Ashdin Publishing
Journal of Evolutionary Medicine
Vol. 1 (2012), Article ID 235553, 13 pages
doi:10.4303/jem/235553

Research Article

THEORY

Modeling Disease Evolution with Multilevel Selection: HIV as a Quasispecies Social Genome

Peter Nonacs[1] and Karen M. Kapheim[1,2]

[1]*Department of Ecology and Evolutionary Biology, University of California, 621 Charles E Young Dr South, Los Angeles, CA 90095, USA*
[2]*Department of Entomology and Institute for Genomic Biology, University of Illinois at Urbana-Champaign, 1206 W. Gregory Drive, Urbana, IL 61801, USA*
Address correspondence to Peter Nonacs, pnonacs@biology.ucla.edu

Received 27 February 2012; Accepted 21 September 2012

Abstract How genetically simple pathogens like HIV overwhelm complex immune systems is not fully understood. One unexplored possibility is that epistatic interactions across genetically complementary quasispecies (i.e., group-level social heterosis) allow pathogens to escape immune suppression. We tested this hypothesis by simulating an "HIV-like" pathogen under simultaneous individual- and group-level selection. Results matched in vivo HIV infection patterns, with virulence correlating to intrahost viral diversity patterns, variable times to immune system collapse extrinsic to variation in host immunity, unchanging virulence across infections, repeated evolution of complementary quasispecies clades, and numerous reversion mutations after virus transmission. Thus when group selection drives pathogen evolution, studying individual clones or consensus sequences will miss the possible effects of genetically synergistic "social genomes" on virulence. Furthermore, results suggest that virulence is reducible by manipulating selection to favor individual-level competitive ability in viruses. Overall, social genome models of pathogen interaction could produce novel approaches in studying diseases.

Keywords fitness epistasis; group selection; HIV; phenotypic enhancement; quasispecies; social heterosis; virulence

1 Introduction

In the coevolutionary arms race between pathogens and their human hosts there is one major asymmetry across the opponents. An initial infection can lead to many subsequent generations of pathogens within the host, in which new mutations and recombination radically change the population's genetic composition. Such rapid and continuous appearance of new alleles in pathogen populations cannot be matched on a genetic level by the hosts. Host-pathogen dynamics will, therefore, generally tend to favor pathogens over the course of chronic infections.

In diseases such as malaria, polio, hepatitis C, and human and simian immunodeficiency virus (HIV and SIV), infecting pathogens do often undergo a considerable genetic change over time [30,38,47]. Particularly with RNA viruses, populations can evolve to be composed of many similar, but non-identical, clones (quasispecies [47]). Furthermore, the original infecting genetic clone may be rare or entirely absent by the time health is compromised [37]. Thus, within-host evolution may determine disease progression and host mortality more so than the exact properties of the initial infecting strain.

To understand all the consequences of pathogen within-host evolution, it is necessary to consider the variety of ways pathogens can interact not only with host immune systems, but also with each other. Although pathogens are typically considered as competing entities where individual-level selection dominates [38], this need not always be the case. For example, in vivo quasispecies dynamics suggest that genotypes are non-independent and interact both competitively and cooperatively to affect each other's replication [4].

One specific way by which quasispecies dynamics may shift to favor pathogens is through fitness-enhancing genetic epistasis arising within increasingly diverse pathogen populations. Of particular importance would be intergenomic epistasis where the environmental component determining phenotype includes genetic influences from other individuals ($G \times E$ interactions, where E is the genotypes of other individuals [11,49,60,61,80,102,103]). Social heterosis, as a form of intergenomic epistasis [60,61], occurs when interactions among individuals with different genotypes produce novel effects (e.g., collective action [12]), or specialize on different aspects of host exploitation [16]. For example, one allele could code for a particularly effective protein product. A different allele at the same locus could code for a slightly different protein, but with its own unique, beneficial properties. No single pathogen could simultaneously produce both proteins, but groups potentially could. Such synergistic interactions could allow haploid viruses and bacteria to effectively act (and evolve)

as polyploid entities, as might be the case in examples of increased viral productivity when complementary genotypes of cytomegalovirus [18], nucleopolyhedrovirus [19] or polyomavirus [59] co-infect the same host cells.

Social heterosis is a potentially powerful evolutionary process when two preconditions are met [60,61]. First, the organism's environment must be complex or heterogenous so that no single genotype can optimally exploit every aspect of it. Second, populations must be subdivided into interacting groups that are genetically heterogenous. This permits a group-level selection to be a counter force against within-group directional selection [97,98]. Given these two preconditions, pathogen phenotypes could evolve to act as "social genomes": multiple interacting genomes whose fitness is enhanced in the presence of other specific genotypes.

HIV may serve as a particularly useful test case for social heterosis affecting pathogen evolution. The model's first precondition is met as HIV exists in the human body as part of a very complex ecosystem. HIV infects various human tissues (e.g., brain, spleen, lymph nodes, etc.), and each "habitat" has its own unique sets of environmental challenges [87]. Furthermore, HIV faces a large variety of "prey" and "predators" (e.g., T-cells, macrophages, dendritic cells, and antibodies) that are different in both external and internal composition, and change over location and time [87]. As required by the second precondition for group-level selection, the genetic composition of HIV populations varies at all levels from tissues to cells [39,42, 53,76,81,86,91]. Overall, HIV replication interacts with up to 273 host proteins [10], as might be expected in a population of genetically diverse viruses. Finally, observed HIV evolution within hosts strongly suggests that epistatic interactions across clones do occur [7,21,68].

To investigate how a social genome might arise in a quasispecies population, we simulated within-host evolution of an "HIV-like" pathogen with epistasis both within and across genomes. Although very general in explanatory potential, the utility of such a multilevel approach is evaluated relative to how well simulation results match observed patterns of HIV transmission and progression to AIDS. Like many viruses, HIV replicates quickly after initial infection. This stimulates an immune response which relatively rapidly suppresses HIV, but usually fails to completely eliminate it. A long, generally asymptomatic phase follows in which the HIV population is evolving and genetically changing [5,15], and this is the specific phase of the disease progression modeled.

2 Methods

Individual stochastic simulations tracked the fitness of an infecting pathogen over 5000 discrete generations. Estimates of HIV generation times vary from 0.76 to 2.64 days [33,70,72], with an average of 1.47 days [83]. Therefore by the average value, our model is analogous to tracking 9.3 years of HIV infection.

2.1 Defining the virus

The hypothetical pathogen has nine loci (as does HIV). No specific gene functions are assigned across Loci 1 through 9. Associated with each locus are wild-type alleles (wt), which can mutate to one of nine other alleles (and back mutations are equally probable). Thus, 10^9 unique viral genomes are possible. A genome composed of all wt alleles is arbitrarily defined as always having the highest fitness. Haplotypes with mutations to alleles 1–9 are equally fit relative to each other at each locus, but all nine have lower fitness than wt when comparing individual haplotypes or homogeneous groups (see below). For each simulation, however, two of alleles 1–9 at a given locus may be randomly designated as producing a positive epistatic interaction when co-infecting the same cell (i.e., social heterosis).

2.2 Cell infection and replication

Simulated target cells could be infected by 1–7 randomly chosen viruses (overall mean $= 4$ viruses per cell). Co-infecting viruses can have the same or different haplotypes. These values reflect the observed co-infection with HIV and the similar SIV [17,22,24,42,82]. The viruses replicate until cell death (this is equivalent to one viral "generation" in the model) and their total productivity is the viral burst size (VBS), such that

$$\text{VBS} = 10 + ny - wf.$$

VBS is a group property based on the various alleles that are present within the cell. The net increase in group productivity through epistatic interactions creating social heterosis is given as y (> 0, with $y = 3$ in all cases unless otherwise indicated), multiplied by the number of loci (n) where both complementary alleles are present in the group. The net loss in group productivity due to non-wt alleles is the product of the number of loci (w) at which no member of the group has a wt allele and its cost, f ($= 1$ for each loci in all simulations). However, if social heterosis is occurring at a locus, $f = 0$ for that locus even if no wt is present in the cell at that locus. Both gains and losses are modeled as additive across loci. The maximum number of loci that can potentially interact epistatically varies from 6 to 9 across simulations.

The model therefore has three categories of alleles at a given locus: (1) the wt; (2) two complementary alleles that can produce social heterosis if together in the same group, but are otherwise deleterious; (3) the remaining alleles that are always deleterious. Thus, for cells infected by one wt clone, VBS $= 10$ new viruses. For wt viruses, co-infections of cells are not advantageous, as cells infected by only one or seven pure wt will both have VBS $= 10$.

Co-infections, however, can be advantageous for mutant haplotypes through phenotypic rescue, as observed in HIV [36, 79]. Neither social heterosis nor phenotypic rescue is dosage dependent in the model. One wt allele rescues all other mutant viruses in the group. In the simulations, VBS can range from 1 (a singly-infected cell by a virus with no wt alleles at any loci) to 37 (2 viral haplotypes present with social heterosis through complementary alleles at all 9 loci). In this way, social heterosis extends beyond phenotypic rescue to actual enhancement across all viral clones, including the wt.

If only one virus infects a cell, all the descendants are copies of that virus. When cells are co-infected, all haplotypes have the possibility of being transmitted (as in HIV [36]) but they are not always copied with equal probability. This represents individual-level competition within cells. The probability (p) of any given descendant being a copy of infecting virus, x, is determined by the proportional cost imposed on x by having w non-wt alleles, relative to the summed costs for all co-infecting viruses:

$$p_x = \big(1 - cw_x\big) / \sum_{i=1}^{a} \big(1 - cw_i\big),$$

where the number of co-infecting viruses is a, and c is the competitive cost for each non-wt allele. For most simulations (unless otherwise indicated), $c = 0.02$ for each locus with a non-wt allele. So a virus with non-wt alleles at all 9 loci has only 82% of the probability of a co-infecting pure wt in being copied to produce descendants. Therefore, independent of whether epistatic interactions are present within any heterogenous group, wt alleles will have a reproductive advantage, on average, over non-wt alleles and the pure wt haplotype will always be the fittest haplotype at the individual-level of selection.

The expansion of the social genome could be a stepwise process, where complementary alleles must appear in a particular order. Three such sets of conditions for social heterosis were simulated: (1) the base condition where social genomes had to evolve in order at Loci 1–3 and thereafter the social genome could be elaborated in any order at Loci 4–6 (with no social heterosis possible at Loci 7–9); (2) all nine loci can be part of a social genome in any order, apart from Locus 1 having to be the first to develop social heterosis; (3) all nine loci can be part of a social genome, but must arise in sequential order from Locus 1 to 9.

Recombination was simulated in a subset of runs when a cell was infected by two or more viruses. A break point was randomly chosen somewhere between Loci 1 and 9, and the genetic material below the break point was switched between two viruses. All recombination assumed an equal exchange of complete genes and there was no hotspot on the genome for recombining. Recombination is a common feature of HIV replication [93], and in the model each clone had a 0.045 probability of recombining with a second, randomly chosen clone.

It is important to note that the model's variables (y, f, and c) are simplified representations of potentially many cumulative processes that eventually determine VBS. Hence, testing model predictions will focus on qualitative patterns of VBS over time rather than specific quantitative effects of the variables.

2.3 Tracking disease progression

The above within-cell dynamics describes one model generation, and in each generation it tracks a viral production from a constant "sample" of 120 cells. At the end of each round of reproduction, all the descendant viruses enter a common pool (= the summed viral output of all 120 cells). The infecting, parental viruses were not added to the pool. Thus, a particular haplotype can infect a cell but may not be represented in the next round due to stochasticity in reproductive success. From this viral pool, individuals were drawn (without replacement) to fill the next generation of 120 cells. With a mean group size of four, simulations tracked the reproduction of approximately 480 viruses in each generation. Before being assigned to a group, each chosen virus had an approximately 0.045 probability of experiencing a non-synonymous mutation (based on a probability of a mutation rate of 0.005 per gene loci per generation). This created, on average, 19.2 mutant viruses per generation in the tracked population. Mutations were distributed randomly across the nine loci, with equal transition probabilities across the 10 possible alleles at each locus. Thus mutation produced, on average, 0.43 alleles with the potential for intergenomic epistasis per generation at a given locus (19.2 mutations $\times 2$ possible complementary alleles/10 possible alleles/9 loci). This gives $a > 95\%$ probability that at least one such allele will appear across every 5 generations.

The model tracks pathogen genetic diversity in three different ways: (1) the number of unique haplotypes present in each sample, scored by the Shannon diversity H-index [52]; (2) the H-index diversity values across alleles at each of the 9 individual gene loci; (3) the mean similarity as the number of alleles held in common at loci that could potentially produce social heterosis, averaged across the four most common haplotypes in a given sample. This mimics data that would most likely result from sampling only subsets of entire viral populations.

It is assumed that mean VBS per infected cell affects the circulating viral load in a host and the number of future infected cells. With HIV, an increase in viral load in the asymptomatic phase strongly correlates with the collapse of the immune system and the onset of AIDS [5]. This does not imply, however, that replication within cells is the only point

of viral life history where positive epistasis could be advantageous. Social heterosis could also play a role in gaining entry into cells, exploiting a variety of cells, or diminishing the efficiency of the immune response. Thus, the results in terms of VBS level are better considered as proxy measures for how epistatic interactions evolve effective escapes from immune systems.

2.4 Pathogen transmission

Each simulation began with only one viral haplotype present in Generation 1 to reflect the strong genetic bottleneck observed in HIV transmission [27,50]. This infecting haplotype was randomly selected from the previous simulation's viral population when VBS in that population reached a defined threshold (i.e., when one or four loci exhibited social heterosis to reflect transmission at earlier or later stages of infection). If the criterion was not reached, the haplotype was randomly chosen from the population at Generation 5000. Each set of conditions was replicated 21 times sequentially to generate a longitudinal series of 20 transmission events (the first simulation initiated with all wt viruses, and was discarded from analyses).

2.5 Immune systems

The immune system of a given host is treated as a constant over the course of an infection in order to isolate and focus on the ramifications of pathogen genetic evolution. In the pathogen-host arms race, human immune systems are genetically fixed with a finite limit to plasticity of response. They cannot reciprocally evolve to produce genetically novel defenses in response to genetically changing disease populations. Thus, the model assumes that some synergistic combinations of viral genotypes (i.e., particular social genomes) overcome all possible responses of a host's immune system, resulting in deleterious viral load levels. Dynamic immune systems have been repeatedly modeled [23,25,63,64,65,66,74,84,99,100,101] and could be simulated by changing the relative fitness of haplotypes, where as a particular haplotype becomes common, the immune system adjusts to reduce its fitness. By assigning the wt allele the highest level of fitness under all circumstances, however, our simulation parameters are analogous to this concept. From a mathematical standpoint, the same defined fittest haplotype repeatedly sweeping through the population is equally probable and identical to multiple sweeps with new allele combinations. Moreover, whether the fittest haplotype changes over time or not does not affect the premise of the model, which is to understand the relative power of individual- versus group-level selection. Although a given immune system is held constant, immune variation across hosts is included by randomly specifying different sets of alleles as producing social heterosis.

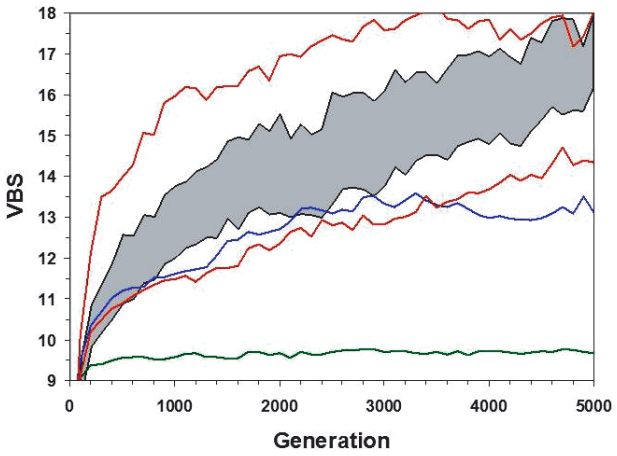

Figure 1: Increase in mean viral burst size (VBS) over time in the simulations. The gray area is the 95% confidence interval for the condition where 6 loci can be involved with social heterosis, with the first 3 having to accumulate mutations in a specific order (with $y = 3$, $c = 0.02$). The blue line is the mean when recombination also occurs. The green line is the mean with no social heterosis ($y = 0$). The red lines show the mean VBS when all 9 loci can be involved in social heterosis. In the lower line, all 9 loci have to accumulate in order. In the upper line, Locus 1 has to first develop social heterosis, but the other 8 loci can develop social heterosis in any order ($N = 20$ simulations for each line).

3 Results

Simulations yield 14 specific result predictions about disease progressions. Eleven of these could be compared to some degree with data on HIV progression and the remaining ones generate novel testable predictions.

3.1 Model results: high virulence can take many generations to evolve

When across-haplotype epistatic synergism is absent ($y = 0$), no between-group selection is possible. Mean VBS across infected cells remains low and constant across viral generations (Figure 1) and wt alleles dominate the population (Figure 2). In contrast when social heterosis is possible, between-group level selection can exceed within-group selection and evolution of higher VBS becomes probable, although individual-level selection can dominate in the population for extended periods of time (i.e., VBS remains low). On average, in the base condition it takes 2715 generations until 4 loci are involved in social heterosis. Model results do depend on key parameters. When $y = 2$ or $c = 0.03$, social genomes producing high VBS almost never develop. Conversely, if mutations are more likely to result in epistatically-interacting alleles, or y is increased, or c decreased, high virulence arises over relatively fewer

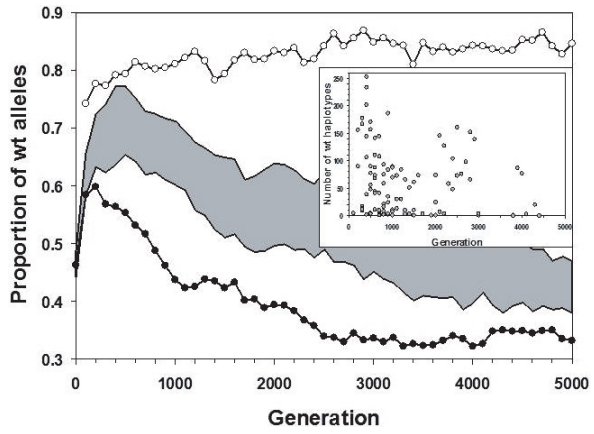

Figure 2: Proportion of all alleles that are wild-type (wt). The gray area is the 95% confidence interval for the condition where 6 loci can be involved with social heterosis, with the first 3 having to accumulate mutations in a specific order ($y = 3$, $c = 0.02$). For the line with filled circles, Locus 1 has to first develop social heterosis, but the other 8 loci can develop social heterosis in any order. The line with open circles has no social heterosis ($y = 0$). ($N = 20$ simulations for each line.) Inset: the number of pure wt haplotypes recorded across the sampled generations. Pure wt haplotypes are not transmitted and have to re-evolve in each simulation run.

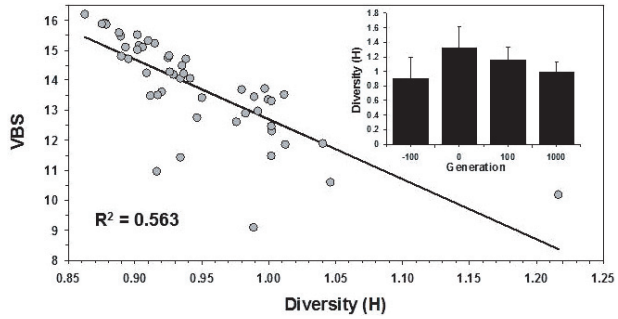

Figure 3: Relationship between allelic diversity and viral burst size (VBS) at Locus 1 in the simulations. Inset: diversity at Locus 1 with SD when social heterosis first develops (set as Generation 0), 100 generations before Generation 0, and 100 and 1000 generations after 0. Allelic diversity is measured with the Shannon H-index. ($N = 20$, $y = 3$, and $c = 0.02$.)

generations (results not shown). Also, mean group size has been shown to affect the selective strength of social heterosis (see [60,61]). Thus, neither individual nor group-level selection can be expected to always dominate pathogen evolution.

Empirical evidence. Long asymptomatic periods of time are common prior to AIDS onset [5,31].

3.2 Model results: onset of immune system collapse is highly variable

The appearance of a high VBS occurs over a wide range of simulated generations, from 500 to over 5000 (with 9 simulations less than the mean, and 11 greater than the mean). Mean VBS is also highly variable as individual-level selection would occasionally destabilize the social genome when wt replaced complementary alleles, causing a decline in VBS. Therefore, the mean VBS varied across a range of values rather than equilibrating to a single value (Figure 1).

Empirical evidence. Variability in the progression of HIV is common in human populations [38,87], as evidenced by fast and slow progressors, who become symptomatic with relatively low or high HIV diversity, respectively [63].

3.3 Model results: variance in AIDS onset can be greater than predicted by variance in only immune system characteristics

Because the model treats the immune system as a constant, the above stochasticity reflects only variability in the trajectories of viral evolution across simulations, which are determined by random chance evolutionary events.

Empirical evidence. Associations between disease progression and host immune variations are often found, but with contradictory and inconsistent results [14,73,89], and host genotype account for only about 10% of the variation in AIDS onset [14]. Stochasticity in vivo is also shown by a clinical case where monozygotic twins progressed to AIDS at very different rates, although both twins were infected by the same donor in a short time period [88].

3.4 Model results: correlations between virulence and viral diversity strongly depend on where and when diversity is measured

When no social heterosis is possible, over 80% of alleles are wt, resulting in a high mean similarity across the most common haplotypes (Figure 2). With social heterosis, the average number of unique haplotypes varies from 53 to 110 in the approximately 480 viruses sampled per generation, but does not correlate with increased VBS. Nor is there a significant relationship with allelic diversity across the entire genome and VBS. Allelic diversities at all individual loci, however, do correlate with VBS. For Locus 1 the correlation is significantly negative, but for all other loci it is significantly positive (Figure 3, Table 1). Locus 1 is under the strongest stabilizing selection to maintain complementary alleles for social heterosis, because if epistasis is lost at this locus, it simultaneously disappears at all other loci. The initial appearance of social heterosis at Locus 1 is clearly signified by when VBS exceeds 10, defining this point in time as generation "0". Comparing this point to 100 generations before, and 100 or 1000 generations afterwards reveals

Table 1: Regression summary for relationships of mean viral burst size to mean allelic diversity across loci. Data for loci 4–6 and 7–9 are averaged together within a simulation run. ($N = 50$ measures: one taken every 100 generations and averaged over 20 simulation runs.)

Loci	Intercept	Slope	R^2	P value
1	32.631	−19.933	0.563	< 0.0001
2	1.213	12.639	0.173	0.0016
3	1.868	11.579	0.574	< 0.0001
4–6	1.152	13.037	0.892	< 0.0001
7–9	0.273	15.093	0.684	< 0.0001

that genetic diversity at Locus 1 initially rises after epistatic synergism increases VBS and quickly reaches a maximum (Figure 3, inset). Thereafter diversity declines over time as non-complementary alleles are lost at that locus, leading to an overall negative relationship.

Empirical evidence. Although HIV diversity is viewed as one of the most consistent correlates for imminent immune system collapse, there is also considerable variance in the apparent relationship across patients [5,15,63]. How much of this variance is due to differences in methodology or foci of measurement is not discernible across the data. A specific prediction on diversity that is supported is that in vivo, HIV-1 diversity most often stabilizes or slightly decreases with the onset of AIDS [5,39,71,76,81,85].

3.5 Model results: consistent evolution of distinct and complimentary quasispecies clades

Haplotype diversity tends to evolve into two quasispecies families, designated as clades A or B. (Note that the exact allelic combinations that constitute A and B genomes vary across simulations, as different alleles become linked due to chance.) These clades appear as haplotypes having alternative complementary alleles for epistasis across loci. Nevertheless, there are no arbitrary constraints on haplotypes and individual genomes could be intermediate by having some number of alleles from both families, such as three alleles from each clade. The resulting distribution of haplotypes, however, is clearly not random with respect to possible allele combinations (Figure 4). Genomes are clustered near the prototypical A and B haplotypes—67% of haplotypes generated in the simulations had 4–6 alleles associated with either an A or B family. Very few intermediate haplotypes remain in the population. Also when group-level selection dominates, the mean similarity of the four most common haplotypes steadily declines as haplotypes segregate into two quasispecies families (Figure 5). Overall, the degree to which the four most common haplotypes are dissimilar significantly correlates with VBS (Figure 5, inset).

Empirical evidence. To our knowledge, there are no direct tests for complimentary alleles across quasispecies

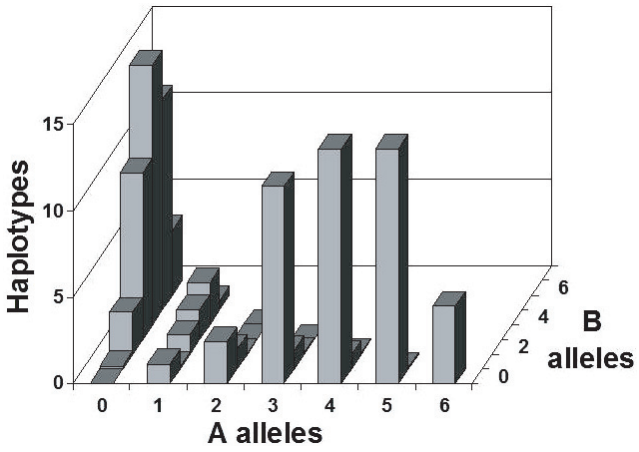

Figure 4: Percentages of viral haplotypes aligning by quasispecies family. Quasispecies families A and B are defined by having alternative alleles involved in social heterosis at the six loci that can produce intergenomic epistasis (with $y = 3$, $c = 0.02$). $N = 20$ simulations, from which a total of 9561 haplotypes were sampled.

families, but distinct quasispecies do consistently appear with functional and non-functional Nef populations across chronically infected individuals [2]. The maintenance of the non-functional Nef is consistent with facilitation through intergenomic epistasis as part of a social genome. Also suggestive is that distinct phylogenetic clusters of HIV clones consistently arose in patients with failing immune systems [41].

3.6 Model results: group-level stabilizing selection is still present as immune system is collapsing

Patterns of genetic diversity that occur as the immune system is collapsing can indicate the selection mechanism. Relaxed selection predicts that random substitutions will accumulate from mutation and genetic diversity should rapidly increase [71]. Alternatively, immune system collapse due to directional selection should favor exploitative genotypes and reduced genetic diversity with non-synonymous mutations for loci increasing replicative fitness. Finally, immune system collapse due to between-group selection for a social genome predicts that stabilizing selection maintains the existing pattern of diversity, as in the results where VBS is rising (Figure 1) while genetic diversity locks into two complementary quasispecies clades (Figure 4).

Empirical evidence. At AIDS onset, genetic diversity tends to either not increase or slightly decline [5,39,71,76,81,85]. Also, in the late stages of disease progression mutations at the Env sequence are both non-random (suggesting selection is still acting) and synonymous (suggesting selection is stabilizing) [95].

3.7 Model results: genetic diversity enhances viral replicative fitness more in late-infection isolates than in early-infection isolates

Isolating clones separates them from the benefits they may receive from a social genome (e.g., if clades A and B in Figure 4 were isolated from each other). As social genomes should be more likely in later stages of infection (Figure 1), separation from the population's genetic diversity should affect those isolates more. A corollary prediction would be that fitness enhancement with genetic diversity should be greatest in the presence of pathogens isolated from the same patient, because variation in host immune systems helps determine the specific allele combinations for social heterosis and social genomes (i.e., the A and B clades in Figure 4 were different combinations of alleles across simulations). Thus, overall pathogen evolution can be determined by $G \times E \times E$ interactions, where the E's are the genotypes of the other clones and the host.

Empirical evidence. Not directly tested to our knowledge.

3.8 Model results: loci vary in rate of accumulating mutations over time, creating consistent patterns in how genetic diversity arises

The rate of VBS increase depends on whether epistatic synergisms must evolve in a particular order. When epistasis at Loci 2–9 can evolve in any order, VBS rises very rapidly (Figure 1). In contrast, when epistasis must evolve in a specific sequence across the nine loci, VBS rises at a significantly slower rate.

Empirical evidence. The Env locus in HIV accumulates the most genetic variants over the course of an infection, but it is highly conserved with few variants in the early stages of infection [26,104]. In contrast, the Gag locus tends to be much more conserved overall, yet early in infection there are more mutations found at this locus than at Env [26,62, 104]. Thus, diversity appears to have to be first created at Gag before diversity becomes advantageous at Env, which may reflect that CTL immune responses (towards Gag) exert an initially stronger selection pressure than antibody immune responses (towards Env). Further evidence for the importance of sequential epistatic interactions comes from associations of variation at specific Gag residues found within single strains in vitro. Mutations are thought to arise in other regions of the Gag C-terminus to compensate for the reduction in replicative fitness associated with escape phase mutations in the CTL-binding Gag epitope [55]. If fitness consequences were independent across loci, then mutations should accumulate randomly and in no particular order. Also, HIV-1 infections with apparently absent or defective Nef loci do not cause AIDS [2,45], although this should not reduce the mutation rate at other loci (i.e., high overall levels of genetic diversity in the viral population

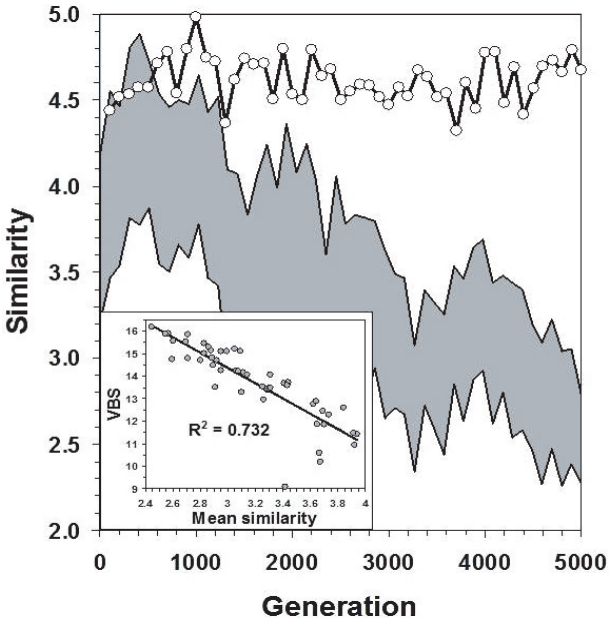

Figure 5: Mean similarity of the most common haplotypes over time. The gray area is the 95% confidence interval zone for similarity. Similarity is defined as the average number of alleles shared by the 4 most common haplotypes in the population, across 6 loci producing social heterosis (with $y = 3$, $c = 0.02$). The line with open circles is mean similarity when no social heterosis is possible, $y = 0$. ($N = 20$ simulations for each set of conditions.) Inset: correlation of mean viral burst size (VBS) to mean similarity as defined in the main figure.

should still be possible). Thus, specific trajectories in the creation of genetic diversity, rather than just diversity per se, appear related to the immune system collapse.

3.9 Model results: recombination decreases the evolution of virulence

Adding recombination into the model slows the rate at which VBS increases (Figure 1). Complementary epistatic genomes are disrupted by recombination that creates intermediate haplotypes that do not gain through epistasis. Maintaining evolutionarily interacting loci in the same genome or the same interacting group requires a strong positive linkage disequilibrium, which also strongly selects against recombination [28,44,61].

Empirical evidence. The majority of epistasis in HIV-1 is positive [7] and recombinant viruses have lower fitness because of disruption of coadapted genetic sequences [35].

3.10 Model results: virulence does not increase across transmissions

Genetically complex social genomes are not transmitted in toto. New infections with only one haplotype resulted in no

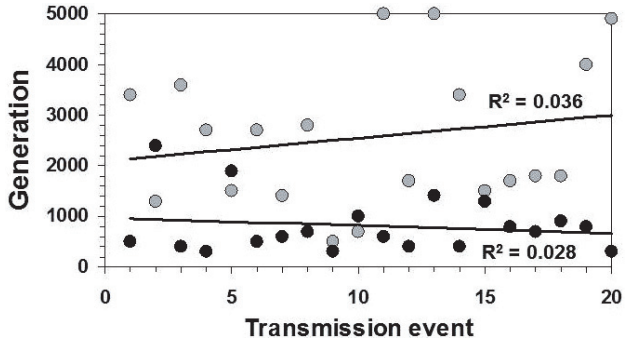

Figure 6: The number of generations required for either 1 (black circle) or 4 (gray circles) loci to first exhibit social heterosis, across a longitudinal series of 20 transmissions. Neither regression (black lines) has a slope significantly different from zero ($P > 0.4$).

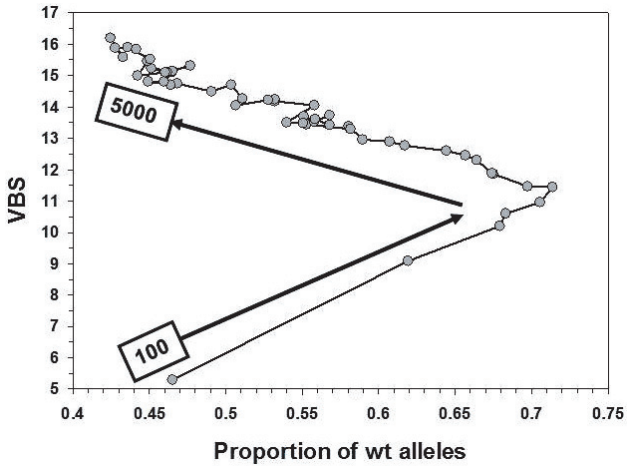

Figure 7: Mean viral burst size (VBS) versus the mean proportion of all alleles that are wt. The line connects points from Generation 100 to 5000. ($N = 20$, $y = 3$, and $c = 0.02$.)

significant change in virulence across a series of 20 transmissions (Figure 6). The results do not differ if transmissions occur early or late in an infection. Neither slope is significantly different from zero, $P > .4$. Evolving social genomes multiple times does not produce haplotypes that are either more or less capable of evolving a new social genome after transmission.

Empirical evidence. The virulence of HIV-1 has been tracked since the beginning of the pandemic, and there is little conclusive evidence for any significant increase or decrease [5,34,53,58]. Additionally, mutants that arise during the pathogenic escape phase do not typically cause a more rapid onset of immune collapse when transmitted [32, 45]. There is further suggestive evidence that certain clones of HIV are more likely to be transmitted, but that these clones are not dominant within the donor at the time of transmission [6,9]. Thus, the strains most effective at overcoming the immune system are not necessarily equally most transmissible.

3.11 Model results: virulence evolves more rapidly with multiple co-infection

Unlike the simulations for this model, if multiple variants with a full range of genetic diversity were transmitted, this would potentially allow group-level benefits of social heterosis to evolve more rapidly.

Empirical evidence. Transmission causes severe genetic bottlenecks for HIV-1, as less than 1% of the standing genetic variability appears to be transmitted in each new infection [9,27,50]. This is independent of whether transmission is vertical or horizontal. However, approximately 25% of HIV infections transmit multiple variants [9,43, 77], and in most cases these come from a single donor [1, 6,9,43,77]. Significant positive correlations are found between rapid disease progression and multiple variant transmission [1,75].

3.12 Model results: wild-type mutations are common in early stages of infection

Consistent with individual-level selection dominating early in an infection, wt alleles increase rapidly after transmission (Figure 2), and the pure wt haplotype often reappears and reaches high population frequencies (Figure 2, inset).

Empirical evidence. After transmission, HIV-1 populations initially evolve back towards a general, consensus genotype following transmission to a new host [3,37,48]. Similar phenomena also occur with HIV-2 and SIV [32]. If AIDS onset is simply a function of viral diversity exceeding levels that can be suppressed by the immune system [23,63,64,65,66], there would be no expectation for specific evolutionary trajectories to be commonly repeated (i.e., quasispecies genetic diversity should arise idiosyncratically across hosts). Evolutionary reversion towards a predictable genotype is consistent with individual-level selection consistently favoring a particularly effective combination of alleles.

3.13 Model results: early in infection virulence negatively correlates with genetic diversity, but the relationship reverses over the length of the infection

Early in infections individual-level selection should dominate and alleles that increase haplotype competitive ability should be favored. When group-level selection begins to dominate later in the infection, the correlation should reverse and virus production should positively correlate with the degree of divergence from the consensus genome (Figure 7).

Empirical evidence. Not directly tested to our knowledge.

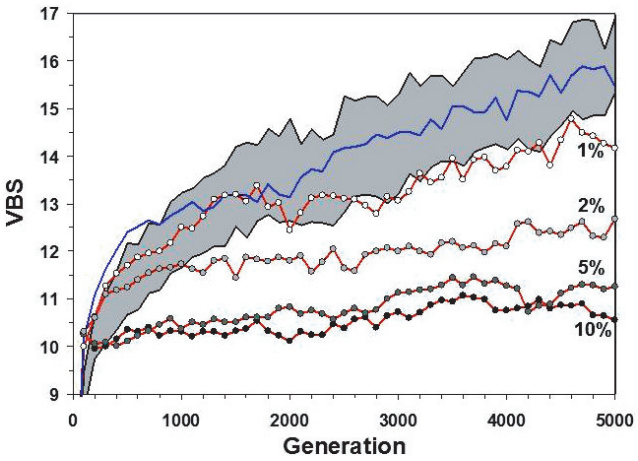

Figure 8: Effect on mean viral burst size (VBS) over time with the addition of other haplotypes. The gray zone is as in Figure 1 (i.e., no addition of haplotypes). The blue line is the mean VBS when random haplotypes replace 5% of the viral population. Red lines are the mean VBS when pure wt clones replace 1–10% of the population. ($N = 20$, $y = 3$, and $c = 0.02$ for all lines.)

3.14 Model results: re-emergence of or re-infection by the wild-type can reduce virulence

High VBS requires the presence of a social genome, which is possible only when group-level selection is stronger than individual-level selection. This suggests that VBS can be reduced by destabilizing social genomes through increasing individual-level selection. The effect of inoculating an infected host with either pure wt viruses or viruses with a randomly-generated genotype (on average, such genomes contained one wt allele) was simulated in the model. The inoculations replaced 1–10% of the evolving virus population every 100 generations ($\approx$ 4 to 48 viruses replaced every 100th generation) as long as VBS > 10. Such inoculations can also be viewed as analogous to the occasional in vivo re-emergence of latent virus from a dormant reservoir. Replacing 5% of the virus population with random genomes is a control for the direct effect of replacing a proportion of the existing population. It has no effect on the rate of increase in VBS (Figure 8). Inoculating with at least 2% of pure wt virus, however, significantly slowed or entirely prevented a large increase in VBS.

Empirical evidence. Not directly tested to our knowledge.

4 Discussion

Evolutionary approaches to medicine continue to illuminate the complex interactions between hosts and their pathogens. There is a growing important realization that disease outcomes are not due solely to host-pathogen interactions, but are also determined by between-pathogen interactions in evolving populations [4,40,92]. Especially in chronic infections, host immune systems can face ever-changing populations of quasispecies families. To understand the progression of such diseases it is required to consider the specific evolutionary mechanisms through which quasispecies can interact. Here we show that intergenomic epistasis as social heterosis [60] has the potential to strongly affect the evolution of virulence when pathogen populations are subdivided in heterogenous groups within a complex host ecosystem. Our model results generate 14 predictions about how an "HIV-like" chronic infection may evolve. Of those predictions, 11 are at least partially consistent with general characteristics of HIV-1 infections. Most notable are matches to heretofore inadequately explained phenomena, such as the high variance in time to collapse of immune systems; variance in AIDS onset that is not correlated to differences in immune systems; shifting correlations between HIV genetic diversity and immune system viability; why mutations may accumulate in specific orders across HIV loci; the lack of significant change in HIV virulence over the course of the epidemic; why reversion mutations in HIV are common after transmission events.

The model also makes a number of novel predictions for future testing. (1) Time to AIDS onset will not be easily predicted by either initial host or viral genotypes due to the inherently large stochastic variation in when group level selection will predominate across pathogen populations. (2) HIV diversity and viral load will be negatively correlated early in infections due to individual-level selection, but positively correlated as immune system functioning declines as group-level selection dominates. (3) Replicative fitness of specific HIV clones should be higher when measured in the presence of other clones found at the same time in the host, than when measured as single-clone isolates. This difference in replicative fitness should increase over the length of the infection because of the increased likelihood that social genomes are affecting HIV clone fitness. (4) Randomly assembled populations of HIV clones may have higher genetic diversity than co-evolved quasispecies clades but their replicative fitness will be lower because they will not represent social genomes. (5) Increasing recombination rates in HIV will slow further virulence increase at later stages in infections because it will create non-complementary clones. (6) Introduction of clones that approximate consensus genotypes, or re-emergence of the initial infecting clones will reduce viral loads in hosts because it will break up the social genome and individual-level selection will have greater influence on the evolutionary dynamics of the pathogen population.

Our model does not explicitly consider dynamic responses by host immune systems in order to more clearly focus on the potential effects of within-host pathogen interaction and evolution. The theoretical ramifications

of a dynamic immune response have been extensively explored, although most often with an implicit assumption that pathogens experience only individual-level selective pressures [23,25,63,64,65,66,74,84,99,100,101], but see [4,7,69,92]. Linking the evolutionary biology of pathogens to immune response models could generate future novel insights for disease treatments. Nevertheless, if group selection on pathogen interactions significantly affects disease progression, then there are already clear methodological consequences for research. For example, isolating individual or incomplete subsets of haplotypes from late-stage patients [5,34,90,94] and testing their replicative fitness, infectiveness, and resistance to drug properties in a simple and ideal growth medium could be very misleading [2]. Particularly suggestive is that the target of selection may be the quasispecies, in that in vitro fitness recovery in debilitated HIV populations occurs more rapidly with higher genetic diversity [8,51].

If epistatic interactions have strong fitness effects, the entire host and social environment must be tested, because genotype expression will depend strongly on the social environment [74,103]. Whole-genome association studies could identify mutation patterns across co-occurring quasispecies and determine when some sets of mutations appear together more often than chance would predict (as possibly in [2,3,39,67,85]). Single-genome amplification (SGA) might also identify social genomes by accurate sequencing of multiple viral genomes within a single sample [13,77, 78]. With repeated sampling over an extended length of infection, SGA could identify complementary alleles and track their frequency within the larger pathogen population.

Finally, specific epistatic interactions that evolve during an infection may differ relative to a patient's genetic makeup and the particular characteristics of their immune response [5,46,57,96]. Therefore, averaging disease progression across many genetically-variable individuals or considering only a few HLA loci could obscure critical interactions [29,46,89]. Strong covariance between viral and host genotypes would have to be discerned by matching allelic variation across hosts and their viral populations.

5 Conclusion

A continuing paradox in the study of infectious diseases is how very simple organisms like HIV (composed of 9000 bases and 9 genes that produce 15 proteins) can eventually overwhelm an entity as complex and redundant as the human immune system [20]. Social heterosis raises the real possibility that even the simplest genomes can function like socially cooperative organisms. Thus, rather than infections being a population of competing entities, they may be "clouds" of synergistic, interactive genomes [30,92] where group-level selection can strongly affect virulence [56]. If HIV can act as a social genome, its ability to effectively manipulate hundreds of different human proteins [10] becomes less puzzling. Furthermore, HIV is far from unique in being able to evolve within its host. Other types of viruses [18,19,59] and diseases such as hepatitis C, polio, malaria (*Plasmodium falciparum*) and hoof-and-mouth disease [37,54,92] also go through many generations within a single host. Certainly any disease with the potential to evolve within a host could become a cooperative social genome with debilitating consequences.

Acknowledgments The authors thank S. Gilboa, J. Pepper and anonymous reviewers for many very helpful comments on earlier versions of this paper. P. Nonacs was partially supported by NSF grant IOS-0642085.

References

[1] M. R. Abrahams, J. A. Anderson, E. E. Giorgi, C. Seoighe, K. Mlisana, L. H. Ping, et al., *Quantitating the multiplicity of infection with human immunodeficiency virus type 1 subtype C reveals a non-poisson distribution of transmitted variants*, J Virol, 83 (2009), 3556–3567.

[2] A. Ali, S. Realegeno, O. O. Yang, and M. J. Lewis, *Simultaneous assessment of CD4 and MHC-I downregulation by Nef primary isolates in the context of infection*, J Virol Methods, 161 (2009), 297–304.

[3] T. M. Allen, M. Altfeld, S. C. Geer, E. T. Kalife, C. Moore, K. M. O'sullivan, et al., *Selective escape from CD8+ T-cell responses represents a major driving force of human immunodeficiency virus type 1 (HIV-1) sequence diversity and reveals constraints on HIV-1 evolution*, J Virol, 79 (2005), 13239–13249.

[4] J. Arbiza, S. Mirazo, and H. Fort, *Viral quasispecies profiles as the result of the interplay of competition and cooperation*, BMC Evol Biol, 10 (2010), 137.

[5] K. K. Ariën, G. Vanham, and E. J. Arts, *Is HIV-1 evolving to a less virulent form in humans?*, Nat Rev Microbiol, 5 (2007), 141–151.

[6] D. I. Boeras, P. T. Hraber, M. Hurlston, T. Evans-Strickfaden, T. Bhattacharya, E. E. Giorgi, et al., *Role of donor genital tract HIV-1 diversity in the transmission bottleneck*, Proc Natl Acad Sci U S A, 108 (2011), E1156–E1163.

[7] S. Bonhoeffer, C. Chappey, N. T. Parkin, J. M. Whitcomb, and C. J. Petropoulos, *Evidence for positive epistasis in HIV-1*, Science, 306 (2004), 1547–1550.

[8] A. V. Bordería, R. Lorenzo-Redondo, M. Pernas, C. Casado, T. Alvaro, E. Domingo, et al., *Initial fitness recovery of HIV-1 is associated with quasispecies heterogeneity and can occur without modifications in the consensus sequence*, PLoS One, 5 (2010), e10319.

[9] C. L. Boutwell, M. M. Rolland, J. T. Herbeck, J. I. Mullins, and T. M. Allen, *Viral evolution and escape during acute HIV-1 infection*, J Infect Dis, 202 (2010), S309–S314.

[10] A. L. Brass, D. M. Dykxhoorn, Y. Benita, N. Yan, A. Engelman, R. J. Xavier, et al., *Identification of host proteins required for HIV infection through a functional genomic screen*, Science, 319 (2008), 921–926.

[11] E. D. Brodie III, *Why evolutionary genetics does not always add up*, in Epistasis and the Evolutionary Process, J. B. Wolf, E. D. Brodie III, and M. J. Wade, eds., Oxford University Press, Oxford, 2000, 3–19.

[12] S. P. Brown, M. E. Hochberg, and B. T. Grenfell, *Does multiple infection select for raised virulence?*, Trends Microbiol, 10 (2002), 401–405.

[13] Z. L. Brumme and B. D. Walker, *Tracking the culprit: HIV-1 evolution and immune selection revealed by single-genome amplification*, J Exp Med, 206 (2009), 1215–1218.

[14] M. Carrington, G. W. Nelson, M. P. Martin, T. Kissner, D. Vlahov, J. J. Goedert, et al., *HLA and HIV-1: heterozygote advantage and B*35-Cw*04 disadvantage*, Science, 283 (1999), 1748–1752.

[15] E. Castro-Nallar, M. Pérez-Losada, G. F. Burton, and K. A. Crandall, *The evolution of HIV: inferences using phylogenetics*, Mol Phylogenet Evol, 62 (2012), 777–792.

[16] L. Chao, K. A. Hanley, C. L. Burch, C. Dahlberg, and P. E. Turner, *Kin selection and parasite evolution: higher and lower virulence with hard and soft selection*, Q Rev Biol, 75 (2000), 261–275.

[17] J. Chen, Q. Dang, D. Unutmaz, V. K. Pathak, F. Maldarelli, P. D., et al., *Mechanisms of nonrandom human immunodeficiency virus type 1 infection and double infection: preference in virus entry is important but is not the sole factor*, J Virol, 79 (2005), 4140–4149.

[18] L. Cicin-Sain, J. Podlech, M. Messerle, M. J. Reddehase, and U. H. Koszinowski, *Frequent coinfection of cells explains functional in vivo complementation between cytomegalovirus variants in the multiply infected host*, J Virol, 79 (2005), 9492–9502.

[19] G. Clavijo, T. Williams, D. Muñoz, P. Caballero, and M. López-Ferber, *Mixed genotype transmission bodies and virions contribute to the maintenance of diversity in an insect virus*, Proc Biol Sci, 277 (2010), 943–951.

[20] J. Cohen, *Virology: HIV gets by with a lot of help from human host*, Science, 319 (2008), 143–144.

[21] J. da Silva, M. Coetzer, R. Nedellec, C. Pastore, and D. E. Mosier, *Fitness epistasis and constraints on adaptation in a human immunodeficiency virus type 1 protein region*, Genetics, 185 (2010), 293–303.

[22] Q. Dang, J. Chen, D. Unutmaz, J. M. Coffin, V. K. Pathak, D. Powell, et al., *Nonrandom HIV-1 infection and double infection via direct and cell-mediated pathways*, Proc Natl Acad Sci U S A, 101 (2004), 632–637.

[23] R. J. de Boer and M. C. Boerlijst, *Diversity and virulence thresholds in AIDS*, Proc Natl Acad Sci U S A, 91 (1994), 544–548.

[24] A. Del Portillo, J. Tripodi, V. Najfeld, D. Wodarz, D. N. Levy, and B. K. Chen, *Multiploid inheritance of HIV-1 during cell-to-cell infection*, J Virol, 85 (2011), 7169–7176.

[25] N. M. Dixit and A. S. Perelson, *HIV dynamics with multiple infections of target cells*, Proc Natl Acad Sci U S A, 102 (2005), 8198–8203.

[26] C. T. Edwards, E. C. Holmes, O. G. Pybus, D. J. Wilson, R. P. Viscidi, E. J. Abrams, et al., *Evolution of the human immunodeficiency virus envelope gene is dominated by purifying selection*, Genetics, 174 (2006), 1441–1453.

[27] C. T. Edwards, E. C. Holmes, D. J. Wilson, R. P. Viscidi, E. J. Abrams, R. E. Phillips, et al., *Population genetic estimation of the loss of genetic diversity during horizontal transmission of HIV-1*, BMC Evol Biol, 6 (2006), 28.

[28] I. Eshel and M. W. Feldman, *On the evolutionary effect of recombination*, Theor Popul Biol, 1 (1970), 88–100.

[29] J. Fellay, K. V. Shianna, D. Ge, S. Colombo, B. Ledergerber, M. Weale, et al., *A whole-genome association study of major determinants for host control of HIV-1*, Science, 317 (2007), 944–947.

[30] S. L. Fishman and A. D. Branch, *The quasispecies nature and biological implications of the hepatitis C virus*, Infect Genet Evol, 9 (2009), 1158–1167.

[31] C. Fraser, T. D. Hollingsworth, R. Chapman, F. de Wolf, and W. P. Hanage, *Variation in HIV-1 set-point viral load: epidemiological analysis and an evolutionary hypothesis*, Proc Natl Acad Sci U S A, 104 (2007), 17441–17446.

[32] T. C. Friedrich, E. J. Dodds, L. J. Yant, L. Vojnov, R. Rudersdorf, C. Cullen, et al., *Reversion of CTL escape-variant immunodeficiency viruses in vivo*, Nat Med, 10 (2004), 275–281.

[33] Y. X. Fu, *Estimating mutation rate and generation time from longitudinal samples of DNA sequences*, Mol Biol Evol, 18 (2001), 620–626.

[34] Y. Gali, B. Berkhout, G. Vanham, M. Bakker, N. K. Back, and K. K. Ariën, *Survey of the temporal changes in HIV-1 replicative fitness in the Amsterdam Cohort*, Virology, 364 (2007), 140–146.

[35] A. Galli, M. Kearney, O. A. Nikolaitchik, S. Yu, M. P. Chin, F. Maldarelli, et al., *Patterns of human immunodeficiency virus type 1 recombination ex vivo provide evidence for coadaptation of distant sites, resulting in purifying selection for intersubtype recombinants during replication*, J Virol, 84 (2010), 7651–7661.

[36] H. C. Gelderblom, D. N. Vatakis, S. A. Burke, S. D. Lawrie, G. C. Bristol, and D. N. Levy, *Viral complementation allows HIV-1 replication without integration*, Retrovirology, 5 (2008), 60.

[37] B. T. Grenfell, O. G. Pybus, J. R. Gog, J. L. Wood, J. M. Daly, J. A. Mumford, et al., *Unifying the epidemiological and evolutionary dynamics of pathogens*, Science, 303 (2004), 327–332.

[38] A. L. Hill, D. I. Rosenbloom, and M. A. Nowak, *Evolutionary dynamics of HIV at multiple spatial and temporal scales*, J Mol Med (Berl), 90 (2012), 543–561.

[39] M. D. Hill and W. Hernndez, *Nucleotide and amino acid mutations in human immunodeficiency virus corresponding to CD4+ decline*, Arch Virol, 151 (2006), 1149–1158.

[40] E. C. Holmes, *The evolutionary genetics of emerging viruses*, Annu Rev Ecol Evol Syst, 40 (2009), 353–372.

[41] S. Hong, J. Cao, and Y. T. Tu, *Evolution of HIV-1 in a patient population failing multiple-drug therapy*, Microbiol Immunol, 53 (2009), 535–539.

[42] A. Jung, R. Maier, J. P. Vartanian, G. Bocharov, V. Jung, U. Fischer, et al., *Recombination: multiply infected spleen cells in HIV patients*, Nature, 418 (2002), 144.

[43] B. F. Keele, E. E. Giorgi, J. F. Salazar-Gonzalez, J. M. Decker, K. T. Pham, M. G. Salazar, et al., *Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection*, Proc Natl Acad Sci U S A, 105 (2008), 7552–7557.

[44] J. K. Kelly, *Epistasis, linkage, and balancing selection*, in Epistasis and the Evolutionary Process, J. B. Wolf, E. D. Brodie III, and M. J. Wade, eds., Oxford University Press, Oxford, 2000, 146–157.

[45] J. T. Kimata, *HIV-1 fitness and disease progression: insights from the SIV-macaque model*, Curr HIV Res, 4 (2006), 65–77.

[46] J. Lalić and S. F. Elena, *Epistasis between mutations is host-dependent for an RNA virus*, Biol Lett, (2012).

[47] A. S. Lauring and R. Andino, *Quasispecies theory and the behavior of RNA viruses*, PLoS Pathog, 6 (2010), e1001005.

[48] B. Li, A. D. Gladden, M. Altfeld, J. M. Kaldor, D. A. Cooper, A. D. Kelleher, et al., *Rapid reversion of sequence polymorphisms dominates early human immunodeficiency virus type 1 evolution*, J Virol, 81 (2007), 193–201.

[49] T. A. Linksvayer, *Ant species differences determined by epistasis between brood and worker genomes*, PLoS One, 2 (2007), e994.

[50] M. Lipsitch and J. J. O'Hagan, *Patterns of antigenic diversity and the mechanisms that maintain them*, J R Soc Interface, 4 (2007), 787–802.

[51] R. Lorenzo-Redondo, A. V. Bordería, and C. Lopez-Galindez, *Dynamics of in vitro fitness recovery of HIV-1*, J Virol, 85 (2011), 1861–1870.

[52] A. E. Magurran, *Ecological Diversity and Its Measurement*, Princeton University Press, Princeton, NJ, 1988.

[53] R. B. Markham, W. C. Wang, A. E. Weisstein, Z. Wang, A. Munoz, A. Templeton, et al., *Patterns of HIV-1 evolution in individuals with differing rates of CD4 T cell decline.*, Proc Natl Acad Sci U S A, 95 (1998), 12568–12573.

[54] V. Martín and E. Domingo, *Influence of the mutant spectrum in viral evolution: focused selection of antigenic variants in a reconstructed viral quasispecies*, Mol Biol Evol, 25 (2008), 1544–1554.

[55] J. Martinez-Picado, J. G. Prado, E. E. Fry, K. Pfafferott, A. Leslie, S. Chetty, et al., *Fitness cost of escape mutations in p24 Gag in association with control of human immunodeficiency virus type 1*, J Virol, 80 (2006), 3617–3623.

[56] R. Miralles, A. Moya, and S. F. Elena, *Is group selection a factor modulating the virulence of RNA viruses?*, Genet Res, 69 (1997), 165–172.

[57] C. B. Moore, M. John, I. R. James, F. T. Christiansen, C. S. Witt, and S. A. Mallal, *Evidence of HIV-1 adaptation to HLA-restricted immune responses at a population level*, Science, 296 (2002), 1439–1443.

[58] V. Müller, B. Ledergerber, L. Perrin, T. Klimkait, H. Furrer, A. Telenti, et al., *Stable virulence levels in the HIV epidemic of Switzerland over two decades*, AIDS, 20 (2006), 889–894.

[59] M. R. Myhre, G. H. Olsen, R. Gosert, H. H. Hirsch, and C. H. Rinaldo, *Clinical polyomavirus BK variants with agnogene deletion are non-functional but rescued by trans-complementation*, Virology, 398 (2010), 12–20.

[60] P. Nonacs and K. M. Kapheim, *Social heterosis and the maintenance of genetic diversity*, J Evol Biol, 20 (2007), 2253–2265.

[61] P. Nonacs and K. M. Kapheim, *Social heterosis and the maintenance of genetic diversity at the genome level*, J Evol Biol, 21 (2008), 631–635.

[62] V. Novitsky, R. Wang, L. Margolin, J. Baca, L. Kebaabetswe, R. Rossenkhan, et al., *Timing constraints of in vivo gag mutations during primary HIV-1 subtype C infection*, PLoS One, 4 (2009), e7727.

[63] M. Nowak and R. M. May, *AIDS pathogenesis: mathematical models of HIV and SIV infections*, AIDS, 7 (1993), S3–S18.

[64] M. A. Nowak, R. M. Anderson, A. R. McLean, T. F. Wolfs, J. Goudsmit, and R. M. May, *Antigenic diversity thresholds and the development of AIDS*, Science, 254 (1991), 963–969.

[65] M. A. Nowak and R. M. May, *Coexistence and competition in HIV infections*, J Theor Biol, 159 (1992), 329–342.

[66] M. A. Nowak, R. M. May, and R. M. Anderson, *The evolutionary dynamics of HIV-1 quasispecies and the development of immunodeficiency disease*, AIDS, 4 (1990), 1095–1103.

[67] D. H. O'Connor, A. B. McDermott, K. C. Krebs, E. J. Dodds, J. E. Miller, E. J. Gonzalez, et al., *A dominant role for $CD8^+$-T-lymphocyte selection in simian immunodeficiency virus sequence variation*, J Virol, 78 (2004), 14012–14022.

[68] M. Parera, N. Perez-Alvarez, B. Clotet, and M. A. Martínez, *Epistasis among deleterious mutations in the HIV-1 protease*, J Mol Biol, 392 (2009), 243–250.

[69] J. W. Pepper, *Defeating pathogen drug resistance: guidance from evolutionary theory*, Evolution, 62 (2008), 3185–3191.

[70] A. S. Perelson, A. U. Neumann, M. Markowitz, J. M. Leonard, and D. D. Ho, *HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time*, Science, 271 (1996), 1582–1586.

[71] T. M. Riddle, N. J. Shire, M. S. Sherman, K. F. Franco, H. W. Sheppard, and J. A. Nelson, *Sequential turnover of human immunodeficiency virus type 1 env throughout the course of infection*, J Virol, 80 (2006), 10591–10599.

[72] A. G. Rodrigo, E. G. Shpaer, E. L. Delwart, A. K. Iversen, M. V. Gallo, J. Brojatsch, et al., *Coalescent estimates of HIV-1 generation time in vivo*, Proc Natl Acad Sci U S A, 96 (1999), 2187–2191.

[73] S. Rowland-Jones, S. Pinheiro, and R. Kaul, *New insights into host factors in HIV-1 pathogenesis*, Cell, 104 (2001), 473–476.

[74] S. M. Roy and D. Wodarz, *Infection of HIV-specific CD4 T helper cells and the clonal composition of the response*, J Theor Biol, 304 (2012), 143–151.

[75] M. Sagar, L. Lavreys, J. M. Baeten, B. A. Richardson, K. Mandaliya, B. H. Chohan, et al., *Infection with multiple human immunodeficiency virus type 1 variants is associated with faster disease progression*, J Virol, 77 (2003), 12921–12926.

[76] M. Sala, G. Zambruno, J. P. Vartanian, A. Marconi, U. Bertazzoni, and S. Wain-Hobson, *Spatial discontinuities in human immunodeficiency virus type 1 quasispecies derived from epidermal Langerhans cells of a patient with AIDS and evidence for double infection*, J Virol, 68 (1994), 5280–5283.

[77] J. F. Salazar-Gonzalez, E. Bailes, K. T. Pham, M. G. Salazar, M. B. Guffey, B. F. Keele, et al., *Deciphering human immunodeficiency virus type 1 transmission and early envelope diversification by single-genome amplification and sequencing*, J Virol, 82 (2008), 3952–3970.

[78] J. F. Salazar-Gonzalez, M. G. Salazar, B. F. Keele, G. H. Learn, E. E. Giorgi, H. Li, et al., *Genetic identity, biological phenotype, and evolutionary pathways of transmitted/founder viruses in acute and early HIV-1 infection*, J Exp Med, 206 (2009), 1273–1789.

[79] K. Salzwedel and E. A. Berger, *Complementation of diverse HIV-1 Env defects through cooperative subunit interactions: a general property of the functional trimer*, Retrovirology, 6 (2009), 75.

[80] M. Santos and S. E., *Genetic hitchhiking can promote the initial spread of strong altruism*, BMC Evol Biol, 8 (2008), 281.

[81] H. Schuitemaker, M. Koot, N. A. Kootstra, M. W. Dercksen, R. E. de Goede, R. P. van Steenwijk, et al., *Biological phenotype of human immunodeficiency virus type 1 clones at different stages of infection: progression of disease is associated with a shift from monocytotropic to T-cell-tropic virus population*, J Virol, 66 (1992), 1354–1360.

[82] A. Schultz, S. Sopper, U. Sauermann, A. Meyerhans, and R. Suspène, *Stable multi-infection of splenocytes during SIV infection—the basis for continuous recombination*, Retrovirology, 9 (2012), 31.

[83] T. K. Seo, J. L. Thorne, M. Hasegawa, and H. Kishino, *Estimation of effective population size of HIV-1 within a host: a pseudomaximum-likelihood approach*, Genetics, 160 (2002), 1283–1293.

[84] L. Sguanci, F. Bagnoli, and P. Liò, *Modeling HIV quasispecies evolutionary dynamics*, BMC Evol Biol, 7 (2007), S5.

[85] R. Shankarappa, J. B. Margolick, S. J. Gange, A. G. Rodrigo, D. Upchurch, H. Farzadegan, et al., *Consistent viral evolutionary changes associated with the progression of human immunodeficiency virus type 1 infection*, J Virol, 73 (1999), 10489–10502.

[86] D. Shriner, Y. Liu, D. C. Nickle, and J. I. Mullins, *Evolution of intrahost HIV-1 genetic diversity during chronic infection*, Evolution, 60 (2006), 1165–1176.

[87] M. Stevenson, *HIV-1 pathogenesis*, Nat Med, 9 (2003), 853–860.

[88] L. Tazi, H. Imamichi, S. Hirschfeld, J. A. Metcalf, S. Orsega, M. Pérez-Losada, et al., *HIV-1 infected monozygotic twins: a tale of two outcomes*, BMC Evol Biol, 11 (2011), 62.

[89] A. Telenti and D. B. Goldstein, *Genomics meets HIV-1*, Nat Rev Microbiol, 4 (2006), 865–873.

[90] R. M. Troyer, K. R. Collins, A. Abraha, E. Fraundorf, D. M. Moore, R. W. Krizan, et al., *Changes in human immunodeficiency virus type 1 fitness and genetic diversity during disease progression*, J Virol, 79 (2005), 9006–9018.

[91] G. van Marle and C. Power, *Human immunodeficiency virus type 1 genetic diversity in the nervous system: Evolutionary epiphenomenon or disease determinant?*, J Neurovirol, 11 (2005), 107–128.

[92] M. Vignuzzi, J. K. Stone, J. J. Arnold, C. E. Cameron, and R. Andino, *Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population*, Nature, 439 (2006), 344–348.

[93] N. N. Vijay, A. R. Vasantika, A. S. Perelson, and N. M. Dixit, *Recombination increases human immunodeficiency virus fitness, but not necessarily diversity*, J Gen Virol, 89 (2008), 1467–1477.

[94] Y. Voronin, J. Overbaugh, and M. Emerman, *Simian immunodeficiency virus variants that differ in pathogenicity differ in fitness under rapid cell turnover conditions*, J Virol, 79 (2005), 15091–15098.

[95] S. Williamson, S. M. Perry, C. D. Bustamante, M. E. Orive, M. N. Stearns, and J. K. Kelly, *A statistical characterization of consistent patterns of human immunodeficiency virus evolution within infected patients*, Mol Biol Evol, 22 (2005), 456–468.

[96] C. C. Wilson, S. A. Kalams, B. M. Wilkes, D. J. Ruhl, F. Gao, B. H. Hahn, et al., *Overlapping epitopes in human immunodeficiency virus type 1 gp120 presented by HLA A, B, and C molecules: effects of viral variation on cytotoxic T-lymphocyte recognition*, J Virol, 71 (1997), 1256–1264.

[97] D. S. Wilson, *The Natural Selection of Populations and Communities*, Series in Evolutionary Biology, Benjamin/Cummings Pub., Menlo Park, CA, 1980.

[98] D. S. Wilson, *Weak altruism, strong group selection*, Oikos, 59 (1990), 135–140.

[99] D. Wodarz, *On the relative fitness of early and late stage Simian immunodeficiency virus isolates*, Theor Popul Biol, 72 (2007), 426–435.

[100] D. Wodarz and D. N. Levy, *Multiple HIV-1 infection of cells and the evolutionary dynamics of cytotoxic T lymphocyte escape mutants*, Evolution, 63 (2009), 2326–2339.

[101] D. Wodarz and M. A. Nowak, *Mathematical models of HIV pathogenesis and treatment*, Bioessays, 24 (2002), 1178–1187.

[102] J. B. Wolf, E. D. Brodie, and A. J. Moore, *Interacting phenotypes and the evolutionary process. II. Selection resulting from social interactions*, Am Nat, 153 (1999), 254–266.

[103] J. B. Wolf, E. D. Brodie III, J. M. Cheverud, A. J. Moore, and M. J. Wade, *Evolutionary consequences of indirect genetic effects*, Trends Ecol Evol, 13 (1998), 64–69.

[104] L. Q. Zhang, P. MacKenzie, A. Cleland, E. C. Holmes, A. J. Brown, and P. Simmonds, *Selection for specific sequences in the external envelope protein of human immunodeficiency virus type 1 upon primary infection*, J Virol, 67 (1993), 3345–3356.