SELECTIVE FORCES THAT SHAPE THE VLS ANTIGENIC VARIATION SYSTEM IN BORRELIA BURGDORFERI

Wei Zhou

A DISSERTATION

in

Biology

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2016

Supervisor of Dissertation

_____

Dustin Brisson, Associate Professor of Biology, University of Pennsylvania

Graduate Group Chairperson

_____

Michael Lampson, Associate Professor of Biology, University of Pennsylvania

Dissertation Committee:

Dustin Brisson, Associate Professor of Biology, University of Pennsylvania

Erol Akçay, Assistant Professor of Biology, University of Pennsylvania

Paul Sniegowski, Professor of Biology, University of Pennsylvania

Rahul Kohli, Assistant Professor of Medicine, University of Pennsylvania

Tim Linksvayer, Assistant Professor of Biology, University of Pennsylvania

SELECTIVE FORCES THAT SHAPE THE VLS ANTIGENIC VARIATION SYSTEM IN BORRELIA BURGDORFERI

ACKNOWLEDGMENT

This dissertation would not have been possible without my advisor, Dustin Brisson. Dustin is a very open-minded and creative scientist. Dustin has given me much freedom to develop my research project and has encouraged me to think independently. I am very grateful for his advices and supports.

I would like to thank the present and former members of my thesis committee - Erol Akçay, Junhyong Kim, Paul Sniegowski, Rahul Kohli and Tim Linksvayor for their constructive and inspiring suggestions that have hugely impacted the thesis and enriched my knowledge as well as ways of thinking.

I am very grateful to the present and former members of the EEDS lab - Alexander Berry, Camilo E. Khatchikian, Christopher J. Graves, Damien Lekkas, Erica Foley, Godfrey Devevey, Jill Devine, Ozlem Onder, Stephanie Seifert and Tam Tran. All of them have contributed to my thesis project, intellectually or experimentally. More importantly, they are friendly, energetic and have made the lab a great place to work and stay in.

I want to thank my cohorts and people in the biology department for organizing a variety of gatherings and events, making my 5-year graduate school experience very enjoyable and memorable.

I would like to thank my parents who have been very patient and caring in the last 5 years. They congratulated me for however small progress I have made, encouraged me for whatever obstacles I have met and have been supportive for decisions I have made.

ABSTRACT

SELECTIVE FORCES THAT SHAPE THE VLS ANTIGENIC VARIATION SYSTEM IN BORRELIA BURGDORFERI

Wei Zhou

Dustin Brisson

Evolutionary success of microbial pathogens requires survival within hosts, despite the rapidly changing and lethal immune response. Pathogens such as the Lyme disease bacteria *Borrelia burgdorferi* have evolved antigenic variation systems that are necessary for survival within the adverse immune environment. Although antigenic variation systems are essential to both microbial pathogenesis and microbial evolution, it is largely unclear what selective forces have influenced the evolution of antigenic variation systems. In this thesis, we investigate evolution of the *vls* antigenic variation system in *B. burgdorferi* by asking two major questions: First, what traits relevant to the *vls* antigenic variation system have natural selection acted on? Second, how did the selective forces shape the genetic sequences of the *vls* antigenic variation systems? We characterize sources of natural selection using mathematical modeling, computational simulation and mutagenesis experiments. Our findings show that natural selection has promoted diversity among VlsE variants on both sequence and structure by organizing the variable sites in the *vls* unexpressed cassettes. We also show that the level of diversity among the VlsE variants may strongly influence the within-host dynamics of Bb, an important fitness component of *B. burgdorferi*. Finally, our results indicate that diversity among

VlsE variants might be constrained by purifying or stabilizing selections on translational efficiency and structural stability of the VlsE variants.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

Chapter 1. Introduction

1.1. Background: Rapidly changing host immune environment

All reproducing organisms are faced with changing environments and have to cope with the potential decrease in fitness caused by environmental changes. Although environmental changes with minimal fitness impact may be simply tolerated, evolutionary success often requires changes in morphology, physiology or behavioral traits to maintain high fitness in the face of large environmental changes [1-3]. The ability to respond to environmental changes is especially important for the evolutionary success of plants and microbes because they have limited mobility and are less capable of physically escaping from changed environments [4]. A fundamental question in microbial evolution is how microbes manage to survive within rapidly changing environments.

The host immune system is a particularly adverse and rapidly fluctuating environment for microbial parasites. The host's immune response to an invading parasite consists of a variety of components that are spatially and temporally variable. One typical mechanism of the response is the antibody-mediated adaptive immune response. Clones of antibody-producing B cells are selected from a pool of up to 10^9 clones to maximize affinity of antibodies to the invading parasite. Selected clones of B cells then secrete antibodies that bind to parasitic antigens with a high affinity. This process happens repeatedly to keep the antibody population highly reactive to newly expressed parasitic antigens. Moreover, changes in B cell populations can happen very rapidly: in young rats, newly proliferated B cells can replace the peripheral B cell pool within days [5]. The high-affinity antibodies

bind to parasitic antigens, recruiting immune effector cells and immune effector molecules to efficiently remove the antibody-bound parasite cells. Therefore, antibody response creates a rapidly changing and adverse environment for microbial parasites.

1.2. Background: Antigenic variation systems in microbial parasites

The parasites need to persist within hosts despite the lethal antibody response so as to elevate within-host density of the parasites and to increase the probability of transmission to naïve hosts [6, 7], both of which are vital fitness components. The ability of microbial parasites to introduce antigenic diversity is essential to surviving the antibody response. The reasons are twofold: First, due to specificity of the antibody response, diversity among the parasitic antigens is needed to facilitate the evasion of antibody recognition and killing. It has been suggested that the probability that a subset of the parasite population can evade the antibody response is a function of the diversity of antigens within the host [8]. Second, the parasite population continues to lose standing variation due to a dynamic selection imposed by the antibody population. That is, only a subset of parasitic antigens can evade each antibody, and if new variation is not generated in time, that subset will also become extinct when the antibody population evolves to target the subset.

As an adaptation to the immune environment, a lot of microbial parasites have evolved mechanisms to promote diversity at a rate higher than expected given the genome-wide mutation rate. These mechanisms include hypermutation at contingency loci, phase variation and antigenic variation [9, 10]. Contingency loci are regions of highly mutable

DNA which often encode proteins located at the interface between the microbe and the environment [7]. These loci usually mutate in a stochastic manner (ie. without sensing the environment) to rapidly and continuously generate genetic variations, and thus arguably facilitate adaptation to changing environments by increasing evolvability [11-13]. Phase variation describes a reversible and often random switch between gene expression profiles which can occur at a rate as fast as 1 switch per 10 replications [10]. Microbial parasites switch between different phases during different stages of pathogenesis due to significant environmental changes [14-16]. Although similar to normal gene regulation, phase variation is distinctive in that it involves some forms of heritability, either genetic or epigenetic, of the altered phenotype [10].

Antigenic variation refers to the alternation of a particular surface antigen through DNA rearrangement [10, 17]. In terms of molecular mechanisms, such rearrangement is commonly realized through homologous recombination between an expression locus and a library of unexpressed, silent paralogs. As an example, *Borrelia burgdorferi* (Bb), the causative agent of Lyme disease, takes advantage of a cassette-based mechanism to evade the host immune response against its surface antigen VlsE (Figure 1-1) [18, 19]. VlsE is a 35kDa surface lipoprotein that is antigenically dominant when the bacteria infect vertebrate hosts. Each *Bb* genome has an array of 10-20 unexpressed cassettes that are paralogous to *vlsE* and are located within the linear plasmid lp28-1. DNA sequences of the *vls* cassettes are very similar to the cassette region of *vlsE*, but also have mutations that are concentrated in six variable regions. Random segments of the silent cassettes are continuously introduced into *vlsE* through homologous and nonreciprocal recombination,

altering the sequence and the structure of the VlsE antigen [18-20]. It has been shown that without the *vls* antigenic variation system, the infection duration and the ability of *Bb* to disseminate to different tissues of laboratory mice are dramatically reduced [21-24].

Microbial pathogens benefit from antigenic variation in several ways. First, antibodies that specifically target the original form of an antigen may have a decreased affinities to the protein variants [8, 9, 17, 25]. Second, antigenic variation increases polymorphism in the transmittable pool of pathogens, which increases the success rate in reinfection or transmission [8, 25]. Third, when a host primed by the original form of an antigen is met with a variant of the same antigen, it tends to activate a memory response against the original antigen instead of a primary response against the variant [25]. In this case, the memory response will likely be inefficient because it does not specifically target the antigenic variant. This phenomenon, known as the Hoskins effect, infers that polymorphism in a transmittable pool can not only permit reinfection, but might also render the host more susceptible after a primary infection [25]. Finally, variability of antigens may help counter the variability in immune systems among hosts with different genetic backgrounds [25].

In addition to its importance in microbial infection, antigenic variation systems are a suitable model to study microbial responses to rapidly fluctuating environments. For recombination-based antigenic variation systems such as the *vls* antigenic variation system, the genetic information for all possible antigenic variants is encoded in the sequences of the paralogs [9, 10, 26]. Therefore, the antigenic variants that a parasite can express can be explored by analyzing the sequences of the antigenic paralogs. For the *vls*

antigenic variation system in *Bb*, antigenic variants of *vlsE* continuously sample DNA segments from the unexpressed cassettes. Therefore, selective pressure on the antigenic variants imposed by the immune system is in turn reflected in the sequences of the unexpressed cassettes. Earlier work has shown evidence of strong diversifying selection in the unexpressed cassettes of the *vls* antigenic variation system: the dN/dS ratio among the unexpressed cassettes is exceptionally high, which indicates that the unexpressed cassettes are under strong selection for the probability of amino acid alterations within the expression site [27]. This is likely because diversity among the unexpressed cassettes need to be elevated to ensure that the newly generated VlsE variants are divergent enough from the parental VlsE targeted by the antibodies, so as to evade the recognition of the extant antibodies.

Despite the strong diversifying selection, not all sites are variable even within the variable regions of *vlsE*. The existence of invariant sites suggests that either variation at the invariant sites has little effect on immune recognition, or purifying selection has limited the variability at these sites. Purifying selections may act on protein-coding genes, including antigens of parasites, because changes that disrupt the structure or function of the protein are detrimental. Although the full function of VlsE is yet unclear, it has been suggested that VlsE can protect cell surface proteins from immune responses because of its immunodominance [28], high level of expression within vertebrate hosts [29, 30] and co-localization with other cell surface proteins [21]. Excessive alteration in the sequence of the antigen can potentially undermine its efficiency in expression, correct folding, structural stability and resistance to proteases [31]. Therefore, antigenic variation systems,

including *vls*, should be selected to promote antigenic diversity, but only within constraints of the structural and functional integrity of the antigen [10].

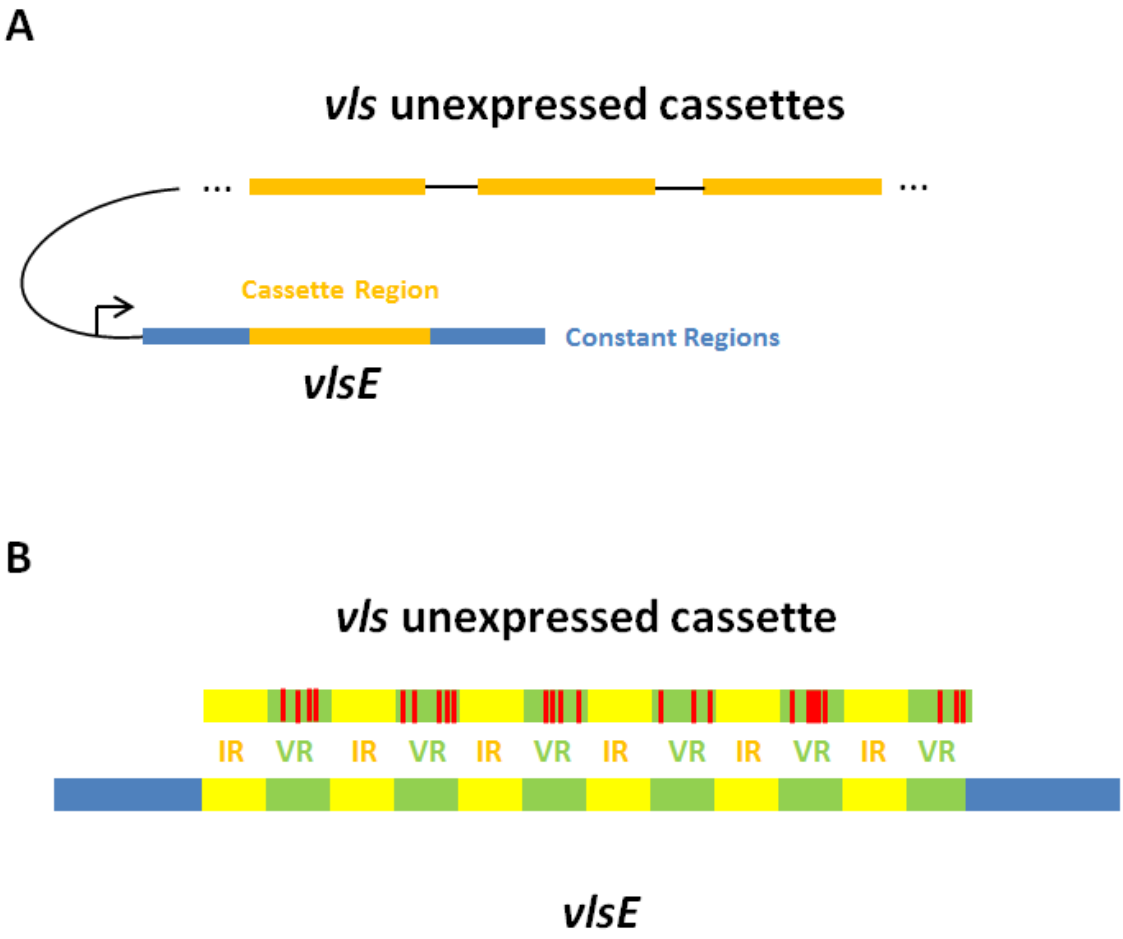


Figure 1-1. *vls* antigenic variation system in *Bb*. (A) the *vlsE* expression site and an array of unexpressed *vls* cassettes are located on the linear plasmid lp28-1 in *Bb*. (B) the cassette region of *vlsE* and the *vls* unexpressed cassettes are composed of variable regions (VR) with a high concentration of mutations and invariable regions (IR) that are highly conserved among the cassettes.

1.3. Scientific questions and dissertation structure

Despite the importance of antigenic variation systems in microbial pathogenesis and microbial evolution, it is largely unknown how natural selection, including diversifying selection and purifying selection, organize the genetic sequences of antigenic variation systems. For example, in the *vls* antigenic variation system, it is unclear what sources of selection have acted on the VlsE variants and how these selective forces have influenced the sequences of the *vls* unexpressed cassettes. This dissertation investigates potential sources of natural selection that have organized variable sites within the *vls* unexpressed cassettes.

Chapter 2 presents a study of potential sources of natural selection affecting the nucleotide identities at, and positions of, the variable sites within the variable regions of the unexpressed *vls* cassettes. The study uses computational modeling and simulation to investigate how perturbations in the variable and invariant nucleotides within the variable regions of the unexpressed cassettes alter diversity and translational efficacy among the resulting VlsE variants. The results suggest that the nucleotide identities and positions of the variable sites are selected for higher sequence diversity among the VlsE variants within strong constraints on protein translational ability.

Chapter 3 examines the influence of diversity among variants of VlsE on Bb dynamics within hosts. Dynamics models are used to describe temporal changes in VlsE variant types, antibody response, immune effector cells and *Bb* abundance within-host. Numerical simulation based on the models predicts that diversity among antigenic variants significantly influence *Bb dynamics* when *Bb* is strongly suppressed by the immune system.

Chapter 4 investigates sources of natural selection acting on structure-level properties of the VlsE antigen. We test the hypothesis that variable amino acid residues in the variable regions of VlsE are more likely targeted by antibodies than the invariable residues in the variable regions of VlsE. A correlation between variability and antigenic importance of amino acid residues within the variable regions of VlsE is supported by computational analysis, but not consistently supported by experimental analysis. We discuss possible mechanisms that may lead to the discrepancy.

Chapter 2. Selection for sequence diversity and translational efficacy among VlsE variants

This chapter was adapted from:

Zhou, W. and D. Brisson. 2014. Potentially conflicting selective forces that shape the vls antigenic variation system in *Borrelia burgdorferi*. Infect Genet Evol. 27:559-65.

2.1. Introduction

The persistence within vertebrate hosts in the face of the potentially lethal environmental conditions imposed by the immune system is a primary constituent of the evolutionary fitness of many microbial pathogens [6, 8, 32, 33]. These strong selective pressure imposed by the immune response has resulted in antigenic variation mechanisms evolved to cope with this rapidly changing environment [6-8, 25, 33].

Antigenic variation systems alter surface antigens of pathogens, giving rise to subpopulations of pathogens with distinct antigenic variants that are not recognized by antibodies targeting previously detected antigens [10]. Antigenic variation systems that more efficiently alter surface antigens of the pathogen are likely to be selectively advantageous as they promote greater residence time within hosts and transmission to naïve hosts, both of which are primary components of pathogen fitness.

One common molecular mechanism of antigenic variation involves recombining gene fragments from unexpressed, paralogous cassettes into an expression site, thereby altering the sequence of the expressed antigen. In these types of recombination-based antigenic

variation systems, which are common in several bacterial genera [18, 34, 35], the ability to alter the sequence of the expressed antigen is correlated with the amount of diversity among the unexpressed cassettes. Thus, natural selection should favor ever greater diversity among unexpressed cassettes to promote ever greater divergence among expressed antigens [27, 36]. However, the extent of the divergence among cassettes can be constrained by other features of the system [37] . Here, we use the well-characterized *vls* antigenic variation system in the Lyme disease bacterium, *Borrelia burgdorferi* (Bb), as a model system to investigate the interactions between selection favoring greater antigenic divergence and other potential constraints on antigenic variation systems.

Bb requires continuous alteration of the highly-expressed VlsE antigen for long-term survival within hosts [18, 22-24, 38, 39]. A fragment of an unexpressed *vls* cassette can be introduced into the *vlsE* expression site through nonreciprocal recombination, thus changing, adding, or removing nucleotides in sequence of the expression site resulting in the expression of a divergent VlsE antigen. However, altering the sequence in the expression site could potentially reduce the ability to translate a functional protein– by introducing stop codons or frameshift mutations – or reduce translational efficiency and accuracy– by introducing non-preferred codons [20, 40]. Little is currently known about how selection on both antigenic divergence and translational ability or efficiency constrains the nucleotide identities at the variable sites, positions of the variable sites and positions of the insertion/deletion mutations.

Here we evaluated the effects of the identity of nucleotides at variable sites, positions of the variable sites, and position of insertion/deletion mutations in the unexpressed

cassettes on the divergence among VlsE variants as well as their translational ability and translational efficiency. We ask if the organization of variable sites and insertion/deletion mutations in the unexpressed cassettes of multiple natural strains promotes sequence divergence, translational ability, and translational efficiency in the VlsE variants. We used *in silico* simulation models to test if perturbing the observed variable sites leads to a decrease in sequence divergence, translational ability and translational efficiency.

2.2. Materials and Methods

2.2.1. Sequence analysis of vlsE and the unexpressed cassettes

The sequences of the unexpressed cassettes from six strains of *Bb* and one *B. afzelii* strain were used to investigate how diversifying selection and translational selection constrain identities and locations of variable sites among the unexpressed cassettes (Table 2-1). Each of the unexpressed cassettes within each strain was aligned using ClustalW [41] with default parameters. The unexpressed *vls* cassettes from all strains have six or seven variable regions in which variable sites are concentrated as described by Zhang et al [18]. Unexpressed cassettes that did not include all variable regions were not analyzed.

Strains	Number of polymorphic sites	Number of variable regions	Reference
B31	191	6	Zhang et al. [20] Schutzer, Fraser-Liggett [42]
JD1	166	6	Schutzer, Fraser-Liggett [42]
WI9123	115	6	Schutzer, Fraser-Liggett [42]
29805	140	6	Schutzer, Fraser-Liggett [42]
Bol26	234	6	Schutzer, Fraser-Liggett [42]
Zs7	90	6	Schutzer, Fraser-Liggett [42]
B. afzelii	223	7	Wang, Botkin [43]

Table 2-1. Unexpressed cassettes in six strains of *Bb* and in *B. afzelii*

2.2.2. In silico perturbation of unexpressed cassettes

For each set of natural cassettes, three perturbation models were generated using the three algorithms (δNuc, δPos, and δInDel) described below and in Figure 2-1. The perturbation models have altered either a) nucleotide identity at each variable site (δNuc), b) the locations of the variable sites within the variable regions (δPos), or c) the locations of insertion/deletion mutations within the variable regions (δInDel). All perturbation models were run independently on each strain to generate 50 iterations of perturbed cassettes.

The maximum-likelihood nucleotide substitution rates at the *vls* unexpressed cassettes in each strain was estimated under the General Time Reversible (GTR) nucleotide substitution model [44], which allows for different nucleotide frequencies and nucleotide substitution rates, using PhyML [45]. The maximum-likelihood substitution model was used to estimate the stationary frequencies of nucleotides in the unexpressed cassettes,

which was subsequently compared to the observed frequencies of nucleotides in the unexpressed cassettes. If the two frequency distributions were not significantly different, we concluded that nucleotide substitution has reached equilibrium among the unexpressed cassettes. To show that the model predictions are robust to the choice of nucleotide substitution model, perturbed cassettes that have altered nucleotide identities at the variable sites (δNuc) or locations of the variable sites (δPos) were generated by sampling nucleotides from both the maximum-likelihood stationary distribution and a uniform distribution assuming equal rate of substitution between any pair of nucleotides.

2.2.2.1. δNuc algorithm

The δNuc algorithm converts the nucleotides observed at every variable site in the cassettes of each natural strain to an alternative nucleotide (Figure 2-1A). That is, all nucleotides of identity X are converted to identity Y (for example, all adenines at a given variable site are converted to cytosines). The identity of the nucleotide to replace the original nucleotide is sampled either at random or from the maximum-likelihood stationary nucleotide frequency at the locus for each polymorphic site in each iteration of the model. Nucleotide conversion is bijective – all nucleotides at a variable site of nucleotide identity X will be converted to identity Y, and nucleotides of identity Y will only be used to replace identity X at that variable site. The δNuc algorithm only replaces nucleotides that differ from that observed in the *vlsE* sequence such that the total number of nucleotides that differ from the parental *vlsE* is not altered.

2.2.2.2. δPos algorithm

The δPos algorithm relocates the position of all variable sites of a given strain except

those sites that contain insertion/deletion mutations, to a random position within the variable regions of the unexpressed cassettes (Figure 2-1B). The nucleotide identities at the relocated variable sites are altered following the δNuc algorithm (2.2.2.1).

2.2.2.3. δInDel algorithm

The δInDel algorithm relocates the position of all variable sites, including those that contain insertion/deletion mutations, to random positions within the variable regions of the unexpressed cassettes (Figure 2-1C). The nucleotide identities at the relocated sites are sampled at random as in the δNuc algorithm (2.2.2.1), with the addition of gap as a fifth possible type of 'nucleotide' that can be replaced and can be used for replacement.

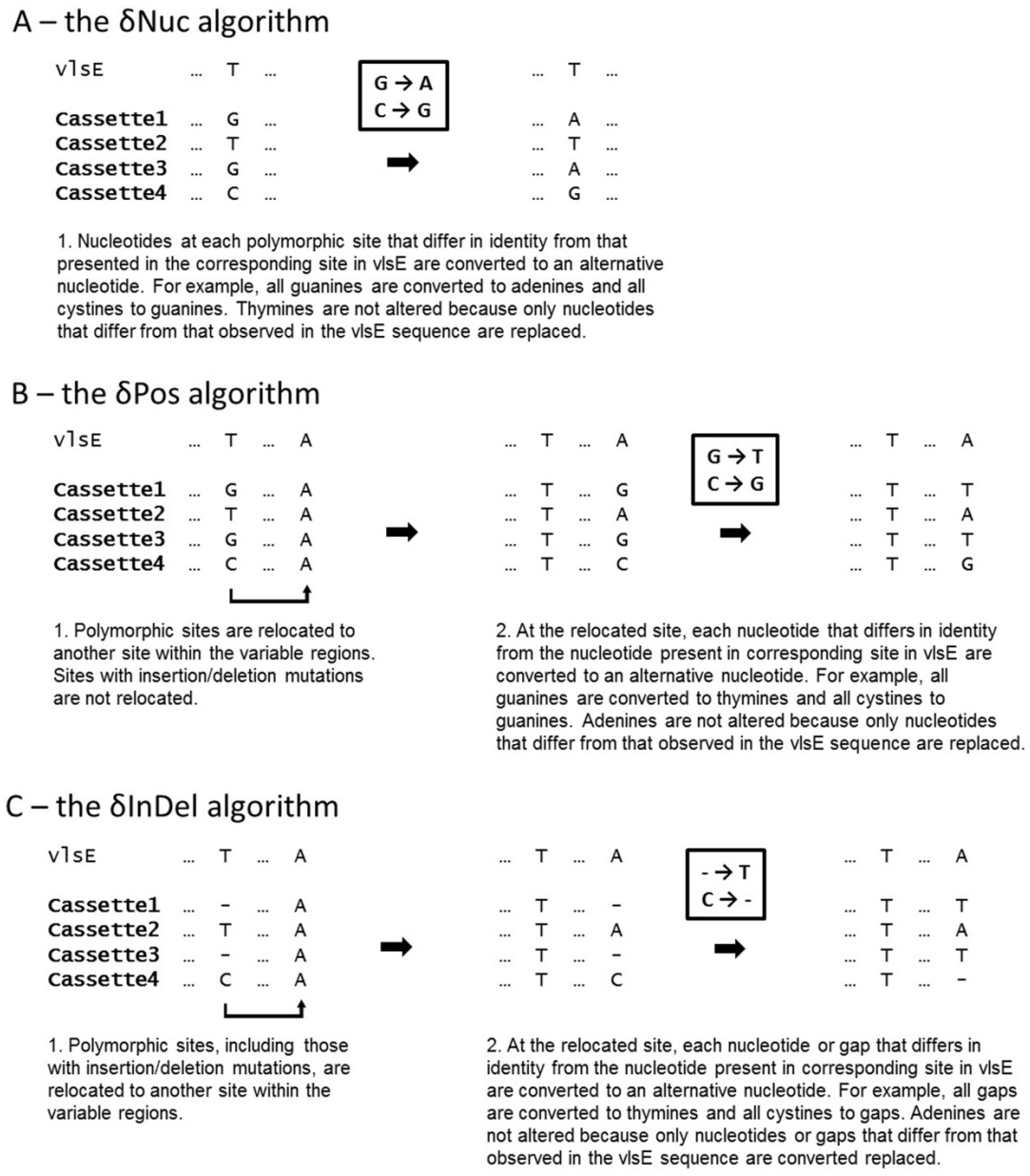


Figure 2-1. Examples of algorithms perturbing the organizations of polymorphic sites. (A) δNuc converts the polymorphic nucleotides to alternative nucleotides. (B) δPos relocates polymorphic sites within the variable regions. The relocated polymorphic sites are then perturbed as in δNuc. (C) δInDel relocates the polymorphic sites including those with

insertion/deletion mutations within the variable regions. The new polymorphic sites are then perturbed as in δNuc, with the addition of gap as a fifth type of 'nucleotide'.

2.2.3. In silico simulation of recombination events

Five hundred VlsE variants were generated from the natural unexpressed cassettes from each of the seven *Borrelia* genotypes analyzed, as well as from each of the *in silico*-perturbed sets of unexpressed cassettes. Each VlsE variant was generated by simulating one recombination event between the unexpressed cassettes and the natural *vlsE* sequence. For each recombination event, a segment was selected at random from the set of unexpressed cassettes and used to replace the homologous segment in the *vlsE* sequence observed in the natural strain. The length of the selected segment was chosen at random from an even distribution between 2 to 200bp. The range and distribution of segment lengths was derived from lengths of recombining fragments observed in *Bb* strains infecting laboratory mice [20].

2.2.4. Sequence divergence, translational ability, and translational efficiency of VlsE variants

We evaluated sequence divergence, translational ability, and translational efficiency of each of the VlsE variants generated from the *in silico* recombination events from each of the natural and perturbed unexpressed cassette.

2.2.4.1. Estimation of sequence divergence

Amino acid sequence divergence is correlated with antigenic divergence [25, 46, 47]. Amino acid sequence divergence between the *in silico*-generated VlsE variant and the natural VlsE sequence was estimated by calculating the ratio of non-identical amino acids in the alignment of the two sequences (amino acid weight matrix=GONNET, gap opening penalty=10, gap extension penalty=0.1).

2.2.4.2. Estimation of translational ability

Translational ability was defined as sequences that do not encode premature stop codons or frameshifts that are not corrected prior to the terminal stop codon. *vlsE* sequences that contained either premature stop codons or an unreverted frameshift had zero translational ability while those that could produce a complete protein had translational ability equal to one.

2.2.4.3. Estimation of translational efficiency

Although translational efficiency is determined by multiple factors, variations in elongation rate among VlsE variants may be especially important in the translational efficiency of this highly expressed protein because the recombination mechanism does not alter the 5’ or 3’ sequences of the expression site and thus does not alter initiation and termination rates [18]. Codons with more abundant tRNAs are likely to be translated with a greater elongation rate and with greater accuracy [40]. Thus, optimality of codon usage with respect to tRNA abundance was used as a proxy for translational efficiency. Codon optimality was estimated using two statistics, average tRNA copy number and tRNA

adaptation index, both of which have been shown to correlate with tRNA abundance [48-51] and codon decoding rate in many prokaryotes [52].

tRNA copy number, *E(T)*, was estimated as the number of copies of tRNA genes that pair to a codon, averaged over all codons in the *vlsE* sequence:

$$\mathrm{E(T)} = \frac{1}{n}\sum_{i=1}^{n} T_i$$

where n is the total number of codons in a *vlsE* variant, T_i is the number of copies of tRNA genes in the genome that pair with the i^{th} codon [53].

tRNA adaptation index, *tAI*, was calculated according to dos Reis et al [52]. First, the absolute adaptiveness value Wi for the i^{th} codon was calculated:

$$\mathrm{Wi} = \sum_{j=1}^{I_i}(1 - S_{ij})tGCN_{ij}$$

where I_i is the total number of tRNA isotypes that pair with the i^{th} codon, tGCN_{ij} is the numer of copies of the jth tRNA isotype genes in the genome that pair with the i^{th} codon, and S_{ij} is a selective constraint on the efficiency of the codon–anticodon coupling. Second, the relative adaptiveness value w_i of the i^{th} codon was calculated:

$$wi = \begin{cases} W_i/W_{mean} & if\ Wi \neq 0 \\ w_{mean} & if\ Wi = 0 \end{cases}$$

where W_{max} is the maximum W_i value and w_{mean} is the geometric mean of all w_i with $W_i \neq 0$. Finally, tAI was calculated as the geometric mean of the relative adaptiveness values of all codons in a *vlsE* variant:

$$\mathrm{tAI} = \left(\prod_{i=1}^{n} w_i\right)^{1/n}$$

where *n* is the total number of codons in a *vlsE* variant.

2.3. Results

2.3.1. Effect of altering the nucleotide identities observed in the variable sites

The maximum-likelihood stationary distribution of the four types of nucleotides was estimated at the unexpressed cassettes in each of the seven *Borrelia* genotypes respectively. The maximum likelihood stationary distributions of nucleotides were not significantly different from the observed distribution of nucleotides in the unexpressed cassettes of any of the seven genotypes (Table 2-2).

The nucleotide identities observed in variable sites of the unexpressed cassettes of natural strains result in the greatest sequence divergence, translational ability, and translational efficiency in VlsE variants. Changing nucleotides at the sites that are naturally polymorphic among the unexpressed cassettes (δNuc) reduced the sequence divergence among VlsE variants as well as the translational ability and translational efficiency of the variants (Figure 2-2). The reduction in sequence divergence, translational ability, tRNA gene copy number and tRNA adaptation index due to nucleotide alterations were statistically significant ($p<0.05$) in all genotypes except for tRNA gene copy number in BoI26 and tRNA adaptation index in 29805 when nucleotides are sampled from the maximum likelihood stationary distribution, as well as tRNA adaptation index in 29805 and zs7 when nucleotide types are sampled from a uniform distribution.

	Frequency A	Frequency T	Frequency G	Frequency C	p (chi-square)
B31 (ML)	0.25174	0.24123	0.40096	0.10607	>0.99
B31 (OB)	0.26222	0.23898	0.38755	0.11125	
JD1 (ML)	0.26558	0.24924	0.37682	0.10836	>0.99
JD1 (OB)	0.26032	0.25714	0.37490	0.10764	
WI9123 (ML)	0.29318	0.21727	0.38468	0.10487	>0.99
WI9123 (OB)	0.27846	0.22986	0.38887	0.10281	
29805 (ML)	0.21181	0.28739	0.38064	0.12016	0.98
29805 (OB)	0.21976	0.27708	0.37411	0.12905	
Bol26 (ML)	0.24390	0.27059	0.37647	0.10904	0.99
Bol26 (OB)	0.23457	0.27132	0.37353	0.12059	
Zs7 (ML)	0.25265	0.23737	0.39438	0.11559	0.99
Zs7 (OB)	0.26279	0.23704	0.38718	0.11299	
***B. afzelii* (ML)**	0.25697	0.24668	0.38509	0.11125	0.96
***B. afzelii* (OB)**	0.23912	0.26598	0.38075	0.11415	

Table 2-2. Maximum likelihood (ML) and observed (OB) nucleotide frequencies at the *vls* loci in seven genotypes of *Borrelia* used in the study

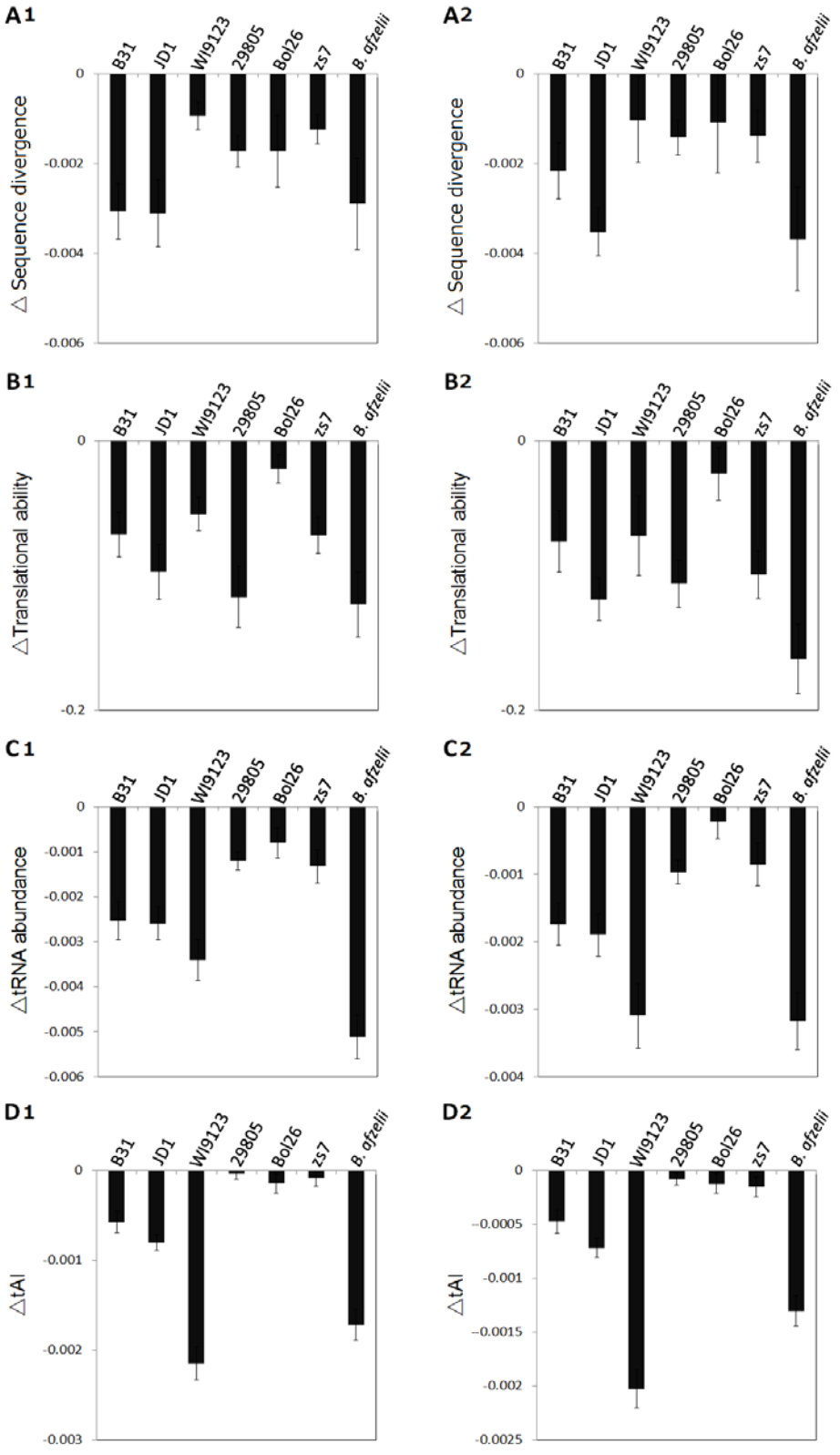

A1
A2
B1
B2
C1
C2
D1
D2
B31
JD1
Wl9123
29805
Bol26
zs7
B. afzelii
△ Sequence divergence
△Translational ability
△tRNA abundance
△tAI
0
-0.002
-0.004
-0.006
-0.2
-0.001
-0.003
-0.005
-0.0005
-0.0015
-0.0025

Figure 2-2. Changes in sequence divergence, translational ability, tRNA gene copy number and tRNA adaptation index of VlsE variants after altering the identities of the variable nucleotides in unexpressed cassettes (δNuc). The y-axis shows difference in sequence divergence (A1, A2), translational ability (B1, B2), tRNA gene copy number (C1, C2) and tRNA adaptation index (D1, D2) from the values of VlsE variants generated by the natural unexpressed cassettes after perturbing the variable nucleotides by sampling nucleotides randomly (A1, B1, C1, D1) or from a maximum-likelihood stationary distribution (A2, B2, C2, D2). Negative values on the y- axis mean a reduction after perturbation; zero means no change after perturbation. Each datapoint corresponds to the mean of 500 vlsE variants generated from one iteration of the perturbation model. Error bars represent 95% confidence intervals of the mean.

2.3.2. Effect of altering the positions of the variable sites

Altering the positions of the variable sites in the variable regions of the cassettes (δPos) resulted in a reduction in the translatability and translational efficiency of VlsE variants compared to the VlsE variants generated from the natural unexpressed cassettes in all analyzed strains (Figure 2-3B1, 2-3B2, 2-3C1 and 2-3C2). However, regardless of how the nucleotides were sampled to generate the perturbed cassettes, the translational ability of VlsE variants produced from the cassettes with altered variable locations was significantly lower than VlsE variants produced from cassettes with altered nucleotides (δNuc) in only two of the seven genotypes (WI9123 and Bol26), tRNA gene copy

number was significantly lower in only two of the seven genotypes (29805 and *B. afzelii*) and tRNA adaptation index was significantly lower in five of the seven genotypes (B31, JD1, 29805, BoI26 and *B. afzelii*) (Figure 2-2B1, 2-2B2, 2-2C1, 2-2C2, 2-3B1, 2-3B2, 2-3C1, 2-3C2). Thus, it is not apparent that altering the positions of the variable sites compounds the reduction in translational ability and efficiency of VlsE variants caused by altering the nucleotides alone. It is important to note that altering the positions of the variable sites must be accompanied by alterations in nucleotides, resulting in a reduction in power to detect effects of the position of the variable sites on translational ability and translational efficiency of VlsE variants independent of the large effect of nucleotide identity.

Sequence divergence among VlsE variants generated from the unexpressed cassettes with computationally altered positions of variable sites (δPos) was significantly lower than the variants produced by the natural cassettes in five of the seven strains (B31, JD1, 29805, zs7 and *B. afzelii*) (Figure 2-3A1 and 2-3A2). However, the divergence among VlsE variants generated from cassettes with positionally altered variable sites (δPos) was significantly lower than the difference produced from altering the nucleotide type (δNuc) in none of the seven genotypes (Figure 2-2A1, 2-2A2, 2-3A1 and 2-3A2). Thus, altering the positions of the variable sites does not cause a greater reduction in sequence divergence than altering the nucleotide identities alone. Further, alteration of the position of variable sites in one genotype (Bol26) resulted in a significantly *greater* sequence difference of the VlsE variants than the variants produced from the natural cassettes. Although the potential for VlsE variant diversity was significantly elevated by altering

the positions of the variable sites in this strain, the sequences in the expression site that resulted from these recombination events had low translational ability and efficiency (Figure 2-3B1, 2-3B2, 2-3C1 and 2-3C2) which are likely to constrain changes in the location of the variable sites.

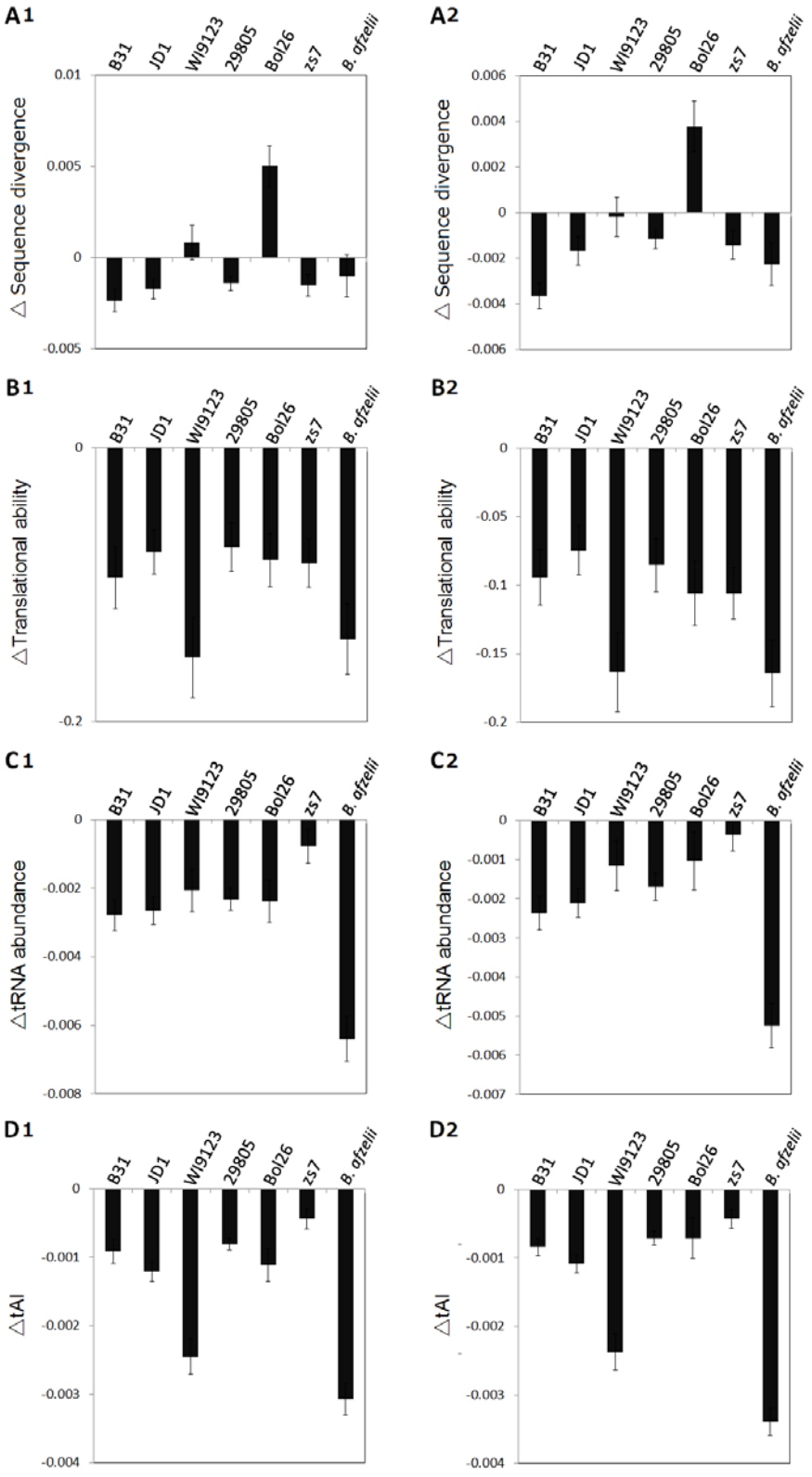

A1
B31
JD1
WI9123
29805
Bol26
zs7
B. afzelii
△ Sequence divergence
0.01
0.005
0
-0.005
A2
△ Sequence divergence
0.006
0.004
0.002
0
-0.002
-0.004
-0.006
B1
△Translational ability
0
-0.2
B2
△Translational ability
0
-0.05
-0.1
-0.15
-0.2
C1
△tRNA abundance
0
-0.002
-0.004
-0.006
-0.008
C2
△tRNA abundance
0
-0.001
-0.002
-0.003
-0.004
-0.005
-0.006
-0.007
D1
△tAI
0
-0.001
-0.002
-0.003
-0.004
D2
△tAI
0
-0.001
-0.002
-0.003
-0.004

Figure 2-3. Changes in sequence divergence, translational ability, tRNA gene copy number and tRNA adaptation index of VlsE variants after altering the positions of the variable sites in unexpressed cassettes (δPos). The y-axis shows difference in sequence divergence (A1, A2), translational ability (B1, B2), tRNA gene copy number (C1, C2) and tRNA adaptation index (D1, D2) from the values of VlsE variants generated by the natural unexpressed cassettes after perturbing the positions of the variable sites. Nucleotides at the relocated variable sites were sampled randomly (A1, B1, C1, D1) or from a maximum-likelihood stationary distribution (A2, B2, C2, D2). Negative values on the y- axis mean a reduction after perturbation; zero means no change after perturbation. Each datapoint corresponds to the mean of 500 vlsE variants generated from one iteration of the perturbation model. Error bars represent 95% confidence intervals of the mean.

2.3.3. Effect of altering the locations of insertion/deletion mutations

The locations of insertion/deletion mutations strongly affect the translational ability of VlsE variants. Altering the location of all variable sites, including those with insertion/deletion variants, significantly decreases translational ability compared to both the VlsE variants produced by the natural cassettes and the variants produced by cassettes with altered locations of variable sites (δPos) (Figure 2-3B1, 2-3B2 and 2-4B). The decreases in translational ability are dramatic, with as many as half of the variants containing at least one stop codon (Figure 2-4B). The VlsE variants also show significantly lower sequence divergence and lower translational efficiency than those

generated by the natural unexpressed cassettes in all strains (Figure 2-4A and 2-4C). These results indicate that the patterns of insertion/deletion occurrence are strongly constrained by sequence divergence, translational efficiency, and especially translational ability.

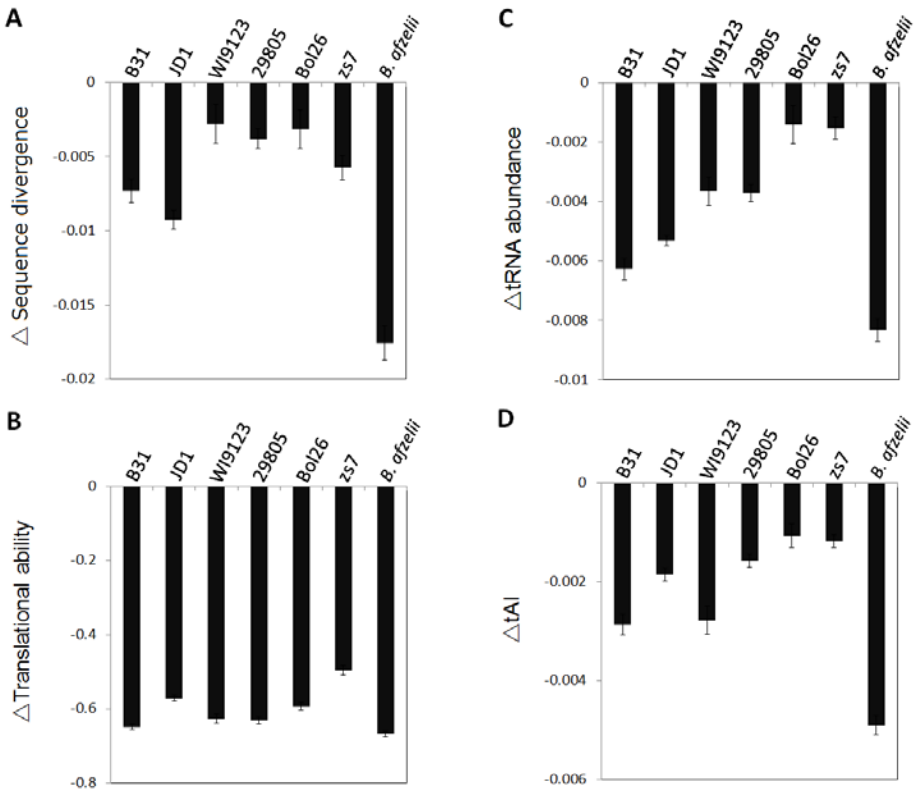


Figure 2-4. Changes in sequence divergence, translational ability, tRNA gene copy number and tRNA adaptation index of VlsE variants after perturbing the insertion/deletion mutations in unexpressed cassettes (δInDel). The y-axis shows difference in sequence divergence (A), translational ability (B), tRNA gene copy number (C) and tRNA adaptation index (D) from the values of VlsE variants generated by the

natural unexpressed cassettes after perturbing the insertion/deletion mutations. Negative values on the y- axis means a reduction after perturbation; zero means no change after perturbation. Each datapoint corresponds to the mean of 500 vlsE variants generated from one iteration of the perturbation model. Error bars represent 95% confidence intervals of the mean.

2.4. Discussion

Antigenic variation systems experience strong selection to evade the rapidly changing and lethal host immune environment. Thus, there is a premium for ever greater antigenic divergence among protein variants generated by antigenic variation systems. However, divergence among antigenic variants as well as the organization of the antigenic variation system as a whole can be constrained by selection for basic molecular functions such as translational ability or translational efficiency. The computational analyses presented here demonstrate that the *vls* antigenic variation system that generates antigenic diversity in natural strains of *Bb* during infections of vertebrate hosts is shaped by forces of natural selection acting on greater antigenic divergence, translational ability, and translational efficiency.

Bb strains generate antigenically distinct protein variants by recombining segments of diverse, homologous, unexpressed cassettes into the expression site of the antigen. In all natural strains analyzed, the sequence divergence among expressed proteins, as well as the translational ability and translational efficiency of those proteins, is highest by using

the nucleotides observed at the sites that are variable among the unexpressed cassettes. Altering the nucleotides used at these sites can reduce the sequence divergence, introduce stop codons, and reduce the elongation rate by using non-preferred codons. Although many of the *Bb* genomic codons are chosen because of mutational biases or replicational selection and are not optimized for translation [54], our results show that codon usage in *vlsE* is selected for translational efficiency. Codon optimality may be particularly important to the elongation rates and translational accuracy of VlsE variants as there are many variable codons that are tandemly arrayed in the unexpressed cassettes. Tandem codons using non-preferred codons can act synergistically to reduce translational rate and accuracy [55, 56]. A hypothesized function of the VlsE protein, to protect other surface-exposed proteins from antibodies by hiding these potential antigens [21], requires highly expressed and accurately translated proteins for appropriate antigenicity. It is worth noting that, although the statistical difference in codon optimality between natural and perturbed cassettes is consistent across *Bb* strains, the quantitative influence of such difference on VlsE concentration is not tested and beyond the scope of this thesis. Future analysis in the translational dynamics of VlsE may shed light on this question.

Altering the position of the variable sites within the variable regions significantly reduced both the translational ability and translational efficiency of VlsE variants generated. The sequence divergence of VlsE variants generated from positionally altered cassettes, however, was significantly lower than that from the natural cassettes in only five of the seven strains. Further, divergence among the VlsE variants generated after altering the positions of the variable sites in one strain (Bol26) was significantly *greater* than those generated from the natural cassettes. Strain BoI26 is a unique case where many codons

contained multiple variable nucleotides. These variable sites become separated on different codons after positional perturbation, leading to greater amino-acid-level divergence after recombination.

The organization and identity of variable sites in the unexpressed cassettes of natural strains resulted in high amino acid divergence among the VlsE variants, suggesting that the sequences of the unexpressed cassettes observed in natural strains strongly promotes antigenic divergence. Amino acid sequence difference among antigens is negatively correlated with cross-reactivity with polyclonal antibodies [25, 46, 47]. This reduction in binding affinity to antibodies is crucial for immune evasion and longevity of *Bb* in infected hosts, a key factor in the fitness of *Bb*. Although altering the positions of variable sites in one strain (Bol26) increases sequence differences of the VlsE variants, which may provide greater ability to evade the immune response, many of these VlsE variants contained stop codons, frameshifts, and low translational efficiency. Therefore, the potential advantage of greater antigenic diversity is likely constrained by selection for efficient expression of functional proteins.

Perturbing both nucleotide identities and the positions of variable sites did not result in a greater reduction in antigenic divergence, translational ability, and translational efficiency of VlsE variants than altering only the nucleotide identities. Although the positions of the variable sites in the unexpressed cassettes may be selectively important, the absence of an observed effect may be attributed to two factors. First, the positions of variable sites cannot be altered without also altering the nucleotides at those positions. Thus, the effect of the position of the variable sites cannot be disentangled from the effects of nucleotide identities at these sites, resulting in low statistical power to detect the effect of the

variable positions. Second, the positions of variable sites might be important for traits that were not evaluated. For example, amino acids at the positions that are naturally variable may be more antigenic than those at other positions or may impact the three dimensional structure of VlsE.

Length differences among unexpressed cassettes occur almost exclusively in triplets in all natural strains, resulting in very low levels of frameshift mutations during recombination. The triplet pattern of insertion/deletion mutations may result from slipped-strand mispairing induced by trinucleotide repeats that are common in the variable regions of all strains [27]. The δInDel algorithm, however, randomized the locations of insertion/deletion mutations, disrupting the triplets and dramatically increasing the rate of frameshift mutations in the VlsE variants. These frameshift mutations result in a large number of premature stop codons in *vlsE*. This observation is consistent with the hypothesis that natural selection can increase off-frame stop codons because they prevent resource waste and potentially cytotoxic products when frameshifts occur [57].

In recombination-based antigenic variation systems, the genetic sequences of the unexpressed cassettes determine the possible degree of diversity among antigenic variants during a host infection. The analyses presented in this chapter demonstrate that the *vls* antigenic variation systems of *Bb* is organized to promote antigenic divergence among proteins variants without significantly sacrificing translational ability and translational efficiency. The selection on translational properties in unexpressed regions was statistically significant and in some cases appeared to constrain the selection for greater antigenic diversity. These analyses identified properties of the antigenic variation system

that appear to be under selective constraints which can be experimentally investigated in future studies. The selection pressures described here could be common in the diverse range of pathogens that generate antigenic variants by intragenomic recombination such as *Neisseria gonorrhoeae* and *Mycoplasma synoviae* [34, 35]. In these pathogens, similar computational analyses to those conducted in this study can be used to establish the sets of selection pressures that affect the evolution of unexpressed cassettes.

Chapter 3. Predicted effects of VlsE diversity on *Borrelia burgdorferi* dynamics within hosts

3.1. Introduction

Lyme disease is the most commonly reported tick-borne illness in North America. The causative agent, *Borrelia burgdorferi* (Bb), establishes long-term, multisystemic infections despite active adaptive immune responses from vertebrate hosts [18, 58]. Bb evades host immune responses through multiple mechanisms including the vital and well-characterized *vls* antigenic variation system [18]. The *vls* antigenic variation system alters the sequence of the immunodominant surface-exposed antigen, VlsE, resulting in VlsE sequence diversity within hosts [18, 19]. Subpopulations of Bb with altered VlsE sequences evade antibody recognition, thus permitting the characteristic long-term infections [18, 22-24, 38, 39]. The quantitative effect of the generation of VlsE diversity on the probability of immune evasion or on Bb population sizes within hosts has not been investigated.

The within-host population dynamics of Bb are controlled by the adaptive immune response. In immunocompetent mice, Bb rapidly proliferates in the initial phases of the infection before inducing a strong antibody response against several surface exposed proteins [59], predominantly VlsE, around day 7 of the infection, leading to a rapid reduction in abundance [60]. Yet, Bb is rarely cleared in immunocompetent mice likely due to the down regulation of most surface exposed antigens and the antigenic variation generated at VlsE [18, 60]. The degree of diversity among VlsE variants expressed in the

Bb population is expected to be essential to avoiding clearance by the antibody response of the host [20, 27]. Evolutionary studies demonstrated that natural selection selects for greater potential to generate diversity at VlsE and that the variable sites within the cassettes are organized to maximize diversity among VlsE variants [27, 61]. Yet it remains unclear how the diversity among VlsE variants facilitates within-host persistence or affects Bb population dynamics within hosts.

In this study, we investigated the effects of antigenic diversity at VlsE on the population dynamics of Bb within infected hosts. The population dynamic models were parameterized using relevant experimental data and the results were compared to data on Bb population dynamics of Bb from experimental infections. Importantly, we compared the effect of a strain's ability to generate diversity at VlsE on the population dynamics of the pathogen. In addition, the models make predictions that can be experimentally tested to estimate the likelihood of mechanisms described in the models.

3.2. Materials and Methods

3.2.1. Dynamic model

The dynamic model used to assess Bb population dynamics incorporates the presence and abundance of each VlsE variant as well as the specificities, affinities and abundances of antibodies. Bb population dynamics - including growth, recombination and immune clearance - are described by:

$$\frac{dBb(i)}{dt} = \omega\beta Bb(i)\left(1 - \frac{\sum_i Bb(i)}{K}\right) - rBb(i) - \mu \times Abrxn(i)Bb(i) + \nu Bb^*(i) \qquad \text{eq. 3.1}$$

$$\frac{dBb^*(i)}{dt} = \mu \times Ab\text{rxn}(i)Bb(i) - \nu Bb^*(i) - rBb^*(i) - \rho Bb^*(i) \qquad \text{eq. 3.2}$$

Such that Bb expressing VlsE variant type *i* (*Bb(i)*) grows logistically with an intrinsic growth rate β. Antibody-independent immune effects are described as an adjustment in the growth rate by a factor ω. A carrying capacity *K* for the Bb population is estimated from *in vitro* growth data [62]. *Bb(i)* recombines with the unexpressed *vls* cassettes at rate *r*. *Bb(i)* interacts with antibodies at association rate μ. *Abrxn*(*i*) represents the binding capacity of all antibodies present in the system to *Bb(i)* and is a function of the concentration of each type of antibody, corrected by the specificity of that antibody for the VlsE variant, summed across all antibodies present (detailed in 3.2.1.3). *Bb(i)* transitions to the susceptible state $Bb^{*}(i)$ when bound by an antibody, which is then removed by the immune system at constant rate ρ. $Bb^{*}(i)$ can dissociate from antibodies at rate ν.

3.2.1.1 Antigenic variation by recombination

Bb infecting immunocompetent mice remain at densities much lower than the carrying capacity due to the host antibody responses [60]. However, Bb are not eliminated due to the sequence changes in VlsE introduced by recombination, which reduces antibody killing efficacy. Pairwise sequence similarity between VlsE variants is expected to be a function of the sequence diversity among the *vls* cassettes that recombine into *vlsE* and the number of recombination events that have occurred since the variants shared a common ancestor. We estimated the distribution of pairwise sequence similarities between VlsE variants by simulating recombination between *vls* cassettes and VlsE. To do this, we first generated 1000 *vlsE* parental variants ($vlsE^{n}$). Each parental variant was generated by simulating 100 sequential recombination events starting from the *vlsE*

sequence of Bb clone 5A3, as described in Chapter 2.2.3. From each of the 1000 parental *vlsE*^{n} sequences, we simulated 20 daughter sequences (vlsE$_1^{n}$-vlsE$_{20}^{n}$) that experienced between 1 and 20 recombination events in order to estimate sequence similarity between *vlsE* variants as a function of the number of recombination events (Dr). Pairwise sequence similarity between each VlsE^{n} and each of its daughter sequences (VlsE$_1^{n}$-VlsE$_{20}^{n}$) was calculated as amino acid sequence identity. The expectation (*M(Dr)*) and variance ($\sigma(Dr)$) in sequence similarities between VlsE variants separated by *Dr* recombination events was calculated as the mean and standard deviation of the 1000 sequence similarity values for each value of *Dr*.

Two sets of *vls* unexpressed cassettes were used to estimate the distribution of pairwise sequence similarities between VlsE variants: the natural *vls* unexpressed cassettes observed in Bb clone 5A3 and a set of unexpressed cassettes generated *in silico* by altering the nucleotides at the variable sites as described in 2.2.2.1. The *in silico* generated unexpressed cassettes produces VlsE variants with lower diversity than the natural unexpressed cassettes.

3.2.1.1.2. Sequence similarities among VlsE variants within infections

In the model of Bb population dynamics, recombination between *vlsE* and the unexpressed *vls* cassettes results in a novel *vlsE* variant, similar to an infinite alleles model. Each new *vlsE* variant can be represented by a node in a tree graph, connected to its direct parental *vlsE* by one edge. The number of recombination events (*Dr*) between VlsE variants *i* and *j* is equivalent to the number of edges connecting the nodes

representing variants *i* and *j*. For VlsE variants *i* and *j* that are separated by over 20 recombination events, *Dr* is set to 20. This approximation utilizes the diminishing effect of recombination events (*Dr*) on sequence similarity (section 3.3.5). Sequence similarities between VlsE variant type *i* and *j* (*S(i,j)*) within an infection are determined by sampling from the distribution of pairwise sequence similarities conditional on *Dr*:

$$P(S(i,j)|Dr) = N(M(Dr), \sigma(Dr)^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(S(i,j)-M(Dr))^2}{2\sigma(Dr)^2}} \quad \text{eq. 3.3}$$

The sequence similarities among variants created by recombination are relevant to the degree of reactivity of antibodies specific to variant *i* have for variant *j*.

3.2.1.2 Antibody kinetics

The concentration of antibodies targeting VlsE variants was modeled similar to Michaelis-Menten kinetics [62, 63]. Antibodies specifically targeting VlsE variant type *i* are triggered when the abundance of the *Bb(i)* subpopulation is greater than 1000 individuals. The magnitude of the antibody response that is specific to VlsE type *i* (*Abspecific(t,i)*) is estimated by:

$$Abspecific(t,i) = \begin{cases} 0 & t - t_{trig} < \Delta t_{lag} \\ \frac{Ab_{max}\,(\,t - t_{trig} - \Delta t_{\text{lag}})^n}{t_{1/2}^2 + (t - t_{trig} - \Delta t_{\text{lag}})^n} & t - t_{trig} > \Delta t_{\text{lag}} \end{cases} \quad \text{eq. 3.5}$$

When $\sum_i Abspeicifc(t,i) \leq Abmax$

Where t_{trig} is the time when the antibody specific to variant type *i* is triggered. If the sum of the concentrations of all antibody types exceeds the experimentally observed

maximum antibody concentration (*Abmax*) [63, 64], the concentration of each antibody type was normalized to *Abmax*.

$$Abspecific(t,i) = Ab(t)\frac{Abspecific(t,i)}{\sum_i Abspecific(t,i)} \quad \text{eq. 3.6}$$

When $\sum_i Abspeicifc(t,i) > Abmax$

3.2.1.2.1 Immunodominant antibody kinetics

A potential constraint on antibody dynamics is the number of types of antibodies that can be induced per unit of time. Although the naïve antibody repertoire is highly heterogeneous, only a small subset of the B cell clones will proliferate significantly during an infection - a well-characterized phenomenon called antibody immunodominance [25, 65]. We extended the dynamics model to include antibody immunodominance to account for this constraint. For the sake of simplicity, we allowed only one immunodominant antibody type to be actively produced at a given time. The type of antibody that was activated was chosen based on the relative abundance of the corresponding VlsE variant.

Each dominant antibody is actively produced for *Td* hours before the next type of dominant antibody was allowed to expand, eventually replacing the previous dominant antibody. The dynamics of each type of dominant antibody was described by:

$$Abspecific_{\mathrm{j}}(t,i) = \begin{cases} 0 & t - t_{trig} < \Delta t_{\mathrm{lag}} \\ \frac{Ab_{max}\,(t - t_{trig} - \Delta t_{\mathrm{lag}})^n}{t_{1/2}^2 + (t - t_{trig} - \Delta t_{\mathrm{lag}})^n} & \Delta t_{lag} \le t - t_{trig} < Td \\ \frac{Ab_{max}\,(t - \Delta t_{\mathrm{lag}})^n}{t_{1/2}^2 + (t - \Delta t_{\mathrm{lag}})^n} - Abspecific_{j+1}(t,i) & t - t_{trig} \ge Td \end{cases} \quad \text{eq. 3.11}$$

Where $Abspecific_j$ represents the concentration of the *j*th activated immunodominant antibody.

3.2.1.3 Antibody effects on Bb dynamics

Specific and non-specific but cross-reactive antibodies negatively impact the abundance of Bb. The quantitative impact of antibodies specific to VlsE variant *j* on the dynamics of Bb expressing VlsE variant *i* is a function of the abundance of antibody *j* (*Abspecific(t,j)*) and the degree of cross-reactivity between antibody *j* and VlsE variant *i* (*f(S(i,j))*). The degree of cross reactivity between antibody *j* and VlsE variant *i* is dependent on the similarity between VlsE types *i* and *j*:

$$Abrxn(t,i) = \sum_j Abspecific(t,j)\, f(S(i,j)) \qquad \text{eq. 3.7}$$

Where $f(S(i,j))$ is a function that modulates antibody reactivity via the sequence similarity between the VlsE variants *i* and *j*. By definition, this function should satisfy *f(1)=1* and *f(0)=0*. For simplicity, cross-reactivity is modeled using a power function.

$$f\big(S(i,j)\big) = S(i,j)^x \qquad \text{eq. 3.8}$$

Where larger values of *x* correspond to higher antibody specificity and lower cross-reactivity.

3.2.1.4 The immune effector cells

The rate at which antibody-bound Bb cells are eliminated depends on the concentration of immune effector cells such as phagocytes [62]. Therefore, population dynamics of immune effector cells may also affect Bb population dynamics. Immune effector cell

dynamics are modeled similarly to Binder et al [62]. Immune effector cells are assumed to behave qualitatively similar to phagocytes in the sense that they have an intrinsic migration rate, a recruitment rate in response to the presence of antigens, and an intrinsic death rate:

$$\frac{dM}{dt} = \phi + \frac{\psi \sum_i (Bb(i)+Bb^*(i))}{\sum_i (Bb(i)+Bb^*(i))+C} M - \theta M \qquad \text{eq. 3.9}$$

Where ϕ is the intrinsic rate of recruitment of immune effectors to an infection sites; ψ is the recruitment rate of immune effectors in response to infection; θ is the death rate of the immune effectors; and C is the number of Bb that corresponds to half maximum recruitment rate.

When immune effector cell dynamics are included in the model, eq. 3.2 becomes:

$$\frac{dBb^*(i)}{dt} = \mu \times Abrxn(i)Bb(i) - \nu Bb^*(i) - rBb^*(i) - \rho^0 Bb^*(i)M \qquad \text{eq. 3.10}$$

Where the rate of removal of the antibody-bound Bb follows the law of mass action.

3.2.2. Numerical simulation based on the dynamics model

Bb population dynamics were simulated numerically to allow stochastic events to impact the qualitative and quantitative outcomes of the system. Parameter values used in numerical simulation were estimated based on literature or by fitting experimental observations to minimize the sum of squared errors between observations and predictions. The numerical simulation was repeated 20 times with each set of parameter values.

3.3. Results

3.3.1. Bb dynamics cannot be explained by the dynamics model without including antibody immunodominance or dynamics of immune effector cells

Regardless of the parameters used, the Bb population dynamics observed in experimental infections could not be replicated using the simple population dynamic model that does not explicitly account for either antibody immunodominance or for immune effector cell dynamics. Simulations across the simple model using either low removal rates of antibody-coated Bb (ρ) or high antibody specificity (x) permitted persistent infections but the dramatic decline in Bb population densities observed in experimental infections after the initial proliferation could not be reproduced (Figure 3-1). By contrast, high removal rates or low antibody specificity recreated the rapid decrease in population densities after the initial proliferation but populations were consistently eliminated from the host (Figure 3-1). Parameter values used for numerical simulation are listed in Table 3-1.

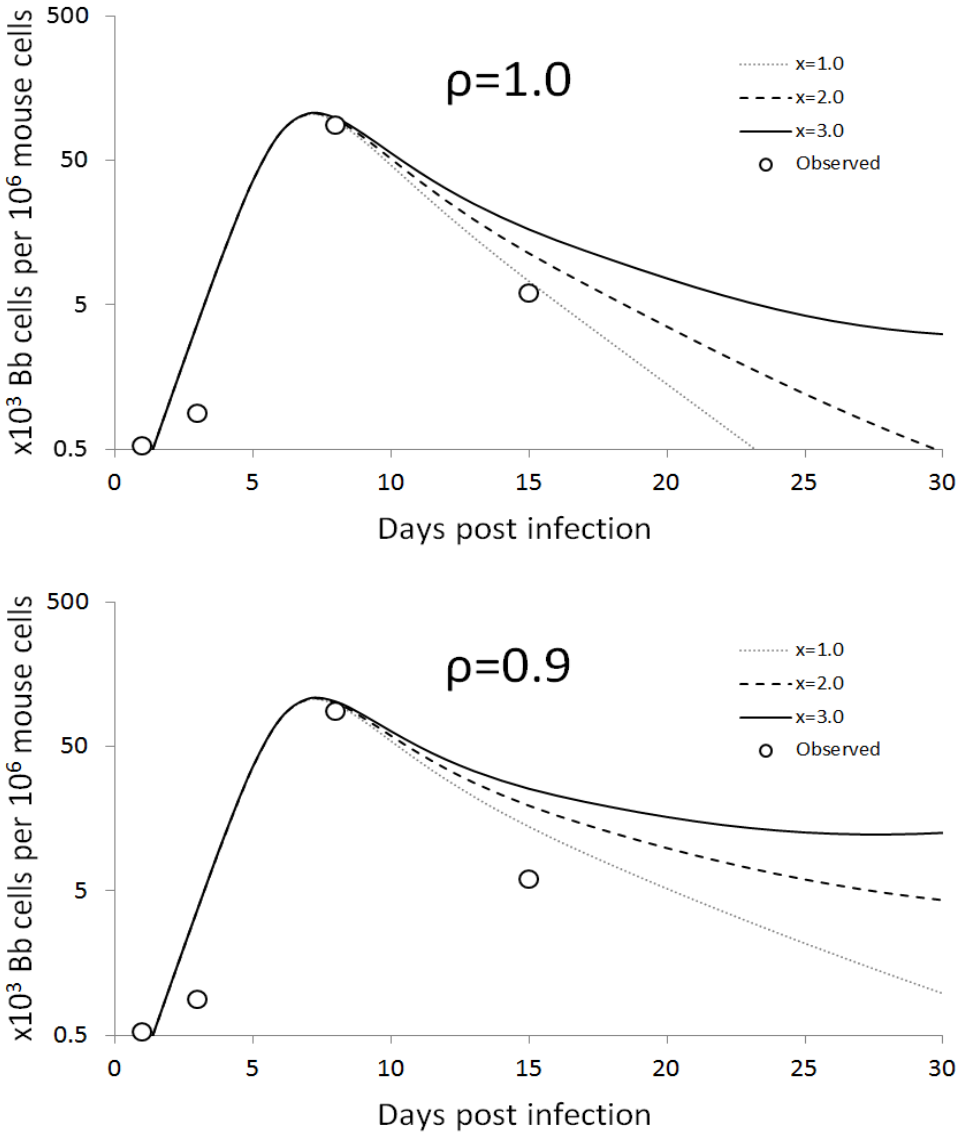


Figure 3-1. Numerical simulations of the Bb population dynamic models that did not include antibody immunodominance nor immune effector cell dynamics could not reproduce experimental time series of Bb population dynamics in mice. Bb population dynamics are simulated with different values of the removal rate (ρ) and antibody specificity (x). Open circle, Bb levels experimentally observed by Pahl et al. [60].

Parameter	Unit	Value	Derived from	Description
β	$hour^{-1}$	0.06	[66]	Bacterial growth rate
K	cells	150000	[60]	Bacterial carrying capacity
ρ^0	$cells^{-1}\ hour^{-1}$	0.0082	[62]	Immune effector cells killing rate
Ab_{max}	$\mu g\ ml^{-1}$	8.5	[59]	Maximum total antibody concentration
ω	-	0.9	Estimated	Decrease in bacterial growth rate due to Innate immune effect
r	$hour^{-1}$	0.0059	[20]	Recombination rate
μ	$ml\ \mu g^{-1}\ hour^{-1}$	0.0085	[62]	Antibody association rate
ν	$hour^{-1}$	1	[62]	Antibody dissociation rate
Δt_{lag}	hour	72	[63]	Time until antibody response starts
x	-	7.5(immune effector dynamics) 2.25(antibody immunodominance)	Estimated	Antibody specificity
ϕ	$cells\ hour^{-1}$	0.1	[62]	Intrinsic recruitment rate of immune effector cells
ψ	$hour^{-1}$	0.025	Estimated	Bb-dependent recruitment rate of immune effector cells
C	cells	35000	Estimated	Bb population size that yields half-maximal recruitment of immune effector cells
θ	$hour^{-1}$	0.08	[67]	Immune effector cells death rate
Td	hour	360	Arbitrary	Time when one type of antibody is dominantly expressed

Table 3-1. Parameters used in the numerical simulations.

3.3.3. Including antibody immunodominance in the model can account for experimentally observed Bb population dynamics within hosts

The Bb dynamics observed in experimentally infected mice can be reproduced by including immunodominance - where only one type of antibody is actively expressed at a time - into the population dynamic model (Table 3-1). Bb populations rapidly increase in density followed by a dramatic decline caused by antibodies specific to the parental VlsE variant (Figure 3-2A). However, a sufficient number of recombinant VlsE variants are different enough from the parental VlsE to allow continued persistence in the host, similar to the observed data. The relative abundance of Bb cells expressing the parental VlsE decreased rapidly following the specific antibody response as was observed in experimentally infected mice (Figure 3-2B). The predicted Bb population dynamics and VlsE variant dynamics that most closely resembled the empirical data from experimental infections assumed that antibody specificity was related to sequence diversity as $S(i,j)^{2.25}$ (Table 3-1).

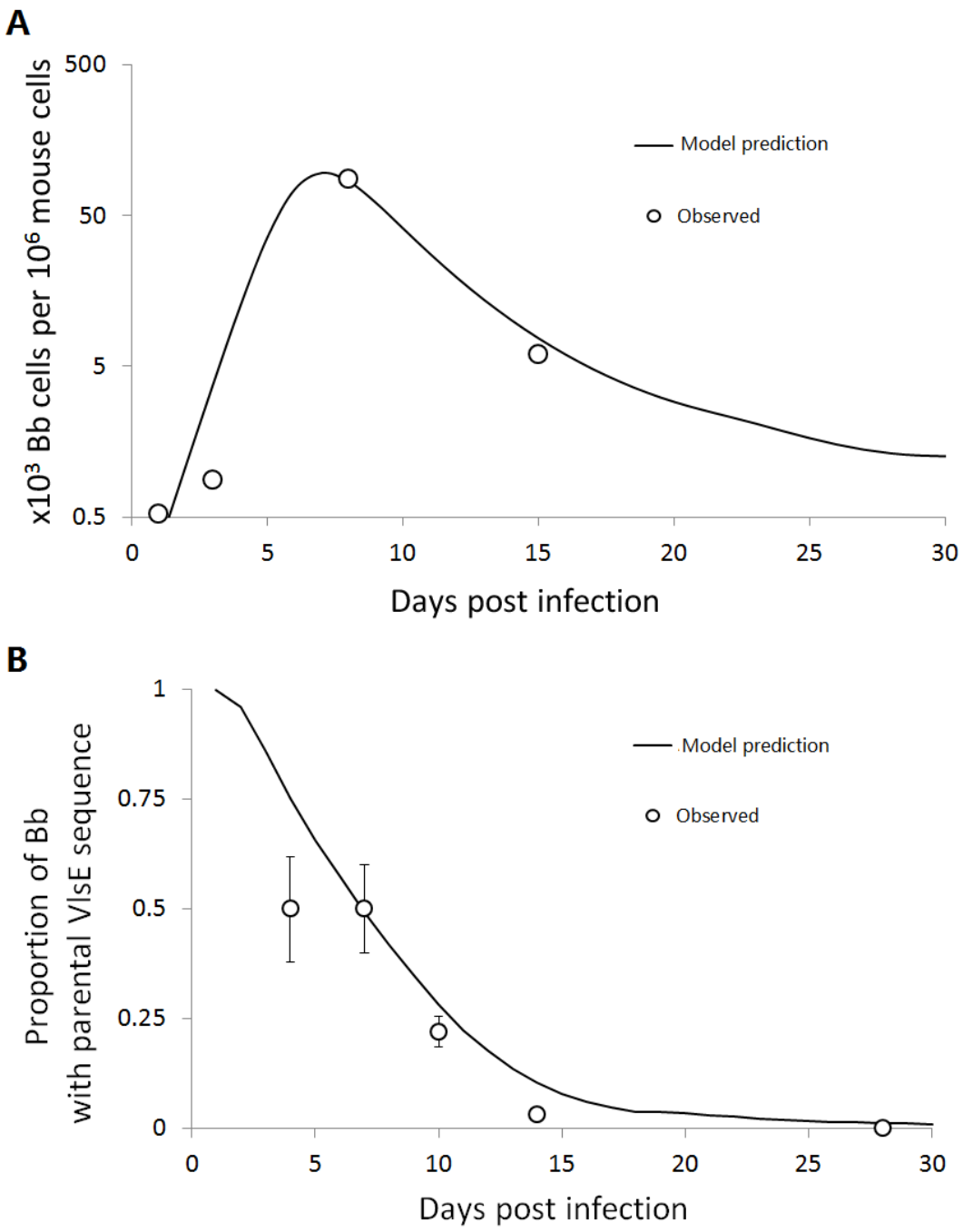


Figure 3-2. Numerical simulation of Bb population dynamics with antibody immunodominance included in the model can reproduce the experimental time series. (A) The Bb population dynamics predicted by the model that includes antibody

immunodominance replicates the data from experimental infections. Open circle, Bb levels experimentally observed by Pahl et al. [60]. (B) Additionally, the proportion of Bb cells retaining the parental VlsE sequences predicted by the numerical simulation was similar to the proportion retaining the parental sequence in experimental infections observed by Coutte et al. [20] in C3H mice (open circles).

The mean antibody pressure (*Abrxn*) across extant VlsE variants differed considerably between the models that included and excluded immunodominance (Figure 3-3). In the models including antibody immunodominance, the average antibody pressure increased each time a new type of antibody was generated and declined thereafter as the proportion of specific variants declined (Figure 3-3 A), demonstrating a fluctuating and dynamic effect on the Bb population. The immunodominant antibody has high affinity to one VlsE variant, which is rapidly removed from the population, and distinct affinities to the other VlsE variants, resulting in a much larger variance in antibody pressure among VlsE types than in models that allow multiple antibodies, each with affinities to different VlsE types, simultaneously (Figure 3-3B). Antibody pressure in models that excluded immunodominance was remained stable after the initial increase (Figure 3-3A).

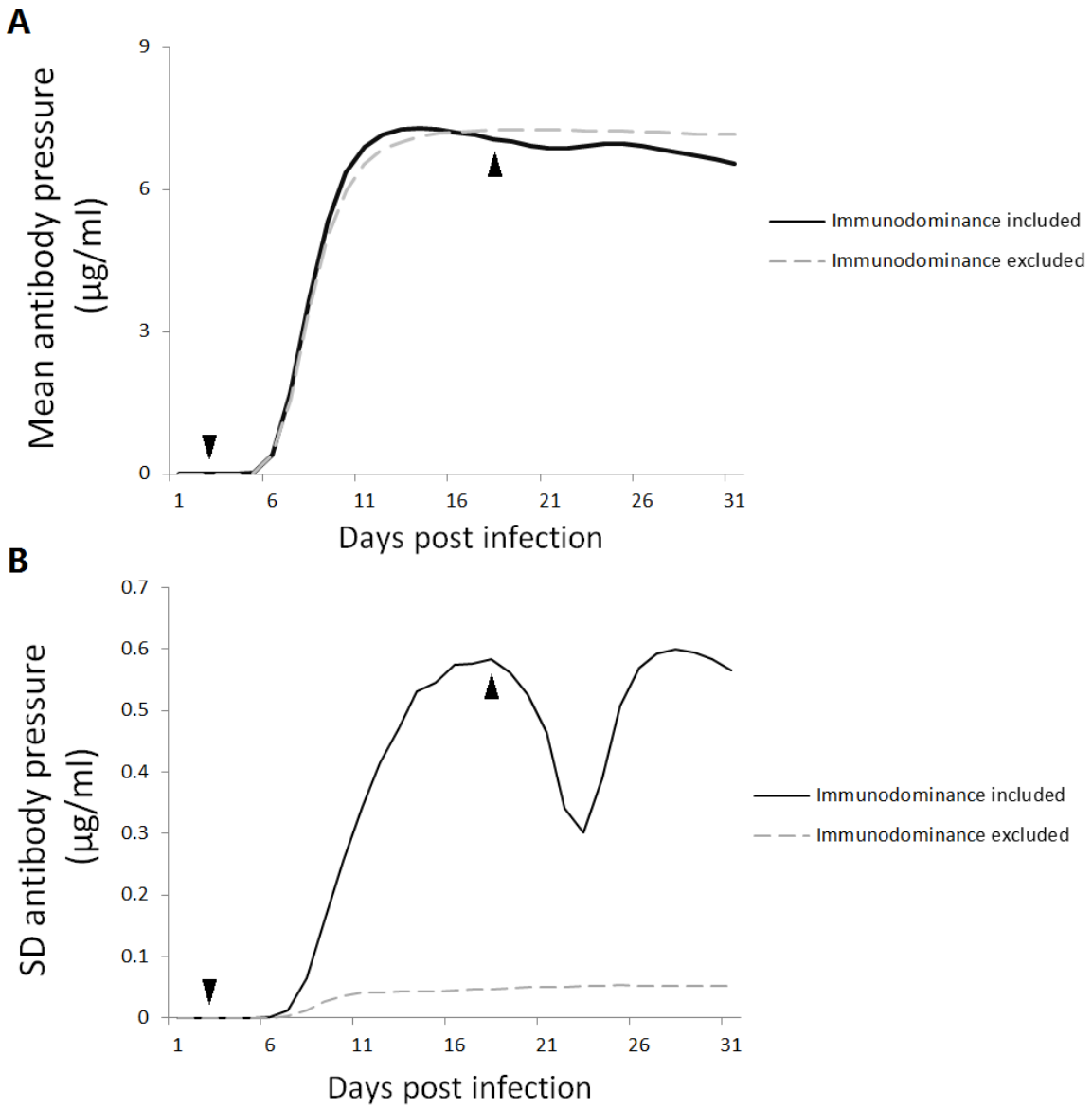
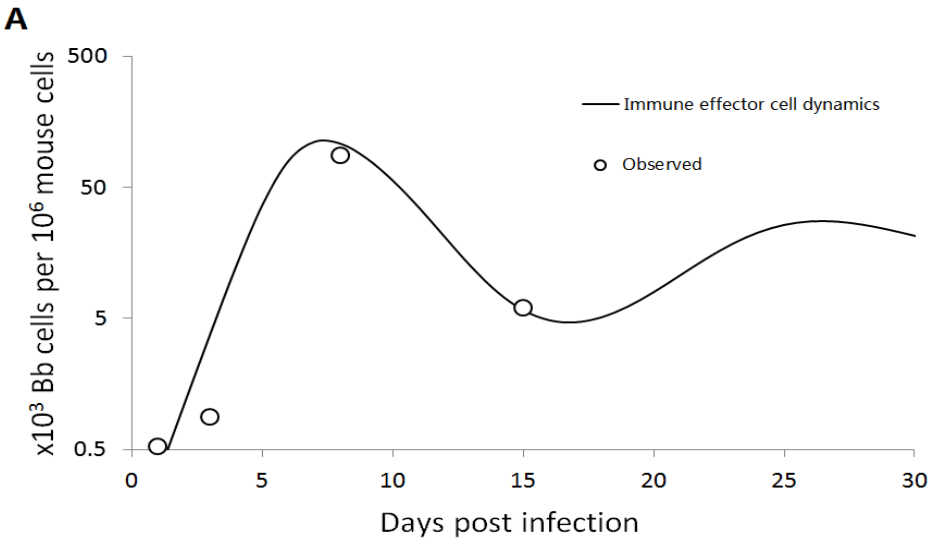


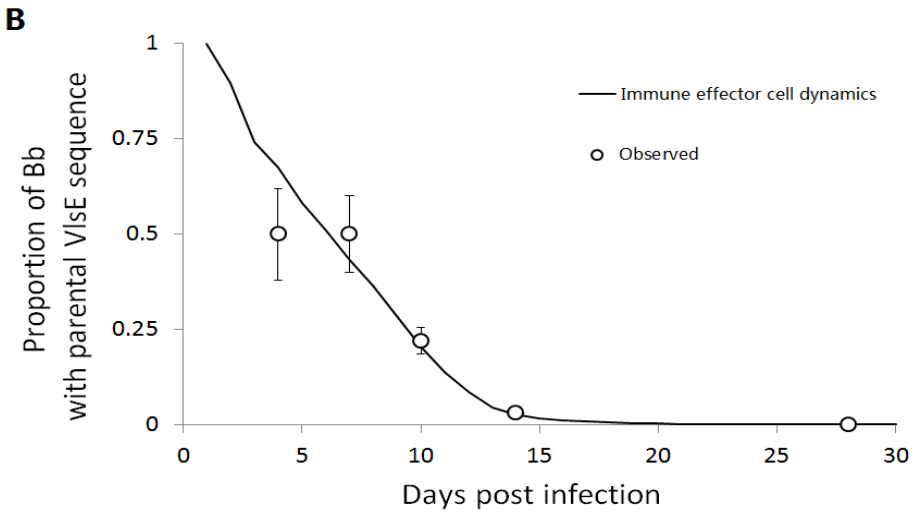
Figure 3-3. Antibody immunodominance creates temporal variation in antibody pressure ($Abrxn$) on Bb within hosts. (A) The mean antibody pressure across all extant VlsE variants increases rapidly after the first antibody is expressed (arrow at day 3) in all models. If antibody immunodominance is not included, the mean antibody pressure remains constant while new antibodies are continuously added. Including

immunodominance results in declining mean antibody pressure until the expansion of a subsequent antibody (arrow at day 18) causes the mean antibody pressure to increase. (B) The difference in the standard deviation in antibody pressure across the extant VlsE variants between models that include immunodominance and those that do not is dramatic. In models that include immunodominance, variance in antibody pressure increases rapidly as the dominant antibody targeting the dominant VlsE variant proliferates, then decreases due to the reduction in the density of the dominant VlsE variant until the introduction of the subsequent antibody. In models that exclude immunodominance, there are many antibodies that each react with each extant VlsE variant, such that each VlsE variant experience similar antibody pressure.

3.3.4. Models including immune effector cell dynamics can reproduce the experimentally observed Bb dynamics

Explicitly modeling immune effector cell dynamics, while excluding immunodominance, resulted in Bb population dynamics similar to observed experimentally (Figure 3-4A). Antibody specificity estimated from the model was very high $S(i,j)^{7.5}$. Similar to the models including immunodominance, the relative abundance of Bb cells expressing the parental VlsE decreased rapidly following the specific antibody response as was observed in experimentally infected mice (Figure 3-4B). The simulation described temporal fluctuations in abundances of both Bb populations and of immune effector cells (Figure 3-4C).


A
500
50
5
0.5
x10³ Bb cells per 10⁶ mouse cells
Immune effector cell dynamics
Observed
0
5
10
15
20
25
30
Days post infection


B
1
0.75
0.5
0.25
0
Proportion of Bb with parental VlsE sequence
Immune effector cell dynamics
Observed
0
5
10
15
20
25
30
Days post infection

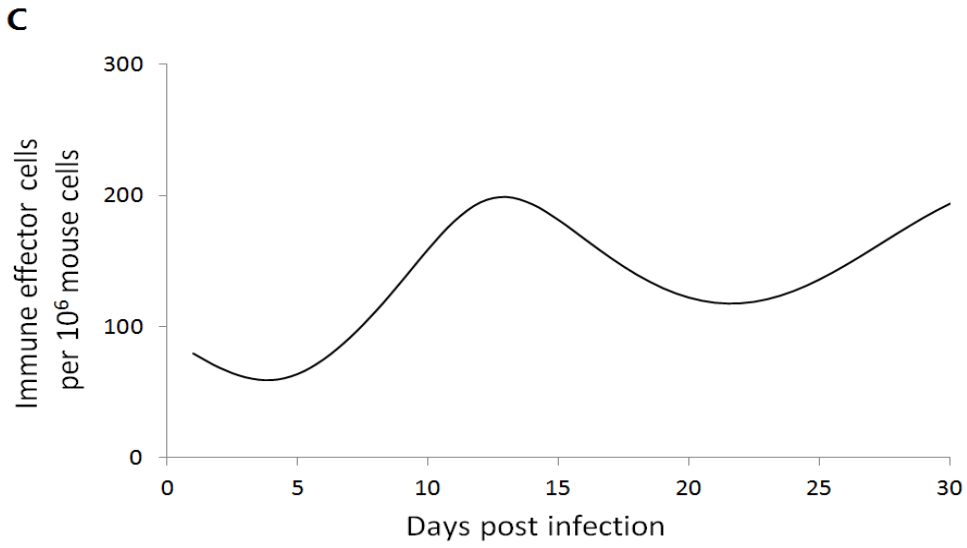

C
300
200
100
0
Immune effector cells per 10⁶ mouse cells
0
5
10
15
20
25
30
Days post infection

Figure 3-4. Numerical simulations of Bb population dynamic models that include immune effector dynamics can reproduce the data from experimental infections in mice. (A) The Bb population dynamics predicted by the model that includes immune effector cell dynamics replicates the data from experimental infections. Open circle, Bb levels experimentally observed by Pahl et al. [60]. (B) The proportion of Bb cells retaining the parental VlsE sequence predicted by the numerical simulation resembles the data observed by Coutte et al. [20] in C3H mice. (C) The predicted dynamics of immune effectors is out-of-phase with the Bb abundances, resembling classical predator-prey dynamics.

3.3.5. VlsE diversity significantly influences Bb levels

Pairwise sequence similarities among VlsE variants ($S(i,j)$) generated from the natural *vls* cassettes are significantly lower than those among VlsE variants generated by the perturbed *vls* unexpressed cassettes (Figure 3-5). Pairwise sequence similarities among variants generated by both the natural and the perturbed *vls* unexpressed cassettes are negatively correlated with the number of recombination events that separate the variants (Dr) (Figure 3-5). Because VlsE contains only a finite number of variable sites, the rate of decrease in $S(i,j)$ diminishes over time (Figure 3-5).

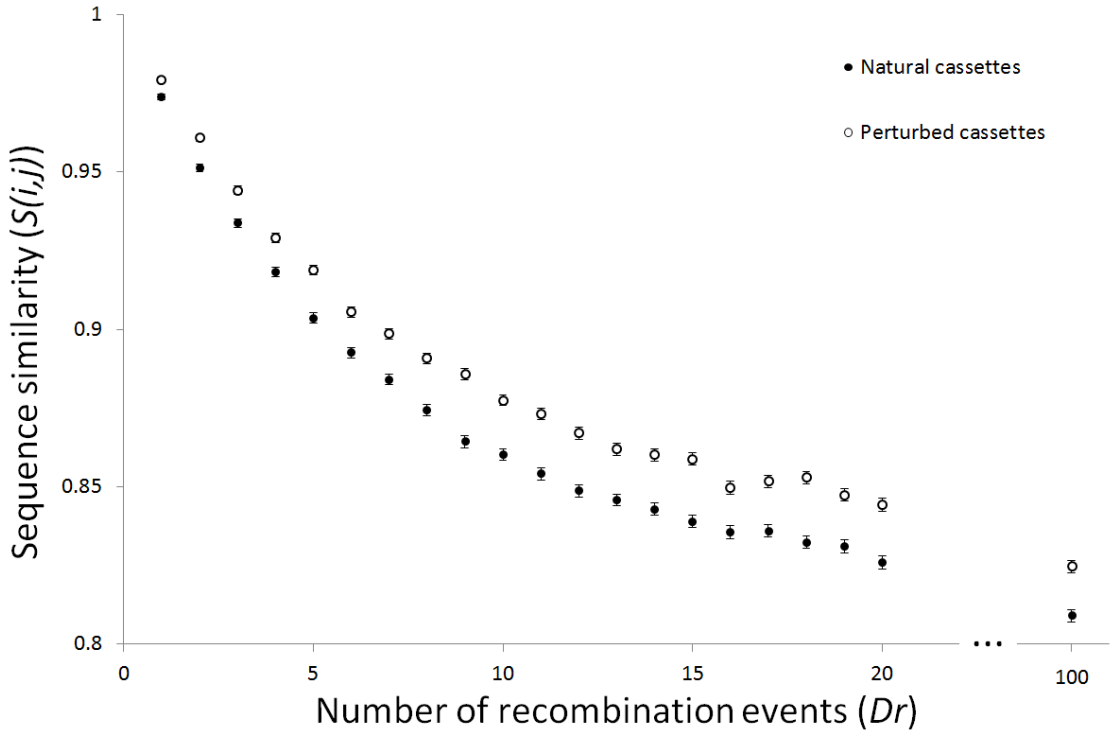


Figure 3-5. Pairwise sequence similarity (*S(i,j)*) is greater among VlsE variants generated by perturbed than natural *vls* unexpressed cassettes. Filled circle, VlsE variants generated by the natural unexpressed cassettes; open circle, VlsE variants generated by the perturbed unexpressed cassettes.

In both the antibody immunodominance and the immune effector dynamics models, higher diversity among VlsE variants results in higher Bb density when immune suppression impacts Bb population dynamics (Figure 3-6). When Bb population abundances are strongly suppressed by antibodies, Bb with the natural cassettes were up to 26% more abundant than Bb with the perturbed cassettes in the model that included immune effector dynamics. In the model that included antibody immunodominance, Bb

with the natural cassettes were up to 200% more abundant than Bb with the perturbed cassettes.

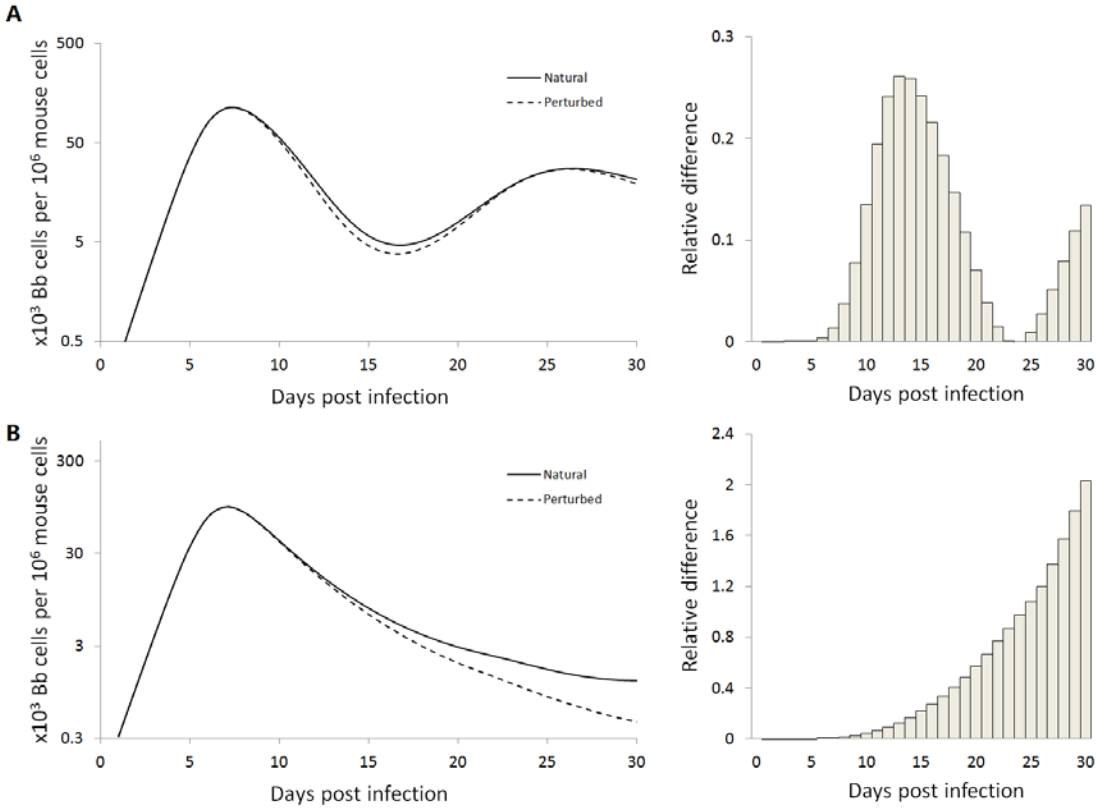


Figure 3-6. Greater diversity among VlsE variants results in substantially higher densities of Bb within hosts. The population dynamics of Bb with natural *vls* cassettes, and thus greater within-host VlsE diversity, were substantially different that the dynamics of Bb with perturbed cassettes in simulations that included immune effector dynamics (A) and those that included antibody immunodominance (B). The left panels show the dynamics of Bb populations with natural (solid lines) and perturbed (dotted line) *vls* cassettes

during the initial stages of infection. The right panels show the relative difference in abundance of Bb populations with natural and perturbed *vls* unexpressed cassettes.

3.4 Discussion

A simple model of the Bb population dynamics in the face of antibody pressure cannot account for Bb persistence within hosts following the rapid, antibody-mediated decline in population sizes observed in experimental infections, regardless of the parameters used in the model. These results suggest that mechanisms in addition to VlsE antigenic variation may be necessary to explain the persistent within-host survival of Bb or that the efficacy of the adaptive immune response wanes quickly in the early stages of infection. Alternatively, two well-characterized mechanisms - antibody immunodominance or immune effector cell dynamics - can account for the Bb population dynamics observed in experimental infections [25, 62].

Including either immunodominance or fluctuation in immune effector cell abundances into the Bb population dynamic model can replicate the Bb population dynamics observed in experimental infections. However, the antibody specificity values (x) necessary to recreate the observed dynamics vary between the models. Antibody specificity can be relatively mild in the model including immunodominance while the model with immune effector cell dynamics required little cross-reactivity to recreate the Bb dynamics. Further, the models result in considerably different predictions of population dynamics in Bb as well as differences in the effect of VlsE variability on within-host Bb abundance. That is, models with antibody immunodominance predicts Bb

abundance stabilizes after the early stages of infection, the level of which is largely affected by the diversity at VlsE (Figure 3-6), while models including immune effector cell dynamics predict continuous fluctuations in Bb abundance with only small effects of VlsE diversity occurring only when population sizes are small (Figure 3-6). The differences in the predictions of the models can be tested with experimental data to estimate the likelihoods of processes occurring within hosts.

The effect of diversity among VlsE variants on the abundance of Bb within-hosts is most apparent at low population densities. Low population density is a sign of strong immune suppression, under which the ability to evade immune targeting is vital. Besides, at low densities, VlsE sequences experience fewer recombination events per unit time, such that high diversity- generating abilities of the unexpressed cassettes are vital for the recombinants to be different enough from the parental VlsE to escape the immune response. It is surprising, however, that small differences in sequence diversity have a large effect on Bb dynamics. For instance, pairwise sequence difference among VlsE variants produced by the two sets of unexpressed cassettes explored here differ by less than 1% per recombination event, but the resulting difference in Bb abundance differed by more than 200% for models that included immunodominance and as much as 26% for models that included immune effector dynamics. The differences in Bb abundance likely have effects on key evolutionary and pathological traits of Bb such as the probability of dissemination to internal organs and the probability of transmission to feeding ticks.

The mechanisms resulting in the rapid decrease in Bb abundance followed by within-host persistence also differ among the models. In the models that include immunodominance, a single antibody type can rapidly eliminate the parental VlsE after the specific antibody

response is initiated. This causes a rapid decrease in the Bb population that is composed primarily of the parental VlsE. However, the lag time needed to initiate the immune response allows a sufficient amount of recombination to occur in a fraction of the population that is able to evade the antibody targeting the parental VlsE variant. Subsequent antibody responses, coupled with the limit amount of recombination occurring in the small Bb population, can control the Bb abundances within the host. The models that include immune effector dynamics permit persistence due to predator-prey like dynamics. The recruitment rate of immune effector cells is a function of the abundance of Bb leading to elevated killing when Bb level is high and reduced killing when Bb is rare [68]. The immune effector dynamics deviate from the pure Lotka-Voltera model in that there is an intrinsic rate of recruitment of the immune effector cells, which can eliminate Bb at large values.

The models described in this chapter suggest testable hypotheses that may account for the population dynamics of Bb within hosts. However, the quantitative predictions are likely affected by several models assumptions. For example, the models were parameterized to fit a sparse time series and thus likely suffer from over-fitting. Further, it is likely that other unexplored mechanisms could also replicate the observed data assuming reasonable parameters. Nevertheless, experimental data can differentiate the likelihoods of the mechanisms proposed in each of the explored models given their divergences in parameter values as well as the predictions of the relative effect of VlsE diversity on Bb abundance.

Chapter 4. Correlation between antigenicity and variability of amino acid residues in VlsE

4.1. Introduction

The continual arms race between host immune responses and antigenic variation of microbial parasites results in strong natural selection shaping the molecular interactions at the antigen-antibody interface. For example, in the Lyme bacteria *Borrelia burgdorferi* (Bb), continuous alteration of the surface antigen, VlsE, is required for the spirochete to escape antibody recognition and establish long-term infections in vertebrate hosts [18, 21-24, 38]. Novel variants of VlsE are generated throughout vertebrate infections by unidirectional recombination of a random segment from one of several unexpressed, paralogous *vls* cassettes into the *vlsE* expression site [18, 19]. Variations among the unexpressed *vls* cassettes are concentrated in six regions that correspond to the six antibody-accessible, surface-loop structures of the VlsE protein [18, 69]. The probability of evading detection by antibodies targeting previously expressed VlsE is expected to correlate with the sequence variability of VlsE. As the sequence variability of VlsE is correlated with the variability among the unexpressed *vls* cassettes, natural selection should favor maximal among-cassette diversity in order to maximize alterations in the VlsE sequence [27]. However, only a subset of the sites within the loop structures are variable [18, 61]. The absence of variability at the remaining sites in the antigenically-important loop structures suggests that these sites are either rarely targeted by antibodies

or are constrained to be invariant by other selective forces such as maintenance of protein structure or function.

The degree to which an antigen must change to escape binding by an antibody depends largely on the physicochemical properties of the amino acids that are recognized by the antibody. Antigen-antibody contact regions consist of an average of 18 amino acids in the antigen [70], each of which has a quantitatively different effect on antibody-antigen reactivity. Variation among amino acids in their effect on binding affinity has been most strongly correlated with the solvent accessible surface areas of amino acid residues [71, 72]. Other physicochemical or statistical properties of amino acid residues including hydrophilicity [73, 74] and statistical epitope propensity [75] are also correlated with reactivity between residues and antibodies. Thus, the amino acid sites in VlsE that have large solvent accessible surface areas, are hydrophilic, and have high epitope propensity scores are likely to have the largest effect on immune evasion when mutated, and are thus the sites that are expected to be variable among *vls* cassettes.

In this study, we investigated the structural and physicochemical properties of amino acids in the antigenically-important surface loops of VlsE to assess if variable amino acid sites have a larger effect on antibody reactivity. We addressed this hypothesis through computational structural analyses of the VlsE antigen and by experimentally testing the effect of altering variable or invariant sites on antibody reactivity.

4.2. Materials and Methods

4.2.1. Correlations between amino acid variability and metrics of antibody reactivity

The published VlsE crystal structure from *Bb* clone 5A3 ([PDB:1L8W], chain B) [69] was used to computationally investigate the hypothesis that amino acid residues that are more reactive to antibodies are more likely to be altered during antigenic variation. Variable and invariant amino acid sites in the surface exposed loops of VlsE were identified as previously described [18, 61]. We used three metrics to assess reactivity to antibodies; 1. solvent accessible surface area, 2. hydrophilicity, and 3. statistical epitope propensity. Solvent accessible surface areas of each amino acid residue in the six surface loops in VlsE were calculated based on the published VlsE crystal structure using the POPS algorithms [76]. The hydrophilicity of each amino acid was estimated using the Parker hydrophilicity scale [73]. Statistical epitope propensities (the log-odds ratio of the frequencies of a type of amino acid appearing in known epitopes) were estimated as previously described [75]. The solvent accessibility, hydrophilicity, and epitope propensities of invariant VlsE amino acids – those that are never altered by recombination with any *vls* cassette – were compared to the solvent accessibility, hydrophilicity, and epitope propensities of variable amino acids.

4.2.2. Antibody binding sites predicted by computational docking analysis

Antibody reactivity of amino acid residues in VlsE was also computationally predicted by conducting protein-protein docking analyses. 394 antibody light or heavy chains whose structures have been solved were extracted from the Protein DataBase as described by Kunik et al (structural consensus among antibodies). Automated protein docking was performed between VlsE and each of the 394 antibodies light or heavy chains using the ClusPro server under *Antibody Mode*. For each VlsE-antibody pair, the docking model

with the highest score was used to identify the antibody-reactive amino acid residues in VlsE. In each model, residues in VlsE that have any atoms that are less than 4A away from any atoms in the antibody were identified as antibody-reactive. The number of times a VlsE residue is identified as antibody-reactive in the 394 VlsE-antibody docking models was used to infer the antibody reactivity of the residue.

4.2.3. Experimental effect of altering amino acid residues on antibody reactivity

The effect of changing either naturally variable or invariant amino acid sites in the surface exposed loops of VlsE on antibody reactivity was investigated using recombinant wild-type and mutant VlsE proteins. The DNA sequence of the wild-type *vlsE* was amplified from *Bb* (clone 5A3) using primers 5A3F (CGGGGATCCCAGCCAAGTTGCTGATAAGGACGACCC) and 5A3R (CGGAAGCTTCAATCATGAGGGCATAGTCGTGTCCATACA). The ends of the amplicon were digested with BamHI and HindIII prior to ligation into expression vector pET45(b) (Novagen). The wild-type VlsE protein was expressed in *E. coli* BL21 (DE3), purified using Ni-NTA Superflow Resin (Qiagen), and dialyzed overnight in 100mM NaCl, 50mM Tris-8 and 0.1% Tween-20. Polyclonal antibodies against the wild-type VlsE protein were generated in two rabbits at Anaspec, Inc. (Fremont, CA).

VlsE mutant proteins were generated by alanine scanning mutagenesis [77]. A single amino acid was changed to an alanine using the Quikchange mutagenesis kit (Agilent Technologies) in each of 28 VlsE mutant proteins, 18 of which are found in loop 1 and 10 in loop 4 (Figure 4-1). Loop 1 and Loop 4 were chosen because approximately half of the

amino acid sites are variable in each of the two loops. VlsE mutant proteins were expressed and purified as described for the wild-type VlsE protein.

LOOP1

Name	Sequence
WT	SSGTAAIGEVVADADAAKVADKASVK
L1M140	ASGTAAIGEVVADADAAKVADKASVK
L1M141	SAGTAAIGEVVADADAAKVADKASVK
L1M142	SSATAAIGEVVADADAAKVADKASVK
L1M143	SSGAAAIGEVVADADAAKVADKASVK
L1M146	SSGTAAAGEVVADADAAKVADKASVK
L1M147	SSGTAAIAEVVADADAAKVADKASVK
L1M148	SSGTAAIGAVVADADAAKVADKASVK
L1M149	SSGTAAIGEAVADADAAKVADKASVK
L1M150	SSGTAAIGEVAADADAAKVADKASVK
L1M152	SSGTAAIGEVVAAADAAKVADKASVK
L1M154	SSGTAAIGEVVADAAAAKVADKASVK
L1M157	SSGTAAIGEVVADADAAAVADKASVK
L1M158	SSGTAAIGEVVADADAAKAADKASVK
L1M160	SSGTAAIGEVVADADAAKVAAKASVK
L1M161	SSGTAAIGEVVADADAAKVADAASVK
L1M163	SSGTAAIGEVVADADAAKVADKAAVK
L1M164	SSGTAAIGEVVADADAAKVADKASAK
L1M165	SSGTAAIGEVVADADAAKVADKASVA

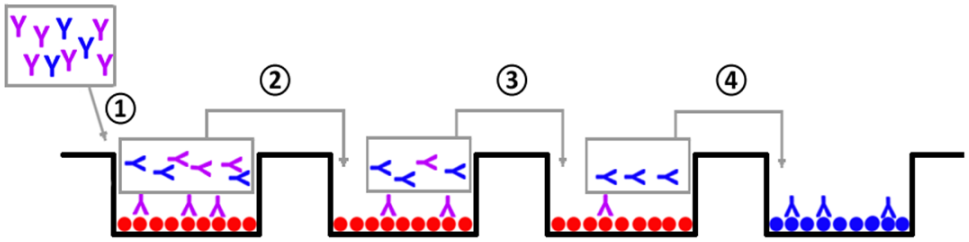
LOOP4

Name	Sequence
WT	EQDGKKPEEAK
L4M243	AQDGKKPEEAK
L4M244	EADGKKPEEAK
L4M245	EQAGKKPEEAK
L4M246	EQDAKKPEEAK
L4M247	EQDGAKPEEAK
L4M248	EQDGKAPEEAK
L4M249	EQDGKKAEEAK
L4M250	EQDGKKPAEAK
L4M251	EQDGKKPEAAK
L4M253	EQDGKKPEEAA

Figure 4-1. The 28 VlsE mutant proteins used in this study. The amino acids replaced by alanine in each of the mutant VlsE are shown in red. Sites coding for alanine in the wild-type protein were not altered. The amino acids that are naturally variable in surface exposed Loops 1 and 4 are shaded in the wild-type (WT) amino acid sequences. Mutant VlsE proteins are named after the location of the mutation in the wild-type protein. For example, the mutation in mutant protein L1M140 is at the 140th amino acid residue which falls in loop 1.

ELISA was used to assess the effect of each mutation on antibody reactivity. However, background noise associated with cross-reactive antibodies that bind both the wild-type and mutant antigens in areas unaffected by the single amino acid substitution hindered measurements of the effect of the mutated residue. In this study, background reactivity was reduced by negative purification of the antisera (Figure 4-2). Negative purification was performed by coating wells of standard ELISA plates with the focal VlsE mutant (50μg/mL) diluted in PBS (4°C overnight), blocking with BSA diluent/blocking solution (KPL; 4°C overnight), and then incubating rabbit antisera in wells for 90 minutes at room temperature. The antibodies generated against the wild-type VlsE that bind to the focal mutant, and are thus cross-reactive, were removed by repeating this process six times, each time transferring the antisera into a new well coated with the same focal VlsE mutant and incubating for 90 minutes (Figure 4-2A). After six rounds of negative purification, only antibodies that bind to the wild-type but not to the mutant VlsE are expected to remain in the antisera (Figure 4-2B).

A



● Wild type VlsE
● VlsE mutant
Y Antibodies that do not cross-react with the VlsE mutant
Y Antibodies that cross-react with the VlsE mutant

① Incubate polyclonal antibodies raised against wild-type VlsE with a VlsE mutant to reduce the concentration of cross-reactive antibodies
② Incubate unadsorbed antibodies with fresh VlsE mutant to further reduce the amount of cross-reactive antibodies
③ Repeat step ② multiple times
④ The remaining unadsorbed antibodies are used in a standard ELISA with wild-type VlsE. The concentration of antibody remaining after repeated adsorbtion with mutant VlsE is proportional to the effect of the mutation on antibody reactivity.

B

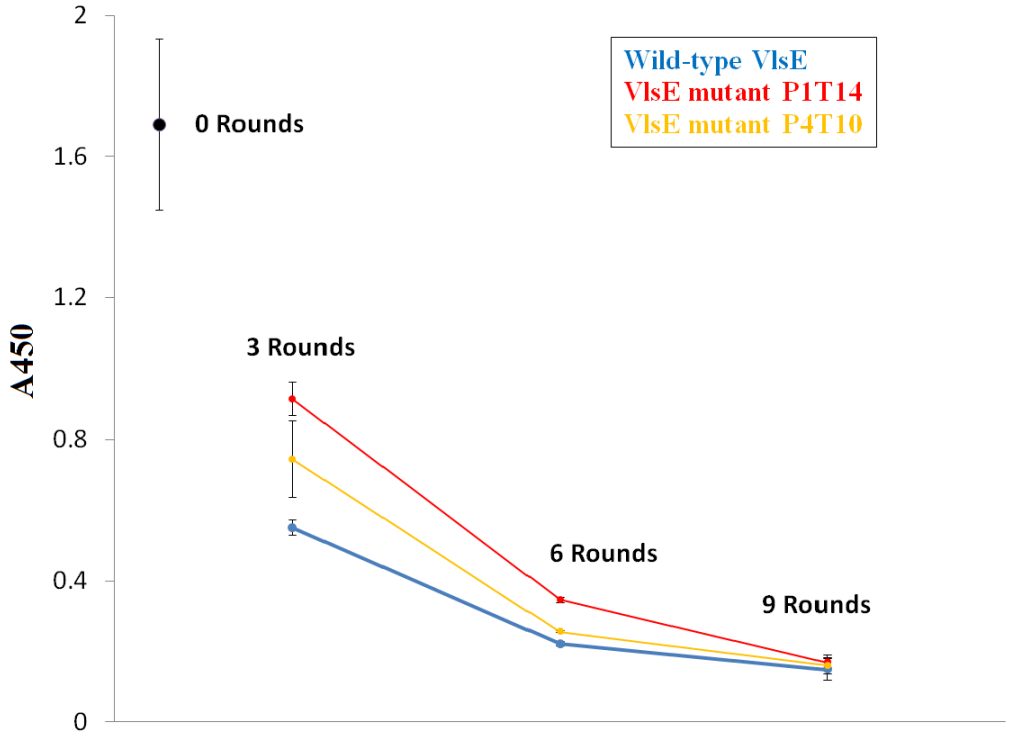


Figure 4-2. Negative purification of antisera for high-sensitivity ELISA by adsorption with a focal VlsE mutant. (A) Cross-reactive antibodies are removed by adsorption to the focal mutant antigen. The quantities of remaining antibodies, which are not cross-reactive, are measure by ELISA using the wild-type antigen. (B) Negative purification of antisera increases sensitivity in immuno-assays. Negative purification consistently reduced the background noise in immuno-assays resulting in high precision data. In the figure, antisera generated by rabbit II were progressively adsorbed with the wild-type or one of two VlsE mutants. “0 Rounds” shows the reactivity of the antisera to the wild-type VlsE without prior negative purification.

Post negative-purification antisera were used in a standard indirect ELISA using wild-type VlsE as the target antigen under standard procedures recommended by abcam, with a final wild-type VlsE concentration of 7.5μg/mL used to coat ELISA wells. Goat-anti-rabbit HRP was used as the secondary antibody and Turbo TMB was used as the chromogen. ELISA signals were measured by reading absorbance at 450nm (A450). The ELISA signals reflect the amount of antisera that did not bind to the focal VlsE mutant used for negative-purification, which correspond to the proportion of antibodies that do not bind to the mutant VlsE due to the mutation (Figure 4-2A). That is, the quantity of the antibodies remaining after negative-purification shows the effect of the amino acid substitution on antibody reactivity (Figure 4-2B). All absorbance readings were normalized by the absorbance reading from a control well on each ELISA plate to account for among-plate variation. The control well used antisera adsorbed with the wild-

type VlsE. Thus, mutations that do not affect antibody reactivity result in a normalized score (A450(mutant) / A450(wt)) equal to 1.

4.3. Results

4.3.1. Computationally-determined correlation between antibody reactivity and variability of amino acid residues in VlsE

Amino acid sites that are variable during antigenic variation in all surface loops have significantly larger solvent accessible surface areas, which is strongly associated with antibody reactivity, than the invariant amino acid sites (Figure 4-3A). Further, surface accessibility is greater at variable amino acid sites than at invariant sites that use the same amino acid residue (Figure 4-3B). Amino acid sites that are variable during antigenic variation also have significantly larger hydrophilicity and epitope propensity scores (Figure 4-3C,D), both of which are associated with greater targeting by antibodies. Unlike solvent accessible surface area, the hydrophilicity index and epitope propensity score are not conditional on the three-dimensional structure of the protein. Consistently, antibodies bind to variable amino acid sites significantly more often than the invariant amino acid sites, as is demonstrated by the antibody-antigen docking analyses (Figure 4-3E).

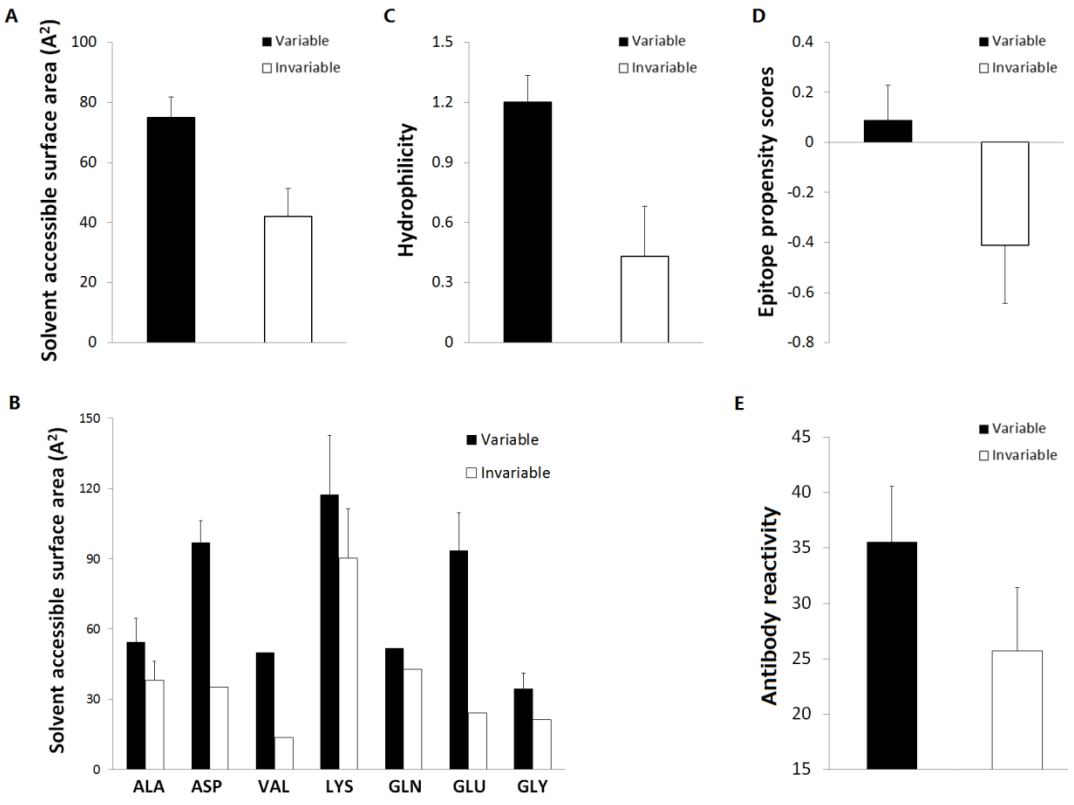


Figure 4-3. Metrics of antigenicity are significantly greater at variable amino acid sites than at invariant sites. (A) Solvent accessible surface areas of the variable amino acid sites are significantly greater than at the invariant sites in all surface loops. (B) Even among sites that use the same amino acid, the solvent accessible surface area is greater at variable sites than invariant sites. Only the seven types of amino acids that are found in both the variable and invariant sites are shown. Hydrophilicity (C) and statistical epitope propensity (D) are also significantly greater at variable sites than at the invariant sites in all surface loops. (E) Antibody docking analysis also suggests that variable amino acid residues have higher reactivity to antibodies.

4.3.2. Inconsistent experimental correlation between antibody reactivity and variability at amino acid sites in VlsE

Changes in antibody reactivity caused by experimentally altering variable or invariant amino acid residues in two surface exposed loop regions are summarized in Table 4-1. Although the amino acids that are altered during antigenic variation have greater solvent accessibilities, are more hydrophilic, and have greater epitope propensities than invariant sites, experimentally changing amino acids at variable sites to alanine does not result in a greater reduction in antibody reactivity than experimentally changing invariant sites (Figure 4-4). In fact, several invariant amino acids have large impacts on antibody reactivity when mutated (Table 4-1). Among variable sites, there is a statistically insignificant positive correlation between the changes in antibody reactivity due to a mutation and accessible surface area (Figure 4-5). Changes in antibody reactivity caused by each experimentally-introduced mutation were varied considerably among the antisera from different rabbits, thus resulting in inconsistent correlations between antibody reactivity and solvent accessibility (Figure 4-6).

Mutant	Solvent accessible surface area of the mutated residue (A^2)	A450(mutant) / A450(wt) Rabbit I	A450(mutant) / A450(wt) Rabbit II
L1M140	21.58	1.116±0.066	1.202±0.097
L1M141	55.93	1.200±0.096	1.222±0.087
L1M142	19.53	1.390±0.097	1.490±0.059
L1M143	114.28	1.818±0.094	1.967±0.055
L1M146	17.91	3.086±0.148	1.380±0.037
L1M147	22.84	1.929±0.093	2.055±0.046
L1M148	43.95	1.683±0.205	1.982±0.128
L1M149	24.52	1.342±0.026	1.513±0.121
L1M150	24.17	1.800±0.229	1.436±0.086
L1M152	97.37	1.822±0.128	1.362±0.077
L1M154	147.16	1.815±0.034	1.427±0.079
L1M157	105.54	1.066±0.029	1.240±0.073
L1M158	56.74	1.590±0.064	1.382±0.033
L1M160	35.15	3.208±0.155	1.502±0.047
L1M161	125.25	2.801±0.120	1.670±0.094
L1M163	11.86	1.204±0.025	1.362±0.062
L1M164	13.54	2.760±0.142	1.340±0.096
L1M165	62.07	1.476±0.126	1.806±0.180
L4M243	151.12	1.101±0.116	1.667±0.254
L4M244	42.81	1.034±0.128	1.559±0.216
L4M245	77.87	1.104±0.063	1.494±0.185
L4M246	21.14	1.208±0.042	1.570±0.097
L4M247	68.7	1.769±0.088	1.344±0.022
L4M248	43.6	1.315±0.048	1.657±0.153

L4M249	13.01	1.692±0.069	1.842±0.161
L4M250	36.63	1.268±0.052	1.396±0.071
L4M251	83.14	2.863±0.269	1.502±0.085
L4M253	76.93	1.373±0.147	1.415±0.079

Table 4-1. Antibody reactivity of mutated amino acid residues in 28 VlsE mutants.

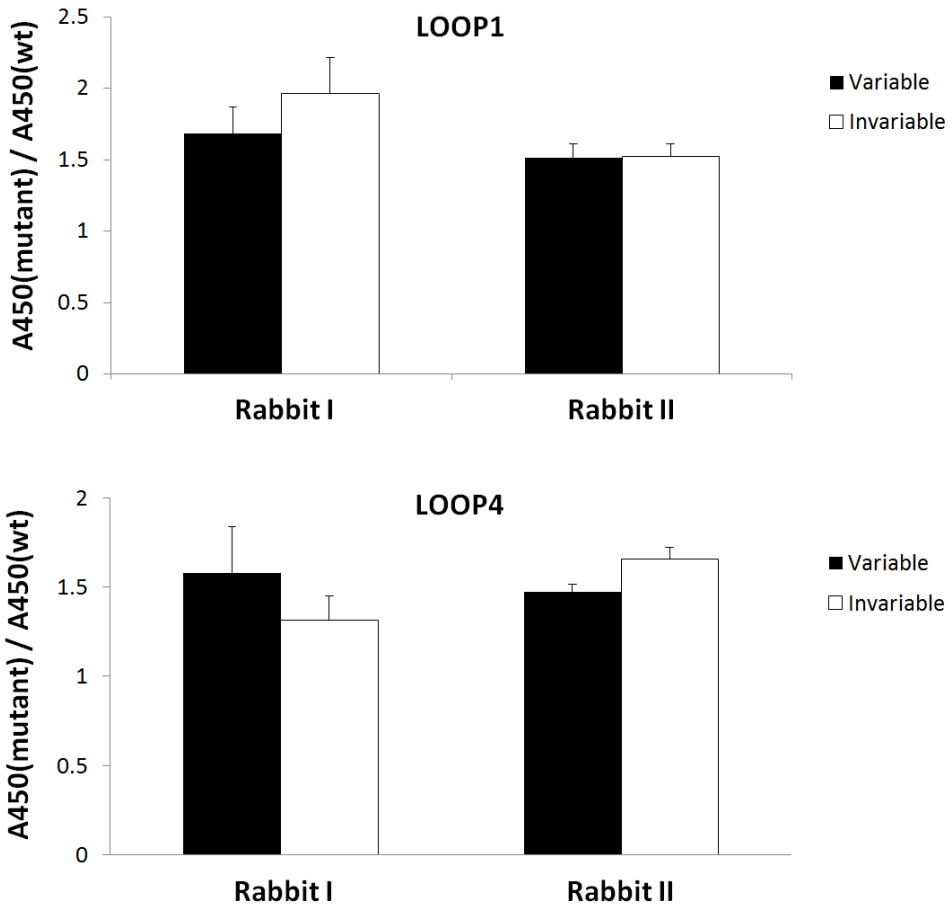


Figure 4-4. Changes in antibody reactivity caused by mutating variable or invariant amino acid residues. Mutations at variable amino acid sites do not provide a greater reduction in antibody reactivity than mutations at invariant sites. The expected decrease in antibody binding caused by mutations in variable sites was detectable only in the antibodies targeting loop 4 from rabbit I but not antibodies from rabbit II. Mutations in the invariant sites of loop 1 had a greater effect on antibody reactivity than mutations to variable sites in antibodies from rabbit I but not rabbit II.

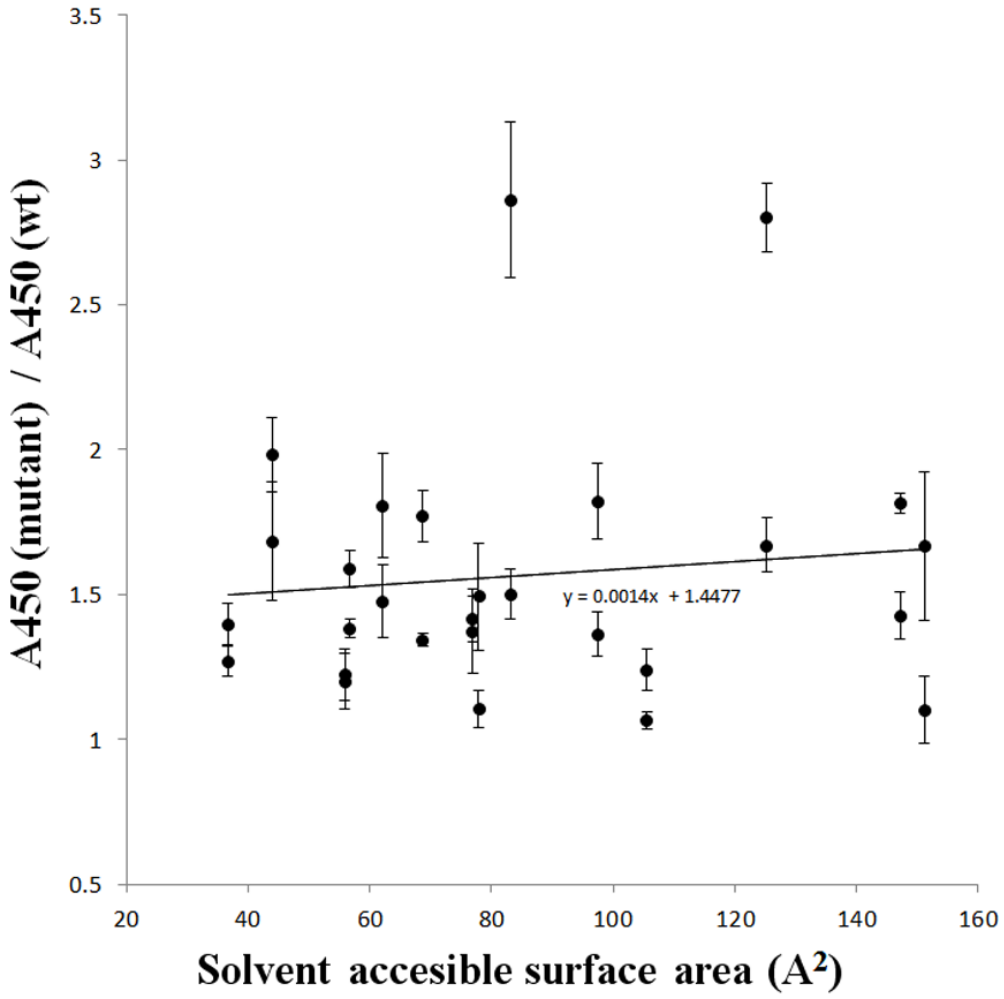


Figure 4-5. Correlation between antibody reactivity and the solvent accessible surface area of variable amino acid residues. Change in antibody reactivity caused by changing a single residue at a variable site is weakly and insignificantly correlated with the solvent accessible surface area of the amino acid residue.

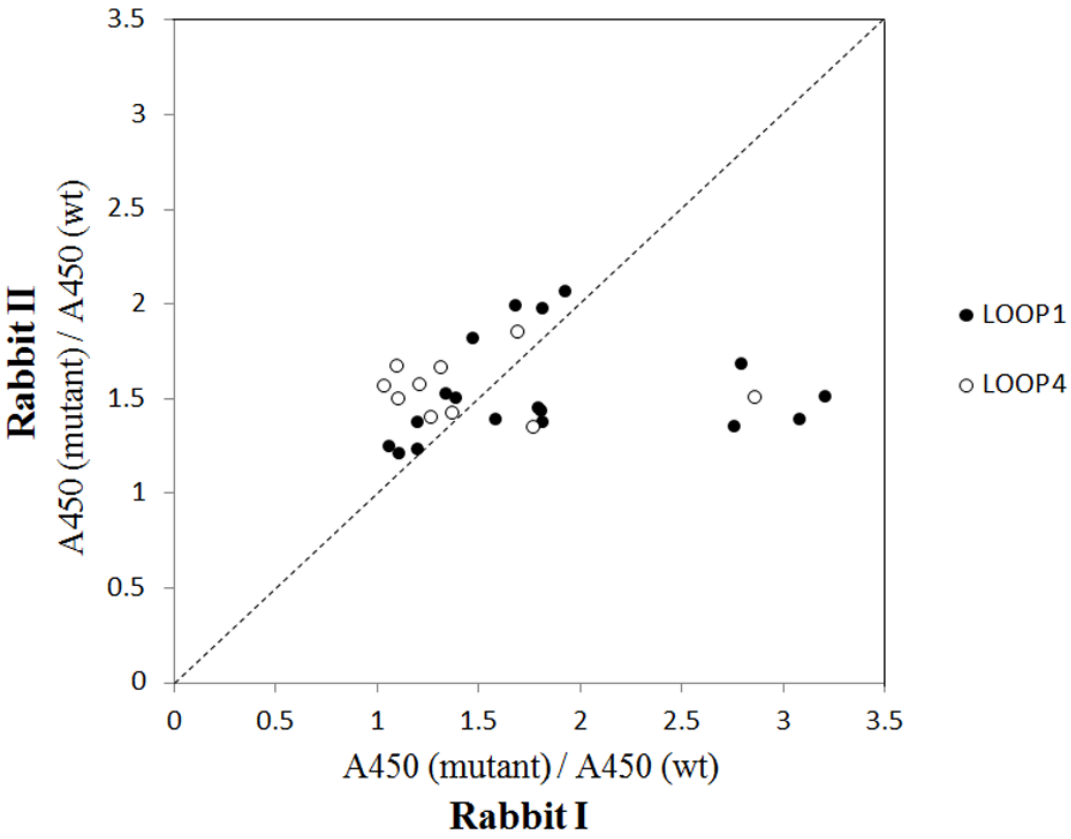


Figure 4-6. Immune reactivity to the same VlsE mutants was inconsistent among antisera raised in different rabbits. ELISA results were shown for each VlsE mutant reacting with negatively purified antisera generated from rabbit I and rabbit II.

4.3.3. Data reliability after negative purification

Negative purification successfully reduced background variation in ELISA and the results were repeatable. For example, antisera (from rabbit II) adsorbed using the wild-type VlsE

or two mutant VlsE variants (L1M160 or L4M253) exhibited decreasing variance in ELISA readings with increasing rounds of negative purification (Figure 4-2B). Adsorption with wild-type VlsE followed by ELISA using wild-type VlsE as the target antigen consistently resulted in the weakest signal, as expected. Adsorption with mutant L1M160 followed by ELISA using wild-type VlsE as the target antigen consistently resulted in a greater ELISA signal than adsorption with mutant L4M253, regardless of the number of rounds of negative purification. Interestingly, antibody reactivity to each VlsE mutant differed qualitatively among antisera from different rabbits (Figure 4-6). That is, a VlsE mutant that resulted in a large reduction in antibody reactivity in one rabbit may have had only a moderate impact on antibody reactivity using the antisera from a different rabbit.

4.4. Discussion

Antigenic variation systems in pathogens generate divergent antigens in order to evade recognition by host antibodies. Natural selection is expected to favor alterations at amino acid positions that effectively reduce antibody reactivity [27, 61]. Thus, naturally occurring changes in antigens are expected to occur at amino acid positions that are commonly and effectively targeted by antibodies [78, 79]. By computationally analyzing the structural features of VlsE, we found that amino acid positions that are most likely to be variable in a primary surface antigen of *Bb*, VlsE, are those with properties associated with high antibody reactivity including larger accessible surface areas, higher hydrophilicity, and higher epitope propensity scores (Figure 4-3). The correlation between accessible surface area and the antigenic importance of amino acids is well-

established and is one of the most important variables used in epitope prediction [71, 72]. Indeed, the correlation between variability and antigenicity is further supported by the results from antibody docking analysis. However, prediction from accessible surface area and antibody docking relies on the accuracy of the crystal structures and is thus subject to crystallization artifacts. Hydrophilicity and statistical epitope propensity scores, on the other hand, are crystal structures-independent and were also greater at variable sites than invariant sites. The agreement among the crystal structure-dependent and independent results is consistent with the hypothesis that variable amino acids are more likely to be targeted by antibodies than invariant amino acids.

Nonetheless, the support for the hypothesis that variable amino acids in VlsE have a greater effect on antibody reactivity is weak and inconsistent in the experimental data (Figure 4-4, 4-5). Despite limited surface exposure and no potential to vary in the face of antibody pressure, many invariant amino acids had as great an effect on antibody reactivity as did variable sites (Figure 4-4). For example, mutations at two invariable sites (L1M146 and L1M160) cause the largest reductions in antibody reactivity in rabbit I antisera among all mutants (Table 4-1). One explanation for the discrepancy in the computational and experimental results is that the computational results are artifacts. However, this explanation is not likely given the consistency among the results from four computational matrices. Another possible explanation is that altering naturally invariant amino acids, which have less surface exposure and are more hydrophobic, destabilizes protein structure [80] resulting in conformational changes that strongly affect antibody binding [81]. The effect of altering invariant amino acids on the conformational stability

and function of VlsE is likely to be highly detrimental and thus selected against [31], despite the potential advantages for antigenic escape. Future investigations into the structural and molecular changes caused by mutating variable and invariant sites may identify targets of natural selection that have organized the locations and composition of residues at sites that are variable among the *vls* cassettes.

The weak and inconsistent experimental patterns in antibody reactivity reductions may also be caused by the stochasticity in immune development among animals, the limited effects of alanine on antibody reactivity, or the limited effect of changing only one amino acid on antibody reactivity. Antisera from the different rabbits often had different reactivity to the same VlsE mutant (Figure 4-6). For example, one mutation in loop 1 (L1M160) resulted in the most dramatic decline in antibody reactivity among all mutations from rabbit I antisera but only moderate declines in reactivity from rabbit II antisera. Variation among individuals in epitope targeting is common (for example see [82]), suggesting that antisera from a large number of experimental animals may be necessary to deduce general trends.

The inconsistent empirical patterns may also be attributed to the use of alanine scanning mutagenesis. Alanine scanning is a conservative technique used to understand the functional effects of individual sites in a protein because alanine is likely to have the least effect on protein structure due to limited steric and charge effects [83]. Although conservative, alanine scanning mutagenesis did reveal an empirical relationship between antibody reactivity and accessible surface area among the variable amino acid sites. This pattern was, however, much less clear than the correlation between accessible surface

areas and variability demonstrated computationally. Experimentally replacing the focal amino acids with residues that are found in naturally occurring VlsE variants may reveal a clearer pattern consistent with the hypothesis that natural selection influences both the sites that are variable and the types of residues used at those sites [61]. However, invariant sites do not have alternative residues in naturally occurring VlsE variants. Thus, only conservative techniques such as alanine scanning allow for comparisons of the effects of altering variable and invariant sites.

Investigating changes in antibody reactivity by altering amino acids individually may have resulted in empirical patterns that were less consistent than computational predictions due to the limited impact of individual amino acids. In natural *Bb* infections, an average of 5 to 89 nucleotides are altered during each recombination event [20], suggesting that changing more than one amino acid may be needed to detect consequential reductions in antibody reactivity. Further, the combination of altered amino acids may have an epistatic effect on antibody reactivity that would not be detected with the alanine scanning approach. Although changes in antibody reactivity caused by one amino acid substitution may be difficult to detect empirically, removal of the background noise by negative purification of antisera enabled precise estimates of changes in antibody reactivity. Negative purification is a useful tool to increase sensitivity of ELISA and other antibody-based immunoassays that can be coupled with semi-rational mutagenesis methods for high-throughput mapping of polyclonal epitopes.

Chapter 5. Discussion

The ability of microbes to maintain high diversity is important for survival within adverse and rapidly changing environments [84, 85]. As an example, the *vls* antigenic variation system in the Lyme disease bacteria *Borrelia burgdorferi* (Bb) continuously generates diversity at a site coding for an immunodominant surface antigen, VlsE, facilitating evasion from host antibody recognition and is required for establishing long term infection [18, 22-24, 38, 39]. The *vls* antigenic variation system is not only an essential component in the pathogenesis of Bb but also a typical molecular mechanism bacteria use to cope with rapidly changing environments. Recognizing the clinical and evolutionary importance of the system, we studied the forces of natural selection that drive the evolution of the *vls* antigenic variation system in the work presented in Chapter 2, Chapter 3 and Chapter 4 of this thesis. Below we discuss the major findings, broad significance, applications and limitations of the studies presented in each chapter.

In Chapter 2 we studied selection for sequence-level diversity among VlsE variants and translational optimality of VlsE variants. The recombination events between *vlsE* and the *vls* unexpressed cassettes were modeled according to experimentally derived data. By modeling and simulating antigenic variation, we described the array of VlsE variants that can be generated by the natural *vls* antigenic variation system and compared those to the VlsE variants generated from cassettes that were computationally perturbed. This computational method can be applied to identify evidence of natural selection in similar systems that alter antigenic sequences using recombination-based mechanisms, such as

Plasmodium falciparum [86] and *Neisseria meningitides* [87]. The predictions made by the computational method can be empirically tested by competitive, mixed infection of vertebrate hosts using the natural strain of Bb and genetically modified Bb strains that carry perturbed unexpressed cassettes.

In this Chapter, we showed that the nucleotide identities at and positions of the variable sites within the variable regions of the *vls* antigenic variation system are organized to promote diversity among VlsE variants, consistent with the hypothesis that higher diversity should be favorable within an adversely changing environment. In addition, we found that translational ability and efficiency of the VlsE variants may constrain the selection for higher sequence diversity. Although these findings are statistically significant, in terms of effect size, the difference in VlsE sequence diversity and codon optimality between natural and perturbed antigenic variation systems is small. Therefore, Chapter 2 provides statistical evidence that both diversifying and stabilizing selection has acted on the *vls* antigenic variation system, but the fitness impact of such selections is unclear. It is also unclear how selection affects biological traits of Bb, such as the dynamics of Bb within vertebrate hosts.

In Chapter 3, we investigated the biological significance of the diversity among VlsE variants. Specifically, we studied how strongly the sequence diversity among VlsE variants can affect Bb dynamics within vertebrate hosts. First, we mathematically modeled Bb dynamics within vertebrate hosts and then asked if the dynamic models, accounting for different microbiological and immunological mechanisms, can replicate experimental data. We showed that to reproduce the experimental time series of *Bb*

abundance observed in laboratory mice, the dynamic models need to include mechanisms supporting immune evasion or fluctuating immune pressure, in addition to antigenic variation in VlsE. We reproduced the experimental data by including two mechanisms – antibody immunodominance and fluctuation of immune effector cells – independently into the dynamic model. The two mechanisms are well-characterized and commonly happen during early stages of bacterial infections [25, 68]. It is important to note that the dynamic models are not suitable for making quantitative predictions because, first, the two mechanisms are not mutually exclusive; second, the experimental time series may also be reproduced under other mechanisms such as dissemination of the bacteria to different infection sites. However, the two dynamics models make qualitatively different predictions of the Bb dynamics when accounting for the two mechanisms: The model that includes antibody immunodominance predicts that the Bb abundance will decrease to a stable level following the initial expansion of antibodies, while the model including immune effector cells predicts that the Bb abundance will fluctuate continuously. Such difference can be tested experimentally to compare the likelihood of the two mechanisms. Finally, we used the dynamics model to explain the biological significance of the diversifying selection among VlsE variants identified in Chapter 2. We demonstrated that a small difference in VlsE diversity can translate into significantly different Bb level when immune clearance is efficient, suggesting that a small difference in genetic diversity could result in substantial variation in a relevant trait value.

In Chapter 4, we studied natural selection that acts on structure-level properties of VlsE. Specifically, we tested the hypothesis that the variable amino acid residues in VlsE are

more likely targeted by antibodies than the invariable residues. Variability of amino acid residues in VlsE was positively correlated to four different matrices that predict antibody reactivity of the residues, consistently supporting the hypothesis that variable amino acid residues in VlsE are more likely targeted by antibodies than the invariable residues. However, this hypothesis was not supported by experimental epitope mapping data: Mutating a naturally variable residue did not always result in a greater reduction in antibody reactivity than mutating a naturally invariable residue. There are several mechanisms that could explain such discrepancies. First, the computationally determined correlations between variability and antigenicity may be artifacts. Second, the biological signal could be obscured by noise from alanine-scanning mutagenesis and diversity between the immune systems of the rabbits used to raise the antibodies. Third, the discrepancy could be due to purifying selection acting on the invariable residues to prevent structural perturbation of the VlsE protein. In particular, the third potential mechanism described a set of conflicting selective forces that promote antigenic diversity while maintaining the structure of the protein. The hypothesis is biologically intriguing and can be tested by analyzing the structural effect of mutating naturally variable or invariable amino acid residues in VlsE.

Consistent with the expectations, our findings suggest that rapidly changing environments select for higher diversity at the locus encoding for molecules at the interface of the microbe and the environment, such as the variable protein antigen VlsE [12, 84, 85]. Each variant of the molecule permits survival within only a subset of the temporally variable environmental states. For example, each VlsE variant can evade

certain types of antibodies but is effectively targeted by other antibodies. Therefore, microbial populations can track and survive multiple environmental states by maintaining high diversity at such loci [85, 88]. It is worth noting that diversity can be measured in different traits: from genetic diversity (diversity on sequence level) to diversity at different levels of organization such as protein structure, antibody reactivity and microbial fitness. The thesis demonstrates that, although the traits at these levels of organization are derived from the same genetic basis, their patterns of diversity may differ. For instance, in Chapter 3 we showed that small sequence diversity among VlsE variants can translate into substantially larger difference in Bb abundance; in Chapter 4 we showed that mutating a single amino acid residue at different positions in different VlsE variants, resulting in the same degree of sequence difference between the mutants and the wild-type VlsE, can lead to substantially different interactions with the same polyclonal antibodies. The different patterns of diversity may result from the fact that traits are connected nonlinearly by biological processes such as gene expression, protein folding and localization, antibody-antigen interaction and immune selection. Moreover, diversity in different traits is likely constrained by various types of stabilizing or purifying selections. For instance, this thesis suggests that the diversity of VlsE variants at the DNA-sequence level is likely constrained by translational efficiency of the VlsE variants, and that diversity at protein structure level might be constrained by natural selection for the structural stability of the protein.

BIBLIOGRAPHY

1. Lynch M, Lande, R. Evolution and extinction in response to environmental change. In: *Biotic interactions and global change*. Edited by Kareiva P, Kingsolver, J., Huey, R. Sunderland, MA: Sinauer; 1993. pp. 234-250.
2. Burger R, Lynch M. Evolution and Extinction in a Changing Environment - a Quantitative-Genetic Analysis. *Evolution* 1995,**49**:151-163.
3. Thomas CD, Cameron A, Green RE, Bakkenes M, Beaumont LJ, Collingham YC*, et al.* Extinction risk from climate change. *Nature* 2004,**427**:145-148.
4. Lopez-Maury L, Marguerat S, Bahler J. Tuning gene expression to changing environments: from rapid responses to evolutionary adaptation. *Nat Rev Genet* 2008,**9**:583-593.
5. Macallan DC, Wallace DL, Zhang Y, Ghattas H, Asquith B, de Lara C*, et al.* B-cell kinetics in humans: rapid turnover of peripheral blood memory cells. *Blood* 2005,**105**:3633-3640.
6. Deitsch KW, Moxon ER, Wellems TE. Shared themes of antigenic variation and virulence in bacterial, protozoal, and fungal infections. *Microbiol Mol Biol Rev* 1997,**61**:281-293.
7. Moxon ER, Rainey PB, Nowak MA, Lenski RE. Adaptive evolution of highly mutable loci in pathogenic bacteria. *Curr Biol* 1994,**4**:24-33.
8. Brunham RC, Plummer FA, Stephens RS. Bacterial antigenic variation, host immune response, and pathogen-host coevolution. *Infect Immun* 1993,**61**:2273-2276.
9. Deitsch KW, Lukehart SA, Stringer JR. Common strategies for antigenic variation by bacterial, fungal and protozoan pathogens. *Nat Rev Microbiol* 2009,**7**:493-503.
10. van der Woude MW, Baumler AJ. Phase and antigenic variation in bacteria. *Clin Microbiol Rev* 2004,**17**:581-611, table of contents.

11. Moxon R, Bayliss C, Hood D. Bacterial contingency loci: the role of simple sequence DNA repeats in bacterial adaptation. *Annu Rev Genet* 2006,**40**:307-333.

12. Sniegowski PD, Murphy HA. Evolvability. *Curr Biol* 2006,**16**:R831-834.

13. Beaumont HJ, Gallie J, Kost C, Ferguson GC, Rainey PB. Experimental evolution of bet hedging. *Nature* 2009,**462**:90-93.

14. Robertson BD, Meyer TF. Genetic variation in pathogenic bacteria. *Trends Genet* 1992,**8**:422-427.

15. Gally DL, Bogan JA, Eisenstein BI, Blomfield IC. Environmental regulation of the fim switch controlling type 1 fimbrial phase variation in Escherichia coli K-12: effects of temperature and media. *J Bacteriol* 1993,**175**:6186-6193.

16. Henderson IR, Owen P, Nataro JP. Molecular switches--the ON and OFF of bacterial phase variation. *Mol Microbiol* 1999,**33**:919-932.

17. Barbour AG. Antigenic variation of a relapsing fever Borrelia species. *Annu Rev Microbiol* 1990,**44**:155-171.

18. Zhang JR, Hardham JM, Barbour AG, Norris SJ. Antigenic variation in Lyme disease borreliae by promiscuous recombination of VMP-like sequence cassettes. *Cell* 1997,**89**:275-285.

19. Zhang JR, Norris SJ. Genetic variation of the Borrelia burgdorferi gene vlsE involves cassette-specific, segmental gene conversion. *Infect Immun* 1998,**66**:3698-3704.

20. Coutte L, Botkin DJ, Gao L, Norris SJ. Detailed analysis of sequence changes occurring during vlsE antigenic variation in the mouse model of Borrelia burgdorferi infection. *PLoS Pathog* 2009,**5**:e1000293.

21. Bankhead T, Chaconas G. The role of VlsE antigenic variation in the Lyme disease spirochete: persistence through a mechanism that differs from other pathogens. *Mol Microbiol* 2007,**65**:1547-1558.

22. Labandeira-Rey M, Skare JT. Decreased infectivity in Borrelia burgdorferi strain B31 is associated with loss of linear plasmid 25 or 28-1. *Infect Immun* 2001,**69**:446-455.

23. Purser JE, Norris SJ. Correlation between plasmid content and infectivity in Borrelia burgdorferi. *Proc Natl Acad Sci U S A* 2000,**97**:13865-13870.

24. Rogovskyy AS, Bankhead T. Variable VlsE is critical for host reinfection by the Lyme disease spirochete. *PLoS One* 2013,**8**:e61226.

25. Frank SA. In: *Immunology and Evolution of Infectious Disease*. Princeton (NJ); 2002.

26. Borst P, Greaves DR. Programmed gene rearrangements altering gene expression. *Science* 1987,**235**:658-667.

27. Graves CJ, Ros VI, Stevenson B, Sniegowski PD, Brisson D. Natural selection promotes antigenic evolvability. *PLoS Pathog* 2013,**9**:e1003766.

28. Philipp MT, Bowers LC, Fawcett PT, Jacobs MB, Liang FT, Marques AR*, et al.* Antibody response to IR6, a conserved immunodominant region of the VlsE lipoprotein, wanes rapidly after antibiotic treatment of Borrelia burgdorferi infection in experimental animals and in humans. *J Infect Dis* 2001,**184**:870-878.

29. Crother TR, Champion CI, Whitelegge JP, Aguilera R, Wu XY, Blanco DR*, et al.* Temporal analysis of the antigenic composition of Borrelia burgdorferi during infection in rabbit skin. *Infect Immun* 2004,**72**:5063-5072.

30. Liang FT, Yan J, Mbow ML, Sviat SL, Gilmore RD, Mamula M*, et al.* Borrelia burgdorferi changes its surface antigenic expression in response to host immune responses. *Infect Immun* 2004,**72**:5759-5767.

31. Parsell DA, Sauer RT. The structural stability of a protein is an important determinant of its proteolytic susceptibility in Escherichia coli. *J Biol Chem* 1989,**264**:7590-7595.

32. Combes C. Fitness of parasites: pathology and selection. *Int J Parasitol* 1997,**27**:1-10.

33. Schmid-Hempel P. Immune defence, parasite evasion strategies and their relevance for 'macroscopic phenomena' such as virulence. *Philos Trans R Soc Lond B Biol Sci* 2009,**364**:85-98.

34. Hagblom P, Segal E, Billyard E, So M. Intragenic recombination leads to pilus antigenic variation in Neisseria gonorrhoeae. *Nature* 1985,**315**:156-158.

35. Noormohammadi AH, Markham PF, Kanci A, Whithear KG, Browning GF. A novel mechanism for control of antigenic variation in the haemagglutinin gene family of mycoplasma synoviae. *Mol Microbiol* 2000,**35**:911-923.

36. Lipsitch M, O'Hagan JJ. Patterns of antigenic diversity and the mechanisms that maintain them. *J R Soc Interface* 2007,**4**:787-802.

37. Haydon DT, Woolhouse ME. Immune avoidance strategies in RNA viruses: fitness continuums arising from trade-offs between immunogenicity and antigenic variability. *J Theor Biol* 1998,**193**:601-612.

38. McDowell JV, Sung SY, Hu LT, Marconi RT. Evidence that the variable regions of the central domain of VlsE are antigenic during infection with lyme disease spirochetes. *Infect Immun* 2002,**70**:4196-4203.

39. Bykowski T, Babb K, von Lackum K, Riley SP, Norris SJ, Stevenson B. Transcriptional regulation of the Borrelia burgdorferi antigenically variable VlsE surface protein. *J Bacteriol* 2006,**188**:4879-4889.

40. Hershberg R, Petrov DA. Selection on codon bias. *Annu Rev Genet* 2008,**42**:287-299.

41. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H*, et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* 2007,**23**:2947-2948.

42. Schutzer SE, Fraser-Liggett CM, Casjens SR, Qiu WG, Dunn JJ, Mongodin EF*, et al.* Whole-genome sequences of thirteen isolates of Borrelia burgdorferi. *J Bacteriol* 2011,**193**:1018-1020.

43. Wang D, Botkin DJ, Norris SJ. Characterization of the vls antigenic variation loci of the Lyme disease spirochaetes Borrelia garinii Ip90 and Borrelia afzelii ACAI. *Mol Microbiol* 2003,**47**:1407-1417.

44. Lanave C, Preparata G, Saccone C, Serio G. A new method for calculating evolutionary substitution rates. *J Mol Evol* 1984,**20**:86-93.

45. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* 2010,**59**:307-321.

46. Benjamin DC, Berzofsky JA, East IJ, Gurd FR, Hannum C, Leach SJ*, et al.* The antigenic structure of proteins: a reappraisal. *Annu Rev Immunol* 1984,**2**:67-101.

47. Prager EM. The sequence-immunology correlation revisited: data for cetacean myoglobins and mammalian lysozymes. *J Mol Evol* 1993,**37**:408-416.

48. Ikemura T. Correlation between the abundance of Escherichia coli transfer RNAs and the occurrence of the respective codons in its protein genes. *J Mol Biol* 1981,**146**:1-21.

49. Kanaya S, Yamada Y, Kudo Y, Ikemura T. Studies of codon usage and tRNA genes of 18 unicellular organisms and quantification of Bacillus subtilis tRNAs: gene expression level and species-specific diversity of codon usage based on multivariate analysis. *Gene* 1999,**238**:143-155.

50. Percudani R, Pavesi A, Ottonello S. Transfer RNA gene redundancy and translational selection in Saccharomyces cerevisiae. *J Mol Biol* 1997,**268**:322-330.

51. Duret L. tRNA gene number and codon usage in the C. elegans genome are co-adapted for optimal translation of highly expressed genes. *Trends Genet* 2000,**16**:287-289.

52. dos Reis M, Savva R, Wernisch L. Solving the riddle of codon usage preferences: a test for translational selection. *Nucleic Acids Res* 2004,**32**:5036-5044.

53. Chan PP, Lowe TM. GtRNAdb: a database of transfer RNA genes detected in genomic sequence. *Nucleic Acids Res* 2009,**37**:D93-97.

54. McInerney JO. Replicational and transcriptional selection on codon usage in Borrelia burgdorferi. *Proc Natl Acad Sci U S A* 1998,**95**:10698-10703.

55. McNulty DE, Claffee BA, Huddleston MJ, Porter ML, Cavnar KM, Kane JF. Mistranslational errors associated with the rare arginine codon CGG in Escherichia coli. *Protein Expr Purif* 2003,**27**:365-374.

56. Kim S, Lee SB. Rare codon clusters at 5'-end influence heterologous expression of archaeal gene in Escherichia coli. *Protein Expr Purif* 2006,**50**:49-57.

57. Seligmann H, Pollock DD. The ambush hypothesis: hidden stop codons prevent off-frame gene reading. *DNA Cell Biol* 2004,**23**:701-705.

58. Rupprecht TA, Koedel U, Fingerle V, Pfister HW. The pathogenesis of lyme neuroborreliosis: from infection to inflammation. *Mol Med* 2008,**14**:205-212.

59. Wooten RM, Ma Y, Yoder RA, Brown JP, Weis JH, Zachary JF*, et al.* Toll-like receptor 2 is required for innate, but not acquired, host defense to Borrelia burgdorferi. *J Immunol* 2002,**168**:348-355.

60. Pahl A, Kuhlbrandt U, Brune K, Rollinghoff M, Gessner A. Quantitative detection of Borrelia burgdorferi by real-time PCR. *J Clin Microbiol* 1999,**37**:1958-1963.

61. Zhou W, Brisson D. Potentially conflicting selective forces that shape the vls antigenic variation system in Borrelia burgdorferi. *Infect Genet Evol* 2014.

62. Binder SC, Telschow A, Meyer-Hermann M. Population Dynamics of Borrelia burgdorferi in Lyme Disease. *Front Microbiol* 2012,**3**:104.

63. Barthold SW, Persing DH, Armstrong AL, Peeples RA. Kinetics of Borrelia burgdorferi dissemination and evolution of disease after intradermal inoculation of mice. *Am J Pathol* 1991,**139**:263-273.

64. Wooten RM, Ma Y, Yoder RA, Brown JP, Weis JH, Zachary JF*, et al.* Toll-like receptor 2 plays a pivotal role in host defense and inflammatory response to Borrelia burgdorferi. *Vector Borne Zoonotic Dis* 2002,**2**:275-278.

65. Rao KV. Selection in a T-dependent primary humoral response: new insights from polypeptide models. *APMIS* 1999,**107**:807-818.

66. Barbour AG. Isolation and cultivation of Lyme disease spirochetes. *Yale J. Biol. Med.* 1983,**57**:521-525.

67. van Furth R, Diesselhoff-Den Dulk MM. The kinetics of promonocytes and monocytes in the bone marrow. *J Exp Med* 1970,**132**:813-828.

68. Hirschfeld M, Kirschning CJ, Schwandner R, Wesche H, Weis JH, Wooten RM*, et al.* Cutting edge: inflammatory signaling by Borrelia burgdorferi lipoproteins is mediated by toll-like receptor 2. *J Immunol* 1999,**163**:2382-2386.

69. Eicken C, Sharma V, Klabunde T, Lawrenz MB, Hardham JM, Norris SJ*, et al.* Crystal structure of Lyme disease variable surface antigen VlsE of Borrelia burgdorferi. *J Biol Chem* 2002,**277**:21691-21696.

70. Stave JW, Lindpaintner K. Antibody and antigen contact residues define epitope and paratope size and structure. *J Immunol* 2013,**191**:1428-1435.

71. Novotny J, Handschumacher M, Haber E, Bruccoleri RE, Carlson WB, Fanning DW*, et al.* Antigenic determinants in proteins coincide with surface regions accessible to large probes (antibody domains). *Proc Natl Acad Sci U S A* 1986,**83**:226-230.

72. Thornton JM, Edwards MS, Taylor WR, Barlow DJ. Location of 'continuous' antigenic determinants in the protruding regions of proteins. *EMBO J* 1986,**5**:409-413.

73. Parker JM, Guo D, Hodges RS. New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: correlation of predicted surface residues with antigenicity and X-ray-derived accessible sites. *Biochemistry* 1986,**25**:5425-5432.

74. Sun J, Xu T., Wang, S., Li, G., Wu, D., Cao, Z. Does difference exist between epitope and non-epitope residues? Analysis of the physicochemical and structural properties on conformational epitopes from B-cell protein antigens. *Immunome Research* 2011,**7**:11.

75. Haste Andersen P, Nielsen M, Lund O. Prediction of residues in discontinuous B-cell epitopes using protein 3D structures. *Protein Sci* 2006,**15**:2558-2567.

76. Cavallo L, Kleinjung J, Fraternali F. POPS: A fast algorithm for solvent accessible surface areas at atomic and residue level. *Nucleic Acids Res* 2003,**31**:3364-3366.

77. Cunningham BC, Wells JA. High-resolution epitope mapping of hGH-receptor interactions by alanine-scanning mutagenesis. *Science* 1989,**244**:1081-1085.

78. Ye J, Xu Y, Harris J, Sun H, Bowman AS, Cunningham F*, et al.* Mutation from arginine to lysine at the position 189 of hemagglutinin contributes to the antigenic drift in H3N2 swine influenza viruses. *Virology* 2013,**446**:225-229.

79. Air GM, Laver WG, Webster RG. Mechanism of antigenic variation in an individual epitope on influenza virus N9 neuraminidase. *J Virol* 1990,**64**:5797-5803.

80. Pace CN, Fu H, Fryar KL, Landua J, Trevino SR, Shirley BA*, et al.* Contribution of hydrophobic interactions to protein stability. *J Mol Biol* 2011,**408**:514-528.

81. Collawn JF, Wallace CJ, Proudfoot AE, Paterson Y. Monoclonal antibodies as probes of conformational changes in protein-engineered cytochrome c. *J Biol Chem* 1988,**263**:8625-8634.

82. Muhle C, Schulz-Drost S, Khrenov AV, Saenko EL, Klinge J, Schneider H. Epitope mapping of polyclonal clotting factor VIII-inhibitory antibodies using phage display. *Thromb Haemost* 2004,**91**:619-625.

83. Morrison KL, Weiss GA. Combinatorial alanine-scanning. *Curr Opin Chem Biol* 2001,**5**:302-307.

84. Hedrick PW. Genetic variation in a heterogeneous environment. II. Temporal heterogeneity and directional selection. *Genetics* 1976,**84**:145-157.

85. Steiner CF. Environmental noise, genetic diversity and the evolution of evolvability and robustness in model gene networks. *PLoS One* 2012,**7**:e52204.

86. Ferreira MU, da Silva Nunes M, Wunderlich G. Antigenic diversity and immune evasion by malaria parasites. *Clin Diagn Lab Immunol* 2004,**11**:987-995.

87. Cahoon LA, Seifert HS. Focusing homologous recombination: pilin antigenic variation in the pathogenic Neisseria. *Mol Microbiol* 2011,**81**:1136-1143.

88. Levins R. Theory of fitness in a heterogeneous environment. VI. The adaptive significance of mutation. *Genetics* 1967,**56**:163-178.