

# Role of stochastic processes in maintaining discrete strain structure in antigenically diverse pathogen populations

Caroline O. Buckee<sup>a,1</sup>, Mario Recker<sup>b</sup>, Eleanor R. Watkins<sup>b</sup>, and Sunetra Gupta<sup>b</sup>

<sup>a</sup>Center for Communicable Disease Dynamics, Department of Epidemiology, Harvard School of Public Health, Boston, MA 02115; and <sup>b</sup>Department of Zoology, University of Oxford, Oxford OX1 3PS, United Kingdom

Edited\* by Robert M. May, University of Oxford, Oxford, United Kingdom, and approved August 3, 2011 (received for review February 15, 2011)

**Many highly diverse pathogen populations appear to exist stably as discrete antigenic types despite evidence of genetic exchange. It has been shown that this may arise as a consequence of immune selection on pathogen populations, causing them to segregate permanently into discrete nonoverlapping subsets of antigenic variants to minimize competition for available hosts. However, discrete antigenic strain structure tends to break down under conditions where there are unequal numbers of allelic variants at each locus. Here, we show that the inclusion of stochastic processes can lead to the stable recovery of discrete strain structure through loss of certain alleles. This explains how pathogen populations may continue to behave as independently transmitted strains despite inevitable asymmetries in allelic diversity of major antigens. We present evidence for this type of structuring across global meningococcal isolates in three diverse antigens that are currently being developed as vaccine components.**

deterministic model | *Neisseria meningitidis* | microbial population dynamics | stochastic model

**M**any pathogen species, including those causing AIDS, malaria, and influenza, exhibit high levels of diversity among genes encoding antigenic proteins, enabling immune escape at both the individual and population levels. Many of the relevant antigens also have important functional roles in the life history of the pathogen, however. The potential antigenic diversity of pathogen populations can thus be represented by either a continuous or discrete “strain space” whose form and boundaries derive from their biological roles as well as considerations of molecular structure. The behavior of the pathogen population within this antigenic strain space is determined by the degree to which infection of a host with one antigenic type or strain generates cross-protective immune responses against other strains. This cross-protection may be (i) strain-transcending, where infection with one strain provides an equal level of protection against all other strains (1, 2); (ii) strain-specific, where strains have well-defined antigenic relationships with each other (3–7); or (iii) a combination of both (8, 9). In cases ii and iii, competition for hosts when immunological cross-protection is strong can cause the pathogen population to become polarized in strain space. In models in which immunity is directed against a specific number of antigenic loci, each with a finite number of allelic variants, this can manifest as the stable coexistence of a subset of discrete discordant strains with nonoverlapping antigenic variant repertoires (6, 7).

The emergence of strains with completely nonoverlapping combinations of antigenic alleles within these systems is a consequence of the inbuilt symmetry imposed by assuming equal levels of allelic diversity among loci, however. Theoretical studies have shown that the signature of immune selection can become considerably more complex in the absence of this symmetry (10, 11), such that dominant strains within pathogen populations may show substantial allelic overlap. It is therefore surprising that evidence of discrete strain structure has been found in several pathogen species. For example, the pathogenic bacterial species *Neisseria meningitidis* exhibits a clear nonoverlapping structure among hypervariable regions of immunodominant subcapsular

antigens at the population level (7, 12, 13). Populations of the relapsing fever spirochete *Borrelia hermsii* are also characterized by discrete independent strains in which repertoires of immunodominant outer membrane proteins in wild isolates are either identical or completely distinct from each other (14). The long-term population structure of two outer membrane proteins of global Group A *Streptococcus* exhibit distinct combinations of alleles with minimal overlap between strains (14, 15). Similarly, recent analysis of the tick-borne bacteria *Anaplasma marginale*, a prevalent livestock pathogen, has shown that the coinfection of cattle with different strains relies on the expression of non-overlapping variants of the major antigen family by superinfecting strains (16). For the human malaria parasite, *Plasmodium falciparum*, analysis of serological responses to disease-causing isolates in children indicates that parasites are behaving phenotypically as independent strains, despite the variety of diverse antigenic determinants expressed during infection (17–21). Although the specific epitopes involved in these responses have not yet been elucidated, wild parasite isolates in areas of intense transmission do not appear to share genetic sequences among the *var* gene family encoding the immunodominant variant surface antigen *P. falciparum* Erythrocyte Membrane Protein 1 (22). Recent analysis of a malaria parasite protein and vaccine candidate, Merozoite Surface Protein-3, has also shown this type of discrete pattern among its relevant epitope regions (23).

Given that immunodominant antigens are likely to contain epitopes with varying levels of diversity, imposed by their structural position and molecular role within the protein (24), the conformation of real pathogen populations to theoretical predictions based on assumptions of allelic symmetry is difficult to explain. Here, we show that stochasticities, inherent to dynamic processes of infection and immunity but often ignored in mathematical frameworks, can explain this paradox. We explore the effects of immune selection on a pathogen population defined by antigenic loci with varying degrees of functional constraints, as reflected by asymmetries in allelic diversity, using a stochastic model that explicitly includes recombination and mutation, and permits extinction of alleles. We show that these processes qualitatively change the resulting pathogen population structure and allow for the emergence of simple discrete strain structure over a wide range and variability of functional constraints on the relevant antigens. We examine a global database of meningococcal isolates and find that regional diversities are similar at three different antigenic loci, despite substantial differences in their diversities. Our results provide an important example in which a stochastic individual-based model is more successful in reconciling pattern with process than its deterministic counterpart (25).

Author contributions: C.O.B. and S.G. designed research; C.O.B., M.R., and E.R.W. performed research; C.O.B., M.R., and E.R.W. analyzed data; and C.O.B., M.R., and S.G. wrote the paper.

The authors declare no conflict of interest.

\*This Direct Submission article had a prearranged editor.

<sup>1</sup>To whom correspondence should be addressed. E-mail: cbuckee@hsph.harvard.edu.

This article contains supporting information online at [www.pnas.org/lookup/suppl/doi:10.1073/pnas.1102445108/-DCSupplemental](http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1102445108/-DCSupplemental).

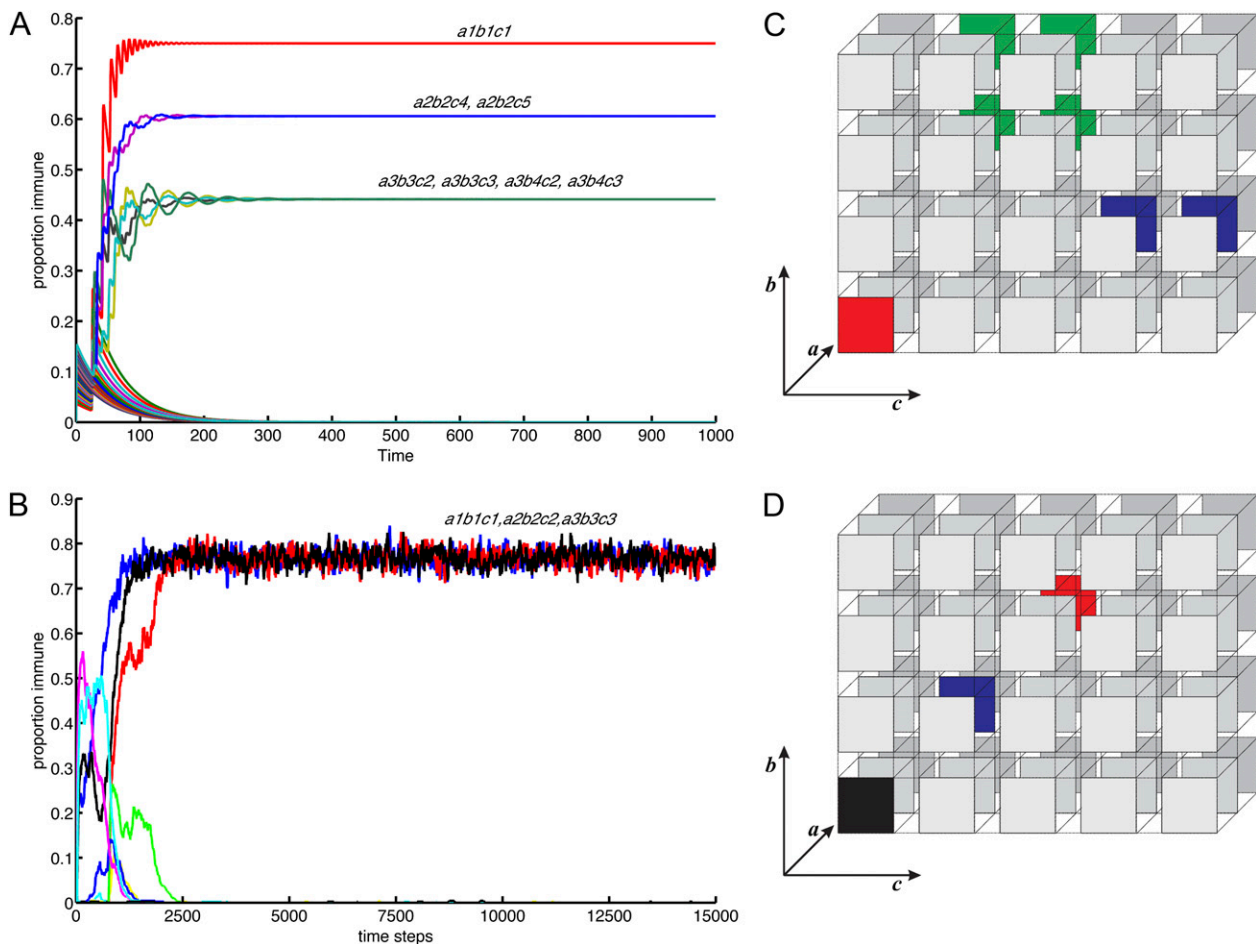
## Results

We developed stochastic analogs of multilocus models (6) in which pathogens were defined by two or three antigenic loci, with each locus having the same (symmetric) or different (asymmetric) number of possible alleles. To compare the deterministic and stochastic model outcomes, we used two metrics of population structure: diversity,  $D$ , and discordance,  $f^*$ , measuring the distribution of strain prevalence and the level of relatedness (i.e., overlap in variant repertoires) in the circulating strains, respectively (more details are presented in *Methods*). For symmetric systems, the general dynamics of both deterministic and stochastic models are equivalent to those described by Gupta et al. (7), and there is a clear relationship between the diversity and discordance of the pathogen population and the strength of immunological cross-protection ( $\gamma$ , the original model dynamics are illustrated in Fig. S1). When  $\gamma$  is low, all pathogen strains can be found at equal prevalence, resulting in maximum diversity ( $D = 1$ ). Because strains with overlapping antigenic determinants coexist, the degree of nonoverlapping structure or discordance ( $f^*$ ) is also very low. When  $\gamma$  is high, diversity drops and the degree of discordance rises as a subset of nonoverlapping strains emerges (Fig. S2).

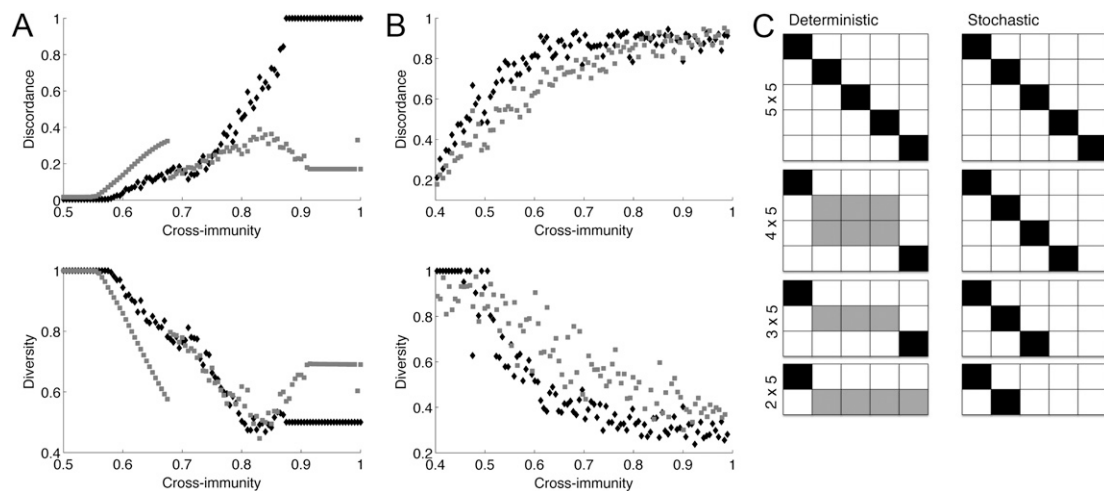
Asymmetries in the number of alleles at each locus introduce important differences between the deterministic and stochastic models, however; Fig. 1 *A* and *B* show simulations for the deterministic and stochastic models, respectively, of a multilocus

system under strong immune selection with three, four, and five alleles at each of three loci. For the deterministic model (Fig. 1*A*), the asymmetric case results in a highly structured equilibrium in which dominant and subdominant sets of strains are maintained and strains with entirely nonoverlapping combinations of alleles cannot exclude all others because of the geometry of the system. Here, alleles at the least diverse locus are associated with more than one allele at the more diverse loci because these have an equal competitive advantage. Fig. 1*C* illustrates the dimensions of this strain space at equilibrium. The stochastic model simulation (Fig. 1*B*) diverges significantly from its deterministic counterpart: Only 3 of the possible 60 strains continue to coexist. In other words, under high levels of immune selection, most of the alleles initially present (or generated by means of mutation) at the more diverse loci are driven to extinction. Fig. 1*D* illustrates the corresponding dimensions of strain space for the stochastic simulation, showing the contrast between the highly structured patterns of dominant and subdominant strains in the deterministic system and the simple nonoverlapping structures in the stochastic system.

The impact of these contrasting dynamics on the discordance and diversity of the pathogen populations within the model systems is shown in Fig. 2. For the deterministic model, the asymmetric case results in a number of complex and variably structured populations in which dominant and subdominant sets of strains are maintained at equilibrium as cross-immunity ( $\gamma$ ) increases. As



**Fig. 1.** Examples of time series for deterministic (*A*) and stochastic (*B*)  $3 \times 4 \times 5$  allele models showing the proportion of the population immune to particular strains. Each strain is labeled according to the locus (*a*, *b*, and *c*), with a subscript representing the allele (between 1 and 5). Parameter values for the deterministic model:  $\beta = 4$ ,  $\sigma = 1$ ,  $\mu = 0.02$ . (*C*) Dimensions of strain space at equilibrium for the deterministic model corresponding to *A*. (*D*) Dimensions of strain space at equilibrium for the stochastic model corresponding to *B*. In both, the three-allele locus is represented as layers one behind the other and the strains are colored according to the allele expressed at this locus (e.g., the green squares in *C* all share allele 3 at locus *a*).



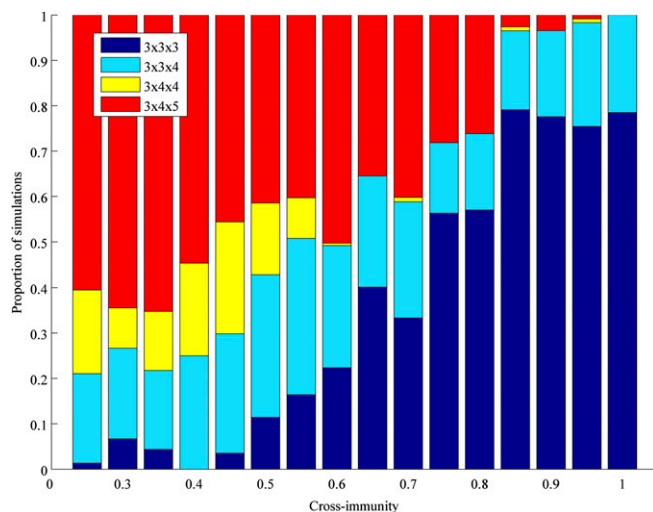
**Fig. 2.** Comparison of the relationship between diversity/discordance and cross-immunity for the symmetric  $5 \times 5$  system (black diamonds) and an asymmetric  $2 \times 5$  system (gray squares) for the deterministic (*A*) and stochastic (*B*) models. Each point in the stochastic figures represents the mean of simulations for which cross-immunity fell within a bin of 0.02. Parameter values for the deterministic model:  $\beta = 4$ ,  $\sigma = 1$ ,  $\mu = 0.02$ . (*C*) Patterns of dominance observed at high levels of cross-immunity for equivalent deterministic and stochastic two-locus systems. Alleles of locus 1 are represented by rows, and alleles of locus 2 are represented by columns, such that each square in the matrix represents one allelic combination. Black squares represent dominance at equilibrium of particular combinations, and white squares represent the absence of particular combinations. Gray squares represent subdominant (in terms of prevalence) allelic combinations.

a result, higher levels of diversity can be maintained even when  $\gamma$  is high. Fig. 2*A* illustrates this result, showing the difference in the diversity and discordance profiles for the symmetric and asymmetric deterministic systems for the two-locus case, where (*i*) each locus has five alleles (in black) and (*ii*) one locus has two alleles and one locus has five alleles (in gray). Fig. 2*B* shows the same metrics for the stochastic version. It is clear from these figures that the stochastic model diverges significantly from its deterministic counterpart as immune selection ( $\gamma$ ) increases. Although the  $2 \times 5$  system maintains slightly more diversity and shows more overlap among strains than the symmetric system at intermediate levels of immune selection, when immune selection is strong, only 2 of the possible 10 strains continue to coexist. Thus, both the symmetric and asymmetric systems show the same high degree of discordance and low levels of diversity under very strong immune selection, because the  $2 \times 5$  system effectively collapses to a symmetric  $2 \times 2$  system. This reversion to symmetry facilitates the dominance of completely nonoverlapping subsets of strains and is stable even in the presence of extremely high rates of recombination and mutation. Fig. 2*C* illustrates the structure of strain space for different deterministic and stochastic systems at equilibrium when cross-immunity is strong (and see Figs. S3 and S4).

The extinction of alleles occurring in the stochastic system allows the pathogen population to reduce competition among strains, albeit at the expense of diversity. Fig. 3 shows the distribution of equilibrium outcomes for increasing levels of immune selection within the stochastic framework. As cross-immunity increases, pathogen populations become increasingly dominated by strains with entirely nonoverlapping antigen repertoires. The dimensions of strain space among these populations were therefore determined by the number of distinct allelic forms possible at the most constrained locus (in this case, the locus with 3 allelic variants), and the emergence of diversity among the less constrained loci becomes unfeasible with increasing competition. An exploration of the effects of host contact network structure was performed to determine how local mixing might affect the structure of these pathogen populations. Consistent with our previous work (26), although a host population characterized only by local mixing produces less structured pathogen populations, only a few long-range connections (coupled with strong cross-immunity) are sufficient to structure the entire pathogen population into symmetric nonoverlapping combinations of alleles (Fig. S5). Importantly, the stochastic model is necessarily constrained in its

population size. We have shown that increasing the population size does not change the outcome at equilibrium (Fig. S6), but we do expect that the speed at which equilibrium is reached will be increased in small populations.

To determine whether the structuring observed in our model output was evident in real pathogen populations, we interrogated a global database of meningococcal isolates ([www.pubmlst.org](http://www.pubmlst.org)). We compiled a list of allelic variants at two meningococcal outer membrane antigens that exhibit considerable diversity and have been included in strain-specific vaccines because of their immunogenicity (27, 28). For each region, we examined the hypervariable regions (VR1 and VR2) of the PorA locus, which may be considered as separate epitopes because they recombine freely (12), and the diverse FetA locus. We have previously shown nonoverlapping structure among these antigens in one region and explored the dynamics of these antigens over time with respect to housekeeping



**Fig. 3.** Relationship between the dimensions of strain space and cross-immunity for the three-locus  $3 \times 4 \times 5$  stochastic model. Stacked bars represent the proportion of simulations with particular dimensions at equilibrium.



**Table 1.  $f^*$  values calculated for the dominant strains (those shown in the darker shades in Fig. 4B) from each country, with sample sizes ( $n$ ) also shown**

	Sample size ( $n$ )	$f^*$ PorA VR1/VR2	$f^*$ PorA VR1/FetA	$f^*$ PorA VR2/FetA
United Kingdom	335	0.690	0.253	0.30
France	126	1	0.519	1
Germany	265	0.688	0.376	0.498
The Netherlands	277	0.782	1	1
Brazil	77	0.809	1	0.712
United States	117	0.797	1	1
Cuba	408	0.957	0.712	0.951
Czech Republic in 2000s	657	0.548	0.274	0.272
Czech Republic in 1990s	297	0.721	0.662	0.911
Czech Republic in 1970s	286	0.939	0.475	0.541
Mean	285	0.793	0.627	0.719

global isolates (Fig. 4B). For PorA VR1, the number of alleles was more or less constant for up to 90% of all isolates. We believe this pattern is consistent with our hypothesis that the diversity of PorA VR1 may be constraining the diversities of PorA VR2 and FetA.

## Discussion

Understanding the evolutionary forces shaping the structure of pathogen populations is an important goal of mathematical models. Although deterministic models provide important insights into these mechanisms, stochastic processes, such as extinction, can play a crucial role in the real world. Deterministic frameworks (10, 11) have indicated that asymmetries in the allelic diversities of different immunodominant antigens, attributable to such factors as varying functional constraints, can lead to a higher degree of antigenic overlap among prevailing strains within pathogen populations than expected when loci are equally diverse. Even when immune selection is strong, these allelic asymmetries can prevent the emergence of entirely nonoverlapping strains in the deterministic model. However, a stochastic version of the model illustrates how the extinction of alleles at more diverse loci can lead to the recovery of a symmetric strain space characterized by entirely nonoverlapping strains. Many important pathogens have multiple immunodominant antigenic determinants, and it is extremely likely that some will be more functionally or otherwise restricted in their diversities. The results presented here may provide the answer to how so many such pathogen populations continue to behave as independently transmitted strains despite these inevitable asymmetries. We propose that the stochastic processes delineated here can explain the success of simple symmetric and deterministic models in describing the population structure of pathogens with multiple immunodominant loci ranging from *N. meningitidis* to *P. falciparum* and *A. marginale*.

There have been various theoretical studies on the effect of host immunity on the population structures of antigenically diverse pathogens. The study presented here concentrates on systems in which the relationship between any two pathogenic variants or strains is clearly defined by an overlap in their respective antigenic repertoire at loci under immune selection. As such, two strains will interfere with each other's transmission through the action of cross-immunity if they have one or more alleles in common, and the actual degree of overlap (in terms of common/shared alleles) does not alter the degree of interference. It will be interesting to see how our results apply to systems where the degree of antigenic overlap also determines the degree of cross-immunity (11, 29); we may hypothesize, however, that these systems will also tend toward complete discordance when stochastic extinction occurs.

Our results indicate that the extinction of alleles at more diverse antigenic loci will occur more rapidly in small or spatially structured host populations. The effect of population size on

pathogen diversity has previously been demonstrated, for example, by Abu-Raddad and Ferguson (1), although the system considered in their study allowed for unconstrained genetic/antigenic evolution and the results crucially depended on relative pathogen fitness and the rate at which strains are introduced and replaced within the population. In contrast, our results do not depend on those parameters and are robust to changes in the shape and dimension of strain-space.

The data presented here provide evidence for the signature of these processes occurring within a range of globally distributed populations of *N. meningitidis*. Our model framework can explain how the enormous diversity of certain meningococcal antigens may be constrained by the levels of diversity present at another antigenic locus. Note that we have previously shown that the nonoverlapping structure at these antigenic loci is not found among housekeeping genes in the same isolates (12), highlighting the fact that recombination can allow for divergent evolutionary trajectories among genes within the same isolates.

Our framework also has important ramifications for invasion of endemic populations by new strains, because pathogens must acquire different epitopes or alleles at every immunodominant locus to invade successfully. The emergence of a new strain of meningococcal bacteria in the Czech Republic provides a good example of this phenomenon. Against a background of stable and nonoverlapping combinations of antigenic variants at three immunodominant loci circulating in the meningococcal population in the 1970s and 1980s, a new lineage emerged in the population in 1993 expressing previously unobserved antigenic variants at all three loci in our study (PorA VR1, VR2, and FetA), rising to a relatively high prevalence and causing significant amounts of disease (12).

The dynamic implications of the stochastic model framework extend also to such pathogens as influenza, in which strains appear to emerge sequentially along a diagonal pathway through antigen space (30). The exploration of strain space by the pathogen population in our simulations follows a trajectory in which new strains can successfully emerge only when they acquire different alleles (or variants) at all loci (or epitope regions). This behavior is consonant with models of influenza that either rely on restrictions on epitope diversity within a discrete framework, as suggested by Recker et al. (31), or propose a more continuous structure to strain space with cross-immunity decaying along a number of different epitope axes (32). Such factors as short-term cross-immunity (33) are likely to reinforce this behavior and, together with a high level of redundancy in genotype-phenotype mapping (34), may be necessary to generate the phylogenetic patterns characteristic of influenza. An important future direction for this work will be to explore the relationship between these model frameworks and phylogenetic data for a variety of other pathogen systems.

## Methods

**Deterministic Model.** The model is a direct extension of the one introduced by Gupta et al. (6) incorporating asymmetric numbers of alleles. Here, pathogen strains are defined by up to three immunodominant loci,  $a$ ,  $b$ , and  $c$ , with  $i$ ,  $j$ , and  $k$  alleles, respectively:  $a_1 \dots a_i$ ,  $b_1 \dots b_j$ , and  $c_1 \dots c_k$ . This allows for a maximum of  $i*j*k$  different combinations of alleles that constitute a strain. The host population is divided into three partially overlapping compartments for each strain  $s$ : the proportion infectious with strain  $s$ ,  $y_s$ ; the proportion immune (here, infected or previously exposed) to strain  $s$ ,  $z_s$ ; and the proportion immune to any strain that shares alleles with  $s$ ,  $w_s$ . The rate of change in these compartments is given by the following set of differential equations:

$$\begin{aligned} \frac{dz_s}{dt} &= \beta y_s (1 - z_s) - \mu z_s \\ \frac{dw_s}{dt} &= \sum_s \beta y_s (1 - w_s) - \mu w_s \\ \frac{dy_s}{dt} &= \beta y_s ((1 - w_s) + (1 - \gamma)(w_s - z_s)) - \sigma y_s \end{aligned}$$

For simplicity, it is assumed that sharing one allele has the same protective effect as sharing more than one. The strength of cross-immunity is determined by the parameter  $\gamma$ , and all strains are assumed to have the same transmissibility,  $\beta$ . The duration of infectiousness is  $1/\sigma$  and is short compared with

the average lifespan  $1/\mu$ , (i.e.,  $\sigma \gg \mu$ ); immunity is assumed to be lifelong.  $R_0$ , the basic reproduction ratio of a strain, is simply given by  $\beta/\sigma$ . Recombination is not explicitly modeled, and all possible strains are present from the start.

**Stochastic Model.** An individual-based stochastic model was also developed, in which a population of hosts is infected by a directly transmitted and antigenically diverse pathogen species. Hosts are connected to each other on a dynamic random-mixing network (26), which is seeded randomly with different strains of the pathogen. Hosts remain infectious (for a period  $1/\sigma$ , on average, where  $\sigma$  is the rate of loss of infectiousness) before recovering and maintain allele-specific immunity for a given duration after this ( $1/\zeta$ , on average, where  $\zeta$  is the rate of loss of immunity). Pathogen strains are defined by antigenic alleles at either two or three loci. Each locus has a defined number of alleles, such that strain space may be envisaged as either a 2D matrix or a 3D rectangular box of allelic associations. Infection with each strain generates allele-specific immunity within the host. The susceptibility of a host to infection by a new strain is then dependent on how many alleles it shares with previously seen strains as well as its transmissibility, similar to previous models (6, 26). It is assumed that all strains are equally transmissible, and two strains coinfecting the same host could undergo recombination with a defined probability. In addition, each locus is subject to mutation, allowing for all possible areas of strain space to be explored during a simulation and every strain to emerge if it could be competitive. Table S1 shows the range of parameters explored. A total of 2,500 simulations were run in which the pathogen population is defined by three loci with 3, 4, and 5 alleles, respectively, just as in the deterministic model. This yielded a strain space described by a  $3 \times 4 \times 5$  matrix of allelic associations. A further 3,600 simulations were run for a two-locus model, 1,000 with a symmetric number of alleles (5 at each locus), 600 with 2 alleles at one locus and 5 alleles at another, and 2,000 with 10 alleles at one locus and 25 alleles at another. These simulations explored the effects of increased numbers of alleles, as well as changing host population size. To explore spatial effects within the host population and the impact of population size, we ran 7,000 simulations with varying host contact network structure and population size, as described in *SI Text*.

**Metrics for the Evaluation of Diversity and Discordance.** To describe the pathogen population, we used a measure of diversity ( $D$ ) based on the

Shannon–Weaver index of diversity, previously used for this kind of system (26). This measure describes how evenly the pathogen population is distributed across all possible strains and is given as

$$D = -\frac{1}{\log(N)} \sum_{s=1..N} p_s \log(p_s),$$

where  $N$  is the total number of possible strains and  $p_i$  is the proportionate prevalence of strain  $i$  in the population (i.e.,  $p_s = y_s/y$  with  $y = \sum_s y_s$ ).

In addition, a metric for discordance or lack of overlap,  $f^*$ , was used to measure the extent to which pathogen strains share alleles (12). Discordance can be calculated by evaluating the mean “dominance” of the most prevalent composition of a particular allele averaged over all alleles, where dominance refers to the relative prevalence of strains expressing particular allelic combinations. That is, for each allele  $i$  (or  $j$  or  $k$ ), we find the most prevalent strain containing allele  $i$  and calculate its dominance over all other strains containing this particular allele. In the context of a two-locus system, this can be understood as calculating the “row” and “column” dominance for each allele within a matrix of allelic associations; in the three-locus model, we also take into account each allele’s relative dominance with respect to all three axes (row, column, and “depth” within a 3D box of allelic associations). As a three-locus example, let  $a_0 b_j c_k$  be the most prevalent of all strains containing allele  $a_0$ . Its dominance compared with all other strains containing  $a_0$  or  $b_j c_k$  can be calculated as

$$f_{a_0} = \frac{Y_{a_0 b_j c_k}}{\sum_i Y_{a_i b_j c_k}} \cdot \frac{Y_{a_0 b_j c_k}}{\sum_m Y_{a_0 b_m c_k}} \cdot \frac{Y_{a_0 b_j c_k}}{\sum_n Y_{a_0 b_j c_n}} \cdot \frac{Y_{a_0 b_j c_k}}{\sum_l \sum_m \sum_n Y_{a_l b_m c_n}}$$

As the averaged mean of all alleles,  $f^*$  is then given as

$$f^* = \frac{1}{i+j+k} \left( \sum_i f_{a_i} + \sum_j f_{b_j} + \sum_k f_{c_k} \right).$$

Therefore,  $f^*$  scores close to 1 indicate that combinations of alleles at three loci share very few alleles, whereas lower scores are indicative of overlapping associations.

- Abu-Raddad LJ, Ferguson NM (2004) The impact of cross-immunity, mutation and stochastic extinction on pathogen diversity. *Proc Biol Sci* 271:2431–2438.
- Andreasen V, Lin J, Levin SA (1997) The dynamics of cocirculating influenza strains conferring partial cross-immunity. *J Math Biol* 35:825–842.
- Gog JR, Grenfell BT (2002) Dynamics and selection of many-strain pathogens. *Proc Natl Acad Sci USA* 99:17209–17214.
- Gog JR, Swinton J (2002) A status-based approach to multiple strain dynamics. *J Math Biol* 44:169–184.
- Gomes MG, Medley GF, Nokes DJ (2002) On the determinants of population structure in antigenically diverse pathogens. *Proc Biol Sci* 269:227–233.
- Gupta S, Ferguson N, Anderson R (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. *Nat Med* 2:437–442.
- Gupta S, Galvani A (1999) The effects of host heterogeneity on pathogen population structure. *Philos Trans R Soc Lond B Biol Sci* 354:711–719.
- Dawes JH, Gog JR (2002) The onset of oscillatory dynamics in models of multiple disease strains. *J Math Biol* 45:471–510.
- Minayev P, Ferguson N (2009) Improving the realism of deterministic multi-strain models: Implications for modelling influenza A. *J R Soc Interface* 6:509–518.
- Recker M, Gupta S (2005) A model for pathogen population structure with cross-protection depending on the extent of overlap in antigenic variant repertoires. *J Theor Biol* 232:363–373.
- Buckee CO, Gupta S, Kriz P, Maiden MC, Jolley KA (2010) Long-term evolution of antigen repertoires among carried meningococci. *Proc Biol Sci* 277:1635–1641.
- Buckee CO, et al. (2008) Role of selection in the emergence of lineages and the evolution of virulence in *Neisseria meningitidis*. *Proc Natl Acad Sci USA* 105:15082–15087.
- Jódar L, Feavers IM, Salisbury D, Granoff DM (2002) Development of vaccines against meningococcal disease. *Lancet* 359:1499–1508.
- Johnson DR, Kaplan EL, VanGheem A, Facklam RR, Beall B (2006) Characterization of group A streptococci (*Streptococcus pyogenes*): Correlation of M-protein and emm-gene type with T-protein agglutination pattern and serum opacity factor. *J Med Microbiol* 55:157–164.
- Futse JE, Brayton KA, Dark MJ, Knowles DP, Jr., Palmer GH (2008) Superinfection as a driver of genomic diversification in antigenically variant pathogens. *Proc Natl Acad Sci USA* 105:2123–2127.
- Bull PC, et al. (2000) *Plasmodium falciparum*-infected erythrocytes: Agglutination by diverse Kenyan plasma is associated with severe disease and young host age. *J Infect Dis* 182:252–259.
- Bull PC, Lowe BS, Kortok M, Marsh K (1999) Antibody recognition of *Plasmodium falciparum* erythrocyte surface antigens in Kenya: Evidence for rare and prevalent variants. *Infect Immun* 67:733–739.
- Marsh K, Howard RJ (1986) Antigens induced on erythrocytes by *P. falciparum*: Expression of diverse and conserved determinants. *Science* 231:150–153.
- Newbold CI, Pinches R, Roberts DJ, Marsh K (1992) *Plasmodium falciparum*: The human agglutinating antibody response to the infected red cell surface is predominantly variant specific. *Exp Parasitol* 75:281–292.
- Vigan-Womas I, et al. (2011) Allelic diversity of the *Plasmodium falciparum* erythrocyte membrane protein 1 entails variant-specific red cell surface epitopes. *PLoS ONE* 6:e16544.
- Bull PC, et al. (2005) *Plasmodium falciparum* variant surface antigen expression patterns during malaria. *PLoS Pathog* 1:e26.
- Polley SD, et al. (2007) *Plasmodium falciparum* merozoite surface protein 3 is a target of allele-specific immunity and alleles are maintained by natural selection. *J Infect Dis* 195:279–287.
- Frank SA, Bush RM (2007) Barriers to antigenic escape by pathogens: Trade-off between reproductive rate and antigenic mutability. *BMC Evol Biol* 7:229.
- Levin SA, Grenfell B, Hastings A, Perelson AS (1997) Mathematical and computational challenges in population biology and ecosystems science. *Science* 275:334–343.
- Buckee CO, Koelle K, Mustard MJ, Gupta S (2004) The effects of host contact network structure on pathogen diversity and strain structure. *Proc Natl Acad Sci USA* 101:10839–10844.
- Bennett JS, Thompson EA, Kriz P, Jolley KA, Maiden MC (2009) A common gene pool for the *Neisseria Feta* antigen. *Int J Med Microbiol* 299:133–139.
- Urwin R, et al. (2004) Distribution of surface protein variants among hyperinvasive meningococci: Implications for vaccine design. *Infect Immun* 72:5955–5962.
- Calvez V, Korobeinikov A, Maini PK (2005) Cluster formation for multi-strain infections with cross-immunity. *J Theor Biol* 233:75–83.
- Smith DJ, et al. (2004) Mapping the antigenic and genetic evolution of influenza virus. *Science* 305:371–376.
- Recker M, Pybus OG, Nee S, Gupta S (2007) The generation of influenza outbreaks by a network of host immune responses against a limited set of antigenic types. *Proc Natl Acad Sci USA* 104:7711–7716.
- Kryazhimskiy S, Dieckmann U, Levin SA, Dushoff J (2007) On state-space reduction in multi-strain pathogen models, with an application to antigenic drift in influenza A. *PLOS Comput Biol* 3:e159.
- Ferguson NM, Galvani AP, Bush RM (2003) Ecological and immunological determinants of influenza evolution. *Nature* 422:428–433.
- Koelle K, Cobey S, Grenfell B, Pascual M (2006) Epochal evolution shapes the phylogenetics of inter-pandemic influenza A (H3N2) in humans. *Science* 314:1898–1903.