Supplementary Information

A random walk in Hamming space (Fig. S1) is described by the following equation:

$$\frac{\partial P_n}{\partial t} = P_{n-1}\frac{N-n+1}{N} + P_{n+1}\frac{n+1}{N} - P_n$$

where P_n is the probability of being at the Hamming distance n at time t (at $t=0$ $P_n = \delta_{n,0}$). The solution approaches the binomial distribution as $t \to \infty$ (see Figure S2):

$$P_n(t) = 2^{-N}\frac{N!}{(N-n)!n!}\left(1-e^{-2t/N}\right)^n\left(1+e^{-2t/N}\right)^{N-n}$$

It should be noted that a self-avoiding random walk (which occurs when a host develops a complete immunity against the previously encountered strains) closely approximates the random walk considered above for parameters that we use in the paper (the genome size is $N=30$, and the total number of infections is about 100).
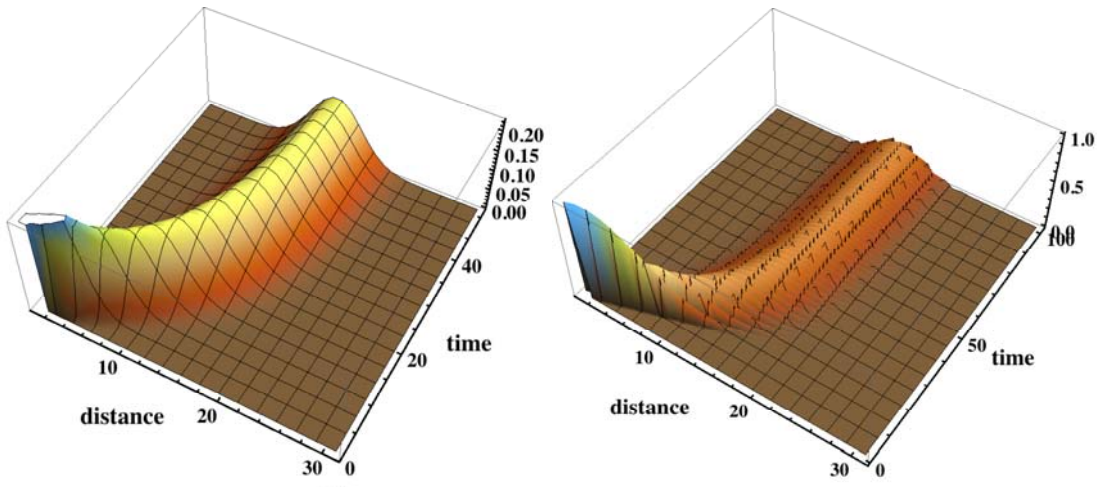
Figure S1. Random walk in Hamming space. Left: continuous time steps, right: discrete time steps.
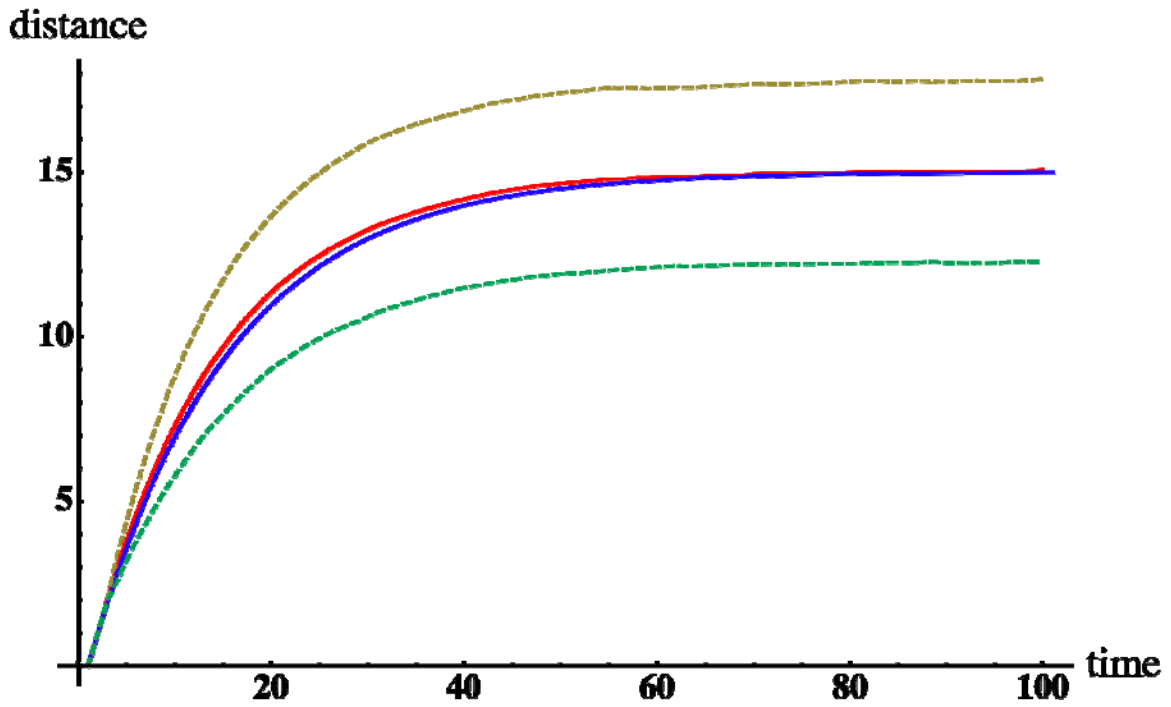
Figure S2. Self-avoiding random walk in Hamming space. x-axis denotes time steps and y-axis denotes the mean distance from origin (averaged over 1000 trials). Blue and red curves correspond to the random and self-avoiding random walks, respectively. Dashed lines correspond to the standard deviation.