Science
AAAS

Evidence for Positive Epistasis in HIV-1
Sebastian Bonhoeffer, *et al.*
Science **306**, 1547 (2004);
DOI: 10.1126/science.1101786

This copy is for your personal, non-commercial use only.

If you wish to distribute this article to others, you can order high-quality copies for your colleagues, clients, or customers by clicking here.

Permission to republish or repurpose articles or portions of articles can be obtained by following the guidelines here.

The following resources related to this article are available online at www.sciencemag.org (this information is current as of March 17, 2010):

A correction has been published for this article at:
http://www.sciencemag.org/cgi/content/full/sci;320/5877/746c

Updated information and services, including high-resolution figures, can be found in the online version of this article at:
http://www.sciencemag.org/cgi/content/full/306/5701/1547

Supporting Online Material can be found at:
http://www.sciencemag.org/cgi/content/full/306/5701/1547/DC1

A list of selected additional articles on the Science Web sites **related to this article** can be found at:
http://www.sciencemag.org/cgi/content/full/306/5701/1547#related-content

This article **cites 25 articles**, 11 of which can be accessed for free:
http://www.sciencemag.org/cgi/content/full/306/5701/1547#otherarticles

This article has been **cited by** 77 article(s) on the ISI Web of Science.

This article has been **cited by** 30 articles hosted by HighWire Press; see:
http://www.sciencemag.org/cgi/content/full/306/5701/1547#otherarticles

This article appears in the following **subject collections**:
Evolution
http://www.sciencemag.org/cgi/collection/evolution

Downloaded from www.sciencemag.org on March 17, 2010

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published weekly, except the last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. Copyright 2004 by the American Association for the Advancement of Science; all rights reserved. The title *Science* is a registered trademark of AAAS.

volcanic systems, it is unlikely that such a pressure is sustained over such a large area inside the brittle volcanic deposits surrounding Ruapehu. However, because our data cannot rule out this possibility, we consider it a potential explanation. Both explanations involve the magma system of Ruapehu in connection with an eruption as a source for the changes in anisotropy, and the deduced assumptions about eruption forecasting are similar for both mechanisms.

We conducted a three-dimensional numerical stress calculation (*17*) to check quantitatively whether our proposed dike model can explain the required stress changes. Results show that stress changes inflicted by the proposed dike system can be strong enough to influence anisotropy. Other studies also find temporal and spatial variations in stress around active volcanoes, which suggests that favorable stress conditions may be common at other volcanoes. At Spurr Volcano, the direction of σ_H, as determined by focal mechanism analysis, changed by 90° (*34*) as a result of an inflating magma dike before the eruption. At Unzen Volcano (*35*), σ_H is spatially rotated by 90°, which indicates a local stress field with dimensions similar to those in our study, produced by pressurized volcanic gas or magma. At Vesuvius volcano, splitting parameters determined from a local earthquake swarm (*36*) show a slight increase in δt and minor variations of Φ before the time of the largest earthquake (M3.6) of the swarm, interpreted as a stress-change caused by the earthquake. All these findings and their interpretations are consistent with our proposed model. Because the observations from Mt. Spurr show a 90° rotation of the stress direction, they do not provide evidence for the 90°-flip model (*30*), which involves a rotation of Φ but not of the stress direction.

If the anisotropy changes recur before and after eruptions, they could be used for midterm forecasting of eruptions. Once the time intervals between changes and eruption, or the existence of a certain "stress threshold" before an eruption, have been established by further monitoring, predictions can be made for the onset of new eruptions. Achievable warning times could be months to a few years in advance, therefore possibly filling a gap in the available forecasting methods. Additionally, the changes between 1998 and 2002 suggest that the technique can be used to monitor real-time stress changes in and around magma chambers that are more subtle than those caused by a large eruption. For other areas in geophysics, the evidence presented in this paper suggests that renewed attempts at using anisotropy for stress monitoring associated with other activities, such as reservoir loading, mining, or even natural changes associated with earthquake activity, could be fruitful.

References and Notes

1. D. W. Peterson, *Studies in Geophysics: Active Tectonics* (National Academy Press, Washington, DC, 1986), pp. 2311–2346.
2. D. Dzurisin, *Rev. Geophys.* **41**, 1 (2003).
3. C. Bryan, S. Sherburn, *J. Volcanol. Geotherm. Res.* **90**, 1 (1999).
4. M. Nakagawa, K. Wada, T. Thordarson, C. P. Wood, J. A. Gamble, *Bull. Volcanol.* **61**, 15 (1999).
5. D. Johnston, D. Paton, B. Houghton, V. Neall, K. Ronan, *Bull. Geol. Soc. Am.* **112**, 720 (2000).
6. A. Nur, G. Simmons, *J. Geophys. Res.* **74**, 6667 (1969).
7. S. Crampin, D. C. Booth, *Geophys. J. R. Astron. Soc.* **87**, 75 (1985).
8. V. Babuska, M. Cara, *Seismic Anisotropy in the Earth* (Kluwer Acad., Norwell, Mass., 1991).
9. S. Crampin, S. Chastin, *Geophys. J. Int.* **155**, 221 (2003).
10. M. K. Savage, X. Shih, R. Meyer, R. Aster, *Tectonophysics* **165**, 279 (1989).
11. M. K. Savage, *Rev. Geophys.* **37**, 65 (1999).
12. Below a depth of a few hundred meters, the minimum stress is typically horizontal and therefore causes cracks with a vertical crack plane (*37*). This system yields a hexagonal or orthorhombic symmetry system with a horizontal symmetry axis. The fast direction is commonly observed parallel to σ_H.
13. V. Miller, M. Savage, *Science* **293**, 2231 (2001).
14. K. Gledhill, *J. Geophys. Res.* **96**, 21,503 (1991).
15. C. Munson, C. Thurber, Y. Li, P. Okubo, *J. Geophys. Res.* **100**, 20,367 (1995).
16. T.-C. Chen, *Geophys. J. R. Astron. Soc.* **91**, 287 (1987).
17. Materials and methods are available as supporting material on *Science* Online.
18. There are few earthquakes between 35 km and 55 km in depth in the region, leading to only two measurements. A complete list of all individual measurements of all deployments is available (*19*).
19. A. Gerst, thesis, Victoria University of Wellington, New Zealand (2003).
20. D. Darby, C. Meertens, *J. Geophys. Res.* **100**, 8221 (1995).
21. At the Far West T-Bar station (FWYZ), which usually shows strong scatter, this trend is visible when events at frequencies higher than 3.5 Hz are excluded (*19*).
22. Throughout this study, scatter of the delay times was more than 10 times higher than scatter of the fast directions. Further, the average delay times depend on the observed frequencies and depths, and thus on processing techniques (e.g., frequency filters) and earthquake magnitudes.
23. E. Audoine, M. Savage, K. Gledhill, *J. Geophys. Res,* in press.
24. T. Parsons, G. A. Thompson, *Science* **253**, 1399 (1991).
25. J. H. Latter, *J. Volcanol. Geotherm. Res.* **10**, 125 (1981).
26. W. R. Hackett, thesis, Victoria University of Wellington, New Zealand (1985).
27. I. A. Nairn, T. Kobayashi, M. Nakagawa, *J. Volcanol. Geotherm. Res.* **86**, 19 (1998).
28. Cracks that are aligned perpendicular to the new σ_H were forced to close, with their pore fluid migrating into cracks that are aligned parallel to the new σ_H (which were previously closed). Effectively, the alignment of the cracks adjusted to the new stress field and became parallel to the new σ_H.
29. S. V. Zatsepin, S. Crampin, *Geophys. J. Int.* **129**, 477 (1997).
30. S. Crampin, T. Volti, S. Chastin, A. Gudmundsson, R. Stefansson, *Geophys. J. Int.* **151**, F1 (2002).
31. Y. Liu, S. Crampin, I. Main, *Geophys. J. Int.* **130**, 771 (1997).
32. E. Angerer, S. Crampin, X. Y. Li, T. L. Davis, *Geophys. J. Int.* **149**, 267 (2001).
33. J. C. Zinke, M. D. Zoback, *Bull. Seismol. Soc. Am.* **90**, 1305 (2000).
34. D. Roman, S. Moran, J. Power, K. Cashman, *Bull. Seismol. Soc. Am.* (2004).
35. K. Umakoshi, H. Shimizu, N. Matsuwo, *J. Volcanol. Geotherm. Res.* **112**, 117 (2001).
36. E. D. Pezzo, F. Bianco, S. Petrosino, G. Saccorotti, *Bull. Seismol. Soc. Am.* **94**, 439 (2004).
37. S. Crampin, *Geophys. J. Int.* **118**, 428 (1994).

38. We thank Earthquake Commission, Marsden, Foundation for Research Science and Technology, Deutsches Akademisches Austausch Dienst (German Academic Exchange Service), and Planet Earth fund for funding this study. Thanks to J. Neuberg and G. Stuart from Leeds University for data and to Institute of Geological and Nuclear Sciences for providing instruments and resources. We also thank S. Hofmann, K. Gledhill, T. Hurst, M. Hagerty, J. Gamble, J. Townend, E. Smith, T. Stern, V. Miller, and F. Wenzel for invaluable help. Thanks to S. Toda and J. Park for help with their software. Maps were generated with Generic Mapping Tools by Wessel and Smith. Thanks to H. Keys and Department of Conservation for field support.

Supporting Online Material

www.sciencemag.org/cgi/content/full/306/5701/1543/DC1
Materials and Methods
Figs. S1 to S7
Tables S1 and S2
References

30 July 2004; accepted 14 October 2004

Evidence for Positive Epistasis in HIV-1

Sebastian Bonhoeffer,[1]* Colombe Chappey,[2] Neil T. Parkin,[2] Jeanette M. Whitcomb,[2] Christos J. Petropoulos[2]

Reproductive strategies such as sexual reproduction and recombination that involve the shuffling of parental genomes for the production of offspring are ubiquitous in nature. However, their evolutionary benefit remains unclear. Many theories have identified potential benefits, but progress is hampered by the scarcity of relevant data. One class of theories is based on the assumption that mutations affecting fitness exhibit negative epistasis. Retroviruses recombine frequently and thus provide a unique opportunity to test these theories. Using amino acid sequence data and fitness values from 9466 human immunodeficiency virus 1 (HIV-1) isolates, we find in contrast to these theories strong statistical evidence for a predominance of positive epistasis in HIV-1.

One of the most fundamental questions in biology is why sexual reproduction and recombination are so widespread. Meiotic recombination and sexual reproduction both induce the shuffling of parental genomes for the production of offspring. Intuitively, shuffling might appear to be beneficial, because it promotes genetic diversity among the offspring and thus allows for a faster rate of adaptation. However, this explanation has several shortcomings.

Downloaded from www.sciencemag.org on March 17, 2010

First, sexual reproduction and recombination do not always increase genetic variation. Second, even when they do, it is not clear why greater genetic variation should generally be adaptive. Third, recombination may not only create but also destroy favorable combinations of mutations. Therefore, what is the evolutionary benefit of shuffling two genotypes that are proven viable in their current environment?

Many theories have been proposed to account for the possible advantages of recombination [see reviews in (*1*–*4*)]. Currently favored are population genetic theories that are based on interactions between selection and genetic variation. These theories can be grouped into two classes: (i) those based on the effect of genetic drift in populations that are limited in size and (ii) those based on the effects of fitness interactions between alternative alleles at different loci (epistasis). Although there is growing belief that an explanation for the evolution of recombination could emerge from one or a combination of the existing theories, the key problem that has hampered progress is the limited amount of data that allows the testing of these theories.

The population genetic theories based on epistasis require a particular form of interaction for recombination to be advantageous. Beneficial mutations need to interact antagonistically (that is, they need to increase fitness less than multiplicatively with increasing number of mutations) and detrimental mutations need to act synergistically (they need to decrease fitness more than multiplicatively with increasing number of mutations) (*5*–*7*). In both cases, the interaction is characterized by negative epistasis, which is a measure of the deviation from multiplicativity of the fitness effect caused by the individual mutations. In a two-locus–two-allele model, epistasis can be defined as $E = w_{ab} + w_{AB} - w_{aB} - w_{Ab}$, where *a*/*A* and *b*/*B* are the alternative alleles at the two loci and w_{**} is the log fitness of the corresponding genotype.

The appeal of theories about the evolution of recombination based on negative epistasis (in particular between deleterious mutations) is that recombination can efficiently eliminate deleterious mutations under these circumstances. Provided that the deleterious mutation rate is sufficiently high, this could be a strong selective advantage for all living organisms, and hence these theories could account for the ubiquity of recombination and sexual reproduction in nature (*6*, *8*). Many studies have tested for epistasis [reviewed in (*9*)]. However, no clear evidence has yet been found for an overall predominance of positive or negative epistasis (*10*–*16*), probably because of limited statistical power. Moreover, few of these studies have been carried out in organisms with high rates of sexual recombination, and often epistasis was measured with regard to fitness-related properties rather than total fitness (*15*, *17*). Only one recent study of RNA bacteriophages inferred positive epistasis by accumulation of mutations in lineages with different fitness (*18*).

Retroviruses such as human immunodeficiency virus 1 (HIV-1) provide a unique opportunity to test theories about the evolution of recombination, because their replication cycle can be viewed as a primitive form of sexual replication (*19*). Retroviruses package two full-length copies of the RNA genome. After infection of a cell, the viral polymerase [reverse transcriptase (RT)] engages one copy of the genomic RNA and converts the sequence into proviral DNA. During this process, the RT carrying the nascent DNA provirus may disengage from the first RNA template and reengage the second. If the infecting virion is heterozygous (that is, if it carries two distinct genomic RNA strands), this process of template switching may lead to the production of a recombinant provirus. Heterozygous virions are produced when cells are simultaneously infected by two or more distinct proviruses, which occurs frequently in HIV-1 infection (*20*). Recent estimates suggest that in HIV-1, the RT switches RNA templates approximately 10 times per replication cycle (*21*), yielding an average of about one recombination event per 1000 base pairs.

To test whether negative epistasis predominates in HIV-1, we analyzed 9466 virus samples derived from HIV-1–infected patients for routine drug-resistance testing. Viral fitness was assayed based on the construction of HIV-derived test vectors (*22*). In essence, patient virus–derived amplicons representing all of protease (PR) and most of RT are inserted into the backbone of a resistance test vector. This vector is based on the NL4-3 molecular HIV clone and has been modified so that it can undergo only a single round of replication. The fitness assay then quantifies the total production of infectious progeny virus after a single round of infection of the patient-derived virus relative to that of an NL4-3–based control virus. The fitness of the NL4-3–based control virus thus equals 1. Note that although most mutations in our data set evolved in patients in the context of drug therapy, fitness is measured in vitro in the absence of drugs. In addition to the fitness measurement, all of PR and amino acids 1 to 305 of RT were sequenced by population sequencing.

The distribution of log fitness values ranges over three orders of magnitude and has a long tail extending to low fitness values (Fig. 1A), which likely reflects the fact that mutations acquired during drug therapy are often detrimental in the absence of drugs. One commonly used method to test for epistasis is to plot the logarithm of fitness as a function of the number of mutations relative to a reference strain (*10*, *11*). If there is no epistasis, the fitness effects of the individual mutations contribute multiplicatively to fitness, and thus log fitness decreases lin-

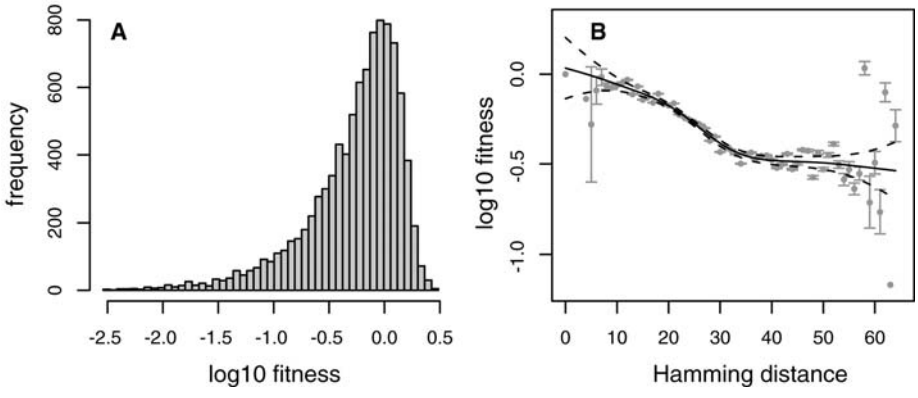

Fig. 1. (**A**) Distribution of relative fitness values of all 9466 sequences included in the data set. (We use $\log_{10}$-transformed data throughout.) The fitness is based on a recombinant virus assay (*22*), which measures the total production of infectious progeny virus after one complete round of replication relative to a reference control virus. (**B**) Mean and standard error (gray dots and bars) of log fitness as a function of the number of amino acids differing from the reference virus (Hamming distance) for all sequences in the data set. The lines represent fitted values (solid lines) and 95% confidence intervals (dashed lines) of a nonparametric regression based on cubic splines (using the implementation of generalized additive models in the R statistical software package) (*28*). For small (<10) and large (>50) Hamming distances, the large standard errors are due to the small number of sequences in each Hamming distance class. Missing error bars indicate that there is only one sequence in this Hamming distance class. In the intermediate range of Hamming distances (10 to 50), standard errors are low because all classes are represented by between 36 and 498 sequences.

[1]Ecology and Evolution, ETH Zurich, ETH Zentrum NW, CH-8092 Zurich, Switzerland. [2]ViroLogic, 345 Oyster Point Boulevard, South San Francisco, CA 94080–1913, USA.

*To whom correspondence should be addressed. E-mail: seb@env.ethz.ch

Downloaded from www.sciencemag.org on March 17, 2010

early with increasing number of mutations. A less than linear decrease suggests the predominance of positive epistasis and a more than linear decrease suggests negative epistasis. Figure 1B shows the mean and standard error of log fitness as a function of the number of amino acid differences compared to the NL4-3 reference virus (the Hamming distance). Because there are limited numbers of sequences in the data set with Hamming distances smaller than 10 or larger than 50, the fluctuations in mean and standard error are large in these Hamming ranges. In the intermediate range, however, the standard errors are small because of the large number of sequences in each Hamming distance class. The 95% confidence interval of a nonparametric fit shows that there is a highly significant deviation from linearity for large Hamming distances, which is suggestive of positive epistasis.

Alternatively, the less than linear decrease could be explained by a bias in the data set against sequences with very low fitness. Such a general bias is unlikely in our data set, because these viral sequences are generally derived from patients receiving drug therapy, whereas fitness is measured in the absence of drugs. Mutations that confer high fitness in the presence of drugs therefore have an increased probability of existing in our data set. However, these mutations are included in the data set irrespective of their effect on fitness in the absence of drugs, and it is therefore unlikely that they are biased against low fitness in our assay. Nevertheless, for very large Hamming distance it is possible that the less than linear decrease of log fitness is a result of the exhaustion of those mutants that have high fitness in the presence but low fitness in the absence of drugs, and an underrepresentation of sequences with low fitness in both environments.

A more direct test for epistasis is to measure fitness interactions between pairs of alternative amino acids at different positions in the aligned sequence set. Assume we have two alternative amino acids a and A at position i and b and B at position j. Provided that there are sequences in the data set for all four possible combinations of the two amino acids at both positions, we estimate epistasis as $E = \overline{w}_{ab} + \overline{w}_{AB} - \overline{w}_{aB} - \overline{w}_{Ab}$, where $\overline{w}_{ab}$ denotes the mean log fitness of all sequences that have amino acid a at position i and amino acid b at position j in the aligned sequence set. The definition of epistasis requires the selection of a reference genotype. That is, it is necessary to define which of the four genotypes represents ab, because otherwise the sign of E is arbitrary. Commonly, ab is defined as the fittest genotype. However, if there is measurement error in fitness, it can be shown that defining ab as the fittest genotype results in a bias toward positive epistasis. Similarly, defining AB as the least fit genotype results in a bias toward negative epistasis. Measuring epistasis with regard to a predefined genotype does not result in a bias. Therefore we choose as the reference genotype the amino acid combination found in the NL4-3 strain, because fitness is measured relative to this strain. (Repeating the analyses below with the consensus sequence as the reference sequence had negligible effects, because the NL4-3 strain and the consensus sequence differ in only six amino acid positions.)

A computer program identified 103,286 pairs of alternative amino acids in the data set, where one genotype represented the amino acid combination found in the NL4-3 strain. The distribution of epistasis values between these pairs extends to both positive and negative values of epistasis (Fig. 2A). The mean of the distribution is 0.052, and 60% of all pairs have positive epistasis. Clearly, the measurements of epistasis are not all independent, because pairs of mutations may be linked to each other. To test whether the observed mean epistasis is significantly different from zero, we randomized the association between sequence and corresponding fitness value and repeated the analysis 100 times. This randomization procedure revealed that the difference in mean epistasis between the real and randomized data is highly significant (Fig. 2B). The observation that the mean epistasis for the randomized data set did not differ significantly from 0 (t test, $P = 0.5$) shows that measuring epistasis with regard to a predefined genotype does not introduce a bias toward nonzero epistasis.

To test whether the observed positive mean epistasis resulted from random measurement errors, we restricted the analysis only to those positions at which we were able to detect amino acid variants that significantly affect fitness (*22*). A comparison of the 59 identified positions with data downloaded from the Stanford HIV Drug Resistance Database (http://hivdb.stanford.edu/) showed that all identified sites exhibit known drug treatment–associated polymorphisms (*22*). Repeating the epistasis analysis restricted to the selected 59 positions, we obtained a mean epistasis of 0.109 (Fig. 2C). Restricting the analysis to positions with significant fitness effects thus shifted the mean epistasis to a more positive value.

These analyses provide strong statistical evidence for the predominance of positive epistasis in the genome of an organism undergoing frequent recombination and thus challenge hypotheses about the evolution of recombination based on negative epistasis. Our findings relate to the larger context of the evolution of recombination as follows. First, as in most other studies, fitness is not measured in the natural environment. However, we see no reason why the in vitro estimates of epistasis should be shifted toward positive values in comparison to the natural environment. Both PR and RT activity are essential for viral replication, and mutations affecting these viral enzymes probably have similar effects in vivo and in vitro. Second, our analysis is restricted to mutations in PR and RT. The genomic region surveyed constitutes around 15% of the coding HIV-1 genome. Because the prevailing form of epistasis may depend on the biological process that is affected (*23*), it is conceivable that positive epistasis prevails only among mutations in PR and RT. Whether it also prevails in other regions of the HIV genome will require additional investigation. The predominance of positive epistasis among drug resistance mu-

Downloaded from www.sciencemag.org on March 17, 2010

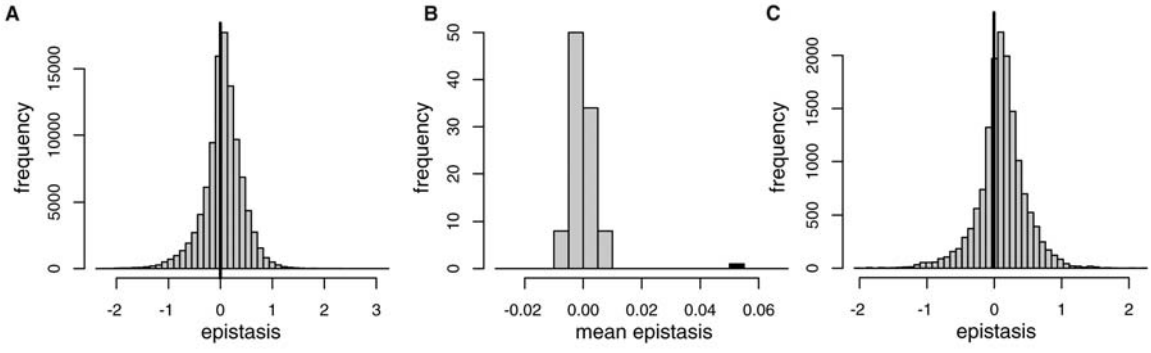

Fig. 2. (**A**) Distribution of epistasis values between all possible pairs of alternative amino acids across the aligned sequence set (n = 103,286). The solid vertical line indicates zero epistasis. To test whether the mean of this distribution (0.052) is significantly different from zero, we repeated the analysis 100 times based on randomized data sets. (**B**) The mean epistasis observed for the real data (black bar) is highly significantly different from the mean epistasis for the 100 randomized data sets (gray bars). (**C**) Distribution of epistasis values when the analysis is restricted to positions that have a highly significant effect on fitness (*22*). Restricting the analysis to these 59 positions shifts the distribution toward more positive values (mean = 0.109).

tations may have implications for drug therapy, because computer simulations suggest that under these conditions, recombination impairs rather then facilitates the evolution of drug resistance in HIV (*24*). Third, it is unclear whether the beneficial effects of recombination for negative epistasis are negated by detrimental effects reflecting an equal degree of positive epistasis. However, approximate calculations assuming weak linkage between the locus coding for recombination rate and the loci under selection suggest that variation in the epistatic interactions weakens the selection for recombination and that recombination is selected against for positive mean epistasis (*25*). Finally, it remains to be shown whether the pattern of epistasis found in HIV-1 is representative of that in other organisms (in particular of that in eukaryotes). However, in contrast to studies measuring epistasis in *Escherichia coli*, yeast, or *Caenorhabditis elegans* (*10*, *13*, *15*), we have measured epistasis in an organism in which recombination occurs frequently. Therefore, it may in fact be more appropriate to extrapolate from retroviruses than from organisms that are effectively asexual.

The predominance of interactions with positive epistasis in HIV-1 raises the question of why retroviruses have evolved the capacity to recombine (*24*). Whether drift-based explanations for the benefits of recombination are applicable to HIV-1 remains to be investigated. Recent studies suggest that drift can favor the evolution of recombination even for very large populations, as long as there are a sufficient number of loci under selection (*26*). An alternative explanation is the repair of single-strand breaks (*27*). Retroviruses have single-stranded RNA genomes that are susceptible to ribonucleases and other agents during the viral life cycle. Template switching by RT in retroviruses could extend transcription beyond breakage points and could thus explain why retroviruses evolved to carry two complete genomes. According to this hypothesis, however, recombination in retroviruses would be the consequence but not the cause of the evolution of template switching.

Note added in proof: After this paper had been accepted, a related paper or epistasis in the vesicular stomatitis virus by Sanjuán *et al.* appeared (*29*).

References and Notes

1. J. Maynard Smith, Ed., *The Evolution of Sex* (Cambridge Univ. Press, Cambridge, 1976).
2. R. E. Michod, B. R. Levin, Eds., *The Evolution of Sex* (Sinauer Press, Sunderland, MA, 1988).
3. N. H. Barton, B. Charlesworth, *Science* **281**, 1986 (1998).
4. S. P. Otto, T. Lenormand, *Nature Rev. Genet.* **3**, 252 (2002).
5. M. W. Feldman, F. B. Christiansen, L. D. Brooks, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 4838 (1980).
6. A. S. Kondrashov, *Nature* **336**, 435 (1988).
7. N. H. Barton, *Genet. Res.* **65**, 123 (1995).
8. A. S. Kondrashov, *J. Hered.* **84**, 372 (1993).
9. W. R. Rice, *Nature Rev. Genet.* **3**, 241 (2002).
10. S. F. Elena, R. E. Lenski, *Nature* **390**, 395 (1997).
11. S. F. Elena, *J. Mol. Evol.* **49**, 703 (1999).
12. M. de la Pena, S. F. Elena, A. Moya, *Evolution* **54**, 686 (2000).
13. A. D. Peters, P. D. Keightley, *Genetics* **156**, 1635 (2000).
14. M. C. Whitlock, D. Bourguet, *Evolution* **54**, 1654 (2000).
15. D. M. Wloch, K. Szafraniec, R. H. Borts, R. Korona, *Genetics* **159**, 441 (2001).
16. C. L. Burch, P. E. Turner, K. A. Hanley, *J. Evol. Biol.* **16**, 1223 (2003).
17. A. Rivero, F. Balloux, S. A. West, *Evolution* **57**, 1698 (2003).
18. C. L. Burch, L. Chao, *Genetics* **167**, 559 (2004).
19. H. M. Temin, *Trends Genet.* **7**, 71 (1991).
20. A. Jung *et al.*, *Nature* **418**, 144 (2002).
21. D. N. Levy, G. M. Aldrovandi, O. Kutsch, G. M. Shaw, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 4204 (2004).
22. Materials and methods are available as supporting material on *Science* Online.
23. E. Szathmary, *Genetics* **133**, 127 (1993).
24. M. T. Bretscher, C. L. Althaus, V. Müller, S. Bonhoeffer, *BioEssays* **26**, 180 (2004).
25. S. P. Otto, M. W. Feldman, *Theor. Popul. Biol.* **51**, 134 (1997).
26. M. M. Iles, K. Walters, C. Cannings, *Genetics* **165**, 2249 (2003).
27. J. M. Coffin, *J. Gen. Virol.* **42**, 1 (1979).
28. The R Project for Statistical Computing (www.r-project.org).
29. R. Sanjuán, A. Moya, S. F. Elena, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 15376 (2004).

30. We acknowledge the ViroLogic Clinical Reference Laboratory for their efforts in generating the replication capacity and PR/RT sequence data and C. Althaus, V. Müller, S. Otto, T. Pfeiffer, and M. Salathé for valuable comments. S.B. acknowledges the Swiss National Science Foundation for financial support. The development of the drug resistance database and replication capacity assay used in this study was funded in part by Small Business Innovative Research–Advanced Technology Grants from the National Institute of Allergy and Infectious Diseases, NIH (R43 AI057068 and R43 AI050321). S.B. is an equity holder in ViroLogic, Inc.

Supporting Online Material

www.sciencemag.org/cgi/content/full/306/5701/1547/DC1
Materials and Methods
References

22 June 2004; accepted 14 October 2004

Femtomolar Sensitivity of a NO Sensor from *Clostridium botulinum*

Pierre Nioche,[1] Vladimir Berka,[2] Julia Vipond,[3] Nigel Minton,[4] Ah-Lim Tsai,[2] C. S. Raman[1]*

Nitric oxide (NO) is extremely toxic to *Clostridium botulinum*, but its molecular targets are unknown. Here, we identify a heme protein sensor (SONO) that displays femtomolar affinity for NO. The crystal structure of the SONO heme domain reveals a previously undescribed fold and a strategically placed tyrosine residue that modulates heme-nitrosyl coordination. Furthermore, the domain architecture of a SONO ortholog cloned from *Chlamydomonas reinhardtii* indicates that NO signaling through cyclic guanosine monophosphate arose before the origin of multicellular eukaryotes. Our findings have broad implications for understanding bacterial responses to NO, as well as for the activation of mammalian NO-sensitive guanylyl cyclase.

Nitric oxide (NO) is a small, short-lived, and highly reactive gaseous molecule. In mammals, NO is biosynthesized from L-arginine by nitric oxide synthases, and it plays a key role in many and disparate cellular responses including host defense against microbial pathogens (*1*). Denitrifying bacteria generate NO (1 to 70 nM) by a unique mode of respiration in which nitrogen oxides (NO_3^-, NO_2^-, NO, and N_2O) are reduced to N_2 (*2*). Nondenitrifiers, like *Escherichia coli*, can produce NO during anaerobic nitrate respiration ($NO_3^- \rightarrow NO_2^- \rightarrow NH_4^+$), although at much reduced levels (*3*). *Clostridium botulinum*, a strict anaerobe, is neither a denitrifier nor capable of making NO via other mechanisms. It is the etiological agent of botulism and produces the most toxic substance known to humans [median lethal dose (LD_{50}) $\approx$ 0.2 ng per kilogram of body weight]. Since the late 1920s sodium nitrite, with NO as the antimicrobial principle, has been used to inhibit the growth of heat-resistant *C. botulinum* spores and toxin production in cured meats (*4*–*6*). However, the molecular strategies used by this bacterium to recognize and avoid NO in its native and host environments have remained elusive.

To identify candidate NO sensors in *C. botulinum*, we investigated the hypothesis (*7*, *8*) that a prokaryotic counterpart to the mammalian NO receptor, soluble guanylyl cyclase (sGC) (*9*, *10*), exists and that a bacterial NO sensor may communicate with the chemotaxis machinery. Thus, we screened the genome of *C. botulinum* (www.sanger.ac.uk/Projects/C_botulinum/) for an ortholog

[1]Structural Biology Research Center and Department of Biochemistry and Molecular Biology, [2]Division of Hematology, Internal Medicine, University of Texas Medical School, Houston, TX 77030, USA. [3]Health Protection Agency, Porton Down, Salisbury, Wiltshire SP4 0JG, UK. [4]Center of Biomolecular Sciences and Institute of Infection, Immunity and Inflammation, University of Nottingham, Nottingham NG7 2RD, UK.

*To whom correspondence should be addressed: E-mail: c.s.raman@uth.tmc.edu

Downloaded from www.sciencemag.org on March 17, 2010

ERRATUM

Post date 9 May 2008

Reports: "Evidence for positive epistasis in HIV-1" by S. Bonhoeffer *et al.* (26 November 2004, p. 1547). The HIV sequence studies in this paper were not deposited in GenBank at the time of publication but are now available there. The accession numbers are EU606356 to EU615822.

Downloaded from www.sciencemag.org on March 17, 2010