bacterial strains tested, along with and human DNA samples (detail in Supplementary information 3).

For clinical diagnosis and treatment, it is essential to understand the *Borrelia* load present in the patient, which requires absolute quantification. To develop an absolute quantification assay, we cloned the relevant *terL* fraction into a plasmid (Ter-plasmid) and carried out the Ter-qPCR assay with samples containing background human DNA, this in order to mimic the real clinical samples. We observed a strong linear relationship between the concentration of the Ter-plasmid and Cq ($R^2=0.99$) with an amplification efficiency of 99.58% (Fig. 3A). B31 DNA dilutions also displayed a robust linear association with Cq values ($R^2=0.999$) with an amplification efficiency of 98.64% (Fig. 3B). This demonstrates the high efficiency of the Ter-qPCR. A close to 100% amplification efficiency from both serial dilution experiments confirmed that our standard curve assay was robust and repeatable⁵².

In this study, we defined the analytical limit of detection (LoD) as the lowest concentration where at least 95% of the technical replicates were positive in the Ter-qPCR⁵³. The LoD was calculated from serial dilutions of Ter-plasmids. The proportion of Ter-qPCR positive among replicates was directly correlated with the number of plasmid copies per reaction (Supplementary Fig. 2). For example, one copy of the Ter-plasmid led to two positives out of 10 replicates, while 20 and 40 copies generated 9 and 10 positives out of 10 replicates, respectively. Probit analysis via the SPSS package was used to estimate the LoD and was found to be 22 copies per PCR⁵⁴.

DNA extraction methods matter: *in vitro* spiked blood samples

To gain an insight into the potential performance of the Ter-qPCR with patients, the Ter-qPCR was applied to *Borrelia*-spiked blood samples. Consistent copy numbers from technical repeats were recovered from samples with ≥ 1 spike-in *Borrelia* cell (Fig. 4A). This indicates that the Ter-qPCR can potentially detect as low as one *Borrelia* cell from a blood sample. In contrast, the blood sample with 0.1 of a *Borrelia* cell (mimicking the scenario of an extremely low level of *Borrelia* presence in the blood) displayed one copy number out of six repeats (Fig. 4A). It is already established that the number of *Borrelia* cells circulating in the blood is extremely low and is often at the lower end of the detection limit of qPCR²⁶. Therefore, it is common to get one PCR amplification out of technical repeats due to stochastic effect when in low concentration⁵⁵. To reflect the low and random distribution nature of *Borrelia* cells in blood, we adopted the following rule for recording copy numbers: the replicate that did not generate a copy number (displayed as 'failed qPCR') was scored 'zero'^{55,56}. The 'zero' *Borrelia* presence in blood was manifested in our later study analysing LD patient samples.

To better understand the reliability of the Ter-qPCR, simple linear regression analysis was performed. As seen in Fig. 4B, a linear association ($R^2=0.9891$) was observed between the amount of spiked *Borrelia* cells and the resulting Cq values, which demonstrates that the signal intensity of Ter-qPCR correlates with the '*Borrelia* load'. In other words, it appears that the Ter-qPCR can detect as low as one *Borrelia* per 300 μ l blood, which is equivalent to 3.3 *Borrelia* cells per ml of blood. This is really promising, bearing in mind that evidence from many published studies indicates that *Borrelia* presence in LD patients can range from 1-100 cells/ml^{15,25,38,57}. Therefore this single cell sensitivity allows the Ter-qPCR test to

revolutionise the way *Borrelia* infection is diagnosed⁵⁸. To put this in context, the current practice to circumvent the bottleneck of low numbers of *Borrelia* circulating in blood is either to culture prior to qPCR or to sample large volumes of blood to artificially increase the amount of PCR templates. Both methods have obvious drawbacks^{15,26,59}. Intuitively, by targeting endogenous multi-copy genes, the Ter-qPCR assay offers a much more elegant and reliable way of increasing the amount of PCR template.

We initially chose the phenol method to extract DNA from blood as it is the method of choice when it comes to extracting phage DNA^{60,61}. Although the phenol extraction method is in many ways a gold standard, it is cumbersome. To improve the scaling of the assay, we tested commercial DNA extraction kits, such as the DNeasy Blood & Tissue Kit (a column filtration system) and Maxwell RSC Viral Total Nucleic Acid Purification Kit (a magnetic beads-based system) to see whether they could replace the solvent-based method.

Despite a significant effort, including comparisons with bacterial chromosome-targeting 16S qPCR, the phenol extraction method outperformed the other two commercial methods. Thus, the phenol method was used throughout this study. As shown in Fig. 5, regardless of qPCR methods, the phenol approach generated a consistently higher copy number compared to the other two DNA extraction methods. The Ter-qPCR coupled with the phenol method produced significantly higher copy numbers than their counterparts of the 16S qPCR from all spiked samples. Additionally, the Ter-qPCR generated robust copy numbers (three positives out of three replicates) from blood spiked with one *Borrelia* cell (Fig. 5), but the 16S qPCR only displayed amplification once from the triplicate repeats of the same sample. The outstanding sensitivity of Ter-qPCR can also be seen from the markedly lower Cq values (therefore high amounts of PCR target) of the Ter-qPCR than that of the 16S qPCR when both qPCRs were targeting the same *B31* genomic DNA (Fig. 6B). The competitive advantage of targeting multicopy genes can also be seen in that the copy numbers determined by the Ter-qPCR were consistently higher than the number of spiked *Borrelia*, while the copy numbers established via the 16S qPCR were numerically about the same as the input *Borrelia* number (Fig. 5).

Spontaneous prophage induction provides unique discriminatory power

It is logical that multicopy PCR targets will lead to higher sensitivity when compared to a single-copy PCR target⁶². We determined whether the prophage encoded genes would provide additional sensitivity to detect infections. Since some prophages can escape from their bacterial hosts either by chemical induction or spontaneous prophage induction (SPI), it is entirely possible that multiple *Borrelia* prophages could be released into the blood where they magnify the diagnostic signals and thus can be detected. In the case of *Borrelia*, *terL*-carrying cp32 prophages have been demonstrated to be susceptible to 1-methyl-3-nitroso-nitroguanidine (MNNG) induction during *in vitro* culturing^{32,35,63}. We hypothesise that *Borrelia* prophages are also capable of SPI which increases the chances of *Borrelia* prophages being induced in the human body.

To prove this SPI hypothesis, we conducted morphological and molecular studies to detect phage presence in cell-free filtrates of *Borrelia* cultures prior and post 36 h incubation. As shown in Fig. 6A, transmission electron microscope (TEM) analysis revealed particles morphologically resembling podoviruses (icosahedral heads and very short tails)⁶⁴ in the

polyethylene glycol (PEG)-purified phage fraction derived from *Borrelia* cultures post 36 h incubation. The PEG product was positive to the Ter-qPCR but negative to the *Borrelia* 16S qPCR (Fig. 6B). The Ter-qPCR positive result therefore indicates the presence of cp32 DNA (prophage) in the PEG product. Meanwhile, the PEG product derived from *Borrelia* culture prior incubation showed no phage-like particles and was negative to both the Ter-qPCR and the *Borrelia* 16S qPCR.

Our spontaneous induction data led us to predict that the prophage-based Ter-qPCR would generate a much stronger *Borrelia* signal than bacteria-based qPCR, because LD *Borrelia* species are tissue bound and thus can only circulate in blood transiently and in very low numbers²⁸. In contrast, under the scenario of induction, prophages are released into bloodstream, and thus can be a 'marker' indicating the presence of *Borrelia*, even though *Borrelia* cells can be hiding and not circulating in the blood. This situation of detecting free phage DNA from human blood bears some resemblance to identifying cell-free circulating DNA (cfDNA) as with cancer diagnosis⁶⁵. Interestingly, the same challenge also stands in terms of the method of choice for isolating cfDNA^{66,67}. Phenol chloroform was also highly efficient in recovering cfDNA from clinical samples⁶⁷. Free phages have been discovered from a range of clinical samples, including blood^{68,69}. Their biological significance remains to be understood. Considering the fact that metabolic active bacteria can better support phage reproduction⁷⁰, a high level of certain phages in the blood would indicate active infections of their respective bacteria, i.e. a strong *terL* signal would implicate active *Borrelia* infection. Given the fact that most pathogenic bacteria carry many inducible (including SPI) prophages, the rational of diagnosing bacteria by detecting their 'breakaway' prophages could revolutionise the current bacteria-focused paradigm of detecting bacterial infections.

Performance of the Ter-qPCR against clinical samples

As a final important validation step of this work, to determine the potential diagnostic value in a human setting, the Ter-qPCR was applied to 78 individuals belonging to three categories (early LD, late LD patients and healthy volunteers) who were diagnosed by Dr Louis Teulières^{6,23,24}. We intended to establish the feasibility of the Ter-qPCR to detect LD. Copy numbers were determined from both blood and serum samples with three technical repeats for each sample type. The technical repeats that did not show detectable copy numbers were scored 'zero'. The raw data was presented in the Supplementary Table. Visually, there are many more 'zero' scores from serum than blood samples, and the mean and median values obtained from blood samples in each category are much higher than those from sera (Supplementary Table). The higher copy numbers determined from the whole blood as compared to the serum samples reflects *Borrelia*'s intracellular life cycle and the fact that Ter-qPCR can detect both prophages that are inside or outside of the *Borrelia* cells (probably due to SPI). Therefore, whole blood is a more robust sample source from which to diagnose LD using PCR assay.

Overall, most of samples showed a low *terL* copy number (<10), which is consistent with the current estimation of a low concentration of *Borrelia* circulating in the blood (Supplementary Table). It is also common to have variation among technical repeats, for example, three technical repeats of the patient No. 25 showed copy numbers of 10, 2.8, and 0, respectively (Supplementary Table). This qPCR variability is due to the stochastic effect of having a low number of PCR templates, and has been observed in PCR detection of low level

bacteria in blood before^{53,55}. Markedly differences were observed in the mean copy number of early LD (2.4), and late LD (6.9), as well as healthy volunteers (0.8), which suggests a potential positive correlation between the severity of LD and the copy numbers, i.e. higher copy numbers in late LD patients, in contrast to lower copy numbers in early LD patients and healthy volunteers. Also, very encouragingly, the median values from the early (2.0) and late (1.8) LD patients are numerically higher than those of the healthy volunteers (0.7) (Supplementary Table). The fact that some healthy volunteers showed positive with copy numbers, indicates possible asymptomatic *Borrelia* carriage⁵⁵. Most importantly, the apparent marked difference in copy numbers from the three categories is supported by statistical analysis. As shown in Fig. 7, Mann-Whitney *U* test revealed significant differences between early LD, late LD patients and healthy volunteers in terms of *terL* levels determined from whole blood, which offers objective evidence to prove that LD is a fact, and that the early and late LD stages do exist. This statistics-backed difference also suggests the potential application of the Ter-qPCR to distinguish early LD from healthy asymptomatic *Borrelia* carrier.

We conclude that our prophage marker has the ability to differentiate between early and late LD patients. This is the first time such a sensitive and specific test has been developed for LD, and as mentioned, can be greatly beneficial in providing more effective treatments. For example, the Ter-qPCR could be used to monitor LD treatment outcomes, to indicate which treatment option may work best, and to help clinicians measure recovery. We are currently validating a *terL*-based assay targeting RF and *B. miyamotoi*, respectively, working towards a 'multiplex' PCR aiming to detect and differentiate LD *Borrelia* spp., RF *Borrelia* spp., and *B. miyamotoi* from a single test.

Rapid and accurate detection of microbial pathogens in blood using PCR methods is promising, but hampered by the low-level presence of bacteria in circulating blood, false signals and reduced sensitivity due to unspecific amplification of human DNA. The data presented in this study demonstrates for the first time that targeting phage DNA in blood could offer a rapid diagnosis of bacterial infections and could change the paradigm in the field of PCR detection of bacteria in general.

Materials and methods:

Phylogenetic analysis

Phylogenetic analysis was constructed using the program Molecular Evolutionary Genetics Analysis (MEGA) 7 according to the previous established method^{71,72}. Neighbour-Joining (NJ) with a maximum composite likelihood model and Maximum Likelihood (ML) based on the Tamura-Nei model analyses were conducted on a nucleotide data set of *terL* genes. Support for clades were estimated using a bootstrap analysis implemented in MEGA using 3, 000 replicates. The trees were rooted with phage Lambda (NC_001416) as an outgroup.

Borrelia strains and cultures

The *Borrelia* strains used in this study are listed in Table 1. Ten strains were provided by Professor Sven Bergström, Department of Molecular Biology, Umea University, Sweden. Seven strains were purchased from the Pasteur Institute and DSMZ (German Collection of Microorganisms and Cell Cultures GmbH). Two strains were provided by the Centre for Disease Control and Prevention (CDC), USA, and two by Cecilia Hizo-Teufel from the German

National Reference Centre for *Borrelia*. *Borrelia* cells were grown in 15 ml Falcon™ conical tubes with a culture volume of 14 ml of Barbour-Stoenner-Kelly (BSK) II medium with 7 % rabbit serum (referred to as complete BSKII or c-BSKII) at 35 °C without agitation as previously reported⁷³. All culture media was filter sterilised via 0.22 µm pore size filters. Visualisation and counting of *Borrelia* was performed using phase contrast microscopy (Ceti Magnum Trinocular) and a Fuchs Rosenthal Disposable Counting Chamber (C-Chip, NanoEnTek).

DNA extraction methods

A modified phenol-chloroform method was used to extract DNA from blood and serum samples. In brief, samples were treated with ammonium hydroxide⁷⁴ followed by the classic phenol chloroform DNA extraction method⁷⁵. The resulting DNA pellet was air dried for 5 min, dissolved in 40 µl Tris-Cl (10 mM, pH 8.5), and kept at -20°C. DNeasy Blood & Tissue Kit and Maxwell RSC Viral Total Nucleic Acid Purification Kit were used according to the respective manufacturer instructions. The Thermo Scientific™ NanoDrop™ One Spectrophotometer was used to measure the quantity and quality of DNA samples.

The Ter-qPCR

Primers and probe were designed using the PrimerQuest® Tool based on an alignment of phage *terL* genes from 13 cp32 plasmids (cp32-1 to cp32-13) carried by *B. burgdorferi* s.l strains. The resulting Ter-qPCR amplified a 147 bp target region⁷⁶. The fluorogenic probe was labelled with 6-carboxyfluorescein (FAM) fluorescent reporter dye at the 5'-end, an internal ZEN™ Quencher and an Iowa Black Fluorescent Quencher (IBFQ) to the 3' (5'FAM/ZEN/3'IBFQ). To rule out PCR inhibition and avoid false negatives, the Ter-qPCR was duplexed with an internal amplification control (IAC) qPCR that generated a 145 bp PCR product⁷⁷. The IAC DNA (accession number FJ357008.1) was synthesised by IDT and added to each Ter-qPCR template. The IAC probe was fluorescently labelled with the fluorescent dye of JOE at the 5'-end. The primers and probe targeting the *Borrelia* 16S rRNA gene were adopted from a published paper⁷⁸. All the primers, probes and PrimeTime Gene Expression Master Mix were supplied by IDT.

Construction of the standard DNA

For absolute quantification, a plasmid carrying the phage *terL* gene fragment (named as Ter-plasmid) was constructed and used as the standard curve. Specifically, a pair of PCR primers was designed using Primer Blast to amplify a 721 bp region of the phage *terL* gene (GenBank accession NC_000948), embracing the 147-bp Ter-qPCR product region. The primers were FTer721:AGACTAAGATGCGGGCAAGA and RTer721:TTGCATCAAGAGCGTCATCA. PCRs were carried out in a LabCycler (SensoQuest GmbH) in a total volume of 50 µl, containing 0.25 mM dNTPs, 3 mM MgCl₂, 3 µM primers, 50 ng template DNA, 0.5 unit of Taq polymerase (Bioline), and 5 µl 10× Taq buffer (Bioline). Amplification conditions were: 94°C for 2 min, 30 cycles of 94°C for 30 sec, 50°C for 30 sec, 72°C for 1 min, with a final extension of 10 min at 72°C. PCR products were gel-purified using a Qiagen gel extraction kit, and subjected to cloning using the NEB® PCR Cloning Kit according to the manufacturer's instructions. The recombinant Ter-plasmid DNA carrying the 721-bp *terL* gene was purified using the Qiagen Plasmid Kit from positive clones. The concentration of the Ter-plasmid was converted into DNA copy number⁷⁹.

PCR setup

The Ter-qPCR duplexed with the IAC assay was conducted in a 20 µl final reaction volume containing 10 µl 2X PrimeTime Master Mix, with each primer and each probe at a final concentration of 0.5 and 0.25 µM, respectively, 4 µl template DNA, 2 µl IAC (200 genome copies) and nuclease-free water. Standard thermal cycling conditions (Applied Biosystems™ 7500 Fast Real-Time PCR System) were followed with an initial step of 3 min at 95°C (polymerase activation), 45 cycles of 15 sec at 95°C (denaturation) and 1 min at 60°C (annealing/extension). The *Borrelia* 16S qPCR setup was carried out according to the previous report⁷⁸

A non-template control (NTC), a positive control of 10 ng *B. burgdorferi* B31 (labelled as B31) DNA and a standard curve made of a series of five tenfold dilutions of the Ter-plasmid DNA (10^6 - 10^2) were included in each run. The qPCR result was analysed and quantified according to the standard curve using FAST7500 software v2.3. For the Ter-qPCR to be valid, the IAC signal should always be produced regardless of the presence or absence of template DNA. All samples were tested in triplicate.

Analytical specificity and sensitivity

The analytical specificity of the Ter-qPCR assay was determined using both *in silico* and *in vitro* analyses. *In silico* analysis of the primer and probe set was carried out using BLAST and Primer-BLAST⁸⁰, and UCSC *In-Silico* PCR (<https://genome.ucsc.edu/cgi-bin/hgPcr>). For *in vitro* analysis, the Ter-qPCR was applied to DNA extracted from a panel of *Borrelia* strains that cause LD and RF (Table 1), and microbial species that have been used in the lab, including *Clostridium difficile*, *Clostridium perfringens*, *Escherichia coli*, *Pseudomonas aeruginosa*, *Streptococcus pneumoniae*, *Staphylococcus aureus*, *Burkholderia thailandensis*, *Burkholderia pseudomallei*, *Haemophilus influenzae* and *Salmonella enterica*. Additionally, human female and male DNA (Promega, G1521 and G1471), and DNA extracted using the phenol method from human whole blood (Cat#: SER-WB10ML from Cambridge Biosciences) were also tested using the Ter-qPCR.

The analytical sensitivity of the Ter-qPCR was firstly evaluated with five tenfold dilutions of the Ter-plasmid (10^6 to 10^2) and B31 DNA (40 ng to 4 pg), respectively. Each dilution was tested with three replicates to determine the PCR linearity and amplification efficiency. The limit of detection (LoD) was then estimated by testing Ter-plasmid dilutions from 1000 to 100, 80, 60, 40, 20, 10, 5, and 1 copies/PCR. Ten replicates were used for each dilution. To mimic the real situation of analysing DNA extracted from human samples, all Ter-plasmid serial dilution experiments were conducted with the presence of background human DNA 125 ng per PCR. Probit analysis via the SPSS software was performed to calculate the LoD with 95% probability⁸¹.

Spiked blood samples

Actively growing B31 cultures in the early exponential phase (around 10^6 spirochaetes/ml) was used to spike human whole blood (in duplicate) to generate a final amount of 10^5 , 10^4 , 10^3 , 10^2 , 10, 1, and 0.1 spike-in B31 cells in 300 µl of blood. DNA extraction was carried out using the modified phenol-chloroform method. To compare the different DNA extraction methods, a subset of the B31-spiked blood samples (corresponding to 10^3 , 10^2 , 10, and 1 B31 cells per 300 µl of blood) were also used for DNeasy and Maxwell DNA extractions,

respectively. All the resulting DNAs were analysed by both the Ter-qPCR and the *Borrelia* 16S qPCR.

Spontaneous phage induction and TEM

14 ml of early exponential phase, actively growing B31 cultures (7×10^5 - 10^6 cells/ml, dominated by free spirochaetal forms) were spun down (6 000 *g* for 20 min) and washed twice with sterile PBS. The resulting pellets were resuspended in 14 ml c-BSKII and incubated for 36 h at 35 °C. Portions of the B31 suspension prior and post 36 h incubation were centrifuged down, the resulting supernatants were filtered through 0.1 µm pore filters. Extraction of phages from the filtrates were carried out using PEG precipitation⁶¹. The PEG-purified phage product was used for DNA extraction⁸² and TEM at the Core Biotechnology Services at the University of Leicester⁷². The resulting DNA was examined by the Ter-qPCR and the *Borrelia* 16S qPCR assays.

Ethic statement and clinical samples

This study was carried out in accordance with protocols reviewed and approved by the Ethics Committee, Comités de protection des personnes (CPP) with investigator reference of Etude Phelix 01617 V1 and CPP reference of 17031. All patients were diagnosed by Dr Louis Teulière according to the ILADS guidelines. A total of 312 samples (156 whole blood and 156 serum samples) were collected from 78 individuals between April-June 2017 (23 healthy volunteers with no 'identifiable' LD symptoms, 13 early stage and 42 late stage LD patients). For everyone involved, two tubes of serum and two tubes of Ethylenediaminetetraacetic acid (EDTA)-treated whole blood (approximately one ml in each tube) were provided. Samples were provided in a coded, de-identified manner to preserve patient anonymity. Informed consent was obtained from all participants. The modified phenol chloroform DNA extraction method was applied to all the samples in a duplicate manner. Therefore, there were four DNA samples generated from one individual, two from the whole blood and two from the serum. Triplicate qPCR was applied to each DNA sample, which led to 12 Ter-qPCR data (six from whole blood, six from serum) expressed in copy numbers according to the standard curve for any one individual.

Statistical analysis

Graphpad Prism 8.4.3. was used for statistical analysis. Descriptive statistics and the D'Agostino-Pearson normality test were used to assess the data distribution. Mann-Whitney U tests were used to determine the significance of the difference between early, late LD patients and healthy volunteers. Furthermore, Probit analysis was carried out via the SPSS software suite (IBM SPSS Statistics 25) to estimate the LoD. The differences were not considered to be significant when the p-values were greater than 0.05. Linear regressions were used to establish correlations between the serial dilution of Ter-plasmids/B31 DNAs and Cq values.

Data availability

The Ter-qPCR assay includes a set of oligonucleotide primers and Taqman® probes and plasmid DNA as the standard for *in vitro* quantitative detection of *Borrelia* species causing LD. All primers and probes are described in the patent application Ref. P184103.EP.01/T. Other relevant data supporting the findings of the study are available in this article and its Supplementary Information files, or from the corresponding author upon request.

Acknowledgements

We would like to thank Dr Andrew Millard and Stacy Guiock for their help in proofreading the manuscript. We gratefully acknowledge the main funding received towards the study from the Phelix Research and Development (Phelix R&D, 37 Langton Street, SW10 0JL London, UK, the Charity Number 1156666), the 'Gift' funding from Lymefonds, the Netherlands (ANBI-number: 858578438), and the University of Leicester Drug Discovery and Diagnostics (LD3) spring fund 2018.

Author contributions

JS and YJ contributed to this work equally. JS and MRJC co-conceived the initial idea. JS expanded the initial idea into a coherent scientific project, designed primers and probes and performed in-depth data analysis. YJ was responsible for experimentation and optimisation, data collecting and initial data interpretation. JS and YJ co-wrote the manuscript. MRJC proofread the manuscript and provided valuable comments and suggestions. FP helped with bioinformatic and phylogenetic analyses. LT carried out the ethical application, clinical samples selection and some data analysis.

Competing interests

JS, LT and MRJC are listed as inventors in the patent application No. PCT/GB2017/053323. The remaining authors declare no competing interests.

References

- 1 Petrulionienė, A. *et al.* Epidemiology of Lyme disease in a highly endemic European zone. *Medicina (Kaunas)* **56**, 115, doi:10.3390/medicina56030115 (2020).
- 2 Rebman, A. W. & Aucott, J. N. Post-treatment Lyme disease as a model for persistent symptoms in Lyme disease. *Front. Med.* **7**, 57-57, doi:10.3389/fmed.2020.00057 (2020).
- 3 Pavia, C. S. & Plummer, M. M. Transfusion-associated Lyme disease - although unlikely, it is still a concern worth considering. *Front. Microbiol.* **9**, 2070-2070, doi:10.3389/fmicb.2018.02070 (2018).
- 4 Stanek, G., Wormser, G. P., Gray, J. & Strle, F. Lyme borreliosis. *Lancet* **379**, 461-473 (2012).
- 5 Chaaya, G., Jaller-Char, J. J. & Ali, S. K. Beyond the bull's eye: Recognizing Lyme disease. *J. Fam. Pract.* **65**, 373-379 (2016).
- 6 Cruickshank, M., O'Flynn, N. & Faust, S. N. Lyme disease: summary of NICE guidance. *BMJ* **12** (2018).
- 7 Gerstenblith, T. A. & Stern, T. A. Lyme disease: a review of its epidemiology, evaluation, and treatment. *Psychosomatics* **55**, 421-429, doi:10.1016/j.psych.2014.02.006 (2014).
- 8 Wormser, G. P. Clinical practice. Early Lyme disease. *N Engl J Med* **354**, 2794-2801, doi:10.1056/NEJMc061181 (2006).
- 9 Hatchette, T. F., Davis, I. & Johnston, B. L. Lyme disease: clinical diagnosis and treatment. *Can Commun Dis Rep* **40**, 194-208, doi:10.14745/ccdr.v40i11a01 (2014).
- 10 Agüero-Rosenfeld, M. E. & Wormser, G. P. *Lyme disease: diagnostic issues and controversies.* (Expert Rev Mol Diagn. 2015 Jan;15(1):1-4. doi: 10.1586/14737159.2015.989837. Epub 2014 Dec 8.).
- 11 Bergström, S. & Normark, J. Microbiological features distinguishing Lyme disease and relapsing fever spirochetes. *Wien. Klin. Wochenschr.* **130**, 484-490, doi:10.1007/s00508-018-1368-2 (2018).
- 12 Koetsveld, J. *et al.* *In vitro* susceptibility of the relapsing-fever spirochete *Borrelia miyamotoi* to antimicrobial agents. *Antimicrob. Agents Chemother.* **61**, doi:10.1128/aac.00535-17 (2017).

- 13 Prevention, C. f. D. C. a. *National Center for Emerging and Zoonotic Infectious Diseases (NCEZID), Division of Vector-Borne Diseases (DVBD)*,
<<https://www.cdc.gov/lyme/treatment/index.html>; <https://www.cdc.gov/relapsing-fever/clinicians/index.html>> (2019).
- 14 Anderson, B. E. & Neuman, M. A. *Bartonella* spp. as emerging human pathogens. *Clin. Microbiol. Rev.* **10**, 203-219 (1997).
- 15 Schutzer, S. E. *et al.* Direct diagnostic tests for Lyme disease. *Clin. Infect. Dis.* **68**, 1052-1057 (2019).
- 16 Waddell, L. A. *et al.* The accuracy of diagnostic tests for lyme disease in humans, a systematic review and meta-analysis of north American research. *PLoS One* **11** (2016).
- 17 Moore, A., Nelson, C., Molins, C., Mead, P. & Schriefer, M. Current guidelines, common clinical pitfalls, and future directions for laboratory diagnosis of lyme disease, United States. *Emerg. Infect. Dis.* **22**, 1169-1177 (2016).
- 18 Leeflang, M. M. *et al.* The diagnostic accuracy of serological tests for Lyme borreliosis in Europe: a systematic review and meta-analysis. *BMC Infect. Dis.* **16**, 016-1468 (2016).
- 19 Marques, A. R. Laboratory diagnosis of Lyme disease: advances and challenges. *Infect. Dis. Clin. North Am.* **29**, 295-307 (2015).
- 20 Magnarelli, L. A., Anderson, J. F. & Johnson, R. C. Cross-reactivity in serological tests for Lyme disease and other spirochetal infections. *J Infect Dis* **156**, 183-188, doi:10.1093/infdis/156.1.183 (1987).
- 21 Theel, E. S., Aguero-Rosenfeld, M. E., Pritt, B., Adem, P. V. & Wormser, G. P. Limitations and confusing aspects of diagnostic testing for neurologic Lyme disease in the United States. *J. Clin. Microbiol.* **57**, 01406-01418 (2019).
- 22 van den Brand, M. *et al.* Evaluation of a real-time PCR assay for detection and quantification of bacterial DNA directly in blood of preterm neonates with suspected late-onset sepsis. *Crit. Care* **22**, 018-2010 (2018).
- 23 Cameron, D. J., Johnson, L. B. & Maloney, E. L. Evidence assessments and guideline recommendations in Lyme disease: the clinical management of known tick bites, erythema migrans rashes and persistent disease. *Expert. Rev. Anti. Infect. Ther.* **12**, 1103-1135 (2014).
- 24 Miller, J. M. *et al.* A guide to utilization of the microbiology laboratory for diagnosis of infectious diseases: 2018 update by the infectious diseases society of America and the American society for microbiology. *Clin. Infect. Dis.* **67**, e1-e94 (2018).
- 25 Liveris, D. *et al.* Comparison of five diagnostic modalities for direct detection of *Borrelia burgdorferi* in patients with early Lyme disease. *Diagn. Microbiol. Infect. Dis.* **73**, 243-245 (2012).
- 26 Brettschneider, S., Bruckbauer, H., Klugbauer, N. & Hofmann, H. Diagnostic value of PCR for detection of *Borrelia burgdorferi* in skin biopsy and urine samples from patients with skin borreliosis. *J. Clin. Microbiol.* **36**, 2658-2665 (1998).
- 27 Lohr, B., Fingerle, V., Norris, D. E. & Hunfeld, K.-P. Laboratory diagnosis of Lyme borreliosis: Current state of the art and future perspectives. *Crit. Rev. Clin. Lab. Sci.* **55**, 219-245, doi:10.1080/10408363.2018.1450353 (2018).
- 28 Liang, L. *et al.* Rapid clearance of *Borrelia burgdorferi* from the blood circulation. *Parasit Vectors* **13**, 191, doi:10.1186/s13071-020-04060-y (2020).
- 29 Argov, T. *et al.* Coordination of cohabiting phage elements supports bacteria-phage cooperation. *Nat. Commun.* **10**, 5288, doi:10.1038/s41467-019-13296-x (2019).
- 30 Casjens, S. R. *et al.* Plasmid diversity and phylogenetic consistency in the Lyme disease agent *Borrelia burgdorferi*. *BMC Genomics* **18**, 165, doi:10.1186/s12864-017-3553-5 (2017).
- 31 Brisson, D., Zhou, W., Jutras, B. L., Casjens, S. & Stevenson, B. Distribution of cp32 prophages among Lyme disease-causing spirochetes and natural diversity of their lipoprotein-encoding *erp* loci. *Applied and environmental microbiology* **79**, 4115-4128, doi:10.1128/aem.00817-13 (2013).

- 32 Eggers, C. & Samuels, D. S. Molecular Evidence for a New Bacteriophage of *Borrelia burgdorferi*. *J. Bacteriol.* **181**, 7308-7313 (2000).
- 33 Grimm, D., Elias, A. F., Tilly, K. & Rosa, P. A. Plasmid stability during in vitro propagation of *Borrelia burgdorferi* assessed at a clonal level. *Infect. Immun.* **71**, 3138-3145, doi:10.1128/jai.71.6.3138-3145.2003 (2003).
- 34 Casjens, S. R. *et al.* Primordial origin and diversification of plasmids in Lyme disease agent bacteria. *BMC Genom.* **19**, 218, doi:10.1186/s12864-018-4597-x (2018).
- 35 Eggers, C. H. *et al.* Transduction by phiBB-1, a bacteriophage of *Borrelia burgdorferi*. *J. Bacteriol.* **183**, 4771-4778, doi:10.1128/jb.183.16.4771-4778.2001 (2001).
- 36 Casjens, S. R. *et al.* Genome stability of Lyme disease spirochetes: comparative genomics of *Borrelia burgdorferi* plasmids. *PLoS One* **7**, 14 (2012).
- 37 Opota, O., Jatton, K. & Greub, G. Microbial diagnosis of bloodstream infection: towards molecular diagnosis directly from blood. *Clin. Microbiol. Infect.* **21**, 323-331 (2015).
- 38 Gosiewski, T. *et al.* Comparison of methods for isolation of bacterial and fungal DNA from human blood. *Curr. Microbiol.* **68**, 149-155, doi:10.1007/s00284-013-0451-1 (2014).
- 39 Karumaa, S., Kärpänoja, P. & Sarkkinen, H. PCR identification of bacteria in blood culture does not fit the daily workflow of a routine microbiology laboratory. *J. Clin. Microbiol.* **50**, 1031-1033, doi:10.1128/JCM.01271-11 (2012).
- 40 Yamamoto, Y. PCR in diagnosis of infection: detection of bacteria in cerebrospinal fluids. *Clin. Diagn. Lab. Immunol.* **9**, 508-514, doi:10.1128/cdli.9.3.508-514.2002 (2002).
- 41 Minasyan, H. Sepsis: mechanisms of bacterial injury to the patient. *Scand. j. trauma resusc. emerg. med.* **27**, 19-19, doi:10.1186/s13049-019-0596-4 (2019).
- 42 Sun, S. *et al.* Structure and function of the small terminase component of the DNA packaging machine in T4-like bacteriophages. *Proc. Natl. Acad. Sci. U S A* **109**, 817-822, doi:10.1073/pnas.1110224109 (2012).
- 43 Pratama, A. A., Chaib De Mares, M. & van Elsas, J. D. Evolutionary history of bacteriophages in the genus *Paraburkholderia*. *Front. Microbiol.* **9**, doi:10.3389/fmicb.2018.00835 (2018).
- 44 Carrias, A. *et al.* Comparative genomic analysis of bacteriophages specific to the channel catfish pathogen *Edwardsiella ictaluri*. *Virol. J.* **8**, 6-6, doi:10.1186/1743-422X-8-6 (2011).
- 45 Takano, A. *et al.* Isolation and characterization of a novel *Borrelia* group of tick-borne borreliae from imported reptiles and their associated ticks. *Environ. Microbiol.* **12**, 134-146, doi:10.1111/j.1462-2920.2009.02054.x (2010).
- 46 Fukunaga, M. *et al.* 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 (1995).
- 47 Telford, S. R., 3rd *et al.* *Borrelia miyamotoi* disease: neither Lyme disease nor relapsing fever. *Clin. Lab. Med.* **35**, 867-882 (2015).
- 48 Molloy, P. J. *et al.* *Borrelia miyamotoi* disease in the northeastern United States: A Case Series. *Ann. Intern. Med.* **163**, 91-98 (2015).
- 49 Han, S. *et al.* Vertical transmission rates of *Borrelia miyamotoi* in *Ixodes scapularis* collected from white-tailed deer. *Ticks Tick Borne Dis.* **10**, 682-689 (2019).
- 50 Ravagnan, S. *et al.* First detection of *Borrelia miyamotoi* in *Ixodes ricinus* ticks from northern Italy. *Parasit Vectors* **11**, 130, doi:10.1186/s13071-018-2713-z (2018).
- 51 Kralik, P. & Ricchi, M. A basic guide to real time PCR in microbial diagnostics: definitions, parameters, and everything. *Front. Microbiol.* **8** (2017).
- 52 Bustin, S. A. *et al.* The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin. Chem.* **55**, 611-622, doi:10.1373/clinchem.2008.112797 (2009).
- 53 Wei, B. *et al.* Development of a PCR Assay to detect low level *Trypanosoma cruzi* in blood specimens collected with PAXgene blood DNA tubes for clinical trials treating chagas disease. *PLoS Negl. Trop. Dis.* **10**, e0005146, doi:10.1371/journal.pntd.0005146 (2016).

- 54 Pavšič, J., Žel, J. & Milavec, M. Assessment of the real-time PCR and different digital PCR platforms for DNA quantification. *Anal. Bioanal. Chem.* **408**, 107-121 (2016).
- 55 Primus, S. *et al.* Efficient detection of symptomatic and asymptomatic patient samples for *Babesia microti* and *Borrelia burgdorferi* infection by multiplex qPCR. *PloS one* **13**, e0196748-e0196748, doi:10.1371/journal.pone.0196748 (2018).
- 56 Ellison, S. L. R., English, C. A., Burns, M. J. & Keer, J. T. Routes to improving the reliability of low level DNA analysis using real-time PCR. *BMC Biotechnol* **6**, 33-33, doi:10.1186/1472-6750-6-33 (2006).
- 57 Eshoo, M. W. *et al.* Direct molecular detection and genotyping of *Borrelia burgdorferi* from whole blood of patients with early Lyme disease. *PLoS One* **7**, e36825-e36825, doi:10.1371/journal.pone.0036825 (2012).
- 58 Buchanan, C. M. *et al.* Rapid separation of very low concentrations of bacteria from blood. *J. Microbiol. Methods* **139**, 48-53, doi:10.1016/j.mimet.2017.05.004 (2017).
- 59 Tilly, K., Rosa, P. A. & Stewart, P. E. Biology of infection with *Borrelia burgdorferi*. *Infect. Dis. Clin. North. Am.* **22**, 217-v, doi:10.1016/j.idc.2007.12.013 (2008).
- 60 Pickard, D. J. Preparation of bacteriophage lysates and pure DNA. *Methods. Mol. Biol.* **502**, 3-9, doi:10.1007/978-1-60327-565-1_1 (2009).
- 61 Sambrook, J. & Russell, D. W. *Molecular cloning: a laboratory manual*. Vol. 1 (Cold Spring Harbor Laboratory, 2001).
- 62 Luo, R. F., Scahill, M. D. & Banaei, N. Comparison of single-copy and multicopy real-time PCR targets for detection of *Mycobacterium tuberculosis* in paraffin-embedded tissue. *J. Clin. Microbiol.* **48**, 2569-2570, doi:10.1128/JCM.02449-09 (2010).
- 63 Eggers, C. H. & Samuels, D. S. Molecular evidence for a new bacteriophage of *Borrelia burgdorferi*. *J. Bacteriol.* **181**, 7308-7313 (1999).
- 64 Ackermann, H. W. Bacteriophage observations and evolution. *Res. Microbiol.* **154**, 245-251 (2003).
- 65 Phallen, J. *et al.* Direct detection of early-stage cancers using circulating tumor DNA. *Sci. Transl. Med.* **9**, doi:10.1126/scitranslmed.aan2415 (2017).
- 66 Bronkhorst, A. J., Ungerer, V. & Holdenrieder, S. The emerging role of cell-free DNA as a molecular marker for cancer management. *Biomol Detect Quantif* **17**, 100087, doi:<https://doi.org/10.1016/j.bdq.2019.100087> (2019).
- 67 Hufnagl, C., Stöcher, M., Moik, M., Geisberger, R. & Greil, R. A modified phenol-chloroform extraction method for isolating circulating cell free DNA of tumor patients. *J. Nucleic Acids Invest.* **4**, e1, doi:10.4081/jnai.2013.4282 (2013).
- 68 Brown-Jaque, M., Muniesa, M. & Navarro, F. Bacteriophages in clinical samples can interfere with microbiological diagnostic tools. *Sci. Rep.* **6**, 33000, doi:10.1038/srep33000 (2016).
- 69 Pacifico, C. *et al.* Natural occurrence of *Escherichia coli*-infecting bacteriophages in clinical samples. *Front. Microbiol.* **10**, doi:10.3389/fmicb.2019.02484 (2019).
- 70 Burns, N., James, C. E. & Harrison, E. Polylysogeny magnifies competitiveness of a bacterial pathogen *in vivo*. *Evol. Appl.* **8**, 346-351 (2015).
- 71 Shan, J. *et al.* Temperature dependent bacteriophages of a tropical bacterial pathogen. *Front. Microbiol.* **5**, 599-599, doi:10.3389/fmicb.2014.00599 (2014).
- 72 Shan, J. *et al.* Prophage carriage and diversity within clinically relevant strains of *Clostridium difficile*. *Appl. Environ. Microbiol.* **78**, 6027-6034 (2012).
- 73 Zuckert, W. R. Laboratory maintenance of *Borrelia burgdorferi*. *Curr. Protoc. Microbiol.* **12** (2007).
- 74 Santino, I., Berlutti, F., Pantanella, F., Sessa, R. & Del Piano, M. Detection of *Borrelia burgdorferi* sensu lato DNA by PCR in serum of patients with clinical symptoms of Lyme borreliosis. *FEMS Microbiol. Lett.* **283**, 30-35, doi:10.1111/j.1574-6968.2008.01134.x (2008).

- 75 Sambrook, J. & Russell, D. W. Purification of Nucleic Acids by Extraction with Phenol:Chloroform. *Cold Spring Harbor Protocols* **2006**, pdb.prot4455, doi:10.1101/pdb.prot4455 (2006).
- 76 Shan, J., Clokie, M. R. & Teulières, L. Phage-based detection of borreliosis and means therefor. (2018).
- 77 Deer, D. M., Lampel, K. A. & Gonzalez-Escalona, N. A versatile internal control for use as DNA in real-time PCR and as RNA in real-time reverse transcription PCR assays. *Lett. Appl. Microbiol.* **50**, 366-372 (2010).
- 78 O'Rourke, M. *et al.* Quantitative detection of *Borrelia burgdorferi* sensu lato in erythema migrans skin lesions using internally controlled duplex real time PCR. *PLoS One* **8**, e63968, doi:10.1371/journal.pone.0063968 (2013).
- 79 Staroscik, A. *Calculator for determining the number of copies of a template*, <<http://cels.uri.edu/gsc/cndna.html>> (2004).
- 80 Ye, J. *et al.* Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics* **13**, 134, doi:10.1186/1471-2105-13-134 (2012).
- 81 Forootan, A. *et al.* Methods to determine limit of detection and limit of quantification in quantitative real-time PCR (qPCR). *Biomol. Detect. Quantif.* **12**, 1-6, doi:10.1016/j.bdq.2017.04.001 (2017).
- 82 Sambrook, J. & Russell, D. W. Purification of nucleic acids by extraction with phenol:chloroform. *Cold Spring Harb. Protoc.* **2006**, pdb.prot4455, doi:10.1101/pdb.prot4455 (2006).
- 83 Schutzer, S. E. *et al.* Whole-genome sequences of thirteen isolates of *Borrelia burgdorferi*. *J. bacteriol.* **193**, 1018, doi:10.1128/jb.01158-10 (2011).

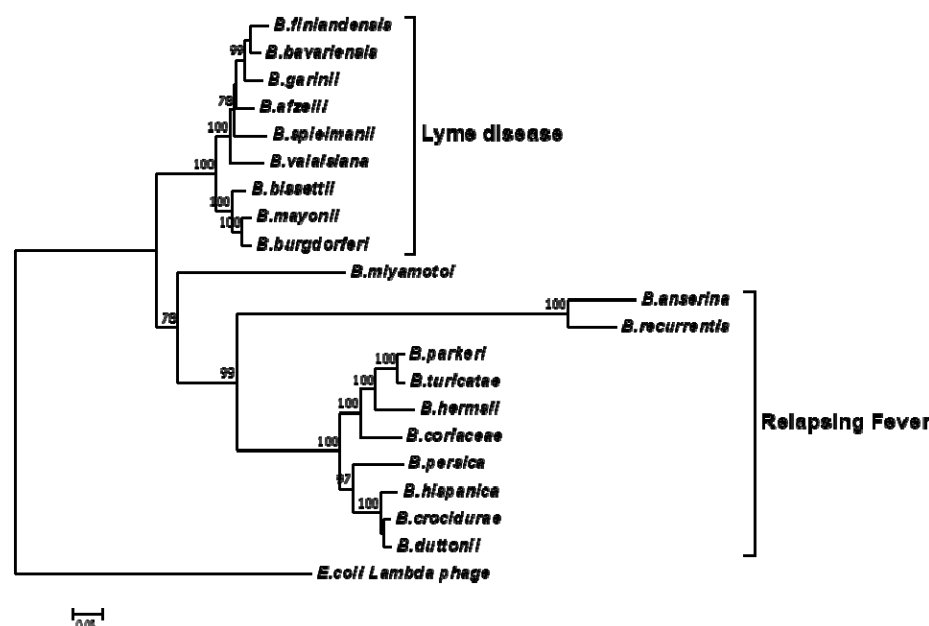


Fig. 1. A phylogenetic tree constructed based on the *terL* nucleotide sequence. The LD and RF *Borrelia* species and *B. miyamotoi* are separated into three well-supported clades. The Molecular Evolutionary Genetics Analysis (MEGA) package version 7 was used, and the tree was constructed via the neighbour-joining method. Scale bar represents the units of the number of base substitutions per site. Support for the clades was estimated via bootstrap analysis in MEGA with 3,000 replicates,

values are indicated at the nodes (only values greater than 75 are displayed). The *terL* sequence from phage lambda was used as an outgroup to root the tree.

```

cp32-12    AAACCTCTATAAAGATATACCATCATATAAAGCAAGAGTTTTGCTAGGTGAGTGGATAGCA
cp32-9     AAACCTCTATAAAGATATACCATCATATAAAGCAAGAGTTTTGCTAGGTGAGTGGATAGCA
cp32-13    AAACCTCTATAAAGATATACCATCATATAAAGCAAGAGTTTTGCTAGGTGAGTGGATAGCA
cp32-2     AAGCTATATAAAGATATACCATCATATAAAGCAAGAGTTTTGCTAGGTGAATGGATAGCA
cp32-10    AAGCTATATAAAGATATACCATCATATAAAGCAAGAGTTTTGCTAGGTGAGTGGATAGCA
cp32-11    AAGCTATATAAAGATATACCATCATATAAAGCAAGAGTTTTGCTAGGTGAGTGGATAGCA
cp32-8     AAGCTATATAAAGATATACCATCATATAAAGCAAGAGTTTTGTTAGGTGAGTGGATAGCA
cp32-1     AAGCTATATAAAGATATACCATCATATAAAGCAAGAGTTTTGTTAGGTGAGTGGATAGCA
cp32-5     AAGCTATATAAAGATATACCATCATATAAAGCAAGAGTTTTGTTAGGTGAGTGGATAGCA
cp32-6     AAGCTATATAAAGATATACCATCATATAAAGCAAGAGTTTTGTTAGGTGAGTGGATAGCA
cp32-3     AAGCTATATAAAGATATACCATCATATAAAGCAAGAGTTTTGTTAGGTGAGTGGATAGCA
cp32-4     AAGCTATATAAAGATATACCATCATATAAAGCAAGAGTTTTGTTAGGTGAGTGGATAGCA
cp32-7     AAGCTATATAAAGATATACCATCATATAAAGCAAGAGTTTTGTTAGGTGAGTGGATAGCA
          ** * *****
cp32-12    AGCACTGATTCGAATTTTACACAAATAAATAATTACTGATGATTATGTATTTACTAGCCCCG
cp32-9     AGCACTGATTCGAATTTTACACAAATAAATAATTACTGATGATTATATTTACTAGCCCCG
cp32-13    AGCACTGATTCGAATTTTACACAAATAAATAATTACTGATGATTATGTATTTACTAGTCCA
cp32-2     AGCACCATTTCGAATTTTACACAAATAAATAATTACTGATGATTATGTATTTACTAGCCCCG
cp32-10    AGCACTGATTCGAATTTTACACAAATAAATAATTACTGATGATTATGTATTTACTAGCCCCG
cp32-11    AGCACTGATTCGAATTTTACACAAATAAATAATTACTGATGATTATGTATTTACTAGCCCCG
cp32-8     AGCACTGATTCGAATTTTACACAAATAAATAATTACTGATGATTATGTATTTACTAGCCCCG
cp32-1     AGCACTGATTCGAATTTTACACAAATAAATAATTACTGATGATTATGTATTTACTAGCCCCG
cp32-5     AGCACTGATTCGAATTTTACACAAATAAATAATTACTGATGATTATGTATTTACTAGCCCCG
cp32-6     AGCACTGATTCGAATTTTACACAAATAAATAATTACTGATGATTATGTATTTACTAGCCCCG
cp32-3     AGCACCATTTCGAATTTTACACAAATAAATAATTACTGATGATTATGTATTTACTAGCCCCG
cp32-4     AGCACTGATTCGAATTTTACACAAATAAATAATTACTGATGATTATGTATTTACTAGCCCCG
cp32-7     AGCACTGATTCGAATTTTACACAAATAAATAATTACTGATGATTATGTATTTACTAGCCCCG
          **** *****
cp32-12    ATAGCATATTTAGACCCAGCAATTTAGTGTGGCGGGGATAAAGTGCATTATGTGTTATG
cp32-9     ATAGCATATTTAGACCCAGCAATTTAGTGTGGAGGGGATAAAGTGCATTATGTGTTATG
cp32-13    ATAGCATATTTAGACCCAGCAATTTAGTGTGGAGGGGATAAAGTGCATTATGTGTTATG
cp32-2     ATAGCATATTTAGACCCAGCAATTTAGTGTGGAGGGGATAAAGTGCATTATGTGTTATG
cp32-10    ATAGCATATTTAGACCCAGCAATTTAGTGTGGCGGGGATAAAGTGCATTATGTGTTATG
cp32-11    ATAGCATATTTAGACCCAGCAATTTAGTGTGGCGGGGATAAAGTGCATTATGTGTTATG
cp32-8     ATAGCATATTTAGACCCAGCAATTTAGTGTGGAGGGGATAAAGTGCATTATGTGTTATG
cp32-1     ATAGCATATTTAGACCCAGCAATTTAGTGTGGCGGGGATAAAGTGCATTATGTGTTATG
cp32-5     ATAGCATATTTAGACCCAGCAATTTAGTGTGGCGGGGATAAAGTGCATTATGTGTTATG
cp32-6     ATAGCATATTTAGACCCAGCAATTTAGTGTGGAGGGGATAAAGTGCATTATGTGTTATG
cp32-3     ATAGCATATTTAGACCCAGCAATTTAGTGTGGCGGGGATAAAGTGCATTATGTGTTATG
cp32-4     ATAGCATATTTAGACCCAGCAATTTAGTGTGGCGGGGATAAAGTGCATTATGTGTTATG
cp32-7     ATAGCATATTTAGACCCAGCAATTTAGTGTGGCGGGGATAAAGTGCATTATGTGTTATG
          ***** ** *****
cp32-12    GAGCGAGTTGATGATAAGTATTATGCTTTTGTATTTCAAGACCAAGACGAGCTAATGAT
cp32-9     GAGCGAGTTGATGATAAGTATTATGCTTTTGTATTTCAAGACCAAGACGAGCTAATGAT
cp32-13    GAGCGAGTTGATGATAAGTATTATGCTTTTGTATTTCAAGACCAAGACGAGCTAATGAT
cp32-2     GAGCGAATTGATGATAAGTATTATGCTTTTGTATTTCAAGACCAAGACGAGCTAATGAT
cp32-10    GAGCGAGTTGATGATAAGTATTATGCTTTTGTATTTCAAGACCAAGACGAGCTAATGAT
cp32-11    GAGCGAGTTGATGATAAGTATTATGCTTTTGTATTTCAAGACCAAGACGAGCTAATGAT
cp32-8     GAGCGAGTTGATGATAAGTATTATGCTTTTGTATTTCAAGACCAAGACGAGCTAATGAT
cp32-1     GAGCGAGTTGATGATAAGTATTATGCTTTTGTATTTCAAGACCAAGACGAGCTAATGAT
cp32-5     GAGCGAGTTGATGATAAGTATTATGCTTTTGTATTTCAAGACCAAGACGAGCTAATGAT
cp32-6     GAGCGAGTTGATGATAAGTATTATGCTTTTGTATTTCAAGACCAAGACGAGCTAATGAT
cp32-3     GAGCGAGTTGATGATAAGTATTATGCTTTTGTATTTCAAGACCAAGACGAGCTAATGAT
cp32-4     GAGCGAGTTGATGATAAGTATTATGCTTTTGTATTTCAAGACCAAGACGAGCTAATGAT
cp32-7     GAGCGAGTTGATGATAAGTATTATGCTTTTGTATTTCAAGACCAAGACGAGCTAATGAT
          ***** *****

```

Fig. 2 Alignment of the *terL* gene sequence located on the 13 cp32 plasmids (cp32-1 to cp32-13). Identical nucleotides are indicated with an asterisk. Conserved regions of the forward and reverse primers, and probe are highlighted in yellow, green and purple, respectively. Identical primer/probe sequences are present in nine of the *terL* genes, except for those of the cp32-2, cp32-3, cp32-6 and cp32-13 plasmids.

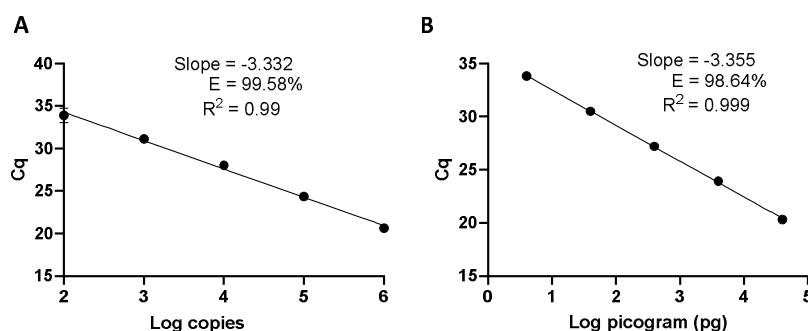


Fig. 3 The Ter-qPCR against serial dilutions of the Ter-plasmid and *B. burgdorferi* B31 DNA to measure the LoD and PCR efficiency: (A) Ter-plasmid and (B) B31 DNA. Cq values were plotted against the log values of serial dilutions of the Ter-plasmid and B31 DNA, respectively. Simple regression analysis was carried out using the Graphpad Prism 8.4.3 software. The slope, coefficient of correlation (R^2) and efficiency of the reaction (E) are shown. Each dot represents the average value from triplicate amplifications, along with the SD.

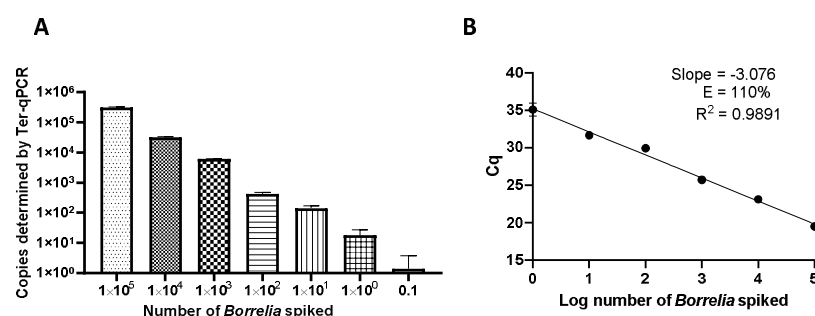


Fig. 4 Performance of the Ter-qPCR estimated by examining human blood spiked with tenfold serial dilutions of *Borrelia* cells (10^5 to 0.1). **A:** bar graph illustrating the number of spike-in *Borrelia* cells and the resulting copy numbers determined by the Ter-qPCR; **B:** Linear regression analysis between the known amount of spiked *Borrelia* cells and the resulting Cq values revealed a strong linear association. The slope, coefficient of correlation (R^2) and efficiency of the reaction (E) are shown. Each dot represents the average value from triplicate repeats along with SD values obtained from two independent experiments.

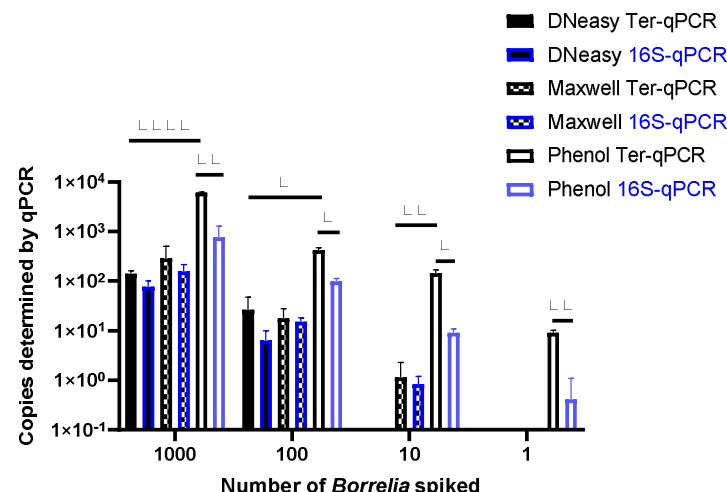


Fig. 5 The Ter-qPCR (bars with black border) and 16S qPCR (bars with blue border) against DNA extracted using three methods from human whole blood spiked with tenfold serial dilutions of *Borrelia* cells (10³ to 1). The phenol method, DNeasy Blood & Tissue Kit and the Maxwell RSC Viral Total Nucleic Acid Purification Kit were compared for DNA extraction. The Phenol method coupled with the Ter-qPCR displayed consistently and significantly higher quantitation compared to the other two methods. Copy numbers obtained from the Ter-qPCR were significantly higher than those from the 16S qPCR. Values shown are the means from triplicate repeats along with SD values obtained from two independent experiments.

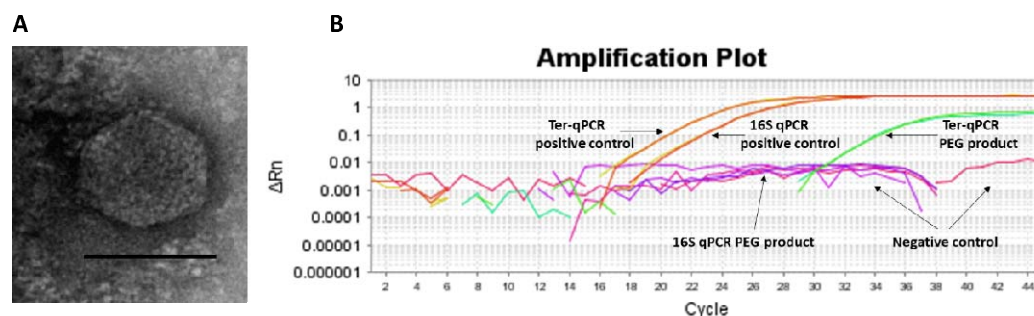


Fig. 6 Spontaneous prophage release from *B. burgdorferi* B31 cultures. **A:** Image of the phage-like particles visualised during TEM analysis of the PEG purified cell-free culture filtrates. **B:** DNA extracted from the PEG product tested positive during the Ter-qPCR, but negative to the 16S qPCR, indicating the spontaneous release of cp32 prophages. 1 ng of DNA from B31 was used as a positive control (as annotated in panel B).

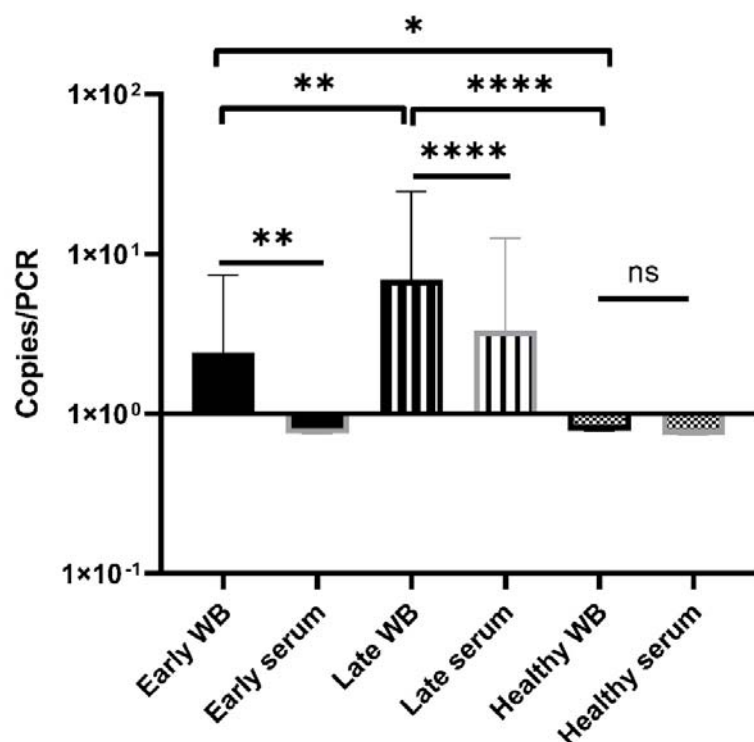


Fig. 7 Bar graph displaying the mean *terL* levels (with SD values) of the three respective study groups from whole blood (black border) and serum (grey border) samples. The Mann-Whitney *U* test was used to compare differences between early LD patients, late LD patients, and healthy volunteers. The *terL* copy numbers obtained from the blood samples of late LD patients are significantly higher larger than those of the early LD patients and healthy volunteers, while the *terL* copy numbers of early LD patients are significantly higher larger than those of healthy volunteers. Statistical significance is denoted as ns ($p > 0.05$), * ($p < 0.05$), ** ($P < 0.01$), *** ($P < 0.001$), and **** ($P < 0.0001$).

Table 1 The prevalence of the *terL* homologs among plasmids residing *Borrelia* species causing Lyme disease (smiley face and 'x' to denote presence and absence of the *terL* homologs, respectively)

Plasmid	<i>B. burgdorferi</i>	<i>B. afzelii</i>	<i>B. garinii</i>	<i>B. bissettii</i>	<i>B. mayonii</i>	<i>B. bavariensis</i>	<i>B. valaisiana</i>	<i>B. finlandensis</i>	<i>B. spielmanii</i>
cp32-1	☺	☺	x	x	☺	☺	x	x	x
cp32-2	☺	x	x	x	x	x	x	x	x
cp32-3	☺	☺	☺	☺	☺	☺	x	☺	x
cp32-4	☺	☺	x	☺	☺	x	x	☺	x
cp32-5	☺	☺	☺	☺	x	☺	☺	x	x
cp32-6	☺	x	☺	☺	☺	x	x	x	x
cp32-7	☺	☺	x	☺	x	☺	x	☺	x
cp32-8	☺	x	x	x	x	x	x	x	x

cp32-9	☺	☺	x	x	x	x	x	x	x
cp32-10	☺	x	☺	x	x	x	☺	x	x
cp32-11	☺	☺	x	☺	x	x	x	x	x
cp32-12	☺	☺	x	x	x	x	x	☺	x
cp32-13	☺	x	x	x	☺	x	x	x	x
cp32-2-7 ^a	x	x	x	x	x	x	☺	x	x
cp32-5+1 ^b	☺	x	x	x	x	x	x	x	x
cp32-5-1 ^c	☺	x	x	x	x	x	x	x	x
cp32-3+10 ^d	☺	x	x	x	x	x	x	x	x
cp30	x	☺	x	x	x	x	x	x	x
lp17	☺	x	x	x	x	x	x	x	x
cp18	☺	x	x	x	x	x	x	x	x
lp25	x	x	x	x	x	☺	x	x	x
lp28-4	x	x	x	x	x	☺	x	x	x
lp54	☺	x	x	x	x	x	☺	☺	☺
lp56	☺	x	x	x	x	x	x	x	x
Total hits	147	25	14	8	8	6	5	5	1

- cp32-2-7 refers to cp32-2 and cp32-7 as both plasmids have the same compatibility³⁰.
- cp32-1+5 refers to a fusion plasmid discovered from *B. burgdorferi* JD1 that is made up of two different full-length cp32-1 and cp32-5 fused together to form a circular plasmid³⁶.
- cp32-1-5 is another version of cp32-1-5 referring to the fusion of cp32-1 and cp32-5 into a single circular replicon³¹.
- cp32-3+10 refers to a fusion plasmid discovered from *B. burgdorferi* B31 that is made up of two different full-length cp32-3 and cp32-10 fused together to form a circular plasmid^{36,83}.

Table 2 *Borrelia* strains used in this study

Lab number	Isolate name	Scientific name	Source	Disease
B31	B31	<i>Borrelia burgdorferi</i>	Sven Bergström	LD
1	1120	<i>Borrelia duttonii</i>	Sven Bergström	RF
2	Her HS1	<i>Borrelia hemsii</i>	Sven Bergström	RF
3	VS185 P9	<i>Borrelia burgdorferi</i>	Sven Bergström	LD
4	NE218	<i>Borrelia valasiana</i>	Sven Bergström	LD
5	ACA1	<i>Borrelia afzelii</i>	Sven Bergström	LD
6	UK filtered	<i>Borrelia burgdorferi</i>	Sven Bergström	LD
7	190 P9	<i>Borrelia garinii</i>	Sven Bergström	LD
8	China23	<i>Borrelia burgdorferi</i>	Sven Bergström	LD
9	CA128	<i>Borrelia bisettii</i>	Sven Bergström	LD
10	FR64b	<i>Borrelia miyamotoi</i>	CDC, USA	RF
11	HT31	<i>Borrelia miyamotoi</i>	CDC, USA	RF

12	CIP 109134	<i>Borrelia burgdorferi</i>	Institut Pasteur	LD
13	CIP 108855T	<i>Borrelia spielmanii</i>	Institut Pasteur	LD
14	CIP 105366T	<i>Borrelia lusitaniae</i>	Institut Pasteur	LD
15	DSM 21467	<i>Borrelia valasiana</i>	DSMZ	LD
16	DSM 10508	<i>Borrelia afzelii</i>	DSMZ	LD
17	DSM 10534	<i>Borrelia garinii</i>	DSMZ	LD
18	DSM 23469	<i>Borrelia bavariensis</i>	DSMZ	LD
19	Pbi	<i>Borrelia bavariensis</i>	Cecilia Hizo-Teufel	LD
20	NT54	<i>Borrelia bavariensis</i>	Cecilia Hizo-Teufel	LD

LD: Lyme disease; RF: Relapsing fever