

Fitness Epistasis and Constraints on Adaptation in a Human Immunodeficiency Virus Type 1 Protein Region

Jack da Silva,^{*,1} Mia Coetzer,[†] Rebecca Nedellec,[†] Cristina Pastore[†] and Donald E. Mosier[†]

^{*}*School of Molecular and Biomedical Science, The University of Adelaide, Adelaide, South Australia 5005, Australia and*

[†]*Department of Immunology and Microbial Science, The Scripps Research Institute, La Jolla, California 92037*

Manuscript received November 25, 2009

Accepted for publication February 8, 2010

ABSTRACT

Fitness epistasis, the interaction among alleles at different loci in their effects on fitness, has potentially important consequences for adaptive evolution. We investigated fitness epistasis among amino acids of a functionally important region of the human immunodeficiency virus type 1 (HIV-1) exterior envelope glycoprotein (gp120). Seven mutations putatively involved in the adaptation of the second conserved to third variable protein region (C2–V3) to the use of an alternative host-cell chemokine coreceptor (CXCR4) for cell entry were engineered singly and in combinations on the wild-type genetic background and their effects on viral infectivity were measured. Epistasis was found to be common and complex, involving not only pairwise interactions, but also higher-order interactions. Interactions could also be surprisingly strong, changing fitness by more than 9 orders of magnitude, which is explained by some single mutations being practically lethal. A consequence of the observed epistasis is that many of the minimum-length mutational trajectories between the wild type and the mutant with highest fitness on cells expressing the alternative coreceptor are selectively inaccessible. These results may help explain the difficulty of evolving viruses that use the alternative coreceptor in culture and the delayed evolution of this phenotype in natural infection. Knowledge of common, complex, and strong fitness interactions among amino acids is necessary for a full understanding of protein evolution.

FITNESS epistasis refers to the interaction among alleles at different loci in their effects on fitness. The importance of such interactions to adaptation has been controversial. WRIGHT (1932) argued that fitness epistasis would cause multi-peaked fitness landscapes and that these would constrain adaptive evolution by attracting populations to local peaks. FISHER (1930), on the other hand, argued against the likelihood of such rugged fitness landscapes. And, although there is some indirect evidence of multi-peaked fitness landscapes (*e.g.*, NIJHUIS *et al.* 1999; BURCH and CHAO 2000), it is difficult to demonstrate the existence of such landscapes conclusively. A direct demonstration would require analyzing all possible interactions in an entire genome because it is always possible that a mutation at an unstudied locus may generate a genotype that spans a fitness valley (WHITLOCK *et al.* 1995). However, even on a single-peaked fitness landscape, epistasis may produce minimum-length mutational trajectories that are unlikely to be realized during adaptive evolution

because they include neutral or deleterious mutations (WEINREICH *et al.* 2006). This will arise if the sign of the fitness effect of a mutation depends on its genetic background (sign epistasis) (WEINREICH *et al.* 2005). Fitness epistasis may also generate linkage disequilibrium (KIMURA 1956; LEWONTIN and KOJIMA 1960), with potentially important consequences for the efficiency of natural selection and the evolutionary maintenance of recombination (FELSENSTEIN 1988; KONDRASHOV 1993).

Notwithstanding theoretical developments, the nature of fitness epistasis remains poorly understood. Early quantitative genetics experiments on the viability effects of epistasis in *Drosophila* showed these to be only weak to moderate (SPASSKY *et al.* 1965; TEMIN *et al.* 1969). More recent observations on microbes adapting to antimicrobial drugs (reviewed by MAISNIER-PATIN and ANDERSSON 2004) and responding to other selection pressures (*e.g.*, POON and CHAO 2005; SANJUAN *et al.* 2005) suggest strong compensatory effects of epistasis. Much of the recent work on fitness epistasis at the molecular level has involved the analysis of intergenic or intragenic interactions in microbes through the study of standing genetic variation or spontaneous mutation (BONHOEFFER *et al.* 2004; MAISNIER-PATIN *et al.* 2005; BERSHTEIN *et al.* 2006), engineered site-specific mutations (SANJUAN *et al.* 2004; LUNZER *et al.*

Supporting information is available online at <http://www.genetics.org/cgi/content/full/genetics.109.112458/DC1>.

¹*Corresponding author:* School of Molecular and Biomedical Science, The University of Adelaide, Adelaide SA 5005, Australia.
E-mail: jack.dasilva@adelaide.edu.au

-----C2-----|-----V3-----|
 1 NTSTITQACPKVSFEPIPIHYCTPAGFAILK...CTRPNNNTRKSIHIGPGRAFYTTEIIGDIRQAHC
 D.....T.....H.....R.....R.....EK.....

FIGURE 1.—The HIV-1 ADA C2 (partial) and V3 gp120 protein regions. Engineered mutations are numbered and shown below the sequence. Underlined residues are putative N-linked glycosylation motifs.

2005; WEINREICH *et al.* 2006; PEPIN and WICHMAN 2007), and a combination of both types of data (POON and CHAO 2005; SANJUAN *et al.* 2005; POON and CHAO 2006). However, a systematic study of the nature and consequences of fitness epistasis within a protein has yet to be conducted. We have investigated the nature and consequences of interactions among amino acid mutations on fitness in a functionally important protein region of human immunodeficiency virus type 1 (HIV-1).

The use of site-directed mutagenesis to measure the fitness effects of mutations singly and in combination on a standard genetic background has the advantage over other approaches of allowing unambiguous attribution of fitness interactions to specific combinations of mutations. This approach was used in a recent study of protein evolution involved in the switch by HIV-1 from using its primary host-cell chemokine coreceptor to an alternative chemokine coreceptor (PASTORE *et al.* 2006). Considerable attention has been focused on this question because HIV-1 preferentially uses the primary coreceptor in early infection, but switches to the alternative coreceptor late in infection in about 50% of patients, and this switch is associated with disease progression (PHILPOTT 2003). The entry of an HIV-1 virus particle (virion) into a host cell requires that the exterior envelope glycoprotein, gp120, on the surface of the virion, interact with two cell-surface receptors: CD4 and one of two chemokine coreceptors, either CCR5 or CXCR4 (WYATT and SODROSKI 1998). Binding of gp120 to CD4 is thought to cause a conformational change to gp120 that exposes its third variable region (V3), allowing it to bind to one of the chemokine coreceptors (HUANG *et al.* 2005; HUANG *et al.* 2007). This conformational change to gp120 appears to involve its second variable region (V2), which may shield V3 (WYATT and SODROSKI 1998). V3 determines which coreceptor is used by the virus (HWANG *et al.* 1991) through its sequence variation (DITTMAR *et al.* 1997; SPECK *et al.* 1997; CORMIER and DRAGIC 2002), although PASTORE *et al.* (2006) have shown that the contiguous first and second variable regions (V1/V2) and the second conserved region (C2), which separates V1/V2 from V3, may also be involved.

PASTORE *et al.* (2004) used selection experiments to identify the amino acid changes putatively involved in coreceptor switching. They selected for CXCR4 use in CCR5-adapted HIV-1 strains by passaging (serially transferring) virus through cell cultures containing progressively increasing proportions of CXCR4-expressing cells. They also sequenced the V1/V2 or C2 region and

the V3 region for several isolates as they adapted to CXCR4. Amino acid replacements observed in these experiments were then engineered, using site-directed mutagenesis, on to CCR5-adapted wild-type genetic backgrounds and tested for their effects singly and in combination on the coreceptor usage of the virus (PASTORE *et al.* 2006). They found that mutations in V3 are necessary for coreceptor switching, but generally reduced viral infectivity, and that mutations in V1/V2 and C2 may compensate for this loss of fitness.

Here, we investigate the nature of this epistasis and its consequences for adaptation. We have assayed the fitness of a set of HIV-1 envelope glycoprotein clones constructed by engineering amino acid mutations in the C2–V3 region singly and in combination on a CCR5-adapted wild-type background (PASTORE *et al.* 2006). These mutations correspond to the amino acid changes of one of the CXCR4-adapted isolates evolved from a CCR5-adapted strain by PASTORE *et al.* (2004). Thus, we describe in detail the frequency, level, direction, form, magnitude, and consequences of fitness epistasis in the context of HIV-1 adapting to CXCR4. We found that fitness epistasis is common and often very strong and compensatory and that it may be complex, involving not only pairwise interactions between residues, but also higher-order interactions. We also show that sign epistasis constrains adaptation in the C2–V3 protein region through the production of selectively inaccessible minimum-length mutational trajectories. Such constraints may help explain the difficulty of evolving CXCR4 use in culture and the delayed evolution of this phenotype in natural infection. Fitness epistasis must be considered for a full understanding of protein evolution.

MATERIALS AND METHODS

Fitness assay: We assayed the fitness of a set of envelope glycoprotein clones that had been constructed by engineering amino acid mutations in the C2–V3 region singly and in combination on the CCR5-adapted ADA strain wild-type background (PASTORE *et al.* 2006). These mutations correspond to the seven amino acid changes of the CXCR4-adapted ADA-1 isolate evolved from ADA (PASTORE *et al.* 2004). Five of the these changes occurred in V3 and two in C2 (Figure 1). The available mutants included all single mutations, all possible combinations of the five V3 mutations, and a subset of combinations involving the two C2 mutations, including all seven mutations, for a total of 53 mutant envelopes (Figure 2). PASTORE *et al.* (2006) did not construct all possible $2^7 - 1 = 127$ mutants as they were attempting to engineer the minimum number of envelopes required to explain the evolution of ADA-1.

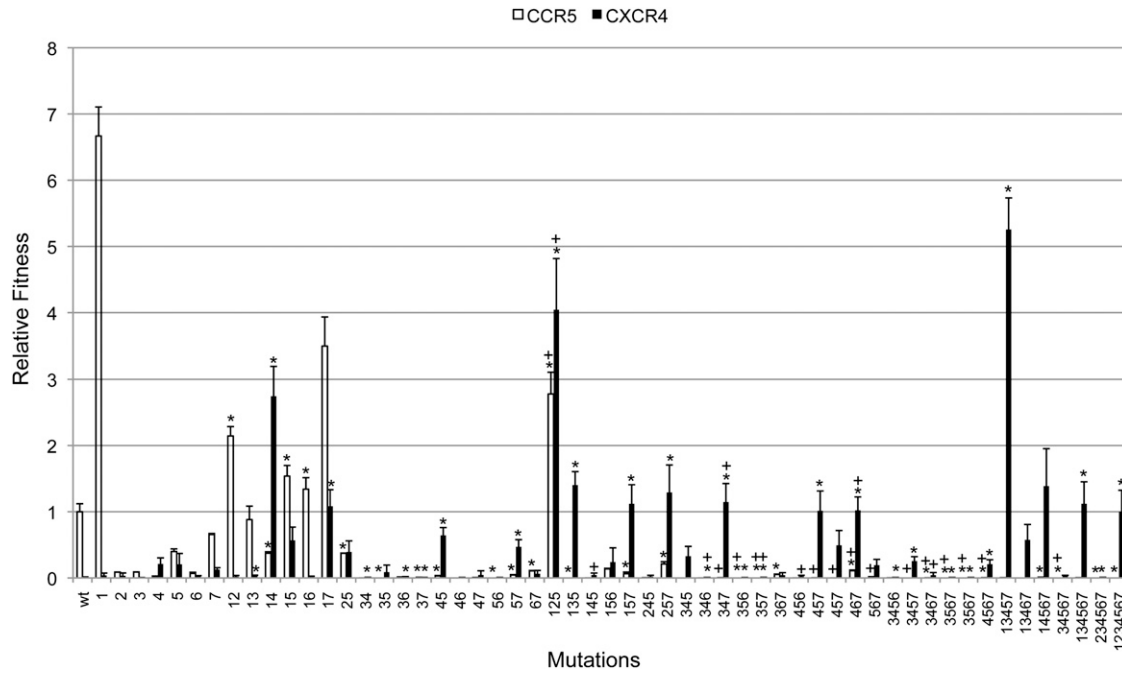


FIGURE 2.—The relative fitness of each engineered envelope assayed on cells expressing CCR5 or CXCR4. Error bars indicate 1 standard deviation. An asterisk (*) indicates statistically significant overall epistasis (ϵ), and a plus sign (+) indicates significant net epistasis for higher-order interactions (ϵ').

Fitness was measured with a single-cycle pseudovirus infection assay, as in PASTORE *et al.* (2006). Briefly, mutant envelope clones inserted into the pSVIII plasmid were cotransfected with *env*-negative, luciferase-positive (NL4-3-Luc+E-R-) reporter plasmids into 293T cells, and the resulting pseudoviruses were harvested, standardized for p24 content, and used to infect either CCR5-expressing cells or CXCR4-expressing cells. The luciferase activity from triplicate wells of a multiwell plate was measured on a luminometer after 48 hr of culture (supporting information, Figure S1). Assays were conducted in triplicate for each host-cell type: cells expressing CCR5 or CXCR4. We used NP-2-CD4-CCR5 and NP-2-CD4-CXCR4 cell lines because these cells do not express endogenous chemokine coreceptors (SODA *et al.* 1999) that mediate entry of some viruses via GPR1 and GPR15 (EDINGER *et al.* 1998), thus forcing entry through CCR5 or CXCR4. Unlike other measures of fitness, such as resistance to an antiviral drug, which may trade off against other components of fitness (MAMMANO *et al.* 2000), the rate of cell infection, or infectivity, is the appropriate measure of the effect on fitness of the C2-V3 protein region. This is because the sole function of this region is in cell entry, which is unlikely to trade off against other components of fitness under culture conditions, or other steps in the replication cycle, which are controlled by other proteins (COFFIN 1999). Furthermore, the interaction between V3 and the chemokine coreceptor affects the rate-limiting step in cellular infection (PLATT *et al.* 2005). However, this protein region is targeted by antibodies (ZOLLA-PAZNER 2004), and here we restrict our analysis to effects on the infectivity component of fitness.

Fitness epistasis: Relative fitness, w , was calculated by dividing absolute fitness by the absolute fitness of the CCR5-adapted wild-type isolate, ADA, for assays with cells expressing CCR5, and by dividing absolute fitness by the absolute fitness of the CXCR4-adapted ADA-1 mutant isolate for assays with cells expressing CXCR4.

The fitness effect of the overall interaction among a set of amino acid mutations was calculated as the epistatic deviation,

$$\epsilon_M = w_M - \prod_{i \in M} w_i, \quad (1)$$

where M is the set of amino acid mutations, w_M is the relative fitness of a mutant containing the mutations in set M , and w_i is the relative fitness of a mutant containing a single mutation from set M . Identifying the mutant containing all of the mutations in set M as M , Equation 1 measures the deviation from the observed fitness of mutant M to its expected fitness in the absence of epistasis. The expected fitness in the absence of interactions is the product of the fitnesses of the mutants carrying each mutation from set M singly.

To test the statistical significance of overall epistasis, the variance in epistatic deviation was estimated. The variance of the product of single-mutation envelope fitnesses was estimated using the formula for the variance of a product of independent random variables (GOODMAN 1962),

$$\text{Var}\left(\prod_{i \in M} w_i\right) = \prod_{i \in M} (\text{Var}(w_i) + w_i^2) - \prod_{i \in M} w_i^2, \quad (2)$$

where $\text{Var}(w_i)$ is the variance in relative fitness of a single-mutation envelope clone estimated from triplicate assays of fitness. Equation 2 assumes that the relative fitnesses, w_i , are independent. Then, the variance of epistatic deviation was estimated using the formula for the variance of a difference of random variables (SOKAL and ROHLF 1995),

$$\text{Var}(\epsilon_M) = \text{Var}(w_M) + \text{Var}\left(\prod_{i \in M} w_i\right), \quad (3)$$

where $\text{Var}(w_M)$ is the variance in relative fitness of mutant M estimated from triplicate assays of fitness. Equation 3 assumes no correlation between the fitness of mutant M and the

product of the single-mutation envelope fitnesses. This assumption inflates the estimate variance, making statistical tests more conservative.

The net epistatic deviation of a higher-order interaction, involving three or more mutations, was calculated by removing the net effects of all lower-order interactions involving the mutations,

$$\varepsilon'_M = \varepsilon_M - \sum_{X=2}^{n-1} \sum_{j=1}^C \varepsilon'_{M_{Xj}}, \quad (4)$$

where n is the number of mutations in set M and M_{Xj} is the subset of mutations containing the j th of C combinations of n mutations taken X at a time. The number of combinations of n mutations taken X at a time is $C = n!/[X!(n-X)!]$. For example, in the case of a three-way interaction, the net epistatic deviation was calculated by subtracting the epistatic deviations of all contributing pairwise interactions,

$$\varepsilon'_{klm} = \varepsilon_{klm} - (\varepsilon_{kl} + \varepsilon_{km} + \varepsilon_{lm}), \quad (5)$$

where k , l , and m are individual mutations (*i.e.*, k , l , and m are the elements of M). Note that for pairwise interactions $\varepsilon = \varepsilon'$ because there are no lower-order interactions. For interactions among four mutations, this would involve subtracting the net epistatic deviations of all contributing three-way interactions as well as the epistatic deviations of all contributing pairwise interactions. The variance of the net epistatic deviation for a higher-order interaction was estimated as follows:

$$\text{Var}(\varepsilon'_M) = \text{Var}(\varepsilon_M) + \sum_{X=2}^{n-1} \sum_{j=1}^C \text{Var}(\varepsilon'_{M_{Xj}}). \quad (6)$$

For example, for a three-way interaction, this variance would be

$$\text{Var}(\varepsilon'_{klm}) = \text{Var}(\varepsilon_{klm}) + \text{Var}(\varepsilon_{kl}) + \text{Var}(\varepsilon_{km}) + \text{Var}(\varepsilon_{lm}). \quad (7)$$

Equation 6 assumes that the contributing lower-order net epistatic deviations are independent.

The statistical significance of epistatic deviations ($H_0: \varepsilon = 0$) was determined with a Z-test. To account for multiple comparisons, the experimentwise type I error rate was maintained at $\alpha' = 0.05$ using the sequential Bonferroni method (SOKAL and ROHLF 1995).

The magnitude of epistasis was calculated as the log of the ratio of the fitness of mutant M and the product of the single-mutation fitnesses:

$$E_M = \log_{10} \left(w_M / \prod_{i \in M} w_i \right). \quad (8)$$

This is the preferred measure of the magnitude of epistasis because it provides the order-of-magnitude change in fitness due to epistasis. Note that when applied to higher-order interactions, E is a measure of the magnitude of overall epistasis, not net epistasis. Therefore, for higher-order interactions the sign of E may differ from that of ε' .

RESULTS

Fitness: The seven mutations in the C2–V3 region of the CXCR4-adapted isolate ADA-1 were engineered

singly and in combinations on the genetic background of the CCR5-adapted wild-type strain ADA (Figure 1). The fitness of these mutants was assayed on CCR5-expressing cells and CXCR4-expressing cells. On CCR5 cells, the fitness of most single and multiple mutants was lower than that of ADA (Figure 2 and [supporting information, Table S1](#)). The exceptions are the mutant with mutation 1 and several multiple mutants containing mutation 1. Therefore, ADA does not have the highest fitness with respect to CCR5 use, possibly because of competing selection pressures in natural infection, such as antibody surveillance (PASTORE *et al.* 2006). On CXCR4 cells, several multiple mutants had higher fitness than ADA-1 (Figure 2 and [Table S1](#)), which had been isolated after selection for CXCR4 use (PASTORE *et al.* 2004). These mutants typically contain mutation 1. Mutant 13457 (containing mutations 1, 3, 4, 5, and 7) had the highest fitness on CXCR4 cells. The fact that a mutant with all seven mutations (ADA-1) evolved in response to selection for CXCR4 use suggests that direct mutational trajectories from ADA to mutants with higher fitness on CXCR4 cells may be selectively inaccessible because of epistasis.

Epistatic deviation: Of the 48 interactions tested for overall epistasis on each host-cell type, 27 (56%) were statistically significant on CCR5 cells and the same number were significant on CXCR4 cells, although some of these interactions differed between the two assays (Figure 2 and [Table S1](#)). On CCR5 cells, significant overall epistatic deviations were both positive and negative, and the median epistatic deviation was 0.0002 (range -1.6857 – 2.5460) (Figure 3a and [Table S1](#)). In contrast, significant overall epistatic deviations on CXCR4 cells were exclusively positive, with a median of 0.4444 (0.0074 – 5.2552). Therefore, overall epistasis was common and ranged from -1.6857 to 5.2552 .

To give specific examples, we focus on two statistically significant pairwise interactions, one on CCR5 cells and the other on CXCR4 cells. On CCR5 cells, mutation 1 had fitness (w) 6.6671 (variance = 0.1926) and mutation 5, fitness 0.3998 (0.0015), relative to ADA ($w = 1$) (Figure 2 and [Table S1](#)). The envelope with both mutations had fitness 1.5381 (0.0258), whereas the expected fitness of the double mutant with independent effects of the mutations is $6.6671 \times 0.3998 = 2.6655$, giving an epistatic deviation of $\varepsilon = 1.5381 - 2.6655 = -1.1271$ (0.1252). On CXCR4 cells, mutation 4 had fitness 0.2129 (0.0082) and mutation 5 had fitness 0.2071 (0.0267), relative to ADA-1. Therefore, the expected fitness of the double mutant in the absence of an interaction is 0.0441. However, the double mutant had fitness 0.6439 (0.0138), giving $\varepsilon = 0.5998$ (0.0156).

Of the 31 envelopes engineered with more than two mutations, higher-order net interactions could be tested for only 23 with each cell-type assay because some lower-order mutants were not constructed. Of these 23 higher-order net interactions that could be tested, 14

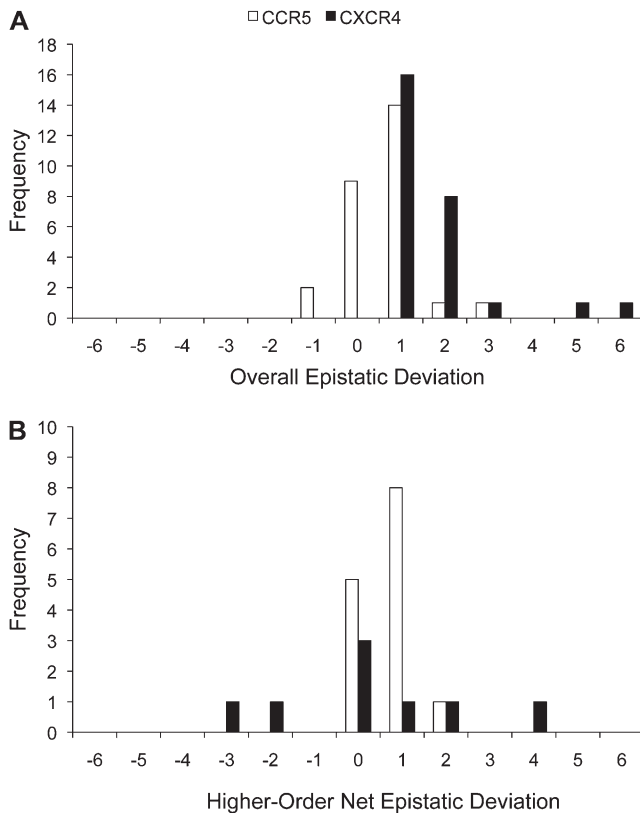


FIGURE 3.—Frequency distributions of epistatic deviation for statistically significant interactions on cells expressing CCR5 or CXCR4. Values on the epistatic deviation axis are upper bounds of the intervals. (A) Overall epistatic deviation. (B) Higher-order net epistatic deviation.

(61%) were statistically significant on CCR5 cells and 8 (35%) were significant on CXCR4 cells (Figure 2 and Table S1). Significant higher-order net interactions on both CCR5 and CXCR4 cells were both positive and negative (Figure 3b and Table S1). For CCR5 cells, the median net epistatic deviation was 0.0604 (range -0.2856 – 1.7724), and for CXCR4 cells, the median net epistatic deviation was -0.2750 (-3.8550 – 3.0768). The highest-order (involving the most residues) significant net interaction occurred on CCR5 cells and involved five mutations (3, 4, 5, 6, and 7) (Figure 2 and Table S1). Therefore, higher-order net interactions were common and could be complex. In addition, higher-order net epistatic deviation on CXCR4 cells had a substantially broader range than on CCR5 cells.

The form of epistasis: The form of fitness epistasis may be classified in various ways according to the signs of the fitness effects of the interacting mutations and the epistatic deviation (PHILLIPS *et al.* 2000). Essentially, if the sign of the fitness effects is opposite to that of the epistatic deviation, epistasis is antagonistic, and if the signs are the same, epistasis is synergistic. Antagonistic epistasis may be further classified as either compensatory or decompensatory. If the fitness effects of individual mutations are negative and epistasis is positive,

then the fitness effect of the combined mutations is less negative than expected from the independent effects of the mutations, and the interaction is compensatory. If the reverse, that is, the fitness effects of the individual mutations are positive and epistasis is negative, then the fitness effect of the combined mutations is less positive than expected from the independent effects of the mutations, and the interaction is decompensatory. A simple classification that accommodates interacting mutations having fitness effects with different signs and that accommodates higher-order interactions was used: if the fitness effects of interacting mutations differ in sign, then, if epistasis is positive, the interaction is compensatory, and if epistasis is negative, the interaction is decompensatory.

On CCR5 cells, fitness was calculated relative to the CCR5-adapted ADA wild type, and the 27 significant cases of overall epistasis were mostly compensatory (59%), although there were also some cases of decompensatory (11%) and negative synergistic (30%) epistasis. On CXCR4 cells, fitness was calculated relative to the CXCR4-adapted mutant ADA-1, and the 27 significant cases of overall epistasis were exclusively compensatory. This reflects the consistently negative fitness effects of interacting mutations and the consistently positive overall epistasis of interactions on CXCR4 cells.

The magnitude of epistasis: The magnitude of epistasis was measured as the number of orders-of-magnitude change in fitness due to overall epistasis. On CCR5 cells, the magnitude for significant overall epistasis ranged from -2.65 to 2.04 (median = 0.46), and on CXCR4 cells, the magnitude ranged from 0.86 to 9.04 (median = 3.15) (Figure 4 and Table S1). These values match the signs and ranges of significant overall epistatic deviations (Figure 3a and Table S1). The large values, especially for effects measured on CXCR4 cells, reflect that although some single mutations are practically lethal, in combinations they produce large increases in fitness (Figure 2 and Table S1). For example, the CXCR4-adapted mutant, ADA-1, which carries all seven mutations, exhibited the highest magnitude on CXCR4 cells (9.04), but four of the mutations had only negligible fitness when measured singly. Therefore, overall epistasis on CXCR4 cells was consistently positive, increasing fitness by a median value of over 3 orders of magnitude, even though the net effects of higher-order interactions were sometimes negative (Figure 3b and Table S1).

The evolutionary consequences of epistasis: The observed common, complex, and strong fitness epistasis is expected to have a significant impact on the dynamics of adaptation. A useful way to identify this impact is to construct the minimum-length mutational trajectories between two alleles. A minimum-length mutational trajectory is one that involves only single-mutation steps and no reversals and is therefore the most direct evolutionary path between alleles. Sign epistasis, in

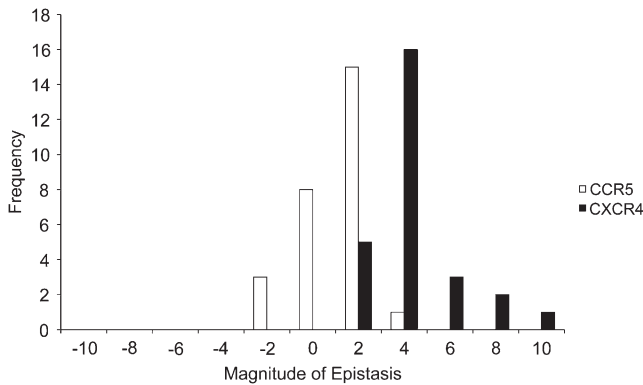


FIGURE 4.—The frequency distribution of the magnitude of significant overall epistasis, E , on cells expressing CCR5 or CXCR4. Values on the magnitude axis are upper bounds of the intervals and are in units of orders of magnitude.

which the sign of a mutation's fitness effect depends on its genetic background, will make a minimum-length mutational trajectory selectively inaccessible, thereby constraining adaptation (WEINREICH *et al.* 2005; WEINREICH *et al.* 2006; DEPRISTO *et al.* 2007). Figure 5 shows the observable minimum-length mutational trajectories on CXCR4 cells, given the constructed mutant envelopes, from the CCR5-adapted wild-type ADA to the envelope with the highest fitness on CXCR4 cells, mutant 13457 (Figure 2 and Table S1). Because the five mutational differences between ADA and mutant 13457 may in principle occur in any order, these mutations produce $5! = 120$ minimum-length trajectories (permutations) between these alleles. However, because not all mutant envelopes were constructed, only 24 complete trajectories are observable. Single-mutation steps in these trajectories were considered to increase fitness relative to the preceding mutant regardless of whether the increase was statistically significant. This makes the identification of mutations that do not increase fitness conservative. Of the 51 single-mutation steps in the observable trajectories from ADA to mutant 13457, 19 do not increase fitness. Indeed, none of the 24 observable minimum-length mutational trajectories from ADA to mutant 13457 are selectively accessible because each contains at least one single-mutation step that does not increase fitness. Other trajectories from ADA to mutant 13457, involving mutants that were not constructed, may be selectively accessible. These trajectories would necessarily involve a four-mutation envelope containing mutation 1 (Figure 5). The mutants are possibly 1345, 1347, 1357, and 1457, some of which may be selectively accessible from the accessible mutants 135, 157, and 457, or other possibly accessible triple mutants that were not constructed. An example would be the trajectory $135 \rightarrow 1345 \rightarrow 13457$. However, any selectively accessible trajectories between ADA and mutant 13457 that may exist could not be traced because not all of the necessary mutants were constructed.

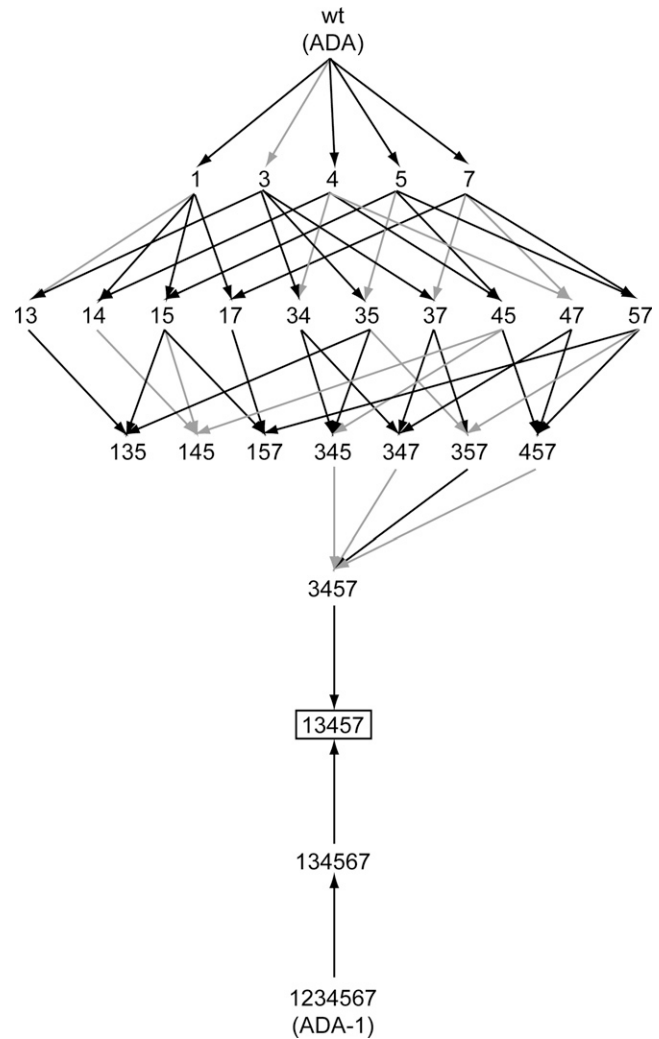


FIGURE 5.—Minimum-length mutational trajectories on CXCR4 cells. The shortest observable mutational trajectories linking the CCR5-adapted wild-type ADA allele (wt), and the CXCR4-adapted ADA-1 allele (1234567), to the allele with the highest fitness when infecting cells expressing CXCR4 (13457). Mutations are numbered as in Figure 1. Only those mutation combinations that were engineered are shown. Solid arrows indicate single mutations that increase fitness. Shaded arrows indicate single mutations that do not increase fitness.

Sign epistasis constrains adaptation for the simple reason that in the absence of sign epistasis the mutations that characterize the allele with highest fitness must be advantageous singly and in any combination. In this case, every minimum-length mutational trajectory to the fittest allele is selectively accessible. Even with magnitude epistasis, in which only the magnitude of the fitness effect (not the sign) depends on the genetic background (the other mutations), every minimum-length trajectory will be selectively accessible (WEINREICH *et al.* 2005). A change in the sign of a mutation's fitness effect dependent on the other mutations with which it is found is equivalent to one or more nonbeneficial mutational steps in a trajectory because the mutation

must be beneficial in the final allele. Sign epistasis is evident in Figure 5. For example, the fitness effect of adding mutation 4 to the wild-type ADA background ($wt \rightarrow 4$) is positive, increasing relative fitness from 0.0109 to 0.2129 (Figure 2 and Table S1). However, the fitness effect of adding mutation 4 to a background containing mutation 7 ($7 \rightarrow 47$) is negative, decreasing fitness from 0.1283 to 0.0459. Therefore, although not all 120 possible minimum-length trajectories between ADA and mutant 13457 could be observed, because not all of the necessary envelopes were constructed, the fact that none of the observable trajectories were selectively accessible indicates severe constraints on adaptation.

Since the variant that evolved in response to selection by CXCR4, ADA-1, contains all seven mutations, in contrast to the observed envelope with the highest fitness on CXCR4 cells, mutant 13457, which contains only five mutations, it is possible that ADA-1 sits on a local fitness peak, implying a fitness landscape with multiple peaks. However, this is not the case since there is a selectively accessible minimum-length mutational trajectory from ADA-1 to mutant 13457 (Figure 5). The evolution of ADA-1 from ADA, together with the numerous selectively inaccessible minimum-length trajectories from ADA to mutant 13457, suggests that ADA-1 lies on an indirect, but selectively accessible, mutational trajectory from ADA to mutant 13457. Such a trajectory would involve reversals of mutations 2 and 6. This trajectory could not be traced with the constructed envelopes. Other envelopes that were not constructed may have even higher fitnesses than mutant 13457. However, the evolution of ADA-1 suggests strong constraints on the evolution of any mutant with higher fitness.

DISCUSSION

Although the statistical test for fitness epistasis was conservative, 56% of interactions tested for overall epistasis on each cell type were statistically significant. Interactions involved not only pairs of amino acids, but also higher-order epistasis above pairwise effects. Previous studies have reported direct evidence of fitness epistasis for *pairs* of residues, either intergenically (e.g., SANJUAN *et al.* 2004; POON and CHAO 2005; SANJUAN *et al.* 2005; POON and CHAO 2006) or intragenically (e.g., BONHOEFFER *et al.* 2004; LUNZER *et al.* 2005; PEPIN and WICHMAN 2007). The present study demonstrates directly and unequivocally net higher-order fitness epistasis among amino acids within a protein region, that is, epistasis among three or more residues in addition to any lower-order interactions occurring among the same residues. The highest-level significant net interaction occurred among five amino acids. We tested whether data from a recent study showing sign epistasis in the evolution of antibiotic resistance in *Escherichia coli*

β -lactamase (WEINREICH *et al.* 2006) also provide evidence of higher-order epistasis. Using our approach, we found that out of 16 mutants with more than two mutations, out of a possible five, 12 exhibit statistically significant higher-order epistasis. The highest-level interactions occurred with quadruple mutants. Therefore, higher-order epistasis may be common among proteins.

Interactions occurred across the 134-amino acid C2–V3 protein region, although five of the seven mutations were within the 35-amino acid V3 region. A high density of interactions has also been reported in a survey of published data on suppressor mutations in viruses, prokaryotes and eukaryotes (POON *et al.* 2005; DAVIS *et al.* 2009). In some cases suppressor mutations recover fitness in individuals with a deleterious mutation by suppressing the phenotypic effect of the deleterious mutation and are therefore compensatory. For viruses, POON *et al.* (2005) estimate approximately nine compensatory mutations for every deleterious mutation and that about 64% of interactions occur intragenically rather than intergenically. In an experimental analysis of the DNA bacteriophage ϕ X174, POON and CHAO (2005) report approximately nine compensatory mutations for each deleterious mutation, that about half of compensatory mutations are intragenic, and that the average intragenic compensatory mutation clusters significantly within 20% of the protein's length from the deleterious mutation. SHAPIRO *et al.* (2006) inferred the coevolutionary history of amino acid replacements in 177 RNA virus genes to detect positive fitness epistasis. They found that interactions most often occur within a distance of 15 amino acids. The concentration of epistasis within short stretches of amino acids suggests that interactions occur directly between residues or indirectly through local conformational effects on protein structure. This conclusion is supported by a recent study of the structural mechanism of epistasis within a protein (ORTLUND *et al.* 2007).

Overall epistasis was both positive and negative on CCR5 cells, but was exclusively positive on CXCR4 cells. The positive epistasis on CXCR4 cells may be explained by the fact that the single mutations analyzed tend to increase fitness relative to ADA, but are deleterious relative to ADA-1 (Figure 2 and Table S1). Fitness on CXCR4 cells was measured relative to that of ADA-1 and deleterious mutations tend to generate positive epistasis in microorganisms (BURCH and CHAO 2004; DE VISSER and ELENA 2007; JASNOS and KORONA 2007; KOUYOS *et al.* 2007; MAISNIER-PATIN *et al.* 2005; SANJUAN and ELENA 2006; SANJUAN *et al.* 2004). An explanation for this may be that mutations within the same gene or protein interact antagonistically on fitness because they are affecting the same functional unit or fitness component. Antagonistic interactions, in which the combined effect of mutations is less than expected from their independent, multiplicative effects, generate

positive epistasis with deleterious mutations and negative epistasis with beneficial mutations (PHILLIPS *et al.* 2000). A similar argument is made by SANJUAN and ELENA (2006) for genome-wide interactions. They argue that for simple, compact genomes, such as those of RNA viruses, antagonistic epistasis is expected to predominate because of the high probability that different mutations will affect the same functional module. The association of positive epistasis with deleterious mutations is also predicted by the biophysics of protein structure (DEPRISTO *et al.* 2005).

The magnitude of overall fitness epistasis could be very high, causing changes in fitness of over 2 orders of magnitude on CCR5 cells and over 9 orders of magnitude on CXCR4 cells. These results contrast with those from early quantitative genetics experiments on the viability effects of epistasis in *Drosophila*, which showed only weak to moderate effects (SPASSKY *et al.* 1965; TEMIN *et al.* 1969). The difference is very likely due to the early experiments dealing with many unknown mutations of small effect and ignoring lethal mutations, whereas in this study we analyzed few mutations of large effect, some of which were effectively lethal. Our results argue for epistasis being a dominant force in adaptive dynamics.

Fitness epistasis may restrict the minimum-length mutational trajectories taken during adaptive evolution. Such constraints will arise when the sign of a mutation's fitness effect depends on its genetic background (WEINREICH *et al.* 2005, 2006). In the present study, the main consequence of such sign epistasis is that all 24 observable minimum-length mutational trajectories from the CCR5-adapted wild-type ADA strain to the engineered envelope with the highest fitness on CXCR4 cells are selectively inaccessible. Although these 24 observable trajectories are only a small portion of the 120 minimum-length trajectories possible involving five mutations, their selective inaccessibility indicates severe constraints on adaptation. Any selectively accessible minimum-length trajectory must have included mutation 1 from C2 in a four-mutation intermediate before reaching the final five-mutation envelope with the highest fitness, mutant 13457. However, the fact that not all envelopes were constructed prevents us from tracing any of the potentially selectively accessible minimum-length trajectories between ADA and mutant 13457. This effect of sign epistasis was first shown for the evolution of antibiotic resistance in *E. coli* β -lactamase, in which only a minority of all minimum-length mutational trajectories are selectively accessible (WEINREICH *et al.* 2006).

Alternatively, sign epistasis may have constrained the evolutionary trajectory from ADA to mutant 13457 to be indirect, involving more than the minimum possible number of mutations. The evolution of ADA-1, the isolate with all seven mutations, from ADA in response to selection by CXCR4 (PASTORE *et al.* 2004) may have occurred as an intermediate in an indirect chain of

selectable mutational steps to mutant 13457. Such a selectively accessible, but indirect, trajectory would require a minimum of four additional mutational steps over a direct trajectory: two additional mutations to reach ADA-1 and two mutational reversions. To reach mutants with even higher fitness, which may exist but were not constructed, would also require an indirect trajectory through ADA-1. Mutational trajectories involving reversions were shown to potentially comprise a large proportion of selectively accessible trajectories in the evolution of *E. coli* β -lactamase antibiotic resistance (DEPRISTO *et al.* 2007).

Nonbeneficial single-mutation steps are unlikely to be circumvented by the spread to fixation of a double mutant, the probability of which is proportional to the product of the square of the mutation rate and the effective population size (GILLESPIE 1984; WEINREICH and CHAO 2005), because of the small within-patient effective population size of HIV-1. Although the within-patient census population size of HIV-1 is large, on the order of 10^7 infected cells (CHUN *et al.* 1997), estimates of the effective population size range from 10^2 to 10^5 (e.g., LEIGH BROWN 1997; ROUZINE and COFFIN 1999; SEO *et al.* 2002; ACHAZ *et al.* 2004; SHRINER *et al.* 2004; KOUYOS *et al.* 2006). A high rate of recombination, as observed for HIV-1 (JUNG *et al.* 2002; LEVY *et al.* 2004), is also unlikely to help because high recombination is expected to reduce the rate of adaptation on rugged fitness landscapes (WEINREICH and CHAO 2005; DE VISSER *et al.* 2009). However, because V3 is the primary target of neutralizing antibodies (ZOLLA-PAZNER 2004), fluctuating selection by antibodies (RICHMAN *et al.* 2003; WEI *et al.* 2003) could change the fitness landscape so that some trajectories become selectively accessible.

Nevertheless, severe constraints on adaptation to CXCR4 are consistent with the apparent difficulty of evolving a CXCR4-utilizing variant in culture (PASTORE *et al.* 2004) and may explain why switching from CCR5 use to CXCR4 use tends to occur only late in infection (PHILPOTT 2003). In selecting for CXCR4 use in CCR5-adapted virus strains, PASTORE *et al.* (2004) report that only one out of four strains evolved exclusive CXCR4 use, the remaining three strains evolving dual coreceptor use only. In addition, evolutionary intermediates in these experiments are more sensitive to cell entry inhibitors (CCR5 or CXCR4 ligands) than either the CCR5-adapted parental viruses or the endpoint viruses (PASTORE *et al.* 2007), suggesting a decrease in fitness along the evolutionary path between phenotypes. This could mean that in an environment with both chemokine coreceptors available, each coreceptor-usage phenotype represents a fitness peak. Furthermore, adaptation to CXCR4 involved on average two to four mutations, mainly within V3, which occurred in parallel for isolates from the same viral strain (PASTORE *et al.* 2004), implying strong constraints on evolutionary

pathways. Finally, although mutations in V3 are necessary for coreceptor switching, they generally reduce viral infectivity and mutations in V1/V2 or C2 may be necessary to compensate for this loss of fitness (PASTORE *et al.* 2006).

In this study, mutation 1 in C2 increased fitness by over sixfold compared to the CCR5-adapted wild-type ADA on CCR5 cells. A similar effect of this mutation is reported by PASTORE *et al.* (2006). This may be explained by mutation 1 altering a putative N-linked glycosylation motif. These motifs are sometimes advantageous under selection by neutralizing antibodies (WEI *et al.* 2003) and are known to affect coreceptor usage (*e.g.*, OGERT *et al.* 2001; POLLAKIS *et al.* 2001). Mutation 1 alone does not appreciably increase fitness on CXCR4 cells in comparison to the wild type, but does increase fitness substantially in combination with other mutations that are moderately beneficial relative to the wild type (*e.g.*, mutants 14 and 125; Figure 2). Measuring fitness relative to the CXCR4-adapted isolate ADA-1, these combinations exhibit significant positive epistasis and are compensatory. This supports our observation that the evolutionary trajectory from the wild type to the variant with the highest fitness on CXCR4 cells must involve mutation 1 and supports the observation by PASTORE *et al.* (2006) that mutation 1 compensates for the loss of infectivity caused by V3 mutations.

Mutation 3 in V3 also eliminates a putative N-linked glycosylation motif, and there is a significant compensatory interaction between mutations 1 and 3 on CXCR4 cells. V3 mutations 4 and 7 have been implicated in affecting coreceptor usage in functional studies (DE JONG *et al.* 1992; FOUCHIER *et al.* 1992; HUNG *et al.* 1999; PASTORE *et al.* 2006) and structural modeling studies (CARDOZO *et al.* 2007; GORRY *et al.* 2007; ROSEN *et al.* 2006). These mutations had significant pairwise and higher-order interactions with mutations 1 and 3, as well as other mutations. The V3 mutations studied here have been reported for subtype B viruses infecting patients (KUIKEN *et al.* 2009), with mutations 4 and 7 more common in CXCR4-using viruses than in CCR5-using viruses (DA SILVA 2006). Mutation 4 has also been reported to be positively selected during the switch from CCR5 use to CXCR4 use in a patient (COETZER *et al.* 2008). In addition, linkage disequilibrium between amino acids at the sites of mutations 4 and 7 and other sites has been reported frequently for HIV-1 subtype B sequences sampled from patients (*e.g.*, BICKEL *et al.* 1996; KORBER *et al.* 1993; POON *et al.* 2007). This disequilibrium may be caused by the fitness epistasis reported here.

We have shown that fitness epistasis is common among the amino acids of a short protein region and that it may be complex, involving not only pairwise interactions but also higher-order interactions. The interactions are mostly compensatory, often very strong and appear to severely constrain the adaptation of the

HIV-1 C2–V3 region to the chemokine coreceptor CXCR4. Sign epistasis may help explain the difficulty in evolving CXCR4 use in culture and the delayed evolution of this phenotype in natural infection. These results support the view that an understanding of protein evolution requires knowledge of the common, complex, and strong fitness interactions among amino acids.

This manuscript was improved by the suggestions of three reviewers. We acknowledge the support of the School of Molecular and Biomedical Science, and its Discipline of Genetics, of the University of Adelaide. This work was supported by grants R01 AI052778 from the National Institute of Allergy and Infectious Diseases (NIAID). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIAID or the National Institutes of Health.

LITERATURE CITED

- ACHAZ, G., S. PALMER, M. KEARNEY, F. MALDARELLI, J. W. MELLORS *et al.*, 2004 A robust measure of HIV-1 population turnover within chronically infected individuals. *Mol. Biol. Evol.* **21**: 1902–1912.
- BERSHTEIN, S., M. SEGAL, R. BEKERMANN, N. TOKURIKI and D. S. TAWFIK, 2006 Robustness-epistasis link shapes the fitness landscape of a randomly drifting protein. *Nature* **444**: 929–932.
- BICKEL, P. J., P. C. COSMAN, R. A. OLSHEN, P. C. SPECTOR, A. G. RODRIGO *et al.*, 1996 Covariability of V3 loop amino acids. *AIDS Res. Hum. Retroviruses* **12**: 1401–1411.
- BONHOEFFER, S., C. CHAPPEY, N. T. PARKIN, J. M. WHITCOMB and C. J. PETROPOULOS, 2004 Evidence for positive epistasis in HIV-1. *Science* **306**: 1547–1550.
- BURCH, C. L., and L. CHAO, 2000 Evolvability of an RNA virus is determined by its mutational neighbourhood. *Nature* **406**: 625–628.
- BURCH, C. L., and L. CHAO, 2004 Epistasis and its relationship to canalization in the RNA virus $\phi 6$. *Genetics* **167**: 559–567.
- CARDOSO, T., T. KIMURA, S. PHILPOTT, B. WEISER, H. BURGER *et al.*, 2007 Structural basis for coreceptor selectivity by the HIV type 1 V3 loop. *AIDS Res. Hum. Retroviruses* **23**: 415–426.
- CHUN, T. W., L. CARRUTH, D. FINZI, X. SHEN, J. A. DIGIUSEPPE *et al.*, 1997 Quantification of latent tissue reservoirs and total body viral load in HIV-1 infection. *Nature* **387**: 183–188.
- COETZER, M., R. NEDELLEC, J. SALKOWITZ, S. McLAUGHLIN, Y. LIU *et al.*, 2008 Evolution of CCR5 use before and during coreceptor switching. *J. Virol.* **82**: 11758–11766.
- COFFIN, J. M., 1999 Molecular biology of HIV, pp. 3–40 in *The Evolution of HIV*, edited by K. A. CRANDALL. The John Hopkins University Press, Baltimore.
- CORMIER, E. G., and T. DRAGIC, 2002 The crown and stem of the V3 loop play distinct roles in human immunodeficiency virus type 1 envelope glycoprotein interactions with the CCR5 coreceptor. *J. Virol.* **76**: 8953–8957.
- DA SILVA, J., 2006 Site-specific amino acid frequency, fitness and the mutational landscape model of adaptation in human immunodeficiency virus type 1. *Genetics* **174**: 1689–1694.
- DAVIS, B. H., A. F. Y. POON and M. C. WHITLOCK, 2009 Compensatory mutations are repeatable and clustered within proteins. *Proc. R. Soc. B. Biol. Sci.* **276**: 1823–1827.
- DE JONG, J. J., A. DE RONDE, W. KEULEN, M. TERSMETTE and J. GOUDSMIT, 1992 Minimal requirements for the human immunodeficiency virus type 1 V3 domain to support the syncytium-inducing phenotype: analysis by single amino acid substitution. *J. Virol.* **66**: 6777–6780.
- DE VISSER, J. A. G. M., and S. F. ELENA, 2007 The evolution of sex: empirical insights into the roles of epistasis and drift. *Nat. Rev. Genet.* **8**: 139–149.
- DE VISSER, J. A. G. M., S. A. PARK and J. KRUG, 2009 Exploring the effect of sex on empirical fitness landscapes. *Am. Nat.* **174**: S15–S30.

- DEPRISTO, M. A., D. L. HARTL and D. M. WEINREICH, 2007 Mutational reversions during adaptive protein evolution. *Mol. Biol. Evol.* **24**: 1608–1610.
- DEPRISTO, M. A., D. M. WEINREICH and D. L. HARTL, 2005 Missense meanderings in sequence space: a biophysical view of protein evolution. *Nat. Rev. Genet.* **6**: 678–687.
- DITTMAR, M. T., A. MCKNIGHT, G. SIMMONS, P. R. CLAPHAM, R. A. WEISS *et al.*, 1997 HIV-1 tropism and co-receptor use. *Nature* **385**: 495–496.
- EDINGER, A. L., T. L. HOFFMAN, M. SHARRON, B. LEE, B. O'DOWD *et al.*, 1998 Use of GP1, GP15, and STRL33 as coreceptors by diverse human immunodeficiency virus type 1 and simian immunodeficiency virus envelope proteins. *Virology* **249**: 367–378.
- FELSENSTEIN, J., 1988 Sex and the evolution of recombination, pp. 74–86 in *The Evolution of Sex*, edited by R. E. MICHOD and B. R. LEVIN. Sinauer Associates, Sunderland.
- FISHER, R. A., 1930 *The Genetical Theory of Natural Selection*. Clarendon Press, Oxford.
- FOUCHIER, R. A., M. GROENINK, N. A. KOOTSTRA, M. TERSMETTE, H. G. HUISMAN *et al.*, 1992 Phenotype-associated sequence variation in the third variable domain of the human immunodeficiency virus type 1 gp120 molecule. *J. Virol.* **66**: 3183–3187.
- GILLESPIE, J. H., 1984 Molecular evolution over the mutational landscape. *Evolution* **38**: 1116–1129.
- GOODMAN, L. A., 1962 The variance of the product of k random variables. *J. Am. Statist. Assoc.* **57**: 54–60.
- GORRY, P. R., R. L. DUNFEE, M. E. MEFFORD, K. KUNSTMAN, T. MORGAN *et al.*, 2007 Changes in the V3 region of gp120 contribute to unusually broad coreceptor usage of an HIV-1 isolate from a CCR5 $\Delta 32$ heterozygote. *Virology* **362**: 163–178.
- HUANG, C.-C., S. N. LAM, P. ACHARYA, M. TANG, S.-H. XIANG *et al.*, 2007 Structures of the CCR5 N terminus and of a tyrosine-sulfated antibody with HIV-1 gp120 and CD4. *Science* **317**: 1930–1934.
- HUANG, C.-C., M. TANG, M.-Y. ZHANG, S. MAJEED, E. MONTABANA *et al.*, 2005 Structure of a V3-containing HIV-1 gp120 core. *Science* **310**: 1025–1028.
- HUNG, C. S., N. VANDER HEYDEN and L. RATNER, 1999 Analysis of the critical domain in the V3 loop of human immunodeficiency virus type 1 gp120 involved in CCR5 utilization. *J. Virol.* **73**: 8216–8226.
- HWANG, S. S., T. J. BOYLE, H. K. LYERLY and B. R. CULLEN, 1991 Identification of the envelope V3 loop as the primary determinant of cell tropism in HIV-1. *Science* **253**: 71–74.
- JASNOS, L., and R. KORONA, 2007 Epistatic buffering of fitness loss in yeast double deletion strains. *Nat. Genet.* **39**: 550–554.
- JUNG, A., R. MAIER, J. P. VARTANIAN, G. BOCHAROV, V. JUNG *et al.*, 2002 Multiply infected spleen cells in HIV patients. *Nature* **418**: 144.
- KIMURA, M., 1956 A model of a genetic system which leads to closer linkage by natural selection. *Evolution* **10**: 278–287.
- KONDRASHOV, A. S., 1993 Classification of hypotheses on the advantage of amphimixis. *J. Hered.* **84**: 372–387.
- KORBER, B. T., R. M. FARBER, D. H. WOLPERT and A. S. LAPEDES, 1993 Covariation of mutations in the V3 loop of human immunodeficiency virus type 1 envelope protein: an information theoretic analysis. *Proc. Natl. Acad. Sci. USA* **90**: 7176–7180.
- KOUYOS, R. D., C. L. ALTHAUS and S. BONHOEFFER, 2006 Stochastic or deterministic: What is the effective population size of HIV-1? *Trends Microbiol.* **14**: 507–511.
- KOUYOS, R. D., O. K. SILANDER and S. BONHOEFFER, 2007 Epistasis between deleterious mutations and the evolution of recombination. *Trends Ecol. Evol.* **22**: 308–315.
- KUIKEN, C., T. LEITNER, B. FOLEY, B. HAHN, P. MARX *et al.* (Editors), 2009 *HIV Sequence Compendium 2009*. Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos, NM.
- LEIGH BROWN, A. J., 1997 Analysis of HIV-1 *env* gene sequences reveals evidence for a low effective number in the viral population. *Proc. Natl. Acad. Sci. USA* **94**: 1862–1865.
- LEVY, D. N., G. M. ALDROVANDI, O. KUTSCH and G. M. SHAW, 2004 Dynamics of HIV-1 recombination in its natural target cells. *Proc. Natl. Acad. Sci. USA* **101**: 4204–4209.
- LEWONTIN, R. C., and K.-I. KOJIMA, 1960 The evolutionary dynamics of complex polymorphisms. *Evolution* **14**: 458–472.
- LUNZER, M., S. P. MILLER, R. FELSHEIM and A. M. DEAN, 2005 The biochemical architecture of an ancient adaptive landscape. *Science* **310**: 499–501.
- MAISNIER-PATIN, S., and D. I. ANDERSSON, 2004 Adaptation to the deleterious effects of antimicrobial drug resistance mutations by compensatory evolution. *Res. Microbiol.* **155**: 360–369.
- MAISNIER-PATIN, S., J. R. ROTH, A. FREDRIKSSON, T. NYSTROM, O. G. BERG *et al.*, 2005 Genomic buffering mitigates the effects of deleterious mutations in bacteria. *Nat. Genet.* **37**: 1376.
- MAMMANO, F., V. TROUPLIN, V. ZENNOU and F. CLAVEL, 2000 Retracing the evolutionary pathways of human immunodeficiency virus type 1 resistance to protease inhibitors: virus fitness in the absence and in the presence of drug. *J. Virol.* **74**: 8524–8531.
- NIJHUIS, M., R. SCHUURMAN, D. D. JONG, J. ERICKSON, E. GUSTCHINA *et al.*, 1999 Increased fitness of drug resistant HIV-1 protease as a result of acquisition of compensatory mutations during suboptimal therapy. *AIDS* **13**: 2349–2359.
- OGERT, R. A., M. K. LEE, W. ROSS, A. BUCKLER-WHITE, M. A. MARTIN *et al.*, 2001 N-Linked glycosylation sites adjacent to and within the V1/V2 and the V3 loops of dualtropic human immunodeficiency virus type 1 isolate DH12 gp120 affect coreceptor usage and cellular tropism. *J. Virol.* **75**: 5998–6006.
- ORTLUND, E. A., J. T. BRIDGHAM, M. R. REDINBO and J. W. THORNTON, 2007 Crystal structure of an ancient protein: evolution by conformational epistasis. *Science* **317**: 1544–1548.
- PASTORE, C., R. NEDELLEC, A. RAMOS, O. HARTLEY, J. L. MIAMIDIAN *et al.*, 2007 Conserved changes in envelope function during human immunodeficiency virus type 1 coreceptor switching. *J. Virol.* **81**: 8165–8179.
- PASTORE, C., R. NEDELLEC, A. RAMOS, S. PONTOW, L. RATNER *et al.*, 2006 Human immunodeficiency virus type 1 coreceptor switching: V1/V2 gain-of-fitness mutations compensate for V3 loss-of-fitness mutations. *J. Virol.* **80**: 750–758.
- PASTORE, C., A. RAMOS and D. E. MOSIER, 2004 Intrinsic obstacles to human immunodeficiency virus type 1 coreceptor switching. *J. Virol.* **78**: 7565–7574.
- PEPIN, K. M., and H. A. WICHMAN, 2007 Variable epistatic effects between mutations at host recognition sites in ϕ X174 bacteriophage. *Evolution* **61**: 1710–1724.
- PHILLIPS, P. C., S. P. OTTO and M. C. WHITLOCK, 2000 Beyond the average: the evolutionary importance of gene interactions and variability of epistatic effects, pp. 20–38 in *Epistasis and the Evolutionary Process*, edited by J. B. WOLF, E. D. BRODIE and M. J. WADE. Oxford University Press, Oxford.
- PHILPOTT, S. M., 2003 HIV-1 coreceptor usage, transmission, and disease progression. *Curr. HIV Res.* **1**: 217–227.
- PLATT, E. J., J. P. DURNIN and D. KABAT, 2005 Kinetic factors control efficiencies of cell entry, efficacies of entry inhibitors, and mechanisms of adaptation of human immunodeficiency virus. *J. Virol.* **79**: 4347–4356.
- POLLAKIS, G., S. KANG, A. KLIPHUIS, M. I. CHALABY, J. GOUDSMIT *et al.*, 2001 N-linked glycosylation of the HIV Type-1 gp120 envelope glycoprotein as a major determinant of CCR5 and CXCR4 coreceptor utilization. *J. Biol. Chem.* **276**: 13433–13441.
- POON, A., and L. CHAO, 2005 The rate of compensatory mutation in the DNA bacteriophage ϕ X174. *Genetics* **170**: 989–999.
- POON, A., B. H. DAVIS and L. CHAO, 2005 The coupon collector and the suppressor mutation: Estimating the number of compensatory mutations by maximum likelihood. *Genetics* **170**: 1323–1332.
- POON, A. F., and L. CHAO, 2006 Functional origins of fitness effect-sizes of compensatory mutations in the DNA bacteriophage ϕ X174. *Evolution* **60**: 2032–2043.
- POON, A. F., F. I. LEWIS, S. L. POND and S. D. FROST, 2007 An evolutionary-network model reveals stratified interactions in the V3 loop of the HIV-1 envelope. *PLoS Comput. Biol.* **3**: e231.
- RICHMAN, D. D., T. WRIN, S. J. LITTLE and C. J. PETROPOULOS, 2003 Rapid evolution of the neutralizing antibody response to HIV type 1 infection. *Proc. Natl. Acad. Sci. USA* **100**: 4144–4149.
- ROSEN, O., M. SHARON, S. R. QUADT-AKABAYOV and J. ANGLISTER, 2006 Molecular switch for alternative conformations of the HIV-1 V3 region: implications for phenotype conversion. *Proc. Natl. Acad. Sci. USA* **103**: 13950–13955.

- ROUZINE, I. M., and J. M. COFFIN, 1999 Linkage disequilibrium test implies a large effective population number for HIV in vivo. *Proc. Natl. Acad. Sci. USA* **96**: 10758–10763.
- SANJUAN, R., J. M. CUEVAS, A. MOYA and S. F. ELENA, 2005 Epistasis and the adaptability of an RNA virus. *Genetics* **170**: 1001–1008.
- SANJUAN, R., and S. F. ELENA, 2006 Epistasis correlates to genomic complexity. *Proc. Natl. Acad. Sci. USA* **103**: 14402–14405.
- SANJUAN, R., A. MOYA and S. F. ELENA, 2004 The contribution of epistasis to the architecture of fitness in an RNA virus. *Proc. Natl. Acad. Sci. USA* **101**: 15376–15379.
- SEO, T. K., J. L. THORNE, M. HASEGAWA and H. KISHINO, 2002 Estimation of effective population size of HIV-1 within a host: a pseudomaximum-likelihood approach. *Genetics* **160**: 1283–1293.
- SHAPIRO, B., A. RAMBAUT, O. G. PYBUS and E. C. HOLMES, 2006 A phylogenetic method for detecting positive epistasis in gene sequences and its application to RNA virus evolution. *Mol. Biol. Evol.* **23**: 1724–1730.
- SHRINER, D., R. SHANKARAPPA, M. A. JENSEN, D. C. NICKLE, J. E. MITTLER *et al.*, 2004 Influence of random genetic drift on human immunodeficiency virus type 1 *env* evolution during chronic infection. *Genetics* **166**: 1155–1164.
- SODA, Y., N. SHIMIZU, A. JINNO, H.-Y. LIU, K. KANBE *et al.*, 1999 Establishment of a new system for determination of coreceptor usages of HIV based on the human glioma NP-2 cell line. *Biochem. Biophys. Res. Commun.* **258**: 313–321.
- SOKAL, R. R., and F. J. ROHLF, 1995 *Biometry: The Principles and Practice of Statistics in Biological Research*. Freeman, New York.
- SPASSKY, B., T. DOBZHANSKY and W. W. ANDERSON, 1965 Genetics of natural populations. XXXVI. Epistatic interactions of the components of the genetic load in *Drosophila pseudoobscura*. *Genetics* **52**: 653–664.
- SPECK, R. F., K. WEHRLY, E. J. PLATT, R. E. ATCHISON, I. F. CHARO *et al.*, 1997 Selective employment of chemokine receptors as human immunodeficiency virus type 1 coreceptors determined by individual amino acids within the envelope V3 loop. *J. Virol.* **71**: 7136–7139.
- TEMIN, R. G., H. U. MEYER, P. S. DAWSON and J. F. CROW, 1969 The influence of epistasis on homozygous viability depression in *Drosophila melanogaster*. *Genetics* **61**: 497–519.
- WEI, X., J. M. DECKER, S. WANG, H. HUI, J. C. KAPPES *et al.*, 2003 Antibody neutralization and escape by HIV-1. *Nature* **422**: 307–312.
- WEINREICH, D. M., and L. CHAO, 2005 Rapid evolutionary escape by large populations from local fitness peaks is likely in nature. *Evolution* **59**: 1175–1182.
- WEINREICH, D. M., N. F. DELANEY, M. A. DEPRISTO and D. L. HARTL, 2006 Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science* **312**: 111–114.
- WEINREICH, D. M., R. A. WATSON and L. CHAO, 2005 Sign epistasis and genetic constraint on evolutionary trajectories. *Evolution* **59**: 1165–1174.
- WHITLOCK, M. C., P. C. PHILLIPS, F. B. MOORE and S. J. TONSOR, 1995 Multiple fitness peaks and epistasis. *Ann. Rev. Ecol. Systematics* **26**: 601–629.
- WRIGHT, S., 1932 The roles of mutation, inbreeding, crossbreeding and selection in evolution, pp. 356–366 in *The Sixth International Congress of Genetics*, edited by D. F. JONES. Brooklyn Botanic Garden, Menasha, WI.
- WYATT, R., and J. SODROSKI, 1998 The HIV-1 envelope glycoproteins: fusogens, antigens, and immunogens. *Science* **280**: 1884–1888.
- ZOLLA-PAZNER, S., 2004 Identifying epitopes of HIV-1 that induce protective antibodies. *Nat. Rev. Immunol.* **4**: 199–210.

Communicating editor: J. LAWRENCE

GENETICS

Supporting Information

<http://www.genetics.org/cgi/content/full/genetics.109.112458/DC1>

Fitness Epistasis and Constraints on Adaptation in a Human Immunodeficiency Virus Type 1 Protein Region

**Jack da Silva, Mia Coetzer, Rebecca Nedellec, Cristina Pastore
and Donald E. Mosier**

Copyright © 2010 by the Genetics Society of America
DOI: 10.1534/genetics.109.112458

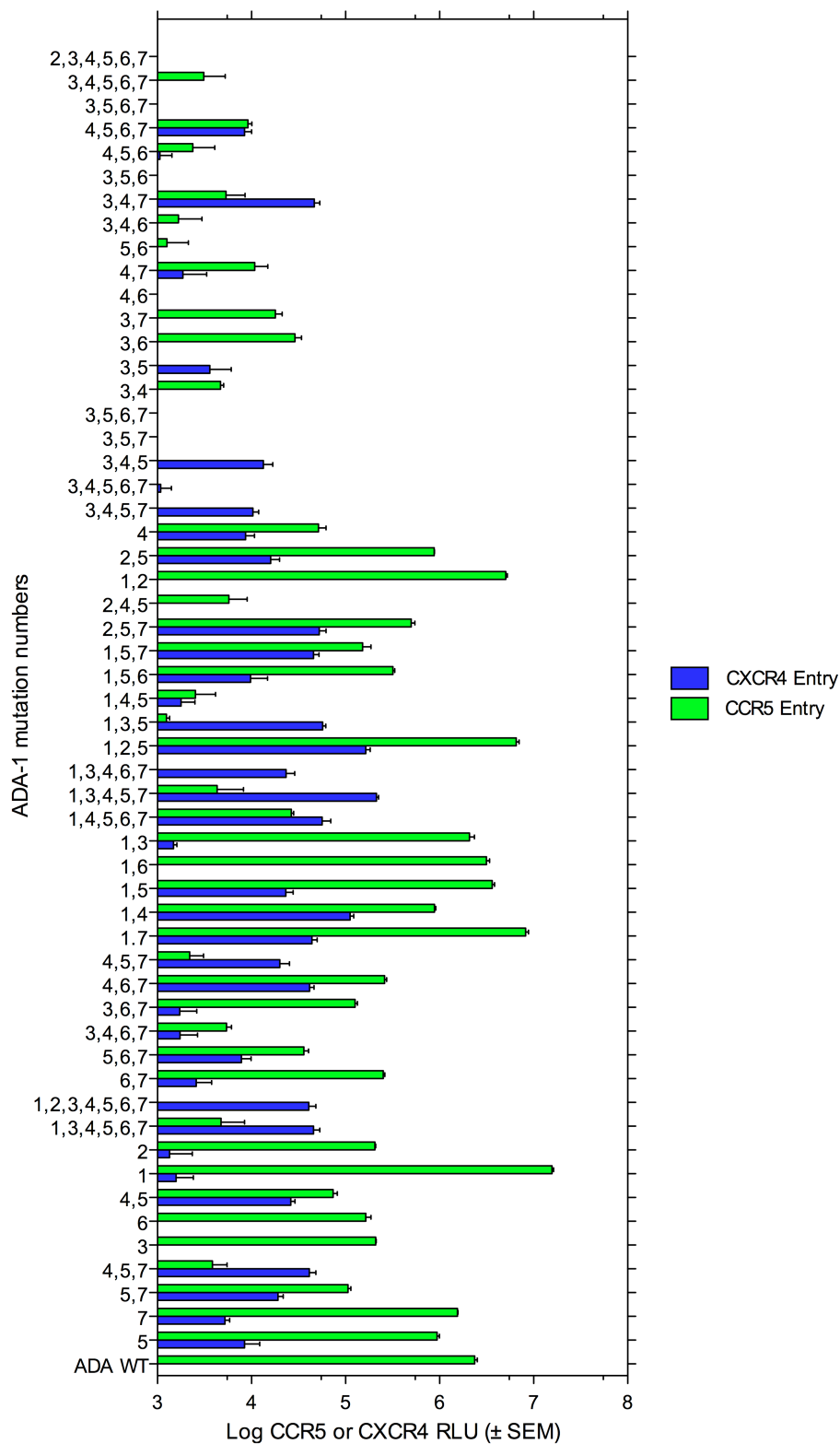


FIGURE S1.—The entry fitness of HIV-1 C2-V3 mutants on CCR5 and CXCR4 host cells in relative light units (RLU). Mutants were engineered with mutations observed for the ADA-1 CXCR4-adapted mutant on the ADA background. The method used to determine entry fitness is identical to that described in (PASTORE *et al.* 2006), with the exception that NP-2.CD4 target cells were used instead of U87.CD4 cells. Briefly, equivalent p24 levels of pseudotyped virus particles were added to each target cell line, and the luciferase activity of triplicate wells determined 48 hr. later. Experiments were repeated three times for all mutations that scored above background (~ 1000 RLU).

PASTORE, C., R. NEDELLEC, A. RAMOS, S. PONTOW, L. RATNER *et al.*, 2006 Human immunodeficiency virus type 1 coreceptor switching: V1/V2 gain-of-fitness mutations compensate for V3 loss-of-fitness mutations. *Journal of Virology* **80**: 750-758.

TABLE S1
Means and variances in fitness and epistatic deviation for assays on CCR5 and CXCR4 target cells

Mutations	CCR5										CXCR4									
	<i>W</i>	<i>w</i>	Var(<i>w</i>)	ϵ	Var(ϵ)	Sig.	E	ϵ'	Var(ϵ')	Sig.	<i>W</i>	<i>w</i>	Var(<i>w</i>)	ϵ	Var(ϵ)	Sig.	E	ϵ'	Var(ϵ')	Sig.
wt	2357403.7	1.0000	0.0147								441.1	0.0109	0.0000							
1	15717078.9	6.6671	0.1926								1584.3	0.0390	0.0012							
2	205870.3	0.0873	0.0000								1346.8	0.0331	0.0018							
3	209335.1	0.0888	0.0000								232.0	0.0057	0.0000							
4	51631.0	0.0219	0.0001								8656.0	0.2129	0.0082							
5	942373.0	0.3998	0.0015								8420.5	0.2071	0.0267							
6	163853.1	0.0695	0.0003								893.1	0.0220	0.0002							
7	1548272.5	0.6568	0.0002								5216.2	0.1283	0.0007							
12	5055976.9	2.1447	0.0202	1.5625	0.0219	*	0.5663				841.4	0.0207	0.0003	0.0194	0.0003		1.2050			
13	2078231.0	0.8816	0.0416	0.2895	0.0433		0.1729				1478.8	0.0364	0.0000	0.0362	0.0000	*	2.2137			
14	880760.1	0.3736	0.0005	0.2276	0.0032	*	0.4080				111577.6	2.7446	0.2023	2.7363	0.2024	*	2.5195			
15	3625942.7	1.5381	0.0258	-1.1271	0.1252	*	-0.2387				23136.3	0.5691	0.0398	0.5610	0.0399		1.8482			
16	3161239.0	1.3410	0.0301	0.8776	0.0436	*	0.4615				717.9	0.0177	0.0001	0.0168	0.0001		1.3144			
17	8238111.1	3.4946	0.1935	-0.8842	0.2846		-0.0980				44040.8	1.0833	0.0625	1.0783	0.0625	*	2.3358			
25	879624.7	0.3731	0.0000	0.3382	0.0000	*	1.0289				15978.0	0.3930	0.0279	0.3862	0.0280		1.7580			
34	4677.9	0.0020	0.0000	0.0000	0.0000		0.0087				356.4	0.0088	0.0000	0.0076	0.0000	*	0.8582			
35	480.4	0.0002	0.0000	-0.0353	0.0000	*	-2.2410				3612.0	0.0888	0.0112	0.0877	0.0112		1.8760			
36	28996.4	0.0123	0.0000	0.0061	0.0000		0.2995				646.2	0.0159	0.0000	0.0158	0.0000	*	2.1031			
37	17917.1	0.0076	0.0000	-0.0507	0.0000	*	-0.8850				329.9	0.0081	0.0000	0.0074	0.0000	*	1.0446			
45	73844.4	0.0313	0.0000	0.0226	0.0000	*	0.5536				26176.8	0.6439	0.0138	0.5998	0.0156	*	1.1644			
46	323.9	0.0001	0.0000	-0.0014	0.0000		-1.0445				351.2	0.0086	0.0000	0.0040	0.0000		0.2664			
47	10832.0	0.0046	0.0000	-0.0098	0.0000		-0.4956				1867.2	0.0459	0.0040	0.0186	0.0041		0.2256			
56	1265.5	0.0005	0.0000	-0.0272	0.0001	*	-1.7140				328.3	0.0081	0.0000	0.0035	0.0000		0.2491			
57	106531.5	0.0452	0.0000	-0.2174	0.0007	*	-0.7642				19148.7	0.4710	0.0117	0.4444	0.0122	*	1.2485			

67	252541.4	0.1071	0.0000	0.0615	0.0002	*	0.3705				2583.2	0.0635	0.0026	0.0607	0.0026		1.3530			
125	6550730.2	2.7788	0.1077	2.5460	0.1085	*	1.0770	1.7724	0.2556	*	164389.1	4.0436	0.5978	4.0434	0.5978	*	4.1796	3.0768	0.6661	*
135	1244.4	0.0005	0.0000	-0.2361	0.0008	*	-2.6516	0.6367	0.1693		57010.6	1.4023	0.0416	1.4023	0.0416	*	4.4835	0.7174	0.0927	
145	2542.6	0.0011	0.0000	-0.0573	0.0005		-1.7334	0.8196	0.1289		1781.2	0.0438	0.0010	0.0421	0.0010		1.4064	-3.8550	0.2588	*
156	320045.8	0.1358	0.0001	-0.0495	0.0026		-0.1350	0.2273	0.1715		9794.4	0.2409	0.0458	0.2407	0.0458		3.1331	-0.3406	0.0858	
157	152494.3	0.0647	0.0006	-1.6857	0.0448	*	-1.4323	0.5429	0.4553		45575.9	1.1211	0.0826	1.1200	0.0826	*	3.0344	-0.9638	0.1972	
245	5752.6	0.0024	0.0000	0.0017	0.0000		0.5040				883.5	0.0217	0.0004	0.0203	0.0004		1.1724			
257	499694.0	0.2120	0.0011	0.1890	0.0011	*	0.9659				52481.4	1.2909	0.1730	1.2901	0.1730	*	3.1662			
345	449.3	0.0002	0.0000	-0.0006	0.0000		-0.6106	0.0121	0.0001		13367.1	0.3288	0.0220	0.3286	0.0220		3.1161	-0.3665	0.0487	
346	1677.2	0.0007	0.0000	0.0006	0.0000		0.7212	-0.0042	0.0000		321.7	0.0079	0.0000	0.0079	0.0000	*	2.4719	-0.0194	0.0000	*
347	5376.8	0.0023	0.0000	0.0010	0.0000		0.2518	0.0615	0.0000	*	46615.1	1.1466	0.0777	1.1465	0.0777	*	3.8666	1.1129	0.0819	*
356	322.8	0.0001	0.0000	-0.0023	0.0000	*	-1.2557	0.0541	0.0001	*	310.8	0.0076	0.0000	0.0076	0.0000	*	2.4690	-0.0993	0.0112	
357	353.0	0.0001	0.0000	-0.0232	0.0000	*	-2.1923	0.2802	0.0008	*	366.5	0.0090	0.0000	0.0089	0.0000	*	1.7741	-0.5306	0.0234	*
367	126438.6	0.0536	0.0000	0.0496	0.0000	*	1.1216	0.0327	0.0002		1733.2	0.0426	0.0014	0.0426	0.0014		3.4233	-0.0413	0.0040	
456	2382.2	0.0010	0.0000	0.0004	0.0000		0.2203	0.0065	0.0001		1064.9	0.0262	0.0002	0.0252	0.0002		1.4320	-0.5821	0.0159	*
457	3851.5	0.0016	0.0000	-0.0041	0.0000		-0.5465	0.2005	0.0008	*	41248.0	1.0146	0.0895	1.0090	0.0896	*	2.2536	-0.0539	0.1215	
457	2198.6	0.0009	0.0000	-0.0048	0.0000		-0.7900	0.1998	0.0008	*	20073.9	0.4938	0.0494	0.4881	0.0495		1.9408	-0.5747	0.0814	
467	260575.8	0.1105	0.0001	0.1095	0.0001	*	2.0436	0.0592	0.0003	*	41540.0	1.0218	0.0407	1.0212	0.0407	*	3.2311	0.9379	0.0475	*
567	36089.6	0.0153	0.0000	-0.0029	0.0000		-0.0763	0.1802	0.0010	*	7841.0	0.1929	0.0077	0.1923	0.0077		2.5190	-0.3164	0.0225	
3456	3115.5	0.0013	0.0000	0.0013	0.0000		1.3884	-0.0320	0.0004		320.9	0.0079	0.0000	0.0079	0.0000	*	3.1546	0.3569	0.1027	
3457	466.8	0.0002	0.0000	-0.0003	0.0000		-0.4114	-0.2636	0.0025	*	10389.4	0.2556	0.0044	0.2555	0.0044	*	3.8984	-0.8115	0.3030	
3467	5438.3	0.0023	0.0000	0.0022	0.0000	*	1.4147	-0.1527	0.0008	*	1737.7	0.0427	0.0016	0.0427	0.0016		4.0962	-2.0614	0.1417	*
3567	436.1	0.0002	0.0000	-0.0014	0.0000	*	-0.9425	-0.2856	0.0030	*	359.3	0.0088	0.0000	0.0088	0.0000	*	3.4237	0.3770	0.0872	
3567	334.0	0.0001	0.0000	-0.0015	0.0000	*	-1.0583	-0.2856	0.0030	*	333.0	0.0082	0.0000	0.0082	0.0000	*	3.3907	0.3763	0.0872	
4567	9145.7	0.0039	0.0000	0.0035	0.0000	*	0.9871	-0.2711	0.0032	*	8461.9	0.2081	0.0046	0.2080	0.0046	*	3.2239	-0.6482	0.2265	
13457	4311.6	0.0018	0.0000	-0.0016	0.0000		-0.2698				213643.4	5.2552	0.2281	5.2552	0.2281	*	6.6208			
13467	559.1	0.0002	0.0000	-0.0004	0.0000		-0.3972				23371.1	0.5749	0.0547	0.5749	0.0547		6.6342			
14567	26450.9	0.0112	0.0000	0.0086	0.0000	*	0.6244				56417.9	1.3878	0.3193	1.3878	0.3193		5.4571			
34567	386.0	0.0002	0.0000	0.0001	0.0000	*	0.6640	0.3745	0.0143	*	1084.1	0.0267	0.0002	0.0267	0.0002		4.5751	1.7838	1.2635	

134567	4734.1	0.0020	0.0000	0.0018	0.0000		0.9288	45549.8	1.1204	0.1099	1.1204	0.1099	*	7.6078
234567	310.2	0.0001	0.0000	0.0001	0.0000	*	1.6279	340.4	0.0084	0.0000	0.0084	0.0000	*	5.5518
1234567	423.2	0.0002	0.0000	0.0002	0.0000	*	0.9389	40653.7	1.0000	0.1050	1.0000	0.1050	*	9.0382

W , absolute fitness; w , relative fitness; Var, variance; ϵ , overall epistatic deviation; ϵ' , net epistatic deviation; E, epistasis magnitude; Sig., significant epistatic deviation at an experimentwise type-1 error rate of $\alpha' = 0.05$.