

Comment on “Evidence for Positive Epistasis in HIV-1”

Kai Wang, John E. Mittler, Ram Samudrala*

Bonhoeffer *et al.* (Reports, 26 November 2004, p. 1547) presented evidence for positive epistasis in a clinical data set of HIV-1 mutants and corresponding fitness values. We demonstrate that biases in the original and simulated data sets may lead to erroneous evidence for epistasis. More rigorous statistical tests must be used to account for such biases before one can infer epistasis.

Bonhoeffer *et al.* reported “strong statistical evidence” for positive epistasis in a clinical sample of 9466 HIV-1 sequences (1). This is of potential importance because it contradicts theories that negative epistasis helps explain the evolution of recombination (2, 3). Their evidence for positive epistasis is derived from a plot showing a decelerating decline in log fitness with the number of mutations [figure 1B in (1)] and from a demonstration that the mean epistasis value for all possible pairs of alternative amino acids was significantly greater than zero [figure 2 in (1)].

Bonhoeffer *et al.* argue that their results are unlikely to be due to a paucity of viruses with low fitness in the absence of drugs, because these viruses were “generally derived” from patients on antiretroviral therapy. This argument assumes that fitness of HIV-1 in the absence of drugs is completely unrelated to fitness in the presence of drugs, which is contradicted by several sources of evidence. First, studies have shown that viral fitness in the absence of drugs gradually increases as viruses acquire secondary/compensatory mutations during therapy (4–6), resulting in positive correlation between fitness values in the presence and absence of drugs for clinical samples. Second, a positive correlation between drug hypersusceptibility and reduced fitness in the absence of drugs has been observed in clinical data sets (7, 8). Third, viruses with greatly impaired enzyme function have extremely low fitness in the presence or absence of drugs, so they will be underrepresented in clinical data sets. Finally, some of their samples may be obtained from untreated, recently treated, or lightly treated patients. All these factors indicate that viruses with low fitness but a high number of mutations are likely to be underrepresented in their clinical data set. However, they made no attempt to adjust for these biases when analyzing their data.

To demonstrate the magnitude of the effect of data biases on their conclusions, we per-

formed similar analyses using a simulated data set without epistasis. We assumed a simple model that the \log_{10} fitness value Y of a given 20-residue sequence can be written as $Y = -\sum_{i=1}^{20} s_i X_i / 20$, where s_i are constants sampled from a uniform distribution between 0 and 1, and X_i are binary variables indicating presence of mutations in each residue. We generated 10,000 genotype-phenotype pairs by randomly assigning the mutations for each residue in each sequence with a probability of 0.5. We then replicated the analysis in figure 1 in (1), either by using all 10,000 genotype-phenotype pairs (Fig. 1A) or by culling the 5% lowest fitness values (Fig. 1B). We observed a less-than-linear trend for the culled data set. When we replicated Bonhoeffer *et al.*'s all-pairwise experiments on the culled data set, we obtained a mean epistasis value of 0.0026, which is statistically significant (shuffling the culled data set 100 times produced mean epistasis values ranging from -0.00034 to 0.00030). Similar observations were made by culling 25% rather than 5% of the low-fitness viruses (Fig. 1C). These analyses demonstrate that their analysis protocols are very sensitive to data biases and can result in a false signal of positive epistasis.

To determine whether these effects would apply to their original data, Monogram Inc. (formerly Virologic Inc.) ran our software on their data set with arbitrarily scrambled genotypes and phenotypes. Although this process precluded us from determining which mutations contribute to epistasis, it allowed us to evaluate whether their data set is exempt from the effects modeled above. Using our software, we were able to replicate the distribution of fitness values in figure 1A in (1) and the decelerating trend in their figure 1B (see our Fig. 2, A and B). We then discarded either 5% or 25% of the lowest fitness values and made the plot of \log_{10} fitness values versus the number of mutations (Fig. 2, C and D). In both cases, we found a more extreme decelerating trend or even a slightly increasing trend in the tail of the curve, demonstrating artifacts caused by simple data biases. Finally, we made the plots after discarding the 5% or 25% highest fitness values (Fig. 2, E and F) and

found that the slopes in the head of both curves are less steep than that in Fig. 2B, indicating that a paucity of high-fitness viruses could also result in misleading evidence for epistasis. Because of the absence of an unbiased reference data set, we cannot perform the reshuffling procedures needed to evaluate statistically the effect of culling on Bonhoeffer *et al.*'s all-pairwise test for epistasis, as we did on the simulated data set. These analyses show how small biases in real data sets, just as with our simulated data set, can easily result in misleading conclusions regarding epistasis.

In conclusion, we demonstrate that an underrepresentation of low-fitness viruses in

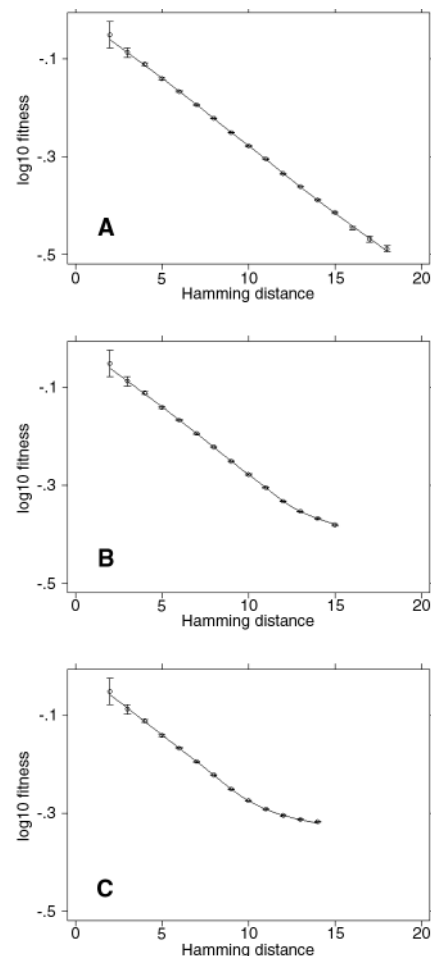


Fig. 1. Mean and standard error (circles and bars) of \log_{10} fitness values versus the number of mutations (Hamming distance) for a simulated data set in which there is no epistasis. The smooth lines here and in Fig. 2 are cubic-spline fits to the data. (A) Plot using all 10,000 samples in our simulated data set. (B) Plot with the 5% lowest fitness values removed from the data set. (C) Plot with the 25% lowest fitness values removed. The decelerating trend in the smooth curve in (B) and (C) [similar to figure 1B in (1)] indicates that selection against viruses with low fitness can result in misleading evidence for positive epistasis.

Department of Microbiology, University of Washington, Seattle, WA 98195, USA.

*To whom correspondence should be addressed. E-mail: ram@compbio.washington.edu

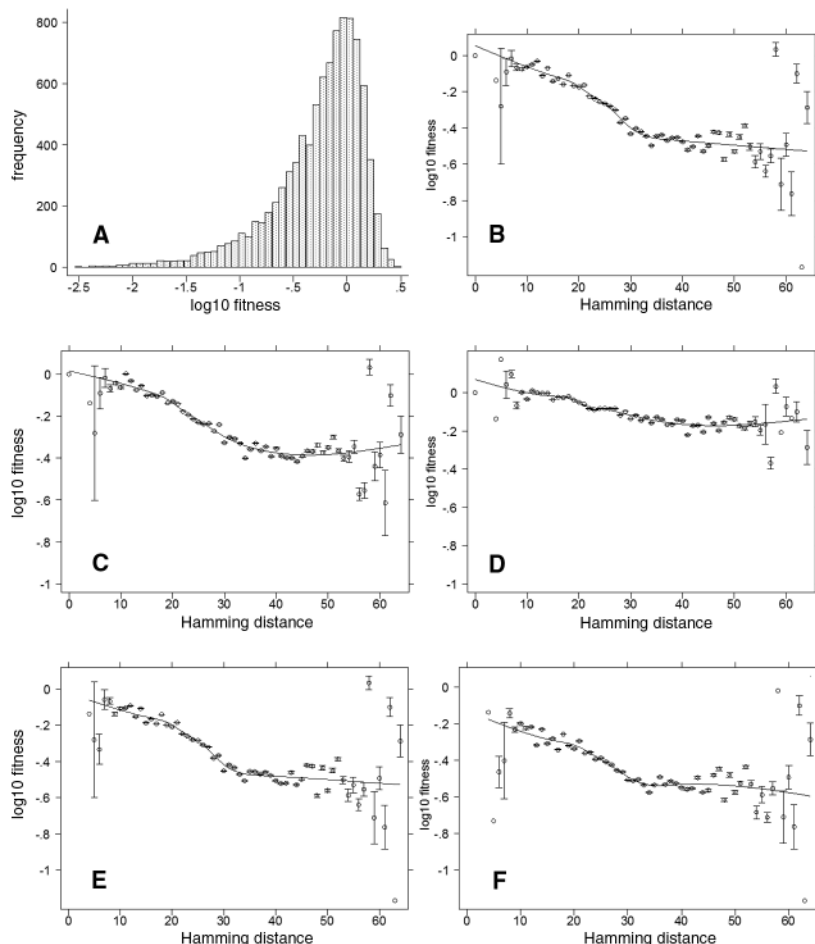


Fig. 2. Evaluation of the phenomenon shown in Fig. 1 for the original Bonhoeffer *et al.* clinical data set. **(A)** Distribution of \log_{10} fitness values of 9466 sequences in the data set, similar to their figure 1A. **(B)** Mean \log_{10} fitness values are plotted against the number of mutations (Hamming distance), similar to their figure 1B. The error bars, which are numerically identical to those in Bonhoeffer *et al.*, are the standard deviation divided by number of observations (SD/n) [mistakenly referred to as the “standard error” (SD/\sqrt{n}) in Bonhoeffer *et al.*] **(C and D)** The \log_{10} fitness values versus Hamming distance after removing samples with the lowest 5% or 25% fitness values, respectively. Compared with **(B)**, we observe a more severe change of slope (or even a reversal of slope) in the tail of the smooth curves in both plots. **(E and F)** The \log_{10} fitness values versus Hamming distance after removing samples with the highest 5% or 25% fitness values, respectively. The heads of the smooth curves (corresponding to samples with less than 20 mutations) are approximately linear curves. The approximate slopes for **(E)** (-0.0085) and **(F)** (-0.0088) at the head of the curves are both less steep than that for **(B)** (-0.013). These results indicate that evidence for epistasis from their data set could in fact be affected by data biases.

simulated or clinical data sets can easily lead to erroneous signals for positive epistasis. Although using clinically derived HIV-1 sequences to test evolutionary theories for recombination is an appealing idea, more rigorous statistical tests must be used to account for such biases before one can infer epistasis.

References and Notes

1. S. Bonhoeffer *et al.*, *Science* **306**, 1547 (2004).
2. A. S. Kondrashov, *Genet. Res.* **40**, 325 (1982).
3. R. E. Michod, B. R. Levin, *The Evolution of Sex* (Sinauer, Sunderland, MA, 1988).
4. M. Nijhuis *et al.*, *AIDS* **13**, 2349 (1999).
5. M. Nijhuis, S. Deeks, C. Boucher, *Curr. Opin. Infect. Dis.* **14**, 23 (2001).
6. M. E. Quiñones-Mateu *et al.*, *AIDS Rev.* **3**, 223 (2001).
7. A. J. Leigh-Brown *et al.*, *J. Virol.* **78**, 2242 (2004).
8. J. Martinez-Picado *et al.*, *J. Virol.* **79**, 5907 (2005).
9. We gratefully acknowledge support from the Searle Scholar's Program (R.S.), NSF grant DBI-0217241 (R.S.), and NIH grants GM068152 (R.S.), R01-HL72631 (J.E.M.), and R03-AI055394 (J.E.M.).

18 January 2006; accepted 27 March 2006
 10.1126/science.1109904