Supporting Information

Buckee et al. 10.1073/pnas.1102445108

SI Text

Exploring an Increased Number of Alleles at Antigenic Loci. For both the stochastic and deterministic models, we explored the effects of increasing the numbers of alleles at each locus. Fig. S3 illustrates the final dimensions of strain space from stochastic and deterministic versions of models in which one antigen has 10 alleles and the other has 25. The same dynamics are observed as in the lower dimensional systems, with the stochastic extinction of alleles at the most diverse locus reducing the complexity of antigen space in the stochastic model and the highly structured maintenance of allelic diversity in the deterministic model. Fig. S4 illustrates a range of equilibrium outcomes for the deterministic model when considering different combinations of diversities among the antigens.

Increasing Population Size Affects the Time to Equilibrium but Not Model Outcomes. To determine how population size might affect the stochastic model outcomes, we explored 2,000 simulations with population sizes varying between 500 and 1,500 hosts, keeping all other parameters the same (C = 12, $1/\mu = 7$, $1/\zeta = 29$, $\beta = 0.5$, $R = 0.05$, $\tau = 0.025$, $\gamma = 3.8$). We found that although there was a nonsignificant increase in the time taken to reach equilibrium, the overall outcome was not affected by population size (within this range). Fig. S6 shows how the mean discordance for simulations varied with increasing population size, showing a slight but nonsignificant decrease with population size. Although running much larger simulations was beyond the scope of this work, we believe that greatly increased population sizes may show variable dynamics and reach equilibrium more slowly. In real pathogen populations, therefore, we expect nonoverlapping structure to be most pronounced in small populations, where competition and herd immunity are most intense.

Effects of Host Contact Network Structure on Strain Structure. Lastly, we explored how the underlying host contact structure might affect the stochastic model by running 5,000 simulations varying the contacts among hosts using the small-world networks developed by Watts and Strogatz (1) and modified for this type of system earlier (2). Here, the parameter ρ corresponds to the probability that an individual's contacts are random in the population as opposed to local; hence, $\rho = 0$ is an entirely locally mixing population and $\rho = 1$ is an entirely randomly mixed population. Fig. S5 illustrates the effects of changing ρ from local to random (with a small-world structure between) on the strain structure of the pathogen. This figure was generated from 5,000 simulations run with strong cross-immunity and the same parameter values in each simulation, with 10 alleles at one locus and 25 at the other, with only host contact structure changing [here, parameters are as follows: C = 12, $1/\mu = 7$, $1/\zeta = 29$, $\beta = 0.5$, $R = 0.05$, $\tau = 0.025$, $\gamma = 3.8$, host population size (P) = 1,500]. As ρ increases, the mean discordance of the pathogen population increased rapidly and then remained relatively constant, and the time to reach this f^* value decreased. Thus, the pathogen population took a longer time to equilibrate with localized host contacts, and the pathogen population was more overlapping than would be expected in the random mixing case. Note that, as seen in our previous studies, even a relatively low ρ resulted in a strong nonoverlapping structure (about 0.2 in these simulations). The rise in nonoverlapping structure appears to be delayed in this system, however, compared with our previous study (occurring when ρ is greater than 0.1). This suggests that the asymmetrical diversities of antigenic loci here do hinder the emergence of discrete strains when hosts mix very locally.

1. Watts DJ, Strogatz SH (1998) Collective dynamics of 'small-world' networks. *Nature* 393:440–442.

2. 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.

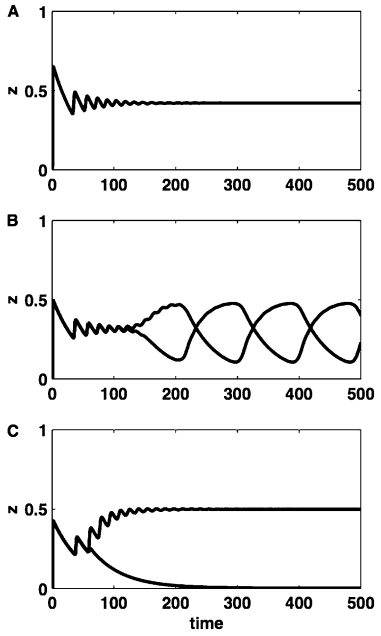

Fig. S1. Model dynamics of the original framework under low cross-immunity (here, z on the *y* axis), producing no strain structure, where all strains exist at the same prevalence (*A*); medium cross-immunity (cyclical or chaotic strain structure) (*B*); and high (discrete strain structure) cross-immunity (*C*). CSS, cyclical or chaotic strain structure; DSS, discrete strain structure; NSS, no strain structure. Strains are defined by two loci each with two alleles (2×2 system). The two lines represent different pairs of nonoverlapping strains, for example a1b1 and a2b2, where a and b are the loci and 1 and 2 are the alleles. The second line represents the complementary pair of strains, in this case a1b2 and a2b1. The *y* axis shows the prevalence.

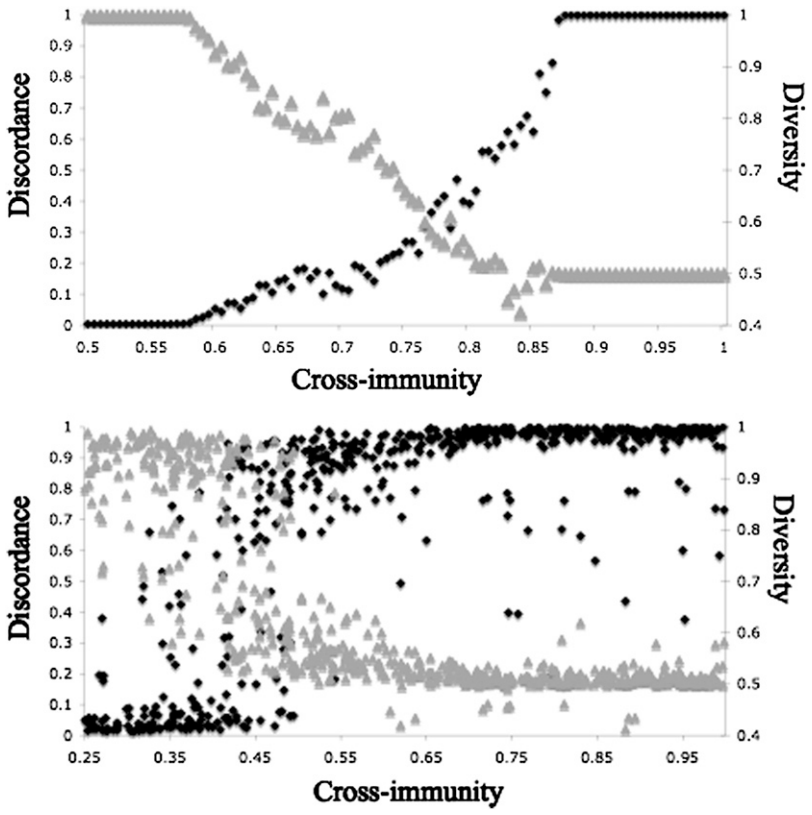

Fig. S2. Diversity and discordance profiles for deterministic (*Upper*) and stochastic (*Lower*) models for the symmetrical two-loci, five-allele (5×5) case. Black diamonds represent the discordance, and gray triangles represent the diversity. For the stochastic model, each point represents one simulation. Parameter values for the deterministic model: $\beta = 4$, $\sigma = 1$, $\mu = 0.02$.

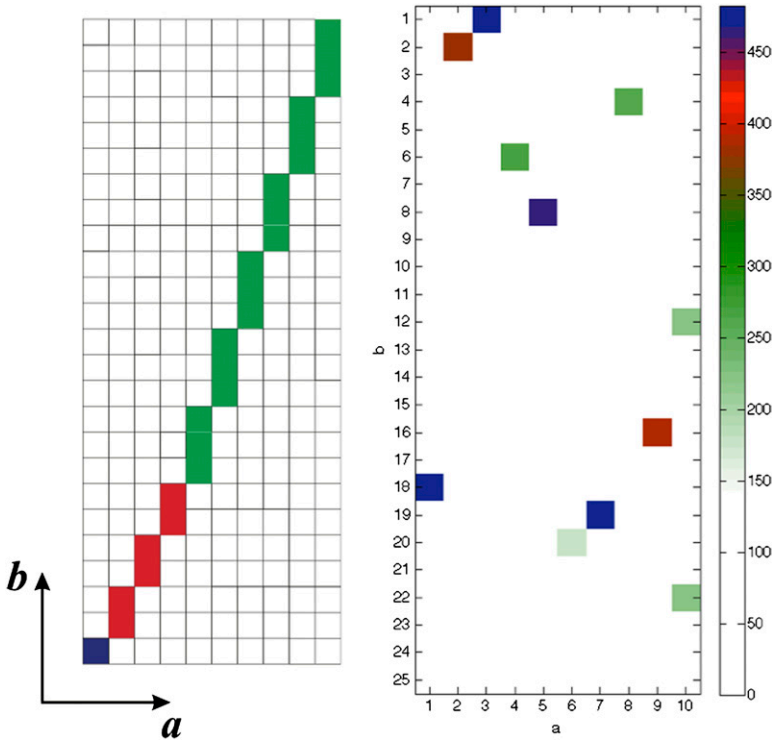

Fig. S3. Strain space for increased numbers of alleles, 10 at locus *a* and 25 at locus *b*. (*Left*) Deterministic model strain space at equilibrium. (*Right*) Stochastic model strain space after 15,000 time steps, with the color bar showing numbers of hosts immune in colors roughly corresponding to the proportions immune shown on the left.

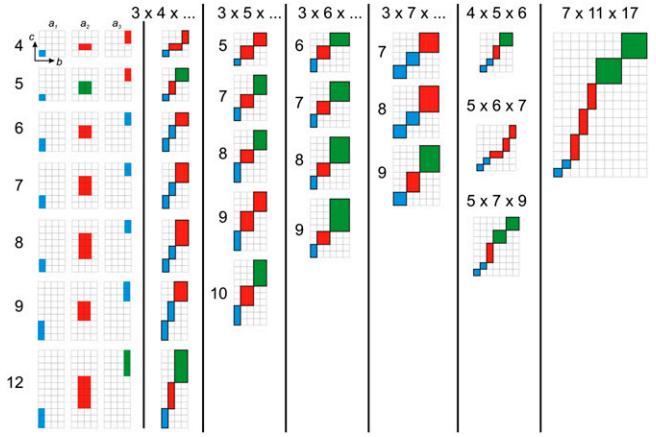

Fig. S4. Equilibrium strain spaces for different combinations of alleles for the deterministic model. Prevalence is indicated by color, with blue > red > green.

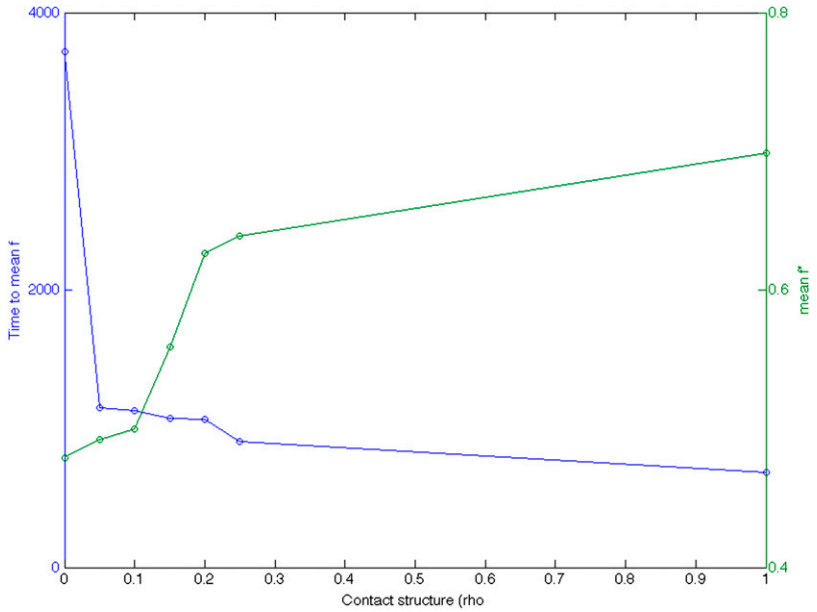

Fig. S5. Effect of host contact network structure on pathogen population structure. The blue line shows the first time step in the simulation, where the population structure reaches or exceeds the mean f^* value for that simulation (all populations start with a very low f^* value in general). The green line shows the mean f^*. Interestingly, the two lines are not exact mirrors of each other.

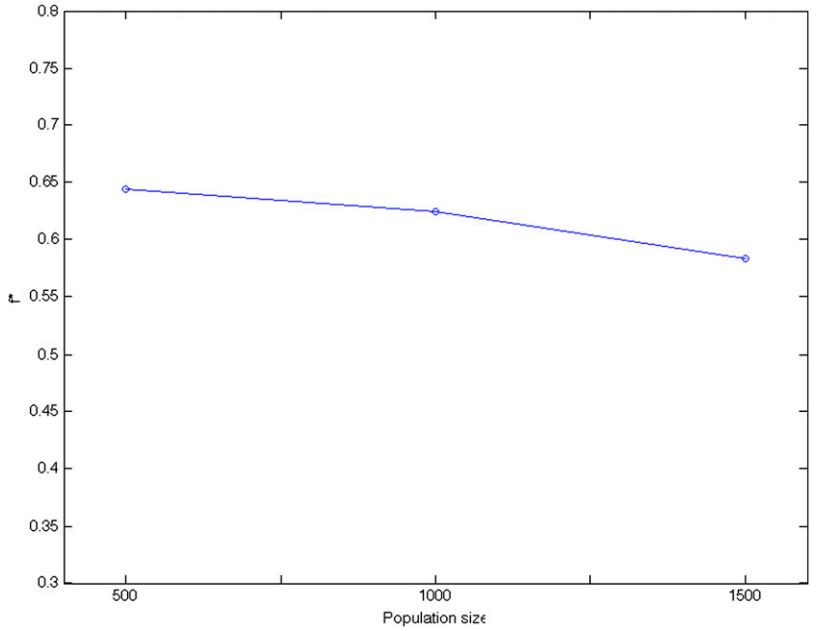

Fig. S6. Effect of population size on pathogen population structure in the stochastic model.

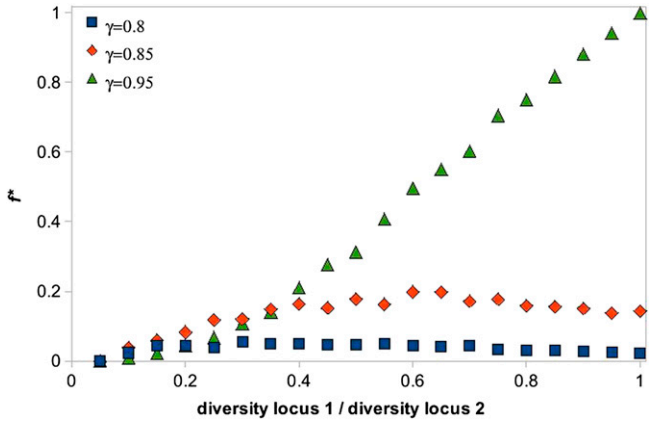

Fig. S7. Values of *f** for the deterministic model with 20 alleles at the least diverse locus, showing the effect of cross-immunity and the diversity of the second locus on the discordance of the pathogen population. Note that all three values of cross-immunity are very high.

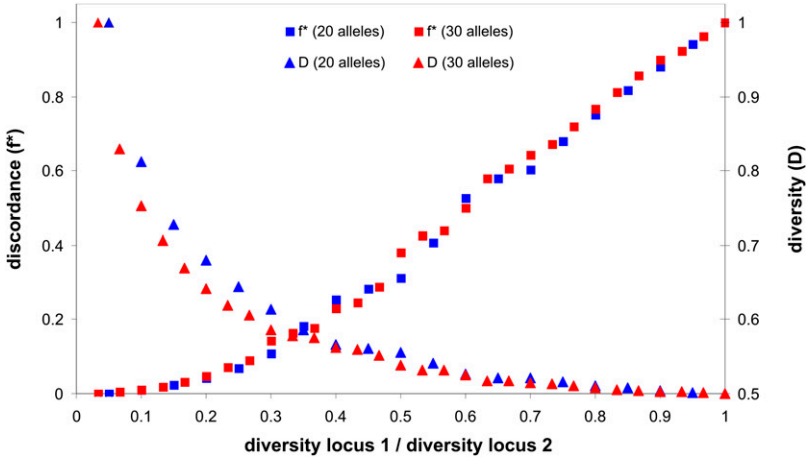

Fig. S8. Discordance values are not dependent on the number of alleles at the least diverse locus. Here, cross-immunity is 0.95.

Table S1. Parameter ranges explored in the stochastic model

Parameter	Description	Range (minimum/maximum)
C	Mean number of contacts per host	(8:20)
$1/\mu$	Average duration of infection, d	(3:50)
$1/\zeta$	Average duration of immunity, d	(10:100)
β	Probability of transmission (to a completely susceptible host)	(0.2:0.8)
R	Probability of recombination per allele	(0.01:0.1)
τ	Probability of allelic mutation per allele	(0.001:0.005)
γ	Degree of cross-immunity	(0:1)
P	Host population size	(500:1,500)