Published in final edited form as:
Cell Microbiol. 2013 December ; 15(12): . doi:10.1111/cmi.12182.

Antigenic variation and transmission fitness as drivers of bacterial strain structure

Guy H. Palmer and **Kelly A. Brayton***
Paul G. Allen School for Global Animal Health, Department of Veterinary Microbiology and Pathology, Washington State University, Pullman, Washington 99164-7090.

Summary

Shifts in microbial strain structure underlie both emergence of new pathogens and shifts in patterns of infection and disease of known agents. Understanding the selective pressures at a population level as well as the mechanisms at the molecular level represent significant gaps in our knowledge regarding microbial epidemiology. Highly antigenically variant pathogens, which are broadly represented among microbial taxons, are most commonly viewed through the mechanistic lens of how they evade immune clearance within the host. However, equally important are mechanisms that allow pathogens to evade immunity at the population level. The selective pressure of immunity at both the level of the individual host and the population is a driver of diversification within a pathogen strain. Using *Anaplasma marginale* as a model highly antigenically variable bacterial pathogen, we review how immunity selects for genetic diversification in alleles encoding outer membrane proteins both within and among strains. Importantly, genomic comparisons among strains isolated from diverse epidemiologic settings elucidates the counterbalancing pressures for diversification and conservation, driven by immune escape and transmission fitness, respectively, and how these shape pathogen strain structure.

Introduction

Microbes exhibit a tremendous range in their degree of genetic diversity within a single pathogen species. At one end of the spectrum are RNA viruses that can rapidly generate a population of clearly related but genetically distinct viruses, often designated as “quasi-species”. In contrast, agents such as *Bacillus anthracis*, in which there is marked conservation in the genome among isolates, characterize the opposite end of the range. While the fidelity of replication, very low in RNA viruses as compared to more complex organisms such as bacteria or eukaryotic parasites, affects the rate at which new genetic variants arise within a given pathogen species, strain structure itself is shaped by selective pressures. Analysis of different members of the *Parvoviridae* illustrate this principle: as RNA viruses, all viruses in this family are capable of rapid genetic change but individual members of the family display very different breadth of diversity in nature (reviewed in Servant-Delmas *et al*., 2010). These same patterns exist for bacterial pathogens. Perpetuation of *B. anthracis* is principally defined by its ability to sporulate and survive in the environment; thus once this phenotype was acquired, selection for additional genetic changes is either weak or only operative over very long periods associated with environmental conditions (Keim et al., 2009). In contrast, *Streptococcus agalactiae*, possessing the archetypical “open-core” genome, reveals marked genomic diversity in its strain structure, reflective of its ability to infect many different animal species and occupy

*Corresponding Author, Paul G. Allen School for Global Animal Health, Washington State University, Pullman, Washington 99164-7090, Tel: (509) 335-6033, Fax: (509) 335-6328, gpalmer@vetmed.wsu.edu.

distinct niches (reviewed in Fraser-Liggett, 2005). *Anaplasma marginale*, the protagonist of this review, also reflects a broad strain structure: over 100 genetically distinct strains have been identified and additional unique genotypes continue to be detected and reported (reviewed in Kocan *et al.*, 2010). Unlike *S. agalactiae* however, *A. marginale* genotypic diversity occurs in the context of an overall "closed-core" genome in which the gene content itself is highly conserved among all *sensu stricto* strains (Brayton *et al.*, 2005; Dark *et al.*, 2009). Research by ourselves, colleagues, and other groups has elucidated the mechanisms and identified the selective pressures that underlie *A. marginale* genetic strain structure. Understanding both the mechanisms and pressures shaping bacterial strain structure, using *A. marginale* as a model, is relevant to how pathogen strains emerge, predominate, and recede—shifts that are reflected in patterns of disease incidence and severity.

The centrality of persistent infection

The capacity to establish persistent infection in immunocompetent hosts is the central force shaping the *A. marginale* genome. Following infection of a mammalian host (*A. marginale* naturally infects both wild and domestic ruminants), bacteremia can exceed 10^9 organisms per ml in the acute phase followed by life-long persistence (reviewed in Palmer *et al.*, 2000). This persistence is characterized by cyclic waves of bacteremia between 10^2–10^7 organisms per ml of blood (French *et al.*, 1998). Persistent bacteremia is critical for ongoing transmission as infection is non-contagious and requires tick feeding to acquire and, following subsequent feeding on a susceptible host, transmit *A. marginale*. As the potential for tick feeding is episodic due to factors such seasonal and climatic fluctuations, persistence maximizes transmission potential. Importantly, *A. marginale* is maintained only in the blood and blood-rich organs, most notably the spleen, and is thus continually exposed to immune effectors but, at the same time, occupies a niche protected from direct competition with other microbes. This is reflected in the genome as there are no apparent microcins or bacteriocins to defend against bacterial competitors nor evidence of lateral gene transfer associated with sharing a common microbial niche (Brayton *et al.*, 2005). In contrast, the striking feature of the genome is the diversity of alleles encoding the immunodominant outer membrane proteins, designated Msp2 and Msp3 (Brayton *et al.*, 2001, 2005). This allelic diversity underlies the mechanism of persistent infection.

Antigenic variation and persistence in the individual host

Unlike other outer membrane proteins (Omps) in *A. marginale*, Msp2 and Msp3 each have a single expression site but multiple alleles distributed throughout the chromosome (Eid *et al.*, 1996; Alleman *et al.*, 1997; Barbet *et al.*, 2000; Brayton *et al.* 2001, 2005). The alleles themselves are in "silent" loci, lacking promoter and other regulatory elements, and encode only the central domains of Msp2 and Msp3 flanked by 5' and 3' sequences that are identical with but truncated relative to the sequences in the single expression sites (Brayton *et al.*, 2001, 2005). This structure facilitates efficient recombination in which an allele from a "silent" locus is inserted into the expression site. The individual alleles encode unique central Msp2/Msp3 extracellular domains; these domains are highly immunodominant both within the full-length Msp2/Msp3 proteins and among all Omps (McGuire *et al.*, 1991; French *et al.*, 1999; Abbott *et al.*, 2004). Consequently, induction of antibody against Msp2 and Msp3, which becomes detectable only after the bacteremic wave peaks, results in clearance of only the specific variant population; recombination of a new allele from a silent locus into the expression site via gene conversion results in expression of new variants with unique central domains (termed the hypervariable region, HVR), and immune escape (French *et al.*, 1998, 1999; Brayton *et al.*, 2003; Meeus *et al.*, 2003). Over time this results in induction of a broad population of antibodies recognizing the full repertoire of encoded Msp2/Msp3 variants.

Importantly, the capacity of *A. marginale* for life-long persistence could not be met solely by the limited number of individual *msp2/msp3* alleles, fewer than 10 each per genome (Brayton *et al.*, 2001; Dark *et al.*, 2009; Herndon *et al.*, 2010). The additional capacity is generated by a second-level mechanism in which only an oligonucleotide segment of an individual allele, rather than the entire allele, is recombined into the expression site (Barbet *et al.*, 2000; Brayton *et al.*, 2001, 2002). This process of segmental gene conversion (illustrated in Palmer *et al.*, 2009) results in a unique variant represented only in the expression site and not by any single allele in its silent locus. This "mix and match" strategy in which an expressed HVR can be derived from as many as four different donor alleles, combined with the intrinsic diversity of the encoded HVRs among alleles, provides the hundreds to thousands of variants required for long-term persistence (Brayton *et al.*, 2001; Futse *et al.*, 2005). The similarities to the diversification mechanism for mammalian immunoglobulin, the effector from which *A. marginale* needs to escape, are striking and point to commonality among biological mechanisms (Kato et al., 2012).

Variant structure and bacterial fitness

The earliest phases of bacteremia, especially the first two months following initial infection, are characterized by "simple" variants—those derived from a single recombination event of the complete allele (or a single segment) from its silent locus (Brayton *et al.*, 2003; Futse *et al.*, 2005). However as the immune response evolves over time, the variant population is increasingly characterized by "complex" variants generated by multiple segmental gene conversions utilizing multiple individual alleles (Futse *et al.*, 2005). That this progression is driven by the mammalian immune response is evident when this selective pressure is removed: inoculation of a population of complex variants into an immunologically naïve host results in rapid reversion to an acute bacteremia characterized by simple variants (Palmer *et al.*, 2007). More relevant to naturally occurring selective pressures, simple variants predominate very early after ticks, the natural transmission vector, feed on animals with a persistent bacteremia of complex variants (Löhr *et al.*, 2002). Free of the selective pressure of the mammalian immune response, simple variants rapidly emerge in the tick midgut. These simple variants are maintained in the tick and then transmitted onward to a naïve host, when again simple variants form the primary bacteremia and then undergo progression to complex variants concomitant with development of variant-specific antibodies (Palmer *et al.*, 2007).

Predominance of simple variants in the absence of the strong selective pressure of the mammalian immune response may be attributable to two different mechanisms. The first is that bacteria expressing the simple variants themselves have a significant intrinsic growth fitness advantage. Consistent with a marked fitness advantage is the significantly higher bacteremia level, on 2–6 $\log_{10}$ more bacteria per ml, during early phases of mammalian infection when simple variants predominate as compared to periods of complex variant predominance (Palmer *et al.*, 2000; Futse *et al.*, 2005). Unlike the complex variants, which are only transiently maintained in the expression site and thus subject only to short-term selection, the simple variants are encoded within the genome itself and presumably selected over a longer term for growth fitness as well as the capacity to evade immune recognition. Detailed analysis of allelic usage during infection clearly indicated that selection is at the level of the allele (and thus the encoded simple variant) rather than due to locus structure or position (Futse *et al.*, 2009). The second, non-mutually exclusive, explanation for simple variant predominance is that the recombination mechanism itself favors insertion of a complete allele. The 5' and 3' regions flanking the variable domains are identical among the alleles and the expression site, thus providing extensive homology, which has been shown to be a determinant of recombination frequency (Brayton *et al.*, 2001). In contrast, segmental gene conversion is tethered by homology at the 5' or 3' end, but not both, and with only

minimal homology at the internal recombination site (Futse *et al.*, 2005). Comparative data from *Trypanosoma brucei* supports recombinatorial advantage for *vsg* sequences, whether complete or segmental, based on homology (Barnes & McCulloch, 2007; Hall *et al.*, 2013). As noted above, the two mechanisms are not mutually exclusive—a higher recombination rate combined with a fitness advantage may underlie the rapid switch to predominance of simple variants in the absence of the selective pressure of the immune response (Palmer *et al.*, 2007).

Strain-specific allelic diversity and superinfection

A. marginale has a "closed-core" genome: gene content is highly conserved among all *sensu stricto* strains and sequencing additional strains has not revealed additional genes (Dark *et al.*, 2009). Notably however, the genetic differences among strains are concentrated in the *msp2* and *msp3* alleles: pairwise strain comparisons revealed markedly greater diversity in the *msp2* and *msp3* alleles ($p<0.0006$) as compared to core housekeeping genes (Futse *et al.*, 2008). This level of diversity was unexpected and led to hypotheses regarding the evolutionary basis for the high degree of allelic diversity among strains. The first explanation was simply that there is more than one set of alleles that allow immune evasion yet retain growth fitness, essentially more than one evolutionary pathway. However pair-wise examination of the allelic repertoires of multiple strains revealed that the differences in encoded HVRs between strains—and notably, co-circulating strains—were as great as those within a strain (Futse *et al.*, 2008). Given that the selective pressure for unique HVRs encoded by the allelic repertoire within a strain is to be antigenically distinct, this suggested that there was a similar pressure for strains to be able to express an antigenically unique repertoire as compared to co-circulating strains (Futse *et al.*, 2008). This hypothesis has a basis in the paradoxical observation that while infected animals cannot clear their own persistent bacteremia, they are resistant to a new infection with the same strain. The hypothesis was tested in a series of experiments in which animals were infected with strain A and allowed to progress through multiple bacteremic cycles, ensuring exposure to and induction of antibody against the complete repertoire of HVRs encoded by the strain A alleles, and were then challenged with strain B after >12 months of persistence (Futse *et al.*, 2008). Challenge was by feeding infected ticks, representing a natural mode of transmission and, critically, delivering simple variants. When strain B had a completely distinct allelic repertoire, the allelic source of the expressed HVR at the time of infection was essentially random, consistent with any of the simple variants being able to escape the pre-existing immune response against strain A (Futse *et al.*, 2008). However, if strain B differed by only a single allele—the remainder of the alleles being shared between the two strains—only the unique HVR encoded by the differing allele was expressed on the bacteria able to establish infection (Futse *et al.*, 2008). Thus the *msp2/msp3* allelic content can be seen as a deterministic strain characteristic, responsible for the capacity to evade the existing strain-specific immunity.

This principle of "strain superinfection", the ability of a second strain to establish infection in a host that has already been infected and mounted an immune response to a primary strain of the same pathogen, has a clear basis in the epidemiology of *A. marginale* infection. In tropical and subtropical regions where infection is highly endemic, most animals become infected early in life and, although persistently infected, develop a broad population immunity to a predominant strain, the theoretical strain A. While persistent infection allows for vector ticks to acquire the strain with remarkable efficiency (Scoles *et al.*, 2005), the presence of broad population immunity would leave very few suitable hosts for onward transmission of strain A. In contrast, emergence of a strain B, bearing at least one allele encoding a sufficient antigenically distinct HVR, could either infect the relatively few truly naïve animals or superinfect those already infected with strain A (or any other distinct

strain). Initial support for this scenario came from the observation that when animals carried more than a single strain, each strain had a unique allelic repertoire (Palmer *et al.*, 2004; Rodriguez *et al.*, 2005). More conclusively, when infected animals in regions with high prevalence (and thus few naïve individuals) were compared to those in regions of low prevalence (with a majority of naïve individuals), the *A. marginale* populations in the highly endemic regions had a dramatically and significantly greater diversity of variant alleles and, specifically, the encoded HVRs (Ueti *et al.*, 2012).

How frequently or rapidly this allelic diversification occurs, even under strong selective pressure in highly endemic regions, is yet unresolved. However, genome analysis has provided clues as to how diversity may be generated with the hypothesized mechanism being gene duplication followed by either mutation or mismatch repair. The presence of two identical alleles has been identified in multiple strains, suggestive of a gene duplication event (Brayton *et al.*, 2005; Dark *et al.*, 2009). This process would provide the existing allele required to generate sufficient variants to maintain persistent infection and provide a second copy on which introduced genetic changes could be "tested" for competitive advantage. There are at least three levels of evidence for introduction of genetic change in a duplicated allele. The first is that two alleles within a strain may differ by only a single HVR oligonucleotide segment, consistent with recombination into a duplicated allele (Dark *et al.*, 2009). Secondly, examination of over 1,300 expressed variants revealed that approximately 1% of the variants contained unique sequences not present in any pre-existing allele; these variants differed by small insertions or deletions but maintaining the reading frame required for a full-length protein (Futse *et al.*, 2005). These are presumed to have arisen via mismatch repair during recombination, as the insertions have not been identified elsewhere in the genome, and represent *de novo* introduced sequence. The third source of allelic diversification, only very recently uncovered via next-generation sequencing, appears to be via individual base mutations, again maintaining the reading frame and providing a template upon which selection can act. How these mechanisms, individually or collectively, work in allelic diversification represents a gap in knowledge broadly applicable to understanding pathogen evolution and strain emergence.

Transmission fitness and the limits to strain chaos

The above scenario of strong selective pressure for strain diversification, if unchecked, would result in "strain chaos" in which highly endemic regions would be characterized by dozens to hundreds of competing strains and with strains bearing alleles so divergent as to be unrecognizable as *msp2/msp3*. However, neither predicted result is observed in endemic regions. In contrast, all studies to date have identified dominant strains in endemic regions (Palmer *et al.*, 2001, 2004; Ueti *et al.*, 2012). Where infection prevalence and, thus population immunity, is low, there may be only a single circulating strain (Palmer *et al.*, 2001; Ueti *et al.*, 2012). However even in regions where infection prevalence is high and, accordingly, superinfection is common, the number of strains is limited and there are clearly predominant strains (Palmer *et al.*, 2004; de la Fuente *et al.*, 2004; Ruybal *et al.*, 2009; Ueti *et al.*, 2012). There is supporting evidence that this predominance reflects competitive transmission advantage effected by greater fitness within the natural tick vector. In contrast to fitness within mammalian reservoir hosts, where there is no clear difference among strains in either the amplitude or duration of the bacteremic cycles, the ability to establish infection in the vector, bacterial levels in the midgut and salivary glands (the two primary tick organs where replication occurs), and levels secreted into the saliva at the time of transmission can differ dramatically (Ueti *et al.*, 2007, 2009). We hypothesize that specific Msp2/3 proteins promote better growth within the tick and thus strains bearing these alleles have a competitive advantage. When endemicity is relatively low and there are a large number of naïve hosts available for transmission, a strain bearing these alleles may be the

sole circulating strain (Palmer *et al.*, 2001; Ueti *et al.*, 2012). However under conditions of high endemicity and population immunity, there is a balancing set of competitive fitness characteristics: colonization and growth within the tick vector remains important but strains with lower fitness levels can be maintained in the population given their ability to superinfect animals already carrying the primary strain (Palmer *et al.*, 2004; Galletti et al., 2009; Ueti *et al.*, 2012). Below a certain level of transmission fitness, regardless of how antigenically diverse the variant repertoire may be, a strain is no longer maintained. Thus the strain structure is shaped by competing selective pressures, which themselves may be ecologically dynamic based on the levels of endemicity and population immunity, vector prevalence, and availability of reservoir mammalian hosts (Estrada-Peña *et al.*, 2009).

What underlies the differential fitness of the strains at the molecular level? Structural modeling of simple Msp2 variants suggests function as porin, consistent with its outer membrane location and experimental evidence of porin function for the closely related *A. phagocytophilum* Msp2 (Huang *et al.*, 2007). We hypothesize that specific alleles encode simple variants with superior porin function, either in diffusion rate or breadth of substrates, and thus have a competitive growth advantage within the intracellular niches within the tick midgut and salivary gland. Diversification provides the structural differences required for immune escape and persistence in an individual host and, by allowing strain superinfection, at the population level. However this diversification for antigenic variation is predicted to result in a growth fitness cost due to diminished porin function. This hypothesis is supported by the preferential expression of specific alleles during replication in the tick vector (Rurangirwa *et al.*, 1999; Palmer *et al.*, 2007) but requires additional experimental verification correlating transmission fitness with porin function.

Conclusion

Understanding pathogen strain structure and how it changes, continually or episodically, requires knowledge of the selective pressures acting upon the pathogen, its hosts, and transmission routes. *A. marginale*, as a model highly antigenically variant pathogen, illustrates the trade-off among its primary selective pressures, immune escape for persistence—both within the individual and within the population—and transmission fitness. Determining how the antigenic variant repertoire and growth fitness intersect with population immunity and transmission fitness to shape pathogen strain structure over space and time, an approach which will require iterative combinations of modeling and field testing, will address key knowledge gaps in how pathogens shift disease patterns.

Acknowledgments

The authors thank Shira Broschat, Minerva Camacho, Jacqueline Castañeda-Ortiz, Eduardo Vallejo Esquerra, James Futse, Telmo Graça, David Herndon, Svetlana Lockwood, Susan Noh, and Varda Shkap, Massaro Ueti, among other colleagues and collaborators, for valuable discussions. The original work in the authors' laboratories was supported by NIH R37 AI044005, BARD 4187-09C, the Wellcome Trust grant GM075800M, and USDA-ARS 5348-3200-033.

References

Abbott JR, Palmer GH, Howard CJ, Hope JC, Brown WC. *Anaplasma marginale* major surface protein 2 $CD4^+$ T-cell epitopes are evenly distributed in conserved and hypervariable regions (HVR), whereas linear B-cell epitopes are predominantly located in the HVR. Infect Immun. 2004; 72:7360–7366. [PubMed: 15557669]

Barbet AF, Lundgren A, Yi J, Rurangirwa FR, Palmer GH. Antigenic variation of *Anaplasma marginale* by expression of MSP2 sequence mosaics. Infect Immun. 2000; 68:6133–6138. [PubMed: 11035716]

Barnes RL, McCulloch R. *Trypanosoma brucei* homologous recombination is dependent on substrate length and homology, though displays a differential dependence on mismatch repair as substrate length decreases. Nucleic Acids Res. 2007; 35:3478–3493. [PubMed: 17478508]

Brayton KA, Knowles DP, McGuire TC, Palmer GH. Efficient use of a small genome to generate antigenic diversity in tick-borne ehrlichial pathogens. Proc Natl Acad Sci USA. 2001; 98:4130–4135. [PubMed: 11274438]

Brayton KA, Palmer GH, Lundgren A, Yi J, Barbet AF. Antigenic variation of *Anaplasma marginale msp2* occurs by combinatorial gene conversion. Mol Microbiol. 2002; 43:1151–1159. [PubMed: 11918803]

Brayton KA, Meeus PFM, Barbet AF, Palmer GH. Simultaneous variation of the immunodominant outer membrane proteins MSP2 and MSP3 during *Anaplasma marginale* persistence *in vivo*. Infect Immun. 2003; 71:6627–6632. [PubMed: 14573687]

Brayton KA, Kappmeyer LS, Herndon DR, Dark MJ, Tibbals DL, Palmer GH, et al. Complete genome sequencing of *Anaplasma marginale* reveals that the surface is skewed to two superfamilies of outer membrane proteins. Proc Natl Acad Sci USA. 2005; 102:844–849. [PubMed: 15618402]

Dark MJ, Herndon DR, Kappmeyer LS, Gonzales MP, Nordeen E, Palmer GH, et al. Conservation in the face of diversity: multistrain analysis of an intracellular bacterium. BMC Genomics. 2009; 10:16–28. [PubMed: 19134224]

De la Fuente J, Passos LM, Van Den Bussche RA, Ribeiro MF, Facury-Filho EJ, Kocan KM. Genetic diversity and molecular phylogeny of *Anaplasma marginale* isolates from Minas Gerais, Brazil. Vet Parasitol. 2004; 121:307–316. [PubMed: 15135871]

Eid G, French DM, Lundgren A, Barbet AF, McElwain TF, Palmer GH. Expression of Major Surface Protein 2 antigenic variants during acute *Anaplasma marginale* rickettsemia. Infect Immun. 1996; 64:836–841. [PubMed: 8641789]

Estrada-Pena A, Naranjo V, Acevedo-Whitehouse K, Mangold AJ, Kocan KM, de la Fuente J. Phylogeographic analysis reveals association of tick-borne pathogen, *Anaplasma marginale* MSP1a sequences with ecological traits affecting tick vector performance. BMC Biol. 2009; 7:57. [PubMed: 19723295]

Fraser-Liggett CM. Insights on biology and evolution from microbial genome sequencing. Genome Res. 2005; 15:1603–1610. [PubMed: 16339357]

French DF, McElwain TF, McGuire TC, Palmer GH. Expression of *Anaplasma marginale* Major Surface Protein 2 variants during persistent cyclic rickettsemia. Infect Immun. 1998; 66:1200–1207. [PubMed: 9488414]

French DM, Brown WC, Palmer GH. Emergence of *Anaplasma marginale* antigenic variants during persistent rickettsemia. Infect Immun. 1999; 67:5834–5840. [PubMed: 10531237]

Futse JE, Brayton KA, Knowles DP, Palmer GH. Structural basis for segmental gene conversion in generation of *Anaplasma marginale* outer membrane protein variants. Mol Microbiol. 2005; 57:212–221. [PubMed: 15948961]

Futse JE, Brayton KA, Dark MJ, Knowles DP, Palmer GH. Superinfection as a driver of genomic diversification in antigenically variant pathogens. Proc Natl Acad Sci USA. 2008; 105:2123–2127. [PubMed: 18252822]

Futse JE, Brayton KA, Nydam SD, Palmer GH. Generation of antigenic variants by gene conversion: Evidence for recombination fitness selection at the locus level in *Anaplasma marginale*. Infect Immun. 2009; 77:3181–3187. [PubMed: 19487473]

Galletti MFBM, Ueti MW, Knowles DP, Brayton KA, Palmer GH. Independence of high and low transmission efficiency *Anaplasma marginale* strains in the tick vector following simultaneous acquisition by feeding on a superinfected mammalian reservoir host. Infect Immun. 2009; 77:1459–1464. [PubMed: 19188360]

Hall JPJ, Wang H, Barry JD. Mosaic VSGs and the scope of *Trypanosoma brucei* antigenic variation. PLoS Pathogens. 2013; 9 e1003502.

Herndon DR, Palmer GH, Shkap V, Knowles DP, Brayton KA. Complete genome sequence of *Anaplasma marginale ss. centrale*. J Bacteriol. 2010; 192:379–380. [PubMed: 19854912]

Huang H, Wang X, Kikuchi T, Kumagai Y, Rikihisa Y. Porin activity of *Anaplasma phagocytophilum* outer membrane fraction and purified P44. J Bacteriol. 2007; 189:1998–2006. [PubMed: 17172334]

Kato L, Stanlie A, Begum NA, Kobayashi M, Aida M, Honjo T. An evolutionary view of the mechanism for immune and genome diversity. J Immunol. 2012; 188:3559–3566. [PubMed: 22492685]

Keim P, Gruendike JM, Klevytska AM, Schupp JM, Challacombe J, Okinaka R. The genome and variation of *Bacillus anthracis*. Mol Aspects Med. 2009; 30:397–405. [PubMed: 19729033]

Kocan KM, de la Fuente J, Blouin EF, Coetzee JF, Ewing SA. The natural history of *Anaplasma marginale*. Vet Parasitol. 2010; 167:95–107. [PubMed: 19811876]

Löhr CV, Rurangirwa FR, McElwain TF, Stiller D, Palmer GH. Specific expression of *Anaplasma marginale* major surface protein 2 salivary gland variants occurs in the midgut and is an early event during tick transmission. Infect Immun. 2002; 70:114–120. [PubMed: 11748171]

McGuire TC, Davis WC, Brassfield AL, McElwain TF, Palmer GH. Identification of *Anaplasma marginale* long-term carrier cattle by detection of serum antibody to isolated MSP-3. J Clin Microbiol. 1991; 29:788–793. [PubMed: 1890178]

Meeus PFM, Brayton KA, Palmer GH, Barbet AF. Conservation of a gene conversion mechanism in two distantly related paralogues of *Anaplasma marginale*. Mol Microbiol. 2003; 47:633–643. [PubMed: 12535066]

Palmer GH, Brown WC, Rurangirwa FR. Antigenic variation in the persistence and transmission of the ehrlichia *Anaplasma marginale*. Microb Infect. 2000; 2:167–176.

Palmer GH, Rurangirwa FR, McElwain TF. Strain composition of the ehrlichia *Anaplasma marginale* within persistently infected cattle, a mammalian reservoir for tick-transmission. J Clin Microbiol. 2001; 39:631–635. [PubMed: 11158120]

Palmer GH, Knowles DP, Rodriguez JL, Gnad DP, Hollis LC, Marston T, Brayton KA. Stochastic transmission of multiple genotypically distinct *Anaplasma marginale* strains within an endemic herd. J Clin Microbiol. 2004; 42:5381–5384. [PubMed: 15528749]

Palmer GH, Futse JE, Leverich CK, Knowles DP, Rurangirwa FR, Brayton KA. Selection for simple Msp2 variants during *Anaplasma marginale* transmission to immunologically naïve animals. Infect Immun. 2007; 75:1502–1506. [PubMed: 17178787]

Palmer GH, Bankhead T, Lukehart SA. Nothing is permanent but change: Antigenic variation in persistent bacterial pathogens. Cell Microbiol. 2009; 11:1697–1705. [PubMed: 19709057]

Rodriguez JL, Palmer GH, Knowles DP, Brayton KA. Distinctly different *msp2* pseudogene repertoires in *Anaplasma marginale* strains that are capable of superinfection. Gene. 2005; 361:127–132. [PubMed: 16202540]

Rurangirwa FR, Stiller DS, French DM, Palmer GH. Restriction of Major Surface Protein 2 (MSP2) variants during tick transmission of the ehrlichiae *Anaplasma marginale*. Proc Natl Acad Sci USA. 1999; 96:3171–3176. [PubMed: 10077656]

Ruybal P, Moretta R, Perez A, Petrigh R, Zimmer P, Alcarez E, et al. Genetic diversity of *Anaplasma marginale* in Argentina. Vet Parasitol. 2009; 162:176–180. [PubMed: 19285808]

Scoles GA, Broce AB, Lysyk TJ, Palmer GH. Relative efficiency of biological transmission of *Anaplasma marginale* (Rickettsiales: Anaplasmataceae) by *Dermacentor andersoni* Stiles (Acari: Ixodidae) compared to mechanical transmission by the stable fly, *Stomoxys calcitrans* (L.) (Diptera: Muscidae). J Med Entomol. 2005; 42:668–675. [PubMed: 16119558]

Servant-Delmas A, Lefrere JJ, Morinet F, Pillet S. Advances in human B19 erythrovirus biology. J Virol. 2010; 84:9658–9665. [PubMed: 20631151]

Ueti MW, Reagan JO, Knowles DP, Scoles GA, Shkap V, Palmer GH. Identification of midgut and salivary gland as specific and distinct barriers to efficient tick-borne transmission of *Anaplasma marginale*. Infect Immun. 2007; 75:2959–2964. [PubMed: 17420231]

Ueti MW, Knowles DP, Davitt CM, Scoles GA, Baszler TV, Palmer GH. Quantitative differences in salivary pathogen load during tick transmission underlie strain-specific variation in transmission efficiency of *Anaplasma marginale*. Infect Immun. 2009; 77:70–75. [PubMed: 18955472]

Ueti MW, Tan Y, Broschat SL, Castañeda Ortiz EJ, Camacho-Nuez M, Mosqueda JJ, et al. Expansion of variant diversity associated with high prevalence of pathogen strain superinfection under conditions of natural transmission. Infect Immun. 2012; 80:2354–2360. [PubMed: 22585962]

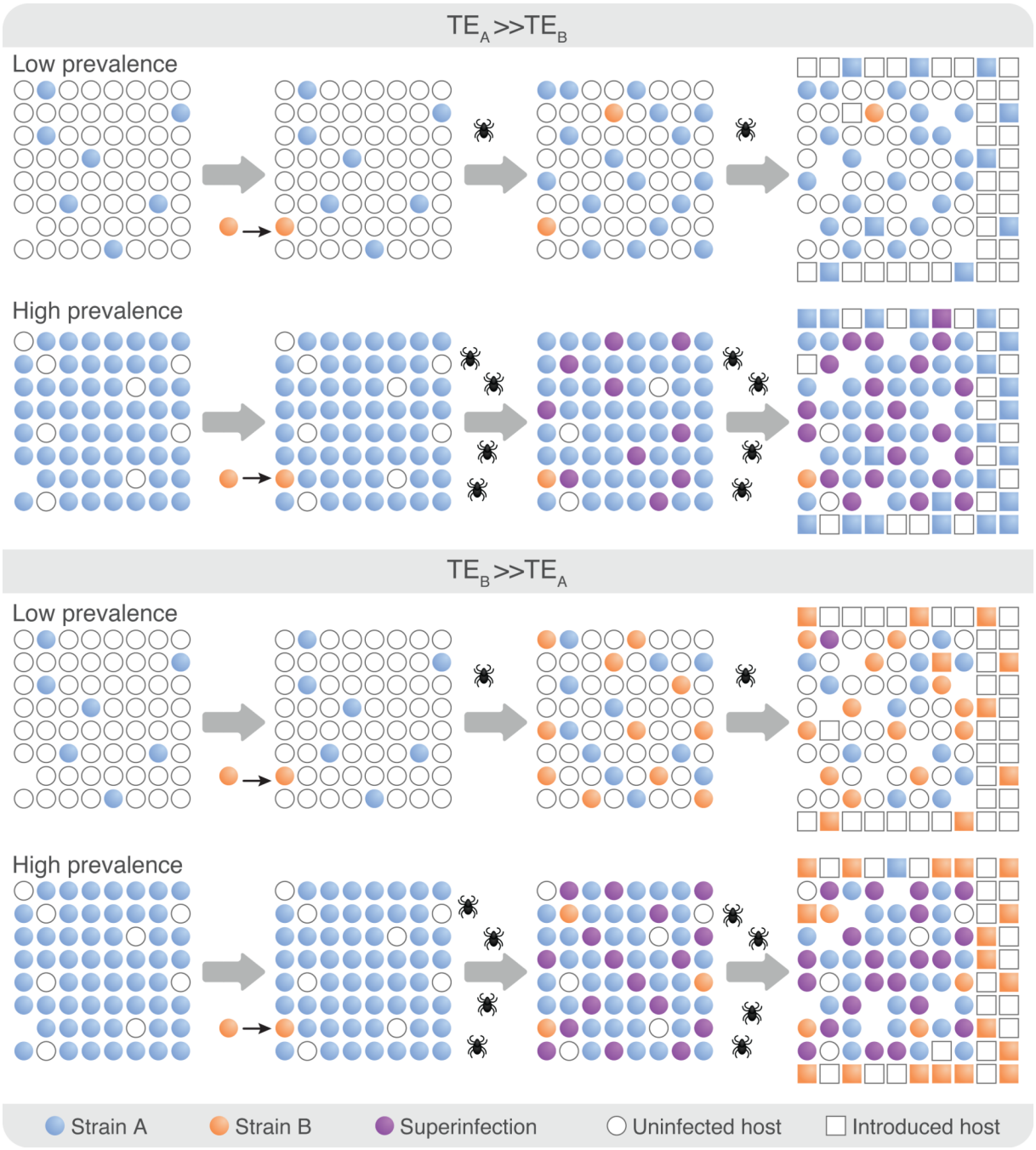

Figure 1. The interplay of pathogen strain-specific transmission efficiency and population immunity defines unique patterns of strain structure

Circles indicate the existing animal population at T_0: white represent uninfected and immunologically naïve hosts; blue represents hosts carrying strain A; orange, strain B; and purple represents hosts superinfected with strains A and B. Squares represent individual hosts introduced to the population by birth or immigration.

Top Panel: The intrinsic transmission efficiency is greater for strain A than strain B ($TE_A >> TE_B$). Under conditions of low prevalence of infection (and hence low population immunity) and low vector presence, strain A is predominant. Following introduction of strain B, its transmission is at a strong disadvantage and there is minimal selective pressure for strain B

superinfection. Consequently, strain A predominance is maintained over time. Under conditions of high prevalence of infection (and high population immunity) and abundant vector presence, strain A is predominant but there is strong selective pressure for strain B superinfection. Strain A transmission is favored for newly introduced naïve hosts and thus remains predominant but accompanied by prevalent superinfection.

Bottom Panel: The intrinsic transmission efficiency is greater for strain B than strain A ($TE_B >> TE_A$). Under conditions of low prevalence of infection (and low population immunity) and low vector presence, the introduction of strain B results in preferential transmission and strain B replaces strain A over time. Under conditions of high prevalence of infection (and high population immunity) and abundant vector presence, the introduction of strain B results in high levels of superinfection. Strain B transmission is favored for newly introduced naïve hosts and over time becomes the predominant strain but in the face of widespread superinfection.