

1 Development of recombinant monoclonal antibodies targeting  
2 conserved VlsE epitopes in Lyme disease pathogens

3

4 Li Li<sup>1</sup>, Lia Di<sup>2</sup>, Saymon Akther<sup>1</sup>, Brian M. Zeglis<sup>1,3,4</sup>, Weigang Qiu<sup>1,2,6,\*</sup>

5

6 <sup>1</sup>Graduate Center, City University of New York, NY, USA

7 <sup>2</sup>Department of Biological Sciences, Hunter College, City University of New York, NY, USA

8 <sup>3</sup>Department of Chemistry, Hunter College, City University of New York, NY, USA

9 <sup>4</sup>Department of Radiology, Weill Cornell Medical College, NY, USA

10 <sup>5</sup>Department of Radiology, Memorial Sloan Kettering Cancer Center, NY, USA

11 <sup>6</sup>Department of Physiology and Biophysics & Institute for Computational Biomedicine, Weill  
12 Cornell Medical College, NY, USA

13

14 \* Correspondence: Weigang Qiu, Email: [wqiu@hunter.cuny.edu](mailto:wqiu@hunter.cuny.edu). Telephone: 1-212-896-0445

15

16 Competing Interests statement:

17 The following authors declare potential conflicts of interest: BMZ (patents) and WGQ (patents).

18 All other authors declare no conflicts of interest in relation to this study.

19

20

21 **Abstract**

22 VlsE (variable major protein-like sequence, expressed) is an outer surface protein of the  
23 Lyme disease pathogen (*Borrelia* species) and a key diagnostic biomarker of Lyme disease.  
24 However, the high sequence variability of VlsE poses a challenge to the development of  
25 consistent VlsE-based diagnostics and therapeutics. In addition, the standard diagnostic protocols  
26 detect immunoglobins elicited by the Lyme pathogen, not the presence of pathogen or its derived  
27 antigens. Here we describe the development of recombinant monoclonal antibodies (rMAbs) that  
28 bind specifically to conserved epitopes on VlsE. We first quantified amino-acid sequence  
29 variability encoded by the *vls* genes from thirteen *B. burgdorferi* genomes by evolutionary  
30 analyses. We showed broad inconsistencies of the sequence phylogeny with the genome  
31 phylogeny, indicating rapid gene duplications, losses, and recombination at the *vls* locus. To  
32 identify conserved epitopes, we synthesized peptides representing five long conserved invariant  
33 regions (IRs) on VlsE. We tested the antigenicity of these five IR peptides using sera from three  
34 mammalian host species including human patients, the natural reservoir white-footed mouse  
35 (*Peromyscus leucopus*), and VlsE-immunized New Zealand rabbits (*Oryctolagus cuniculus*). The  
36 IR4 and IR6 peptides emerged as the most antigenic and reacted strongly with both the human  
37 and rabbit sera, while all IR peptides reacted poorly with sera from natural hosts. Four rMAbs  
38 binding specifically to the IR4 and IR6 peptides were identified, cloned, and purified. Given  
39 their specific recognition of the conserved epitopes on VlsE, these IR-specific rMAbs are  
40 promising diagnostic and theragnostic agents for direct detection of Lyme disease pathogens  
41 regardless of strain heterogeneity.

## 42 Introduction

43 Lyme disease is a multistage, tick-transmitted infection caused by spirochetes of the  
44 bacterial species complex *Borrelia burgdorferi* *sensu lato* (*Bbsl*), known more concisely (albeit  
45 controversially) as a new genus *Borrelia* (Barbour and Qiu, 2019; Margos *et al.*, 2020). Lyme  
46 disease is the most common tick-borne disease in regions of North America, Europe, and Asia  
47 (Stanek *et al.*, 2012; Kilpatrick *et al.*, 2017). In the United States, approximately 476,000 cases  
48 are diagnosed annually (Kugeler *et al.*, 2021). The majority of Lyme disease cases in the US are  
49 caused by the single species *B. burgdorferi* and transmitted by the hard-bodied *Ixodes scapularis*  
50 or *I. pacificus* ticks, although the same tick vectors carry other *Borrelia* species as well as  
51 *Borrelia* species closely related to relapsing fever spirochetes (Barbour *et al.*, 2009; Pritt *et al.*,  
52 2016; Schwartz *et al.*, 2021). *B. burgdorferi* cause multisystemic manifestations in humans  
53 including erythema migrans (EM) at early stages, arthritis, carditis, and neuroborreliosis in late  
54 stages, and chronic symptoms associated with persistent infections (Stanek *et al.*, 2012; Sharma  
55 *et al.*, 2015; Feng *et al.*, 2019).

56 Antigenic variation via continuously altering the sequences of surface antigens during  
57 infection is a common strategy that microbial pathogens employ to escape adaptive immune  
58 responses of vertebrate hosts (Vink, Rudenko and Seifert, 2012; Palmer, Bankhead and Seifert,  
59 2016). In the two closely spirochetal groups of *Borrelia* causing relapsing fever and *Borrelia*  
60 causing Lyme disease, two homologous but distinct molecular systems have evolved facilitating  
61 continuous antigenic variation through recombination between an expressed locus and silent  
62 archival loci during persistent infection within the vertebrate hosts (Norris, 2006). In *B.*  
63 *burgdorferi*, the molecular system able to generate antigenic variation consists of one expression  
64 site (*vlsE*, variable major protein-like sequence, expressed) and a set of tandemly arranged silent

65 cassettes (*vlsS*) that share more than 90% similarities to the central cassette region of *vlsE*  
66 (Zhang *et al.*, 1997; Norris, 2014; Verhey, Castellanos and Chaconas, 2019) (Fig 1). During  
67 mammalian infection, *vlsE* continuously expresses and undergoes random segmental  
68 recombination with the silent cassettes, generating a considerable number of new VlsE antigen  
69 variants to prolong spirochete infection in hosts (Norris, 2006; Verhey, Castellanos and  
70 Chaconas, 2019).

71 The *vlsE* gene encodes a 36 kD lipoprotein that is anchored to the outer membrane on the  
72 cell surface. The primary structure of VlsE comprises the N- and C-terminal domains as well as  
73 the central cassette which consists of six highly variable regions (VR1-VR6) interspersed with  
74 six conserved invariant regions (IR1-IR6) (Fig 1). The N- and C-terminal regions do not undergo  
75 antigenic variation and are thought to be important in maintaining the functional structure of the  
76 molecule (Norris, 2014). The VRs of the cassette are the sequences that undergo antigenic  
77 variation during infection, while the IRs are conserved among *B. burgdorferi* strains (Liang *et al.*,  
78 1999). The crystal structure of recombinant VlsE protein revealed that the six VRs constitute  
79 loop structures and form a “dome” on the membrane distal surface exposed to the host  
80 environment, which may shield the IRs from antibody binding (Eicken *et al.*, 2002).

81 VlsE elicits strong humoral responses that can be detected throughout the course of Lyme  
82 disease, making it a powerful antigen in serologic assays of Lyme disease diagnosis (Lawrenz *et*  
83 *al.*, 1999; Bacon *et al.*, 2003; Elzbieta *et al.*, 2016). Contrary to the established paradigm of weak  
84 immunogenicity of the conserved regions of bacterial surface proteins, the conserved IR6 elicits  
85 immunodominant antibody responses during human infection despite the region being largely  
86 inaccessible on the intact VlsE molecule (Liang *et al.*, 1999; Chandra *et al.*, 2011; Elzbieta *et al.*,  
87 2016). The surprising finding of immunodominance of IR6 in human patients is hypothesized to

88 be a result of antigen processing of the VlsE proteins in non-reservoir host species (Embers *et al.*,  
89 2007).

90 A 26-amino acid peptide that reproduces the IR6 sequence, known as C6 peptide, is used in  
91 commercial diagnostic tests of Lyme disease (Bacon *et al.*, 2003; Wormser *et al.*, 2013). The  
92 standardized two-tiered testing (STTT) for Lyme disease diagnosis includes a screening enzyme  
93 immunoassay (EIA) with the whole cell sonicate and a subsequent confirmatory Western blot  
94 assay for the presence of both IgM and IgG antibodies against ten *Borrelia* antigens (CDC,  
95 1995; Moore *et al.*, 2016). Recently, a modified two-tiered testing (MTTT) protocol using two  
96 sequential EIAs with the C6 peptide or the whole VlsE protein has been developed. MTTT  
97 improved sensitivity and specificity relative to STTT, especially in Lyme patients with early-  
98 stage manifestations (Branda *et al.*, 2017). Nevertheless, the overall sensitivity for early-stage  
99 diagnosis remains low, ranging from 36% to 54%, even with MTTT (Pegalajar-Jurado *et al.*,  
100 2018). In addition, both diagnostic assays are indirect tests and do not distinguish between active  
101 infection and past exposure. In sum, there is a need to simplify the testing protocol for Lyme  
102 disease, improve testing sensitivity in the early infection stage, and detect the presence of Lyme  
103 pathogen or its derivative antigens directly.

104 During the transmission cycle of *B. burgdorferi*, the *vls* locus is expressed during the late-  
105 stage persistent infection within the mammalian host, in contrast to genes like *ospA* (encoding  
106 outer surface protein A) expressed within the ticks, and genes like *ospC* expressed exclusively  
107 during a short window of time when the spirochetes begin migrate from the tick to the  
108 mammalian host (Samuels, 2011; Tilly, Bestor and Rosa, 2013). As a multi-copy gene family  
109 and driven by adaptive amino-acid substitutions, the *vls* cassettes exhibit high sequence  
110 variability not only between *B. burgdorferi* strains but also within the same genome (Glöckner *et*

111 *al.*, 2004; Schutzer *et al.*, 2011; Graves *et al.*, 2013). In the present study, we developed a  
112 bioinformatics workflow to facilitate the automated identification of *vls* sequences from the  
113 sequenced *Borrelia* genomes. We quantified evolutionary rates at individual amino acid sites  
114 of the *vls* coding sequences identified from thirteen *B. burgdorferi* genomes. Extending the  
115 previous analysis of mechanisms of evolution at the *vls* locus (Graves *et al.*, 2013; Schwartz *et*  
116 *al.*, 2021), we explored the evolution mechanisms by comparing the *vls* gene phylogeny with the  
117 genome-derived strain phylogeny. Our experimental investigations of the immunogenicity of the  
118 *VlsE* protein confirmed the immunodominance of the IR6 peptide and discovered the similar  
119 immunodominance of the IR4 peptide in human patients and immunized rabbits but not the  
120 reservoir hosts. Finally, we identified, cloned, and purified four recombinant IR-specific  
121 monoclonal antibodies (rMAbs) that are promising theragnostic agents for direct assay of *B.*  
122 *burgdorferi* infection in clinical samples and model organisms of Lyme disease.

123 **Materials and Methods**

124 **Identification of *vls* cassette sequences and evolutionary analysis**

125 We downloaded the whole genome sequences of 12 *B. burgdorferi* strains from NCBI  
126 GenBank (Schutzer *et al.*, 2011). The *vlsS* sequences of B31-5A3 clone (GenBank accession  
127 U76406) (Zhang *et al.*, 1997) were used as the queries to searched for sequences homologous to  
128 the *vls* cassette sequences using HMMER (version 3.3.2) (Eddy, 2011). A customized web-based  
129 software tool was developed to identify and extract individual *vls* sequences given a *B.*  
130 *burgdorferi* replicon sequence (<http://borreliabase.org/vls-finder>). The identified *vlsS* and *vlsE*  
131 sequences were translated, aligned, and converted into a codon alignment using MUSCLE  
132 (version 3.8.31) (Edgar, 2004) and the *bioaln* utility (–*dna2pep* method) of the BpWrapper  
133 toolkit (Hernández *et al.*, 2018). A maximum likelihood tree was subsequently

134 inferred using IQ-TREE (version 1.6.1) with the best-fit nucleotide substitute model KOSI07 and  
135 1000 bootstrap replicates (Nguyen *et al.*, 2015). Branches with lower than 80% bootstrap support  
136 were collapsed using the *biotree* utility (-D method) of the BpWrapper utility (Hernández *et al.*,  
137 2018). The tree was rendered using the R package *ggtree* (Version 2.2.4) (Yu *et al.*, 2017). To  
138 quantify the sequence conservation, evolutionary rates at individual amino acid positions were  
139 estimated using Rate4Site (version 3.0.0) with the protein alignment and the phylogenetic tree as  
140 inputs and the B31-5A3 VlsE sequence (GenBank accession U76405) as the reference (Pupko *et*  
141 *al.*, 2002). Sequence conservation at the IRs was further quantified and visualized with WebLogo  
142 (Crooks *et al.*, 2004).

#### 143 [Synthesis of peptides representing conserved epitopes of VlsE](#)

144 The preparation of the peptides was based on the annotation of B31-5A3 VlsE protein  
145 sequence in the literature (Zhang *et al.*, 1997). Five invariant regions, IR1, IR2, IR4, IR5, and  
146 IR6, were tested for antigenicity using sera from three host species. IR3 (AGKLFVK), the  
147 shortest IR, was excluded from the antigenicity test. Extra flanking amino acids were added to  
148 IR2, IR4, and IR5 to meet the minimum length for peptide synthesis. Peptides were  
149 commercially synthesized and biotin-labeled on the N-terminus using Fmoc chemistry  
150 (GenScript, Piscataway, NJ, USA). Sequences of these peptides are shown in Table 1.

#### 151 [Sera collection from naturally infected hosts](#)

152 The 56 serum samples, consisting of Lyme patient and control sera provided by the US  
153 Center for Disease Control and Prevention (CDC,  $n = 40$ ), Lyme patient sera provided by Dr  
154 Maria Gomes-Solecki (University of Tennessee Health Science Center,  $n = 6$ ), and sera from the  
155 reservoir hosts (white-footed mouse, *Peromyscus leucopus*) provided also by Dr Maria Gomes-  
156 Solecki ( $n = 10$ ), have been used and described in previous publications (Ivanova *et al.*, 2009;

157 Molins *et al.*, 2014; Di *et al.*, 2021). Briefly, among the human samples, 25 serum samples were  
158 derived from patients with early-stage Lyme disease including those diagnosed as having the  
159 skin symptom erythema migran (EM) or as EM convalescence. Seventeen human sera samples  
160 were from patients displaying late-stage Lyme disease symptoms including arthritic, cardiac, and  
161 neurological Lyme diseases. Four human sera samples were from healthy individuals as controls.

162 [Cloning, over-expression, and purification of recombinant VlsE protein](#)

163 Recombinant VlsE protein from the B31 strain was cloned, over-expressed, and purified  
164 using a protocol described previously (Di *et al.*, 2021). Briefly, the 585-bp *vlsE* cassette region  
165 (including the direct repeat regions on both ends) of B31-5A3 clone was codon-optimized,  
166 synthesized, and cloned into the pET24 plasmid vector which then transfected *Escherichia coli*  
167 BL21 cells. A10 × Histidine-tag was added on the N-terminus of the construct to facilitate the  
168 downstream purification. All cloning work was performed by a commercial service  
169 (GeneImmune Biotechnology Corp., Rockville, MD, USA). The *E. coli* strain that contained a  
170 cloned *vlsE* cassette was cultured in Luria-Bertani (LB) broth containing 0.4% glucose and 50  
171 µg/ml Ampicillin. When the culture reached exponential growth, we induced the expression of  
172 the cloned *vlsE* cassette by adding isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final  
173 concentration of 0.25 mM and by incubation overnight at 25 °C. Cells were collected and then  
174 lysed by lysozyme and sonication. The lysate supernatant, containing the recombinant VlsE  
175 protein, was purified using nickel sepharose beads (Ni-NTA, Thermo Fisher Scientific, Waltham,  
176 MA, USA) following the manufacturer's protocol. The identity and concentration of the purified  
177 protein was examined and quantified using the sodium dodecyl sulphate polyacrylamide gel  
178 electrophoresis (SDS-PAGE) and the Pierce Bradford Protein Assay Kit (Thermo Fisher  
179 Scientific, Waltham, MA, USA).

180 Immunization of rabbits and preparation of polyclonal and monoclonal antibodies

181 Antibody preparation was conducted with a commercial service GenScript (Piscataway,  
182 NJ, USA). Briefly, the project consisted of four stages. In Stage 1, animals were immunized and  
183 polyclonal antibodies were obtained. Specifically, four New Zealand rabbits (*Oryctolagus*  
184 *cuniculus*) were immunized with 100 µg purified recombinant VlsE protein on Days 1, 14, and  
185 28. The rabbits were bled for antiserum collection one week after the third immunization. The  
186 antisera were subsequently purified by affinity chromatography to obtain polyclonal antibodies  
187 (pAbs), which were assayed for anti-VlsE activity. In Stage 2, monoclonal antibodies (MAbs)  
188 were identified via single B cell sorting. Peripheral blood mononuclear cells (PBMC) were  
189 collected from the two selected immunized rabbits one week after a booster dose with the  
190 purified VlsE. Plasma B cells (CD138+) were isolated and enriched using a commercial kit. B  
191 cells were then screened for VlsE-specific cell lines using ELISA. The supernatants of positive  
192 cell lines were used to test for binding with VlsE and positive cell lines were chosen for mAb  
193 production. In Stage 3, the variable domains of the light and heavy chains of the VlsE-binding  
194 antibodies were sequenced. Total RNA was isolated from the VlsE-binding B cell lines and  
195 reverse-transcribed into cDNA using universal primers. Antibody fragments of the heavy chain  
196 and the light chain were amplified and sequenced. In Stage 4, the recombinant MAbs (rMAbs)  
197 were produced. The amplified antibody variable fragments were cloned into plasmid vector  
198 pcDNA3.4 which then transfected mouse cells for expression. Supernatants of cell cultures were  
199 harvested continuously. The rMAbs were purified using the Protein A/G affinity chromatography  
200 (with immobilized Protein A and G from *Staphylococcus aureus*) followed by size exclusion  
201 chromatography (SEC).

202 Identification of IR-specific mAbs with ELISA

203 We tested sera from naturally infected hosts for reactivity to the IR peptides (Table 1) and  
204 recombinant VlsE protein with ELISA using a protocol described previously (Di *et al.*, 2021).  
205 Briefly, a 96-well MICROLON 600 plate (USA Scientific, Inc., Ocala, FL, USA) was incubated  
206 with 10 µg/ml of antigen overnight at 4 °C. Serum samples diluted between 1:100 to 1:1000 were  
207 applied after blocking with 5% milk and were incubated for 2 h at 37 °C, followed by application  
208 of horseradish peroxidase (HRP)-conjugated secondary antibodies. We used the Goat Anti-  
209 Human IgG/IgM (H + L) (Abcam, Cambridge, UK) 1:40,000 for assays of human sera and the  
210 Goat Anti-*P. leucopus* IgG (H + L) (SeraCare Life Sciences, MA, USA) 1:1000 for assays of *P.*  
211 *leucopus* sera. The antigen-antibody reaction was probed by TMB ELISA Substrate Solution  
212 (Invitrogen eBioscience) and was terminated with 1M sulfuric acid after 15 minutes. Binding  
213 intensities were measured at the 450 nm wavelength using a SpectraMax i3 microplate reader  
214 (Molecular Devices, LLC, CA, USA).

215 The purified anti-VlsE pAbs, the supernatants of selected B cell cultures, and the purified  
216 mAbs were tested for reactivity to the IR peptides and the purified recombinant VlsE protein  
217 with ELISA using the same protocol as described above. Mouse Anti-Rabbit IgG Fr secondary  
218 antibody (GenScript, Piscataway, NJ) 1:30,000 was used for assays of these rabbit-derived  
219 samples. Serial dilutions of MAbs by factors from 1,000 to 512,000 were tested with ELISA to  
220 quantify the binding activities.

221 Protein structure visualization

222 The PDB file of VlsE protein structure (accession 1L8W) was downloaded from the  
223 protein data bank (PDB) (<https://www.rcsb.org/>) (Eicken *et al.*, 2002). The PDF file describes a

224 tetramer of VlsE. We used Chimera (version 1.15) (Pettersen *et al.*, 2004) to visualize the protein  
225 structure in ribbon and surface-filled formats and to color the six invariable regions (IR1-6).

226 [Animal care, data visualization, statistical analysis, and data and code availability](#)

227 Antibody production from the New Zealand rabbits followed the protocols approved by  
228 the Office of Laboratory Animal Welfare (OLAW) Assurance and the Institutional Animal Care  
229 and Use Committee (IACUC) of the vendor (GenScript, Piscataway, NJ).

230 Data visualization and statistical analysis were performed in the R statistical computing  
231 environment (R Core Team, 2013) accessed with RStudio. The alignment of translated *vls*  
232 sequences, ELISA readings, and R scripts are publicly available on Github at  
233 <https://github.com/weigangq/vls-mabs>.

234 **Results**

235 [Phylogenetic inconsistencies indicate duplications and losses, sequence divergence, and](#)  
236 [recombination at the \*vls\* locus](#)

237 We identified 194 *vls* cassette sequences from 13 *B. burgdorferi* strains and inferred a  
238 maximum likelihood tree of the cassette (Fig 2). These *B. burgdorferi* strains have been  
239 classified into four phylogenetic groups (A-D) based on chromosomal single-nucleotide  
240 polymorphisms (SNPs) (Mongodin *et al.*, 2013). The *vls* gene phylogeny consists of eight major  
241 clades and is consistent with a previously published *vls* cassette phylogeny (Graves *et al.*, 2013).  
242 Here we analyzed the *vls* gene phylogeny in the broader context of strain phylogeny.  
243 Phylogenetic inconsistencies between gene and strain trees may result from – and thus indicate  
244 the occurrence of – horizontal gene transfers between strains, ancestral gene duplications  
245 followed by the loss of duplicated copies, and incomplete lineage sorting when strains rapidly  
246 diverge from one other (Rogers *et al.*, 2017; Kundu and Bansal, 2018).

247 The *vls* sequences from the two SNP group D strains (JD1 and 156a) formed a  
248 monophyletic group consistent with the strain phylogeny. However, within this major clade the  
249 *vls* sequences did not separate into two strain-specific clades, a phylogenetic inconsistency that  
250 could be caused by gene duplications in a common ancestor followed by losses of duplicated  
251 copies or by incomplete lineage sorting, but unlikely by horizontal gene exchanges which would  
252 have introduced *vls* sequences from other SNP groups.

253 In contrast, the *vls* sequences from the strains belonging to the SNP groups A, B, and C  
254 all formed paraphyletic groups, each of which contained multiple clades highly divergent from  
255 one another than one would expect from the strain phylogeny (Fig 2). In the SNP group A, the  
256 *vls* sequences from the strain BOL26 formed a clade highly divergent from the *vls* sequences  
257 from the strains B31, PABe, 64B, and ZS7. In the SNP group B, the *vls* sequences from three  
258 strains (WI91-23, N40, and 29805) formed three strain-specific clades. The *vls* sequences from  
259 strains belonging to the SNP group C were split into two clades, one consisting of the sequences  
260 from the strain 94a and the other consisting of sequences from the strains 72a and 118a.

261 As in the group D, the *vls* sequences within the SNP groups A, B, and C did not sort into  
262 strain-species clades, indicating frequent gene duplications, rapid gene losses, and fast sequence  
263 divergence within each of these phylogenetic groups. Indeed, it has been shown that the rapid  
264 sequence evolution of the *vls* cassettes was driven by adaptive differentiation evidenced by the  
265 accelerated nonsynonymous nucleotide substitutions (i.e., a high *dN/dS* ratio) (Graves *et al.*,  
266 2013).

## 267 Evolutionary rates and molecular structure of *vls* cassettes

268 Rates of amino-acid substitutions are not uniform along the translated *vls* sequence,  
269 which consists of mostly fast-evolving variant regions (VRs) interspersed with six short

270 conserved invariant regions (IR1-6) (Zhang *et al.*, 1997). Here we quantified *vls* variability at  
271 individual amino-acid sites among the 13 *B. burgdorferi* strains using the 194 *vls* sequences  
272 including both the expressed and unexpressed cassettes. Conserved regions were detected by  
273 computing the relative evolutionary rate of each amino-acid site in the multiple sequence  
274 alignment, with the average variability score scaled to zero (Fig 3). Most residues in the IRs  
275 showed negative variability scores, indicating below-average evolutionary rates. The mean  
276 variability score for each IR was shown in Table 1. Among the IRs, IR1 was the least conserved  
277 followed by IR4. The IR2, IR3, and IR5 were conserved but relatively short. The IR6 was highly  
278 conserved at all 25 residues.

279 We further mapped the IRs to a published three-dimensional structure of the VlsE protein  
280 (from the strain B31) (Eicken *et al.*, 2002) (Fig 4). All the IRs formed alpha helices, as the  
281 ribbon model showed (Fig 4A). The space-filled model showed that the IR1, IR2, and IR4 were  
282 partially surface exposed while the IR2, IR5, and IR6 exhibited limited surface exposure (Fig  
283 4B). The VlsE molecules likely form dimers on the spirochete cell surface (Eicken *et al.*, 2002),  
284 which would further shield the invariant regions located on the monomer-monomer interface  
285 (Fig 4C and 4D). Nevertheless, the IR4 and IR6 are partially exposed at the membrane distal  
286 surface even in a dimerized form (Fig 4D).

#### 287 Antigenicity of IRs against host sera

288 We measured antigenicity of the IRs with sera from human patients, white-footed mice,  
289 and immunized rabbits using ELISA. For the 46 human sera, the ELISA result showed an overall  
290 significant difference in the mean OD450 values among the antigens ( $p < 2.2\text{e-}16$  by ANOVA)  
291 (Fig 5, left panel). Reactivities of the IR4 and IR6 peptides with the human sera were  
292 significantly higher than that of BSA ( $p = 2.87\text{e-}13$  and  $3.06\text{e-}13$  by ANOVA, respectively),

293 while reactivities of IR1, IR2, and IR5 were less significant ( $p = 0.034$ ,  $0.034$ , and  $0.0019$  by  
294 ANOVA, respectively) (Fig 5, left panel). Reactivity of VlsE was the strongest ( $p < 2.2\text{e-}16$  by  
295 ANOVA). In addition, reactivities of the IR4 and IR6 with the human sera were weakly although  
296 significantly correlated with those of the VlsE ( $p = 7.6\text{e-}4$  and  $R^2=0.212$  for IR4,  $p = 3.6\text{e-}3$  and  
297  $R^2=0.158$  for IR6, both by linear regression). There was no significant difference in reactivity  
298 between the early and late-stage patient samples ( $p = 0.8654$  by *t*-test).

299 Ten serum samples from whited-footed mice, the natural reservoir host of *B. burgdorferi*  
300 were tested. Reactivities of the IR peptides against the mouse sera showed little differences  
301 among the antigens ( $p = 0.0159$  by ANOVA), with only VlsE showing a significant difference  
302 from the BSA control ( $p = 2.9\text{e-}3$  by ANOVA) (Fig 5, middle panel). These results are consistent  
303 with findings of an earlier study which showed low antigenicity of the IR6 peptide in natural  
304 hosts relative to its antigenicity in humans (Liang *et al.*, 1999).

305 Reactivities of the IRs against the four sera from four immunized rabbits showed a  
306 similar pattern as those against the naturally infected human (Fig 5, right panel). For example,  
307 VlsE, IR4, and IR6 peptides displayed the highest antigenicity ( $p = 6.8\text{e-}10$ ,  $1.2\text{e-}7$ , and  $2.1\text{e-}8$ ,  
308 respectively with an overall  $p = 2.2\text{e-}10$  by ANOVA). Antigenicity of the IR1, IR2, and IR5  
309 peptides against the rabbit polyclonal antibodies did not differ or differed weakly from that of  
310 BSA, the negative control ( $p = 0.855$ ,  $0.011$ , and  $0.236$  by ANOVA, respectively).

311 In sum, these ELISA results suggested that (1) anti-VlsE antibodies were present in  
312 patients throughout different stages of Lyme disease, (2) antibodies against the VlsE IRs were  
313 strongly present in naturally infected or artificially immunized non-reservoir hosts but minimally  
314 present in reservoir hosts, and (3) the IR4 and IR6 peptides were highly immunogenic conserved  
315 epitopes on the VlsE molecule in non-reservoir hosts relative to the IR1, IR2, and IR5 peptides.

316 These results are consistent with conclusions of earlier studies on the antigenicity of VlsE and  
317 conserved epitopes, which established the use VlsE and the C6 peptide (derived from IR6) in  
318 both the standard and modified diagnostics tests of Lyme disease (Liang *et al.*, 1999; Liang and  
319 Philipp, 1999, 2000; McDowell *et al.*, 2002; Price, Dehal and Arkin, 2010; Branda *et al.*, 2017;  
320 Pegalajar-Jurado *et al.*, 2018; Lone and Bankhead, 2020).

321 Here we established that the IR4 peptide was as antigenic as the IR6 peptide. Indeed,  
322 both IR peptides reacted at a level similar to the reactivity of the whole VlsE protein with the  
323 sera from naturally infected and immunized hosts (Fig 5). The use of the highly conserved IR4  
324 and IR6 as targets for theragnostic agents has the advantage that they are expected to exhibit  
325 antigenicity against a broad set of *B. burgdorferi* strains, with the potential to mitigate the  
326 challenge of strain-specific antigenicity of the highly variable antigens including VlsE and OspC  
327 (Bockenstedt *et al.*, 1997; Bhatia *et al.*, 2018).

### 328 Identification and characterization of recombinant IR-specific monoclonal antibodies

329 Recombinant VlsE of the strain B31 was over-expressed, purified, and used to immunize  
330 New Zealand rabbit (Fig 6, gel image). IR-specific antibodies were identified via B cell sorting  
331 and by testing the reactivity of the supernatant of the 20 B cell lines against the five IR peptides  
332 with ELISA. We found that one cell line (1D11) bound specifically to the IR6 peptide and five  
333 cell lines (7C9, 15E2, 17A8, 28D3, and 42G10) specifically to the IR4 peptide in addition to  
334 their binding to the purified VlsE protein (Fig 6, bar plots). Supernatants of the remaining  
335 fourteen B cell lines reacted with the purified VlsE protein but not with the IR peptides,  
336 suggesting that the majority of B cell lines in the immunized rabbit expressed antibodies  
337 recognizing epitopes located on the variable and not the conserved regions.

338 One pair of the most abundant heavy chain and light chain variable region ( $V_H$  and  $V_L$ )  
339 sequences in each of four IR-specific cell lines – including the anti-IR6 1D11 cell line and three  
340 top anti-IR4 cell lines – were identified by pyrosequencing and subsequently cloned and over-  
341 expressed. Specificity of the purified recombinant monoclonal antibodies (rMAb) were validated  
342 using ELISA. The initial rMAb cloned from the 1D11 cell line based on the most abundant  $V_H$   
343 and  $V_L$  sequences was not reactive to the IR6 peptide as the supernatant of the cell line did. A  
344 new rMAb – based on the second most abundant  $V_H$  and  $V_L$  sequences – was re-cloned and over-  
345 expressed and reacted with the IR6 peptide strongly and specifically. The  $V_H$  and  $V_L$  sequences  
346 of the four IR-specific rMAbs and their binding characteristics were obtained by titration  
347 experiments (Fig 7).

348 **Discussion**

349 **Rapid adaptive diversification of *vls* cassettes**

350 The *vls* gene system in *Borrelia* was discovered based on gene sequence homology  
351 with the *vsp/vlp* (variable small and large proteins) system in *Borrelia* spirochetes causing  
352 relapsing-fever (Zhang *et al.*, 1997; Norris, 2006). Since then, the molecular mechanism of  
353 segmental recombination between the expression site and the archival cassettes has been well  
354 characterized in *B. burgdorferi* B31, the type strain (Coutte *et al.*, 2009; Verhey, Castellanos and  
355 Chaconas, 2019; Chaconas, Castellanos and Verhey, 2020). In parallel, genome-based  
356 comparative analysis of the *vls* system among *Borrelia* species and among strains of the same  
357 species uncovered rapid evolution in sequence, copy number, and genomic location of the *vls*  
358 cassettes (Glöckner *et al.*, 2004; Graves *et al.*, 2013; Schwartz *et al.*, 2021).

359 In the present study, we showed pervasive phylogenetic inconsistencies between the *vls*  
360 gene tree and the genome-based strain tree, suggesting frequent gene duplications, gene losses,

361 and gene exchanges, in addition to adaptive sequence evolution at the locus (Fig 2). The highly  
362 divergent *vls* cassette sequences between phylogenetic sister strains are reminiscent of the rapid  
363 amino-acid sequence diversification at the locus encoding the outer surface protein C (*ospC*),  
364 another immunodominant antigen of *B. burgdorferi* (Barbour and Travinsky, 2010). Protein  
365 sequences of major *ospC* alleles diverge in a strain-specific fashion with an average sequence  
366 identity of ~75.9% among *B. burgdorferi* strains in the Northeast US, due to a history of  
367 recombination among coexisting strains and diversifying selection driven by host immunity and  
368 possibly host-species preferences (Wang *et al.*, 1999; Brisson and Dykhuizen, 2004; Haven *et al.*,  
369 2011; Di *et al.*, 2021). In contrast, the coding sequences of the *vls* cassettes vary at a significantly  
370 higher level between the eight major sequence clusters (~56.3% average sequence identity),  
371 while varying at a high level between copies of the genome as well (e.g., 81.0% for B31, 76.5%  
372 for N40, and 76%.4 for JD1) (Fig 2). The much greater sequence diversity among the *vls* alleles  
373 than diversity among the *ospC* alleles indicates more rapid evolution driven by more intense  
374 immune selection at the *vls* locus.

375 As more *Borrelia* genomes are sequenced, the bioinformatics workflow including the  
376 customized web-based tool (<http://borreliabase.org/vls-finder>) established in the present study  
377 will facilitate large-scale automated identification of *vls* sequences and a quantification of the  
378 rates of gene duplication, losses, exchanges, and sequence divergence in this key adaptive  
379 molecular system in *Borrelia*.

380 **Immunogenicity of the IRs in non-reservoir hosts**

381 The VlsE and its derivative C6 peptide (based on IR6) are key diagnostic antigens in  
382 serological tests of Lyme disease (Bacon *et al.*, 2003; Marques, 2015; Branda *et al.*, 2017). In the  
383 present study, we confirmed the predominant immunogenicity of IR6 in serum samples from

384 human patients and VlsE-immunized rabbits (Fig 5). In contrast to an earlier study but consistent  
385 with another one (Liang *et al.*, 1999; Chandra *et al.*, 2011), the IR4 peptide showed as a similar  
386 level of antigenicity as the IR6 peptide in all three host species. Indeed, epitopes on IR4 might be  
387 more immunodominant than the IR6 epitopes in rabbits, as we obtained five anti-IR4 cell lines  
388 and only one anti-IR6 cell line out of a total of twenty randomly selected VlsE-reactive B cell  
389 lines (Fig 6). The IR1, IR2 and IR5 appeared to be barely immunogenic in reservoir as well as  
390 non-reservoir hosts (Fig 6).

391 Epitope mapping studies suggested that the IR6 may function as a single conformational  
392 epitope (Liang and Philipp, 2000). On an intact VlsE molecule (or its dimerized structure), the  
393 IR6 is almost entirely buried underneath the membrane surface and immunofluorescence assays  
394 demonstrated that the IR6 was inaccessible to antibodies on intact spirochetes (Liang, Nowling  
395 and Philipp, 2000; Embers *et al.*, 2007; Elzbieta *et al.*, 2016). It appears paradoxical that IR4 and  
396 IR6, two highly conserved and mostly buried regions on VlsE, contain immunodominant  
397 epitopes in human patients. Evolutionary arms races drive co-diversification of the antigen  
398 sequences in microbial pathogens along with the sequences of antigen-recognition proteins in  
399 vertebrate hosts through population mechanisms like negative frequency-dependent selection  
400 (Schierup, Mikkelsen and Hein, 2001; Haven *et al.*, 2011; Papkou *et al.*, 2019). Regions on  
401 antigen molecules shielded from host immune systems, like the IRs on VlsE, are not under such  
402 diversifying selection and thereby expected to be conserved in molecule sequences. The paradox  
403 resolves itself however when one considers that the IRs were indeed weakly immunogenic in the  
404 infected mice that belong to the natural reservoir species of *B. burgdorferi* (Fig 6). Indeed, as the  
405 whole VlsE molecule elicits significant antibody responses in the infected *P. leucopus* mice,

406 such immunogenicity is likely due to epitopes on the variable regions as expected from the  
407 pathogen-host co-evolutionary arms race (McDowell *et al.*, 2002) (Fig 6).

408 It is likely that *B. burgdorferi* is well adapted to infecting the natural hosts and able to  
409 maintain a high level of cell integrity including intact VlsE molecules on the cell surfaces  
410 throughout the infection cycle with a reservoir host. Indeed, *B. burgdorferi* expresses cell surface  
411 proteins binding specifically to proteins of the host complement system to down-regulate innate  
412 and adaptive host immunity (Kraiczy *et al.*, 2006; Samuels, 2011; Hallström *et al.*, 2013;  
413 Hammerschmidt *et al.*, 2014). On an intact spirochete cell surface, the VlsE molecules can  
414 further shield other surface antigens from being recognized by antibodies (Lone and Bankhead,  
415 2020).

416 In non-natural hosts such as humans and rabbits to which *B. burgdorferi* is poorly  
417 adapted, however, the pathogen may lose or diminish its ability to inhibit host immune responses  
418 and is thus more easily recognized by the host immune system. Upon cell disintegration and  
419 degradative processing of the surface antigens including VlsE by the major histocompatibility  
420 complex (HMC), the IR6 would be exposed along with other epitopes and elicit strong antibody  
421 responses. Since the IRs are conserved among the *vls* alleles and, unlike the VRs, their total  
422 amount remains stable during antigenic shift during infection, the IRs would result in stronger  
423 and more long-lasting host responses and become immunodominant in non-natural hosts  
424 including humans.

425 Future work: needs for *in vitro* and *in vivo* validation

426 *B. burgdorferi* infection is characterized by a low number of colonizing spirochetes. It is  
427 difficult to directly detect the pathogen through culture or PCR approaches due to the extreme  
428 scarcity of the organism in infected hosts (Marques, 2015). Current diagnostic assays of Lyme

429 disease, targeting the anti-VlsE or anti-C6 antibodies, do not distinguish between active and past  
430 infections (Wormser *et al.*, 2013; Pegalajar-Jurado *et al.*, 2018). Using recombinant monoclonal  
431 antibodies to directly detect the presence of spirochetes is a solution to the problem of low  
432 spirochete counts in human patients.

433 While we were able to obtain four IR4- and IR6-specific rMAbs, we recognize that the  
434 current study – which demonstrated their binding specificities to immobilized synthetic IR  
435 peptides using ELISA – is a key but only the first step. To validate the utility of these rMAbs as  
436 diagnostic and theragnostic agents, it is necessary to perform *in vitro* testing using cultured *B.*  
437 *burgdorferi* cells followed by *in vivo* testing using a mouse model of Lyme disease. We  
438 anticipate a number of biological and technical challenges during *in vitro* and *in vivo* validation  
439 testing of the rMAbs. For *in vitro* testing using cultured spirochetes, first, it is unclear if the  
440 rMAbs would bind VlsE anchored on the surface of live *B. burgdorferi* cells because of limited  
441 surface accessibility of the IRs at native conformations, even though the rMAbs reacted strongly  
442 with VlsE molecules fixed on a ELISA plate (Fig 2) (Liang and Philipp, 2000; McDowell *et al.*,  
443 2002). Molecular conformation is expected to differ among the synthetic peptides, the IRs in  
444 human sera, and the IRs on intact VlsE molecules anchored to the outer membrane of spirochete  
445 cells. Second, *B. burgdorferi* does not constitutively express a large quantity of VlsE during *in*  
446 *vitro* culture and supplementing the standard media with human tissue cells may be necessary to  
447 increase VlsE expression for *in vitro* validation of rMAb binding (Hudson *et al.*, 2001). Third,  
448 both the IR4 and IR6 sequences vary slightly among *B. burgdorferi* strains despite high sequence  
449 conservation (Fig 3). The affinity of these rMAbs, which were raised using a single allelic  
450 variant (the B31 VlsE), is expected to vary among the *B. burgdorferi* strains. Effects of sequence  
451 variability to rMAb affinity could be quantified with ELISA using synthetic peptides

452 representing the IR variants. Ideally, amino acid residues essential for the rMAb binding could  
453 be accurately pin-pointed with systematic epitope mapping (Chandra *et al.*, 2011).

454 For *in vivo* testing of the rMAbs binding to the spirochetes, it is necessary to first to  
455 develop a live-imaging technology based on labeled rMAbs. Second, it is necessary to establish a  
456 murine model of Lyme disease in which laboratory mice (*Mus musculus*) are inoculated by  
457 needle or by infected ticks (Ivanova *et al.*, 2009; Arumugam *et al.*, 2019). For example, we plan  
458 to label the IR-species rMAbs with a radioactive isotope such as zirconium-89 and perform a  
459 positron emission tomography (PET) for the sensitive detection of trace quantities of spirochetes  
460 in experimentally infected mice.

## 461 Author contributions

462 Li Li performed evolutionary analysis, protein purification, and ELISA. Li Li composed  
463 the initial draft. Lia Di developed the bioinformatics pipeline and web tool and participated in  
464 protein purification and ELISA. Saymon Akther participated in protein purification and ELISA  
465 and edited the manuscript. Brian Zeglis and Weigang Qiu conceived of, obtained funding for,  
466 and supervised the project. Weigang Qiu and Brian Zeglis revised the manuscript.

## 467 Acknowledgments

468 This work was supported by the Public Health Service awards AI139782 (WGQ) from  
469 the National Institute of Allergy and Infectious Diseases (NIAID) and EB030275 (BZ and WGQ)  
470 from the National Institute of Bio-medical Imaging and Bioengineering (NIBIB) of the US  
471 National Institutes of Health (NIH). The content of this manuscript is solely the responsibility of  
472 the authors and does not necessarily represent the official views of NIH. Li Li and Saymon  
473 Akther are supported in part by the Doctoral Program in Biology of the Graduate Center, the

474 City University of New York. We thank Dr. Christophe Sexton and Dr. Jeannine Petersen of the  
475 Center for Disease Control and Prevention (CDC) and Dr. Maria Gomes-Solecki (University of  
476 Tennessee Health Science Center) for providing sera samples from human patients and reservoir  
477 mice, respectively.

478 **Reference cited**

479 Arumugam, S. *et al.* (2019) 'A Multiplexed Serologic Test for Diagnosis of Lyme Disease for  
480 Point-of-Care Use', *Journal of Clinical Microbiology*, 57(12). doi:10.1128/JCM.01142-19.

481 Bacon, R.M. *et al.* (2003) 'Serodiagnosis of Lyme disease by kinetic enzyme-linked  
482 immunosorbent assay using recombinant VlsE1 or peptide antigens of *Borrelia burgdorferi*  
483 compared with 2-tiered testing using whole-cell lysates', *J. Infect. Dis.*, 187(8), pp. 1187–1199.  
484 doi:10.1086/374395.

485 Barbour, A.G. *et al.* (2009) 'Niche partitioning of *Borrelia burgdorferi* and *Borrelia miyamotoi*  
486 in the same tick vector and mammalian reservoir species', *The American Journal of Tropical  
487 Medicine and Hygiene*, 81(6), pp. 1120–1131. doi:10.4269/ajtmh.2009.09-0208.

488 Barbour, A.G. and Qiu, W. (2019) 'Borrelia', in *Bergey's Manual of Systematics of Archaea  
489 and Bacteria*. John Wiley & Sons, Inc., in association with Bergey's Manual Trust, pp. 1–22.  
490 doi:10.1002/9781118960608.gbm01525.

491 Barbour, A.G. and Travinsky, B. (2010) 'Evolution and distribution of the *ospC* gene, a  
492 transferable serotype determinant of *Borrelia burgdorferi*', *mBio*, 1(4), pp. e00153-10.  
493 doi:10.1128/mBio.00153-10.

494 Bhatia, B. *et al.* (2018) 'Infection history of the blood-meal host dictates pathogenic potential of  
495 the Lyme disease spirochete within the feeding tick vector', *PLOS Pathogens*, 14(4), p.  
496 e1006959. doi:10.1371/journal.ppat.1006959.

497 Bockenstedt, L.K. *et al.* (1997) 'Borrelia burgdorferi strain-specific OspC-mediated immunity in  
498 mice', *Infection and Immunity*, 65(11), pp. 4661–4667.

499 Branda, J.A. *et al.* (2017) 'Evaluation of Modified 2-Tiered Serodiagnostic Testing Algorithms  
500 for Early Lyme Disease', *Clinical Infectious Diseases: An Official Publication of the Infectious  
501 Diseases Society of America*, 64(8), p. 1074. doi:10.1093/CID/CIX043.

502 Brisson, D. and Dykhuizen, D.E. (2004) 'ospC diversity in *Borrelia burgdorferi* different hosts  
503 are different niches', *Genetics*, 168(2), pp. 713–722. doi:10.1534/genetics.104.028738.

504 CDC (1995) 'Recommendations for test performance and interpretation from the Second  
505 National Conference on Serologic Diagnosis of Lyme Disease', *Morbidity and Mortality Weekly  
506 Report Surveillance Summaries*, 44(31), pp. 590–591.

507 Chaconas, G., Castellanos, M. and Verhey, T.B. (2020) 'Changing of the guard: How the Lyme  
508 disease spirochete subverts the host immune response', *The Journal of Biological Chemistry*,  
509 295(2), pp. 301–313. doi:10.1074/jbc.REV119.008583.

510 Chandra, A. *et al.* (2011) 'Epitope mapping of antibodies to VlsE protein of *Borrelia burgdorferi*  
511 in post-Lyme disease syndrome', *Clinical Immunology*, 141(1), pp. 103–110.  
512 doi:10.1016/j.clim.2011.06.005.

513 Coutte, L. *et al.* (2009) 'Detailed analysis of sequence changes occurring during *vlsE* antigenic  
514 variation in the mouse model of *Borrelia burgdorferi* infection', *PLoS pathogens*, 5(2), p.  
515 e1000293. doi:10.1371/journal.ppat.1000293.

516 Crooks, G.E. *et al.* (2004) 'WebLogo: a sequence logo generator', *Genome research*, 14(6), pp.  
517 1188–1190. doi:10.1101/gr.849004.

518 Di, L. *et al.* (2021) 'Maximum antigen diversification in a lyme bacterial population and  
519 evolutionary strategies to overcome pathogen diversity', *ISME Journal*, (August), pp. 1–18.  
520 doi:10.1038/s41396-021-01089-4.

521 Eddy, S.R. (2011) 'Accelerated profile HMM searches 2011', *PLOS Comp. Biol.*, 7(e1002195).

522 Edgar, R.C. (2004) 'MUSCLE: A multiple sequence alignment method with reduced time and  
523 space complexity', *BMC Bioinformatics*, 5(113). doi:10.1186/1471-2105-5-113.

524 Eicken, C. *et al.* (2002) 'Crystal structure of Lyme disease variable surface antigen VlsE of  
525 *Borrelia burgdorferi*', *The Journal of Biological Chemistry*, 277(24), pp. 21691–21696.  
526 doi:10.1074/jbc.M201547200.

527 Elzbieta, J. *et al.* (2016) 'Epitope-Specific Evolution of Human B Cell Responses to *Borrelia*  
528 *burgdorferi* VlsE Protein from Early to Late Stages of Lyme Disease', *The Journal of*  
529 *Immunology*, 196(3), pp. 1036–1043. doi:10.4049/jimmunol.1501861.

530 Embers, M.E. *et al.* (2007) 'Dominant epitopes of the C6 diagnostic peptide of *Borrelia*  
531 *burgdorferi* are largely inaccessible to antibody on the parent VlsE molecule', *Clin. Vaccine*  
532 *Immunol.*, 14(8), pp. 931–936. doi:10.1128/CVI.00075-07.

533 Feng, J. *et al.* (2019) 'Stationary phase persister/biofilm microcolony of *Borrelia burgdorferi*  
534 causes more severe disease in a mouse model of Lyme arthritis: implications for understanding  
535 persistence, Post-treatment Lyme Disease Syndrome (PTLDS), and treatment failure', *Discovery*  
536 *Medicine*, 27(148), pp. 125–138.

537 Glöckner, G. *et al.* (2004) 'Comparative analysis of the *Borrelia garinii* genome', *Nucleic Acids*  
538 *Research*, 32(20), pp. 6038–6046. doi:10.1093/nar/gkh953.

539 Graves, C.J. *et al.* (2013) 'Natural selection promotes antigenic evolvability', *PLoS Pathog*,  
540 9(11), p. e1003766. doi:10.1371/journal.ppat.1003766.

541 Hallström, T. *et al.* (2013) 'CspA from *Borrelia burgdorferi* inhibits the terminal complement  
542 pathway', *mBio*, 4(4). doi:10.1128/mBio.00481-13.

543 Hammerschmidt, C. *et al.* (2014) 'Versatile Roles of CspA Orthologs in Complement  
544 Inactivation of Serum-Resistant Lyme Disease Spirochetes', *Infection and immunity*, 82(1), pp.  
545 380–392. doi:10.1128/IAI.01094-13.

546 Haven, J. *et al.* (2011) 'Pervasive recombination and sympatric genome diversification driven by  
547 frequency-dependent selection in *Borrelia burgdorferi*, the Lyme disease bacterium', *Genetics*,  
548 189(3), pp. 951–966. doi:10.1534/genetics.111.130773.

549 Hernández, Y. *et al.* (2018) 'BpWrapper: BioPerl-based sequence and tree utilities for rapid  
550 prototyping of bioinformatics pipelines', *BMC Bioinformatics* [Preprint]. doi:10.1186/s12859-  
551 018-2074-9.

552 Hudson, C.R. *et al.* (2001) 'Increased expression of *Borrelia burgdorferi* vlsE in response to  
553 human endothelial cell membranes', *Molecular Microbiology*, 41(1), pp. 229–239.  
554 doi:10.1046/j.1365-2958.2001.02511.x.

555 Ivanova, L. *et al.* (2009) 'Comprehensive seroprofiling of sixteen *B. burgdorferi* OspC:  
556 implications for Lyme disease diagnostics design', *Clinical Immunology (Orlando, Fla.)*, 132(3),  
557 pp. 393–400. doi:10.1016/j.clim.2009.05.017.

558 Kilpatrick, A.M. *et al.* (2017) 'Lyme disease ecology in a changing world: Consensus,  
559 uncertainty and critical gaps for improving control', *Philosophical Transactions of The Royal  
560 Society B Biological Sciences*, 372(1722), p. 20160117. doi:10.1098/rstb.2016.0117.

561 Kraiczy, P. *et al.* (2006) 'Binding of human complement regulators FHL-1 and factor H to  
562 CRASP-1 orthologs of *Borrelia burgdorferi*', *Wiener klinische Wochenschrift*, 118(21–22), pp.  
563 669–676. doi:10.1007/s00508-006-0691-1.

564 Kugeler, K.J. *et al.* (2021) 'Estimating the Frequency of Lyme Disease Diagnoses, United States,  
565 2010–2018', *Emerging Infectious Diseases*, 27(2), pp. 616–619. doi:10.3201/eid2702.202731.

566 Kundu, S. and Bansal, M.S. (2018) 'On the impact of uncertain gene tree rooting on duplication-  
567 transfer-loss reconciliation', *BMC bioinformatics*, 19(Suppl 9), p. 290. doi:10.1186/s12859-018-  
568 2269-0.

569 Lawrenz, M.B. *et al.* (1999) 'Human antibody responses to VlsE antigenic variation protein of  
570 *Borrelia burgdorferi*', *Journal of Clinical Microbiology* [Preprint]. doi:10.1128/jcm.37.12.3997-  
571 4004.1999.

572 Liang, F.T. *et al.* (1999) 'An immunodominant conserved region within the variable domain of  
573 VlsE, the variable surface antigen of *Borrelia burgdorferi*', *Journal of Immunology (Baltimore,  
574 Md.: 1950)*, 163(10), pp. 5566–5573.

575 Liang, F.T., Nowling, J.M. and Philipp, M.T. (2000) 'Cryptic and exposed invariable regions of  
576 VlsE, the variable surface antigen of *Borrelia burgdorferi* sl', *Journal of Bacteriology*, 182(12),  
577 pp. 3597–3601. doi:10.1128/JB.182.12.3597-3601.2000.

578 Liang, F.T. and Philipp, M.T. (1999) 'Analysis of antibody response to invariable regions of  
579 VlsE, the variable surface antigen of *Borrelia burgdorferi*', *Infection and Immunity*, 67(12), pp.  
580 6702–6706. doi:10.1128/iai.67.12.6702-6706.1999.

581 Liang, F.T. and Philipp, M.T. (2000) 'Epitope mapping of the immunodominant invariable  
582 region of *Borrelia burgdorferi* VlsE in three host species', *Infection and Immunity*, 68(4), pp.  
583 2349–2352. doi:10.1128/IAI.68.4.2349-2352.2000.

584 Lone, A.G. and Bankhead, T. (2020) 'The *Borrelia burgdorferi* VlsE Lipoprotein Prevents  
585 Antibody Binding to an Arthritis-Related Surface Antigen', *Cell Reports*, 30(11), pp. 3663–  
586 3670.e5. doi:10.1016/j.celrep.2020.02.081.

587 Margos, G. *et al.* (2020) 'Rejection of the name *Borrelia* and all proposed species comb. nov.  
588 placed therein', *International Journal of Systematic and Evolutionary Microbiology*, 70(5), pp.  
589 3577–3581. doi:10.1099/ijsem.0.004149.

590 Marques, A.R. (2015) 'Laboratory Diagnosis of Lyme Disease: Advances and Challenges',  
591 *Infectious Disease Clinics of North America*, 29(2), pp. 295–307. doi:10.1016/j.idc.2015.02.005.

592 McDowell, J.V. *et al.* (2002) 'Evidence that the variable regions of the central domain of VlsE  
593 are antigenic during infection with lyme disease spirochetes', *Infection and immunity*, 70(8), pp.  
594 4196–4203.

595 Molins, C.R. *et al.* (2014) 'Collection and characterization of samples for establishment of a  
596 serum repository for Lyme disease diagnostic test development and evaluation', *Journal of  
597 Clinical Microbiology*, 52(10), pp. 3755–3762. doi:10.1128/JCM.01409-14.

598 Mongodin, E.F. *et al.* (2013) 'Inter- and intra-specific pan-genomes of *Borrelia burgdorferi* sensu  
599 lato: Genome stability and adaptive radiation', *BMC Genomics* [Preprint]. doi:10.1186/1471-  
600 2164-14-693.

601 Moore, A. *et al.* (2016) 'Current Guidelines, Common Clinical Pitfalls, and Future Directions for  
602 Laboratory Diagnosis of Lyme Disease, United States', *Emerging infectious diseases*, 22(7), pp.  
603 1169–1177. doi:10.3201/EID2207.151694.

604 Nguyen, L.T. *et al.* (2015) 'IQ-TREE: A fast and effective stochastic algorithm for estimating  
605 maximum-likelihood phylogenies', *Molecular Biology and Evolution*, 36(1), pp. 268–274.  
606 doi:10.1093/molbev/msu300.

607 Norris, S.J. (2006) 'Antigenic variation with a twist - The *Borrelia* story', *Molecular  
608 Microbiology*, 60(6), pp. 1319–1322. doi:10.1111/j.1365-2958.2006.05204.x.

609 Norris, S.J. (2014) 'vls Antigenic Variation Systems of Lyme Disease Borrelia: Eluding Host  
610 Immunity through both Random, Segmental Gene Conversion and Framework Heterogeneity',  
611 *Microbiology Spectrum*, 2(6). doi:10.1128/microbiolspec.mdna3-0038-2014.

612 Palmer, G.H., Bankhead, T. and Seifert, H.S. (2016) 'Antigenic Variation in Bacterial  
613 Pathogens', *Microbiology Spectrum*, 4(1). doi:10.1128/microbiolspec.VMBF-0005-2015.

614 Papkou, A. *et al.* (2019) 'The genomic basis of Red Queen dynamics during rapid reciprocal  
615 host-pathogen coevolution', *Proceedings of the National Academy of Sciences of the United  
616 States of America*, 116(3), pp. 923–928. doi:10.1073/pnas.1810402116.

617 Pegalajar-Jurado, A. *et al.* (2018) 'Evaluation of modified two-tiered testing algorithms for  
618 Lyme disease laboratory diagnosis using well-characterized serum samples', *Journal of Clinical  
619 Microbiology*, 56(8), pp. e01943-17. doi:10.1128/JCM.01943-17.

620 Pettersen, E.F. *et al.* (2004) 'UCSF Chimera--a visualization system for exploratory research and  
621 analysis', *Journal of Computational Chemistry*, 25(13), pp. 1605–1612. doi:10.1002/jcc.20084.

622 Price, M.N., Dehal, P.S. and Arkin, A.P. (2010) 'FastTree 2--approximately maximum-  
623 likelihood trees for large alignments', *PloS One*, 5(3), p. e9490.  
624 doi:10.1371/journal.pone.0009490.

625 Pritt, B.S. *et al.* (2016) 'Borrelia mayonii sp. nov., a member of the *Borrelia burgdorferi sensu  
626 lato* complex, detected in patients and ticks in the upper midwestern United States', *International  
627 Journal of Systematic and Evolutionary Microbiology*, 66(11), pp. 4878–4880.  
628 doi:10.1099/ijsem.0.001445.

629 Pupko, T. *et al.* (2002) 'Rate4Site: an algorithmic tool for the identification of functional regions  
630 in proteins by surface mapping of evolutionary determinants within their homologues',  
631 *Bioinformatics*, 18(Suppl 1), pp. S71-7. doi:10.1093/bioinformatics/18.suppl\_1.s71.

632 R Core Team (2013) *R: A Language and Environment for Statistical Computing*. R Foundation  
633 for Statistical Computing. Available at: <http://www.r-project.org/> (Accessed: 12 November  
634 2012).

635 Ritz, C. *et al.* (2015) 'Dose-Response Analysis Using R', *PLOS ONE*, 10(12), p. E0146021.

636 Rogers, J. *et al.* (2017) 'Reconciliation feasibility in the presence of gene duplication, loss, and  
637 coalescence with multiple individuals per species', *BMC bioinformatics*, 18(1), p. 292.  
638 doi:10.1186/s12859-017-1701-1.

639 Samuels, D.S. (2011) 'Gene Regulation in *Borrelia burgdorferi*', *Annual Review of Microbiology*,  
640 65(1), pp. 479–499. doi:10.1146/annurev.micro.112408.134040.

641 Schierup, M.H., Mikkelsen, A.M. and Hein, J. (2001) 'Recombination, balancing selection and  
642 phylogenies in MHC and self-incompatibility genes', *Genetics*, 159(4), pp. 1833–1844.

643 Schutzer, S.E. *et al.* (2011) 'Whole-Genome Sequences of Thirteen Isolates of *Borrelia*  
644 *burgdorferi*', *Journal of Bacteriology*, 193(4), pp. 1018–1020. doi:10.1128/JB.01158-10.

645 Schwartz, I. *et al.* (2021) 'Multipartite Genome of Lyme Disease *Borrelia*: Structure, Variation  
646 and Prophages', in *Lyme Disease and Relapsing Fever Spirochetes: Genomics, Molecular*  
647 *Biology, Host Interactions and Disease Pathogenesis*. Caister Academic Press.  
648 doi:10.21775/9781913652616.02.

649 Sharma, B. *et al.* (2015) 'Borrelia burgdorferi, the Causative Agent of Lyme Disease, Forms  
650 Drug-Tolerant Persister Cells', *Antimicrobial Agents and Chemotherapy*, 59(8), pp. 4616–4624.  
651 doi:10.1128/AAC.00864-15.

652 Stanek, G. *et al.* (2012) 'Lyme borreliosis', *Lancet*, 379(9814), pp. 461–473.

653 Tilly, K., Bestor, A. and Rosa, P.A. (2013) 'Lipoprotein succession in *Borrelia burgdorferi*:  
654 similar but distinct roles for OspC and VlsE at different stages of mammalian infection',  
655 *Molecular Microbiology*, 89(2), pp. 216–227. doi:10.1111/mmi.12271.

656 Verhey, T.B., Castellanos, M. and Chaconas, G. (2019) 'Antigenic variation in the Lyme  
657 spirochete: detailed functional assessment of recombinational switching at vlsE in the JD1 strain  
658 of *Borrelia burgdorferi*', *Molecular Microbiology*, 111(3), pp. 750–763. doi:10.1111/mmi.14189.

659 Vink, C., Rudenko, G. and Seifert, H.S. (2012) 'Microbial antigenic variation mediated by  
660 homologous DNA recombination', *FEMS Microbiology Reviews*, 36(5), pp. 917–948.  
661 doi:10.1111/j.1574-6976.2011.00321.x.

662 Wang, I.-N. *et al.* (1999) 'Genetic diversity of *ospC* in a local population of *Borrelia burgdorferi*  
663 *sensu stricto*', *Genetics*, 151(1), pp. 15–30.

664 Wormser, G.P. *et al.* (2013) 'Single-Tier Testing with the C6 Peptide ELISA Kit Compared with  
665 Two-Tier Testing for Lyme Disease', *Diagn Microbiol Infect Dis.*, 75(1), pp. 9–15.  
666 doi:10.1016/j.diagmicrobio.2012.09.003.

667 Yu, G. *et al.* (2017) 'ggtree: an R package for visualization and annotation of phylogenetic trees  
668 with their covariates and other associated data', *Methods in Ecology and Evolution*, 8(1), pp. 28–  
669 36.

670 Zhang, J.R. *et al.* (1997) 'Antigenic variation in Lyme disease borreliae by promiscuous  
671 recombination of VMP-like sequence cassettes', *Cell*, 89(2), pp. 275–285. doi:10.1016/s0092-  
672 8674(00)80206-8.

673

674

675 Tables

676 Table 1. Peptides used to screen for IR-specific monoclonal antibodies

Peptide <sup>a</sup>	Sequence <sup>b</sup>	Length	Variability (z-score) <sup>c/d/7</sup>
IR1	EVSELLDKLVKAVKTAEGASSG	22	-0.1746
IR2	(ASVK)GIAKGIKEIVEAA(GGSE)	21	-0.6066
IR3 <sup>c</sup>	AGKLFVK	7	-0.6585
IR4	KAAGAVSAVSGEQILSAIV(TAA)	22	-0.5393
IR5	(AEEAD)NPIAAAIG(TTNEDA)	19	-0.7378
IR6	MKKDDQIAAAIALRGMAKDGFKFAVK	25	-0.6932

678 <sup>a</sup> IRs: Invariant regions. Six IRs were identified from an alignment of VlsE proteins from three  
679 strains representing two species (Liang *et al.*, 1999).

680 <sup>b</sup> IR sequences were based on the VlsE protein of the B31 strain (GenBank accession U76405).  
681 Flanking residues (in parentheses) were padded to the IR2 and IR4 to facilitate ELISA. For IR5,  
682 the padded residues were the most common residues flanking this conserved region.

683 <sup>c</sup> Antigenicity of IR3, the shortest IR, was not tested in the present study.

684 <sup>d</sup> Standard deviation from the mean amino-acid substitution rate of zero (see Materials and  
685 Methods)

686 Table 2. Specificity and sequences of IR-specific mAbs

MAb	Specificity	EC50 (ng/mL) <sup>a</sup>	Variable heavy chain (V <sub>H</sub> ) sequence <sup>b</sup>	Variable light chain (V <sub>L</sub> ) sequence
1D11-4	IR6	59.07 (1.75)	MGWSCIILFLVATATGVHSQSLVESGGGLVQP EGSLTLLTCKASGFSFSSGGYDMCWVRQAPGKGL EYIACIDAGDDITHYASWVKGRFTVSKTSSTTV TLQLNSLTVADTATYFCGRFWDLWGPGLTVSS (131 aa)	MGWSCIILFLVATATGVHSSVLQTTPSPVSAAV GGTVTINCQSSQSVYDSTWLWYQQKPGQPPK LLIYKASNLAvgVPSRFKGSQSGTHFTLTISDLE CDDAATYYCVGGYSGSVDNWAFGGGTEVVVK (130 aa)
15E2-1	IR4	125.17 (5.10)	METGLRWLLLVAVLKGVQCQSLEESGGGLFK PTDTLTLTCTVSGIDLSSYAMIWVRQAPGKGL WIGYIWSSGRIWYASWAKGRFTISRTSTTVDL KLASPTTEDTATYFCARLWDIWPGLTVSS (128 aa)	MDTRAPTQLLGLLLWLPGATFAQVLTQTPSSV SAAVGGTVTINCQASQSVSNNNVLAWFQQKPG QPPKLLIFGASNLESGVSSRFRGSGSGTQFTLTIS GVQCDDAATYYCLGEFSCSSADCLAFGGGTEV VVK (135 aa)
17A8-1	IR4	86.14 (2.44)	METGLRWLLLVAVLKGVQCQSVEESGGLVLT PGTPLTLTCTVSGFPLSSYSMAWVRQAPGKGL EYIGFINTDGSAYYASWAKGRITISKTTVEL KITSPTTEDTATYFCGTGNIWGPGLTVSS (127 aa)	MDTRAPTQLLGLLLWLPGATFAQVLTQTPSSV SAAVGGTVTINCQASQSVSNNNVLAWFQQKPG QPPKRLIYSALTLDGVPSRFKGSQSGTHFTLTIS GVQCDDAATYYCAGGYDCSSNDIAFGGGTEV VVK (135 aa)
28D3-1	IR4	245.91 (7.93)	METGLRWLLLVAVLKGVQCQSVEESGGLVLT PGTPLTLTCTVSGFSLSSYSMGWVRQAPGKGL EYIGMIISNNSTYYASWAKGRITISKTTVELK ITSPTTEDTATYFCGTGNIWGPGLTVSS (127 aa)	MDTRAPTQLLGLLLWLPGATFAQVLTQTPAS VSAAVGGTVTINCQASQSTSNNNALAWFQQKPG QPPKRLIYSALTLDGVPSRFKGSQSGTHFTLTIS SGVQCDDAATYYCAGGYDCSSNDICITFGGGTE VVVK (135 aa)

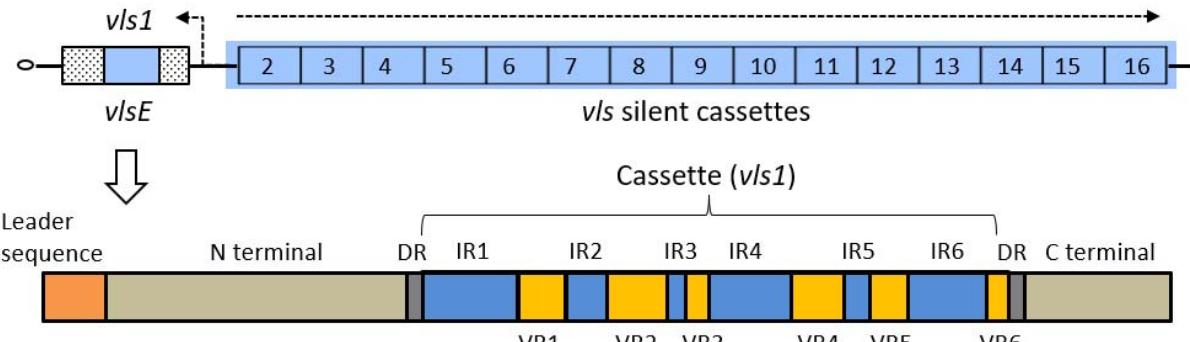
687 <sup>a</sup> EC50: Effective concentration of 50% response level (and standard error) based on titration by ELISA. A lower EC50 indicates  
688 effective binding at a lower concentration (Fig 7).

689 <sup>b</sup> Peptide sequences from the topmost (or 2<sup>nd</sup> topmost) abundant sequences identified in the sorted single B cells producing IR-specific  
690 antibodies.

691

692 Figures

693 **Fig 1.** Genomic and gene structures of the *vls* locus in *B. burgdorferi* strain B31  
694

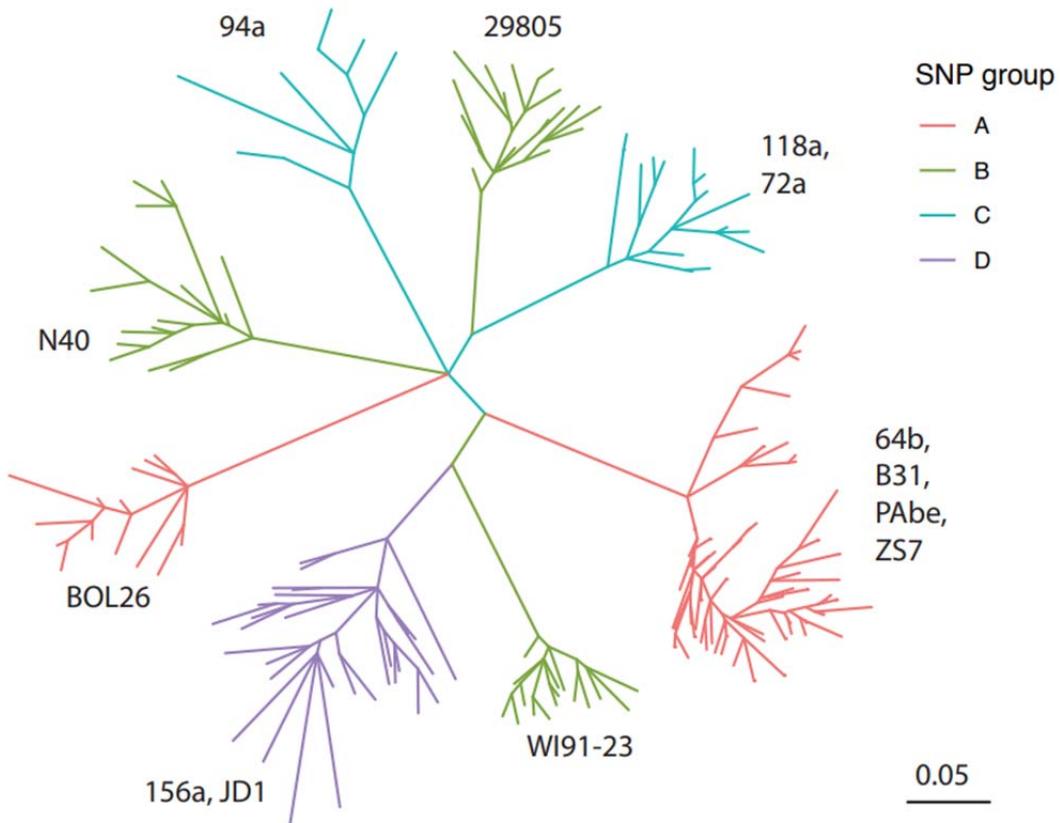


695  
696  
697 (A) The *vls* locus is located close to the telomere of the linear plasmid lp28-1 (GenBank  
698 accession AE000794) in the B31 genome, consisting of cassettes of silent (un-expressed)  
699 open reading frames (ORFs) (*vls2* through *vls16*) and an expressed ORF (*vlsE*) containing  
700 the *vls1* cassette introduced by recombination (Zhang *et al.*, 1997). Dashed arrows indicate  
701 the direction of coding strands.

702  
703 (B) The *VlsE* protein consists of a leader peptide, a N-terminus domain, a cassette flanked by  
704 two direct repeats (DRs), and a C-terminus domain. The central cassette consists of  
705 interspersed variable (VR1-6) and invariant regions (IR1-6).

706

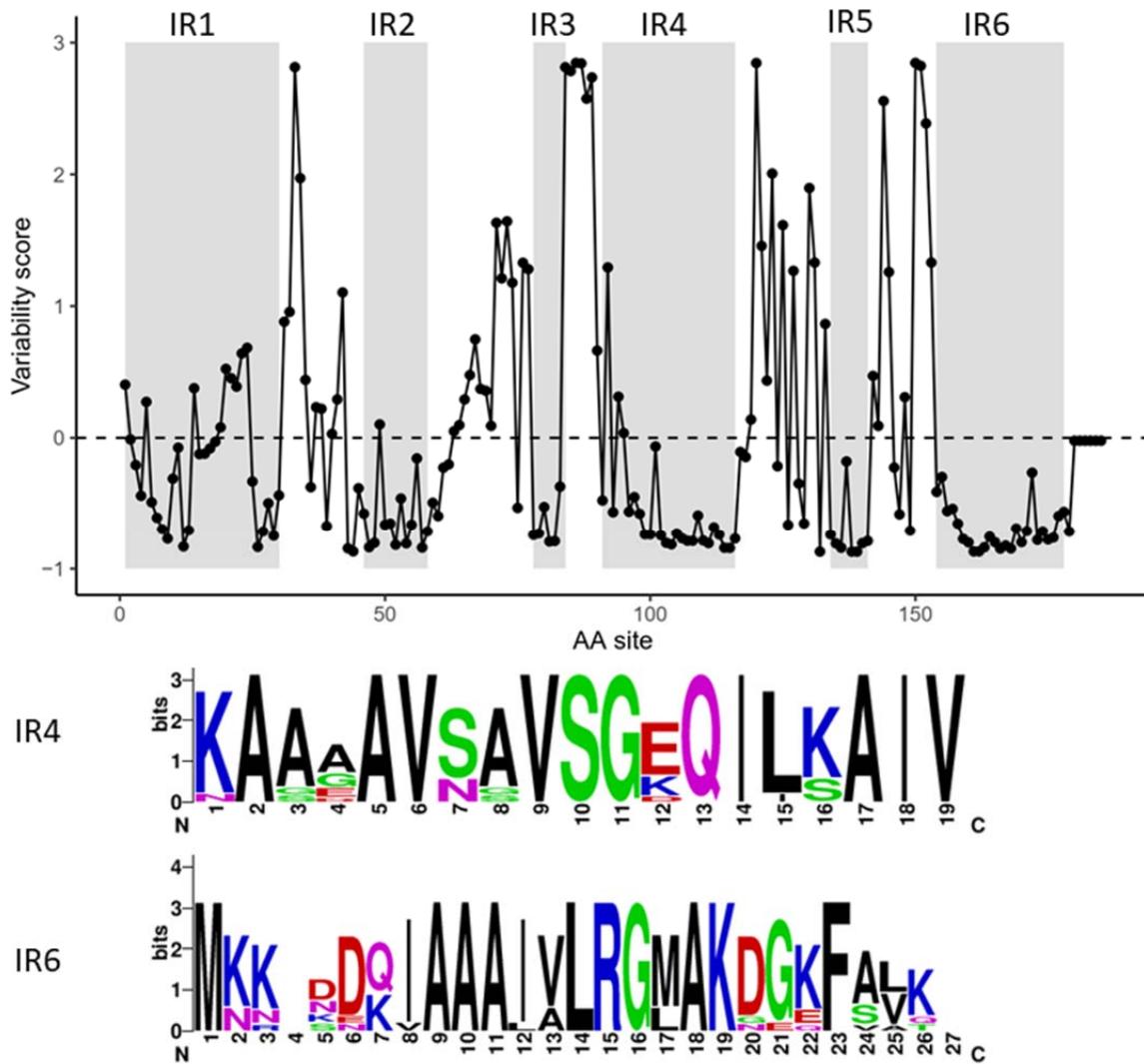
707 **Fig 2.** Sequence diversity of *vls* cassettes of 13 *B. burgdorferi* strains  
708



709  
710 Eight major clades of *vls* alleles were identified based on the codon alignment of 194 cassette  
711 sequences from 12 US *B. burgdorferi* genomes (strain names shown by the clades) (Schutzer *et*  
712 *al.*, 2011). The maximum likelihood tree was inferred using IQ-TREE (version 1.6.1) (Nguyen *et*  
713 *al.*, 2015). All branches were supported by  $\geq 80\%$  bootstrap values. The tree was rendered using  
714 the R package *ggtree* (Version 2.2.4) (Yu *et al.*, 2017). Branches were colored according to the  
715 four phylogenetic groups (A through D) identified based on genome-wide single-nucleotide  
716 polymorphisms (SNPs). SNP groups A, B, C split into multiple clades, indicating rapid *vls*  
717 sequence divergence between closely related strains (Graves *et al.*, 2013). Sequences within the  
718 multi-strain clades (e.g., within D group) did not separate into strain-specific subclades,  
719 suggesting frequent gene duplications and losses.

720

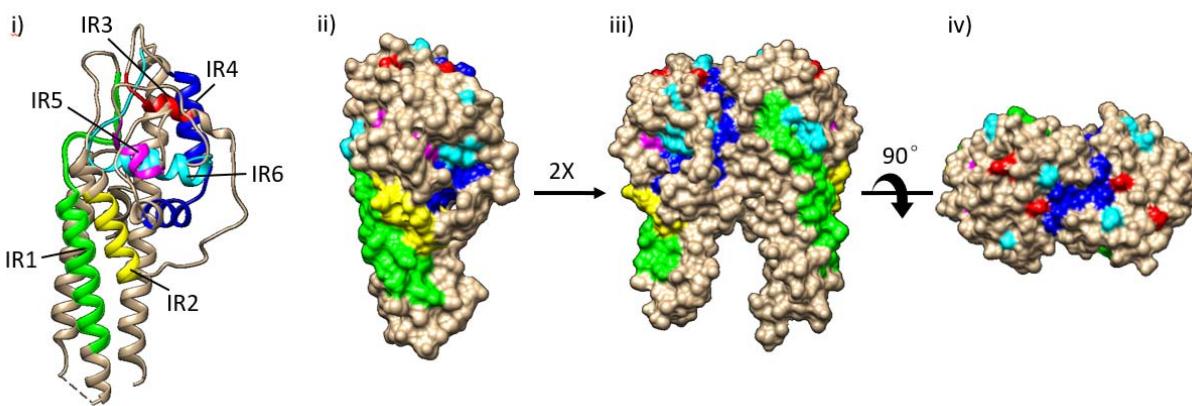
721 **Fig 3.** Site-specific evolutionary rates of *vls* cassettes



724 (Top) Evolutionary rate, denoted as variability score (in the unit of standard deviation, y-axis), at  
725 each amino acid site was estimated by Rate4Site (Version 3.0.0) (Pupko *et al.*, 2002) based on an  
726 alignment of translated sequences of 194 *vls* cassettes and the maximum-likelihood tree (Fig 2).  
727 The dashed line at 0 indicates the average evolutionary rate. The six IRs, showing generally  
728 lower-than-average rates, were shaded in gray. VlsE of B31-5A3 clone (GenBank accession  
729 U76405) was used as the reference for computation and annotation.

730 (Bottom) SeqLogo images of IR4 and IR6 sequences, constructed based on one representative *vls*  
731 allele (translated) from each of the 12 *B. burgdorferi* genomes (Schutzer *et al.*, 2011). Amino  
732 acid residues were colored according to physiochemistry. Letter heights correspond to  
733 information content in bits, a measure of site conservation (Crooks *et al.*, 2004)

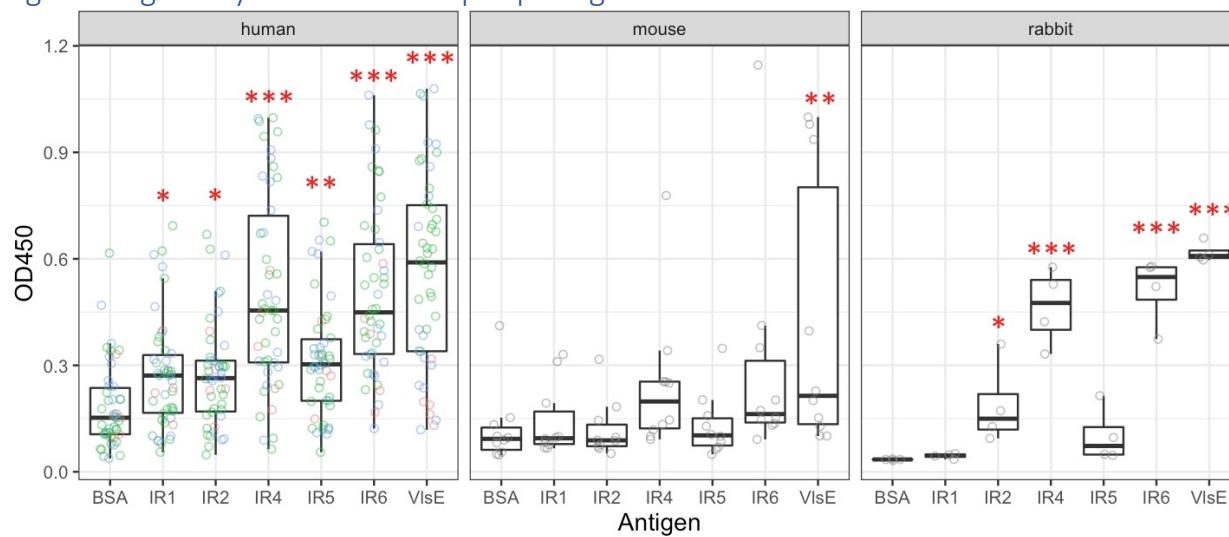
734 **Fig 4.** IR-highlighted three-dimensional structures of VlsE.



735  
736 Structure diagrams of VlsE protein from B31 were prepared in Chimera (Version 1.15)  
737 (Pettersen *et al.*, 2004) based on the PDB file (accession 1L8W) (Eicken *et al.*, 2002). IRs were  
738 highlighted in different colors (IR1 in green, IR2 in yellow, IR3 in red, IR4 in dark blue, IR5 in  
739 magenta, and IR6 in cyan). (i) Ribbon diagram showing that IRs tend to form alpha helices. (ii)  
740 Surface-filled diagram showing membrane surface exposure of IRs in monomeric form. (iii) and  
741 (iv) Dimerized structure models. The structures were oriented to show the membrane proximal  
742 part at the bottom (*i*, *ii*, and *iii*).

743

744 Fig 5. Antigenicity of conserved epitopes against host sera



745 Lyme stage: ○ Control ○ Early ○ Late ○ NA

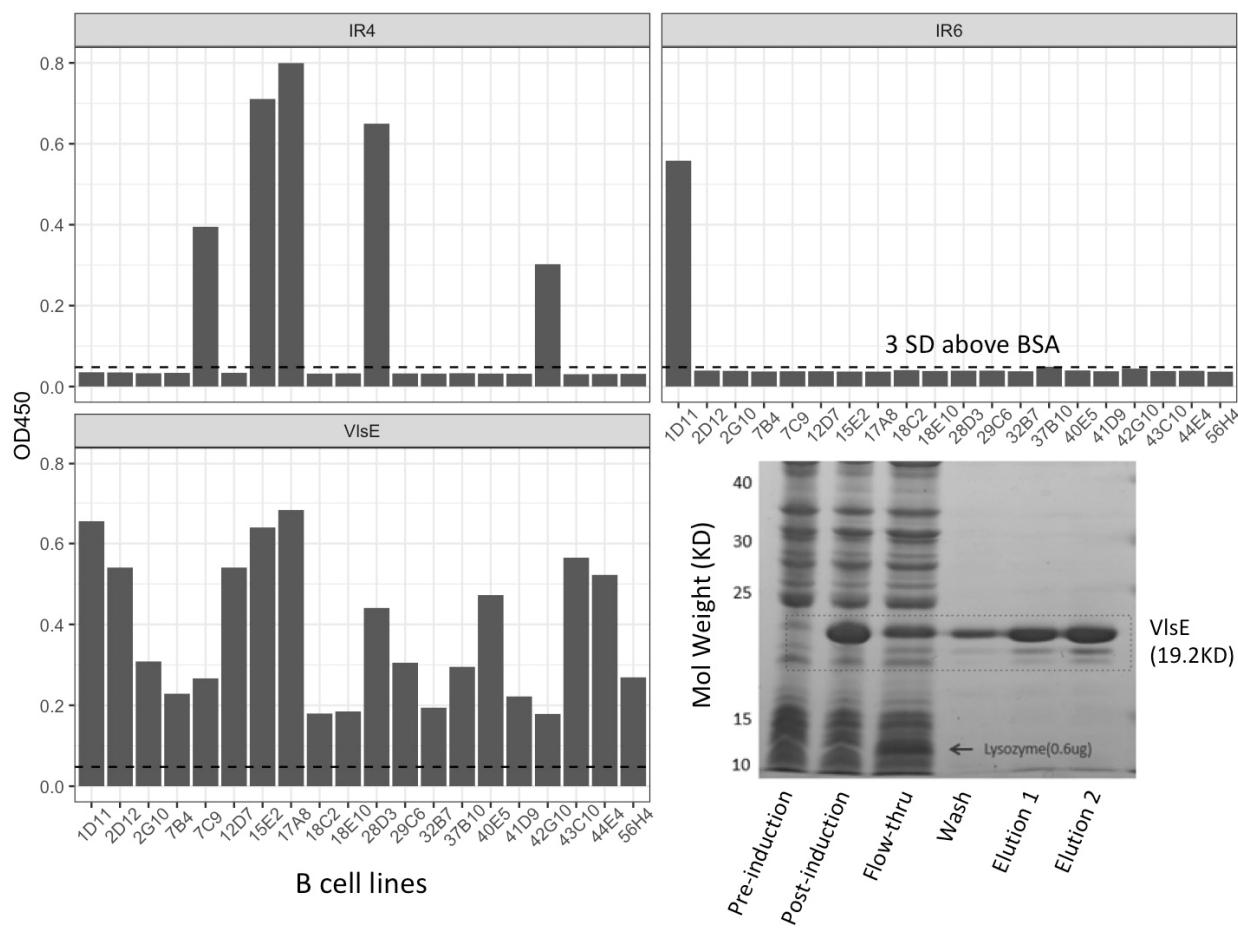
746

747 Antigenicity of five IRs (x-axis) was quantified with ELISA (see Material and Methods). Each  
748 IR peptide was tested for reactivity (OD450, y-axis) with host sera (represented by dots) from 46  
749 human patients (left), 10 natural hosts (white-footed mouse, *Peromyscus leucopus*) (middle), and  
750 4 New Zealand rabbits (*Oryctolagus cuniculus*) (right). Bovine serum albumin (BSA) was used  
751 as the negative control and the purified recombinant VlsE protein (of strain B31) as the positive  
752 control. Asterisks indicate significant differences between antigens and BSA by ANOVA  
753 analysis at varying degrees of confidence: “\*” for  $0.01 < p < 0.05$ , “\*\*” for  $0.001 < p < 0.01$ , and  
754 “\*\*\*” for  $p < 0.001$ .

755

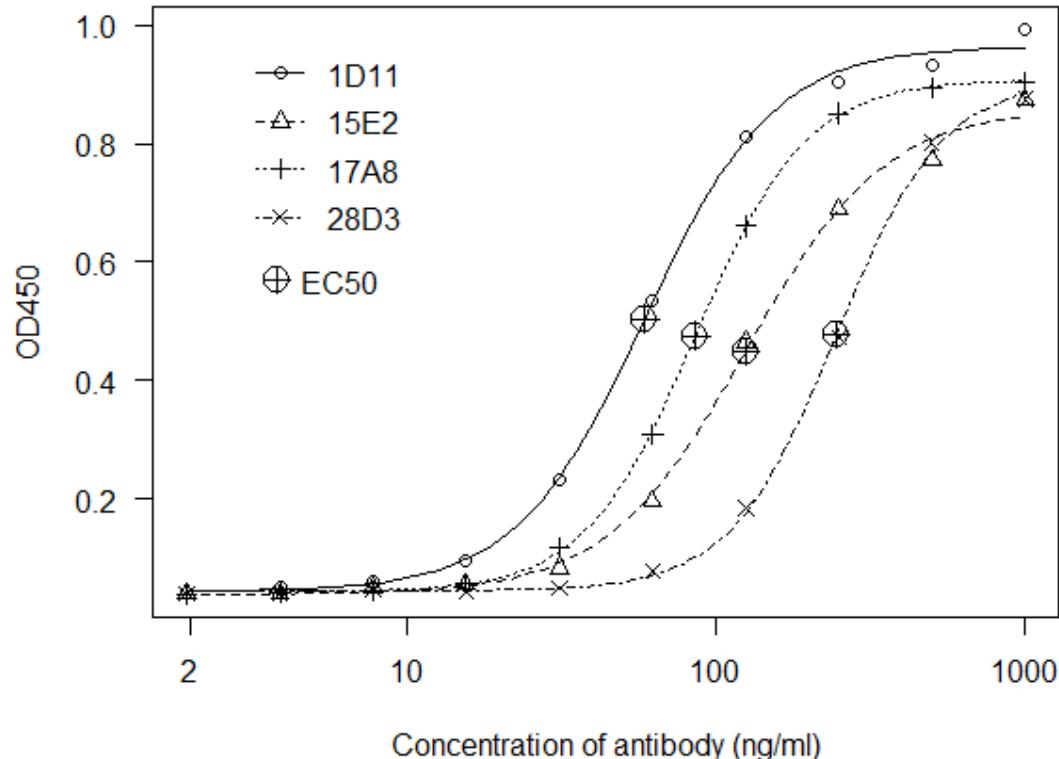
756

757 Fig 6. Identification of B cells producing IR-specific antibodies



758  
759  
760 (Bar plots) Twenty VlsE-positive cell lines (x-axis) were selected with B cell sorting technique  
761 and tested against purified antigens with ELISA. Bovine serum albumin (BSA) was used as the  
762 negative control and the purified recombinant VlsE protein (of strain B31) as the positive control.  
763 An OD450 value (y-axis) greater than 3 standard deviations above the mean BSA reactivity  
764 (dashed lines) was considered to show significant antibody-antigen reactivity. Five cell lines  
765 expressing anti-IR4 antibodies and one cell line expressing anti-IR6 antibodies were identified.  
766 (Image) SDS-PAGE image of induction and purification of the recombinant B31 VlsE. Elution 1  
767 and Elution 2 were combined into a single preparation with an estimated purity of ~65% and a  
768 concentration of ~5.0 mg/ml, which was subsequently used to immunize New Zealand rabbits.  
769

770 Fig 7. Binding characteristics of IR-specific monoclonal antibodies



771  
772

773 Serially diluted preparations of the four affinity-purified rMAbs were tested with ELISA against  
774 their respective IRs (1D11 against IR6; 15E2, 17A8 and 28D3 against IR4). The R package *drc*  
775 (Version 3.0-1) (Ritz *et al.*, 2015) was used to estimate the effective concentration and to plot the  
776 titration curves. EC50 (effective concentration at 50% of the maximum activity) values were  
777 estimated from the fitted curves, with a lower EC50 indicating stronger antigen affinity.

778  
779  
780