PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Experimental evolution recapitulates natural evolution

H. A. Wichman, L. A. Scott, C. D. Yarber and J. J. Bull

Phil. Trans. R. Soc. Lond. B 2000 **355**, 1677-1684
doi: 10.1098/rstb.2000.0731

THE ROYAL SOCIETY

doi 10.1098/rstb.2000.0731

Experimental evolution recapitulates natural evolution

H. A. Wichman[1], L. A. Scott[1], C. D. Yarber[1] and J. J. Bull[2*,3]

[1]*Department of Biological Sciences, University of Idaho, Moscow, ID 83844, USA*
[2]*Section of Integrative Biology, and* [3]*Institute of Cellular and Molecular Biology, University of Texas, Austin, TX 78712-1023, USA*

Genomes of the closely related bacteriophages ϕX174 and S13 are 5386 bases long and differ at 114 nucleotides, affecting 28 amino acids. Both parental phages were adapted to laboratory culture conditions in replicate lineages and analysed for nucleotide changes that accumulated experimentally. Of the 126 experimental substitutions, 90% encoded amino-acid changes, and 62% of the substitutions occurred in parallel in more than one experimental line. Furthermore, missense changes at 12 of the experimental sites were at residues differing between the parental phages; in ten cases the ϕX174 experimental lineages were convergent with the S13 parent, or vice versa, at both the nucleotide and amino-acid levels. Convergence at a site was even obtained in both directions in three cases. These results point to a limited number of pathways taken during evolution in these viruses, and also raise the possibility that much of the amino-acid variation in the natural evolution of these viruses has been selected.

Keywords: molecular evolution; microbial evolution; convergent evolution; parallel evolution; viral evolution; genome

1. INTRODUCTION

The direction and magnitude of adaptation by a population are determined by heritable variation, which ultimately depends on the nature of mutations. Nonetheless, one of the most popular and productive approaches in evolutionary biology is built on the principle of fitness optimization, which implicitly assumes that genetic details can be ignored. This view supposes that natural selection imposed by the environment, the breeding system and other properties of the population is the ultimate determinant of phenotypic evolution, because sufficient genetic variation exists or can arise to evolve in many directions (Maynard Smith 1982, 1989). The success of the optimization school suggests that organisms possess considerable latitude in adaptation. What, then, are the limits and constraints on adaptation?

Convergent and parallel molecular evolution are signatures of common pathways of evolution and may reflect limited avenues for adaptation. We use the term 'parallel evolution' to describe the independent occurrence of the same substitution in two independent lineages. Parallel evolution is a special case of convergent evolution, which describes a substitution to the same nucleotide that is present in another, independently evolved lineage. Parallel and convergent molecular evolution have been observed in both experimental populations (Bull *et al*. 1997; Crill *et al*. 2000; Cunningham *et al*. 1997; Treves *et al*. 1998; Wichman *et al*. 1999) and natural populations (Borman *et al*. 1996; ffrench-Constant 1994; Stewart & Wilson 1987) of viruses, bacteria and even insects and mammals exposed to the same, strong selection pressures. In a similar vein, comparisons of naturally evolved taxa sometimes exhibit a high incidence of amino-acid substitution at a small number of sites, suggesting that a limited number of residues are responding to selection (Bush *et al*. 1999; Crill *et al*. 2000; Fitch *et al*. 1997). Nonetheless, convergent and parallel molecular evolution are still regarded as novelties and are poorly understood, perhaps because of the paucity of systems in which it has been possible to observe DNA sequence changes during evolution.

The present study began as an extension of earlier experimental work on parallel evolution in the bacteriophage ϕX174 (Bull *et al*. 1997; Crill *et al*. 2000; Wichman *et al*. 1999). High levels of parallel evolution were observed in culture when replicate lineages of that phage were adapted to high temperature and novel host. The initial question for the present study was whether the changes accumulating during experimental adaptation depended on the sequence of the starting genome. Specifically, we sought to compare the nucleotide changes accumulating in ϕX174 with those accumulating in the close relative S13 under the same experimental conditions. At the completion of this work, it became clear that the most novel aspect of the study was a comparison of the experimental changes with the differences between the two natural isolates, ϕX174 and S13.

2. MATERIAL AND METHODS

The experimental portion of this study consisted of adapting 12 lineages of phage to bacterial hosts grown at the high temperature of 43.5 °C. Each period of adaptation lasted 10–11 days and was performed in a chemostat that maintained a tube of bacterial hosts free of phage, these hosts being pumped

*Author and address for correspondence (bull@bull.biosci.utexas.edu).

continuously into a second tube containing phage (Bull *et al.* 1997). Six experimental lineages were initiated with a parental isolate of S13 and six experimental lineages from a parental isolate of ϕX174. Three of the S13 lineages were grown on host *Escherichia coli* C (lineages 13C1, 13C2 and 13C3); three of the S13 lineages were grown on host *Salmonella typhimurium* (13S1, 13S2 and 13S3), and similarly for the ϕX174 lineages (XC1, XC2 and XC5; XS3, XSid and XStx). All of the S13 lineages and XC5 were new to this study, whereas the remaining ϕX174 (X) lineages have been published previously (the prefix 'X' was not used in their prior designations: see Bull *et al.* (1997) for C1, C2 and S3, and Wichman *et al.* (1999) for Sid and Stx). The X lineages C1, C2 and S3 were adapted for 11 days by raising the temperature by 1 °C daily from 38 to 43.5 °C with five days at the latter temperature, whereas the other nine lineages were initiated and maintained for ten days at 43.5 °C.

Each lineage is represented by the complete sequence of an isolate obtained at the end of its passage as well as the sequence of the starting (parental) phage. The sequence of our ϕX174 isolate differed from the original Sanger sequence (Sanger *et al.* 1977) at five positions (GenBank accession numbers AF176034 versus V01128, as noted in Bull *et al.* (1997)), whereas the sequence of our parental S13 isolate differed from that of the published sequence (Lau & Spencer 1985) at 22 positions (GenBank accession number AF274751 for our isolate, M14428 for the original S13 sequence). For S13, our isolate showed 11 contiguous differences from the published sequence over a 380-base stretch, for which the published sequence was identical with that of ϕX174 (base positions 512–892 in the ϕX174 genome). The simple interpretation of this difference is that either the previously published S13 isolate was a recombinant with ϕX174 over this region, or our isolate was a recombinant with an uncharacterized phage.

Additional details on the strains and methods of phage propagation and sequencing have been described in previous publications on this system (Bull *et al.* 1997; Crill *et al.* 2000; Wichman *et al.* 1999). For simplicity in this study, the numbering of bases for both ϕX174 and S13 follows the convention for ϕX174 in Sanger *et al.* (1977), and are therefore not consistent with the published numbering for S13 (Lau & Spencer 1985).

3. RESULTS

(a) *Natural isolates exhibited typical ratios of missense to silent variation*

Our parental ϕX174 and S13 isolates differed by seven substitutions in intergenic regions, 78 silent substitutions in coding regions, and 29 substitutions in codons with predicted amino-acid differences between the two phages (28 codons affected). Thus, the ratio of missense substitutions to silent substitutions was 0.37. For the spectrum of nucleotide differences that separated the ϕX174 and S13 isolates, the ratio of missense to silent changes that would be observed in the absence of selection was 2.06, which was 5.7-fold the observed ratio. (To address changes in overlapping genes, a change was considered silent if it did not affect the amino-acid sequence of any gene overlapping the site.) There has probably been purifying selection against missense substitutions to a greater extent than against silent ones, and the ratio of missense to silent changes was not high enough to arouse suspicion that adaptive evolution has had a role in this divergence (Li 1997).

(b) *Experimental lineages acquired mostly missense changes*

Among the 12 experimental lineages, there were 126 substitutions and two identical deletions (Appendix A). A total of 72 sites experienced one or more substitutions. All but ten of these sites encoded missense changes, and the others were silent. All ten silent changes were within coding regions; however, owing to the overlapping nature of some genes, substitutions in one gene can lie in regulatory regions of downstream genes. For example, one of the silent substitutions (at position 324) was in gene C and occurred in four lineages, but was also in the regulatory region of gene D. The ratio of missense to silent changes was therefore 8.7, which was 23-fold that between the parental phages. This result is consistent with the view that most of the changes in chemostats were selected, a view supported by the high level of parallelism and the observed dynamics of changes sweeping though the populations (Bull *et al.* 1997; Crill *et al.* 2000; Wichman *et al.* 1999).

(c) *Significant clustering of parallel changes occurred within host and phage*

In this study we use the term 'parallel evolution' to describe the independent occurrence of the same substitution in two independent lineages. For example, at nucleotide position 4110, the ancestral state in both ϕX174 and S13 is a C, but parallel substitutions to T occurred in four of the S13 experimental lineages and three of the ϕX174 experimental lineages. We use the term 'convergent evolution' to describe a substitution in an experimental lineage that is to the same nucleotide as in the other parental phage. For example, at nucleotide position 1460, the ancestral state in S13 is G and the ancestral state in ϕX174 is C. Two experimental lineages of S13 had G to C changes at 1460, whereas one ϕX174 lineage had a C to G change at this site. The two changes in S13 occurred in parallel (i.e. the same substitution at the same nucleotide site) and converged on the sequence of ϕX174, whereas the single change in ϕX174 converged on the sequence of S13.

Parallel substitutions were observed at 27 out of 72 sites that evolved in the experimental lines, accounting for 62% of all substitutions. The average number of substitutions at these sites was 2.9, but the distribution of parallelisms was broad: 14 of these sites experienced just two parallel changes, seven sites experienced three, four experienced four substitutions, and one site each experienced six and seven substitutions.

Parallel substitutions provide a convenient method of evaluating the effect of environment and genotype on the substitution process. The extent to which the same substitution arises in replicate lineages provides an unambiguous measure of the repeatability of the evolutionary process, one that controls for genome position, mutation process, and so on. However, in interpreting our data we point out that these experiments were performed in two different laboratories by all four authors, and some measured and many unmeasured variables were not equally spread across all phage–host combinations. Five lineages were part of previously published studies, and three of them (XC1, XC2 and XS3) were evolved under a different temperature protocol from the others. Thus,

some clustering of parallel substitutions within a phage type, or within a phage–host combination, could stem from subtle differences in propagation treatments. Examination of the data suggests that this effect is probably small. There is only slightly less parallel evolution between lines adapted under alternative protocols than between lines adapted under the same protocol, but little if any effect of different individuals or laboratories. For example, XC1 and XC2 (adapted under the same protocol in the same laboratory) had parallel changes at five sites, whereas XC5 (adapted under the later protocol in a different laboratory) had parallel changes with XC1 at four sites and XC2 at three sites. Similarly, XSid and XStx (adapted under the same protocol in different laboratories by different individuals) had parallel changes at six sites, whereas XS3 (adapted under the earlier protocol but in the same laboratory and by the same individual as XSid) had parallel changes with XSid at three sites and XStx at four sites. With this caveat, we offer a brief analysis of the clustering of parallel changes within treatments.

Parallel evolution was obtained both at sites whose initial base identity differed and at sites with the same base in the two parental phages. The latter sites offer the most straightforward way of evaluating whether host and/or genotype influenced evolution: if certain phage–host combinations are more likely than others to evolve a particular change, parallel changes should occur in the same phage–host combination more often than at random. For example, with two parallel changes, the chance that both occur in the same phage–host treatment is 2/11, if host and phage genome have no effect. Of the nine sites that experienced two parallel changes and did not differ between the parental phages, seven were confined to a single phage–host treatment (these seven were distributed across three of the four treatments). The probability of this clustering is approximately 0.0001 under the null model. Similarly, with three parallel changes, two of the six cases were also confined to a single phage–host treatment, for $p < 0.005$. There is thus substantial evidence that both phage and host affect the substitution process, subject to the caveat offered above. This conclusion strictly applies to only a subset of the sites included in these tests, as the substitution process at other sites could be insensitive to treatment.

(d) *Parallel and convergent evolution occurred between natural differences and experimental changes*

The most striking result is that, of the 28 amino-acid differences between our parental isolates of ϕX174 and S13, the experimental lineages evolved changes in 12 of these codons (table 1). The total number of nucleotide positions with changes that caused amino-acid substitutions in the experimental lineages was 62. A bootstrap analysis indicated that the expected number of experimental substitutions landing in these 28 codons was approximately 1, and that the expected number of substitutions at the same nucleotide sites was approximately 1/2 ($p \ll 10^{-6}$ for the observed level of convergence between experimental and natural isolates in either case; the bootstrap simulation used the observed substitution matrix and sampling without replacement).

Table 1. *Nucleotide substitutions between ϕX174 and S13 that result in amino-acid replacements*

(All sites of amino-acid replacement between the parental phages ϕX174 and S13 are shown. Amino-acid positions are proceeded by the single-letter designation of the phage protein affected; single-letter amino-acid codes are used. Sites undergoing substitutions in the experimental lineages are shown in bold. Convergent substitutions in the experimental lineages are indicated by a single asterisk (*) if they occurred in one direction only, and by a double asterisk (**) if they occurred in both directions. Substitutions at the same amino-acid residue that were not convergent are indicated by a hash sign (#).)

nucleotide position in ϕX174	amino-acid position	amino-acid residue in ϕX174 (S13)
686	E40	R (K)
690	D101	S (A)
771	D128	C (R)
830	E88	A (V)
870	**J8**	**F (S)***
883	J12	S (P)
1010	**F3**	**V (I)***
1242	**F80**	**R (Q)***
1305	**F101**	**S (G)#**
1460	**F153**	**E(Q)****
1727	**F242**	**F(L)***
1956	F318	A (V)
2085	**F361**	**V(A)****
2452	G20	A (L)
2456	G21	S (T)
2477	G28	L (S)
2731	G113	D (N)
2980	**H17**	**V(A)***
3120	**H64**	**S(P)****
3340	**H137**	**G (D)***
3351	H141	S (A)
3359	H143	I (M)
4203	**A75**	**Y(D)#**
4803	A275	S (T)
5123	B17	F (V)
5163	A395/B30	F (L)/V (A)
5185	**A402**	**M (T)***

The significance of this pattern is enhanced beyond the level indicated by this test in two ways. First, parallel changes were observed in more than one of the experimental lineages at seven out of these 12 sites. Second, convergent changes occurred in both directions at three sites (i.e. the residue in ϕX174 converged on the sequence of S13 and vice versa).

4. DISCUSSION

High rates of parallel substitutions were reported previously in ϕX174 adapted to high temperature in chemostats (Bull *et al.* 1997; Crill *et al.* 2000; Wichman *et al.* 1999). The present study extends those findings in that (i) the set of experimental lineages now includes S13, a close relative of ϕX174; (ii) parallel evolution exhibits significant clustering by phage and host and thus probably depends on the genotype of the starting phage (a host effect was reported previously); and (iii) striking similarities were observed between the substitutions in

experimental lines and the differences between our parental genomes of S13 and ϕX174.

Of the experimental substitutions, 90% were missense. As noted previously, this high rate of missense substitutions, combined with the previous observations of high levels of parallel substitutions and rapid nucleotide replacements in chemostats, indicate that most substitutions in chemostats are adaptive (Crill *et al*. 2000; Wichman *et al*. 1999). In contrast with the high percentage of missense substitutions in experimental lines, only *ca*. 25% of the differences between the parental phages were missense. This level of missense changes is more typical of natural evolution, and would usually be interpreted to suggest that drift and purifying selection of amino-acid substitutions were the predominant processes in the natural evolution of these viruses. However, the extraordinary concordance between experimental and natural missense substitutions raises the possibility that many of the natural amino-acid substitutions were selected.

A surprising result here is that many experimental changes are convergent with one of the parental phages. This type of convergence can be detected only at sites differing between the two parental genomes, and thus could not have been observed in our earlier studies. It means either that the same substitutions occurred in our experiments as in nature or that our experiments reversed a natural change. Three classes of model can be invoked to explain these results. These models extend and complement those proposed to explain the parallelism between experimental replicates (Bull *et al*. 1997).

(i) Population dynamics: large population sizes and high mutation rates allow the phage to explore the adaptive landscape more completely, and thus to converge on the same beneficial substitutions of large effect.
(ii) Genetic constraints: these phages have few residues that can change in response to selection, and those residues respond to a wide range of selective factors.
(iii) Common selection: the selective environment in our chemostats duplicated the selective history of the natural phages, or equivalently, different environmental challenges were selecting for the same phenotypic change in the phage.

These three models propose that convergence stems from extremes in different components of the evolutionary process, and the models are not necessarily exclusive of one another. For example, to suppose that a similar selective environment between our chemostats and the evolutionary history of the parental phages is the cause of convergence also requires either (i) a limited spectrum of beneficial mutations in that environment or (ii) few mutations of large benefit, and also a large population size to ensure the evolution of the best mutations. Any full description of adaptation necessarily includes all three components; the issue here is therefore the quantitative one of whether our system is extreme in any or all of these components. We now discuss each of these possibilities.

(a) *Population dynamics*

Population sizes in the chemostat were large. With typical titres of $10^8\,\mathrm{ml}^{-1}$ and a population doubling 100 times per day, a ten day experiment would pass in excess of 10^{11} phage through the chemostat. If approximately one in every 300 phage has a new mutation (Drake 1991), there would be 10^9 new mutations, or more than 10^5 variants per site, during the course of an experiment. These numbers are perhaps gross approximations—we do not know the effective number of phage in the chemostat, and mutation rates for the same class of transition (e.g. A→G) can vary by three orders of magnitude in a single phage genome (Ronen & Rahat 1976). While it is plausible that most single-site variants arose numerous times during the course of these experiments, some potentially adaptive changes may seldom (or never) occur due to mutational constraints. Thus complete exploration of the adaptive neighbourhood is unlikely even with very large population sizes. Nevertheless, independent populations would have been likely to converge on the same substitutions of large effect, as long as there were few classes of large-benefit mutations available. This model also requires that the ancestral, natural populations of ϕX174 and S13 were large.

(b) *Genetic constraints*

This model assumes that few residues in the genome are free to vary without deleterious consequences, and that these few sites can also respond to multiple selective factors. There is in fact some evidence to support this idea. Fane and co-workers (Ekechukwu & Fane 1995; Fane & Hayashi 1991; Fane *et al*. 1993) identified 16 ϕX174 mutations affecting gpF, obtained either as second-site suppressors of primary phage mutations in internal and external scaffolding proteins, or in response to host defects. Of the 426 residues in gpF, three changes are common to Fane's 16 mutations and the 18 substitutions in gpF in our experimental lines (expected: 0.76; $p = 0.03$ for three or more residues in common under the null model; Poisson test).

One interpretation of this result is that these sites are responsive to multiple aspects of the environment. However, the magnitude of concordance between Fane's sites and ours is much less than that observed between our experimental lines and the parental differences. An alternative interpretation of these results (and this model) is that what seem to us to be very different selective pressures are in fact favouring the same phenotypic response. For example, exposure to a high temperature, a novel host or a fixation of new substitutions might all destabilize the procapsid, and thus select for the same substitutions.

(c) *Common selection*

Our remaining alternative concerns the similarity of selection. It seems incredible that differences in the selective history between ϕX174 and S13 might be recreated in the chemostats. However, some aspects of this model are plausible. S13 was originally isolated on *Salmonella typhimurium* and, after an unknown period of laboratory propagation on that host, was switched to *E. coli* C. Possibly, host is a factor differing in similar ways between the evolutionary history of the natural isolates and in our chemostats (although our *Salmonella* strain, at least, undoubtedly differs from natural strains in some important ways). In support of the host as a common variable across natural and experimental lines, some of the differences

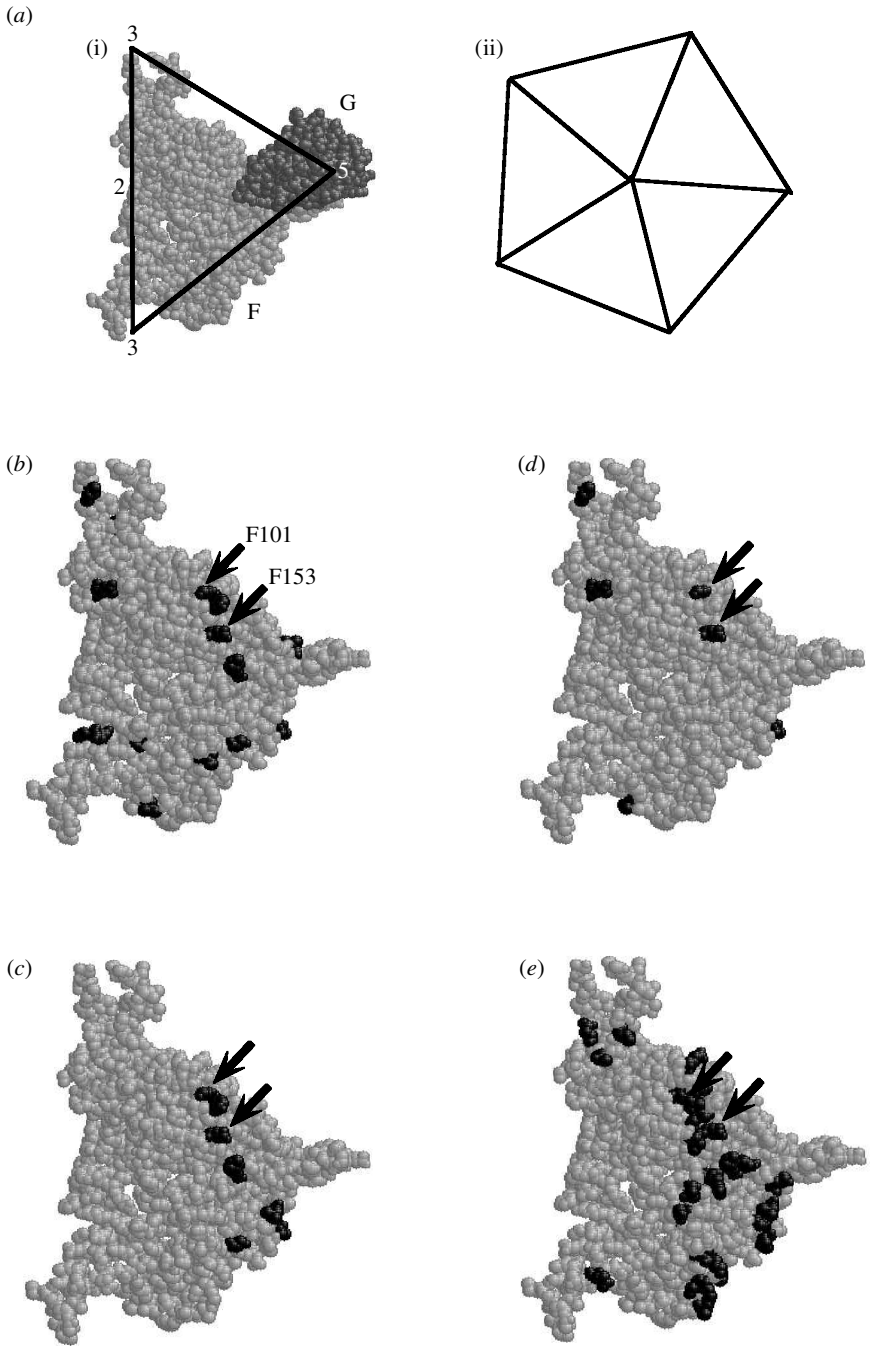

Figure 1. Substitutions in experimental and natural isolates overlaid on the strucuture of gpF. The outer surface of gpF is shown (McKenna *et al*. 1994). Structures are drawn with RasMol. (*a*) Orientation of gpF in the capsid. (i) The outer surface of gpF, the major capsid protein, is shown in relationship to gpG, the major spike protein. The twofold, threefold and fivefold axes of symmetry are marked. (ii) Five copies of gpF and gpG form pentameric units as shown; the gpG spike is in the centre of the pentameric unit. Twelve of these units make up the icosahedral phage (not shown). (*b*) Substitutions in the experimental lines. Residues undergoing substitution in any of the 12 experimental lines are shown in black. Of the 18 sites of substitutions in gpF among all experimental lines, 13 are clearly visible on the outer surface. Two host switching sites, F101 and F153 are labelled. (*c*) Host switching sites. Seven host switching substitutions in gpF are shown in black. F101 and F153 are marked with arrows. Six of the seven sites undergo substitutions when ϕX174 is evolved on a novel host and reversions when evolved on the natural host. The seventh site was identified by convergent substitutions in ϕX174 and S13 when evolved on alternative hosts. (*d*) Differences between natural isolates. Residues in gpF that differ between ϕX174 and S13 are shown in black. F101 and F153 are marked with arrows. (*e*) Differences between natural isolates. Residues in gpF that differ between α3 and ϕK are shown in black. F101 and F153 are marked with arrows.

between the S13 and ϕX174 major capsid gene were shown to affect host-specific adaptation in chemostats (Crill *et al*. 2000). A further selective factor that could be identical between natural and experimental lines is intraspecific competition. Chemostats maintain high densities of phage, such that one of the major selective agents is probably competition between unrelated genomes infecting the same cell. This density-dependent

competition can impose selection that is largely a function of the phage genome itself, regardless of whether the competition occurs in nature or the laboratory.

A comparison of two different natural isolates corroborates this model, although with less force. The two isometric phages α3 and φK both differ from φX174 and S13 at *ca*. 39% of their nucleotides in the major capsid protein (gene F), but differ from each other by only 13%. The major capsid protein (gpF) is easily alignable and is almost the same length in all four phages, so across these regions of known homology, the 36 residues differing between α3 and φK can be compared with the nine differing between φX174 and S13 (figure 1). There are three residues in common, compared with only 0.76 expected ($p < 0.05$), so there is weak evidence that evolution has been restricted to a portion of this molecule, but again the effect does not seem to be large. However, in this case the extensive divergence of the gene between the two pairs of phages weakens any expectation that the same subset of residues would be responsible for most adaptation. Nonetheless, two of the three residues that differ between φX174 and S13 and also differ between α3 and φK are included in the seven residues demonstrated by experimental evolution to be involved in host specificity (F101 and F153; figure 1*c*). These same two residues also changed in our experimental lines.

An elaboration of our understanding of capsid mutations might shed light on these parallelisms and convergences in this gene. We have identified seven residues in gpF that seem to affect fitness on *E. coli* C compared with *S. typhimurium* differentially (six were identified in Crill *et al*. (2000), and we have since identified a seventh because of convergence between S13 selected on *E. coli* C and φX174 selected on *Salmonella* (this study)). These seven host switching sites occur in a band on the surface of gpF (figure 1*c*) near the viral spike formed by gpG (figure 1*a*). In the intact capsid, they form a ring round the spike. Five out of these seven changes are seen in these experimental lines (figure 1*b*), and three of these residues differ between the parental phages, φX174 and S13 (figure 1*d*). Furthermore, this region of gpF has the bulk of the differences between α3 and φK for this protein (figure 1*e*), and it has undergone multiple radical amino-acid substitutions during the evolution of natural isolates characterized so far (Crill *et al*. 2000). These results suggest that selection on multiple hosts might be important in the adaptive divergence of these phages in nature, that changes in response to host switching are clustered on the surface of the phage, and that a few sites might have a major effect on this phenotype.

A related observation has recently been reported for influenza A (Bush *et al*. 1999; Fitch *et al*. 1997). With the use of sequences of the haemagglutinin protein from viral isolates taken over three decades of annual epidemics, this study identified a small set of amino-acid residues exhibiting accelerated rates of molecular evolution. In this case, the sites of interest did not show repeated evolution to the same amino acids, but rather showed multiple changes over time (directional evolution). The presumed basis of selection is attack by the immune system. These residues have the interesting property that they are being used to predict which strains among those currently circulating will die out and which will be the progenitors of circulating strains in future years. Together, the phage and influenza results point to a pattern of molecular evolution (under particular but perhaps broad types of selection) in which substitutions are confined to specific amino-acid residues.

A deliberate comparison between natural and experimental variation was engineered in β-lactamase (Huang *et al*. 1996; Palzkill & Botstein 1992). Site-directed mutagenesis was used to randomize codons systematically, two at a time, across the entire molecule. Variants retaining antibiotic resistance were sequenced and compared with codons observed in natural isolates, revealing a remarkable similarity but with some important differences. In this comparison, the experimental variants were not necessarily selectively advantageous, but were merely shown to maintain activity above a threshold.

It is unknown how similar our parental phages are to the phages originally obtained from nature. Those isolates were obtained half a century ago and are no longer available in original form (having been passaged periodically to maintain high titres). In view of the ease with which phages evolve in culture and the known variation between different stocks originating from the same virus (for example, compare Bull *et al*. (1997), Sanger *et al*. (1977) and Fane & Hayashi (1991)), we can presume that several base substitutions have accumulated during their 50-year maintenance. Could that process explain some of the concordance between experimental substitutions and parental phage differences? The possibility seems remote, because laboratory propagation should already have eliminated some differences between the phages. However, this question is perhaps unanswerable directly until new isolates have been obtained from the wild and adapted to the laboratory. Our findings concern the existing differences between parental S13 and φX174, and their propagation in the laboratory on the same host and under similar conditions might already have eliminated some parental differences that would have contributed to our pattern.

In conclusion, we have a puzzling result: experimental lineages of two bacteriophages evolved substitutions that strongly converged on differences between the parental genotypes. This pattern might have been created through a combination of factors: large population size, constraints on residues capable of responding to selection, and common selection between our experiments and evolution in nature. Recent work on molecular evolution of the haemagglutinin protein of influenza A and the antibiotic resistance gene β-lactamase reveals some parallels to the phage results.

This work was funded by the National Institutes of Health.

APPENDIX A

In table A1, nucleotide positions, numbered according to the positions for φX174 in GenBank accession V01128 (Sanger *et al*. 1977), are shown in column 1. Nucleotide positions that differ between the parental phage are preceded by a symbol indicating the degree of convergent evolution in the experimental lineages: convergent substitutions are indicated by a single asterisk (*) if they occurred in one direction only, and by a double asterisk

Table A1. *Nucleotide substitutions in experimental lineages*

number of substitutions per lineage			Sl3anc	l3Sl (13)	l3S2 (13)	l3S3 (8)	l3Cl (6)	l3C2 (8)	l3C3 (10)	Xanc	X S3 (13)	X Sid (12)	X Stx (13)	X Cl (12)	X C2 (11)	X C5 (7)
4110	A44	H-Y	c	c	c	T	T	T	T	c	T	T	T	c	c	c
4122	A48	D-N	g	A	g	g	g	g	g	g	g	g	g	A	A	A
4168	A63	Q-R	a	a	a	a	a	a	a	a	a	G	G	a	a	a
#4203	A75	Y-H	t	C	t	t	t	t	t	g	g	g	g	g	g	g
4420	A147	D-A	a	C	C	a	a	a	a	a	C	a	a	a	a	a
4430	A150	D-E	t	t	A	t	t	t	t	t	t	t	t	t	t	t
4623	A215	D-N	g	g	g	g	g	g	A	g	g	g	g	g	A	g
4637	(A219)	s	t	t	t	t	t	t	t	t	t	C	t	t	t	t
4657	A226	D-G	a	a	a	a	a	G	a	a	a	a	a	a	a	a
4751	(A257)	s	a	a	a	a	a	a	a	a	a	a	a	G	a	a
4796	(A272)	s	a	a	a	a	a	G	a	a	a	a	a	a	a	a
4861	A294	S-T	g	g	C	g	g	g	g	g	g	g	g	g	g	g
4886	(A302)	s	t	t	C	t	t	t	t	t	t	t	t	t	t	t
5121	A381/B16	L-I/A-D	c	A	c	c	c	c	c	c	c	c	c	c	c	c
*5185	A402 (B37)	M;T-M/s	t	t	t	t	t	t	t	c	c	c	T	T	c	c
5214	A412/B47	G-S/G-E	g	g	g	A	g	g	A	g	g	g	g	g	g	g
5244	A422/B57	L-V/I-S	t	t	G	t	t	t	t	t	t	t	t	t	t	t
5304	A442/B77	V-I/R-H	g	g	g	g	g	g	g	g	g	g	g	g	g	A
5365	A462 (B97)	M-T/s	t	t	t	t	t	t	t	t	t	C	C	t	t	t
28	(A478) B114	s/V-I	g	g	g	g	g	g	g	g	g	g	g	g	A	g
31	(A479) B115	s/N-D	a	a	a	a	a	a	a	a	a	a	a	G	a	a
203	C24	T-M	c	c	T	c	c	c	c	c	c	c	c	c	c	c
226	(C32)	s	t	t	t	t	t	t	t	t	t	C	t	t	t	t
319	C63	V-F	g	g	g	g	g	g	g	g	T	g	T	T	g	g
323	C64	E-G	a	a	a	a	a	G	a	a	a	a	a	a	a	a
324	(C64)	s	c	c	T	c	T	c	T	c	c	c	c	c	c	T
344	C71	L-S	t	C	t	t	t	t	t	t	t	t	t	t	t	t
500	(D37)	s	g	g	g	g	g	g	g	g	T	g	g	g	g	g
554	(D55)	s	t	t	C	t	t	t	t	t	t	t	t	t	t	t
572	(D61) E2	s/V-A	t	C	t	t	t	t	t	t	t	t	t	t	t	t
590	(D67) E8	s/D-G	a	a	a	G	a	a	a	a	a	a	a	a	a	a
756	D123 (E63)	F-L/s	t	t	t	t	t	t	t	t	t	t	t	C	C	C
782	(D132) E72	s/T-I	c	c	c	c	c	c	c	c	c	c	T	c	c	c
789	D134 (E74)	V-L/s	g	g	g	g	g	A	g	g	g	g	g	g	g	g
*870	J8	F-S;S	t	C	C	t	t	t	t	c	c	c	c	c	c	c
872	J9	G-C	g	g	g	g	g	g	g	g	g	g	g	g	T	g
894	J16	Q-R	a	a	a	G	a	a	a	a	a	a	a	a	a	a
903	J19	R-Q	g	g	g	g	g	g	g	g	A	g	g	g	g	g
965–991	—	—										Δ	Δ			
*1010	F3	V-I;I	g	g	g	g	A	g	g	a	a	a	a	a	a	a
1025	F8	E-K	g	g	g	g	g	g	A	g	A	g	A	A	g	g
1180	(F59)	s	t	t	t	t	t	t	t	t	C	t	t	t	t	t
*1242	F80	R-Q;Q	g	g	g	g	g	g	A	a	a	a	a	a	a	a
1300	(F99)	s	t	t	t	t	t	t	t	t	t	C	t	t	t	t
#1305	F101	S;G-D	c	c	c	c	c	c	c	g	g	A	A	g	g	g
1308	F102	Y-C	a	a	a	a	a	a	a	a	a	G	a	a	a	a
**1460	F153	E-Q;Q-E	g	g	g	g	g	C	C	c	c	G	c	c	c	c
1533	F177	T-I	c	c	c	c	c	c	c	c	c	T	T	c	T	c
1613	F204	T-S	a	a	T	T	T	T	T	a	T	a	a	a	a	a
1618	F205	D-E	c	c	c	c	c	c	c	c	G	c	c	c	c	c
*1727	F242	F;L-F	t	t	t	t	t	t	t	c	c	c	T	T	T	T
1748	F249	V-I	g	g	A	g	g	g	g	g	g	g	g	g	g	g
1959	F319	L-S	t	C	t	t	t	t	t	t	t	t	t	t	t	t
2000	F333	V-I	g	A	g	g	g	g	g	g	g	g	g	g	g	g
2009	F336	S-T	t	t	t	t	t	t	t	t	A	t	t	t	t	t
2018	F339	S-P	t	t	t	C	t	t	t	t	t	t	t	t	t	t
**2085	F361	V-A;A-V	t	t	t	t	C	t	t	c	c	c	T	c	c	c
2150	F383	V-I	g	A	g	A	g	g	g	g	g	g	g	g	g	g
2167	F388	H-Q	t	G	t	t	t	t	t	t	G	G	t	t	t	t
2410	G6	I-V	a	a	a	a	a	a	a	a	G	a	a	a	a	a
2591	G66	V-A	t	t	t	t	t	t	t	t	t	t	t	C	t	t

(*Cont.*)

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Table A1. (*Cont.*)

number of substitutions per lineage			S13anc	13S1 (13)	13S2 (13)	13S3 (8)	13C1 (6)	13C2 (8)	13C3 (10)	Xanc	X S3 (13)	X Sid (12)	X Stx (13)	X C1 (12)	X C2 (11)	X C5 (7)
2973	H15	G-S	g	g	g	g	g	g	g	g	g	g	A	g	g	g
2979	H17	A-I	g	g	g	g	g	g	g	g	g	g	g	g	A	g
*2980	H17	V;A-V	t	t	t	t	t	t	t	c	c	c	c	T	T	c
3055	H42	N-S	a	G	G	a	a	a	a	a	a	a	a	a	a	a
3111	H61	V-I	g	g	g	g	g	g	g	g	g	g	g	A	g	A
**3120	H64	S-P;P-S	t	t	t	t	t	t	C	c	c	c	c	c	T	c
3166	H79	A-V	c	c	c	c	c	c	c	c	T	T	T	c	c	c
*3340	H137	G-D;D	g	g	g	A	A	g	A	a	a	a	a	a	a	a
3378	H150	N-H	a	a	a	a	a	a	a	a	a	a	a	C	C	a
3381	H151	Q-E	c	G	G	c	c	c	c	c	c	c	c	c	c	c
3402	H158	Q-E	c	c	c	c	c	G	c	c	c	c	c	c	c	c
3847	H306	A-D	c	c	c	c	c	c	c	c	c	c	c	c	A	c

(**) if they occurred in both directions; substitutions at the same amino-acid residue that were not convergent are indicated by a hash sign (#). The single-letter gene designation and amino-acid residue affected are shown in column 2; residues in which substitutions are synonymous are enclosed in parentheses. The third column shows the amino-acid identities in the parental phage followed by that in the evolved phage. Synonymous sites are indicated by a lower-case 's'. Where reading frames are overlapping, amino-acid identities are shown in the order indicated in column 2, and genes are separated by a solidus (/). Where the parental phages differ, the amino-acid identities for S13 parental and evolved phages are shown first, followed by amino-acid identities for ϕX174; parental types are separated by a semicolon (;). Nucleotide identities are shown in columns 4–17. Parental states are shown in lower-case letters; evolved states are shown in upper-case letters.

REFERENCES

Borman, A. M., Paulous, S. & Clavel, F. 1996 Resistance of human immunodeficiency virus type 1 to protease inhibitors: selection of resistance mutations in the presence and absence of the drug. *J. Gen. Virol.* **77**, 419–426.

Bull, J. J., Badgett, M. R., Wichman, H. A., Huelsenbeck, J. P., Hillis, D. M., Gulati, A., Ho, C. & Molineux, I. J. 1997 Exceptional convergent evolution in a virus. *Genetics* **147**, 1497–1507.

Bush, R. M., Bender, C. A., Subbarao, K., Cox, N. J. & Fitch, W. M. 1999 Predicting the evolution of human influenza A. *Science* **286**, 1921–1925.

Crill, W. D., Wichman, H. A. & Bull, J. J. 2000 Evolutionary reversals during viral adaptation to alternating hosts. *Genetics* **154**, 27–37.

Cunningham, C. W., Jeng, K., Husti, J., Badgett, M., Molineux, I. J., Hillis, D. M. & Bull, J. J. 1997 Parallel molecular evolution of deletions and nonsense mutations in bacteriophage T7. *Mol. Biol. Evol.* **14**, 113–116.

Drake, J. W. 1991 A constant rate of spontaneous mutation in DNA-based microbes. *Proc. Natl Acad. Sci. USA* **88**, 7160–7164.

Ekechukwu, M. C. & Fane, B. A. 1995 Characterization of the morphogenetic defects conferred by cold-sensitive prohead accessory and scaffolding proteins of ϕX174. *J. Bacteriol.* **177**, 829–830.

Fane, B. A. & Hayashi, M. 1991 Second-site suppressors of a cold-sensitive prohead accessory protein of bacteriophage ϕX174. *Genetics* **128**, 663–671.

Fane, B. A., Shien, S. & Hayashi, M. 1993 Second-site suppressors of a cold-sensitive external scaffolding protein of bacteriophage ϕX174. *Genetics* **134**, 1003–1011.

ffrench-Constant, R. H. 1994 The molecular and population genetics of cyclodiene insecticide resistance. *Insect Biochem. Mol. Biol.* **24**, 335–345.

Fitch, W. M., Bush, R. M., Bender, C. A. & Cox, N. J. 1997 Long term trends in the evolution of H(3) HA1 human influenza type A. *Proc. Natl Acad. Sci. USA* **94**, 7712–7718.

Huang, W., Petrosino, J., Hirsch, M., Shenkin, P. S. & Palzkill, T. 1996 Amino acid sequence determinants of β-lactamase structure and activity. *J. Mol. Biol.* **258**, 688–703.

Lau, P. C. & Spencer, J. H. 1985 Nucleotide sequence and genome organization of bacteriophage S13 DNA. *Gene* **40**, 273–284.

Li, W.-H. 1997 *Molecular evolution*. Sunderland, MA: Sinauer Associates.

McKenna, R., Ilag, L. L. & Rossmann, M. G. 1994 Analysis of the single-stranded DNA bacteriophage ϕX174, refined at a resolution of 3.0 Å. *J. Mol. Biol.* **237**, 517–543.

Maynard Smith, J. 1982 *Evolution and the theory of games*. Cambridge University Press.

Maynard Smith, J. 1989 *Evolutionary genetics*. Oxford University Press.

Palzkill, T. & Botstein, D. 1992 Probing β-lactamase structure and function using random replacement mutagenesis. *Proteins* **14**, 29–44.

Ronen, A. & Rahat, A. 1976 Mutagen specificity and position effects on mutation in T4rII nonsense sites. *Mutat. Res.* **34**, 21–34.

Sanger, F., Air, G. M., Barrell, B. G., Brown, N. L., Coulson, A. R., Fiddes, C. A., Hutchison, C. A., Slocombe, P. M. & Smith, M. 1977 Nucleotide sequence of bacteriophage ϕX174 DNA. *Nature* **265**, 687–695.

Stewart, C. B. & Wilson, A. C. 1987 Sequence convergence and functional adaptation of stomach lysozymes from foregut fermenters. *Cold Spring Harb. Symp. Quant. Biol.* **52**, 891–899.

Treves, D. S., Manning, S. & Adams, J. 1998 Repeated evolution of an acetate-crossfeeding polymorphism in long-term populations of *Escherichia coli*. *Mol. Biol. Evol.* **15**, 789–797.

Wichman, H. A., Badgett, M. R., Scott, L. A., Boulianne, C. M. & Bull, J. J. 1999 Different trajectories of parallel evolution during viral adaptation. *Science* **285**, 422–424.