



Potentially conflicting selective forces that shape the *vls* antigenic variation system in *Borrelia burgdorferi*



Wei Zhou*, Dustin Brisson

University of Pennsylvania, 3451 Walnut Street, Philadelphia, PA 19104, United States

ARTICLE INFO

Article history:

Received 14 January 2014
Received in revised form 4 April 2014
Accepted 23 April 2014
Available online 14 May 2014

Keywords:

Environmental change
Antigenic variation
Borrelia burgdorferi
VlsE
Antigenic divergence
Translational selection

ABSTRACT

Changing environmental conditions present an evolutionary challenge for all organisms. The environment of microbial pathogens, including the adaptive immune responses of the infected host, changes rapidly and is lethal to the pathogen lineages that cannot quickly adapt. The dynamic immune environment creates strong selective pressures favoring microbial pathogen lineages with antigenic variation systems that maximize the antigenic divergence among expressed antigenic variants. However, divergence among expressed antigens may be constrained by other molecular features such as the efficient expression of functional proteins. We computationally examined potential conflicting selection pressures on antigenic variation systems using the *vls* antigenic variation system in *Borrelia burgdorferi* as a model system. The *vls* system alters the sequence of the expressed antigen by recombining gene fragments from unexpressed but divergent ‘cassettes’ into the expression site, *vlsE*. The *in silico* analysis of natural and altered cassettes from seven lineages in the *B. burgdorferi* sensu lato species complex revealed that sites that are polymorphic among unexpressed cassettes, as well as the insertion/deletion mutations, are organized to maximize divergence among the expressed antigens within the constraints of translational ability and high translational efficiency. This study provides empirical evidence that conflicting selection pressures on antigenic variation systems can limit the potential antigenic divergence in order to maintain proper molecular function.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

Changes in environmental conditions are a common source of natural selection driving adaptation in natural populations. While adaptation to predictable or cyclic environmental changes has been the focus of numerous studies (for example, Cronin and Schneider, 1990; Erwin, 2009; Merila, 2012), adaptation to unpredictable and rapidly changing environments is less well characterized. However, unpredictable and rapidly changing environments are common and can result in lethal selection pressures, such as the adaptive responses of vertebrate immune systems that shape the evolutionary dynamics of pathogen populations. The persistence of pathogens 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 (Brunham et al., 1993; Combes, 1997; Deitsch et al., 1997; Schmid-Hempel, 2009). These strong selective

pressure imposed by the immune response has resulted in antigenic variation mechanisms evolved to cope with this rapidly changing environment (Brunham et al., 1993; Deitsch et al., 1997; Frank, 2002; Moxon et al., 1994; Schmid-Hempel, 2009).

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 (van der Woude and Baumler, 2004). Evading the antibody response permits longer residence times of the pathogens within the host, thus increasing opportunities for transmission to naïve hosts (Deitsch et al., 1997; Moxon et al., 1994). Antigenic variation systems that more efficiently alter the antigenic surface 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.

Many pathogens utilize antigenic variation systems that alter the genetic sequence in the expression site of the antigenic protein (Deitsch et al., 2009; van der Woude and Baumler, 2004). One common molecular mechanism involves recombining gene fragments from unexpressed, paralogous ‘cassettes’ into an expression site, thereby altering the sequence of the expressed antigen. In these

* Corresponding author. Address: Department of Biology, University of Pennsylvania, Leidy Laboratories, 307A, 433 South University Avenue, Philadelphia, PA 19104-6018, United States. Tel.: +1 (215) 746 1732.

E-mail address: zhouwei2@sas.upenn.edu (W. Zhou).

types of recombination-based antigenic variation systems, which are common in several bacterial genera (Hagblom et al., 1985; Noormohammadi et al., 2000; Zhang et al., 1997), 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 (Graves et al., 2013; Lipsitch and O'Hagan, 2007). However, the extent of the divergence among cassettes can be constrained by other features of the system (Haydon and Woolhouse, 1998). Here, we use the well-characterized *vls* antigenic variation system in the Lyme disease bacterium, *Borrelia burgdorferi*, as a model system to investigate the interactions between selection favoring greater antigenic divergence and other potential constraints on antigenic variation systems.

B. burgdorferi requires continuous alteration of the highly-expressed VlsE antigen for long-term survival within hosts (Bankhead and Chaconas, 2007; Bykowski et al., 2006; Labandeira-Rey and Skare, 2001; McDowell et al., 2002; Purser and Norris, 2000; Rogovskyy and Bankhead, 2013; Zhang et al., 1997). 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 (Coutte et al., 2009; Hershberg and Petrov, 2008). Little is currently known about how selection on translational ability or efficiency constrains the nucleotide identities at the polymorphic sites, positions of the polymorphic sites and positions of the insertion/deletion mutations.

Here we evaluated the effects of the identity of nucleotides at polymorphic sites, positions of the polymorphic sites, and position of insertion/deletion mutations in the unexpressed cassettes on the divergence among antigenic variants as well as their translational ability and translational efficiency. We ask if the organization of polymorphic sites and insertion/deletion mutations in the unexpressed cassettes of multiple natural strains results in the greatest possible antigenic divergence, translational ability, and translational efficiency in the VlsE variants. We used *in silico* simulation models to test if perturbing the observed polymorphic sites leads to a decrease in antigenic divergence, translational ability and translational efficiency.

2. Material and methods

2.1. Sequence analysis of *vlsE* and the unexpressed cassettes

The sequences of the unexpressed cassettes from six strains of *B. burgdorferi* sensu stricto and one *Borrelia afzelii* strain were used to investigate how diversifying selection and translational

selection constrain identities and locations of polymorphism among the unexpressed cassettes (Table 1). Each of the unexpressed cassettes within each strain was aligned using ClustalW (Larkin et al., 2007) with default parameters. The unexpressed *vls* cassettes from all strains have six or seven variable regions in which polymorphic sites are concentrated as described previously (Zhang et al., 1997) (Fig. S1). Unexpressed cassettes that did not include all variable regions were not analyzed (Fig. S1).

2.2. *In silico* perturbation of unexpressed cassettes

For each set of natural cassettes, three perturbation models were generated using the three algorithms (δ Nuc, δ Pos, and δ InDel) described below and in Fig. 1. The perturbation models have altered either (a) nucleotide identity at each polymorphic site (δ Nuc), (b) the locations of the polymorphic sites within the variable regions (δ Pos), or (c) the locations of insertion/deletion mutations within the variable regions (δ InDel). All perturbation models were run independently on each strain.

2.2.1. δ Nuc algorithm

The δ Nuc algorithm converts the nucleotides observed at every polymorphic site in the cassettes of each natural strain to an alternative nucleotide (Fig. 1A). That is, all nucleotides of identity X are converted to identity Y (for example, all adenines at a given polymorphic site are converted to cytosines). The identity of the nucleotide to replace the original nucleotide is chosen at random for each polymorphic site in each iteration of the model. Nucleotide conversion is bijective in that all nucleotides at a polymorphic site of identity X will be converted to identity Y, and Y will only be used to replace nucleotides of identity X at that polymorphic site. The δ Nuc 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. δ Pos algorithm

The δ Pos algorithm relocates the position of all polymorphic 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 (Fig. 1B). The nucleotide identities at the relocated polymorphic sites are altered following the δ Nuc algorithm (Section 2.2.1).

2.2.3. δ InDel algorithm

The δ InDel algorithm relocates the position of all polymorphic sites, including those that contain insertion/deletion mutations, to random positions within the variable regions of the unexpressed cassettes (Fig. 1C). The nucleotide identities at the relocated sites are altered following the δ Nuc algorithm (Section 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.

Each algorithm was used to create 50 iterations for each perturbation model, resulting in a total of 1050 (7 strains \times 3 models \times 50 iterations) sets of perturbed unexpressed cassettes.

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* strains 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

Table 1
Unexpressed cassettes in six strains of *B. burgdorferi* sensu stricto and in *B. afzelii*.

Strains	Number of polymorphic sites	Number of variable regions	Reference
B31	191	6	Zhang et al. (1997)
JD1	166	6	Schutzer et al. (2011)
WI9123	115	6	Schutzer et al. (2011)
29805	140	6	Schutzer et al. (2011)
Bol26	234	6	Schutzer et al. (2011)
Zs7	90	6	Schutzer et al. (2011)
<i>B. afzelii</i>	223	7	Wang et al. (2003)

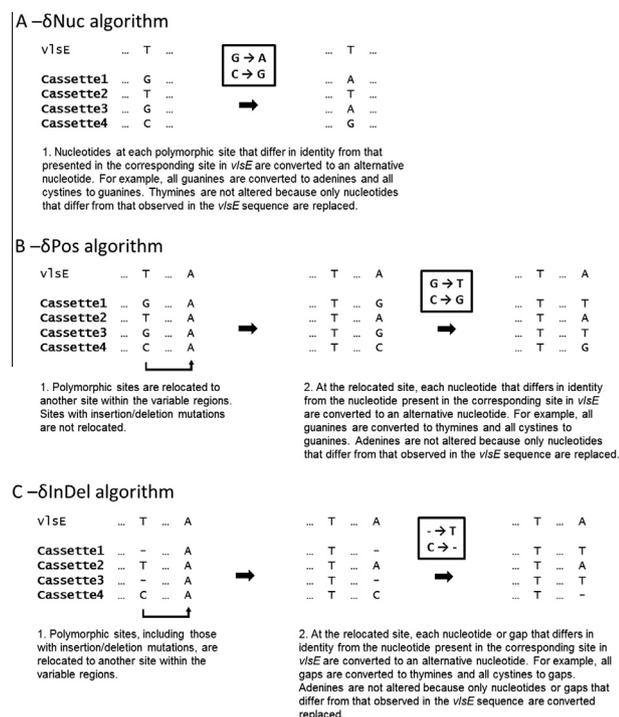


Fig. 1. Examples of algorithms perturbing the nucleotides at polymorphic sites or the positions of the polymorphic sites. (A) δ Nuc converts the polymorphic nucleotides to alternative nucleotides. (B) δ Pos relocates polymorphic sites within the variable regions. The relocated polymorphic sites are then perturbed as in δ Nuc. (C) δ InDel relocates the polymorphic sites, including those with insertion/deletion mutations, within the variable regions. The new polymorphic sites are then perturbed as in δ Nuc, with the addition of gap as a fifth possible type of ‘nucleotide’.

was chosen at random from an even distribution between 2 and 200 bp. The range and distribution of segment lengths was derived from lengths of recombining fragments observed in *B. burgdorferi* strains infecting laboratory mice (Coutte et al., 2009).

2.4. Antigenic divergence, translational ability, and translational efficiency of *vlsE* variants

We evaluated antigenic 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.

2.4.1. Estimation of antigenic divergence

Amino acid sequence divergence is linearly correlated with antigenic divergence (Benjamin et al., 1984; Frank, 2002; Prager, 1993). Thus, antigenic 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).

2.4.2. 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.

2.4.3. 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 (Zhang et al., 1997). Codons with more abundant tRNAs are likely to be translated with a greater elongation rate and with greater accuracy (Hershberg and Petrov, 2008). Thus, optimal codon usage with respect to tRNA abundance was used as a proxy for translational efficiency. Assuming that the copy number of tRNA genes is correlated with tRNA abundance (Dong et al., 1996; Kanaya et al., 1999), codon optimality, $E(T)$, can be estimated as the number of copies of tRNA genes that pair to a codon, averaged over all codons in the *vlsE* sequence.

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

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 (Chan and Lowe, 2009).

2.5. Robustness analyses of the perturbation algorithms

Two types of hypothetical *vlsE* cassettes (negative and positive control cassettes) were generated for each of the algorithms (δ Nuc, δ Pos, and δ InDel) to benchmark their performance. Negative control cassettes represent sequences experiencing no selection for translational ability or efficiency. Each set of negative control cassettes contains 15 cassette sequences, same as the number of cassettes in the strain B31. The sequences of the cassettes were generated by first randomly choosing nucleotides at each site, then creating polymorphic sites at randomly chosen locations. The nucleotide identities (A, T, G, C, or gap) at the polymorphic sites were chosen at random. Positive control cassettes represent sequences experiencing strong selection on translation. Each set of positive control cassettes contains 15 cassette sequences. The sequences of the positive control cassettes were generated by, first, randomly choosing codons sampled from the frequency distribution of its corresponding tRNA genes, then creating polymorphic sites at randomly chosen locations. In each cassette, codons containing the polymorphic nucleotides (A, T, G or C) were sampled according to the frequency of the tRNA genes corresponding to the codon. Two sets of *in silico*-constructed negative and positive control cassettes, each of 630 total base pairs, equivalent in length to strain B31 including insertion/deletion mutations, were generated for benchmarking analyses.

Each set of negative and positive control cassettes were altered by each of the perturbation algorithms (δ Nuc, δ Pos, and δ InDel) to create 50 perturbed cassettes, resulting in a total of 600 (4 pairs \times 3 models \times 50 iterations) sets of perturbed hypothetical cassettes. Each of the perturbed sets of cassettes were evaluated with regard to antigenic divergence, translational ability, and translational efficiency of the *vlsE* variants produced, as described in Section 2.4.

The robustness of the results from cassettes perturbed by δ Pos and δ InDel to the locations to which polymorphic sites are moved was also evaluated. We perturbed the unexpressed cassettes of strain B31 using algorithms similar to δ Pos and δ InDel but allowing polymorphic sites to relocate only to sites within the variable region from which they originated. Using each of the two algorithms, 50 sets of perturbed unexpressed cassettes were created. The resulting perturbed cassettes were evaluated as was described in Section 2.4.

3. Results

3.1. Effect of altering the nucleotide identities observed in the polymorphic sites

The nucleotide identities observed in polymorphic sites of the unexpressed cassettes of natural strains result in the greatest antigenic divergence, translational ability, and translational efficiency in VlsE variants. Changing nucleotides at the sites that are naturally polymorphic (δNuc) among the unexpressed cassettes dramatically reduced the antigenic divergence among VlsE variants as well as the translational ability and translational efficiency of the variants (Fig. 2). The reduction in antigenic divergence, translational ability, and translational efficiency were statistically significant ($p < 0.05$) in all strains except for antigenic divergence in strain Bol26 ($p = 0.057$).

3.2. Effect of altering the positions of the polymorphic sites

Altering the positions of the polymorphic 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 (Fig. 3B and C). However, the translational ability of VlsE variants produced from the cassettes with altered polymorphic locations was significantly lower than VlsE variants produced from cassettes with altered nucleotides (δNuc) in only two of the seven genotypes (WI9123 and Bol26) and translational efficiency was significantly lower in only three of the seven genotypes (29805, Bol26 and *B. afzelii*) (Figs. 2B, C and 3B, C). Thus, it is not apparent that altering the positions of the polymorphic 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 polymorphic sites must be accompanied by alterations in nucleotides, resulting in a reduction in power to detect effects of the position of the polymorphic sites on translational ability and translational efficiency of VlsE variants independent of the large effect of nucleotide identity.

Antigenic divergence among VlsE variants generated from the unexpressed cassettes with computationally altered positions of polymorphic 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*) (Fig. 3A). However, the divergence among VlsE variants generated from cassettes with positionally altered polymorphic sites (δPos) was significantly lower than the divergence produced from altering the nucleotide types (δNuc) only in *B. afzelii* (Figs. 2A and 3A). Thus, altering the positions of the polymorphic sites does not cause a greater reduction in antigenic divergence than altering the nucleotide identities alone. Further, alteration of the position of polymorphic 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 polymorphic sites in this strain, the sequences in the expression site that resulted from these recombination events had low translational ability and efficiency (Fig. 3B and C) which are likely to constrain changes in the location of the polymorphic sites.

The δPos algorithm in which polymorphic sites are only allowed to move within the variable region where they originated leads to qualitatively and quantitatively similar results to those described above (Fig. S5).

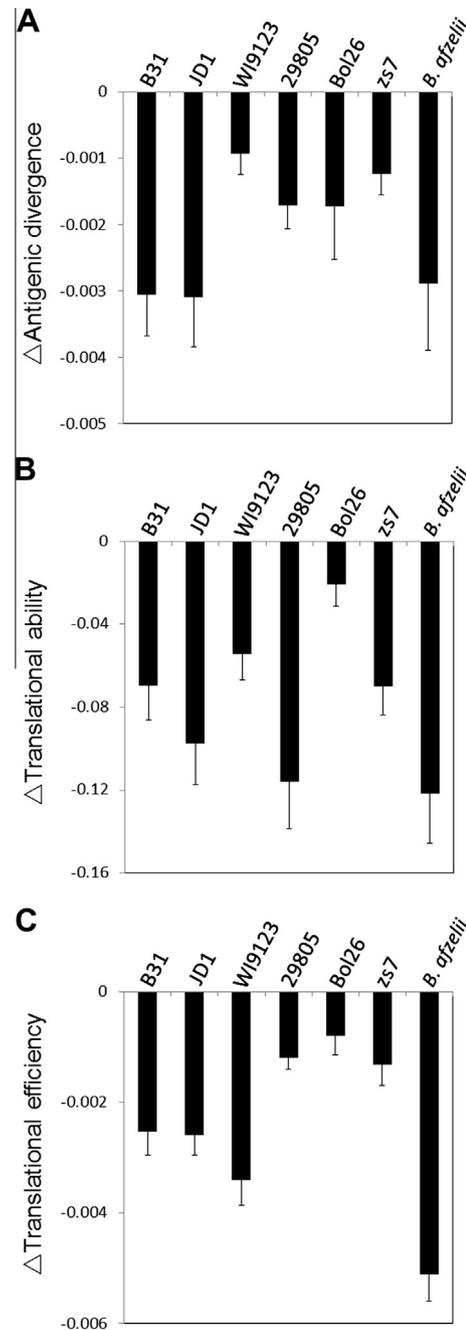


Fig. 2. Changes in antigenic divergence, translational ability, and translational efficiency of VlsE variants after altering the identities of the polymorphic nucleotides in unexpressed cassettes (δNuc). The y-axis shows difference in antigenic divergence (A), translational ability (B), and translational efficiency (C) from the values of VlsE variants generated by the natural unexpressed cassettes after perturbing the polymorphic nucleotides. Negative values on the y-axis indicate reduced antigenic divergence, translational ability, or translational efficiency due to cassette perturbation. Error bars represent 95% confidence intervals of the mean.

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 polymorphic 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 polymorphic sites (δPos) (Figs. 3B and 4B). The decreases in translational ability

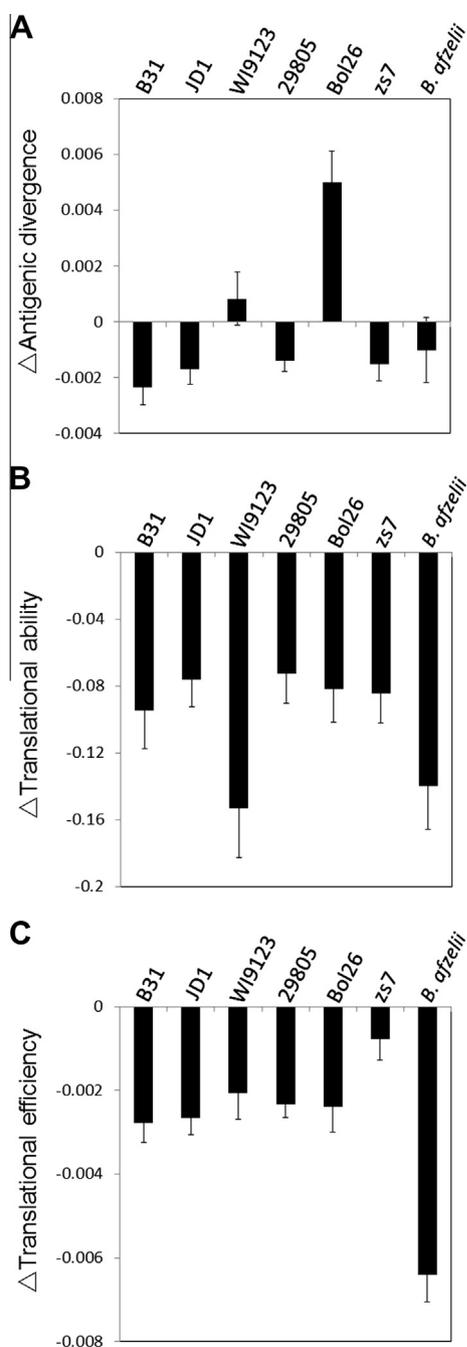


Fig. 3. Changes in antigenic divergence, translational ability and translational efficiency of VlsE variants after perturbing the positions of the polymorphic sites in unexpressed cassettes (δPos). The y-axis shows differences in the antigenic divergence (A), translational ability (B) and translational efficiency (C) from the values of VlsE generated by the natural unexpressed cassettes after perturbing the positions of polymorphic sites. Negative values on the y-axis indicate reduced antigenic divergence, translational ability, or translational efficiency due to cassette perturbation. Error bars represent 95% confidence intervals of the mean.

are dramatic, with as many as half of the variants containing at least one stop codon (Fig. 4B). The VlsE variants also show significantly lower antigenic divergence and lower translational efficiency than those generated by the natural unexpressed cassettes in all strains (Fig. 4A and C). These results indicate that the patterns of insertion/deletion occurrence are strongly constrained by antigenic divergence, translational efficiency, and especially translational ability. The δInDel algorithm in which polymorphic sites are only allowed to move within the variable region where they

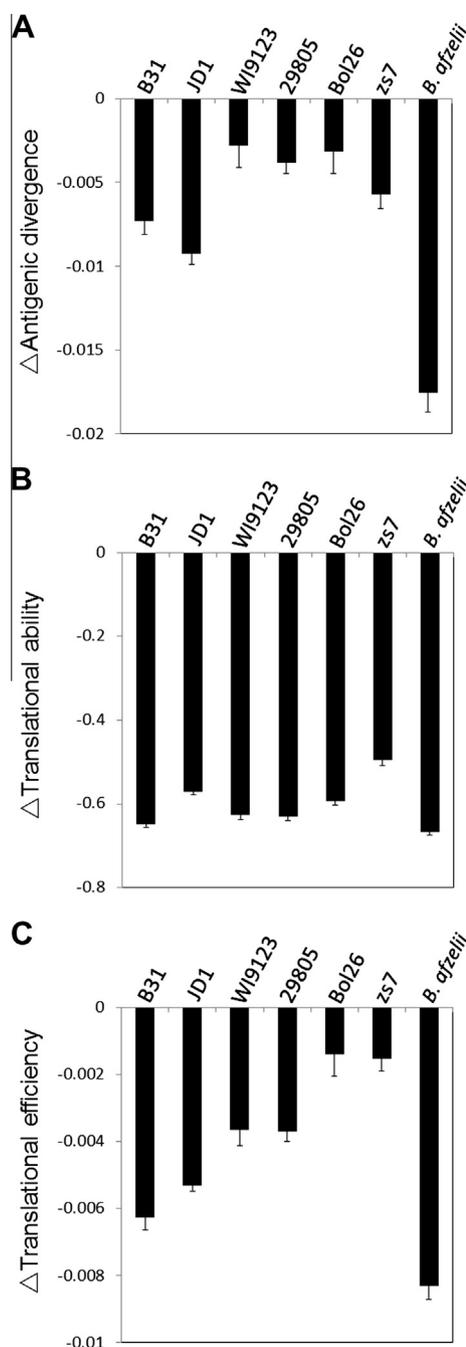


Fig. 4. Changes in antigenic divergence, translational ability and translational efficiency of VlsE variants after perturbing the insertion/deletion mutations in unexpressed cassettes (δInDel). The y-axis shows differences in the antigenic divergence (A), translational ability (B) and translational efficiency (C) from the values of VlsE generated by the natural unexpressed cassettes after perturbing the insertion/deletion mutations. Negative values on the y-axis indicate reduced antigenic divergence, translational ability, or translational efficiency due to cassette perturbation. Error bars represent 95% confidence intervals of the mean.

originated leads to qualitatively and quantitatively similar results to those described above (Fig. S5).

3.4. Robustness analyses of the perturbation algorithms

The robustness of the three perturbation algorithms was confirmed using *in silico*-generated hypothetical cassettes that have no selection for translational properties (negative control cassettes) or strong selection for translational properties (positive

control cassettes). As expected, perturbation of the negative control cassettes using any of the three algorithms does not decrease the antigenic divergence, translational ability, or translational efficiency (Figs. S2–S4). Also as expected, perturbation of the positive control cassettes using any of the three algorithms dramatically decrease the antigenic divergence, translational ability, and translational efficiency because perturbation eliminated some of the translationally-optimal codons.

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 *B. burgdorferi* *sensu lato* during infections of vertebrate hosts is shaped by the sometimes conflicting forces of natural selection acting on greater antigenic divergence, translational ability, and translational efficiency.

B. burgdorferi 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 antigenic divergence among expressed proteins, as well as the translational ability and translational efficiency of those proteins, is maximized by using the nucleotides observed at the sites that are polymorphic among the unexpressed cassettes. Altering the nucleotides used at these sites can reduce the antigenic divergence, introduce stop codons, and reduce the elongation rate by using non-preferred codons. Although many of the *B. burgdorferi* genomic codons are chosen because of mutational biases or replicational selection and are not optimal for translation (McInerney, 1998), our results show that codon usage in *vlsE* is optimized for translational efficiency. Recombination events that use only preferred codons may be particularly important to the elongation rates and translational accuracy of *VlsE* variants as there are many polymorphic codons that are tandemly arrayed in the unexpressed cassettes. Tandem codons using non-preferred codons can act synergistically to reduce translational rate and accuracy (Kim and Lee, 2006; McNulty et al., 2003). A hypothesized function of the *VlsE* protein, to protect other surface-exposed proteins from antibodies by hiding these potential antigens (Bankhead and Chaconas, 2007), requires highly expressed and accurately translated proteins for appropriate antigenicity.

The position of the polymorphic sites within the antigenically important variable regions in the natural strains resulted in the maximal translational ability and translational efficiency of generated *VlsE* variants. Altering the position of the polymorphic sites within the variable regions significantly reduced both the translational ability and translational efficiency of *VlsE* variants generated. The antigenic 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, antigenic divergence among the *VlsE* variants generated after altering the positions of the polymorphic sites in one strain (Bol26) was significantly greater than those generated from the natural cassettes. Strain Bol26 is a unique case where several codons contained multiple polymorphic sites. These polymorphic sites become separated on different codons after positional perturbation, leading to greater amino-acid-level divergence after recombination.

The organization and identity of polymorphic sites in the unexpressed cassettes of natural strains resulted in the maximal amino acid divergence among the *VlsE* variants, suggesting that the sequences of the unexpressed cassettes observed in natural strains maximizes antigenic divergence. Amino acid sequence difference among antigens is negatively correlated with cross-reactivity with polyclonal antibodies (Benjamin et al., 1984; Frank, 2002; Prager, 1993). This reduction in binding affinity to antibodies is crucial for immune evasion and longevity of *B. burgdorferi* in infected hosts, a key factor in the evolutionary fitness of *B. burgdorferi*. Although altering the positions of polymorphic 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.

The primary mechanism by which perturbing the identity of polymorphic nucleotides can decrease the antigenic divergence of *VlsE* variants is by introducing premature stop codons preferentially in divergent variants. While there are almost no stop codons found in the natural cassettes, stop codons are regularly found in perturbed cassettes. Importantly, the number of perturbed sites in the recombined fragment is positively correlated with both the probability of introducing a stop codon into *vlsE* and the degree of antigenic divergence from the original *VlsE*. However, antigenic divergence cannot be calculated for sequences with stop codons. Thus recombination events using fragments with many perturbed polymorphic sites, which would lead to high antigenic divergence, are also the events that are more likely to contain a stop codon where antigenic divergence cannot be calculated. Thus, the average antigenic divergence computed from the *translatable* sequences will decrease. While this is a computational explanation, it is very relevant biologically as the degree of antigenic divergence is important only for translatable sequences.

Perturbing both nucleotide identities and the positions of polymorphic 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 polymorphic 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 polymorphic sites cannot be altered without also altering the nucleotides at those positions. Thus, the effect of the position of the polymorphic sites cannot be disentangled from the effects of nucleotide identities at these sites, resulting in low statistical power to detect the effect of the polymorphic positions. Second, the positions of polymorphic 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*. Future experimental work exploring the mechanistic effects of the position and identity of nucleotides can determine the effects of the position of polymorphic nucleotides in the *vls* antigenic variation system.

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 (Graves et al., 2013). 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 (Seligmann and Pollock, 2004).

In recombination-based antigenic variation systems, the genetic sequences of the unexpressed cassettes determine the degree of antigenic variation possible during a host infection. The analyses presented demonstrate that the *vls* antigenic variation systems of *B. burgdorferi* is organized, both in location of sites that are polymorphic among unexpressed cassettes and the identities of the nucleotides used at the polymorphic sites, to maximize antigenic divergence among protein variants while simultaneously maximizing translational ability and translational efficiency. The selection on translational properties in unexpressed regions was strong and in some cases appeared to constrain the selection for greater antigenic diversity. Antigenic diversity and molecular functions may sometimes act as conflicting fitness-related constraints on the antigenic variation systems necessary to cope with rapid and unpredictable environmental changes within the host. 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* (Hagblom et al., 1985; Noormohammadi et al., 2000). In these pathogens, similar analyses to those conducted in this study can be used to establish the sets of selection pressures that affect the evolution of unexpressed cassettes. Identification of the constraints on evolution of antigenic variation systems may also inform the design of vaccines or other therapies.

Acknowledgments

We thank Steven Norris for assistance with the project, and Junhyong Kim, Paul Sniegowski, Rahul Kohli and Timothy Linksvayer for helpful suggestions.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.meegid.2014.04.020>.

Reference

- Bankhead, T., Chaconas, G., 2007. The role of VlsE antigenic variation in the Lyme disease spirochete: persistence through a mechanism that differs from other pathogens. *Mol. Microbiol.* 65, 1547–1558.
- Benjamin, D.C., Berzofsky, J.A., East, I.J., Gurd, F.R., Hannum, C., Leach, S.J., Margoliash, E., Michael, J.G., Miller, A., Prager, E.M., et al., 1984. The antigenic structure of proteins: a reappraisal. *Annu. Rev. Immunol.* 2, 67–101.
- Brunham, R.C., Plummer, F.A., Stephens, R.S., 1993. Bacterial antigenic variation, host immune response, and pathogen–host coevolution. *Infect. Immun.* 61, 2273–2276.
- Bykowski, T., Babb, K., von Lackum, K., Riley, S.P., Norris, S.J., Stevenson, B., 2006. Transcriptional regulation of the *Borrelia burgdorferi* antigenically variable VlsE surface protein. *J. Bacteriol.* 188, 4879–4889.
- Chan, P.P., Lowe, T.M., 2009. GtRNAdb: a database of transfer RNA genes detected in genomic sequence. *Nucleic Acids Res.* 37, D93–D97.
- Combes, C., 1997. Fitness of parasites: pathology and selection. *Int. J. Parasitol.* 27, 1–10.
- Coutte, L., Botkin, D.J., Gao, L., Norris, S.J., 2009. Detailed analysis of sequence changes occurring during vlsE antigenic variation in the mouse model of *Borrelia burgdorferi* infection. *PLoS Pathog.* 5, e1000293.
- Cronin, T.M., Schneider, C.E., 1990. Climatic influences on species: evidence from the fossil record. *Trends Ecol. Evol.* 5, 275–279.
- Deitsch, K.W., Lukehart, S.A., Stringer, J.R., 2009. Common strategies for antigenic variation by bacterial, fungal and protozoan pathogens. *Nature Rev. Microbiol.* 7, 493–503.
- Deitsch, K.W., Moxon, E.R., Wellems, T.E., 1997. Shared themes of antigenic variation and virulence in bacterial, protozoal, and fungal infections. *Microbiol. Mol. Biol. Rev.* 61, 281–293.
- Dong, H., Nilsson, L., Kurland, C.G., 1996. Co-variation of tRNA abundance and codon usage in *Escherichia coli* at different growth rates. *J. Mol. Biol.* 260, 649–663.
- Erwin, D.H., 2009. Climate as a driver of evolutionary change. *Curr. Biol.* 19, R575–R583.
- Frank, S.A., 2002. *Immunology and Evolution of Infectious Disease*. Princeton University Press, NJ.
- Graves, C.J., Ros, V.I., Stevenson, B., Sniegowski, P.D., Brisson, D., 2013. Natural selection promotes antigenic evolvability. *PLoS Pathog.* 9, e1003766.
- Hagblom, P., Segal, E., Billyard, E., So, M., 1985. Intragenic recombination leads to pilus antigenic variation in *Neisseria gonorrhoeae*. *Nature* 315, 156–158.
- Haydon, D.T., Woolhouse, M.E., 1998. Immune avoidance strategies in RNA viruses: fitness continuums arising from trade-offs between immunogenicity and antigenic variability. *J. Theor. Biol.* 193, 601–612.
- Hershberg, R., Petrov, D.A., 2008. Selection on codon bias. *Annu. Rev. Genet.* 42, 287–299.
- Kanaya, S., Yamada, Y., Kudo, Y., Ikemura, T., 1999. 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* 238, 143–155.
- Kim, S., Lee, S.B., 2006. Rare codon clusters at 5'-end influence heterologous expression of archaeal gene in *Escherichia coli*. *Protein Expr. Purif.* 50, 49–57.
- Labandeira-Rey, M., Skare, J.T., 2001. Decreased infectivity in *Borrelia burgdorferi* strain B31 is associated with loss of linear plasmid 25 or 28-1. *Infect. Immun.* 69, 446–455.
- Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., Higgins, D.G., 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* 23, 2947–2948.
- Lipsitch, M., O'Hagan, J.J., 2007. Patterns of antigenic diversity and the mechanisms that maintain them. *J. R. Soc. Interface/R. Soc.* 4, 787–802.
- McDowell, J.V., Sung, S.Y., Hu, L.T., Marconi, R.T., 2002. Evidence that the variable regions of the central domain of VlsE are antigenic during infection with Lyme disease spirochetes. *Infect. Immun.* 70, 4196–4203.
- McInerney, J.O., 1998. Replicational and transcriptional selection on codon usage in *Borrelia burgdorferi*. *Proc. Natl. Acad. Sci. U.S.A.* 95, 10698–10703.
- McNulty, D.E., Claffee, B.A., Huddleston, M.J., Porter, M.L., Cavnar, K.M., Kane, J.F., 2003. Mistranslational errors associated with the rare arginine codon CGG in *Escherichia coli*. *Protein Expr. Purif.* 27, 365–374.
- Merila, J., 2012. Evolution in response to climate change: in pursuit of the missing evidence. *BioEssays* 34, 811–818.
- Moxon, E.R., Rainey, P.B., Nowak, M.A., Lenski, R.E., 1994. Adaptive evolution of highly mutable loci in pathogenic bacteria. *Curr. Biol.* 4, 24–33.
- Noormohammadi, A.H., Markham, P.F., Kanci, A., Whithear, K.G., Browning, G.F., 2000. A novel mechanism for control of antigenic variation in the haemagglutinin gene family of mycoplasma synoviae. *Mol. Microbiol.* 35, 911–923.
- Prager, E.M., 1993. The sequence-immunology correlation revisited: data for cetacean myoglobins and mammalian lysozymes. *J. Mol. Evol.* 37, 408–416.
- Purser, J.E., Norris, S.J., 2000. Correlation between plasmid content and infectivity in *Borrelia burgdorferi*. *Proc. Nat. Acad. Sci. U.S.A.* 97, 13865–13870.
- Rogovskyy, A.S., Bankhead, T., 2013. Variable VlsE is critical for host reinfection by the Lyme disease spirochete. *PLoS ONE* 8, e61226.
- Schmid-Hempel, P., 2009. Immune defence, parasite evasion strategies and their relevance for 'macroscopic phenomena' such as virulence. *Phil. Trans. R. Soc. B* 364, 85–98.
- Schutzer, S.E., Fraser-Liggett, C.M., Casjens, S.R., Qiu, W.G., Dunn, J.J., Mongodin, E.F., Luft, B.J., 2011. Whole-genome sequences of thirteen isolates of *Borrelia burgdorferi*. *J. Bacteriol.* 193, 1018–1020.
- Seligmann, H., Pollock, D.D., 2004. The ambush hypothesis: hidden stop codons prevent off-frame gene reading. *DNA Cell Biol.* 23, 701–705.
- van der Woude, M.W., Baumler, A.J., 2004. Phase and antigenic variation in bacteria. *Clin. Microbiol. Rev.* 17, 581–611 (table of contents).
- Wang, D., Botkin, D.J., Norris, S.J., 2003. Characterization of the *vls* antigenic variation loci of the Lyme disease spirochaetes *Borrelia garinii* Ip90 and *Borrelia afzelii* ACA1. *Mol. Microbiol.* 47, 1407–1417.
- Zhang, J.R., Hardham, J.M., Barbour, A.G., Norris, S.J., 1997. Antigenic variation in Lyme disease borreliae by promiscuous recombination of VMP-like sequence cassettes. *Cell* 89, 275–285.