Structural evolution of an immune evasion determinant shapes Lyme borreliae host tropism

Ashley L. Marcinkiewicz[1#], Kalvis Brangulis[2,3#]*, Alan P. Dupuis, II[1], Thomas M. Hart[1,4†], Maxime Zamba-Campero[1], Tristan A. Nowak[1,5], Jessica L. Stout[1], Inara Akopjana[2], Andris Kazaks[2], Janis Bogans[2], Alexander T. Ciota[1,5], Peter Kraiczy[6], Sergios-Orestis Kolokotronis[7,8,9,10]*, Yi-Pin Lin[1,5]*

[1]Division of Infectious Diseases, Wadsworth Center, NYSDOH, Albany, NY, USA; [2]Latvian Biomedical Research and Study Centre, Riga, Latvia; [3]Department of Human Physiology and Biochemistry, Riga Stradins University, Riga, Latvia; [4]Department of Biological Sciences, SUNY Albany, Albany, NY, USA; [5]Department of Biomedical Sciences, SUNY Albany, Albany, NY, USA; [6]Institute of Medical Microbiology and Infection Control, University Hospital of Frankfurt, Goethe University Frankfurt, Frankfurt, Germany; [7]Department of Epidemiology and Biostatistics, School of Public Health; [8]Institute for Genomics in Health; [9]Division of Infectious Diseases, Department of Medicine, College of Medicine; [10]Department of Cell Biology, College of Medicine, SUNY Downstate Health Sciences University, Brooklyn, NY, USA.

Short title: Structural and evolutionary insights into pathogen host tropism

Key words: Lyme disease, CspZ, *Borrelia*, Factor H, Complement

#These authors have contributed equally to this work.

†current address: Yale University School of Medicine, New Haven, CT

*correspondence: Kalvis Brangulis, Ph.D.
Latvian Biomedical Research and Study Centre, Riga, Latvia LV-1067
Telephone: +371 6780 8200
Email: kalvis@biomed.lu.lv

*correspondence: Sergios-Orestis Kolokotronis, Ph.D.
Department of Epidemiology and Biostatistics. School of Public Health
SUNY Downstate Health Sciences University
450 Clarkson Avenue, MSC43A, Brooklyn, NY 11203, USA
Telephone: +1 718-270-6741
Email: sok@downstate.edu

*correspondence: Yi-Pin Lin, Ph.D.
Wadsworth Center, NYSDOH, Albany, NY 12208
Telephone: +1 518-402-2233
Email: Yi-Pin.Lin@health.ny.gov

ABSTRACT (186 words)

Lyme disease presents a suitable model for the investigation of host tropism – a pathogen's ability to colonize and survive in different host species – since its causative agent, the spirochete *Borrelia burgdorferi* (*Bb*) is transmitted by ticks and maintained in rodent and bird reservoir hosts.

In order to survive in vertebrates and escape from killing by complement, a first-line host immune defense, *Bb* produces the outer surface protein CspZ that binds the complement inhibitor factor H (FH) to promote infection. Protein sequence conservation seems to be linked to FH-binding activity divergence, raising the hypothesis that even minor variation can confer host-specific, FH-binding-mediated infectivity. Our work shows that that this minor variation is located in a loop in the CspZ protein localized in the CspZ-FH binding interface. Our functional experiments prove that this loop promotes bird- or rodent-specific FH-binding activity and infectivity. Swapping loops between rodent- and bird-associated CspZ variants alters their capability to confer host-specific dissemination. We further investigated the evolutionary mechanisms driving the emergence of the CspZ loop-mediated, host-dependent complement evasion. This multifaceted work demonstrates how a single, short protein motif can significantly impact host tropism.

INTRODUCTION

The emergence of most infectious disease outbreaks often involves changes in host tropism, the preferential adaptation of pathogens to selectively invade and persist in hosts (1, 2). Such host tropism is often the result of ongoing host-pathogen interactions (3). Evolution theoretically favors the emergence of host-specializing pathogens, but host-generalist strategies can be advantageous in environments when pathogens have the potential to regularly interact with multiple hosts (4). Specialism vs. generalism can be attributed to polymorphisms within pathogen proteins that differentially interact with host ligands (3). Such polymorphism-mediated host ranges are often thought to require complex host-specific adaptive mechanisms (e.g., (5, 6)) that can be conferred by merely a few amino acids (7-9). Understanding how such minor differences impact diverse host-adapted phenotypes can elucidate the mechanistic insights for the emergence of modern infectious diseases, allowing the development of earlier and more efficient public health interventions.

Species within the *Borrelia* (also known as *Borreliella*) *burgdorferi* sensu lato genospecies complex cause Lyme disease. These spirochetal bacteria are transmitted by ticks and maintained by several reservoir hosts, primarily rodents and birds (10, 11). Lyme borreliae have been genotyped using different polymorphic loci, such as *ospA*, *ospC*, the 16S-23S rRNA intergenic spacer (RST[yping]), and the loci included in multilocus sequence typing (MLST) (12, 13). Laboratory and field studies have shown that not only *Borrelia* species differ in the host species they infect, but that individual genotypes within single spirochete species display distinct preferential host associations, particularly within *B. burgdorferi* sensu stricto (hereafter *B. burgdorferi*), the primary Lyme disease agent in North America (14, 15). This species- and

genotype -specific host selectivity distinguishes Lyme borreliae as a model system for the study of the molecular basis and evolutionary history of host tropism (14, 15).

Lyme borreliae host associations require the spirochetes to transmit from infected ticks to hosts, establish an infection at tick biting sites, and disseminate hematogenously to persist at distal tissues (14, 15). The survival of Lyme borreliae during these discrete events necessitate evasion of multiple host immune responses, including complement, a first-line innate defense mechanism of vertebrate extracellular fluid and blood (16-18). This powerful immune mechanism can be activated by the classical, lectin, or alternative pathways (19). The activation of these pathways results in the cascading cleavage and recombination of multiple complement components, ultimately resulting in phagocytic clearance, inflammation, and pore formation on the pathogen surface by the C5b-9 protein complex to lyse the cells (20). Complement is downregulated by diverse regulatory proteins to inhibit activation, preventing native cell damage in the absence of pathogens (21). For example, the alternative pathway is inhibited by factor H (FH), which is comprised of 20 individual short consensus repeat (SCR) domains (22, 23).

Similar to many bloodborne pathogens that evolved to circumvent complement-mediated clearance (20), Lyme borreliae produce several complement-inhibitory outer surface proteins (16-18). These proteins bind and recruit complement components and/or regulatory proteins on spirochete surface to inactivate complement (16-18). One of these proteins is a FH-binding protein, CspA (also known as Complement Regulator Acquiring Surface Protein 1, CRASP-1) that we demonstrated to confer survival of spirochetes via complement evasion in the tick bloodmeal (24). However, this protein is downregulated after spirochetes infect hosts suggesting a possibilities of other functionally redundant proteins to facilitate spirochete survival at this infection stage. In fact, CspZ (also known as CRASP-2), which binds to the 6th and 7th domains of SCR from FH (SCR6-

7), is upregulated when spirochetes reside in hosts (25, 26). This protein is highly conserved both between and within different Lyme borreliae species (>85% amino acid (aa) identity between species and 98% between *B. burgdorferi* strains) (27-30). We and others have previously reported that FH-binding activity of one CspZ variant promotes spirochete survival in vertebrate sera, and consequently promotes host infectivity when non-physiologically relevant infection routes are used (i.e., not tick-transmitted) (31, 32). However, CspZ variants differ in their human FH-binding activity (27, 28), and vertebrate complement components or regulatory proteins vary between host taxa (e.g., ~40% aa identity between mammalian and avian FH) (33). These findings raise several intriguing questions: Could such a minor divergence among CspZ variants confer host-specific differences in FH-binding activity, resulting in varying host infectivity phenotypes? Further, if those minor divergences are indeed a determinant of host tropism, how did such divergences evolve to impact Lyme borreliae-host associations?

In this study, we solved the high-resolution structure of the CspZ-human FH complex, identified a polymorphic CspZ motif within the FH-binding interface, and defined its contribution to host-specific FH-binding activity. We then examined the role of this motif in dictating strain-specific, host-dependent dissemination using mice and quail as rodent and avian models, respectively. Paired with evolutionary analyses of *cspZ*, we further elucidated how the evolutionary history behind a minor divergence in an immune evasion determinant can impact host tropism.

RESULTS

A polymorphic CspZ loop structure defines human FH-binding activity. We set to identify the regions driving the differences in human FH-binding activity among CspZ variants. We first

pinpointed the amino acids that are involved in FH-binding by co-crystalizing the SCR6-7 of human FH and a CspZ variant with human FH-binding activity from *B. burgdorferi* strain B408 ($CspZ_{B408}$) (28). The resulting crystal structure at 2.59Å showed an extensive binding interface between $CspZ_{B408}$ and both SCR6 and SCR7 (**Fig. 1A**). Specifically, Asp47, Tyr50, Asn51, Thr54, Asn58, and Thr62 in helix-B, Arg142 in helix-F, Asn183 from helix-G, and Tyr214 in helix-I of $CspZ_{B408}$ interacted with SCR7 (**Fig. 1B, Table S1**). Additionally, Asp71, from the loop between helix-B and -C, and Asp73 and Ser75 from the same region, bound to SCR7 and SCR6, respectively (**Fig. 1B, Table S1**).

To locate the residues that impact CspZ FH-binding activity, we aligned the sequences of $CspZ_{B408}$ with another CspZ variants that binds to human FH (i.e., CspZ from *B. burgdorferi* B31, $CspZ_{B31}$), as well as CspZ that lacks human FH-binding ability from strain B379 ($CspZ_{B379}$)(28). All the aforementioned residues involved in human FH-binding are conserved among all three variants, except Asn51 and Asp71 (**Fig. S1A**). Such conservation also reflects to the three-dimensional overlayed structure of the previously resolved $CspZ_{B31}$ (30) with our newly-resolved crystal structures of $CspZ_{B379}$ and $CspZ_{B408}$ (2.10 and 2.45Å, respectively) (**Fig. 1A)**. For those two non-conserved amino acids, Asn51 of $CspZ_{B408}$ is substituted to serine in $CspZ_{B379}$, but the superimposed structure suggests this would not inhibit human FH-binding in $CspZ_{B379}$ (**Fig. 1B**). Asp71 of $CspZ_{B408}$ interacts with Lys388 and Tyr390 of SCR6, but this residue is part of an insertion that is not present in $CspZ_{B31}$ and $CspZ_{B379}$ (**Fig. 1B**). Therefore, these non-conserved residues alone cannot explain the differential human FH binding ability of these CspZ variants. However, the C-terminal helix-B and the following loop region between helix-B and -C do harbor polymorphisms (**Fig. 1A)**, with the unique duplication of the last residues in helix-B (Ile60, Met61, Thr62, and Tyr63) of $CspZ_{B379}$ (**Fig. S1A**), resulting in an extended helix-B in this variant (**Fig.**

1B). Such a structural extension leads to steric hinderance between $CspZ_{B379}$ and human FH to selectively prevent human FH-binding activity by $CspZ_{B379}$. Overall, our results reveal variant-specific CspZ structural differences in a loop and the adjacent helix-B (hereafter, "loop structures"), providing mechanistic insights into the CspZ-mediated, polymorphic human FH-binding activity.

CspZ loop structures determine host-specific FH-binding activity. Similar to CspZ, SCR6-7 is divergent among vertebrate species (**Fig. S1B**), suggesting a role of the CspZ loop structures in dictating host-specific FH-binding activity. We found that $CspZ_{B31}$ and $CspZ_{B408}$ –but not $CspZ_{B379}$– bind to mouse FH, whereas $CspZ_{B31}$ and $CspZ_{B379}$ –but not $CspZ_{B408}$– bind to quail FH (**Fig. 2A and S2, Table S2**). We then searched for structural evidence of this FH-binding activity by superimposing the complex structure of $CspZ_{B408}$-human FH with $CspZ_{B31}$, $CspZ_{B379}$, as well as the previously resolved structures of mouse SCR6-7, and the AlphaFold-predicted structure of quail SCR6-7 (pLDDT > 70) (34, 35) (**Fig. 2C**). We noted the similar tertiary structures between mouse and human SCR6-7, consistent with the ability of FH from both origins selectively binds to $CspZ_{B31}$ and $CspZ_{B408}$ (28) (**Table S2**). Similar to human FH, the $CspZ_{B379}$-specific duplication by the extended C-terminal helix-B showed potential structural hinderance in the mouse FH-binding interface (**Fig. 2D**). However, compared to a loop region within human and mouse SCR7, the equivalent region in quail SCR7 (median pLDDT: 92) is positioned away from the CspZ-FH-binding interface, allowing sufficient space for extended helix-B of $CspZ_{B379}$ to interact with quail FH (**Fig. 2D**). Further, the crystal structure of $CspZ_{B408}$-SCR6-7 superimposed with the predicted structure of quail SCR6-7 showed that Asp71 within the loop of $CspZ_{B408}$ may collide with the opposite loop region of quail FH, while the equivalent region in mouse SCR7 would not interfere with the binding of $CspZ_{B408}$ (**Fig. 2D**).

These structural differences among FH from mouse/human vs. quail, pairing with the respective impacting AAs from different CspZ variants, suggest a possibility of the loop structures of CspZ to determine CspZ allelically different, host-specific FH-binding activity. To test this possibility, we swapped the loop regions of the two CspZ variants with distinct host-specific FH-binding ability to generate two chimeric proteins: $CspZ_{B379}L_{B408}$ has the backbone of a quail FH-binder ($CspZ_{B379}$) and the loop structures of a mouse FH-binder ($CspZ_{B408}$), whereas $CspZ_{B408}L_{B379}$ has the backbone of $CspZ_{B408}$ and the loop structures of $CspZ_{B379}$ (**Fig. S1A**). We found that $CspZ_{B379}L_{B408}$ and $CspZ_{B408}L_{B379}$ selectively bound to mouse and quail FH, respectively, demonstrating the CspZ loop structures are a determinant of host-specific FH-binding activity (**Fig. 2B and S2**).

The CspZ loop structures dictate spirochete strain- and host-specific complement inactivation. We next examined the host-dependent complement inactivation of these CspZ variants produced on the spirochete surface. We thus obtained a wild-type *B. burgdorferi* strain B31-A3 (WT B31-A3) and a *cspZ*-deficient mutant strain in this background harboring an empty vector (Δ*cspZ*). We complemented this mutant with a plasmid encoding $cspZ_{B31}$ ($pCspZ_{B31}$) (32), $cspZ_{B379}$ ($pCspZ_{B379}$), or $cspZ_{B408}$ ($pCspZ_{B408}$), or one of the two loop-swapped mutants ($pCspZ_{B379}L_{B408}$ and $pCspZ_{B408}L_{B379}$). We verified there was a similar generation time of these strains (**Table S3A**), and that there were no differences in the CspZ surface production level (**Fig. S3B**). Additionally, surface-produced CspZ bound FH in the same host-specific manner as the recombinant proteins (**Fig. S4**). We then determined the ability of each of these CspZ variants and mutants to inactivate mouse and quail complement by measuring the deposition levels of mouse C5b-9 and quail C8, respectively, in the presence of sera from each animal using flow cytometry

(**Fig. 3A-B**). All strains had levels of mouse C5b-9 or quail C8 deposition significantly lower than the high passage, non-infectious, and mouse and quail complement-susceptible control strain *B. burgdorferi* B313 (**Fig. 3C-D**)(26). The Δ*cspZ* strain had significantly greater levels of mouse C5b-9 and quail C8 deposition than WT B31-A3 or pCspZ$_{B31}$ (**Fig. 3C-D**)(32, 36). Compared to Δ*cspZ*, pCspZ$_{B379}$ had indistinguishable levels of mouse C5b-9 deposition but significantly lower levels of quail C8, whereas pCspZ$_{B408}$ harbored significantly lower levels of mouse C5b-9 but indistinguishable levels of quail C8 (**Fig. 3C-D**). Further, pCspZ$_{B379}$L$_{B408}$ recruited significantly lower levels of mouse C5b-9 but indistinguishable levels of quail C8 than Δ*cspZ*, while pCspZ$_{B408}$L$_{B379}$ bound indistinguishable levels of mouse C5b-9 but significantly lower levels of quail C8 than Δ*cspZ* (**Fig. 3C-D**). These results correspond with these variants' host-specific FH-binding activity, suggesting that the CspZ loop structures-driven host-specific FH-binding activity confers host-specific complement inactivation.

We also measured the survival of these strains in rodent and quail sera. Note that sera from white-footed mice (*Peromyscus leucopus*), rather than house mice (*Mus musculus)*, were used to represent rodent sera, as complement of house mouse is labile *in vitro*, leading to inconsistent results (37, 38). Both WT B31-A3 and pCspZ$_{B31}$ survived in white-footed mouse and quail sera more efficiently than Δ*cspZ* (**Fig. 3E-F**) (32). Compared to Δ*cspZ*, pCspZ$_{B379}$ survived at significantly greater levels in quail but not white-footed mouse sera, whereas pCspZ$_{B408}$ survived at significantly higher levels in white-footed mouse but not quail sera (**Fig. 3E-F**). Relative to Δ*cspZ*, pCspZ$_{B379}$L$_{B408}$ survived significantly higher in white-footed mouse but not in quail sera, whereas pCspZ$_{B408}$L$_{B379}$ survived significantly higher in quail but not in white-footed mouse sera (**Fig. 3E-F**). However, these differences were not observed in the presence of sera treated with Cobra Venom factor (CVF) or *Ornithodorus moubata* complement inhibitor (OmCI) that deplete

functional rodent and quail complement, respectively (**Fig. S5**) (39, 40), suggesting that CspZ loop structures determine spirochete host-specific serum survival.

The CspZ loop structures define host-specific, complement-dependent, early onset dissemination. We further evaluated if the CspZ loop structure variants determine host infectivity by allowing *Ixodes scapularis* nymphs carrying similar loads of each of the aforementioned *B. burgdorferi* strains to feed on mice and quail (**Fig. S6A, S6D**). The burdens of Δ*cspZ* were significantly lower than WT B31-A3 at 7- and 10-days post feeding (dpf) in mouse blood, and at 10 dpf in distal tissues (joints, heart, and bladder), and at 9 dpf in quail blood and distal tissues (brain and heart) (**Fig. S7**). There were no differences in burdens by a later time point (i.e., 14 dpf, **Fig. S7**). We thus performed subsequent experiments on 10 and 9 dpf in mouse and quail, respectively, for the other of *cspZ*-complemented strains. In mice, we found indistinguishable spirochete burdens at initial infection sites (tick bite sites) between all seven strains tested at 10dpf (**Fig. 4A**). The strains pCspZ$_{B31}$, pCspZ$_{B408}$, and pCspZ$_{B379}$L$_{B408}$, but neither pCspZ$_{B379}$ nor pCspZ$_{B408}$L$_{B379}$, colonized mouse blood and distal tissues at significantly higher levels than Δ*cspZ* (**Fig. 4B-E**). At 9dpf in quail, we found no significantly different spirochete burdens at tick bite sites among strains. However, pCspZ$_{B31}$, pCspZ$_{B379}$, and pCspZ$_{B408}$L$_{B379}$, but neither pCspZ$_{B408}$ nor pCspZ$_{B379}$L$_{B408}$, colonized quail blood and distal tissues at significantly greater levels than Δ*cspZ* (**Fig. 4F-I**). There were no significantly different burdens of any strain in any tissues from mice and quail with complement deficiency (C3$^{-/-}$ mice or OmCI-treated quail) (**Fig. S8**). These results demonstrate that the CspZ loop structures define complement-dependent, spirochete strain- and host-specific early onset dissemination.

The population-wide CspZ loop structures evolved from a variant with versatile FH-binding ability. To investigate the evolutionary history of the CspZ loop structures, we mined *B. burgdorferi cspZ* from NCBI GenBank and sequence read archive (totaling 174 high-quality *cspZ* isolates originated from ticks, reservoir hosts, and patients, across North America, Europe, and Asia). Phylogenetic analysis revealed three lineages with uncertain grouping among them, suggesting insufficient variation of *cspZ* to allow for well-established phylogenetic groups to emerge (data not shown). However, a haplotype-based phylogenetic network distinctly separated these three individual lineages, each of which contains one of the above-tested *cspZ* alleles: $cspZ_{B31}$, $cspZ_{B379}$, or $cspZ_{B408}$ (**Fig. S9**). Isolates within the same lineage had over 99% sequence identity and the same loop structure compared to $cspZ_{B408}$, $cspZ_{B379}$, or $cspZ_{B31}$ (containing the duplication, insertion, or neither at the loci encoding loop structure, respectively) (**Fig. 5A**). We pinpointed single amino acid polymorphisms (SAPs) in the isolates and found none of the SAPs within each lineage were located in the FH-binding interface (**Table S1**), with the exception of a SAP in one isolate in the $CspZ_{B379}$ lineage (**Table S4**)). Rather, the majority of the SAPs were between lineages, and were located in the loop structures that are driving the FH-binding abilities. While this present study has been the only one to investigate the abilities of CspZ variants to bind house mouse and quail FH, the human FH-binding abilities of 13% of these 174 isolates have been evaluated (27, 28, 32, 41). Incorporating these findings with the phylogeny, we found a 100% correlation of known host-specific FH-binding ability with lineage (**Fig 5A**). These results suggest each CspZ lineage with distinct loop structures can be linked to CspZ-specific, host-dependent FH-binding ability.

Additionally, we explored the evolutionary mechanisms that could have led to this minor loop structure-dependent FH-binding specificity. No evidence was found for recombination or past

changes to the effective population size (data not shown). Although there was no evidence of selection for *cspZ* in its entirety (data not shown), several codons were undergoing either positive or negative selection (**Fig. S1A**). We reconstructed the possible last common ancestor states ("LCAS") of the entire *B. burgdorferi* CspZ population using the chromosome mutation rate estimated in a previous study (42). We found that the diversification of the three lineages was estimated to be 261 to 784 years ago (y.a.) whereas the diversification of the LCAS likely occurred approximately 2166y.a. (896-5718 $HPD_{5\text{-}95\%}$) (**Table S5)**. The LCAS variants have a loop region resembling that of $CspZ_{B31}$ both in AAs and structure (**Fig. S1A, S10**), and versatilely bound to human, mouse, and quail FH, similar to $CspZ_{B31}$ (**Fig. 5B, Table S6**). These results suggest the diversification of CspZ is arisen relatively recently from a generalist variant with versatile FH-binding features.

DISCUSSION

Host tropism is an outcome of ongoing host-pathogen interactions (15). In vector-borne zoonotic pathogens, such interactions include vector-to-host transmission, pathogen dissemination and persistence, and host-to-vector acquisition (43). For Lyme borreliae to survive throughout these infection steps, spirochetes need to overcome numerous host immunological mechanisms, including complement-mediated killing (17, 18, 44). In fact, Lyme borreliae inactivate complement in the tick blood meal for tick-to-host transmission and in the host bloodstream for dissemination and persistence (32, 45-47). In this study, we examined this inactivation conferred by a Lyme borreliae FH-binding protein, CspZ. We showed early colonization defects of a tick-introduced, *cspZ*-deficient mutant at distal tissues, defining this polymorphic protein as a contributor to the early stages of dissemination. Superimposing the crystal structures of three

polymorphic CspZ variants from genotypically-diverse *B. burgdorferi* strains revealed a variable motif of a short AA stretch (30). Our newly-resolved and software-predicted complex structures of CspZ with human, mouse, and quail SCR6-7 further linked those loop-encoded variable residues to distinct host-specific FH binding activity. These results are congruent with the ability of CspZ variants and their associated mutants with swapped loop structures to promote host-specific FH-binding activity, complement evasion, and dissemination, defining this CspZ loop structure as a determinant of host tropism. Our findings thus demonstrate the concept that minor variation could functionally impact host-adapted phenotypes, potentially modulating host tropism (7-9). Our findings can also be attributed to a structurally unique FH-binding mechanism, as the FH-binding interface of CspZ is significantly different from that of the only other structurally-characterized SCR6-7-binding pathogen protein, *Neisseria meningitidis* Fhbp (48)(**Fig. S11**). Taken together, these results provide a platform of using structure-guided approaches in identifying pathogens' host tropism determinants and their biological functions.

We also integrated an evolutionary approach to investigate the emergence of such host tropism determinants, an approach whose importance has been recently demonstrated to understand and track modern disease emergence (e.g., (49, 50)). Our findings demonstrate that the *B. burgdorferi* CspZ variants diversified from a last common ancestor that could bind human, mouse, and quail FH. This diversification occurred relatively recently (**Table S10**), compared to the previous finding showing that *B. burgdorferi* as a species evolved about 60,000 years BP, and the other FH-binding protein, CspA, diversified about 25,000 years y.a. (42, 45). The fact that CspZ has no orthologs in non-*Borrelia* organisms also supports the young age of this protein, suggesting CspZ may be undergoing weaker purifying and more variable selective pressure (30, 51). We did not detect selective pressure on the entirety of the *cspZ* locus, but rather only in a few

individual codons, none of which were in the loop structures driving FH-binding phenotypes. The loop structures, which are the most variable part of these CspZ variants, contain indels, so the detection of gene-wide selective pressures was likely masked by the exclusion of gaps by the selection tests. In fact, other polymorphic anti-complement loci, particularly *ospC*, are undergoing balancing selection (52), (53). Both Multiple Niche Polymorphism (MNP) and Negative Frequency Dependent Selection (NFDS) have been proposed as mechanisms driving this selection. On the one hand, NFDS proposes that a high diversity of alleles emerge through the negative correlation between spirochete fitness and the frequency of any allele in the overall population (54). In Lyme borreliae, NFDS may explain the greater antibody-mediated clearance of spirochete strains harboring *ospC* alleles of higher frequency loci (55, 56). On the other hand, MNP proposes that the diversity of alleles/genotypes is maintained by fitness variation of these spirochetes across reservoir hosts (53, 57). Though it is possible that lack of selective pressure detection reflects the neutral selection experienced in *cspZ*, as a systemically redundant anti-complement protein (58), the possibility that *cspZ* is undergoing the same selective pressures as other functionally redundant proteins (e.g., *ospC*) cannot be excluded. In fact, balancing selection on one locus can increase diversity in another genetically-linked locus (59), and there is genetic linkage between CspZ types and *ospC* types (27, 28, 36). This may raise an intriguing possibility that the balancing selection on *ospC* may allow diversification through drift in other functionally redundant genes like *cspZ* without deleterious effect to the spirochetes, which warrants further investigations.

Some Lyme borreliae species or strains carry determinants that promoting host-specific phenotypes that differ from the known host range of these species or strains (36, 60-62). For example, the *B. burgdorferi* strain 297 is highly infectious in mice, but its CspZ loop-structures are identical to those in $CspZ_{B379}$ from the strain B379 that is not mouse adapted (36, 62). This

reflects the fact that most conclusions from this study were drawn by using spirochete strains with the same genetic background but producing the CspZ variants of interest, which may not account for polygenic contributions of other proteins in different strains (36). One possibility that addresses this discrepancy is that host-adapted phenotypes are conferred by not only CspZ but other anti-complement proteins, and the functional contribution from each of these proteins differ (and/or are host-dependent). In fact, the colonization defects of the *cspZ*-deficient *B. burgdorferi* were not found in the initial infection site at early infection stages, or in the distal tissues at later timepoints, consistent with the concurrent production of other FH-binding proteins (25, 45, 63-65). Additionally, non-FH-binding anti-complement proteins (66-68) and/or tick salivary proteins with anti-complement functions (69-71) could also drive complement-based phenotypes. Further, pathogen evasion to complement-independent mechanisms (e.g., antibodies), or even the host immune evasion-independent adaptive phenotypes (e.g., adhesion), can occur simultaneously with the anti-complement mechanisms (72), all contributing to the overall host-adapted phenotypes (6). All these confounding factors may complicate the delineation of the roles for CspZ at later stages of infection. Thus, our results do not rule out the possibility of CspZ-mediated complement evasion at distal tissues during post-dissemination timepoints, but rather highlight the importance of this protein immediately after spirochetes begin to disseminate. Despite such caveats, the strength of these isogenic strains needs to be emphasized in delineating the roles of a particular pathogen determinant from an enormous complexity of functionally redundant proteins (24, 45, 73, 74). Such potential polygenic or multi-factorial, and host-specific phenotypes should consider the complexity of numerous determinants across different mechanisms in defining the effect on host tropism.

In this study, we used Lyme borreliae as a model to apply structural, microbiological, and evolutionary approaches to identify a determinant of host-adapted phenotypes and the specific mechanism impacting host tropism. It should be noted that distinct *B. burgdorferi* infectivity have been observed for host species within the same taxa (e.g., house mice vs. the reservoir white-footed mice) (6, 75, 76). Thus, we do not intend to extend the host-specific phenotypes from these laboratory models to every individual reservoir animal in the same host taxa. Rather, our results establish a platform with more accessibility of tools to facilitate the identification of host tropism determinants and their underlying mechanisms. Such results build the foundation to further examine similar concepts in reservoir animals to recapitulate field findings in understanding the mechanisms of host tropism in a controllable laboratory setting. The information and the platform established in this multi-disciplinary study establishes greater insights of pathogen-host interactions, facilitating the understanding of host tropism as the cause of newly emergent infectious diseases.

MATERIALS AND METHODS

Ethics statement. All mouse and quail experiments were performed in strict accordance with all provisions of the Animal Welfare Act, the Guide for the Care and Use of Laboratory Animals, and the PHS Policy on Humane Care and Use of Laboratory Animals. The protocol was approved by the Institutional Animal Care and Use Committee (IACUC) of Wadsworth Center, New York State Department of Health (Protocol docket number 19-451). All efforts were made to minimize animal suffering.

Mice, quail, ticks, bacterial strains, animal sera, OmCI, and FH. BALB/c and Swiss Webster mice were purchased from Taconic (Hudson, NY). $C3^{-/-}$ mice in the BALB/c background were generated from the $C3^{-/-}$(C57BL/6) from Jackson Laboratory (Bar Harbor, ME) as described (24). *Coturnix* quail were purchased from Cavendish Game Birds Farm (Springfield, VT). *Ixodes scapularis* tick larvae were purchased from National Tick Research and Education Center, Oklahoma State University (Stillwater, OK) or obtained from the CDC through BEI Resources (Manassas, VA).

The *Escherichia coli*, *Pichia pastoris* and *Borrelia* strains used in this study are described in **Table S7**. *E. coli* strains DH5α, BL21(DE3), and derivatives were grown in LB broth or agar, supplemented with kanamycin (50µg/ml), ampicillin (100µg/ml), or no antibiotics as appropriate. *P. pastoris* strain X-33 was grown on YPD plates supplemented with zeocin (800µg/ml) or BMGY medium supplemented with 1% methanol. All *B. burgdorferi* strains were grown in BSK-II completed medium supplemented with kanamycin (200µg/mL), streptomycin (50µg/mL), or no antibiotics (**Table S7**).

Mouse FH was purchased from MyBiosource (San Diego, CA). Quail FH and recombinant OmCI proteins were generated as described previously (24, 32, 36, 39). The mouse and quail sera were obtained from Southern Biotech, Inc (Birmingham, AL) and Canola Live Poultry Market (Brooklyn, NY), respectively. The sera from white-footed mice were obtained previously (6). Prior to being used, all these sera were screened for antibodies against the C6 peptide of the *B. burgdorferi* protein VlsE (77) with the C6 Lyme ELISA kit (Diamedix, Miami Lakes, FL) to ensure the mice did not have prior exposure to *B. burgdorferi*.

Generation of recombinant CspZ proteins and recombinant human FH SCR6-7. To generate recombinant CspZ proteins for crystallization, *cspZ*$_{B379}$ (GenBank: FJ911671.1) and *cspZ*$_{B408}$ (GenBank: FJ911677.1) were amplified by PCR from the genomic DNA of *B. burgdorferi* strain B379 and B408 using the primers listed in **Table S8**. Note that B408 has two copies of *cspZ*, but they are functionally identical with only a single synonymous SNP at nucleotide 699 (36, 78). Based on the prediction by SignalP 4.1 (79) and according to our previous structural data from CspZ$_{B31}$ (30), the lipoprotein signal peptide (residues 1-22) was excluded from the amplified gene. The introduced *NcoI* and *NotI* restriction sites were used for ligation of the amplified fragments into the pETm-11 expression vector which contains the coding region for an N-terminal 6xHis tag and a tobacco etch virus (TEV) protease cleavage site. Expression in *E. coli*, purification by affinity chromatography, and 6xHis tag cleavage by TEV protease of both proteins CspZ$_{B379}$ and CspZ$_{B408}$ were performed similarly as described previously for CspZ$_{B31}$ (30). The purified and cleaved proteins were buffer exchanged into 10 mM Tris-HCl (pH 8.0) and concentrated to 11 mg/ml using an Amicon centrifugal filter unit (Millipore, Burlington, MA, USA).

To produce recombinant human FH for crystallization, the gene encoding the SCR6-7 of human FH was synthesized by BioCat GmbH (Heidelberg, Germany) and cloned into pPICZαA vector behind the α-factor secretion signal using *XhoI* and *NotI* restriction sites in a way to restore the Kex2 signal cleavage site. The plasmid was linearized with *PmeI* and transformed by electroporation into *Pichia pastoris* (reassigned as *Komagataella phaffii*) strain X-33. Transformants were obtained on YPD agar plates containing 800µg/ml of the antibiotic zeocin. The selected clone was cultivated 24-h in BMGY medium at 30°C with aeration (250rpm) following addition of 1% methanol daily, and cultivation was continued for three more days. The cell pellet was removed by low-speed centrifugation. Supernatant was buffer-exchanged into

50mM sodium phosphate (pH 6.0) by Sephadex G-25 Fine column (bed volume 360ml) (Cytiva, Marlborough, MA, USA) in 100ml portions at a flow rate of 20ml/min. Two liters of supernatant was passed through the CaptoS Improved Resolution column (bed volume 20ml) (Cytiva, Marlborough, MA, USA) and bound material was eluted with a linear salt gradient at a flow rate of 6ml/min. Target protein fractions were selected based on SDS-PAGE. The relevant fractions were pooled and buffer-exchanged into 20mM Tris-HCl (8.0), 50mM NaCl and 10mM NaH_2PO_4 using an Amicon filter device (Millipore, Burlington, MA, USA).

To generate recombinant $CspZ_{B379}$ and $CspZ_{B408}$ for the studies other than crystallization, the region encoding $cspZ_{B379}$ or $cspZ_{B408}$ without the signal peptide was amplified as described above and engineered to encode *BamHI* and *SalI* sites at the 5' and 3' ends, respectively, allowing subsequent cloning into the pJET cloning vector (Thermo Fisher Scientific, Waltham, MA). These pJET-derived plasmids encoding $cspZ_{B379}$ or $cspZ_{B408}$ were used as template for site-directed, ligase-independent mutagenesis (SLIM) (**Table S7**) to generate plasmids producing $CspZ_{B379}$-L_{B408} and $CspZ_{B408}$-L_{B379} (80). After verifying the sequences of all the plasmids (Wadsworth ATGC facility), the DNA fragments were subsequently excised using *BamHI* and *SalI* and then inserted into the same sites in pGEX4T2 (GE Healthcare, Piscataway, NJ) (32). The pGEX4T2-derived plasmids were then transformed into the *E. coli* strain BL21(DE3). The GST-tagged CspZ proteins were produced and purified by affinity chromatography. These proteins were verified for their secondary structures not impacted by the mutagenesis using CD (**Fig. S12**), as described in the section "Circular dichroism (CD) spectroscopy."

To generate recombinant CspZ from the last common ancestor states, pET-28a+ encoding these states flanked by *BamHI* and *SalI* sites at the 5' and 3' ends, respectively, were cloned (Synbio Technologies, Monmouth Junction, NJ). The plasmids were transformed into the *E. coli*

strain BL21(DE3), and the His-tagged CspZ proteins were produced and purified by affinity chromatography.

Crystallization and structure determination. For crystallization of $CspZ_{B379}$ and $CspZ_{B408}$, 96-well sitting drop plates were set using a Tecan Freedom EVO100 workstation (Tecan Group, Männedorf, Switzerland) by mixing 0.4µl of protein with 0.4µl of precipitant using the 96-reagent sparse-matrix screens JCSG+ and Structure Screen 1&2 (Molecular Dimensions, Newmarket, UK). The crystals for $CspZ_{B379}$ were obtained in 0.2M Ammonium citrate and 24% PEG 3350. For $CspZ_{B408}$, the crystals were formed in 0.2M potassium acetate, 0.1M Tris-HCl (pH 8.0) and 28% PEG 3350. Prior to the data collection, the crystals were frozen in liquid nitrogen. An additional 20% glycerol was used as a cryoprotectant for $CspZ_{B379}$ crystals, whereas the respective precipitant with an additional 14% glycerol was used as cryoprotectant for $CspZ_{B408}$ crystals.

$CspZ_{B408}$ (4mg/ml) and human SCR6-7 (3mg/ml) were mixed together at a molar ratio of 1:2 and loaded on a HiLoad 16/600 Superdex 200 prep grade column (GE Healthcare, Chicago, IL, USA) pre-equilibrated with 20mM Tris-HCl (8.0), 50mM NaCl and 10mM NaH_2PO_4. The flow rate was set to 2ml/min. Size exclusion chromatography resulted in one major peak containing the complex, confirmed by SDS-PAGE. Crystallization was set as described earlier for $CspZ_{B379}$ and $CspZ_{B408}$ by mixing 0.4µl of protein with 0.4µl of precipitant and using the 96-reagent sparse-matrix screens. The crystals for $CspZ_{B408}$-SCR6-7 complex were obtained in 0.2M Zinc acetate, 0.1M imidazole (pH 7.4) and 10% PEG 3000. Crystals were frozen in liquid nitrogen by using 20% glycerol as a cryoprotectant.

Diffraction data for $CspZ_{B379}$, $CspZ_{B408}$ and $CspZ_{B408}$-SCR6-7 complex were collected at the MX beamline instrument BL 14.1 at Helmholtz-Zentrum, Berlin (81). Reflections were

indexed by XDS and scaled by AIMLESS from the CCP4 suite (82-84). Initial phases for $CspZ_{B379}$ and $CspZ_{B408}$ were obtained by molecular replacement using Phaser (85), with the crystal structure of the orthologous protein $CspZ_{B31}$ was used as a searching model (97% sequence identity, PDB: 4CBE). For $CspZ_{B408}$-SCR6-7 complex, the phases were determined using $CspZ_{B408}$ (PDB: 7ZJK, RMSD 0.98 Å) and human FH SCR6-7 (PDB: 4AYD-A, RMSD 0.98 Å) as the searching models. After molecular replacement, the protein models were built automatically in BUCCANEER (86). The crystal structures were improved by manual rebuilding in COOT (87). Crystallographic refinement was performed using REFMAC5 (88). A summary of the data collection, refinement and validation statistics for $CspZ_{B379}$, $CspZ_{B408}$ and $CspZ_{B408}$-SCR6-7 complex are given in **Table S9.**

Protein 3D structure prediction using AlphaFold. AlphaFold v2.0 (34) was used to predict the 3D structure for quail FH SCR6-7 extrapolated from the sequences of *Coturnix japonica* complement FH (GenBank: XM_015869474.2). Structure prediction with AlphaFold v2.0 was performed according to the default parameters as indicated at the website https://github.com/deepmind/alphafold/ running on AMD Ryzen Threadripper 2990WX 32-Core; 128 GB RAM; 4 x NVIDIA GeForce RTX 2080, and using the full databases downloaded on 2021-09-25. For further structural analysis, only the predicted structure with the highest confidence was used (as ranked by using LDDT (pLDDT) scores).

Circular dichroism (CD) spectroscopy. CD analysis was performed on a Jasco 810 spectropolarimeter (Jasco Analytical Instrument, Easton, MD) under nitrogen. CD spectra were measured at room temperature (RT, 25°C) in a 1mm path length quartz cell. Spectra of each of the

CspZ proteins (10μM) were recorded in phosphate based saline buffer (PBS) at RT, and three far-UV CD spectra were recorded from 190 to 250nm for far-UV CD in 1nm increments. The background spectrum of PBS without proteins was subtracted from the protein spectra. CD spectra were initially analyzed by the software Spectra Manager Program (Jasco). Analysis of spectra to extrapolate secondary structures were performed using the K2D3 analysis programs (89).

ELISAs. Quantitative ELISA was used to determine FH-binding by CspZ proteins, or ancestral proteins, as described previously (24, 90), with the following modifications: Mouse anti-GST tag or mouse anti-His tag 1:200× (Sigma-Aldrich) and HRP-conjugated goat anti-mouse IgG 1:2,000× (Seracare Life Sciences) were used as primary and secondary antibodies, respectively, to detect the binding of GST- or histidine-tagged proteins.

Surface Plasmon Resonance (SPR). Interactions of CspZ proteins with FH were analyzed by SPR using a Biacore T200 (Cytiva, Marlborough, MA). Ten micrograms of mouse or quail FH were conjugated to a CM5 chip (Cytiva) as described previously (90). For quantitative SPR experiments, 10μL of increasing concentrations (0.08, 0.03125, 0.0125, 0.5, 2μM) of each of the CspZ proteins were injected into the control cell and the flow cell immobilized with FH at 10μl/min, 25°C. To obtain the kinetic parameters of the interaction, sensogram data were fitted by means of BIAevaluation software version 3.0 (GE Healthcare), using the one step biomolecular association reaction model (1:1 Langmuir model), resulting in optimum mathematical fit with the lowest Chi-square values.

Shuttle vector construction and plasmid transformation into *B. burgdorferi*. "Loop swapped" CspZ variants (i.e., $CspZ_{B379}L_{B408}$ and $CspZ_{B408}L_{B379}$) were designed based on the full-length sequences (B379 accession: OM643341; B408: accession: OM643340) and purchased as double-stranded DNA fragments flanked by *BamHI* and *SalI* on the 5' and 3', respectively (Integrated DNA Technologies, Inc., Coralville, IA). B31-A3Δ*cspZ* was complemented with these variants, or with native CspZ from B379 and B408 flanked by the same restriction enzyme sites (**Table S7**), in the same manner as the previously published strains of B31-A3Δ*cspZ*/pKFSS and B31-A3Δ*cspZ*/p$CspZ_{B31}$ (32). The plasmid profiles of these spirochetes were examined to ascertain identical profiles between these strains and their parental strain B31-A3 (91). The generation time of these transformants was calculated as previously described (32).

Flow cytometry. CspZ production and FH-binding on spirochete surface were determined as described (24), including blood-treatment to induce the production of CspZ (32). To determine the levels of mouse C5b-9 or quail C8 deposition on the surface of spirochetes, mouse or quail sera were incubated with $1x10^7$ spirochetes in PBS at a final concentration of 20% at 25ºC for one hour and the detection procedure has been described previously (36). Basically, after incubation, spirochetes were washed then resuspended in HBSC-DB (25mM Hepes acid, 150mM sodium chloride, 1mM $MnCl_2$, 1mM $MgCl_2$, 0.25mM $CaCl_2$, 0.1% glucose, and 0.2% BSA). Rabbit anti-mouse C5b-9 polyclonal IgG (1:250x) (Complement Technology, Tyler, TX) or mouse anti-quail C8 polyclonal sera (1:250x) (36) were used as the primary antibodies. An Alexa 647-conjugated goat anti-rabbit (ThermoFisher) or a goat anti-mouse IgG (ThermoFisher) (1:250x) was used as the secondary antibody. After staining, the spirochetes were fixed with 0.1% formalin. The

resulting fluorescence intensity of spirochetes was measured and analyzed by flow cytometry using a FACSCalibur (BD Bioscience) as described (32, 36).

Serum resistance assays. The serum resistance of *B. burgdorferi* was measured as described with modifications (6, 24, 32). Cultures in mid-log phase of each strain treated with human blood (32), as well as the high passaged, non-infectious, and serum-sensitive human blood-treated *B. burgdorferi* strain B313 (control), were cultivated in triplicate and diluted to a final concentration of 5×10^6 cells/mL in BSK-II medium without rabbit sera. The cell suspensions were mixed with sera collected from naïve white-footed mice or quail (60% spirochetes and 40% sera) in the presence or absence of 2µM of CVF (Complement Technology) or recombinant OmCI, to deplete complement from mouse and quail sera, respectively. Heat-inactivated sera (65°C for 2-h) were also included in each of the aforementioned combinations as a control (all strains survived equally; data not shown). To determine the bacteria survival, the number of motile spirochetes was counted under dark field microscopy at 0- and 4-h following incubation with sera as described (32), as we have shown that motile spirochetes determined using this methodology accurately reflect results using live-dead staining assays (39). The percent survival of *B. burgdorferi* was calculated by the normalization of motile spirochetes at 4-h post incubation to that immediately after incubation with sera.

Mouse and quail infection by ticks. Generating flat, infected *I. scapularis* nymphs has been described previously (24, 92). The infected nymphs were placed in a chamber to feed on 4- to 6-week old male and female BALB/c or $C3^{-/-}$ mice in BALB/c background, or on four- to six-week old male and female untreated or OmCI-treated quail, as described previously (45). For OmCI-

treatment, the quail were subcutaneously injected with OmCI (1mg/kg of quail) one day prior to the nymph feeding. The engorged nymphs were obtained from the chambers at four days post tick feeding. Animals were sacrificed and tissues collected from the mice at 7-, 10-, or 14- (blood, tick bite site of the skin, heart, bladder, tibiotarsus joint), and quail (blood, tick bite site of the skin, heart, brain) at 9-, or 14-, days post nymph feeding. To ensure OmCI was still functional at these timepoints, quail were subcutaneously injected with OmCI (1mg/kg of quail) or PBS buffer (control), and the sera were collected at 10 days post injection (equivalent to 11 days post nymph feeding). The lack of the ability of this serum to kill the sera-sensitive *B. burgdorferi* strain B313 (i.e., to ensure complement was still depleted) was assessed as described in the section "Serum resistance assays." (**Fig. S13**)

Quantification of spirochete burden The DNA from tissues, blood, and ticks was extracted as described previously (45). qPCR was then performed to quantitate bacterial loads. Spirochete genomic equivalents were calculated using an ABI 7500 Real-Time PCR System (ThermoFisher Scientific) in conjunction with PowerUp SYBR Green Master Mix (ThermoFisher Scientific) based on amplification of the Lyme borreliae *recA* gene using primers BBRecAfp and BBRecArp (**Table S8**) with the amplification cycle as described previously (32). The number of *recA* copies was calculated by establishing a threshold cycle (Cq) standard curve of a known number of *recA* gene extracted from *B. burgdorferi* strain B31-A3, then comparing the Cq values of the experimental samples.

Genomic analyses. To generate the *cspZ* phylogenetic trees, we mined all publicly available *cspZ* sequences on NCBI as of September 2021, including assembled genomes, nucleotides, and

unassembled genomes on the SRA. To pull *cspZ* from unassembled genomes, the short reads were aligned to *cspZ* from B31, B379, or B408 with UGENE v39.0 using BWA-MEM at defaults (93, 94). Strains were removed from the analyses if the coverage was too low, there was evidence of PCR/sequencing errors (e.g., non-conserved homopolymer length) or evidence of multiple CspZ variants within one strain. All resulting 174 *cspZ* sequences, plus the outgroup strains (*B. spielmanii* A14s accession: EU272854.1; *B. afzelii* FEM4 accession: OM243915; *B. afzelii* VS461 accession: MN809989.1; *B. garinii* PBr accession: CP001307.1; *B. bissettii* DN127 accession: NC_015916.1; *B. bissettii* CO275 accession: JNBW01000464.1), were aligned by codons using TranslatorX (95). All isolates were collapsed into haplotypes in FaBox v1.61, and these haplotypes were used with the *B. bissettii* outgroup to generate a NeighborNet network in SplitsTree v4 (96, 97). Phylogenetic trees were estimated using likelihood as optimality criterion in IQ-tree v1.6.12 (98) and a full substitution model search procedure in ModelFinder (99). Internode branch support was estimated with 10,000 replicatess of both ultrafast bootstrapping and the SH-aLRT branch test (98-100). All resulting phylogenetic trees were visualized in iTOL v6.4.3 (101). The pairwise sequence similarity for each of the 174 *B. burgdorferi cspZ* isolates relative to $CspZ_{B31}$, $CspZ_{B408}$, or $CspZ_{B379}$ was determined in MEGA-X with default settings (102). Putative recombination breakpoints were analyzed with GARD (103), and evidence of selection for individual codons or branches on the *cspZ* gene tree were inferred with FUBAR (104), FEL (105) and MEME (106), all on the Datamonkey server (107). Selection tests for the entire *cspZ* gene included Tajima's neutrality test in MEGA-X (102) and all tests available in DnaSP v5 with 10,000 coalescent simulations (108), all at default settings. The ancestor state for the entire *B. burgdorferi* ingroup was reconstructed using the LG model in GRASP 2020.05.05 (109, 110), as well as FireProt-ASR with default settings (111) using both full and haplotype phylogenies, a multitaxon outgroup and

solely *B. bissettii* as outgroup. Divergence dating was carried out in BEAST v1.10.4 (112) using the HKY+Γ_4 substitution model(113, 114), a coalescent Bayesian skyline coalescent model, and a strict clock with a uniform prior on the substitution rate using a previously determined rate of 4.75e-06 substitutions/site/year(42). A Markov chain Monte Carlo chain length of 100 million steps was used with a 10,000-step thinning, resulting in effective sample sizes greater than 200, an indication of an adequate chain mixing. The analyses were ran in triplicate and combined after removing a 10% burn-in.

The amino acids encoding SCRs 6-7 from human (GenBank accession U56979.1), mouse (NM_009888.3), or quail (XM_015869474.2) FH or $CspZ_{B31}$, $CspZ_{B379}$, and $CspZ_{B408}$, the loop-swapped variants, and the reconstructed ancestral CspZ sequences were aligned in MEGA-X using ClustalW with default settings, analyzed with ESPript v3.0, and visualized with Jalview v2.11.0 (102, 115, 116).

Statistical analysis. Samples were compared using the Mann-Whitney *U* test or the Kruskal-Wallis test with the two-stage step-up method of Benjamini, Krieger, and Yekutieli (117).

Accession numbers. The coordinates and the structure factors for $CspZ_{B379}$, $CspZ_{B408}$, and human SCR-$CspZ_{B408}$ have been deposited in the Protein Data Bank with accession codes 7ZJJ, 7ZJK, and 7ZJM, respectively.

ACKNOWLEDGEMENTS

We thank Simon Starkey, Quinton Smith, Deirdre Torrisi, and Joey Anderson from the Wadsworth Veterinary Sciences facility for animal husbandry; Richard Marconi, Utpal Pal and Patricia Rosa

for providing the *B. burgdorferi* strains; David Vance and Nicholas Mantis for sharing the *E. coli* strain; Timothy Czajka for the assistance of SPR analysis; Nikhat Parveen, Klemen Strle, and Grace Chen for insightful comments; and Laurel Lown for assistance with running PCRs. We also appreciate Leslie Eisele and Renjie Song of Wadsworth Biochemistry and Immunology Core for CD spectroscopy, SPR, and flow cytometry; and Karen Chave of the Wadsworth Protein Expression Core for the help of the materials to purify factor H. We thank Wadsworth ATGC core for plasmid sequencing. Diffraction data for *B. burgdorferi* CspZ proteins and CspZ-human FH were collected on BL14.1 at the BESSY II electron storage ring operated by the Helmholtz-Zentrum, Berlin. We would particularly like to acknowledge the help and support of Manfred S. Weiss and Jan Wollenhaupt during the diffraction data collection. This work was supported by NSF IOS1754995 (SOK), NSF IOS1755286 (YPL, ALM, TAN, ADII, JS, and AC), DoD TB170111, NIHR21AI144891, NIH R21AI146381, New York State Department of Health Wadsworth Center Start-Up Grant (ALM, TAN, YPL), and LOEWE Center DRUID Novel Drug Targets against Poverty-Related and Neglected Tropical Infectious Diseases, project C3 (PK). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. The authors have declared that no competing interests exist.

REFERENCES

1. Trovato M, Sartorius R, D'Apice L, Manco R, De Berardinis P. Viral Emerging Diseases: Challenges in Developing Vaccination Strategies. Frontiers in immunology. 2020;11:2130.

2. Cui J, Li F, Shi ZL. Origin and evolution of pathogenic coronaviruses. Nature reviews Microbiology. 2019;17(3):181-92.

3. Douam F, Gaska JM, Winer BY, Ding Q, von Schaewen M, Ploss A. Genetic Dissection of the Host Tropism of Human-Tropic Pathogens. Annual review of genetics. 2015;49:21-45.

4. Kurtenbach K, Hanincova K, Tsao JI, Margos G, Fish D, Ogden NH. Fundamental processes in the evolutionary ecology of Lyme borreliosis. Nature reviews Microbiology. 2006;4(9):660-9.

5. Longdon B, Brockhurst MA, Russell CA, Welch JJ, Jiggins FM. The evolution and genetics of virus host shifts. PLoS pathogens. 2014;10(11):e1004395.

6. Lin YP, Tufts, D.M., Combs, M., Dupuis II, A.P., Marcinkiewicz, A.L., Hirsbrunner, A.D., Diaz, A.J., Stout, J.L., Blom, A.M., Strle, K., Davis, A.D., Kramer, L.D., Kolokotronis, S.O., Diuk-Wasser, M.A. Cellular and immunological mechanisms influence host-adapted phenotypes in a vector-borne microparasite. Proc Biol Sci. 2022. 289(1969):20212087.

7. Wollert T, Pasche B, Rochon M, Deppenmeier S, van den Heuvel J, Gruber AD, et al. Extending the host range of *Listeria monocytogenes* by rational protein design. Cell. 2007;129(5):891-902.

8. Viana D, Comos M, McAdam PR, Ward MJ, Selva L, Guinane CM, et al. A single natural nucleotide mutation alters bacterial pathogen host tropism. Nature genetics. 2015;47(4):361-6.

9. Sabrina Pankey M, Foxall RL, Ster IM, Perry LA, Schuster BM, Donner RA, et al. Host-selected mutations converging on a global regulator drive an adaptive leap towards symbiosis in bacteria. Elife. 2017;6.

10. Radolf JD, Strle K, Lemieux JE, Strle F. Lyme Disease in Humans. Curr Issues Mol Biol. 2021;42:333-84.

11. Steere AC, Strle F, Wormser GP, Hu LT, Branda JA, Hovius JW, et al. Lyme borreliosis. Nature reviews Disease primers. 2016;2:16090.

12. Schwartz I, Margos G, Casjens SR, Qiu WG, Eggers CH. Multipartite Genome of Lyme Disease *Borrelia*: Structure, Variation and Prophages. Curr Issues Mol Biol. 2021;42:409-54.

13. Margos G, Gatewood AG, Aanensen DM, Hanincova K, Terekhova D, Vollmer SA, et al. MLST of housekeeping genes captures geographic population structure and suggests a European origin of *Borrelia burgdorferi*. Proceedings of the National Academy of Sciences of the United States of America. 2008;105(25):8730-5.

14. Wolcott KA, Margos G, Fingerle V, Becker NS. Host association of *Borrelia burgdorferi* sensu lato: A review. Ticks and tick-borne diseases. 2021;12(5):101766.

15. Tufts DM, Hart TM, Chen GF, Kolokotronis SO, Diuk-Wasser MA, Lin YP. Outer surface protein polymorphisms linked to host-spirochete association in Lyme borreliae. Molecular microbiology. 2019;111(4):868-82.

16. Skare JT, Garcia BL. Complement Evasion by Lyme Disease Spirochetes. Trends in microbiology. 2020. 28(11):889-899.

17. Lin YP, Diuk-Wasser MA, Stevenson B, Kraiczy P. Complement Evasion Contributes to Lyme Borreliae-Host Associations. Trends in parasitology. 2020;36(7):634-45.

18. Dulipati V, Meri S, Panelius J. Complement evasion strategies of *Borrelia burgdorferi* sensu lato. FEBS letters. 2020;594(16):2645-56.

19. Reis ES, Mastellos DC, Hajishengallis G, Lambris JD. New insights into the immune functions of complement. Nature reviews Immunology. 2019;19(8):503-16.

20. Merle NS, Church SE, Fremeaux-Bacchi V, Roumenina LT. Complement System Part I - Molecular Mechanisms of Activation and Regulation. Frontiers in immunology. 2015;6:262.

21. Blom AM. The role of complement inhibitors beyond controlling inflammation. Journal of internal medicine. 2017;282(2):116-28.

22. Jozsi M. Factor H Family Proteins in Complement Evasion of Microorganisms. Frontiers in immunology. 2017;8:571.

23. Najafpour B, Cardoso JCR, Canario AVM, Power DM. Specific Evolution and Gene Family Expansion of Complement 3 and Regulatory Factor H in Fish. Frontiers in immunology. 2020;11:568631.

24. Hart T, Nguyen NTT, Nowak NA, Zhang F, Linhardt RJ, Diuk-Wasser M, et al. Polymorphic factor H-binding activity of CspA protects Lyme borreliae from the host complement in feeding ticks to facilitate tick-to-host transmission. PLoS pathogens. 2018;14(5):e1007106.

25. Bykowski T, Woodman ME, Cooley AE, Brissette CA, Brade V, Wallich R, et al. Coordinated expression of *Borrelia burgdorferi* complement regulator-acquiring surface proteins during the Lyme disease spirochete's mammal-tick infection cycle. Infection and immunity. 2007;75(9):4227-36.

26. Hartmann K, Corvey C, Skerka C, Kirschfink M, Karas M, Brade V, et al. Functional characterization of BbCRASP-2, a distinct outer membrane protein of *Borrelia burgdorferi* that binds host complement regulators factor H and FHL-1. Molecular microbiology. 2006;61(5):1220-36.

27. Rogers EA, Marconi RT. Delineation of species-specific binding properties of the CspZ protein (BBH06) of Lyme disease spirochetes: evidence for new contributions to the pathogenesis of *Borrelia* spp. Infection and immunity. 2007;75(11):5272-81.

28. Rogers EA, Abdunnur SV, McDowell JV, Marconi RT. Comparative analysis of the properties and ligand binding characteristics of CspZ, a factor H binding protein, derived from *Borrelia burgdorferi* isolates of human origin. Infection and immunity. 2009;77(10):4396-405.

29. Marcinkiewicz AL, Lieknina I, Yang X, Lederman PL, Hart TM, Yates J, et al. The Factor H-Binding Site of CspZ as a Protective Target against Multistrain, Tick-Transmitted Lyme Disease. Infection and immunity. 2020;88(5).

30. Brangulis K, Petrovskis I, Kazaks A, Bogans J, Otikovs M, Jaudzems K, et al. Structural characterization of CspZ, a complement regulator factor H and FHL-1 binding protein from *Borrelia burgdorferi*. The FEBS journal. 2014;281(11):2613-22.

31. Lin T, Gao L, Zhang C, Odeh E, Jacobs MB, Coutte L, et al. Analysis of an ordered, comprehensive STM mutant library in infectious *Borrelia burgdorferi*: insights into the genes required for mouse infectivity. PloS one. 2012;7(10):e47532.

32. Marcinkiewicz AL, Dupuis AP, 2nd, Zamba-Campero M, Nowak N, Kraiczy P, Ram S, et al. Blood treatment of Lyme borreliae demonstrates the mechanism of CspZ-mediated complement evasion to promote systemic infection in vertebrate hosts. Cellular microbiology. 2019;21(2):e12998.

33. Nakao M, Somamoto, T. The Evolution of Complement Sysstem Functions and Pathways in Vetebrates. Malagoli D, editor: Academic Press; 2016.

34. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596(7873):583-9.

35. Tunyasuvunakool K, Adler J, Wu Z, Green T, Zielinski M, Zidek A, et al. Highly accurate protein structure prediction for the human proteome. Nature. 2021;596(7873):590-6.

36. Combs M, Marcinkiewicz AL, Dupuis AP, 2nd, Davis AD, Lederman P, Nowak TA, et al. Phylogenomic Diversity Elucidates Mechanistic Insights into Lyme Borreliae-Host Association. mSystems. 2022:e0048822.

37. Caine JA, Coburn J. A short-term Borrelia burgdorferi infection model identifies tissue tropisms and bloodstream survival conferred by adhesion proteins. Infection and immunity. 2015;83(8):3184-94.

38. Ristow LC, Miller HE, Padmore LJ, Chettri R, Salzman N, Caimano MJ, et al. The beta(3)-integrin ligand of *Borrelia burgdorferi* is critical for infection of mice but not ticks. Molecular microbiology. 2012;85(6):1105-18.

39. Frye AM, Hart TM, Tufts DM, Ram S, Diuk-Wasser MA, Kraiczy P, et al. A soft tick *Ornithodoros moubata* salivary protein OmCI is a potent inhibitor to prevent avian complement activation. Ticks and tick-borne diseases. 2020;11(2):101354.

40. Finnie JA, Stewart RB, Aston WP. A comparison of cobra venom factor-induced depletion of serum C3 in eight different strains of mice. Dev Comp Immunol. 1981;5(4):697-701.

41. Marcinkiewicz AL, Lieknina I, Kotelovica S, Yang X, Kraiczy P, Pal U, et al. Eliminating Factor H-Binding Activity of *Borrelia burgdorferi* CspZ Combined with Virus-Like Particle Conjugation Enhances Its Efficacy as a Lyme Disease Vaccine. Frontiers in immunology. 2018;9:181.

42. Walter KS, Carpi G, Caccone A, Diuk-Wasser MA. Genomic insights into the ancient spread of Lyme disease across North America. Nat Ecol Evol. 2017;1(10):1569-76.

43. Coburn J, Garcia B, Hu LT, Jewett MW, Kraiczy P, Norris SJ, et al. Lyme Disease Pathogenesis. Curr Issues Mol Biol. 2021;42:473-518.

44. Garcia BL, Summers BJ, Lin Z, Ramyar KX, Ricklin D, Kamath DV, et al. Diversity in the C3b [corrected] contact residues and tertiary structures of the staphylococcal complement inhibitor (SCIN) protein family. The Journal of biological chemistry. 2012;287(1):628-40.

45. Hart TM, Dupuis AP, 2nd, Tufts DM, Blom AM, Starkey SR, Rego ROM, et al. Host tropism determination by convergent evolution of immunological evasion in the Lyme disease system. PLoS pathogens. 2021;17(7):e1009801.

46. Lawrenz MB, Wooten RM, Zachary JF, Drouin SM, Weis JJ, Wetsel RA, et al. Effect of complement component C3 deficiency on experimental Lyme borreliosis in mice. Infection and immunity. 2003;71(8):4432-40.

47. Woodman ME, Cooley AE, Miller JC, Lazarus JJ, Tucker K, Bykowski T, et al. *Borrelia burgdorferi* binding of host complement regulator factor H is not required for efficient mammalian infection. Infection and immunity. 2007;75(6):3131-9.

48. Schneider MC, Prosser BE, Caesar JJ, Kugelberg E, Li S, Zhang Q, et al. *Neisseria meningitidis* recruits factor H using protein mimicry of host carbohydrates. Nature. 2009;458(7240):890-3.

49. Singh D, Yi SV. On the origin and evolution of SARS-CoV-2. Exp Mol Med. 2021;53(4):537-47.

50. Xiang Y, White A. Monkeypox virus emerges from the shadow of its more infamous cousin: family biology matters. Emerging microbes & infections. 2022;11(1):1768-77.

51. Vishnoi A, Kryazhimskiy S, Bazykin GA, Hannenhalli S, Plotkin JB. Young proteins experience more variable selection pressures than old proteins. Genome Res. 2010;20(11):1574-81.

52. Wang IN, Dykhuizen DE, Qiu W, Dunn JJ, Bosler EM, Luft BJ. Genetic diversity of *ospC* in a local population of *Borrelia burgdorferi* sensu stricto. Genetics. 1999;151(1):15-30.

53. Brisson D, Dykhuizen DE. ospC diversity in *Borrelia burgdorferi*: different hosts are different niches. Genetics. 2004;168(2):713-22.

54. Haven J, Vargas LC, Mongodin EF, Xue V, Hernandez Y, Pagan P, et al. Pervasive recombination and sympatric genome diversification driven by frequency-dependent selection in *Borrelia burgdorferi*, the Lyme disease bacterium. Genetics. 2011;189(3):951-66.

55. Haven J, Magori K, Park AW. Ecological and inhost factors promoting distinct parasite life-history strategies in Lyme borreliosis. Epidemics. 2012;4(3):152-7.

56. Qiu WG, Martin CL. Evolutionary genomics of *Borrelia burgdorferi* sensu lato: findings, hypotheses, and the rise of hybrids. Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases. 2014;27:576-93.

57. Bown KJ, Lambin X, Ogden NH, Begon M, Telford G, Woldehiwet Z, et al. Delineating *Anaplasma phagocytophilum* ecotypes in coexisting, discrete enzootic cycles. Emerging infectious diseases. 2009;15(12):1948-54.

58. Ghosh S, O'Connor TJ. Beyond Paralogs: The Multiple Layers of Redundancy in Bacterial Pathogenesis. Frontiers in cellular and infection microbiology. 2017;7:467.

59. Charlesworth D. Balancing selection and its effects on sequences in nearby genome regions. PLoS genetics. 2006;2(4):e64.

60. Wang G, Ojaimi C, Iyer R, Saksenberg V, McClain SA, Wormser GP, et al. Impact of genotypic variation of *Borrelia burgdorferi* sensu stricto on kinetics of dissemination and severity of disease in C3H/HeJ mice. Infection and immunity. 2001;69(7):4303-12.

61. Chan K, Awan M, Barthold SW, Parveen N. Comparative molecular analyses of *Borrelia burgdorferi* sensu stricto strains B31 and N40D10/E9 and determination of their pathogenicity. BMC microbiology. 2012;12:157.

62. Masuzawa T, Beppu Y, Kawabata H, Yanagihara Y, Iwamoto Y, Shimizu T, et al. Experimental *Borrelia burgdorferi* infection of outbred mice. Journal of clinical microbiology. 1992;30(11):3016-8.

63. von Lackum K, Miller JC, Bykowski T, Riley SP, Woodman ME, Brade V, et al. *Borrelia burgdorferi* regulates expression of complement regulator-acquiring surface protein 1 during the mammal-tick infection cycle. Infection and immunity. 2005;73(11):7398-405.

64. Hellwage J, Meri T, Heikkila T, Alitalo A, Panelius J, Lahdenne P, et al. The complement regulator factor H binds to the surface protein OspE of *Borrelia burgdorferi*. The Journal of biological chemistry. 2001;276(11):8427-35.

65. Lin YP, Frye AM, Nowak TA, Kraiczy P. New Insights Into CRASP-Mediated Complement Evasion in the Lyme Disease Enzootic Cycle. Frontiers in cellular and infection microbiology. 2020;10:1.

66. Garcia BL, Zhi H, Wager B, Hook M, Skare JT. *Borrelia burgdorferi* BBK32 Inhibits the Classical Pathway by Blocking Activation of the C1 Complement Complex. PLoS pathogens. 2016;12(1):e1005404.

67. Caine JA, Lin YP, Kessler JR, Sato H, Leong JM, Coburn J. *Borrelia burgdorferi* outer surface protein C (OspC) binds complement component C4b and confers bloodstream survival. Cellular microbiology. 2017.

68. Pereira MJ, Wager B, Garrigues RJ, Gerlach E, Quinn JD, Dowdell AS, et al. Lipoproteome screening of the Lyme disease agent identifies inhibitors of antibody-mediated complement killing. Proceedings of the National Academy of Sciences of the United States of America. 2022;119(13):e2117770119.

69. Tyson KR, Elkins C, de Silva AM. A novel mechanism of complement inhibition unmasked by a tick salivary protein that binds to properdin. Journal of immunology. 2008;180(6):3964-8.

70. Schuijt TJ, Coumou J, Narasimhan S, Dai J, Deponte K, Wouters D, et al. A tick mannose-binding lectin inhibitor interferes with the vertebrate complement cascade to enhance transmission of the lyme disease agent. Cell host & microbe. 2011;10(2):136-46.

71. Schuijt TJ, Hovius JW, van Burgel ND, Ramamoorthi N, Fikrig E, van Dam AP. The tick salivary protein Salp15 inhibits the killing of serum-sensitive *Borrelia burgdorferi* sensu lato isolates. Infection and immunity. 2008;76(7):2888-94.

72. Bockenstedt LK, Wooten RM, Baumgarth N. Immune Response to *Borrelia*: Lessons from Lyme Disease Spirochetes. Curr Issues Mol Biol. 2021;42:145-90.

73. Lin YP, Benoit V, Yang X, Martinez-Herranz R, Pal U, Leong JM. Strain-specific variation of the decorin-binding adhesin DbpA influences the tissue tropism of the lyme disease spirochete. PLoS pathogens. 2014;10(7):e1004238.

74. Lin YP, Tan X, Caine JA, Castellanos M, Chaconas G, Coburn J, et al. Strain-specific joint invasion and colonization by Lyme disease spirochetes is promoted by outer surface protein C. PLoS pathogens. 2020;16(5):e1008516.

75. Barthold SW, Persing DH, Armstrong AL, Peeples RA. Kinetics of *Borrelia burgdorferi* dissemination and evolution of disease after intradermal inoculation of mice. The American journal of pathology. 1991;139(2):263-73.

76. Hanincova K, Ogden NH, Diuk-Wasser M, Pappas CJ, Iyer R, Fish D, et al. Fitness variation of *Borrelia burgdorferi* sensu stricto strains in mice. Applied and environmental microbiology. 2008;74(1):153-7.

77. Lawrenz MB, Hardham JM, Owens RT, Nowakowski J, Steere AC, Wormser GP, et al. Human antibody responses to VlsE antigenic variation protein of *Borrelia burgdorferi*. Journal of clinical microbiology. 1999;37(12):3997-4004.

78. Casjens SR, Gilcrease EB, Vujadinovic M, Mongodin EF, Luft BJ, Schutzer SE, et al. Plasmid diversity and phylogenetic consistency in the Lyme disease agent *Borrelia burgdorferi*. BMC Genomics. 2017;18(1):165.

79. Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nature methods. 2011;8(10):785-6.

80. Chiu J, March PE, Lee R, Tillett D. Site-directed, Ligase-Independent Mutagenesis (SLIM): a single-tube methodology approaching 100% efficiency in 4 h. Nucleic acids research. 2004;32(21):e174.

81. Mueller U, Darowski N, Fuchs MR, Forster R, Hellmig M, Paithankar KS, et al. Facilities for macromolecular crystallography at the Helmholtz-Zentrum Berlin. J Synchrotron Radiat. 2012;19(Pt 3):442-9.

82. Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, et al. Overview of the CCP4 suite and current developments. Acta crystallographica Section D, Biological crystallography. 2011;67(Pt 4):235-42.

83. Evans PR. An introduction to data reduction: space-group determination, scaling and intensity statistics. Acta crystallographica Section D, Biological crystallography. 2011;67(Pt 4):282-92.

84. Kabsch W. Xds. Acta crystallographica Section D, Biological crystallography. 2010;66(Pt 2):125-32.

85. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. J Appl Crystallogr. 2007;40(Pt 4):658-74.

86. Cowtan K. The Buccaneer software for automated model building. 1. Tracing protein chains. Acta crystallographica Section D, Biological crystallography. 2006;62(Pt 9):1002-11.

87. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. Acta crystallographica Section D, Biological crystallography. 2004;60(Pt 12 Pt 1):2126-32.

88. Murshudov GN, Vagin AA, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. Acta crystallographica Section D, Biological crystallography. 1997;53(Pt 3):240-55.

89. Kelly SM, Jess TJ, Price NC. How to study proteins by circular dichroism. Biochimica et biophysica acta. 2005;1751(2):119-39.

90. Lin YP, Greenwood A, Nicholson LK, Sharma Y, McDonough SP, Chang YF. Fibronectin binds to and induces conformational change in a disordered region of leptospiral immunoglobulin-like protein B. The Journal of biological chemistry. 2009;284(35):23547-57.

91. Purser JE, Norris SJ. Correlation between plasmid content and infectivity in *Borrelia burgdorferi*. Proceedings of the National Academy of Sciences of the United States of America. 2000;97(25):13865-70.

92. Hart T, Yang X, Pal U, Lin YP. Identification of Lyme borreliae proteins promoting vertebrate host blood-specific spirochete survival in *Ixodes scapularis* nymphs using artificial feeding chambers. Ticks and tick-borne diseases. 2018.

93. Rose R, Golosova O, Sukhomlinov D, Tiunov A, Prosperi M. Flexible design of multiple metagenomics classification pipelines with UGENE. Bioinformatics. 2019;35(11):1963-5.

94. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv. 2013;1303:3997.

95. Abascal F, Zardoya R, Telford MJ. TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. Nucleic acids research. 2010;38(Web Server issue):W7-13.

96. Huson DH, Bryant D. Application of phylogenetic networks in evolutionary studies. Mol Biol Evol. 2006;23(2):254-67.

97. Villesen P. FaBox: an online toolbox for fasta sequences. Mol Ecol Notes. 2007;7(6):965-8.

98. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2015;32(1):268-74.

99. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. Nature methods. 2017;14(6):587-9.

100. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. UFBoot2: Improving the Ultrafast Bootstrap Approximation. Mol Biol Evol. 2018;35(2):518-22.

101. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic acids research. 2021;49(W1):W293-W6.

102. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Mol Biol Evol. 2018;35(6):1547-9.

103. Kosakovsky Pond SL, Posada D, Gravenor MB, Woelk CH, Frost SD. Automated phylogenetic detection of recombination using a genetic algorithm. Mol Biol Evol. 2006;23(10):1891-901.

104. Murrell B, Moola S, Mabona A, Weighill T, Sheward D, Kosakovsky Pond SL, et al. FUBAR: a fast, unconstrained bayesian approximation for inferring selection. Mol Biol Evol. 2013;30(5):1196-205.

105. Kosakovsky Pond SL, Frost SD. Not so different after all: a comparison of methods for detecting amino acid sites under selection. Mol Biol Evol. 2005;22(5):1208-22.

106. Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, Kosakovsky Pond SL. Detecting individual sites subject to episodic diversifying selection. PLoS genetics. 2012;8(7):e1002764.

107. Weaver S, Shank SD, Spielman SJ, Li M, Muse SV, Kosakovsky Pond SL. Datamonkey 2.0: A Modern Web Application for Characterizing Selective and Other Evolutionary Processes. Mol Biol Evol. 2018;35(3):773-7.

108. Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics. 2009;25(11):1451-2.

109. Le SQ, Gascuel O. An improved general amino acid replacement matrix. Mol Biol Evol. 2008;25(7):1307-20.

110. Ross CM, Foley G, Boden M, Gillam EMJ. Using the Evolutionary History of Proteins to Engineer Insertion-Deletion Mutants from Robust, Ancestral Templates Using Graphical Representation of Ancestral Sequence Predictions (GRASP). Methods in molecular biology. 2022;2397:85-110.

111. Musil M, Khan RT, Beier A, Stourac J, Konegger H, Damborsky J, et al. FireProtASR: A Web Server for Fully Automated Ancestral Sequence Reconstruction. Brief Bioinform. 2021;22(4).

112. Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. Virus Evol. 2018;4(1):vey016.

113. Hasegawa M, Kishino H, Yano T. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J Mol Evol. 1985;22(2):160-74.

114. Yang Z. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. J Mol Evol. 1994;39(3):306-14.

115. Robert X, Gouet P. Deciphering key features in protein structures with the new ENDscript server. Nucleic acids research. 2014;42(Web Server issue):W320-4.

116. Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. Bioinformatics. 2009;25(9):1189-91.

117. Benjamini YK, A. M. Yekutieli, D. . Adaptive linear step-up procedures that control the false discovery rate. Biometrika. 2006;93:491-507.

118. Kraiczy P, Skerka C, Brade V, Zipfel PF. Further characterization of complement regulator-acquiring surface proteins of *Borrelia burgdorferi*. Infection and immunity. 2001;69(12):7800-9.

119. Kraiczy P, Skerka C, Kirschfink M, Brade V, Zipfel PF. Immune evasion of *Borrelia burgdorferi* by acquisition of human complement regulators FHL-1/reconectin and Factor H. European journal of immunology. 2001;31(6):1674-84.

120. Hallstrom T, Siegel C, Morgelin M, Kraiczy P, Skerka C, Zipfel PF. CspA from *Borrelia burgdorferi* inhibits the terminal complement pathway. mBio. 2013; 4(4):e00481-13.

121. Elias AF, Stewart PE, Grimm D, Caimano MJ, Eggers CH, Tilly K, et al. Clonal polymorphism of *Borrelia burgdorferi* strain B31 MI: implications for mutagenesis in an infectious strain background. Infection and immunity. 2002;70(4):2139-50.

122. Coleman AS, Yang X, Kumar M, Zhang X, Promnares K, Shroder D, et al. *Borrelia burgdorferi* complement regulator-acquiring surface protein 2 does not contribute to complement resistance or host infectivity. PloS one. 2008;3(8):3010e.

123. Benoit VM, Fischer JR, Lin YP, Parveen N, Leong JM. Allelic variation of the Lyme disease spirochete adhesin DbpA influences spirochetal binding to decorin, dermatan sulfate, and mammalian cells. Infection and immunity. 2011;79(9):3501-9.

124. Frank KL, Bundle SF, Kresge ME, Eggers CH, Samuels DS. *aadA* confers streptomycin resistance in Borrelia burgdorferi. Journal of bacteriology. 2003;185(22):6723-7.

SUPPLEMENTAL TABLES

Table S1: The list of $CspZ_{B408}$ residues that binds to human FH based on CspZB408-human SCR6-7 complex structure and their equivalent residues of $CspZ_{B31}$ and $CspZ_{B379}$.

$CspZ_{B408}$	$CspZ_{B31}$	$CspZ_{B379}$
Asp47	Asp47	Asp47
Tyr50	Tyr50	Tyr50
Asn51	Asn51	Ser51
Thr54	Thr54	Thr54
Asn58	Asn58	Asn58
Thr62	Thr62	Thr62
Asp71	-	-
Asp73[a]	Asp70[a]	Asp74[a]
Ser75[a]	Ser72[a]	Ser76[a]
Arg142	Arg139	Arg143
Asn183	Asn180	Asn184
Tyr214	Tyr211	Tyr215

[a]Interacts with SCR6.

975 **Table S2. CspZ variants differ in binding to Factor H from different animals**

		ELISA[a]	----- Surface Plasmon Resonance[b] -----		
CspZ variant	**Factor H source**	**K_D (μM)**	**K_D (μM)**	k_{on} ($10^3 s^{-1} M^{-1}$)	k_{off} (s^{-1})
$CspZ_{B31}$	**Mouse**	**0.43±0.74[c]**	**0.20±0.02**	45.78±15.36	0.0084±0.0022
	Quail	**0.91±0.08[d]**	**0.81±0.01**	117.43±3.25	0.095±0.0032
$CspZ_{B379}$	**Mouse**	**n.b.[d]**	**n.b.**	n.b.	n.b.
	Quail	**0.59±0.037**	**0.75±0.15**	119.83±7.08	0.088±0.012
$CspZ_{B379}L_{B408}$	**Mouse**	**1.39±0.12**	**0.90±0.12**	21.16±10.58	0.018±0.0094
	Quail	**n.b.**	**n.b.**	n.b.	n.b.
$CspZ_{B408}$	**Mouse**	**0.68±0.05**	**0.20±0.02**	22.13±1.31	0.008±0.002
	Quail	**n.b.**	**n.b.**	n.b.	n.b.
$CspZ_{B408}L_{B379}$	**Mouse**	**n.b.**	**n.b.**	n.b.	n.b.
	Quail	**1.38±0.47**	**0.99±0.06**	72.43±2.14	0.072±0.006
GST[e]	**Mouse**	**n.b.**	**n.d.**	n.d.	n.d.
	Quail	**n.b.**	**n.d.**	n.d.	n.d.

976 All values represent the mean ± SEM of three experiments.
977 [a]Determined using GST tagged CspZ variants or mutant proteins.
978 [b]Determined using untagged CspZ variants or mutant proteins.
979 [c]Reported previously in (41).
980 [d]No binding activity was detected.
981 [e]GST was included as a negative control.

982 **Table S3. The generation time for *B. burgdorferi* strains used in this study.**

Strain	Generation time[a, b]
B31-A3	16.45 ± 1.38
B31-A3Δ*cspZ*/vector	16.42 ± 1.79
B31-A3Δ*cspZ*/pCspZ$_{B31}$	17.12 ± 1.05
B31-A3Δ*cspZ*/pCspZ$_{B379}$	16.62 ± 2.07
B31-A3Δ*cspZ*/pCspZ$_{B379}$L$_{B408}$	16.40 ± 0.83
B31-A3Δ*cspZ*/pCspZ$_{B408}$	17.31 ± 0.62
B31-A3Δ*cspZ*/pCspZ$_{B408}$L$_{B379}$	18.66 ± 2.06

983 [a]The generation time was calculated as described previously (18)
984 [b]There were no significant differences between generation time of any strain (Kruskal-Wallis test with the two-stage step-up
985 method of Benjamini, Krieger, and Yekutieli)
986

Table S4: The percentage of SAPs in CspZ variants

Loop type	Position	AA	%	AA	%
$CspZ_{B408}$	20	Asp	64.00	Asn	36.00
	30	Asp	88.00	Asn	36.00
	41	Val	56.00	Phe	44.00
	84	Phe	52.00	Leu	48.00
	95	Lys	92.00	Asn	8.00
	107	Met	56.00	Ile	44.00
	183	Asp	52.00	Asn	48.00
$CspZ_{B31}$	30	Asn	88.30	Asp	11.70
	66[a]	Gly	80.85	Val	19.15
	68[a]	Phe	80.85	Tyr	19.15
	81	Phe	88.30	Leu	11.70
	88	Val	98.94	Ala	1.06
	131	Val	94.68	Ala	5.32
	154	Ser	98.64	Pro	1.06
	161	Lys	96.81	Glu	3.19
	203	Ser	98.94	Lys	1.06
	204	Arg	98.94	Leu	1.06
	208	Asn	67.02	Asp	32.98
$CspZ_{B379}$	30	Asp	87.50	Asn	12.50
	50[b]	Tyr	97.92	His	2.08
	235	Ile	97.92	Ser	2.08

[a]Part of the loop structures

[b]Predicted to directly interact with FH

Table S5. Estimated diversification times for each lineage.

Lineage	Median[a]	$HPD_{5\%}$[a]	$HPD_{95\%}$[a]
CspZ-B31	784	263	1860
CspZ-B379	261	59	741
CspZ-B408	671	179	1679

[a]Years before present

Table S6. **CspZ variants of last common ancestral states display versatile binding ability to human, mouse, and quail FH**

CspZ variant	Factor H source	K_D (μM)[a]
$CspZ_{B31}$	Human	0.31±0.03
	Mouse	0.38±0.07
	Quail	0.74±0.15
CspZ-LCAS1	Human	0.34±0.10
	Mouse	0.41±0.07
	Quail	0.77±0.09
CspZ-LCAS2	Human	0.23±0.02
	Mouse	0.36±0.05
	Quail	0.94±0.09
CspZ-LCAS3	Human	0.25±0.02
	Mouse	0.53±0.12
	Quail	0.73±0.06
CspZ-LCAS4	Human	0.30±0.01
	Mouse	0.40±0.01
	Quail	0.93±0.06
DbpA[b]	Human	n.b.[c]
	Mouse	n.b.
	Quail	n.b.

All values represent the mean ± SEM of three experiments.
[a]Determined using histidine tagged CspZ variants or mutant proteins.
[b]DbpA was included as a negative control
[c]No binding activity was detected.

1000 **Table S7. The strains and plasmids used in this study.**

Strain or plasmid	Genotype or characteristic	Source
B. burgdorferi		
B313	High-passage *B. burgdorferi* B31 missing lp5, lp17, lp21, lp25, lp28-1, lp28-2, lp28-3, lp28-4, lp36, lp38, lp54, lp56, cp9, cp32-4, cp32-6, cp32-8, cp32-9	(120)
B31-A3	Clone of *B. burgdorferi* B31 missing cp9 RST type 1, *ospC* type A	(121)
B379	Clone of *B. burgdorferi* B379 isolated from humans with erythema migrans. RST Type 2, *ospC* type K	(21)
B408	Clone of *B. burgdorferi* B379 isolated from humans with erythema migrans. RST type 3, *ospC* type K	(21)
B31-A3Δ*cspZ*	B31-A3Δ*cspZ*::KanR[a]	(122)
B31-A3Δ*cspZ*/Vector	B31-A3Δ*cspZ*::KanR carrying plasmid pKFSS	(32)
B31-A3Δ*cspZ*/pCspZ$_{B31}$	B31-A3Δ*cspZ*::KanR complemented with intact *cspZ* driven by the promoter of *cspZ*$_{B31}$	(32)
B31-A3Δ*cspZ*/pCspZ$_{B379}$	B31-A3Δ*cspZ*::KanR complemented with intact *cspZ*$_{B379}$ driven by the promoter of *cspZ*$_{B31}$	This study
B31-A3Δ*cspZ*/pCspZ$_{B408}$	B31-A3Δ*cspZ*::KanR complemented with intact *cspZ*$_{B408}$ driven by the promoter of *cspZ*$_{B31}$	This study
B31-A3Δ*cspZ*/pCspZ$_{B379}$L$_{B408}$	B31-A3Δ*cspZ*::KanR complemented with intact *cspZ* $_{B379}$ except the residues 190 to 216 replaced by residues 190 to 213 from *cspZ*$_{B408}$,driven by the promoter of *cspZ*$_{B31}$	This study

B31-A3Δ*cspZ*/pCspZ$_{B408}$L$_{B379}$	B31-A3Δ*cspZ*::KanR complemented with intact *cspZ* $_{B408}$ except the residues 190 to 213 replaced by residues 190 to 216 from *cspZ*$_{B379}$, driven by the promoter of *cspZ*$_{B31}$	This study
E. coli		
DH5α	F- Φ80lacZΔM15 Δ(lacZYA-argF) U169 recA1 endA1 hsdR17(rk-, mk+) phoA supE44 thi-1 gyrA96 relA1 λ-	ThermoFisher
BL21(DE3)	F−, ompT hsdSB (rB− mB−) gal dcm (DE3)	Novagen
BL21(DE3)/pGEX4T2-CspZ$_{B31}$	BL21(DE3) producing GST-tagged residues 58 to 711 of CspZ$_{B31}$	(32)
BL21(DE3)/pGEX4T2-CspZ$_{B379}$	BL21(DE3) producing GST-tagged residues 58 to 723 of CspZ$_{B379}$	This study
BL21(DE3)/pGEX4T2-CspZ$_{B408}$	BL21(DE3) producing GST-tagged residues 58 to 720 of CspZ$_{B408}$	This study
BL21(DE3)/pGEX4T2-CspZ$_{B379}$L$_{B408}$	BL21(DE3) producing GST-tagged residues 58 to 723 of CspZ$_{B379}$ except the residues 190 to 216 replaced by residues 190 to 213 from CspZ$_{B408}$	This study
BL21(DE3)/pGEX4T2-CspZ$_{B408}$L$_{B379}$	BL21(DE3) producing GST-tagged residues 58 to 720 of CspZ$_{B408}$ except the residues 190 to 213 replaced by residues 190 to 216 from CspZ$_{B379}$	This study
BL21(DE3)/pET15b-DbpA$_{VS461}$	BL21(DE3) producing histidine-tagged residues 22 to 170 of DbpA$_{VS461}$	(123)
BL21(DE3)/pET28a-CspZ-LCAS1	BL21(DE3 producing histidine tagged residue 58 to 711 of CspZ-LCAS1	This study
BL21(DE3)/pET28a-CspZ-LCAS2	BL21(DE3 producing histidine tagged residue 58 to 711 of CspZ-LCAS2	This study

BL21(DE3)/pET28a-CspZ-LCAS3	BL21(DE3 producing histidine tagged residue 58 to 711 of CspZ-LCAS3	This study
BL21(DE3)/pET28a-CspZ-LCAS4	BL21(DE3 producing histidine tagged residue 58 to 711 of CspZ-LCAS4	This study
Rosetta-gami(DE3)	F– ompT hsdSB (rB– mB–) gal dcm lacY1 ahpC (DE3) gor522::Tn10 trxB pRARE (CamR, KanR, TetR)	MilliporeSigma
P. pastoris		
X-33/FH SCR6-7	Wild-type Mut+ Pichia strain for expression of FH SCR6-7 (residues 321 to 444 of human FH)	Invitrogen
Plasmids		
pJET1.2/Blunt	AmpR[a]; PCR cloning vector	ThermoFisher
pGEX4T2	AmpR; GST-tagged protein expression vector	Qiagen
pGEX4T2-CspZ	AmpR; pGEX4T2 encoding GST fusion protein residue 58 to 711 of CspZ	(32)
pGEX4T2-CspZ- $CspZ_{B379}$	AmpR; pGEX4T2 encoding GST fusion protein residue 58 to 723 of $CspZ_{B379}$	This study
pGEX4T2-CspZ- $CspZ_{B408}$	AmpR; pGEX4T2 encoding GST fusion protein residue 58 to 720 of $CspZ_{B408}$	This study
pGEX4T2-CspZ-$CspZ_{B379}L_{B408}$	AmpR; pGEX4T2 encoding GST fusion protein residue 58 to 723 of $CspZ_{B379}$ except the residues 190 to 216 replaced by residues 190 to 213 from $CspZ_{B408}$	This study
pGEX4T2-CspZ-$CspZ_{B408}L_{B379}$	AmpR; pGEX4T2 encoding GST fusion protein residue 58 to 720 of $CspZ_{B408}$ except the residues 190 to 213 replaced by residues 190 to 216 from $CspZ_{B379}$	This study

pET28a	KanR[b]; Histidine-tagged protein expression vector	EMD Millipore
pET28a-CspZ-LCAS1	KanR; pET28a encoding histidine protein residues 58 to 711 of the CspZ-LCAS1	This study
pET28a-CspZ-LCAS2	KanR; pET28a encoding histidine protein residues 58 to 711 of the CspZ-LCAS2	This study
pET28a-CspZ-LCAS3	KanR; pET28a encoding histidine protein residues 58 to 711 of the CspZ-LCAS3	This study
pET28a-CspZ-LCAS4	KanR; pET28a encoding histidine protein residues 58 to 711 of the CspZ-LCAS4	This study
pKFSS-1	StrR[c]; *Borrelia* shuttle vector	(124)
pKFSS/pCspZ$_{B31}$	StrR; pKFSS-1 encoding intact *cspZ*$_{B31}$, driven by the promoter of *cspZ*$_{B31}$	(32)
pKFSS/pCspZ$_{B379}$	StrR; pKFSS-1 encoding intact *cspZ*$_{B379}$, driven by the promoter of *cspZ*$_{B31}$	This study
pKFSS/pCspZ$_{B408}$	StrR; pKFSS-1 encoding intact *cspZ*$_{B408}$, driven by the promoter of *cspZ*$_{B31}$	This study
pKFSS/pCspZ$_{B379}$L$_{B408}$	StrR; pKFSS-1 encoding intact *cspZ*$_{B379}$ except the residues 190 to 216 replaced by residues 190 to 213 from *cspZ*$_{B408}$, driven by the promoter of *cspZ*$_{B31}$	This study
pKFSS/pCspZ$_{B408}$L$_{B379}$	StrR; pKFSS-1 encoding intact *cspZ*$_{B408}$ except the residues 190 to 213 replaced by residues 190 to 216 from *cspZ*$_{B379}$, driven by the promoter of *cspZ*$_{B31}$	This study
pETm-11/pCspZ$_{B379}$	KanR; 6xHis tag expression vector encoding CspZ$_{B379}$ residues 23-236	This study
pETm-11/pCspZ$_{B408}$	KanR; 6xHis tag expression vector encoding CspZ$_{B408}$ residues 23-236	This study
pPICZα-FH		This study

	Zeocin resistant; yeast expression vector encoding SCR6-7 of human FH under αMF secretion signal

[a]Ampicillin resistant
[b]Kanamycin resistant
[c]Streptomycin resistant

1005 **Table S8. Primers used in this study.**

Purpose	Primer	Sequence
qPCR spirochete burden: *recA*	BBRecAfp	GTGGATCTATTGTATTAGATGAGGCTCTCG
	BBRecArp	CAGCAACATGTCTGGCATTAGACAC
Generate $_{B379}L_{B408}$: SLIM step 1	$_{B379}L_{B408}$_A-1_mtsengtd	CAGAAGATGTGTTACCTTCCGAAGTCATATAAGTCATAATATC ATTATATGCTCCTGTA
	$_{B379}L_{B408}$_B-1_mtsengtd	TATGACTTATATGACTTCGGAAGGTAACACATCTTCTGATAAA AGTAAGGTTAATCAAG
	$_{B379}L_{B408}$_C-1_mtsengtd	ATATCATTATATGCTCCTGTA
	$_{B379}L_{B408}$_D-1_mtsengtd	ATAAAAGTAAGGTTAATCAAG
Generate $_{B379}L_{B408}$: SLIM step 2	$_{B379}L_{B408}$_A-2_tsenngtd	TTTTATCAGAATCTGTGTTATTACCTTCCGAAGTATAAGTCATA ATATCATTATATGCTCCTGTA
	$_{B379}L_{B408}$_B-2_tsenngtd	TATGACTTATACTTCGGAAGGTAATAACACAGATTCTGATAAA AGTAAGGTTAATCAAGCTATAT
	$_{B379}L_{B408}$_C-2_tsenngtd	ATATCATTATATGCTCCTGTA
	$_{B379}L_{B408}$_D-2_tsenngtd	GTAAGGTTAATCAAGCTATAT
Generate $_{B379}L_{B408}$: SLIM step 3	$_{B379}L_{B408}$_A-3_sevnnvtd	CAGAATCTGTAACGTTATTAACTTCCGAATAAGTCATAATATCA TTATATGCTCCTGTA
	$_{B379}L_{B408}$_B-3_sevnnvtd	TATGACTTATTCGGAAGTTAATAACGTTACAGATTCTGATAAA AGTAAGGTTAATCAAG
	$_{B379}L_{B408}$_C-3_sevnnvtd	ATATCATTATATGCTCCTGTA
	$_{B379}L_{B408}$_D-3_sevnnvtd	ATAAAAGTAAGGTTAATCAAG
Generate $_{B408}L_{B379}$: SLIM step 1	$_{B408}L_{B379}$_A-1_tsenngtd	CAGAATCTGTACCGTTATTTTCCGAAGTATAAGTCATAATATCA TTATATGCTTCTGTA
	$_{B408}L_{B379}$_B-1_tsenngtd	TATGACTTATACTTCGGAAAATAACGGTACAGATTCTGATAAA AGTAAGGTTAATCAAG
	$_{B408}L_{B379}$_C-1_tsenngtd	ATATCATTATATGCTTCTGTA
	$_{B408}L_{B379}$_D-1_tsenngtd	TAAAAGTAAGGTTAATCAAG
Generate $_{B408}L_{B379}$: SLIM step 2	$_{B408}L_{B379}$_A-2_mtsengts	TTTTATCAGAAGATGTACCGTTTTCCGAAGTCATATAAGTCATA ATATCATTATATGCTTCTGTA

	$_{B408}L_{B379}$_B-2_mtsengts	TATGACTTATATGACTTCGGAAAACGGTACATCTTCTGATAAA AGTAAGGTTAATCAAGCTATAT
	$_{B408}L_{B379}$_C-2_mtsengts	ATATCATTATATGCTTCTGTA
	$_{B408}L_{B379}$_D-2_mtsengts	GTAAGGTTAATCAAGCTATAT
Generate $_{B408}L_{B379}$: SLIM step 3	$_{B408}L_{B379}$_A-3_imtysegts	AAGATGTACCTTCCGAATAAGTCATAATATAAGTCATAATATC ATTATATGCTTCTGTA
	$_{B408}L_{B379}$_B-3_imtysegts	TATGACTTATATTATGACTTATTCGGAAGGTACATCTTCTGATA AAAGTAAGGTTAATC
	$_{B408}L_{B379}$_C-3_imtysegts	ATATCATTATATGCTTCTGTA
	$_{B408}L_{B379}$_D-3_imtysegts	CTGATAAAAGTAAGGTTAATC
Generate $CspZ_{B379}$ residues 23-236: structural studies	*cspZ*$_{B379}$ Forw.	CATGCCATGGGCAGATTAAATCAGAGAAAT
	cspZ$_{B379}$ Rev.	GCTTGCGGCCGCTTATAATAAAGTTTGCTTAAT
Generate $CspZ_{B408}$ residues 23-236: structural studies	*cspZ*$_{B408}$ Forw.	CATGCCATGGGCAGATTAAATCAGAGAAAT
	cspZ$_{B408}$ Rev.	GCTTGCGGCCGCTTATAATAAAGTTTGCTTAAT

1006

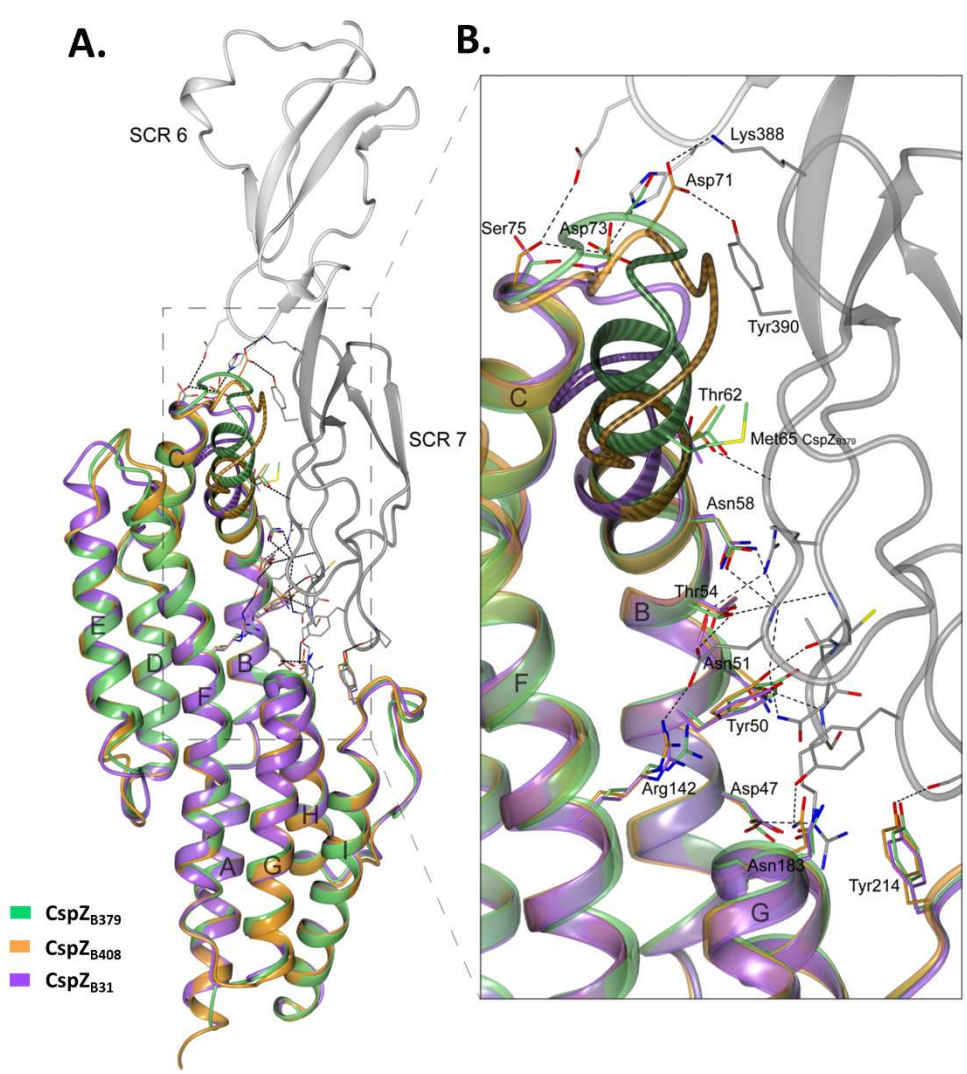
Table S9. Data processing, refinement, and validation statistics of crystal structures.

Dataset	$CspZ_{B408}$	$CspZ_{B379}$	$CspZ_{B408}$-SCR6-7
X-ray diffraction data			
PDB entry	7ZJK	7ZJJ	7ZJM
Beamline	BESSY II beamline 14.1	BESSY II beamline 14.1	BESSY II beamline 14.1
Space group	$P2_1$	$P22_12_1$	$P2_12_12_1$
a, *b*, *c* (Å)	47.63, 87.49, 48.86	53.60, 59.95, 61.05	42.10, 71.07, 147.75
α, *β*, *γ* (°)	90.0, 97.1, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0
Wavelength (Å)	0.9798	0.9762	0.9798
Resolution (Å)	48.49-2.45	61.05-2.10	73.87-2.59
Highest resolution bin (Å)	2.55-2.45	2.16-2.10	2.65-2.59
No. of reflections	97062	127094	101717
No. of unique reflections	14406	11805	14404
Completeness (%)	98.2 (88.4)[a]	98.9 (99.5)	99.8 (99.8)
R_{merge}	0.09 (0.38)	0.10 (0.35)	0.11 (0.38)
$CC_{1/2}$	0.997 (0.940)	0.998 (0.985)	0.988 (0.926)
I/σ (I)	12.1 (4.2)	14.8 (6.3)	12.8 (4.9)
Multiplicity	6.7 (6.2)	10.8 (11.1)	7.1 (7.5)
Refinement			
R_{work}	0.193 (0.248)	0.208 (0.371)	0.217 (0.270)
R_{free}	0.262 (0.399)	0.262 (0.431)	0.275 (0.354)
Average B-factor (Å^2)			
Overall	45.1	33.0	31.9
From Wilson plot	36.9	15.9	24.1
No. of atoms			
Protein	3592	1769	1914
RMS deviations from ideal			
Bond lengths (Å)	0.007	0.008	0.010
Bond angles (°)	1.435	1.541	1.548
Ramachandran outliers (%)			
Residues in most favored regions (%)	93.74	95.31	94.93
Residues in allowed regions (%)	5.10	4.69	4.17

Outliers (%)	1.16	0.00	0.90

[a]Values in parentheses are for the highest resolution bin.

1030 **FIGURES**



1031

1032 **Figure 1. Superimposed crystal structures of CspZ$_{B408}$-SCR6-7 complex, CspZ$_{B31}$, and**
1033 **CspZ$_{B379}$ reveals the FH-binding mechanisms of CspZ variants.** (**A**) The crystal structure of
1034 human FH domains SCR6 (light grey) and SCR7 (dark grey) in complex with CspZ$_{B408}$ (orange)
1035 was superimposed with that of CspZ$_{B31}$ (purple) and CspZ$_{B379}$ (green). (**B**) The FH-binding
1036 interface with the location of the CspZ$_{B408}$ residues involved in human SCR6-7 binding, and the
1037 equvalent residues in CspZ$_{B31}$ and CspZ$_{B379}$. The numbering of the residues is given for CspZ$_{B408}$.
1038 α-helices in CspZ are labelled from A to I starting from the N-terminus. The polymorphic loop

1039 region between α-helices B and C in $CspZ_{B408}$ (residues 60-IMTYSEVNNVTD-71), $CspZ_{B31}$

1040 (residues 60-IMTYSEGT-67) and $CspZ_{B379}$ (residues 60-IMTYIMTYSEGT-71) are indicated.

1041

1042

1043

1044

1045

1046

1047

1048

1049

1050

1051

1052

1053

1054

1055

1056

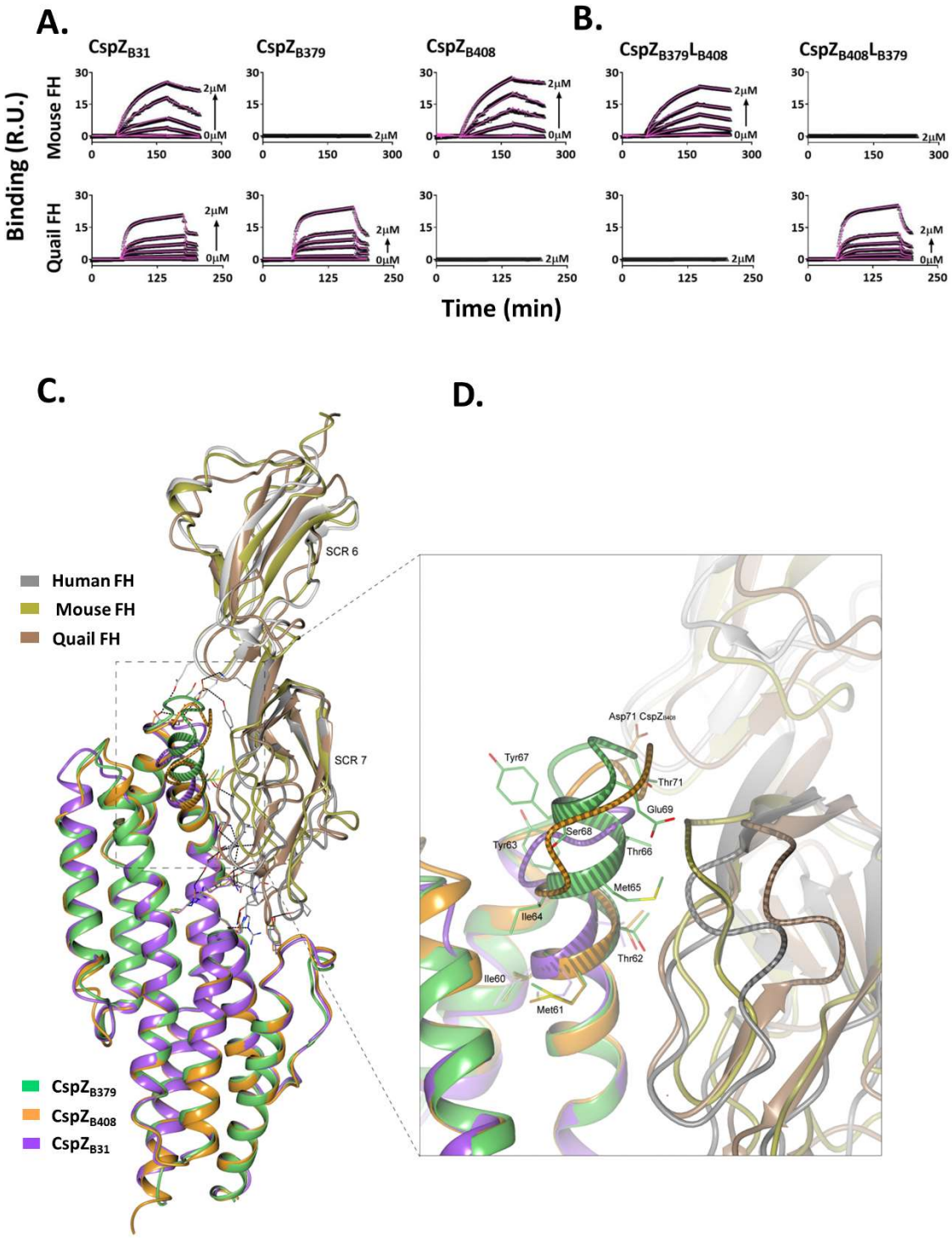
1057

1058

1059

1060

1061



1062

1063 **Figure 2. The polymorphic loop region in CspZ promotes host-specific FH-binding ability.**

1064 **(A, B)** Ten micrograms of mouse or quail FH were conjugated on a SPR chip. 0.008 to 2μM of

1065 untagged **(A)** $CspZ_{B31}$, $CspZ_{B379}$, $CspZ_{B408}$, **(B)** $CspZ_{B379}L_{B408}$, or $CspZ_{B408}L_{B379}$ was flowed over

1066 the chip surface. Binding was measured in response units (RU) by SPR. The experiments were
1067 performed with a single preparation of recombinant proteins tested in three independent replicates
1068 with samples ran in duplicate. The k_{on}, k_{off}, and K_D values were determined from the average of
1069 these three experiments (**Table 1**). Shown is one representative experiment. (**C**) The crystal
1070 structure of $CspZ_{B408}$-SCR6-7 where human SCR6-7 (grey) is superimposed with mouse SCR6-7
1071 (gold, PDB: 2YBY), and the predicted structure of quail SCR6-7 (brown), while $CspZ_{B408}$ (orange)
1072 is superimposed with $CspZ_{B31}$ (purple) and $CspZ_{B379}$ (green). (**D**) The loop region in CspZ proteins
1073 located between α-helices B and C. The Asp71 in $CspZ_{B408}$, and residues 60-IMTYIMTYSEGT-
1074 71 in $CspZ_{B379}$ are highlighted. The polymorphic loop region between α-helices B and C in CspZ
1075 variants, as well as the loop region in human, mouse and quail SCR6-7 located at the interface
1076 with helix-B in CspZ, is showed with a striped pattern.
1077
1078
1079
1080
1081
1082
1083
1084
1085

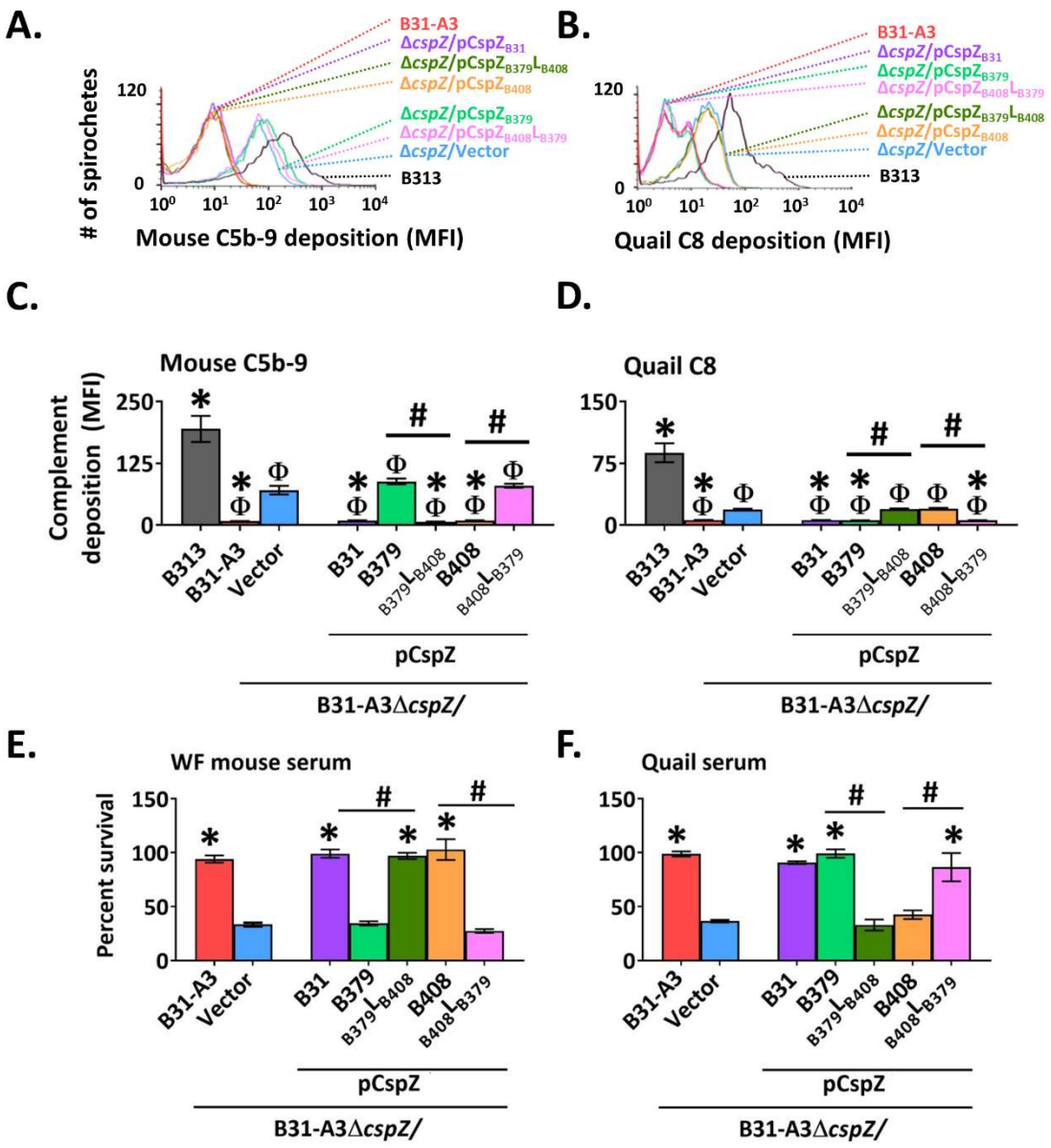


Figure 3. The CspZ loop-driven, variable FH-binding activity confers host-specific complement inactivation on spirochete surface and serum resistance. (A to D) *B. burgdorferi* strains B313 (negative control), B31-A3, B31-A3Δ*cspZ* harboring the empty vector pKFSS ("Δ*cspZ*/Vector"), or this *cspZ* mutant producing $CspZ_{B31}$, $CspZ_{B379}$, $CspZ_{B379}L_{B408}$, $CspA_{B408}$, or $CspZ_{B408}L_{B379}$, were incubated with mouse or quail sera at a final concentration of 20%. The bacteria were stained with antibodies that recognize mouse C5b-9 or quail C8 prior to analysis by flow cytometry. Shown are the representative histograms of the analysis presenting the deposition

1094 levels of **(A)** mouse C5b-9 or **(B)** quail C8 on the surface of the indicated *B. burgdorferi* strains.
1095 The deposition levels of **(C)** mouse C5b-9 or **(D)** quail C8 on the surface of the indicated strains
1096 were measured by flow cytometry and presented as mean fluorescence index (MFI). Each bar
1097 represents the mean of three independent experiments ± SEM. Significant differences ($p < 0.05$,
1098 Kruskal-Wallis test with the two-stage step-up method of Benjamini, Krieger, and Yekutieli) in
1099 the deposition levels of mouse C5b-9 or quail C8 relative to B313 ("Φ"), Δ*cspZ*/Vector ("*"), or
1100 between two strains relative to each other ("#"), are indicated. (**E, F**) The *B. burgdorferi* strains
1101 were incubated for 4-h with (**E)** white-footed mouse sera or **(F)** quail sera, to a final concentration
1102 of 40%. The number of motile spirochetes was assessed microscopically. The precent survival of
1103 the strains was calculated using the number of motile spirochetes at 4-h post incubation normalized
1104 to that prior to incubation with sera. Each bar represents the mean of three independent experiments
1105 ± SEM. Significant differences ($p < 0.05$, Kruskal-Wallis test with the two-stage step-up method
1106 of Benjamini, Krieger, and Yekutieli) in the percent survival of spirochetes relative to the
1107 Δ*cspZ*/Vector ("*") or between two strains relative to each other ("#") are indicated.
1108
1109
1110
1111
1112
1113

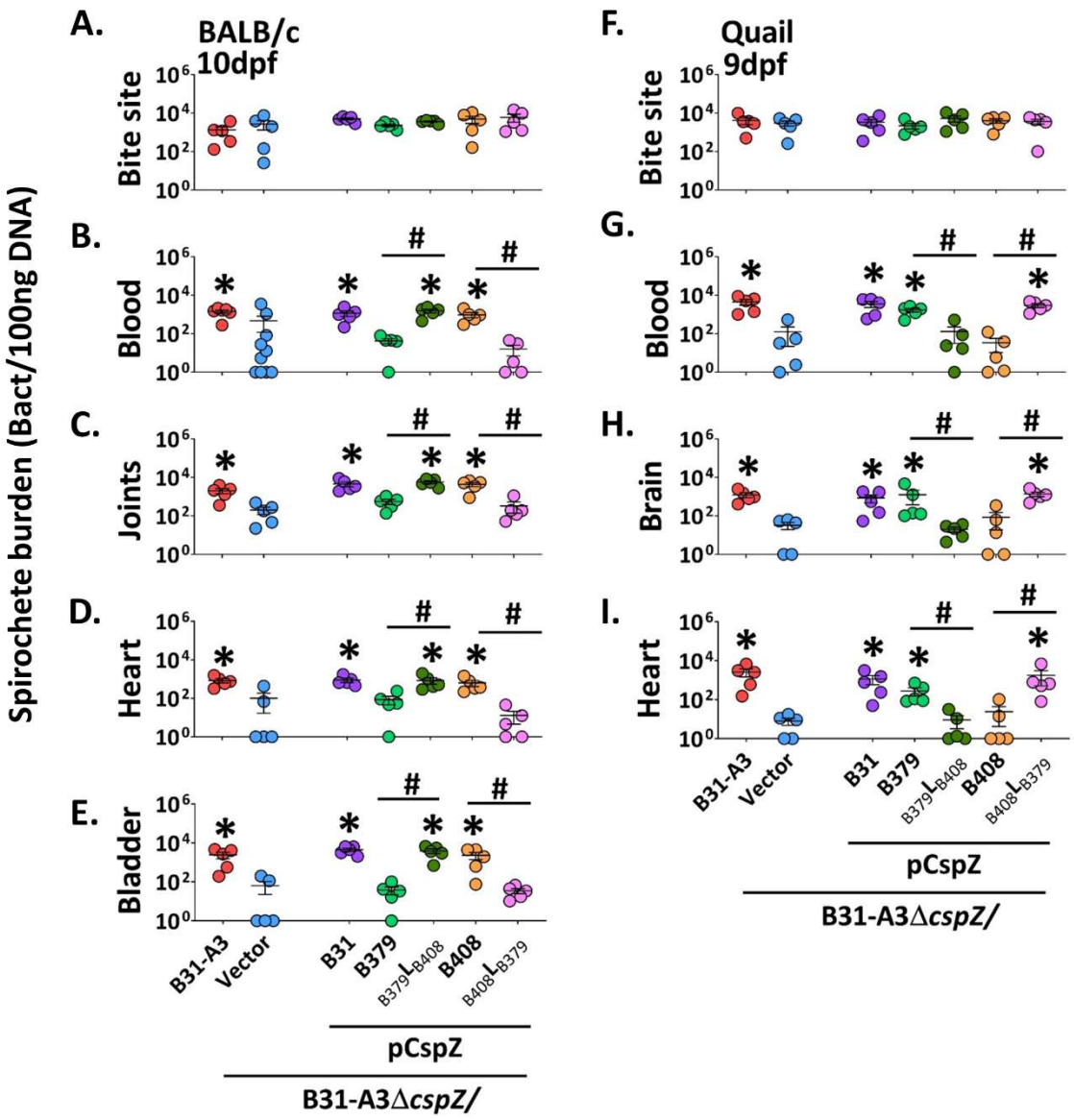
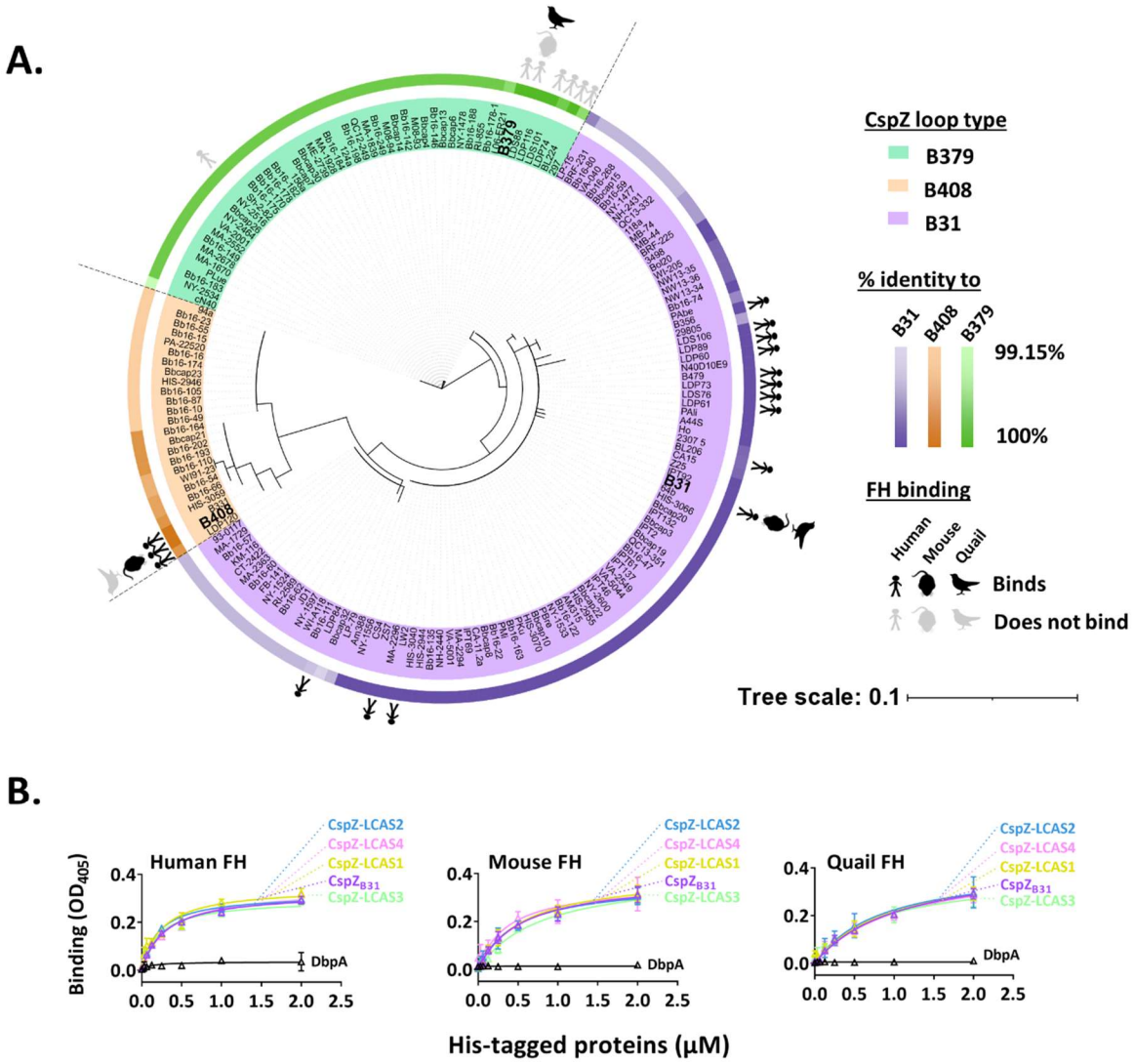


Figure 4. The CspZ loop-mediated FH-binding activity defines host-specific early hematogenous dissemination in a complement-dependent manner. *I. scapularis* nymphs carrying B31-A3, B31-A3Δ*cspZ* harboring the empty vector pKFSS ("Δ*cspZ*/Vector"), or this *cspZ* mutant strain producing $CspZ_{B31}$, $CspZ_{B379}$, $CspZ_{B379}L_{B408}$, $CspZ_{B408}$, or $CspZ_{B408}L_{B379}$ were allowed to feed until repletion on **(A to E)** BALB/c mice or **(F to I)** quail. The bacterial loads in the indicated distal tissues were determined by qPCR at 10 days post nymphs feeding (dpf) in mice and 9dpf in quail. The bacterial loads were normalized to 100ng total DNA. Shown are the

1122 geometric mean of bacterial loads ± SEM of five mice or quail per group, except for the blood
1123 from BALB/c mice, which has nine mice per group. Significant differences ($p < 0.05$, Kruskal-
1124 Wallis test with the two-stage step-up method of Benjamini, Krieger, and Yekutieli) in the
1125 spirochete burdens relative to the Δ*cspZ*/Vector ("*") or between two strains relative to each other
1126 ("#") are indicated.
1127
1128
1129
1130
1131
1132
1133
1134
1135
1136
1137
1138
1139
1140
1141
1142



1143

1144 **Figure 5. Phylogeny and sequence comparisons support polymorphic CspZ loop arising from**
1145 **host-specific adaptation of Lyme borreliae.** (**A**) Unrooted likelihood tree of 174 *B. burgdorferi*
1146 *cspZ* isolates generated with IQ-tree and visualized in iTOL. The names of the isolates are
1147 highlighted based on their CspZ loop type ($CspZ_{B379}$, $CspZ_{B408}$, and $CspZ_{B31}$ as green, orange, and
1148 purple, respectively). The percent identity of each isolate relative to $CspZ_{B31}$ (purple), $CspZ_{B408}$
1149 (orange), or $CspZ_{B379}$ (green) is indicated on the outer ring. The known FH-binding activity is
1150 marked: black or gray stick figures, mice, and quail indicate FH binding or lack of thereof,
1151 respectively, as determined herein and from previously published studies (27-29, 118, 119). **(B)**

1152 The indicated concentrations of recombinant histidine-tagged, predicted last common ancestor
1153 states of CspZ ("CspZ-LCAS1, CspZ-LCAS2, CspZ-LCAS3, and CspZ-LCAS4), $CspZ_{B31}$, or
1154 DbpA (negative control) were added to triplicate wells coated with FH from human, mouse or
1155 quail, and protein binding was quantitated by ELISA. The experiments were performed with a
1156 single preparation of recombinant proteins tested in three independent iterations, in which samples
1157 were ran in duplicate. The K_D values **(Table S6)** representing the FH-binding affinity of each
1158 protein were determined from the average of three experiments. Shown is a representative iteration
1159 averaging the duplicates.
1160
1161
1162
1163
1164
1165
1166
1167
1168
1169
1170
1171
1172
1173
1174

SUPPLEMENTAL FIGURES

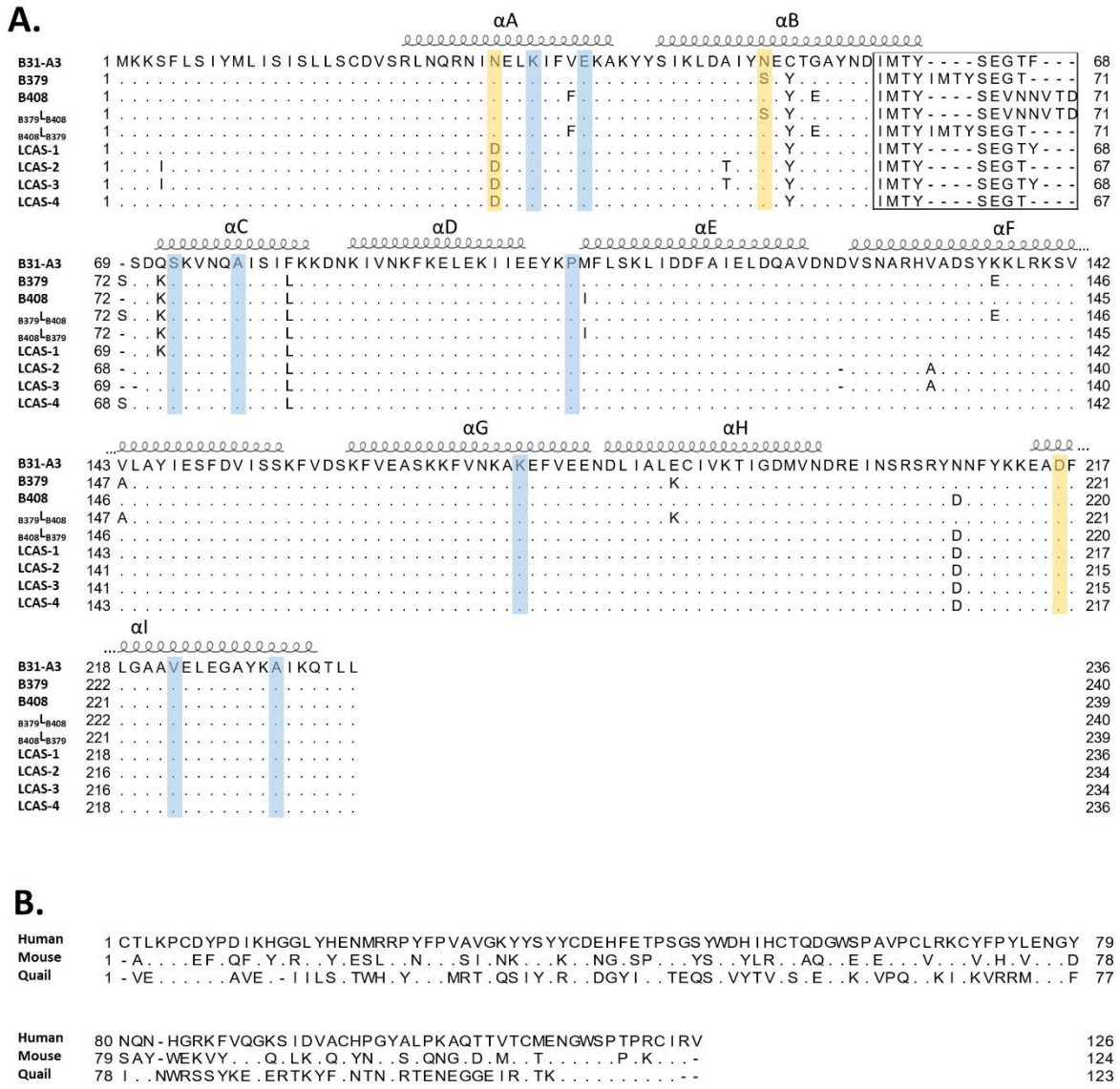


Figure S1. Amino acid alignments of the CspZ and FH variants. Amino acid alignments of **(A)** the indicated CspZ variants, mutants, and reconstructed ancestral states of variants or **(B)** human, mammalian, and avian FH SCR 6-7. (**A**) The CspZ amino acids accounting for the loop structures are indicated with the box, and the alpha-helices labeled above the sequences are extrapolated from the high-resolution structure of $CspZ_{B408}$. The yellow and blue shading are indicative of loci showing evidence of positive and negative selection, respectively.

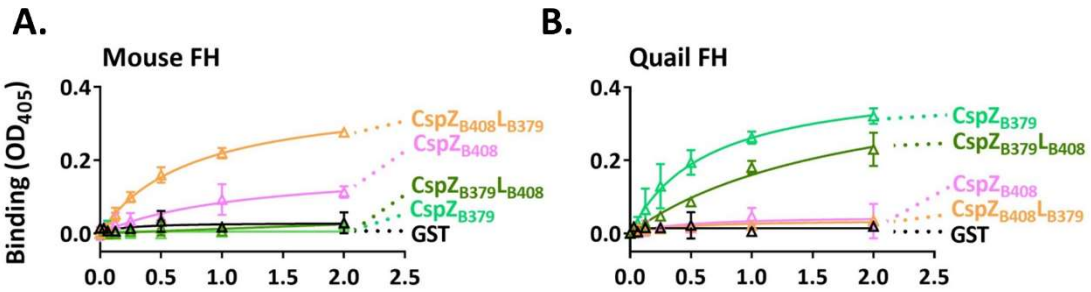


Figure S2. The polymorphic loop structures in recombinant CspZ proteins promote host-specific FH-binding ability determined by ELISA. The indicated concentrations of recombinant GST-tagged $CspZ_{B379}$, $CspZ_{B408}$, $CspZ_{B379}L_{B408}$, or $CspZ_{B408}L_{B379}$, or GST (negative control) were added to triplicate wells coated with FH from mouse or quail, and protein binding was quantitated by ELISA. The experiments were performed with a single preparation of recombinant proteins tested in three independent iterations, in which samples were ran in duplicate. The K_D values **(Table S2)** representing the FH-binding affinity of each protein were determined from the average of three experiments. Shown is a representative iteration averaging the duplicates.

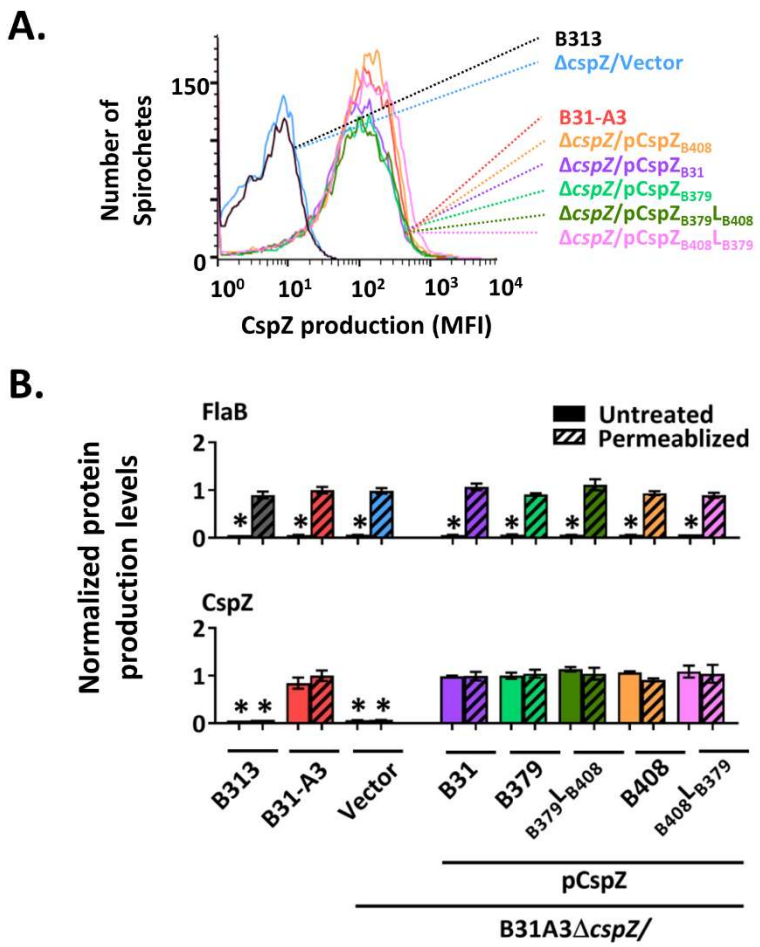


Figure S3. Indistinguishable surface production of CspZ among *B. burgdorferi* strains was observed using flow cytometry. Flow cytometry analysis of CspZ localized on the surface of *B. burgdorferi* strains B31-A3, B31-A3Δ*cspZ* harboring the vector pKFSS ("Δ*cspZ*/Vector"), or this *cspZ* mutant strain producing CspZ$_{B31}$, CspZ$_{B379}$, CspZ$_{B379}$L$_{B408}$, CspA$_{B408}$, or CspZ$_{B408}$L$_{B379}$. **(A)** Representative histograms of flow cytometry analysis showing the levels of CspZ surface production on the indicated strains. **(B)** The production of FlaB (negative control) and CspZ on the surface of indicated *B. burgdorferi* strains was detected by flow cytometry. Values are shown normalized to the production levels of FlaB or CspZ on the surface of permeabilized B31-A3. Each bar represents the mean of four independent experiments ± the standard deviation. An asterisk (*) indicates that relative surface production of the indicated proteins was significantly lower ($p <$

1212 0.05, Kruskal-Wallis test with the two-stage step-up method of Benjamini, Krieger, and Yekutieli)
1213 than that of permeabilized FlaB or CspZ by B31-A3.
1214
1215
1216
1217
1218
1219
1220
1221
1222
1223
1224
1225
1226
1227
1228
1229
1230
1231
1232

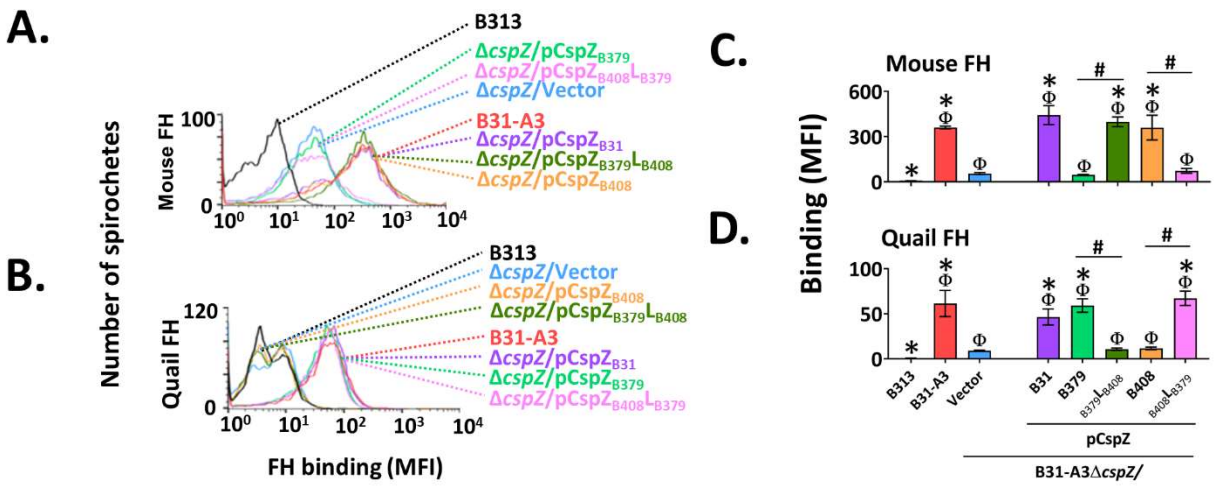
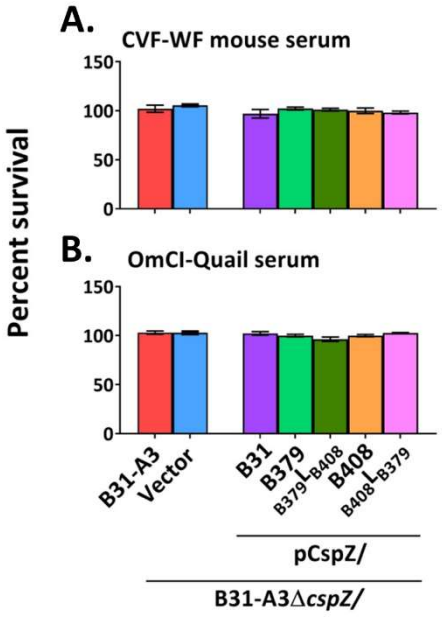


Figure S4. The polymorphic CspZ loop determines the host-specific, allelically variable FH-binding activity on spirochete surface. *B. burgdorferi* strains B313 (negative control), B31-A3, B31-A3Δ*cspZ* harboring the empty vector pKFSS ("Δ*cspZ*/Vector"), or this mutant strain producing $CspZ_{B31}$, $CspZ_{B379}$, $CspZ_{B379}L_{B408}$, $CspZ_{B408}$, or $CspZ_{B408}L_{B379}$, was incubated with mouse or quail FH. The bacteria were stained with antibodies that recognize these FH variants prior to flow cytometry. Shown are the representative histograms of flow cytometry analysis presenting the levels of FH from **(A)** mouse or **(B)** quail binding to each *B. burgdorferi* strain. The levels of **(C)** mouse or **(D)** quail FH-binding were measured by flow cytometry and presented as mean fluorescence index (MFI). Each bar represents the mean of three independent experiments ± SEM. Significant differences ($p < 0.05$, Kruskal-Wallis test with the two-stage step-up method of Benjamini, Krieger, and Yekutieli) in the levels of FH-binding relative to the B313/Vector ("Φ"), Δ*cspZ*/Vector ("*"), or between two strains relative to each other ("#") are indicated.



1249

1250 **Figure S5. The CspZ loop-driven, host-specific serum resistance is recovered in complement-**
1251 **depleted sera.** *B. burgdorferi* strains B313 (negative control), B31-A3, B31-A3Δ*cspZ* harboring
1252 the empty vector pKFSS ("Δ*cspZ*/Vector"), or this mutant strain producing $CspZ_{B31}$, $CspZ_{B379}$,
1253 $CspZ_{B379}L_{B408}$, $CspZ_{B408}$, or $CspZ_{B408}L_{B379}$were incubated for 4-h with (**A)** CVF-treated white-
1254 footed mouse or **(B)** OMCI-treated quail sera, to a final concentration of 40%. The number of
1255 motile spirochetes was assessed microscopically. The precent survival of the *B. burgdorferi* strains
1256 was calculated using the number of motile spirochetes at 4-h post incubation normalized to that
1257 prior to incubation with sera. Each bar represents the mean of three independent experiments ±
1258 SEM. There were no significant differences ($p < 0.05$, Kruskal-Wallis test with the two-stage step-
1259 up method of Benjamini, Krieger, and Yekutieli) between the percent survival of any strain.

1260

1261

1262

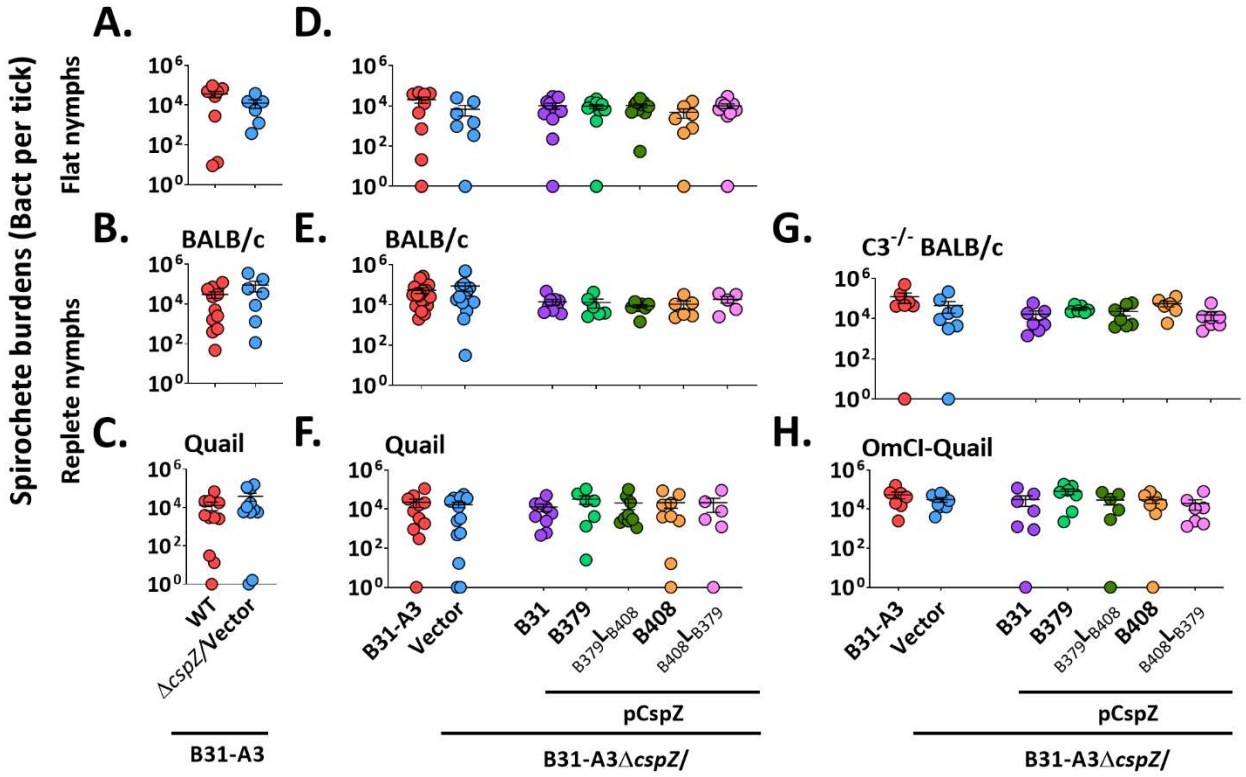


Figure S6. *B. burgdorferi* strains exhibit similar burdens in flat and fed nymphs. *B. burgdorferi*-infected flat nymphs were allowed to feed to repletion on **(B, E)** BALB/c or **(G)** C3-/- BALB/c mice, or **(C, F)** quail or **(H)** OmCI-treated quail. The spirochete loads in **(A, D)** flat or **(B, C, E, F, G, H)** replate nymphs were determined by qPCR. Shown are the geometric mean ± geometric standard deviation of at least five nymphs per group. There was no statistical difference ($p > 0.05$) of the spirochete burdens between different groups of the replete ticks using a **(A to C)** Mann-Whitney test or **(D to H)** Kruskal-Wallis test with the two-stage step-up method of Benjamini, Krieger, and Yekutieli.

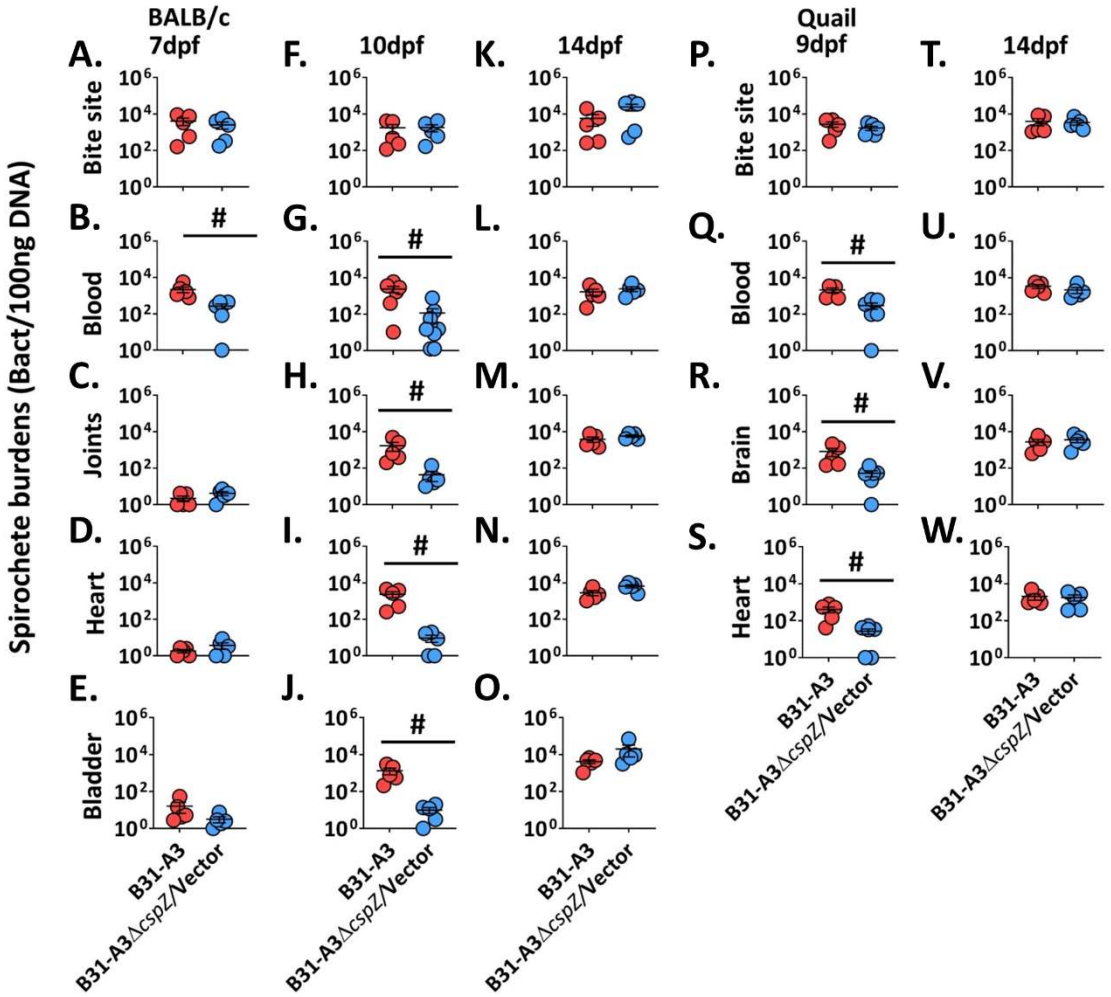


Figure S7. CspZ facilitates early bacteremia and distal tissue colonization during tick infection. The *I. scapularis* nymphs carrying *B. burgdorferi* strain B31-A3 or B31-A3Δ*cspZ* harboring the vector pKFSS ("Δ*cspZ*/Vector") were allowed to feed until repletion on **(A to O)** BALB/c mice or **(P to W)** quail. The mice were euthanized at **(A to E)** 7, **(F to J)** 10, or **(K to O)** 14 days post nymphs feeding (dpf), whereas the quail were euthanized at **(P to S)** 9 or **(T to W)** 14dpf. **(A, F, K)** The site of the skin where nymphs fed ("Bite site"), **(B, G, L)** blood, **(C, H, M)** tibiotarsus joints, **(D, I, N)** heart, and **(E, J, O)** bladder of mice; and **(P, T)** the bite site, **(Q, U)** blood, **(R, V)** brain, and **(S, W)** heart of quail, were collected immediately after euthanasia and spirochete loads were determined by qPCR. The burdens were normalized to 100ng total DNA.

1285 Shown are the geometric mean of bacterial loads ± SEM of five mice or quail per group, except
1286 for samples from the mouse blood at 7 and 10dpf where the results are from six and nine mice,
1287 respectively. Significant differences ($p < 0.05$, Mann-Whitney test) in the spirochete burdens
1288 between two strains relative to each other ("#") are indicated.
1289
1290
1291
1292
1293
1294
1295
1296
1297
1298
1299
1300
1301
1302
1303
1304
1305

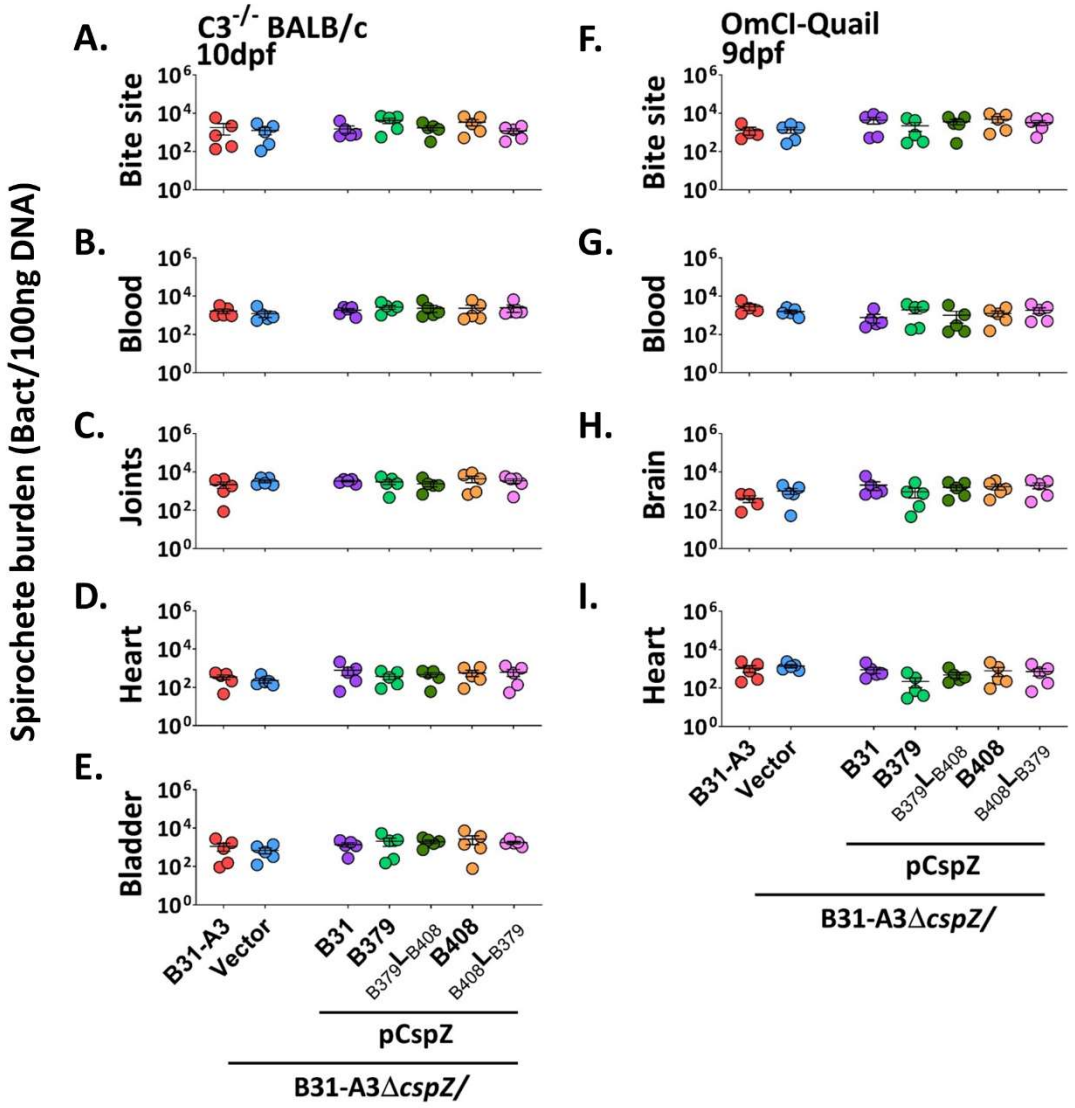


Figure S8. The CspZ loop-mediated early hematogenous dissemination is recovered when complement is depleted from hosts. *I. scapularis* nymphs carrying B31-A3, B31-A3Δ*cspZ* harboring the empty vector pKFSS ("Δ*cspZ*/Vector"), or this *cspZ* mutant strain producing $CspZ_{B31}$, $CspZ_{B379}$, $CspZ_{B379}L_{B408}$, $CspZ_{B408}$, or $CspZ_{B408}L_{B379}$ were allowed to feed until repletion on **(A to E)** $C3^{-/-}$ mice in a BALB/c background or **(F to I)** OMCI-treated quail, both of which deplete complement in the respective hosts. The bacterial loads in the indicated distal tissues were determined by qPCR and were normalized to 100ng total DNA, at 10 days post nymphs feeding (dpf) in mice and 9dpf in quail. Shown are the geometric mean of bacterial loads ± SEM of five

1315 mice or quail per group, except for the blood from BALB/c mice, which has nine mice per group.
1316 There were no significant differences ($p < 0.05$, Kruskal-Wallis test with the two-stage step-up
1317 method of Benjamini, Krieger, and Yekutieli) in the spirochete burdens for any strain.
1318
1319
1320
1321
1322
1323
1324
1325
1326
1327
1328
1329
1330
1331
1332
1333
1334
1335

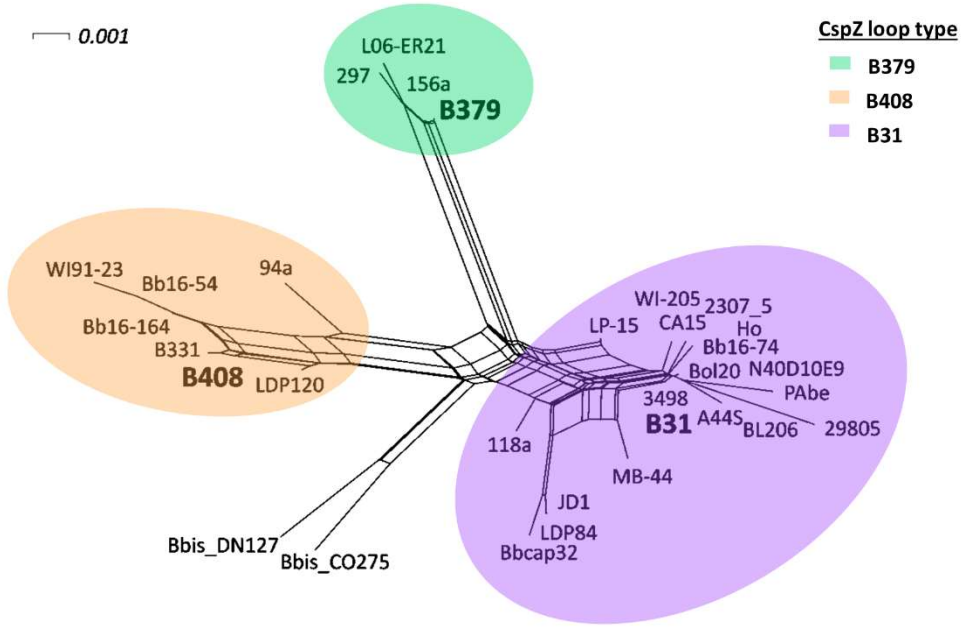


Figure S9. Phylogenetic network of *cspZ* haplotypes. Edges are colored based on their loop structure ($CspZ_{B379}$, $CspZ_{B408}$, and $CspZ_{B31}$ in green, orange, and purple, respectively).

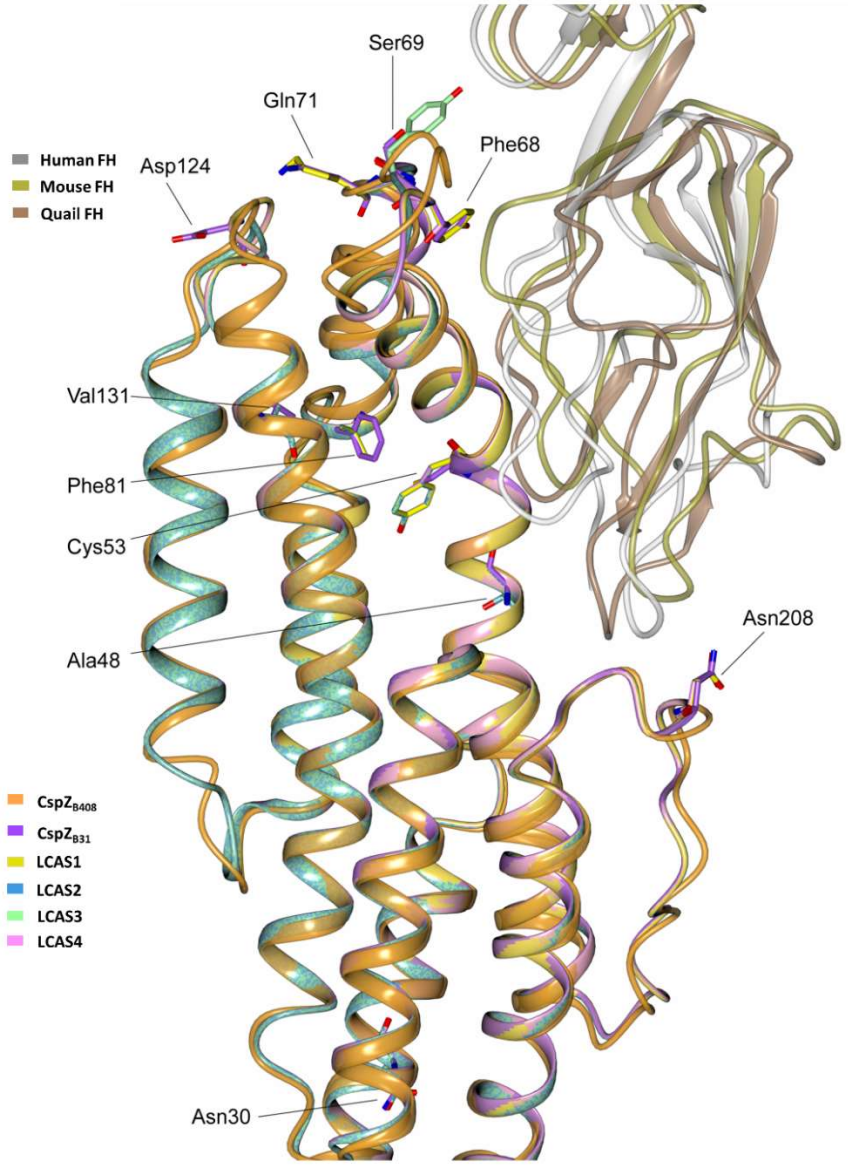
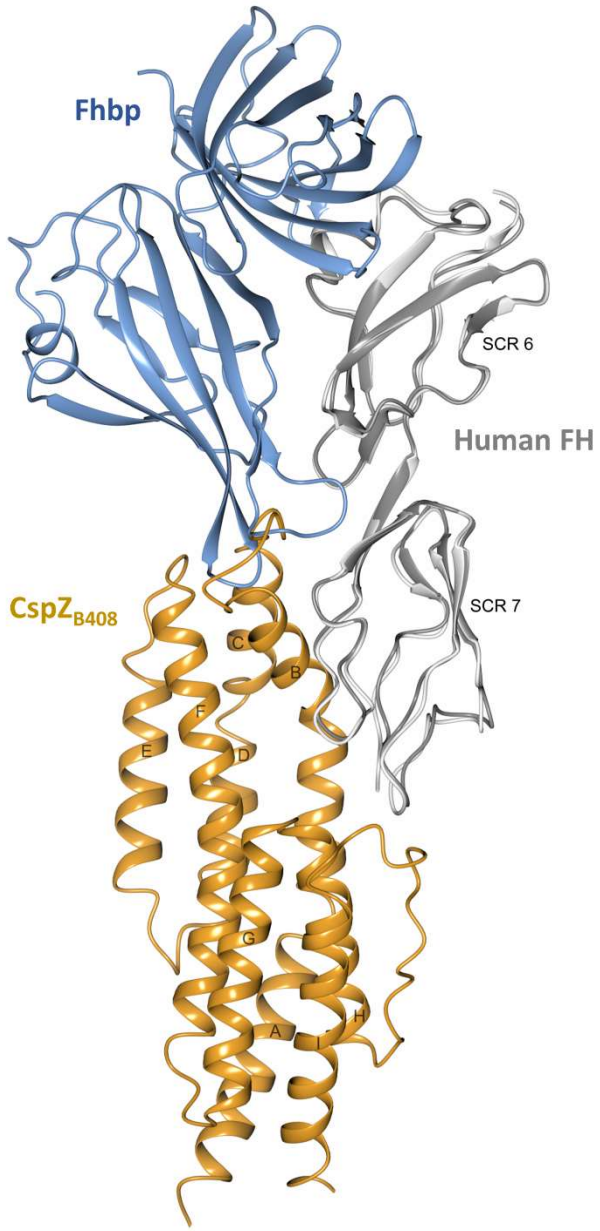


Figure S10. The CspZ last common ancestor states are predicted to versatiley bind FH. The crystal structure of $CspZ_{B408}$-SCR6-7 where human SCR6-7 (grey) is superimposed with mouse FH SCR6-7 (gold, PDB: 2YBY) and the predicted structure of quail SCR6-7 (brown). $CspZ_{B408}$ (orange) is superimposed with $CspZ_{B31}$ (purple) and the last common ancestor states: LCAS1 (yellow), LCAS2 (blue), LCAS3 (green), and LCAS4 (pink). Residues that differ between $CspZ_{B31}$ and the LCAS variants are labelled.



1357

1358 **Figure S11. Structural comparison between CspZ$_{B408}$-human SCR6-7 and the *N. meningitidis***
1359 **Fhbp-human SCR6-7 complexes.** Human FH SCR6-7 (light grey) from the complex structure
1360 with CspZ$_{B408}$ (orange) was superimposed with human FH SCR6-7 (dark grey) from the complex
1361 structure with Fhbp (dark blue, PDB: 2W81), the FH-binding protein from *N. meningitidis.* α-
1362 helices in CspZ$_{B408}$ are labelled from A to I starting from the N-terminus.

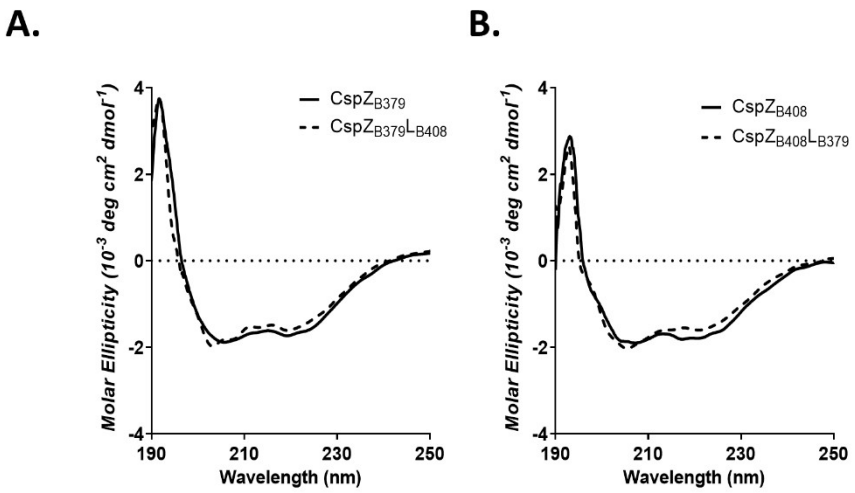
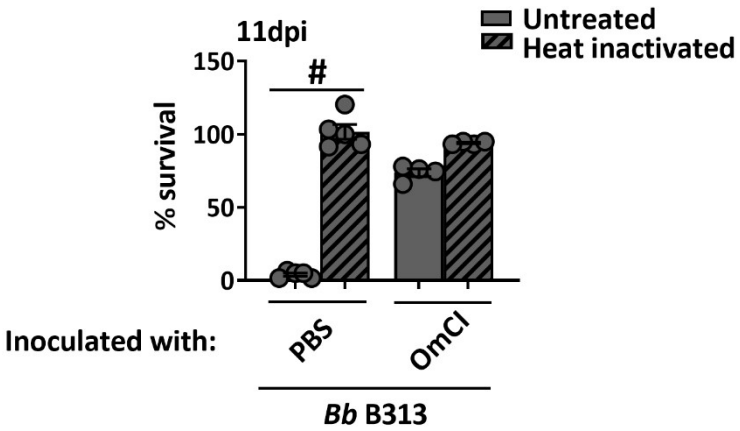


Figure S12. CD spectra demonstrate no impacts of secondary structures by swapping the loops. Far-UV CD analysis of $CspZ_{B379}$, $CspZ_{B408}$, $CspZ_{B379}L_{B408}$, and $CspZ_{B408}L_{B379}$. The molar ellipticity, Φ, was measured from 190 to 250nm for 10μM of each protein in PBS.



1379

1380 **Figure S13. OmCI prevents quail serum-mediated killing of a complement-sensitive**
1381 **spirochete strain at 11 days post injection.** *Coturnix* quail were subcutaneously inoculated with
1382 OmCI (1mg/kg of quail) or PBS buffer. Untreated (filled bars) or heat-treated (hatched bars) sera
1383 collected from these quail at 11 days post inoculation (dpi) were incubated with a serum-sensitive,
1384 highly passaged *B. burgdorferi* strain B313. The number of motile spirochetes were quantified
1385 microscopically and the survival percentage of the spirochetes was calculated using the number of
1386 mobile spirochetes at 4 h post incubation normalized to that at 0 h. Each bar represents the mean
1387 ± SEM of three independent experiments from sera from four quail per group. Significant
1388 differences ($p < 0.05$, Mann-Whitney test) in the percentage survival of spirochetes are indicated
1389 ("#").

1390

1391