a KCl concentration gradient from 100 mM to 1 M in the lysis buffer. Human ESP1 was eluted at about 250 mM KCl. Fractions containing hESP1 were pooled and loaded onto a Superdex 200 column.
34. G. Fang, H. Yu, M. W. Kirschner, *Mol. Cell* **2**, 163 (1998).
35. We thank C. Pfleger and S. Rankin for comments on the manuscript; G. Fang, H. Yu, and other members of the Kirschner laboratory for helpful discussion and technical assistance; C. Pfleger and E. Lee for providing modified pCS2 and pET28a vectors; and Q. Zang for mammalian cell transfection. H.Z. is a fellow of The Jane Coffin Childs Memorial Fund for Medical Research. T.J.M. was supported by a Physician Postdoctoral Grant from the Howard Hughes Medical Institute and a Mentored Clinical Scientist Development Award from the National Heart, Lung, and Blood Institute. Supported by grants GM39023 and GM26875 from NIH awarded to M.W.K.

31 March 1999; accepted 18 May 1999

Different Trajectories of Parallel Evolution During Viral Adaptation

H. A. Wichman,[1]* M. R. Badgett,[2] L. A. Scott,[1] C. M. Boulianne,[1] J. J. Bull[2,3]

The molecular basis of adaptation is a major focus of evolutionary biology, yet the dynamic process of adaptation has been explored only piecemeal. Experimental evolution of two bacteriophage lines under strong selection led to over a dozen nucleotide changes genomewide in each replicate. At least 96 percent of the amino acid substitutions appeared to be adaptive, and half the changes in one line also occurred in the other. However, the order of these changes differed between replicates, and parallel substitutions did not reflect the changes with the largest beneficial effects or indicate a common trajectory of adaptation.

The wealth of molecular data now available reveals that genetic variation is virtually ubiquitous not only between species but also within species. Yet the extent to which this variation is adaptive is enigmatic. This question has taken on new importance as molecular data find increasing use to deduce the mechanism of drug resistance, monitor pathogen populations for incidence of resistance, and compare molecular responses to different drugs and treatment regimes. DNA sequence comparisons are commonly used to address this question, because they carry legacies of their histories that may reflect adaptive change. For example, historical occurrences of adaptive evolution can sometimes be recognized by the high rate of nonsynonymous to synonymous substitution in coding regions, which is a statistical landmark adaptive change (*1*–*4*). However, for two reasons, this statistical approach may be less useful for looking at contemporary adaptation of pests and pathogens. First, adaptive responses to strong selection may require too few changes at the molecular level to leave such "statistical tracks." Second, these methods reveal the occurrence but not the identity of adaptive changes. Parallel evolution—the same change having evolved repeatedly and independently—is also regarded as evidence of adaptation and can reveal the identity of at least some adaptive amino acid substitutions (*5*–*8*). In fact, parallel evolution is a standard criterion used to identify the amino acid substitutions responsible for drug resistance. Although some believe that organisms have too many degrees of freedom to allow this type of predictability, it is not out of the question that the short-term course of adaptation in response to specific selective agents can be predicted, and the short-term course is the one most relevant to human health and infectious disease. It is thus important to determine the extent to which parallel evolution allows us to predict the mechanisms and dynamics of evolution. Here we assess the genomewide dynamics of evolution to examine the molecular basis for and process of adaptation in a viral population under strong selection.

The single-stranded DNA bacteriophage ϕX174 was adapted to high temperature and a novel host (*Salmonella typhimurium*) in a two-stage chemostat for about 1000 population doublings over 10 days. Methods were as described (*9*) except that the temperature was 43.5°C throughout the 10 days of selection.

To guard against contamination, which could be misinterpreted as parallel evolution, we grew replicate lineages in our geographically separated laboratories; we refer to them as the TX and ID replicates. The chemostats were sampled every 24 hours, and isolates from these daily samples were archived into microtiter plates for later analysis. Population sizes in the chemostats were typically in excess of 10^7 except during the first few days of the TX chemostat, when population sizes were closer to 10^4. Our goal was to observe the process of adaptation in a very strong selective environment rather than to dissect responses to specific conditions in the environment. These conditions—large population size and strong selection—are highly favorable to adaptation.

In previous experiments (*9*), we found that adaptation to these conditions resulted in over a dozen substitutions genomewide, with 25 to 50% of them occurring in parallel between any two replicates. On the basis of this parallel evolution alone, we concluded that substitutions were adaptive at more than one-third of the sites where differences were detected. However, the underlying dynamics of evolution and the extent of genetic differentiation within populations was unknown. At one extreme, substitutions might quickly sweep through the population to fixation, so that most genotypes in the population are very similar at any point in time; at the other extreme, there could be multiple competing genotypes present throughout the history of the population. Furthermore, it is not known to what extent high levels of parallel evolution are a signature of similarity of adaptation at the phenotypic level.

Adaptation to chemostat conditions was evident by massive improvements in phage growth rates at the high temperature (Fig. 1). Phage population growth rates were measured as doublings of phage concentration per hour at 43°C under defined conditions (*9*); this assay measures a major component of fitness in the chemostat but does not measure all relevant fitness components. The population growth rate of the ID replicate increased from −5 to 7.1 over the course of the 10 days of selection, with no gain detectable after day 4. This corresponds to a 4000-fold improvement in the number of descendants per hour. The population growth rate of the TX replicate increased from −5 to 12.5 (about 18,000-fold) over the course of the 10 days of selection, with major improvements at several time points. A difference in correlated response to selection was also detected: ID lost its ability to plate on *Escherichia coli* C, whereas TX retained this ability. From this fitness evidence alone, it is obvious that both replicates accumulated changes but that they followed at least somewhat different pathways.

The genetic basis of adaptation was studied at both a nucleotide level and a population level over time. To identify substitutions at high frequency at the end of selection, we obtained complete genome sequences from polymerase chain reaction products amplified

[1]Department of Biological Sciences, University of Idaho, Moscow, ID 83844, USA. [2]Department of Zoology and [3]Institute of Cellular and Molecular Biology, University of Texas, Austin, TX 78712, USA.

*To whom correspondence should be addressed. E-mail: hwichman@uidaho.edu

from aliquots of the day 10 populations. This yields a consensus genotype for the population. Additionally, we sequenced two isolates from each day 10 population completely, revealing two substitutions at low frequency in ID. We then studied the population frequencies of all detected changes by oligonucleotide hybridization of plaque isolates archived daily (*10*). For each substitution, we designed separate oligonucleotides to detect the evolved and the ancestral states. We screened eight isolates from each daily sample to identify when each substitution appeared, and we screened 96 isolates per day across the period of polymorphism.

Fourteen nucleotide substitutions and one intergenic deletion were detected in the TX replicate; 13 substitutions and 1 intergenic deletion were detected in the ID replicate (Table 1). The average fitness increase per substitution is thus about one doubling per hour, which indicates that at least some of the changes carried profound advantages. If we assume a generation time of 20 min, this translates into a relative fitness of about 1.26 per substitution, which is very large in evolutionary terms. These substitutions occurred in 6 of the 11 ϕX174 genes (*11*) in TX (A, A*, C, E, F, and H) and in 4 genes in ID (A, A*, F, and H). Consistent with the other evidence of adaptation in this system, all 14 of the TX substitutions, and 11 of the 13 ID substitutions were nonsynonymous.

Additional information must be used to determine which changes are adaptive. One powerful indication of selection is parallel evolution. About 50% of the changes occurred in parallel between both replicates: six substitutions (F101, F177, H79, A44, A63, and A462) and the 27-base deletion. Furthermore, of the 29 changes in these two replicates, only 5 have not been observed in other lineages we have propagated (*12*): E72, F361, H15, and the silent changes at positions 1300 and 4637. (These 5 changes are referred to below as unique substitutions.) Equally striking, 4 of our 19 different amino acid substitutions are convergent with the closely related bacteriophage S13. S13 was originally isolated on *Salmonella* (*13*) and differs from ϕX174 at 113 sites, including 30 nonsynonymous differences (*14, 15*). One of our four convergences with S13 is the unique substitution F361, suggesting that it, too, is adaptive. This was further supported when selection of S13 on *E. coli* C resulted in convergence with the ϕX174 ancestral sequence at this site (data not shown).

The adaptive value of a change may also be assessed from the dynamics of its ascent in a population. Substitutions occurred as a series of rapid sweeps through both populations (Fig. 2). Genetic drift is not a plausible mechanism for these rapid increases in so large a population, but genetic hitchhiking could allow one beneficial change to drag other, nonadaptive changes through the population. Substitutions that increased in frequency independently of all others can be considered adaptive. Dynamics reveal the adaptive nature of 8 of the 29 changes, including one of the unique changes (E72). In contrast, one of the silent changes in the ID replicate (1300) appeared and swept through in strict coupling with the change at 1460 (F153), which is consistent with genetic hitchhiking of a neutral change. From the combined evidence of parallelism and dynamics, only the silent substitutions and H15 remain as candidate nonadaptive substitutions. Thus, in each replicate, at least 86% of the substitutions appear to be adaptive.

In contrast to the patterns seen for population growth rate, the overall picture of molecular dynamics is remarkably similar between the two replicates, except that the onset of change was delayed somewhat in TX. In both populations, changes swept through rapidly and, except during the initial 2 days of TX, they accumulated throughout the entire 10-day period. Chemostat populations were successively dominated by the new genotypes as changes appeared, instead of generating a stable coexistence of different, complex genotypes as observed in some bacterial chemostats (*16*). Substitutions in F were among the first changes in both replicates and coincided with large increases in population growth rate; additional substitutions in F also arose late in both populations. In two cases, we detected substitutions that entered and were soon lost from the population (C64 in TX; F204 in ID). Other substitutions swept through the population as these changes dropped out. These two losses were detected serendipitously; other losses may have occurred but went undetected.

The high degree of parallel evolution and the superficially similar dynamics suggest that many similarities underlie the evolutionary processes in both replicates. Yet there are surprising differences. The order of appearance of the seven parallel changes differs between TX and ID (*17*). By comparison, the accumulation of mutations in the human immunodeficiency virus (HIV-1) reverse transcriptase or protease gene has been studied under a variety of drug treatments. Parallel changes are frequent, and often there is a

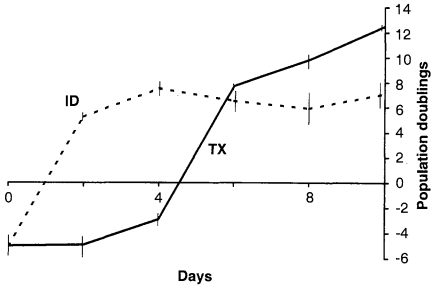

Fig. 1. Phage population growth rate at 43°C. Phage population growth rates were measured as doublings of phage concentration per hour at 43°C under defined conditions (9), except that, in contrast to this earlier assay, aeration was achieved by bubbling to closely match the chemostat environment. The TX replicate is shown as a solid line; the ID replicate is shown as a dashed line. Bars represent one standard error above and below the mean. Each fitness assay included all 11 of the cultures shown here. The values for each assay were standardized by adding a common correction factor to render the mean equal between assays performed at different times. This standardization affects only the variances and not the means.

Table 1. Molecular changes in ϕX174 during adaptation to high temperature and a novel host. Changes are listed in the order in which they appeared in the population. The nucleotide position in the published sequence (GenBank accession no. J02482) is listed, followed by the protein affected, with the number of the amino acid residue, and the nature of the amino acid substitution where applicable (*33*). Protein A functions in viral-strand synthesis; A* shuts off host DNA synthesis; C functions in DNA maturation; E is a lysis protein; F is the major coat protein; H is the minor spike or pilot protein. Parallel changes are shown in boldface, and superscript indicates the order of those changes in the other replicate. Order corresponds to the numbering of changes shown in Fig. 2.

Order	TX site	Amino acid change	ID site	Amino acid change
1	782	E72, T→ I	2167	F388, H→Q
2	1727	F242, L→ F	1613	F204, T→ S
3	2085	F361, A→V	**1533**^{6}	F177, T→ I
4	319	C63, V→ F	1460	F153, Q→ E
5	2973	H15, G→ S	1300	F99, silent
6	323	C64, D→G	**1305**^{3}	F101, G→D
7	**4110**^{3}	A44, H→ Y	1308	F102, Y→C
8	1025	F8, E→ K	**4110**^{1}	A44, H→ Y
9	**3166**^{7}	H79, A→V	4637	A219, silent
10	5185	A402, T→M	**965-91**^{4}	Δ
11	**1305**^{2}	F101, G→D	**5365**^{5}	A462, M→ T
12	**965-91**^{4}	Δ	**4168**^{7}	A63, Q→ R
13	**5365**^{5}	A462, M→ T	**3166**^{2}	H79, A→V
14	**1533**^{1}	F177, T→ I	1809	F269, K→ R
15	**4168**^{6}	A63, Q→ R		

common order to the appearance of these changes (*18, 19*). Substitutions may confer increasing levels of drug resistance per se or may compensate for deleterious pleiotropic effects of earlier resistance mutations (*20*). Compensatory substitutions will, by definition, follow the changes that directly confer resistance, thus giving rise to an order effect. A common order is also expected in large populations on the grounds that mutations with the largest beneficial effect sweep through before those of lesser effect (*21*). However, observation of the evolution of HIV suggests that the effective population size of viruses within a person appears to be orders of magnitude smaller than the total number of viruses, which may account for the lack of order in some treatments (*22*).

The fact that a common order did not prevail in our lines suggests that these parallel changes may provide a benefit in the chemostat that accrues over a wide range of genotypic backgrounds—that the sign of fitness effect (if not the magnitude) is stable across many genotypes. The earliest changes in F coincided with large initial increases in population growth rate in both replicates, but another difference between ID and TX was that the changes in F that underlie these initial fitness increases were not the same. When there are multiple pathways of adaptation, as appears to be the case here, the beneficial effect of a particular substitution may be greater when the virus suffers low fitness than after much fitness has been recovered (*23*).

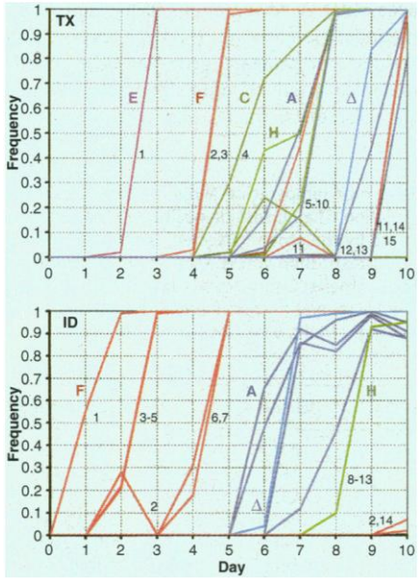

Fig. 2. Population dynamics of molecular evolution during adaptation of ϕX174 to a novel host and high temperature. The frequency of each substitution is shown daily for the TX and ID lineages over the 10-day period of selection. Numbering of changes to the order shown in Table 1. Line and gene designations are color coded; Δ is the 27-base intergenic deletion from 965 to 991.

This model would explain why different early changes cause the largest boosts in fitness and also may explain why mutations conferring a high fitness in one replicate do not always appear in other replicates. Many later substitutions fixed without obvious increase in phage growth rate; their benefit may lie in fitness components not measured by growth rate or may just lie outside the scale of detection (*24*).

Most parallel substitutions did not occur in all five selections carried out under the same experimental conditions (*25*). Why is parallel evolution not complete? Given the population size in the chemostat, all these substitutions would have arisen multiple times during the course of these selections. This variation among replicates suggests that stochastic features, such as the identity of the earliest change to sweep through the population and the order in which substitutions arise, may influence the pattern of adaptation even in systems where parallelism is the rule rather than the exception. And, as in the case of HIV (*22*), the effective population size may be much smaller than the total number of phage in the chemostat at any point in time.

There is considerable precedent for experimental studies of microbial adaptation with patterns of fitness change used to infer the underlying genetic changes (*16, 26–31*). Our work extends those earlier studies by identifying the nucleotide changes genomewide during adaptation. From our observations of parallel evolution and the dynamics of substitutions sweeping through the populations, we were able to conclude that all but 1 of the 25 amino acid substitutions seen in these populations was adaptive. Although our analysis would not have detected most changes that appeared and that were ultimately lost, we nonetheless have observed a complexity that would defy analysis at only the fitness level. Further studies are needed to determine how such patterns are affected by rates of recombination (*32*), mutation, and environmental change.

References and Notes

1. J. H. McDonald and M. Kreitman, *Nature* **351**, 652 (1991).
2. M. F. Taylor, Y. Shen, M. E. Kreitman, *Science* **270**, 1497 (1995).
3. W.-H. Li, *Molecular Evolution* (Sinauer Associates, Sunderland, MA, 1997).
4. E. F. Boyd and D. L. Hartl, *J. Mol. Evol.* **47**, 258 (1998).
5. C. B. Stewart, J. W. Schilling, A. C. Wilson, *Nature* **330**, 401 (1987).
6. A. Molla *et al.*, *Nature Med.* **2**, 760 (1996).
7. W. Messier and C. B. Stewart, *Nature* **385**, 151 (1997).
8. M. M. Canica, C. Y. Lu, R. Krishnamoorthy, G. C. Paul, *J. Mol. Evol.* **44**, 57 (1997).
9. J. J. Bull *et al.*, *Genetics* **147**, 1497 (1997).
10. For oligonucleotide hybridization to detect single-base differences, we followed the TMAC protocol from Ausubel *et al.* [F. M. Ausubel *et al.*, Eds., *Current Protocols in Molecular Biology* (Green Publications/Wiley-Interscience, New York, 1989), pp. 6.4.3–6.4.10].
11. M. Hayashi, A. Aoyama, D. L. R. Jr., M. N. Hayashi, in *The Bacteriophages*, R. Calendar, Ed. (Plenum Press, New York, 1988), vol. 2, pp. 1–71.
12. To date we have carried out over 20 selection experiments with ϕX174 on *S. typhimurium* or *E. coli* C and have seen changes at a total of 2% of sites in the genome; 9 of these lineages were reported in (*9*).
13. J. H. Spencer *et al.*, in *The Single-Stranded DNA Phages*, D. T. Denhardt, D. Dressler, D. S. Ray, Eds. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1978), pp. 87–101.
14. F. Sanger *et al.*, *Nature* **265**, 687 (1977).
15. P. C. Lau and J. H. Spencer, *Gene* **40**, 273 (1985).
16. R. F. Rosenzweig, R. R. Sharp, D. S. Treves, J. Adams, *Genetics* **137**, 903 (1994).
17. Substitutions in F were the earliest parallel changes in ID but these changes did not sweep through until day 10 in TX. The common change in H was the second parallel change in TX but the last one in ID. The three changes in gene A occurred in the same order but were not consecutive. Common order of three changes is expected to occur by chance one in six times.
18. C. A. Boucher *et al.*, *J. Infect. Dis.* **165**, 105 (1992).
19. P. Kellam, C. A. Boucher, J. M. Tijnagel, B. A. Larder, *J. Gen. Virol.* **75**, 341 (1994).
20. A. M. Borman, S. Paulous, F. Clavel, *ibid.* **77**, 419 (1996).
21. P. J. Gerrish and R. E. Lenski, *Genetica* **103**, 127 (1998).
22. M. Nijhuis *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 14441 (1998).
23. S. F. Elena *et al.* [*Genetics* **142**, 673 (1996)] noted that, during experimental evolution of a virus, the gain in fitness was larger the lower the initial fitness of the viral clone. This is an alternative statement of our model.
24. A fitness improvement of 0.28 on the scale in Fig. 1 is sufficient to allow a substitution to increase over 48 hours from a frequency of 0.01 to 0.99, which represents the limits of our ability to detect polymorphism. This magnitude of change is compatible with several of the sweeps we observe, yet it is statistically undetectable in our assays.
25. The five lineages include the two reported here and the three reported in (*9*).
26. J. J. Holland, E. A. Grabau, C. L. Jones, B. L. Semler, *Cell* **16**, 495 (1979).
27. D. E. Dykhuizen, A. M. Dean, D. L. Hartl, *Genetics* **115**, 25 (1987).
28. L. Chao, *Nature* **348**, 454 (1990).
29. R. E. Lenski and M. Travisano, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 6808 (1994).
30. I. S. Novella *et al.*, *ibid.* **92**, 5841 (1995).
31. S. F. Elena, V. S. Cooper, R. E. Lenski, *Science* **272**, 1802 (1996).
32. We cannot unambiguously assess the role of recombination in these selections. Chemostat conditions were likely favorable to ϕX174 recombination, which requires multiple infection and recombination-competent hosts [J. C. Warner and I. Tessman, in *The Single-Stranded DNA Phages*, D. T. Denhardt, D. Dressler, D. S. Ray, Eds. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1978), pp. 417–432]. The concurrent increase of multiple substitutions during selection (Fig. 2), as well as the change in frequency of various single- and double-mutation combinations (data not shown), is consistent with recombination among phages in these cultures. However, recurrent mutation could also create the double-mutant genotypes and (depending on fitness interactions) could lead to similar population dynamics. Failure to observe the slow accumulation of substitutions over this 10-day period may reflect their being driven out or suddenly swept through to fixation by linkage disequilibrium with rapidly advancing changes, but these combinations could arise by mutation, recombination, or both.
33. Abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.

34. Funded by NIH grant GM57755 to J.J.B. C.M.B. was supported by the Sundquist Award for Undergraduate Research. This work was carried out while H.A.W. was on sabbatical leave at the University of Texas.

16 January 1999; accepted 15 June 1999