Infection, Genetics and Evolution xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

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 xxxx

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.

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).

http://dx.doi.org/10.1016/j.meegid.2014.04.020
1567-1348/

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).

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)
B. afzelii	223	7	Wang et al. (2003)

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 Fig. 1. The perturbation models have altered either (a) nucleotide identity at each polymorphic site (δNuc), (b) the locations of the polymorphic 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.

2.2.1. δNuc algorithm

The δNuc 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 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. δPos algorithm

The δPos 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 δNuc algorithm (Section 2.2.1).

2.2.3. δInDel algorithm

The δInDel 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 δNuc 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 × 3 models × 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

Please cite this article in press as: Zhou, W., Brisson, D. Potentially conflicting selective forces that shape the *vls* antigenic variation system in *Borrelia burgdorferi*. Infect. Genet. Evol. (2014), http://dx.doi.org/10.1016/j.meegid.2014.04.020