International Institute for
Applied Systems Analysis
Schlossplatz 1
A-2361 Laxenburg, Austria

Tel: +43 2236 807 342
Fax: +43 2236 71313
E-mail: publications@iiasa.ac.at
Web: www.iiasa.ac.at

Interim Report **IR-09-069**

Antigenic distance and cross-immunity, invisibility and coexistence of pathogen strains in an epidemiological model with discrete antigenic space

Ben Adams (b.adams@bath.ac.uk)
Akira Sasaki (sasaki_akira@soken.ac.jp)

Approved by

Ulf Dieckmann
Leader, Evolution and Ecology Program

June 2010

Contents

Antigenic distance and cross-immunity, invasibility and coexistence of pathogen strains in an epidemiological model with discrete antigenic space

Ben Adams[1,2] and Akira Sasaki[3,4,5]

1. Department of Biology, Kyushu University, Fukuoka 812-8581, Japan
2. Department of Mathematical Sciences, University of Bath, Bath BA2 7AY, UK
3. Department of Evolutionary Study of Biosystems, The Graduate University for Advanced Studies (SOKENDAI), Hayama, Kanagawa 240-0193, Japan
4. Evolution and Ecology Program, International Institute for Applied Systems Analysis, A-2361 Laxenburg, Austria
5. PRESTO, Japan Science and Technology Agency, 4-1-8 Honcho Kawaguchi, Saitama, Japan.

January 12, 2010

Abstract

In models of pathogen interaction and evolution discrete genotypes in the form of bitstrings may be mapped to points in a discrete phenotype space based on similarity in antigenic structure. Cross-immunity between strains, that is the reduction in susceptibility to strain A conferred to a host by infection with strain B, can then be defined for pairs of points in the antigenic space by a specified function. Analysis of an SIR type model shows that, if two strains are at equilibrium, the shape of the cross-immunity function has a strong influence on the invasion and coexistence of a third strain and, consequently, the expected evolutionary pathway. A function that is constant except for discontinuities at the end points is expected to result in the accumulation of diversity until a pair of discordant strains occurs that can, depending on parameter values, exclude all other strains. For a function of the form $f(h) = h^q$ invasion and coexistence is always possible if $q \leq 1$ and little antigenic structure is expected in the pathogen population. However, if $q > 1$ invasion and coexistence may be impossible, depending on parameter values, and the pathogen population is expected to show significant antigenic structuring. In addition to illuminating the role of cross-immunity in pathogen evolution, this analysis indicates that the choice of cross-immunity function, the representation of immunity acquired from multiple previous infections and the number of elements used to characterize the antigenic space must be carefully considered in the development and interpretation of more sophisticated models of pathogen dynamics and evolution.

1. Introduction

Pathogens often occur as a variety of strains with slightly modified antigenic structures and the antibodies that bind most efficiently with one strain may be partially, or completely, ineffective against another strain (Alberts, 2002; Janeway and Janeway, 1999). The prevalence and spread of a pathogen strain is strongly influenced by the immunological state of the host population. For any given pathogen strain this immune landscape is determined by the current and historical prevalence of all other variants and the immune interaction with antibodies raised in response to those variants. Thus epidemiology, cross-immunity and pathogen evolution are tightly interwoven (Adams et al., 2006; Dieckmann, 2002). Central to any study of the relationship between antigenic evolution and epidemiology is the concept of an antigenic space. One phenotypic characteristic of a pathogen strain is its antigenic surface structure. In this article, this phenotype is conceptualized as a point in an antigenic space in which the antigenic distance between strains is a measure of the similarity in their surface structures. Another phenotypic characteristic of a pathogen strain is the cross-reaction between its antigen and antibodies raised in response to infection with another strain. This partial cross-immunity between strains is expected to be a function of their antigenic distance, and can also be used to define antigenic distance directly. Here we consider a discrete antigenic space directly derived from a bitstring, or locus-allele, genotype construct, and examine how the function relating antigenic distance in this space to partial cross-immunity affects invasion, coexistence and, consequently, the evolutionary pathway.

The theory of pathogen evolution is well developed, particularly with respect to the evolution of virulence (Anderson and May, 1982; Anderson and May, 1991; Ewald, 1994). Studies are often based on pairwise invasion analysis, which considers a population in which one strain is at equilibrium and asks when a second mutant strain can invade and replace the existing strain (Dieckmann, 2002). If two strains have identical epidemiological characteristics and cross-immunity is symmetric then many simple models predict coexistence. Here, therefore, we progress directly to considering two strains at equilibrium and examining when a third strain can invade and whether it will exclude either of the existing strains. Where possible we then go on to consider invasion and coexistence when three or more strains already coexist.

It has been shown that, in a one dimensional continuous antigenic space, if the function relating antigenic distance to cross-immunity is linear or strictly concave then the two strain equilibrium can always be invaded by a third strain and all three strains will coexist. However, if the function is strictly convex then invasion, and coexistence, is not always possible (Adams and Sasaki 2007). In the context of the discrete antigenic space employed in this article we consider a discontinuous function relating antigenic distance to cross-immunity, in which cross-immunity can take one of only three values according to whether the antigenic structures of the strains are identical, entirely distinct or related (Gupta et al. 1996, Gupta et al. 1998, Ferguson and Andreasen 2002) and continuously defined functions of the form $f(h) = h^q$ where h is antigenic distance and q is a positive real number. With the discontinuous cross-immunity function an equilibrium composed of two similar strains can always be invaded by a third strain but an equilibrium composed of two discordant strains may be resistant to invasion, depending on the degree of cross-immunity and the basic

reproductive number. With the continuous cross-immunity function, invasion is always possible if $q \leq 1$ but otherwise may be impossible depending on the value of q and the basic reproductive number. By carefully examining the mechanisms that determine whether invasion and coexistence are possible we show that the shape of the cross-immunity function is critical because of its influence on secondary infections while the basic reproductive number particularly influences the impact of tertiary, or subsequent, infections. The number of elements used in the bitstring characterization of the antigenic configuration, which implicitly defines the antigenic change associated with a single mutation, is also important as it influences when branching can occur during a sequence of point mutations. In addition to improving our understanding of how cross-immunity influences pathogen evolution and diversity, this work highlights important factors to consider in the development and interpretation of more complex simulation models of this type of system.

2. Model description

2.1 SIR framework

SIR models (Anderson and May, 1991; Castillo-Chavez, 2002) with multiple pathogen strains interacting by cross-immunity group the host population according to their current infection status and infection history. Since recording all possible infection histories in a system with N strains requires 2^N compartments, models rapidly become very large as N increases (Andreasen et al., 1997; Ferguson and Andreasen, 2002; Gomes et al., 2002), and a number of simplification have been proposed including: status based approaches, whereby an infected individual either gains complete immunity to cross-reactive strains with some probability or else remains completely susceptible; the assumption that immunity reduces infectivity rather than susceptibility; the assumption that many of the strains have identical immune interactions; approximations for the size of host groups that have experienced multiple previous infections (Calvez et al., 2005; Gog and Grenfell, 2002; Gupta et al., 1998; Gupta et al., 1996; Kryazhimskiy et al., 2007). However, since we will focus on situations with just a few strains, the original formulation can be used: the host population is grouped according to all possible infection histories and immunity acts on susceptibility by reducing the probability of subsequent infection.

A complete description of this model can be found in several other places (Andreasen et al., 1997; Ferguson and Andreasen, 2002; Gomes et al., 2002) so only a brief review is provided here. Given a set K of N strains, the population is divided into 2^N compartments S_J where J are all the subsets of K including $\varnothing$ and K. Each compartment S_J represents the proportion of the host population currently or previously infected with all the strains in J. A further N equations record the forces of infection λ_i of strains $i = 1...N$. Cross-immunity is expressed by σ_J^i, where $0 \leq \sigma_J^i \leq 1$ and represents reduction in susceptibility to strain i conferred by infection history J. When $\sigma_J^i = 0$ cross-immunity is perfect. When $\sigma_J^i = 1$ cross-immunity is absent. The host population is assumed to be well mixed and the rate at which hosts with infection history J are infected with strain $i \notin J$ is given by $\sigma_J^i S_J \lambda_i$. Birth and death rates are constant and equal so the total population size is also constant. Infected individuals recover at a constant rate. A non-dimensional form of the equations is used (Andreasen et al., 1997) and so parameters for the birth, death and recovery rates do not appear directly but are compounded into parameters e, the ratio of the duration of infection and life-expectancy, and r which would be the basic reproductive number if

just one strain was introduced into a naïve population. In order to focus on the impact of cross-immunity, these epidemiological parameters, e and r, are assumed to be the same for all strains. Due to non-dimensionalization the total population size is always 1. Differential equations describing the model are:

$$\begin{aligned}
\dot{S}_{\varnothing} &= e - S_{\varnothing}\sum_{i\in K}\lambda_i - eS_{\varnothing} \\
\dot{S}_J &= \sum_{i\in J}\sigma^i_{J/i}\lambda_i S_{J/i} - S_J\sum_{j\notin J}\sigma^j_J\lambda_j - eS_J \\
\dot{\lambda}_i &= \lambda_i(r\sum_{J\subseteq K/i}\sigma^i_J S_J - 1)
\end{aligned} \tag{1}$$

The analysis in this article focuses on the situation when two strains are at equilibrium and a third strain attempts to invade. Therefore it is expedient to dispense with the set theory notation and write $x = S_{\varnothing}$ (never infected, therefore susceptible to primary infection), $y_1 = S$, (currently or previously infected with strain S1, therefore susceptible to secondary infection with strain S2 or S3), $y_2 = S_{\{2\}}$ (currently or previously infected with strain S2, therefore susceptible to secondary infection with strain S1 or S3) $y_{12} = S_{\{12\}}$ (currently or previously infected with strains S1 and S2, therefore susceptible to tertiary infection with strain S3). As described in Section 2.2 the degree of cross-immunity between strain S1 and S2, previously given by the parameter $\sigma_1^2 = \sigma_2^1$,is now given by the function $f(h_{12})$ where h_{12} is antigenic distance. Homologous immunity is assumed to be perfect. Hence the two strain model is described by the equations:

$$\begin{aligned}
\dot{x} &= e(1-x) - (\lambda_1 + \lambda_2)x \\
\dot{y}_1 &= \lambda_1 x - (f(h_{12})\lambda_2 + e)y_1 \\
\dot{y}_2 &= \lambda_2 x - (f(h_{12})\lambda_1 + e)y_2 \\
\dot{y}_{12} &= f(h_{12})(\lambda_1 y_2 + \lambda_2 y_1) - ey_{12} \\
\dot{\lambda}_1 &= \lambda_1[r(x + f(h_{12})y_2) - 1] \\
\dot{\lambda}_2 &= \lambda_2[r(x + f(h_{12})y_1) - 1]
\end{aligned} \tag{2}$$

2.2 Antigenic space and cross-immunity

An individual has one genotype but many phenotypes. One pathogen phenotype is the efficiency with which its antigen binds with antibodies raised in response to infection with another phenotype. This is often measured using HI assays (Smith et al., 2004) and conceptualized as the degree of cross-immunity. Another pathogen phenotype is the physical structure of the antigen determined by the configuration of surface proteins. Antigenic distance may be defined by the similarity between two strains in either of these phenotype spaces (Gupta et al., 2006; Smith et al., 1999; Smith et al., 2004). Models often represent genotypes by sequences of alleles or, sometimes, nucleotides, codons or amino acids. Sophisticated schemes can then be used to map a genotype to its phenotypes (Koelle et al., 2006; Lapedes and Farber, 2001) but simpler approaches are often pragmatic. Some frameworks determine the similarity between genotypes, using the Hamming distance or a related measure, and map this directly to cross-immunity without explicitly considering surface structure similarity (Ferguson and Andreasen, 2002; Girvan et al., 2002; Gupta et al., 1998; Tria et al.,

2005). However, another well established approach is to map the genotype to a point in antigenic space based on surface structure and use another function to determine the immune cross-reaction from the antigenic distance (Andreasen et al., 1997; Ferguson et al., 2003; Gog and Grenfell, 2002; Gomes et al., 2002). In this article adopt the second of these frameworks.

We express the genotype as a sequence of n elements each taking one of m possible values (Andreasen et al., 1997; Calvez et al., 2005; Ferguson and Andreasen, 2002; Ferguson et al., 2003; Girvan et al., 2002; Gupta et al., 1998; Gupta et al., 1996; Recker et al., 2007). Often this is thought of as an n locus, m allele system and the genotype is rendered as a bitstring by setting $m = 2$. The similarity between the genotypes of strains S1 and S2 is expressed in terms of the normalized Hamming distance h_{12}, the proportion of positions at which the two bitstrings have different elements. The antigenic distance, in the structural phenotype space, is then taken to be identical to the normalized Hamming distance. Thus the antigenic distance is 0 when the two strains are identical and 1 when they are entirely distinct, often termed discordant. This antigenic space is high dimensional and discrete. The discreteness is further emphasized by the small number of possible antigenic distances. For example the an 8 element bitstring gives a total of $2^8 = 256$ different strains but there are only 8 possible distances between them. Analysis in this space can be difficult because even when the number of elements is small it has complicated geometry and there may be several possible antigenic locations for a strain S3 with distance h_{13} from strain S1 and h_{23} from strain S2.

Given two strains at known locations in this structural antigenic space, some function f must be used to relate the antigenic distance (i.e. structural similarity) to the degree of cross-immunity. In this article it is assumed that f is a monotonic increasing function of h with $f(0) = 0$ and $f(1) = 1$, implying that cross-protection is perfect between identical strains and entirely absent between discordant strains. In Section 4 we consider a discontinuous function $f(h_{ij}) = 0$ if $h_{ij} = 0$, η if $0 < h_{ij} < 1$ and 1 if $h_{ij} = 1$. This function classifies strains into three groups, identical, related and discordant, simplifying analysis but increasing discreteness. The same framework is often used to map bitstring genotypes directly to the cross-immunity phenotype. In Section 5 we consider continuously defined functions of the form $f(h_{ij}) = h_{ij}^{q}$ where $0 < q < \infty$. The linear form of this function $f(h_{ij}) = h_{ij}$ implies that cross-immunity is directly proportional to antigenic distance. The convex form, for example $f(h_{ij}) = h_{ij}^{2}$, implies that cross-immunity diverges slowly when antigenic distances are small but rapidly when distances are large. The concave form, for example, $f(h_{ij}) = h_{ij}^{1/2}$, implies that cross-immunity diverges slowly when antigenic distances are large but rapidly when distances are small.

To complete the model it is necessary to define how immunity from two or more previous infections affects the probability of subsequent infection by another strain. This is not well understood empirically. Many models assume that only the most closely related previous infection is effective $f(h_{12}, h_{13}) = \min\{f(h_{12}), f(h_{13})\}$ or that the combination of antibodies from previous infections is more effective than any one of them individually $f(h_{12}, h_{13}) = f(h_{12})f(h_{13})$. Further alternatives include assuming that no more than two infections can be experienced, $f(h_{12}, h_{13}) = 0$ (Cummings et al., 2005) or that only the most recent infection is effective (Andreasen and Sasaki, 2006). The minimum function will be used in this study. Other studies have found that

similar steady states arise from both the minimum and multiplicative forms, although transient dynamics may be more variable (Adams and Sasaki, 2007; Gomes et al., 2002).

3. General two strain equilibrium and invasion criterion

Given two strains, S1 and S2, with identical epidemiological parameters the coexistence equilibrium of system (2) has $y_1^* = y_2^* = y^*$, $\lambda_1^* = \lambda_2^* = \lambda^*$ and:

$$\begin{aligned} x^* &= \frac{2f(h_{12})}{(2r+1)f(h_{12}) - 2 + \Omega(h_{12})} \\ y^* &= \frac{(2r-1)f(h_{12}) - 2 + \Omega(h_{12})}{2rf(h_{12})((2r-1)f(h_{12}) + \Omega(h_{12}))} \\ y_{12}^* &= \frac{((2r-1)f(h_{12}) - 2 + \Omega(h_{12}))^2}{4rf(h_{12})((2r-1)f(h_{12}) + \Omega(h_{12}))} \\ \lambda^* &= \frac{((2r-1)f(h_{12}) - 2 + \Omega(h_{12}))e}{4f(h_{12})} \\ \text{where} \quad \Omega(h_{12}) &= \sqrt{4 - 4f(h_{12}) + f(h_{12})^2(2r-1)^2} \end{aligned} \tag{3}$$

Furthermore, since $\lambda^*[r(x^* + f(h_{12})y^*) - 1] = 0$ and $\lambda^* \neq 0$, $x^* + f(h_{12})y^* = 1/r$. Numerical investigation has suggested this coexistence equilibrium is globally stable as long as $r > 1$ and, even if the basic reproductive numbers of the two strains, say r_1 and r_2, are not identical, there is stable coexistence as long as $r_1[1 - (\alpha/(1-\alpha))(r_1 - 1)]^{-1} < r_2 < r_1[1 + \alpha(r_1 - 1)]^{-1}$ (Castillo-Chavez et al. 1989). Given strains S1 and S2 at the symmetric coexistence equilibrium another strain, S3, can invade if the rate of change of the force of infection $\lambda_3 > 0$ when the number of S3 infections is small. The force of infection λ_3 is easily found from equation (1):

$$\dot{\lambda}_3 = \lambda_3[r(x^* + f(h_{13})y_1^* + f(h_{23})y_2^* + \min\{f(h_{13}), f(h_{23}\}y_{12}^*) - 1] \tag{4}$$

So, the invasion function $\xi(h_{13}, h_{23})$ gives the invasion criterion:

$$\xi(h_{13}, h_{23}) = r[x^* + f(h_{13})y_1^* + f(h_{23})y_2^* + \min\{f(h_{13}), f(h_{23}\}y_{12}^*] > 1 \tag{5}$$

Given that $y_1^* = y_2^* = y^*$ and $x^* + f(h_{12})y^* = 1/r$ this can also be written as:

$$\xi(h_{13}, h_{23}) = (f(h_{13}) + f(h_{23}) - f(h_{12}))y^* + \min\{f(h_{13}), f(h_{23}\}y_{12}^* > 0 \tag{6}$$

Additionally, y_{12}^* can be expressed as $y_{12}^* = (1 - x^* - 2y^*)$ or, from (3), $y_{12}^* = [(2r-1)f(h_{12}) - 2 + \Omega]y^*/2$. If it is also assumed that $h_{13} < h_{23}$ then $f(h_{13}) < f(h_{23})$ and an alternative form of the invasion criterion is:

$$\xi(h_{13}, h_{23}) = f(h_{13})[(2r-1)f(h_{12}) + \Omega] + 2[f(h_{23}) - f(h_{12})] > 0 \tag{7}$$

A similar expression is obtained if it is assumed that $h_{23} < h_{13}$.

4. Invasion, coexistence and evolution with a discontinuous cross-immunity function

We now consider in detail when invasion is possible if the cross-immunity function is discontinuous. This structure of f means that the system is independent of the number of bitstring elements n. Each strain has cross-immunity 0 with itself, cross-immunity 1 with exactly one discordant partner and cross-immunity η with all other (related) strains. An equilibrium composed of two related strains can always be invaded. But an equilibrium composed of two discordant strains is resistant to invasion if $r < 1/(2\eta)$ and the intrinsic growth rate is insufficient to compensate for cross-immunity (Ferguson and Andreasen, 2002). This threshold is shown in Figure 1. Consequently, three related strains will always coexist but a trio consisting of a discordant pair plus one related strain will only coexist if r is sufficiently large, otherwise the related strain will be excluded. Examining the contributions of primary, secondary and tertiary infections to the invasion function given in (5) shows why only discordant pairs are resistant and why the threshold for invasion depends on r and η. Assuming that S1 and S2 are the two existing strains and S3 is the invading strain, there are three cases to consider:

Case i. If S1, S2 and S3 are all related then $f(h_{12}) = f(h_{23}) = f(h_{23}) = \eta$. From (6) the invasion function is $\xi(h_{13,}\, h_{23}) = \eta(y^* + y_{12}^*) > 0$ and so the invasion criterion is clearly always satisfied. Considering (5), and using the equilibrium solutions given in (3), the contribution of primary infections is $2\eta r/((2r + 1)\eta - 2 + \Omega)$, the contribution of secondary infections is $1 - 2/((2r - 1)\eta + \Omega)$ and the contribution of tertiary infections is $((2r - 1)\eta - 2 + \Omega)^2/(4(2r - 1)\eta + \Omega)$. The first column of Figure 2 shows how these components depend on r and η. Weaker cross-immunity (larger η) reduces the contribution of primary infections, increases the contribution of secondary and tertiary infections. Overall, the changes in the primary and secondary infections approximately balance but the tertiary infections ensure that invasion is always possible.

Case ii: If S1 and S2 are related and S3 is discordant with S1 then $f(h_{12}) = f(h_{23}) = \eta$ and $f(h_{13}) = 1$. From (6) the invasion function is $\xi(h_{13,}\, h_{23}) = y^* + \eta y_{12}^* > 0$ and the invasion criterion is always satisfied. Considering (5), and using the equilibrium solutions given in (3), the contributions of primary and tertiary infections to invasion are the same as in case i. The contribution of secondary infections is $(1 + \eta)/2\eta - (1 + \eta)/((2r - 1)\eta + \Omega)\eta)$. The second column of Figure 2 shows how these components depend on r and η. In contrast to case i, weaker cross-immunity reduces the contribution of secondary infections unless r is small, in which case there is a slight increase. However, a rapid increase in tertiary infections as η increases still ensures that invasion is always possible.

Case iii: If S1 and S2 are discordant and both related to S3 then $f(h_{12}) = 1$ and $f(h_{13}) = f(h_{23}) = \eta$. The invasion function is $\xi(h_{13,}\, h_{23}) = (2\eta - 1)y^* + \eta y_{12}^* > 0$. Using the equilibrium solutions given in (3) this becomes $2\eta r^2 - (2\eta + 1)r + 1 > 0$ and so invasion is only possible if $r > 1/(2\eta)$. Considering (5), the contribution of primary infections to invasion is $rx^* = r/(2r - 1)$. The contribution of secondary infections is $\eta r y^* = 2\eta(r - 1)/(2r - 1)$ and the contribution of tertiary infections is $\eta r y_{12}{}^* = 2\eta(r - 1)^2/(2r - 1)$. The third column of Figure 2 shows how these components depend on r

and η. The primary component is independent of η but both the secondary and tertiary components increase linearly as η increases. When $\eta = 0$ strain 3 is restricted to primary infections and invasion is always impossible. It only becomes possible when cross-immunity is weaker and there are sufficient secondary and tertiary infections. Increasing the value of r has relatively little impact on the contributions of primary and secondary infections but the contribution of tertiary infections increases significantly. This reduces the value of η at which invasion first becomes possible.

When more than two strains are able to coexist, the equilibrium solution is not always stable and the system may exhibit complex oscillatory behavior (Andreasen et al., 1997; Ferguson and Andreasen, 2002; Gupta et al., 1998). It is not, therefore, possible to examine the invasibility of additional strains analytically but numerical solutions offer some insight. Due to the rapid increase in complexity, we considered a maximum of 6 strains. The system was initialized with between 2 and 6 strains present and iterated to a quasi-equilibrium state. The invasion function for an additional strain was derived by extending (5) to accommodate multiple strains in the obvious way implied by (1). It turns out that the oscillatory nature of the system means that the invasion function may not remain on the same side of the invasion threshold along the entire solution trajectory. Therefore, the system was solved for a further 10^5 time units to determine the total proportion of time for which the invasion criterion was satisfied. For more details, see the caption to Figure 3.

If all the initial strains were related, any other strain could always invade. Results for other strain combinations are shown in Figure 3. However many strains were initially present, the set of strains present at quasi-equilibrium always consisted of either all the initial strains, a single discordant pair of strains or, when r or η were small, two discordant pairs of strains. If the initial collection of up to six strains only included one discordant pair, the regions of the r -η parameter space for which this pair excluded all other strains was identical to the region in which such a discordant pair was shown analytically, and in Figure 1, to be resistant to invasion by any other strain. This result suggests that dominance of a discordant pair is independent of the number of strains initially present. Outside of this region of parameter space, and away from the boundary, any additional strain could always invade. However, in a narrow region close to the boundary the value of the invasion function fluctuated across the threshold of the invasion criterion, indicating that an additional strain could only invade for part of the time. The precise region of parameter space in which this phenomenon was observed depended on how the initial strains were related to one another and to the invading strain. For oscillations in the population dynamics to occur at all, at least three strains are necessary. For the invasion function to oscillate across the invasion threshold at least four strains appear to be necessary, but not sufficient, and the occurrence of such behavior was generally more extensive when the invading strain was not discordant with any of the existing strains. A detailed mathematical analysis carried out for the special case of 4 strains made up of 2 discordant pairs (Dawes and Gog, 2002) has detected a narrow region of parameter space close to the boundary for dominance of the discordant pair in which a non-oscillatory asymmetric equilibrium is stable. The numerical results presented in this article are oscillatory and so do not correspond to this equilibrium. It may, however, be possible to develop similar methods to examine the underlying reasons behind this

ambiguity in the invasion function. While of mathematical interest this will not be pursued further here as the parameter region in which it is observed is very narrow and has limited relevance to this study.

The framework of invasion analysis used here does not explicitly include the process of mutation. It is, however, possible to infer the consequences of mutation from Figure 3 based on the reasonable hypothesis that results will remain similar even when more than six strains are co-circulating. If the basic reproductive number is small or the cross-immunity between similar strains is strong then the system will lie in the black shaded areas of the invasion plots, indicating that a single discordant pair can exclude all other strains. Starting with a single strain we expect the process of mutation to lead to the emergence and coexistence of multiple related strains with diversity continuing to accumulate until a discordant pair appears, excludes all others and forms an evolutionary stable alliance. If the basic reproductive number or cross-immunity are not such that a discordant pair will be dominant, all strains will coexist and the process of mutation is expected to lead from a single initial strain to a highly diverse population without any clear antigenic structure. The discrete antigenic space with discontinuous cross-immunity considered here is very basic but may be considered an abstraction of a portion of a neutral network model proposed for influenza (Koelle et al., 2006). In our model the immune interaction between a mutant and the existing strains is always the same unless it has a rare genotype that forms one half of a discordant pair and thus experiences a significant reduction in host immunity. In the neutral network model, the cross-reaction between a mutant and the existing strains is always the same as long as it remains in the same network but, under certain conditions, a mutant crosses to an adjacent network and the cross-reaction with existing strains becomes radically weaker. Numerical simulations using the neutral network framework predict pathogen evolution characterized by accumulations of diversity punctuated by clade replacement events when a new network is accessed (Koelle et al., 2006). Applying invasion analysis in our simple model to infer the process of mutation and selection leads to a strikingly similar prediction of boom and bust in diversity, and a thorough assessment of any similarity in the underlying mechanisms merits further attention in the future.

5. Invasion, coexistence and evolution with a continuously defined cross-immunity function

We now return to the model with two existing strains at equilibrium and make a detailed examination of when a third strain can invade if cross-immunity is a continuously defined function of the antigenic distance. As before, the existing strains are labeled S1 and S2, the invading strain S3. If strains S1 and S2 are fixed at distance h_{12} then h_{13} and h_{23}, are not independent but, in contrast to a one-dimensional antigenic space, fixing h_{13} does not necessarily uniquely define h_{23}. Using the form of the invasion function given in (6), for fixed h_{12} and h_{13}, the function is smallest, and invasion is most difficult, when h_{23} takes the minimum admissible value h^*_{23}. Here we show that, for $f(h) = h^q$ and fixed h_{12}, the minimum value of the invasion fuction is positive for all values of h_{13} if $q \leq 1$ but could be negative if $q > 1$. This means that invasion is possible for any arrangement of strains in antigenic space if $q \leq 1$. However, invasion may be impossible for some arrangements if $q > 1$ and we consider the case of $q = 2$ as a particular example. For the purposes of this proof we

consider h to be continuous, so that $f(h) = h^q$ is also continuous. However, it should be clear that the result still applies when h is discrete.

We begin by showing that, if h_{12} and h_{13} are fixed and the genotype and structural phenotype are defined by the same binary string of length n, the minimum possible value of h_{23} is $h^*_{23} = h_{23}^{\min} = \max\{h_{12} - h_{13}, h_{13} - h_{12}\}$ and the set of all possible values for h_{23} is $\{ h_{23}^{\max}, h_{23}^{\max} - 2/n, h_{23}^{\max} - 4/n, \ldots, h_{23}^{\min} \}$ where $h_{23}^{\max} = \min\{h_{12} + h_{13}, 2 - (h_{12} + h_{13})\}$. To see this let S1 and S2 be fixed binary strings of length n having nh_{12} positions with different values and n - nh_{12} positions with identical values. Let S3 be another binary string having nh_{13} positions with values different from S1. Then the value of h_{23} is minimized if S3 is constructed from S1 by, so far as possible, switching values at positions that are different between S1 and S2. If $h_{13} \leq h_{12}$ then initially S2 and S3 have nh_{12} different positions but nh_{13} of these are made identical by the switching, leaving $nh_{12} - nh_{13}$ different positions and giving $h_{23} = h_{12} - h_{13}$. If $h_{13} > h_{12}$ then all of the nh_{12} initially different positions are made identical by the switching, but $nh_{13} - nh_{12}$ of the initially identical positions are made different giving $h_{23} = h_{13} - h_{12}$. The maximum possible values are found by a similar argument. Values of h_{23} change in steps of 2 between the maximum and minimum since switching an element of S3 to adjust h_{23} always requires switching an additional element to ensure that h_{13} remains constant.

We now use this expression for minimum possible value for h_{23} in terms of h_{12} and h_{13} together with the invasion function given in (6) to show that invasion is possible for all values of h_{12} and h_{13} if $q \leq 1$ but may not always be possible if $q > 1$. Four cases, of which only three are admissible, describe the antigenic relationships between the three strains. To complement the analysis, these cases are illustrated in Figure 4 with an invasion function evaluated for the specific example of $q = 2$. Note that throughout the following the coexistence equilibrium of the original two strains x^* and y^* is written as x and y to simplify notation.

Case i: Suppose $h_{13} \leq h_{12}$, hence $h^*_{23} = h_{12} - h_{13}$. Suppose also $h_{13} \leq h^*_{23}$, implying that $h_{13} \leq h_{12}/2$. Then, from (6) and using $y_{12} = 1 - x - 2y$, the minimum value of the invasion function as a function of h_{13} is:

$$\xi(h_{13}) = (f(h_{12} - h_{13}) - f(h_{12}))y + f(h_{13})(1 - x - y) \tag{8}$$

Differentiating with respect to h_{13}:

$$\xi'(h_{13}) = -f'(h_{12} - h_{13})y + f'(h_{13})(1 - x - y) \tag{9}$$

$$\xi''(h_{13}) = f''(h_{12} - h_{13})y + f''(h_{13})(1 - x - y) \tag{10}$$

So, ξ has at most one turning point and this occurs when

$$\frac{y}{1 - x - y} = \frac{f'(h_{13})}{f'(h_{12} - h_{13})} = \left(\frac{h_{13}}{h_{12} - h_{13}} \right)^{q-1} \tag{11}$$

Also, from (8), $\xi(0) = 0$ and, substituting $f(h) = h^q$ into (9), the gradient at this point is:

$$\xi'(0) = -qh_{12}^{q-1}y + h_{13}^{q-1}(1-x-y)\Big|_{h_{13}=0} \tag{12}$$

Note that $x + 2y + y_{12} = 1$ means that $y/(1 - x - y) < 1$ and, by assumption, $h_{13}/(h_{12} - h_{13}) \leq 1$. So, if $q \leq 1$, then (11) cannot be satisfied and ξ cannot have a turning point. Furthermore, if $q \leq 1$ then by (12) $\xi'(0) > 0$, so the minimum value of $\xi(h_{13}) > 0$ and invasion is possible for all $0 \leq h_{13} \leq h_{12}/2$. If $q > 1$, (11) may have an admissible solution but this is not guaranteed. However, if a turning point does exist it will be a minimum because $f''(h) = q(q - 1)h^{q-2} > 0$. From (12), if $q > 1$ then $\xi'(0) < 0$ so there must always be some interval $(0, h_x]$ where invasion is impossible although, since h_{13} actually changes in discrete steps of size $1/n$, invasion will only be impossible in practice if $h_x > 1/n$.

Case ii: As before, suppose $h_{13} \leq h_{12}$, hence $h^*_{23} = h_{12} - h_{13}$. Suppose also $h_{13} > h^*_{23}$, implying that $h_{13} > h_{12}/2$. Then a similar analysis to case i shows that, if $q \leq 1$ invasion is possible for all $h_{12}/2 \leq h_{13} \leq h_{12}$ but if $q > 1$ invasion is not possible in an interval $[h_y, h_{12})$.

Case iii: Suppose $h_{13} > h_{12}$ then $h^*_{23} = h_{13} - h_{12}$ and $h_{13} < h^*_{23}$ is impossible.

Case iv: Suppose $h_{13} > h_{12}$, hence $h^*_{23} = h_{13} - h_{12}$. Suppose also $h_{13} \geq h^*_{23}$. Then, using (6) and $y_{12} = 1 - x - 2y$:

$$\xi(h_{13}) = (f(h_{13}) - f(h_{12}))y + f(h_{13} - h_{12})(1-x-y) \tag{13}$$

Differentiating with respect to h_{13}:

$$\xi'(h_{13}) = f'(h_{13})y + f'(h_{13} - h_{12})(1-x-y) \tag{14}$$

Since f is monotonic increasing by assumption, $\xi'(h_{13}) > 0$ for $h_{12} \leq h_{13} \leq 1$. For all values of q, $\xi(h_{12}) > 0$ and so invasion is always possible for $h_{12} \leq h_{13} \leq 1$.

Hence invasion, and coexistence, of a third strain is always possible when $q \leq 1$ but may be impossible for certain parameter sets when $q > 1$. We now consider $q = 2$ as a specific example. Assume, as in case i, that $0 \leq h_{13} \leq h_{12}/2$. Then the minimum value of the invasion function given in (8) becomes:

$$\xi(h_{13}) = h_{13}((h_{13} - 2h_{12})y + h_{13}(1-x-y)) \tag{15}$$

Invasion is possible when ξ given by (15) is positive. That is, when

$$h_{13} > \frac{2h_{12}y}{1-x} = \tau_1(h_{12}) \tag{16}$$

A similar threshold arises from case ii. For $h_{12}/2 \leq h_{13} \leq h_{12}$ invasion is possible when:

$$h_{13} > \frac{h_{12}(1-x-2y)}{1-x} = \tau_2(h_{12}) \qquad (17)$$

From case iv, invasion is always possible when $h_{12} \le h_{13} \le 1$. Recall that these thresholds, shown in Figure 5 for several values of r, correspond to the value of h_{13} at which invasion is possible when h^*_{23} minimizes the invasion function. When these thresholds are satisfied, invasion is possible for all admissible values of h_{23}. When they are not satisfied, invasion is impossible for $h_{23} = h^*_{23}$, but may be possible at other values. From Figure 5, when $r \le 2$, invasion is impossible for $h_{23} = h^*_{23}$ unless $h_{13} > h_{12}$. As r increases, a region of the h_{13} axis centered on $h_{12}/2$ appears where, for large h_{12}, invasion is possible at $h_{23} = h^*_{23}$. For higher values of r this region appears at lower values of h_{12} and spans a broader range of h_{13} values. Invasion for $q = 2$, $n = 10$ and values of h_{23} other than h^*_{23} was explored numerically. When the two existing strains were dissimilar (h_{12} close to 1) invasion was generally restricted to locations close to one of the existing strains and distant from the other. When the existing strains were similar (h_{12} close to 0) invasion was generally possible in all admissible locations except those very close to both existing strains. Invasion was always easier, and occurred for a broader range of the admissible h_{23} values, for larger values of r. Figure 6 shows when invasion is possible for all admissible combinations h_{12}, h_{13} and h_{23} when $r = 1.01$. Systems of up to 6 strains were also explored numerically for $q = ½$, 1 and 2, a range of values of r and all admissible combinations of h_{12}, h_{13} and h_{23}. This indicated that for $q = ½$ or $q = 1$ at least 6 strains will always coexist, although not always at a constant equilibrium. However, for $q = 2$ invasion, and hence coexistence of multiple strains, is not always possible and depends in a complex way on the distribution of strains in antigenic space and the basic reproductive number.

The preceding analysis shows that invasion depends on the shape of the cross-immunity function f. We now investigate why. The reasons for this dependency are most clearly seen when the existing strains S1 and S2 are discordant since, in this case, h_{23} is uniquely determined for any pair h_{12} and h_{13}. Based on (5), Figure 7 shows how the primary, secondary and tertiary (x^*, y^* and $y_{12}{}^*$) components of the invasion function contribute to the total for $q = ½$, 1 and 2 and a range of values of h_{13}. As expected, the contribution of primary infections does not depend on h_{13} since they are not affected by cross-immunity. However, the contribution of tertiary infections depends strongly on h_{13}. For all three forms of f, tertiary infections increase rapidly as h_{13} increases, reaching a maximum at $h_{13} = 1/2$ when strain S3 is equidistant from S1 and S2. The contribution of secondary infections is somewhat different for each functional form. When f is linear, the net contribution of secondary infections does not depend on h_{13}. This is because as h_{13} increases, the number of secondary infections by S3 of hosts previously infected with S1 increase at exactly the same rate as the number of secondary infections of hosts previously infected with S2 decrease. When f is a square-root the contribution of secondary infections increases with h_{13} to a maximum at $h_{13} = 1/2$ then decreases again symmetrically. This is because the number of S1-S3 secondary infections increases more rapidly than the number of S2-S3 secondary infections decreases as h_{13} increases from 0 to $n/2$. When f is parabolic the converse occurs and the contribution of secondary infections deceases to a minimum at $h_{13} = n/2$. So, assuming that S3 is closer to S1 than S2, for linear and square-root forms of f all components of the invasion function are either constant or increasing

with h_{13}. The converse assumption leads to a symmetric result. Since invasion must be neutral when $h_{13} = 0$ (S1 and S3 are identical) it must be possible at all other points. However, when f is parabolic, invasion is only possible when there are sufficient tertiary infections to compensate for the decrease in secondary infections with h_{13}. The contribution of primary and secondary infections depends only weakly on r. However, the contribution of tertiary infections is approximately proportional to r. So larger values of r mean that the increase in tertiary infections compensates for the decrease in secondary infections, and invasion becomes possible, at lower values of h_{13}.

The case with S1 and S2 discordant is instructive, and corresponds closely to results for a continuous one-dimensional antigenic space (Adams and Sasaki, 2007). When S1 and S2 are not discordant the situation is much more complex and difficult to interpret. Figure 8 shows the components of the invasion function when $n = 10$ and $h_{12} = 7/10$. When f is parabolic invasion is possible for some values of h_{23} when $h_{13} \leq 6/10$ but the interactions are clearer in the region $h_{13} < 5/10$. Here, for all forms of f tertiary infections are insensitive to h_{23} because immunity is controlled by the minimum of $f(h_{13})$ and $f(h_{23})$, which is either $f(h_{13})$ or very close to it. When h_{13} is close to 0, each change in h_{13} leads to a large change in $f(h_{13})$ if $q < 1$ but a small change in $f(h_{13})$ if $q > 1$. So the contribution of tertiary infections to invasion increases more rapidly with h_{13}, and is generally more significant, when $q < 1$. The contribution of secondary infections to invasion is more sensitive to h_{23}. When $q > 1$ a small change in h_{23} often leads to a large change in $f(h_{23})$ and there is considerable variation in the contribution of secondary infections for any given value of h_{13}. Conversely, when $q < 1$, $f(h_{23})$ is less sensitive and the variation is much smaller. Combined with the low baseline of tertiary infections, the high variation in secondary infection means that invasion is sometimes impossible when $q > 1$.

The form of the cross-immunity function f plays a critical role in determining the expected evolutionary pathway of the pathogen. When $q \leq 1$ two strains at equilibrium can always be invaded by a third strain. Numerical results indicate that at least 6 strains can always coexist and it seems likely that this extends to any number of strains. Therefore, cross-immunity is not a strong selective force and pathogens will exist as a cloud of strains with little antigenic structure to the population. When $q > 1$ invasion and coexistence are limited and the pathogen population is expected to be highly antigenically structured. To investigate further, we set $q = 2$ and assume that one strain is initially present. Then a mutant strain can arise by the random switching of a single bitstring element. This will lead to two coexisting strains, S1 and S2, with $h_{12} = 1/n$. A third strain S3 can then arise by random switching of a single element in either of these bitstrings. We wish to know if S3 can invade and, if so, whether it will coexist with or replace the existing strains. The process of mutation and replacement will maintain an evolutionary trajectory composed of two distinct branches, while coexistence of the mutant with both existing strains will lead to the establishment of a new branch. This event is our primary focus here and we refer to it as branching. It is expedient to consider the more general situation of two existing strains S1 and S2 with distance h_{12} and a mutant strain S3 resulting from S1 and so having $h_{13} = 1/n$. Clearly a symmetric argument will apply if the mutant strain results from S2. There are two cases to consider:

Case i, step-out mutant: The element j in S1 that is switched to create S3 is such that S1[j] = S2[j]. Then S3[j] ≠ S2[j] but all other elements are unchanged so $h_{23} = h_{12} + 1/n$. Substituting these values into (6) gives the invasion criterion $2(1/n + h_{12})y^* + y^*_{12} > 0$. Clearly this is always satisfied. Furthermore, the invasion criterion if S1 and S3 are at equilibrium and S2 is attempting to invade is $2h_{12}(h_{12} + 1/n)y^* + y^*_{13} > 0$ which is also always satisfied. However, the invasion criterion if S2 and S3 are at equilibrium and S1 is attempting to invade is:

$$h_{12} < \frac{y^*_{23}}{2ny^*} \tag{18}$$

Here $h_{12} \geq 1/n$, otherwise S1 and S2 would be identical. The explicit expressions for y^* and y^*_{23} can be written in terms of h_{12}, r and n and, substituting these into (18) gives:

$$(2r-1)\left(h_{12} + 1/n\right)^2 + \Omega - 2 - 4nh_{12} > 0 \tag{19}$$

where $\Omega = \sqrt{4 - 4(h_{12} + 1/n)^2 + (h_{12} + 1/n)^4 (2r-1)^2}$. This simplifies to the threshold:

$$h_{12} > \rho_1(r,n) = \frac{2n^2 - 3r + 2 + \sqrt{r^2 - 12rn^2 + 8n^2 + 4n^4}}{2n(2r-1)} \tag{20}$$

There is also a lower threshold that results from a negative square root in (20) but it can be shown that this is always less than $1/n$ and so never admissible.

Case ii, step-in mutant: The element j in S1 that is switched is such that S1[j] ≠ S2[j]. Then S3[j] = S2[j] and $h_{23} = h_{12} - 1/n$. Then, using (7), and substituting the explicit expressions for the equilibrium solutions, the invasion criterion is:

$$(2r-1)h_{12}^2 + \Omega + 2 - 4nh_{12} > 0 \tag{21}$$

where $\Omega = \sqrt{4 - 4h_{12}^2 + h_{12}^4(1-2r)^2}$. This simplifies to the threshold:

$$h_{12} > \rho_2(r,n) = \frac{2n^2 + r + \sqrt{r^2 - 12rn^2 + 8n^2 + 4n^4}}{2n(2r-1)} \tag{22}$$

This condition is only valid if $h_{12} \geq 2/n$ since, if $h_{12} = 0$ then $h_{23} = h_{12} - 1/n$ is nonsense and, if $h_{12} = 1/n$ then $h_{23} = 0$ would imply that S2 and S3 are identical. Again, there is a lower threshold that results from a negative square root in (22) but it can be shown that this is always less than $2/n$ and so never admissible. Furthermore, the threshold ρ_2 is clearly always greater than ρ_1 and so we need only consider case i when examining the conditions for branching.

The three invasion conditions derived in case i show that an equilibrium composed of S1 and S2 is locally unstable to invasion by S3, an equilibrium composed of S1 and S3 is locally unstable to invasion by S2 but an equilibrium composed of S2 and S3 is

locally stable to invasion by S1 unless (20) is satisfied. Together, these results mean that a mutant S3 can invade, the existing strain S2 cannot be excluded but the existing strain S1 will be excluded unless there is a bistable state with all three strains coexisting. We do not provide an analytic result to rule out such bistability. However, numerical bifurcation analysis suggests that the threshold for local stability given in (20) is identical to the threshold for the existence of a solution with three strains coexisting, as shown in Figure 9. So, starting with two strains with distance $1/n$, a one step mutant strain can always invade and, since the three strain equilibrium does not exist, it must replace one of the existing strains to form a new pair. This process will continue until the distance between the current pair reaches the threshold ρ_1 given in (20). At that point three strains can coexist and a new evolutionary branch appears. The way in which the threshold distance for branching given by (20) depends on r and n is shown in Figure 9. Branching does not occur for $r < n$. If r is increased branching first occurs when the existing strains are increasingly antigenically similar until, when r is large in comparison to n, branching occurs immediately. The number of elements in the bitstring, n, is important here because it implicitly determines the magnitude of antigenic change associated with a single point mutation. There is a region around each existing strain, the width of which is determined by their antigenic similarity and the basic reproductive number, where invasion and coexistence are impossible. In order for a mutant to establish a new evolutionary branch it must escape this region. When n is small the antigenic change associated with each point mutation is large and such an escape is more likely. Thus branching can occur even when two existing strains are antigenically similar. Conversely, if n is large, each mutation is associated with very little antigenic change and escape is difficult. Thus branching can only occur when the two existing strains are distant and the antigenic influence of the strain most dissimilar to the mutant strain is weak.

The ordinary differential equation model extended to incorporate a maximum of 6 strains was solved numerically to simulate the evolutionary process when $r = 15$. A new strain could occur as the result of switching a random element in an existing strain. A strain was assumed to be extinct if its force of infection became very small. The mutation rate was set sufficiently low to ensure that extinction occurred at a similar rate as the production of new strains and so the total number of strains remained below 6 as long there was no branching. For more details see the caption to Figure 10. Simulations were terminated as soon as 6 strains appeared, usually indicating the existence of three stable strain branches and one mutant strain from each which may or may not persist if the simulation were continued. From (20) and Figure 9, when $r = 15$ branching is predicted to occur immediately for $n = 5$, at $h_{12} = 6/10$ for $n = 10$ and never for $n = 15$ and $n = 20$. As Figure 10 shows, simulation results were largely in agreement although the value of h_{12} at which the simulation was actually terminated was often 1 or 2 mutational steps greater than the predicted branching value because further mutation and extinction events involving the two founder branches occurred before a total of six strains accumulated. There were also a few cases when, by chance, a rapid sequence of mutations led to the premature accumulation of six strains although most of these would be expected to become extinct if the simulation could have been continued.

6. Discussion

In this paper we have shown that, in the context of a discrete bitstring derived antigenic space, the form of the relationship between antigenic distance and cross-

immunity is of critical importance in the invasion, coexistence and predicted evolution of pathogen strains. If the relationship is discontinuous and only differentiates between identical, related or discordant strains then two coexisting related strains can always be invaded by another strain, whether it is related or discordant, mainly as the result of tertiary infections. Two coexisting discordant strains, however, may be resistant to invasion if cross-immunity between related strains is strong or the basic reproductive number of the pathogen is low and secondary or tertiary infections are rare. If more than two strains are initially present the incidence of each may show complex oscillations. However, the cross-immunity and basic reproductive numbers required for dominance of a pair of discordant strains show no change. Thus, if epidemiological characteristics are such that a discordant pair cannot be dominant, then we can infer that the evolutionary pathway will be characterized by the continuous accumulation of diversity without antigenic selection. If a discordant pair can be dominant, we can infer that the evolutionary pathway will characterized by the accumulation of diversity as related strains proliferate followed by widespread extinction when a discordant pair finally arises. Here it is the discordant pair of strains behind the selective sweep, not either strain individually.

If the relationship between antigenic distance and cross-immunity can be continuously defined (say $f(h) = h^q$) then the curvature, (represented by q) has a big impact influence on invasion and coexistence. When the relationship is concave ($q < 1$) cross-immunity is disproportionately weak between strains that are close together. Given two coexisting strains invasion, and coexistence, of a third strain with any antigenic type is possible, mainly as the result of the strong influence of tertiary infections. Extending further, at least 6, and possibly any number, of strains will coexist. Thus cross-immunity is not a strong selective force and pathogens may be expected to exist as clouds of strains with little or no antigenic structuring. When the relationship is convex ($q > 1$) cross-immunity between strains that are close together is disproportionately strong compared to those that are far apart. Given two coexisting strains, invasion of a third strain, and subsequent coexistence, depends on the distribution of the strains in antigenic space and the basic reproductive number. The inter-relationship of these factors is complicated but, generally, invasion and coexistence are more likely when the basic reproductive number is larger. When the existing strains are antigenically dissimilar, invasion is restricted to regions of antigenic space close to one of these strains and distant from the other. When the existing strains are antigenically similar, invasion is possible anywhere except the region of antigenic space that is close to both of them. In both cases, successful invasion may result in the extinction of one of the existing strains. The inferred evolutionary pathway will be characterized by such events. Mutant strains replace members of a dominant pair until the antigenic distance between them is sufficiently great that a subsequent mutant can escape the antigenic shadow of its progenitor and establish a new lineage. The antigenic change associated with a single mutation is important. Smaller changes mean that mutants can only escape the antigenic shadow when the existing strains are further apart. In general, the total number of strains may be limited by both the availability of antigenic niches and the accessibility of these niches by a sequence of relatively small mutations alone. So cross-immunity may be a strong selective force and the pathogen population may show significant antigenic structuring, particularly when the basic reproductive number is small.

Many pathogens exist in populations displaying considerable antigenic variation. Immune escape is likely to be a significant aspect of the evolutionary process and may be particularly important for the management of vaccine preventable diseases such as influenza, *Haemophilus influenzae, Streptococcus pneumoniae, Neisseria meningitides* and *Bordetella pertussis* (Martcheva et al., 2008). Models using discrete antigenic spaces have been already developed for both influenza and *N. meningitidis* (Gupta et al., 1998; Recker et al., 2007) and will undoubtedly be elaborated and applied to other pathogens in the future. The analysis presented here offers general insights into the role of cross-immunity in pathogen evolution and highlights several important factors that must be considered when developing more sophisticated models. The choice of function relating antigenic distance to cross-immunity, the calculation of immunity arising from two or more previous infections and the number of elements used in the bitstring genotype all have a significant impact on invasion, coexistence and the evolutionary trajectory the model predicts whether the antigenic space is continuous and one-dimensional or, as used here, discrete and high dimensional. At present there are few empirical results available to guide the construction and parameterization of these key components and the choice is often rather arbitrary. Such analysis will hopefully become available eventually. In the meantime, its absence need not be an obstacle to the development and application of models for pathogen evolution as long as their predictions are interpreted with an appropriate element of caution.

Acknowledgements
This work was partially funded by a Japanese Society for Promotion of Science Postdoctoral Fellowship to B. A.

Figure captions

Figure 1: Invasibility depending on r and η when cross-immunity is discontinuous, two discordant strains are at equilibrium and the invading strain is related to both of them. Invasion is possible in the white region, impossible in the black region.

Figure 2: Invasion functions and components depending on r and η when the cross-immunity function is discontinuous and two existing strains are at equilibrium. First column: all strains related. Second column: existing strains related, invading strain discordant. Third column: existing strains discordant, invading strain related. Solid black line – total invasion function, grey lines – components: dashed – primary infections, dot-dash – secondary infections, solid – tertiary infections. The dotted black line shows the invasion threshold.

Figure 3: Invasion potential, expressed as the proportion of 10,000 time units for which the invasion criterion for a mutant is satisfied, when different combinations of 2 to 6 existing strains co-circulate at quasi-equilibrium. Black indicates the invasion criterion is never satisfied, white indicates it is always satisfied, shades of grey indicate it is satisfied for some of the time. Hatching indicates parameter combinations that are inadmissible because the required initial qausi-equilibrium does not exist. In each panel the initial strains present are indicated by D and R, for instance D2R3 indicates a discordant pair and 3 related strains. The strain introduced is indicated by the final letter, +D or +R indicating a discordant or related strain respectively. The top left panel, D2+R, is the same as Figure 1, but determined numerically. The initial quasi-equilibrium was found by applying a Runge-Kutta method until $t = 10^5$. The system was then iterated for a further 10^5 time units during which period the invasion criterion for the introduced strain was evaluated at regular, small time intervals.

Figure 4: Minimum value of invasion function ξ for strain S3 as function of h_{13} when existing strains S1 and S2 have antigenic distance $h_{12} = 0.7$ and $r = 5$. For case i $h_{13} \leq h_{12}$, the minimum value of h_{23}, $h^*_{23} = h_{12} - h_{13}$, $h_{13} \leq h^*_{23}$ and equation (8) applies. For case ii $h_{13} \leq h_{12}$, $h^*_{23} = h_{12} - h_{13}$, $h_{13} > h^*_{23}$. For case iv $h_{13} > h_{12}$, the minimum value of h_{23}, $h^*_{23} = h_{13} - h_{12}$, $h_{13} > h^*_{23}$ and equation (13) applies.

Figure 5: Threshold value of h_{13} for invasion when $f(h) = h^2$. Strains S1 and S2, with antigenic distance h_{12}, are at equilibrium. In the black regions, given by $\tau_1 \leq h_{13} \leq h_{12}/2$ and $h_{12}/2 \leq h_{13} \leq \tau_2$ and $h_{12} \leq h_{13}$ invasion is possible for all values of h_{23} associated with the given value of h_{13}. In the white regions invasion is impossible for at least one value of h_{23}.

Figure 6: Numerically evaluated invasion criteria when two strains are initially present $f(h) = h^2$, $n = 10$ and $r = 1.01$. Existing strains S1 and S2 have antigenic distance h_{12} and the distance between S1 and S3 is h_{13}. The exact location of S3 is then determined by h_{23}. Admissible values for h_{23} are shaded black if invasion is possible for S3 at this location and white if invasion is impossible. Inadmissible values for h_{23} are shaded grey.

Figure 7: Invasion functions and components depending on h_{13} with $r = 3$, $n = 10$ and $h_{12} = 10/10$ (S1 and S2 are discordant). Note that h_{23} is uniquely defined by $h_{23} = 1 - h_{13}$. Since h_{13} is discrete, makers donate actual values of the invasion fucntions, lines are only added to aid visualization. Grey lines are the components: circles – primary, triangles – secondary, squares – tertiary component. Black lines and circles are the total invasion functions. Columns from left to right, parabolic, linear and square-root cross-immunity functions

Figure 8: Invasion functions and components depending on h_{13} and h_{23} for $r = 3$, $n = 10$, and $h_{12} = 7/10$. For most values of h_{13} there are several possible values of h_{23}. All of these are marked, with numbers corresponding to n x h_{23}. Grey circles denote the contribution from secondary infections, crosses the contribution from tertiary infections, black circles the total invasion function. The contribution from primary infections is constant in all cases and so has been omitted for clarity.

Figure 9: Threshold distance between two existing strains for a new branch to appear. A third strain S3 can only invade and coexist with two existing strains S1 and S2 when the distance between them $h_{12} > \rho_1(r, n)$. The solid lines show thresholds for mutual invasibility of all strains and strain pairings derived by local stability analysis and given by (20). The dashed grey lines show thresholds for the existence of a solution with three strains coexisting calculated by numerical bifurcation analysis.

Figure 10: Value of h_{12} at which 6 strains first appear in an evolutionary simulation with $r = 15$ and $n = 5$, 10, 15 and 20. For each value of n, 20 independent runs were made. Initially one strain was present. New strains were generated by random switching of a single node in the bitstring of an existing strain with probability 5×10^{-6} per infection per time unit. Strains were assumed to be extinct if the force of infection was less than 10^{-80}.

References

Adams, B., et al., 2006. Cross-protective immunity can account for the alternating epidemic pattern of dengue virus serotypes circulating in Bangkok. Proceedings of the National Academy of Sciences of the United States of America. 103**,** 14234-14239.

Adams, B., Sasaki, A., 2007. Cross-immunity, invasion and coexistence of pathogen strains in epidemiological models with one-dimensional antigenic space. Math Biosci. 210**,** 680-99.

Alberts, B., 2002. Molecular biology of the cell. Garland Science, New York.

Anderson, R. M., May, R. M., 1982. Coevolution of Hosts and Parasites. Parasitology. 85**,** 411-426.

Anderson, R. M., May, R. M., 1991. Infectious diseases of humans : dynamics and control. OUP, Oxford.

Andreasen, V., et al., 1997. The dynamics of cocirculating influenza strains conferring partial cross-immunity. Journal of Mathematical Biology. 35**,** 825-842.

Andreasen, V., Sasaki, A., 2006. Shaping the phylogenetic tree of influenza by cross-immunity. Theoretical Population Biology. 70**,** 164-173.

Calvez, V., et al., 2005. Cluster formation for multi-strain infections with cross-immunity. Journal of Theoretical Biology. 233**,** 75-83.

Castillo-Chavez, C. E., 2002. Mathematical approaches for emerging and reemerging infectious diseases : an introduction. Springer.

Cummings, D. A. T., et al., 2005. Dynamic effects of anti body-dependent enhancement on the fitness of viruses. Proceedings of the National Academy of Sciences of the United States of America. 102**,** 15259-15264.

Dawes, J. H. P., Gog, J. R., 2002. The onset of oscillatory dynamics in models of multiple disease strains. Journal of Mathematical Biology. 45**,** 471-510.

Dieckmann, U., 2002. Adaptive dynamics of infectious diseases : in pursuit of virulence management. IIASA ; Cambridge : Cambridge University Press, Laxenburg.

Ewald, P. W., 1994. Evolution of infectious disease. OUP, Oxford.

Ferguson, N., Andreasen, V., The influence of different forms of cross-protective immunity on the population dynamics of antigenically diverse pathogens. In: C. Castillo-Chavez, (Ed.), Mathematical approaches for emerging and reemerging infectious diseases : models, methods and theory. Springer, New York, 2002.

Ferguson, N. M., et al., 2003. Ecological and immunological determinants of influenza evolution. Nature. 422**,** 428-433.

Girvan, M., et al., 2002. Simple model of epidemics with pathogen mutation. Physical Review E. 65**,** -.

Gog, J. R., Grenfell, B. T., 2002. Dynamics and selection of many-strain pathogens. Proceedings of the National Academy of Sciences of the United States of America. 99**,** 17209-17214.

Gomes, M. G. M., et al., 2002. On the determinants of population structure in antigenically diverse pathogens. Proceedings of the Royal Society of London Series B-Biological Sciences. 269**,** 227-233.

Gupta, S., et al., 1998. Chaos, persistence, and evolution of strain structure in antigenically diverse infectious agents. Science. 280**,** 912-915.

Gupta, S., et al., 1996. The maintenance of strain structure in populations of recombining infectious agents. Nature Medicine. 2**,** 437-442.

Gupta, V., et al., 2006. Quantifying influenza vaccine efficacy and antigenic distance. Vaccine. 24**,** 3881-8.
Janeway, C. A., Janeway, C. A. I., 1999. Immunobiology : the immune system in health and disease. Current Biology Publications, London.
Koelle, K., et al., 2006. Epochal evolution shapes the phylodynamics of interpandemic influenza A (H3N2) in humans. Science. 314**,** 1898-1903.
Kryazhimskiy, S., et al., 2007. On state-space reduction in multi-strain pathogen models, with an application to antigenic drift in influenza A. Plos Computational Biology. 3**,** 1513-1525.
Lapedes, A., Farber, R., 2001. The geometry of shape space: Application to influenza. Journal of Theoretical Biology. 212**,** 57-69.
Martcheva, M., et al., 2008. Vaccine-induced pathogen strain replacement: what are the mechanisms? J R Soc Interface. 5**,** 3-13.
Recker, M., et al., 2007. The generation of influenza outbreaks by a network of host immune responses against a limited set of antigenic types. Proceedings of the National Academy of Sciences of the United States of America. 104**,** 7711-7716.
Smith, D. J., et al., 1999. Variable efficacy of repeated annual influenza vaccination. Proc Natl Acad Sci U S A. 96**,** 14001-6.
Smith, D. J., et al., 2004. Mapping the antigenic and genetic evolution of influenza virus. Science. 305**,** 371-6.
Tria, F., et al., 2005. A minimal stochastic model for influenza evolution. Journal of Statistical Mechanics-Theory and Experiment. -.

Figure 1

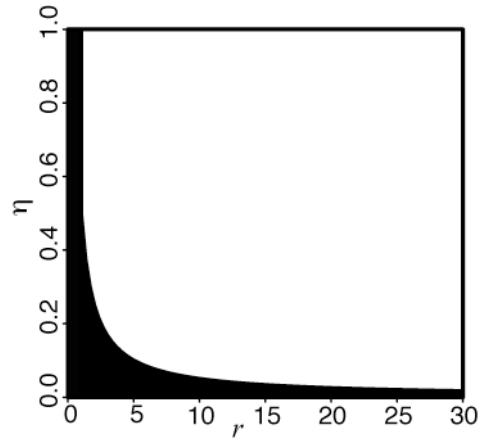


Figure 2

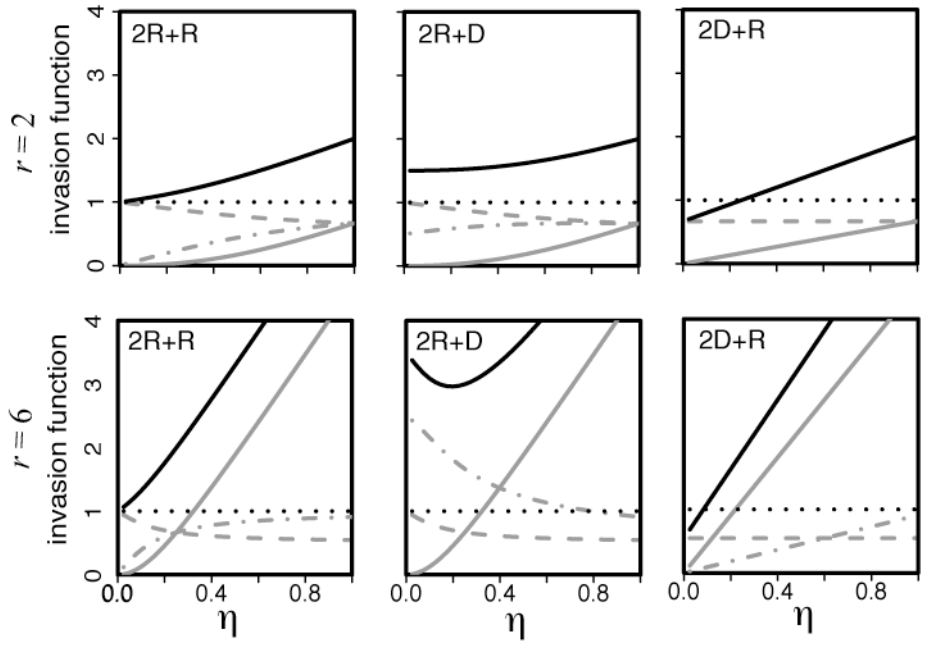

2R+R
2R+D
2D+R
r = 2
invasion function
r = 6
invasion function
0
1
2
3
4
0.0
0.4
0.8
η

Figure 3

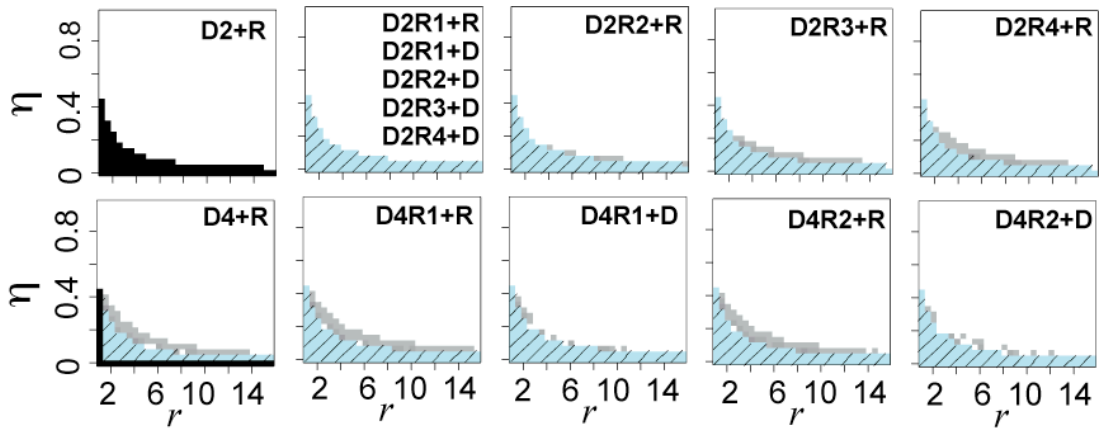

D2+R
D2R1+R
D2R1+D
D2R2+D
D2R3+D
D2R4+D
D2R2+R
D2R3+R
D2R4+R
D4+R
D4R1+R
D4R1+D
D4R2+R
D4R2+D
η
0
0.4
0.8
2
6
10
14
r

Figure 4

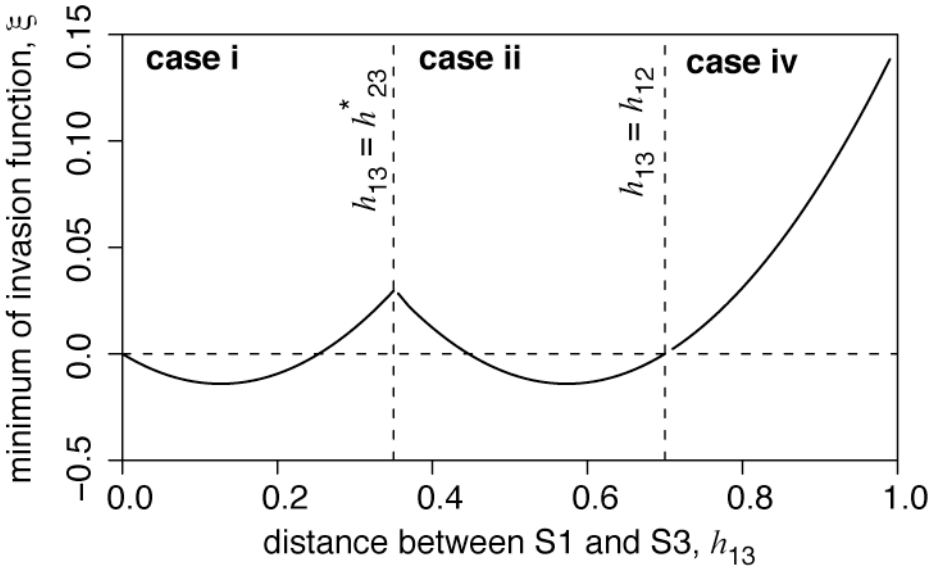

case i
case ii
case iv
$h_{13} = h^*_{23}$
$h_{13} = h_{12}$
minimum of invasion function, ξ
−0.5
0.0
0.05
0.10
0.15
0.0
0.2
0.4
0.6
0.8
1.0
distance between S1 and S3, h_{13}

Figure 5

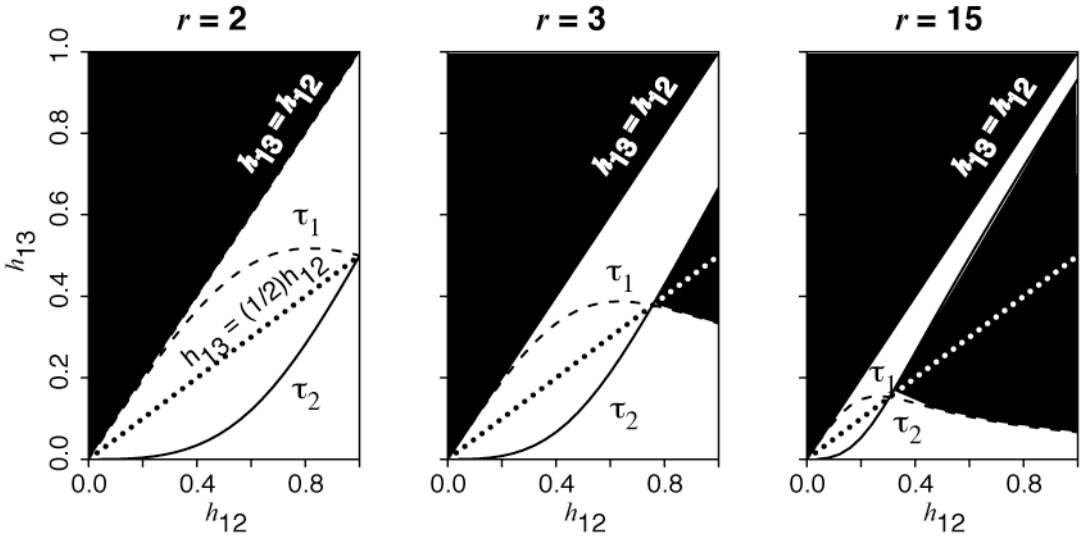


Figure 6

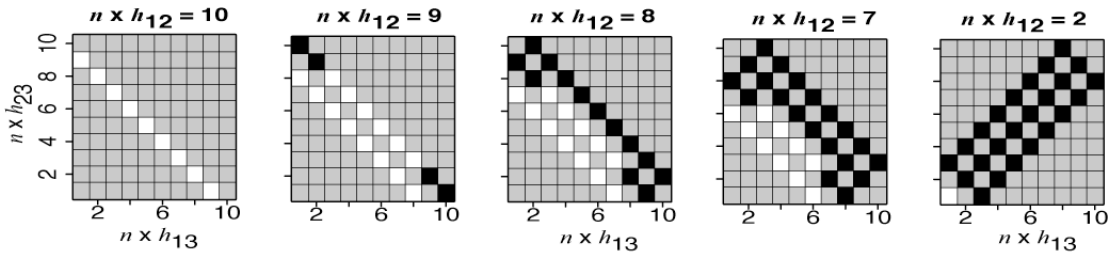

$n \times h_{12} = 10$
$n \times h_{12} = 9$
$n \times h_{12} = 8$
$n \times h_{12} = 7$
$n \times h_{12} = 2$
$n \times h_{23}$
2
4
6
8
10
$n \times h_{13}$

Figure 7

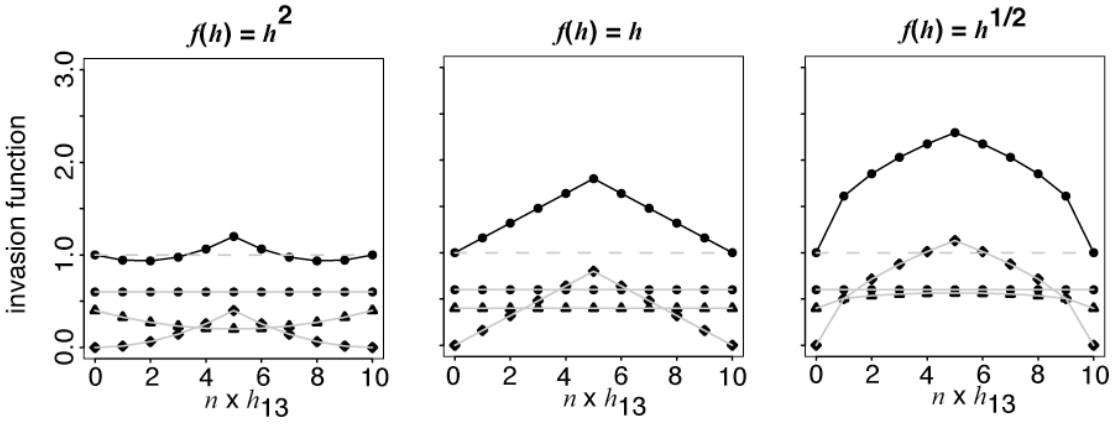


Figure 8

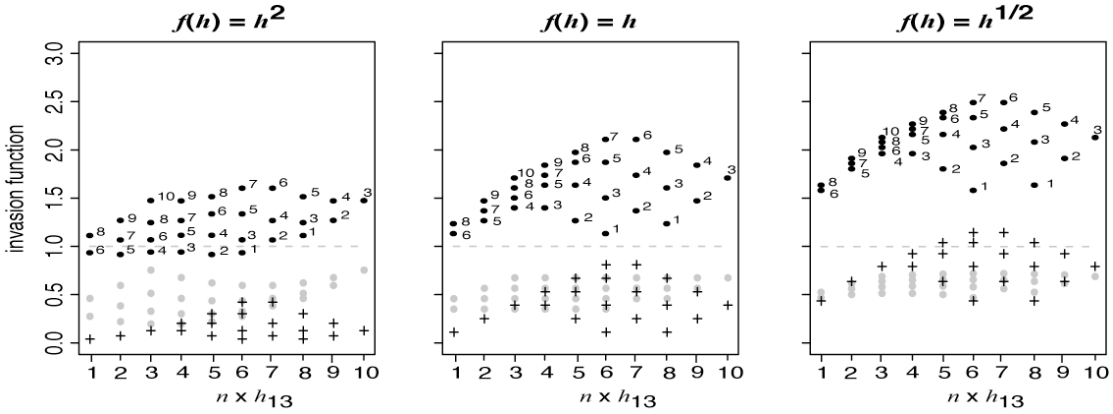

f(h) = h^2
f(h) = h
f(h) = h^1/2
invasion function
0.0
0.5
1.0
1.5
2.0
2.5
3.0
1 2 3 4 5 6 7 8 9 10
n × h13

Figure 9

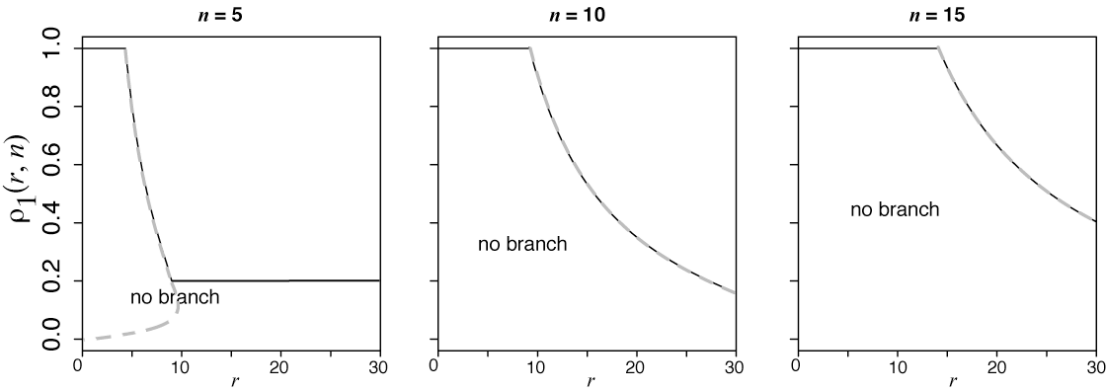
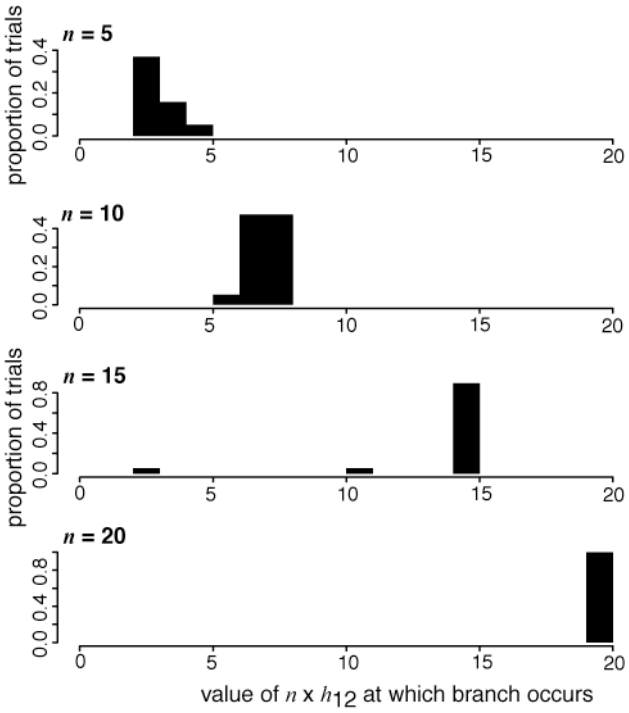

n = 5
n = 10
n = 15
ρ1(r, n)
0.0
0.2
0.4
0.6
0.8
1.0
no branch
no branch
no branch
0
10
20
30
r

Figure 10


proportion of trials
n = 5
0.0 0.2 0.4
0 5 10 15 20
n = 10
0.0 0.2 0.4
0 5 10 15 20
proportion of trials
n = 15
0.0 0.4 0.8
0 5 10 15 20
n = 20
0.0 0.4 0.8
0 5 10 15 20
value of n x h12 at which branch occurs