EXPERIMENTAL EVOLUTION, LOSS-OF-FUNCTION MUTATIONS, AND "THE FIRST RULE OF ADAPTIVE EVOLUTION"

Michael J. Behe

Department of Biological Sciences, Lehigh University, Bethlehem, Pennsylvania 18015 USA

e-mail: mjb1@lehigh.edu

keywords

experimental evolution, adaptation, mutation, loss of function, malaria, *Yersinia pestis*

abstract

Adaptive evolution can cause a species to gain, lose, or modify a function; therefore, it is of basic interest to determine whether any of these modes dominates the evolutionary process under particular circumstances. Because mutation occurs at the molecular level, it is necessary to examine the molecular changes produced by the underlying mutation in order to assess whether a given adaptation is best considered as a gain, loss, or modification of function. Although that was once impossible, the advance of molecular biology in the past half century has made it feasible. In this paper, I review molecular changes underlying some adaptations, with a particular emphasis on evolutionary experiments with microbes conducted over the past four decades. I show that by far the most common adaptive changes seen in those examples are due to the loss or modification of a pre-existing molecular function, and I discuss the possible reasons for the prominence of such mutations.

Adaptation by Gain, Loss, or Modification of Function

In *THE ORIGIN OF SPECIES*, Darwin (1859) emphasized the relentlessness of natural selection:

> [N]atural selection is daily and hourly scrutinising, throughout the world, every variation, even the slightest; rejecting that which is bad, preserving and adding up all that is good; silently and insensibly working, whenever and wherever opportunity offers, at the improvement of each organic being in relation to its organic and inorganic conditions of life. (p. 84)

Yet he realized that the changes that were selected to adapt an organism to its environment did not have to be ones that conferred upon it a new ability, such as sight or flight. For instance, while observing some barnacles, Darwin discovered unexpected cases of the gross simplification of an organism (Stott 2003):

> The male is as transparent as glass . . . In the lower part we have an eye, & great testis & vesicula seminalis: in the capitulum we have nothing but a tremendously long penis coiled up & which can be exserted. There is no mouth no stomach no cirri, no

The Quarterly Review of Biology, December 2010, Vol. 85, No. 4

0033-5770/2010/8504-0002$15.00

> proper thorax! The whole animal is reduced to an envelope . . . containing the testes, vesicula, and penis. (p. 213)

Adaptive evolution can just as easily lead to the loss of a functional feature in a species lineage as to the gain of one. Over the course of evolutionary history, snakes have lost legs, cavefish have lost vision, and the parasitic bacterium *Mycoplasma genitalium* has lost its ability to live independently in the wild, all in an effort to become better adapted to their environments. Whatever variation that sufficiently aids a particular species at a particular moment in a particular environment—be it the gain or loss of a feature, or the simple modification of one—may be selected. Since species can evolve to gain, lose, or modify functional features, it is of basic interest to determine whether any of these tends to dominate adaptations whose underlying molecular bases are ascertainable. Here, I survey the results of evolutionary laboratory experiments on microbes that have been conducted over the past four decades. Such experiments exercise the greatest control over environmental variables, and they yield our most extensively characterized results at the molecular level.

DISTINCTIONS AMONG ADAPTIVE MUTATIONS

Adaptation is often viewed from two distinct aspects: the phenotypic aspect and the molecular aspect. In order to avoid confusion concerning how adaptive evolution proceeds, these two aspects must be kept separate. A phenotypic adaptation might have any number of underlying molecular bases, resulting from either the gain, loss, or modification of a molecular feature. As an example, consider a bacterium that evolves virulence in a mammalian host. If the adaptive phenotype is deemed to be the gain of pathogenicity, then that could arise in several ways. For instance: 1) a gene for a bacterial protein that is the target of the mammalian immune response might be deleted, which would be the loss of a functional molecular feature of the bacterium; 2) a gene for a protein that is the target of the mammalian immune response might sustain a missense mutation, causing an amino acid substitution at a recognized epitope of the protein, thereby decreasing or eliminating the immune response while preserving protein function, which could be considered a simple modification of a pre-existing molecular element; 3) the bacterium might acquire a gene or gene cluster by horizontal transfer that leads to expression of a new surface feature that hides the epitope from the mammalian immune system, which would be a gain of a functional molecular feature. The focus of this review is on the gain, loss, or modification of functional molecular features underlying adaptation, no matter whether the phenotypic manifestation strikes us as a loss or gain of a property.

There are, of course, myriad functional molecular features of a cell, including those of all the major biomolecular categories: nucleic acids, proteins, polysaccharides, lipids, small metabolites, and so forth. However, the functional molecular features of only two categories—nucleic acids and proteins—are directly coded by the genome, and thus only these features can be directly affected by mutation; all other biological features are only affected indirectly by mutations, through their effects on nucleic acids and proteins. In this review, I focus on adaptive evolution by gain, loss, or modification of what I term *Functional Coded elemenTs* (FCTs). An FCT is a discrete but not necessarily contiguous region of a gene that, by means of its nucleotide sequence, influences the production, processing, or biological activity of a particular nucleic acid or protein, or its specific binding to another molecule. Examples of FCTs are: promoters; enhancers; insulators; Shine-Dalgarno sequences; tRNA genes; miRNA genes; protein coding sequences; organellar targeting- or localization-signals; intron/extron splice sites; codons specifying the binding site of a protein for another molecule (such as its substrate, another protein, or a small allosteric regulator); codons specifying a processing site of a protein (such as a cleavage, myristoylation, or phosphorylation site); polyade-

nylation signals; and transcription and translation termination signals. Examples of general nucleic acid features that are not FCTs include base composition, strand asymmetry, distance between coded elements, and the like. They may enter into consideration indirectly only if they cause an FCT to be lost or gained—for example, if altering the base composition also eliminates the activity of an enhancer.

With these considerations in mind, I classify adaptive mutations as belonging to one of three modes:

1) A "loss-of-FCT" adaptive mutation is a mutation that leads to the effective loss of the function of a specific, pre-existing, coded element, while adapting an organism to its environment. The loss of the ability of a frame-shifted gene to produce a functional product, of an altered promoter to bind a transcription factor, or of a mutated protein to bind its former ligand, are examples of loss-of-FCT mutations.

2) A "gain-of-FCT" adaptive mutation is a mutation that produces a specific, new, functional coded element while adapting an organism to its environment. The construction by mutation of a new promoter, intron/exon splice site, or protein processing site are gain-of-FCT mutations. Also included in this category is the divergence by mutation of the activity of a previously duplicated coded element.

3) A "modification-of-function" adaptive mutation is a mutation whose defining property is negative—while adapting an organism to an environment, it does not lead to the loss or gain of a specific FCT. This definition is intended to be broad enough to act as a "catch-all" for anything that falls outside the first two modes. It includes point mutations as well as other mutations that have a quantitative effect on a pre-existing FCT, increasing or decreasing its strength, for instance, or shifting its activity somewhat (such as allowing a protein to bind a structurally-related ligand at the same site as its normal substrate), but without effectively eliminating it. The category "modification-of-function" also includes the simple duplication of features, without divergence of activity, such as a gene or regulatory region, since this process does not by itself produce a new functional element, although it may, like other events classified here as "modification-of-function," be a step to future gain-of-FCT events. It also includes mutations that may act by more amorphous means, such as rearranging gene order, or changing the distance or orientation between two interacting elements (e.g., the deletion of a stretch of DNA that brings two interacting transcription factors closer, or that affects their orientation with respect to each other). This category is dubbed "modification-of-*function*" rather than "modification-of-*FCT*" in order to emphasize that the mutation may not necessarily exert its influence through alteration of a discrete coded element, as in the above amorphous examples.

In order to fall under any of the above three categories, a mutation must be adaptive. Therefore, if, for instance, a mutation causes a transcription factor binding site to be formed or lost, but the mutation is either neutral or deleterious, it is excluded from consideration. Although assignment of an adaptive mutation to one of the modes above may occasionally be ambiguous, the assignment is straightforward in the great majority of cases in which molecular changes are elucidated, as we shall see below.

Illustrations of modes of adaptation

To better our understanding of the kinds of mutations that fall into these three modes, it is useful to look at illustrations from outside the realm of the experimental evolution of microbes. Consider, for instance, human mutations that have been selected in the past 10,000 years under the pressure of the malarial parasite *Plasmodium falciparum*. The reason for using such examples as these is that, because of its threat to human health, the interaction of malaria with humans has been very well-studied at the molecular level.

If a person of typical northern European descent visited a malarious region of the world and contracted the disease, she would likely view the resistance to malaria among many of the native peoples as an

TABLE 1
Human mutations selected for resistance to malaria

Clinical phenotype	Underlying mutation	Mode of adaptation	Reference
Hb S	β6 glu⇨val	G	Cavalli-Sforza et al. (1994)
Hb C	β6 glu⇨lys	M	Modiano et al. (2001)
Hb E	β26 glu⇨lys	M	Weatherall et al. (2002)
Hereditary persistence of fetal hemoglobin	Deletion/point mutations in control regions of hemoglobin gamma chain gene	L	Forget (1998)
Thalassemia	Deletion/point mutations in either α or β hemoglobin genes	L	Flint et al. (1986)
G6PD deficiency	Point mutations, deletions of G6PD gene	L	Ruwende et al. (1995)
Loss of Duffy antigen in red blood cells	Point mutation switching off production of Duffy antigen specifically in red blood cells	L	Pogo and Chaudhuri (2000)
Band 3 protein deficiency	Deletion	L	Kennedy (2002)

Abbreviations:
G - adaptive gain of functional coded element
L - adaptive loss of functional coded element
M - adaptive modification of function

unquestionably positive trait: the ability to withstand a debilitating disease. However, because of the detailed knowledge of the molecular bases of malaria resistance that has been accumulated over past decades, that is not the way I would categorize it here. Many persons indigenous to malarious regions are resistant, but they have acquired their resistance through different molecular mechanisms (Table 1). Although the mutations are all adaptive, distinctions may be made among them. The most well-known population of resistant persons is heterozygous for sickle hemoglobin. In one of the two genes that they inherited from their parents for the β chain of hemoglobin, the sixth codon designates valine rather than the glutamic acid that the other β chain gene codes. When the red blood cell is invaded by the malarial parasite, the mutant, hydrophobic valine residue allows hemoglobin molecules to aggregate with each other into long microtubular structures. This polymerization is associated with a measure of resistance to malaria (Friedman 1978), although the exact mechanism for resistance in vivo is still a matter of debate (Verra et al. 2007). This is an example of an adaptive gain-of-FCT mutation because a codon helping to specify a new, albeit weak—with $K_d \sim 1$ mM (Behe and Englander 1979)—protein binding site has appeared.

Another population indigenous to some malarial regions has a different point mutation in their hemoglobin. In this instance, the sixth codon of the β chain has mutated from glutamic acid to lysine. Although the altered hemoglobin (HbC) does not aggregate as sickle hemoglobin does, it confers resistance to malaria for reasons that are unclear. Because apparently no new, discrete, coded molecular feature has been developed, this is categorized as a "modification-of-function" adaptive mutation; no FCT has been lost or gained.

A third population of natives in some malarial regions is comparatively healthy due to a different kind of mutation. They have a condition called thalassemia, which, like sickle hemoglobin and HbC, also confers a measure of resistance to malaria. In a thalassemic person, however, one of the α or β chain hemoglobin genes that is inherited from a parent is nonfunctional, for any one of a variety of reasons. In some thalassemic individuals, a whole gene has been deleted, whereas, in others, a frameshift mutation has caused the gene to produce a nonfunctional protein. In still others, the genetic control region near

one of the globin genes has suffered a change that makes it nonfunctional, and so it produces no protein. Any of these would be an example of a loss-of-FCT mutation.

Other populations are also malaria-resistant due to loss-of-FCT mutations, but in genes other than the globin genes. Take, for instance, the gene that codes for glucose-6-phosphate dehydrogenase (G6PD). Any one of hundreds of separate known mutations in the coding and control regions of the gene has eliminated the functional enzyme, or has significantly decreased its catalytic activity. Other loss-of-FCT mutations include the deletion of a gene for Band 3 protein, and the loss of a promoter site for expression of Duffy antigen in red blood cells.

In summary, human genetic lines in malarious regions of the world have adapted to an environment that includes the malarial parasite, thereby giving rise to a number of molecular changes (Table 1). They have done so by different modes, which can be classified as adaptive mutations involving loss- or gain-of-FCT, or modification-of-function. As we shall see below, microbes in laboratory evolution experiments also adapt in these ways.

Evolution Experiments with Bacteria

Rather than performing an exhaustive review of laboratory evolution experiments, I will use several recent, influential reviews as jumping-off points for my survey of the relevant literature, starting with "Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation" (Elena and Lenski 2003). In their review, Elena and Lenski (2003) were particularly interested in laboratory studies that were "open-ended and long-term," such as they themselves had conducted. They eschewed reviewing experiments that "targeted specific, often new, metabolic functions, even though this work is of great interest" (Elena and Lenski 2003:458). Instead, for that topic, they referred readers to two previous reviews. Interestingly, even though their own review was quite up-to-date, the two reviews on specific metabolic functions that they cited—"Evolution of new metabolic functions in laboratory organisms" (Hall 1983) and *Microorganisms as Model Systems for Studying Evolution* (Mortlock 1984a)—were both about twenty years old at the time, thus indicating a lack of recent work in that subfield.

metabolism of unusual compounds

I will begin by examining the early work reported in *Microorganisms as Model Systems for Studying Evolution,* edited by Mortlock (1984a) (Table 2). Four of the first five chapters concern several species of bacteria in which cultures are evolved to metabolize an unusual chemical food source that the unevolved species cannot grow upon. In general, in each of the investigations, the first mutation selected is either one that inactivates a repressor gene, thus turning the gene it regulates into a constitutively-expressed gene, or, less often, one that increases the copy number of a particular metabolic gene. (The identities of some mutations in this early work were inferred by the authors, not demonstrated.) The amplified or deregulated gene is one that typically codes for an enzyme that metabolizes a related compound and that has a limited ability to metabolize the novel substrate. The increase in the concentration of enzyme brought about by deregulation or gene amplification allows it to metabolize enough of the novel material to permit the mutated bacterium to grow, albeit sometimes slowly. Similar results were reported by Clarke (1984) in Chapter 7, which focuses on mutations in the metabolic pathway of *Pseudomonas* that metabolizes amides. During the course of many experiments, she obtained a variety of point mutations in the regulatory protein for the amidase gene and in the promoter region for the amidase gene, as well as in the gene itself.

Constitutive mutations are loss-of-FCT mutations because the organism has lost the function of a coded feature that controls synthesis of the enzyme. Once the bacterium has mutated, allowing it to utilize a novel substrate sufficiently enough to grow, further selection can often improve the growth

TABLE 2
Work reported in Microorganisms as Model Systems for Studying Evolution *(Mortlock 1984)*

Chapter	Laboratory phenotype	Underlying mutation	Mode of adaptation	Reference
1	Xyltitol catabolism by *Klebsiella*	Inactivation of a ribitol dehydrogenase repressor gene	L	Mortlock (1984c)
	L-arabitol catabolism by *Klebsiella*	Inactivation of a ribitol dehydrogenase repressor gene	L	
	Increased catabolism of xylitol	Apparent point mutations of ribitol dehydrogenase	M	
	Increased catabolism of xylitol	Apparent alteration of ribitol dehydrogenase gene promoter	M	
	Increased catabolism of xylitol	Apparent increase in copy number of ribitol dehydrogenase gene	M	
	Xylitol transport across membrane	Deregulation of D-arabitol operon	L	
2	Increased xylitol catabolism by *Klebsiella*	Apparent increase in copy number of ribitol dehydrogenase gene	M	Hartley (1984)
	Increased xylitol catabolism by *Klebsiella*	Point mutation in ribitol dehydrogenase protein	M	
	Increased xylitol catabolism by *E. coli*	Transfer of *Klebsiella* rdh operon to *E. coli* plus gene amplification	M	
	Increased xylitol catabolism by *E. coli*	Transfer of *Klebsiella* rdh operon to *E. coli* plus point mutation	M	
4	D-lyxose isomerase activity by *Klebsiella*	Uncertain; perhaps due to constitutive D-mannose isomerase	L	Mortlock (1984b)
	D-lyxose isomerase activity in *E. coli*	Uncertain; perhaps due to constitutive D-mannose isomerase	L	
	D-arabinose isomerase activity by *Klebsiella*	Constitutive L-fucose isomerase	L	
	D-arabinose isomerase activity in *E. coli*	Point mutation-induced L-fucose pathway	M	
5	Growth on L-1,2-propanediol of *E. coli*	Possible alteration of oxidoreductase promoter	M	Lin and Wu (1984)
	Growth on D-arabinose of *E. coli*	Either constitutive expression of fucose genes, or dual control of regulator protein by L-fucose and D-arabinose	L G	
	Growth on D-arabitol of *E. coli*	Constitutive production of oxidoreductase in already-mutated strain 3	L	
	Growth on xylitol of *E. coli*	Constitutive production of D-xylose pathway	L	
	Growth on ethylene glycol of *E. coli*	Constitutive use of L-1,2-propanediol pathway	L	
6	Growth of *E. coli* on lactose (after deletion of *lacZ*)	Point mutation in cryptic β-galactosidase ebg^o; also rendered constitutive by frameshifts and IS-disruption of ebg repressor	M L	Hall (1984)
	Growth on lactulose	Second point mutation in ebg^o	M	
	Increased growth on lactulose	Point mutation in ebg repressor allowing lactulose as inducer	M	
	Growth on lactobionate	Third point mutation in ebg^o	M	

continued

TABLE 2
Continued

Chapter	Laboratory phenotype	Underlying mutation	Mode of adaptation	Reference
7	Growth in presence of butyramide of *P. aeruginosa*	Possible point mutation in regulator protein gene *amiR*	M	Clarke (1984)
	Growth in presence of succinate/lactamide	Possible point mutation in regulator protein gene *amiR* plus catabolite repression mutant	M	
	Increased production of amidase	Mutation in promoter	M	
	Growth on butyramide, valeramide, acetanilide	Point mutations in amidase	M	
8	Resistance of yeast to allyl alcohol	Point mutations in alcohol dehydrogenase	M	Wills (1984)
9	Rescue of *leuD* deletion	Availability of cryptic *leuD*-like gene	M	Kemper (1984)

Abbreviations:
L - adaptive loss of functional coded element
M - adaptive modification of function

rate through point mutations in the gene for the overproduced enzyme. These additional mutations are modification-of-function mutations, because no FCT is gained or lost. Increased-copy-number mutants are also modification-of-function mutants, because there is no evidence that any of the extra copies have diversified, which would then place it in the class of gain-of-FCT mutations.

An interesting variation on this pattern is reported in Chapter 5 by Lin and Wu (1984). Mutants of *E. coli* and *S. typhimurium* (*S. enterica* var *Typhimurium*) gain the ability to metabolize the unusual substrate D-arabinose by altering the specificity of a regulatory protein. Normally, the enzyme fucose isomerase is induced in these bacteria when some fucose enters the cell and binds to a positive regulatory protein, which then turns on the gene for the isomerase. The regulatory protein of some mutants, however, responds to both fucose and D-arabinose. It turns out that the unusual substrate D-arabinose can be metabolized by enzymes of the fucose pathway, and, because the protein has apparently gained an additional binding site for the novel substrate, the mutation is classed as gain-of-FCT.

RECRUITMENT OF CRYPTIC PROTEINS

The evolution of the ability of the bacteria described above to metabolize novel compounds depends on mutations in known proteins of known metabolic pathways. Several chapters in *Microorganisms as Model Systems for Studying Evolution* (Mortlock 1984a), however, describe examples where previously unsuspected, cryptic proteins were recruited to take over the function of a known protein whose gene had been deliberately deleted.

In Chapter 9, Kemper (1984) describes the ability of the product of a gene of *Salmonella typhimurium*, which he called *newD*, to replace a protein coded by the *leuD* gene of the leucine biosynthesis pathway. NewD ordinarily binds tightly to a larger protein subunit that Kemper dubbed SupQ, just as LeuD binds to a larger subunit, LeuC. In order to allow NewD to restore leucine biosynthetic function when *leuD* is deliberately deleted, the gene for SupQ also has to be deliberately knocked out. Kemper argues that *supQ/newD* was not simply a gene duplication of *leuC/leuD* (the genes of the leucine pathway) (Stover et al. 1988). However, at that time, sequence data was more difficult to come by than it is today. It is now known that the *S. typhimurium* genome contains a homolog of LeuD with 38%

amino acid sequence identity. It seems likely that the protein that Kemper discovered that could substitute for LeuD was this homolog. The *S. typhimurium* genome also contains a homolog of LeuC with 52% amino acid sequence identity; it seems likely that SupQ was this homolog. No such LeuC or LeuD homologs exist in *E. coli*. Because newD is a homolog of LeuD that already binds a homolog of LeuC, this is a modification-of-function event.

In Chapter 6, Hall (1984) discusses his work over the previous decade with an enzyme he dubbed Ebg, which stands for evolved β-galactosidase. He had shown that *E. coli* that had had its *lacZ* gene (which codes for a β-galactosidase) intentionally deleted could be mutated so that the altered *Ebg* protein metabolized the sugar lactose in its stead. In subsequent years, however, Hall (1995) showed unexpectedly that Ebg was an unrecognized homolog of LacZ, with 38% sequence identity. The Ebg active site was very similar to that of LacZ, and acquired its ability to metabolize lactose only when the Ebg active site mutated to one identical to that of an active LacZ. Because the unevolved enzyme could already metabolize lactose at a low rate, the adaptive point mutations that increased its activity are categorized as modification-of-function ones. Mutations in the *ebg* repressor that allowed the ebg β-galactosidase protein to be constitutively expressed are loss-of-FCT.

In further experiments, both Kemper (1984) and Hall (2003) deleted the homologous recruited protein genes *ebg* and *newD*, but were unable to recruit any of the thousands of other cellular proteins to replace them. Recently, Patrick et al. (2007) employed functional genomics tools to survey 107 single-gene knockout strains of *E. coli* that could not grow on M9-glucose medium, transfecting each one with a plasmid library that contained every *E. coli* open-reading frame. They discovered that about 20% of the knockout strains could be rescued by overexpression of at least one other *E. coli* gene. Patrick et al. (2007) found that the missing gene product and its suppressor were generally not homologs, unlike the previous cases of LacZ/Ebg and LeuD/NewD.

The work of Patrick et al. (2007) is not in itself an experimental evolution study, as all of the manipulations were performed by the investigators and no spontaneous mutations arose. Nonetheless, it does demonstrate much potential redundancy in the *E. coli* genome that may contribute to its evolvability, as the authors suggest. It may be instructive to consider how newer work would be classified if the events that Patrick et al. (2007) investigated were to occur in the wild. If, to adapt an organism in nature, a gene were initially lost by mutation, that of course would be a loss-of-FCT event. If subsequent rescue in nature required the overexpression of a potentially compensating gene, as in Patrick et al.'s (2007) experimental cases, then that could occur in at least several ways: 1) a repressor element of the compensating gene could be deleted, rendering the gene constitutive, which would be a loss-of-FCT mutation; 2) an existing promoter of the compensating gene could have its sequence altered, binding a transcription factor more strongly and thus leading to overexpression, which would be classed as a modification-of-function mutation, as an existing FCT was altered, not gained or lost; 3) a new promoter could be constructed by mutation to increase the transcription rate of the gene, which would be a gain-of-FCT mutation; 4) the compensating gene could increase in copy number, which by itself would be classed as a modification-of-function mutation; or 5) a copy of the compensating gene could diverge in sequence to more efficiently assume the activity of the deleted gene, which would be classified as a gain-of-FCT mutation. Finally, the effect of the deleted gene may be compensated in the wild not by increasing the expression of its potential complement, but instead by deleting or reducing the expression of one or more other genes, which would be an adaptive loss-of-FCT mutation (this possibility was not investigated in the excellent work of Patrick et al. [2007]). As these examples make clear, any of multiple molecular evolutionary

pathways could underlie the rescued phenotype. To understand whether an adaptation represents a gain, loss, or modification of a function, the molecular events underlying the adaptation must first be understood.

The work of Lenski and colleagues

In the 1990s, investigators began to perform the "open-ended and long-term" evolution experiments that are the main focus of a review by Elena and Lenski (2003). Certainly, the most extensive and longest running investigation has been undertaken by Lenski himself, who has been continuously growing cultures of *E. coli* in his laboratory since the late 1980s and monitoring the facets of their evolution (Lenski 2004). Diluting a portion of the previous day's culture a hundred-fold, each day can potentially see 6 to 7 new generations of bacteria ($2^{6.6} \approx 100$). At intervals, Lenski freezes a portion of a bacterial generation and stores it, so that, later, the descendant generations can be straightforwardly analyzed and compared head-to-head with their ancestors. Over the years, the number of generations has approached 50,000. With a cumulative population size of about 10^{14} cells, Lenski's investigation is large enough and long enough to give solid, reliable answers to many questions about evolution.

The results of a number of Lenski's papers over the term of the experiment are very briefly summarized in Table 3, including his early work (not discussed here) in which the underlying molecular bases of adaptive phenotypes had not yet been identified. Schneider et al. (2000) examined multiple insertion sequence (IS) mutations that became fixed in the bacterial populations within ten thousand generations. The mutations were of the kind caused by IS elements, including insertions and recombinations, that led to genetic inversions and deletions. By examining the DNA sequence of the *E. coli* in the neighborhood surrounding the IS elements, the investigators saw that several genes involved in central metabolism were knocked out, as well as some cell wall synthesis genes and several others. In subsequent work, Cooper et al. (2001) discovered that twelve of twelve cell lines showed adaptive IS-mediated deletions of their *rbs* operon, which is involved in making the sugar ribose. Thus, the adaptive mutations that were initially tracked down all involved loss-of-FCT.

Several years later, when the cultures had surpassed their 20,000th generation, Lenski's group at Michigan State brought more advanced techniques to bear on the problem of identifying the molecular changes underlying the adaptation of the *E. coli* cultures. Using DNA expression profiles, they were able to reliably track down changes in the expression of 1300 genes of the bacterium, and determined that 59 genes had changed their expression levels from the ancestor, 47 of which were expressed at lower levels (Cooper et al 2003). The authors stated that "The expression levels of many of these 59 genes are known to be regulated by specific effectors including guanosine tetraphosphate (ppGpp) and cAMP-cAMP receptor protein (CRP)" (Cooper et al 2003:1074). They also noted that the cellular concentration of ppGpp is controlled by several genes including *spoT*. After sequencing, they discovered a nonsynonymous point mutation in the *spoT* gene. When the researchers examined ten other populations that had evolved under the same conditions for 20,000 generations, they found that seven others also had fixed nonsynonymous point mutations in *spoT*, but with different substitutions than the first one that had been identified, thus suggesting that the mutations were decreasing the protein's activity.

The group then decided to concentrate on candidate genes suggested by the physiological adaptations that the cells had made over 20,000 generations. One such adaptation was a change in supercoiling density; therefore, genes affecting DNA topology were investigated (Crozat et al. 2005). Two of these genes, *topA* and *fis*, had sustained point mutations. In the case of *topA*, the mutation coded an amino acid substitution, whereas, with *fis*, a transversion had occurred at the fourth nucleotide before the starting ATG codon. The *topA*

TABLE 3
Selected results of Richard Lenski's long-term E. coli *evolution project*

Laboratory phenotype	Underlying mutation	Mode of adaptation	Reference
Mean fitness of evolved strains improves by 37% on minimal glucose medium over 2000 generations; most improvement comes early.	unknown	—	Lenski et al. (1994)
Over 10,000 generations, fitness and cell volume increase rapidly, then slow down; fitness increases in discrete steps; authors liken this to "macroevolution."	unknown	—	Lenski and Travisano (1994)
Over 10,000 generations, mutator strains evolve in 3 of 12 populations; probably hitchhike with beneficial mutations.	Probable defect in methyl-directed mismatch repair pathway	—*	Sniegowski et al. (1997)
After 6000 generations, two phenotypes, L and S, coexist in balanced polymorphism. Fitness is frequency dependent.	unknown	—	Rozen and Lenski (2000)
Study of nine IS mutants that are fixed after 10,000 generations. IS had inserted into several genes (*pykF, nadR, pbpA-rodA, hokB/sokB*) or caused rearrangements.	IS mediated insertions, deletions, rearrangements	L,L L,L	Schneider et al. (2000)
12 of 12 lines show deletion of *rbs* operon mediated by IS150. Selective value of deletion is 1%-2%.	Repeated deletion of rbs operon	L	Cooper et al. (2001)
DNA expression arrays showed 59 gene changes in parallel in two strains grown for 20,000 generations. Many genes controlled by CRP and ppGpp.	Eight separate point mutations in *spoT*, which controls cellular concentration of ppGpp	M	Cooper et al. (2003)
Sequence random genes from 10,000 and 20,000 generations	No mutations in nonmutator strains. Some point mutations in mutator strains, but no adaptive effect	—*	Lenski et al. (2003b)
Mutations affecting DNA topology: *topA* mutant decreases protein activity; *fis* mutant decreases amount of protein made	*topA* H33Y; *fis* A-C transversion four nucleotides before start ATG	M M	Crozat et al. (2005)
Four genes (*pykF, nadR, pbpA-rodA, hokB/sokB*) first identified by IS insertions are examined in other populations.	Same four genes repeatedly suffer selectable mutations at different sites	M,M M,M	Woods et al. (2006)
Comparison of protein profiles of evolved *E. coli* vs. ancestor	*malT* locus suffers multiple deletions, substitutions	L	Pelosi et al. (2006)
Citrate utilization under aerobic conditions	unknown	—	Blount et al. (2008)

*Identified or inferred mutations are not categorized because they are neutral, not adaptive.
Abbreviations:
G - adaptive gain of functional coded element
L - adaptive loss of functional coded element
M - adaptive modification of function

mutation decreased the activity of the enzyme, while the *fis* mutation decreased the amount of *fis* gene product produced. Moving the mutations into the ancestor improved its fitness in minimal glucose media.

The genes that had earlier been observed to be disrupted by IS elements, thereby leading to greater fitness in the growth medium, were then sequenced in twelve culture lines that had evolved for 20,000 generations (Woods et al. 2006). All lines were observed to have point mutations in or near one or more of the genes;

however, the genes were mutated at a variety of points. In other words, there was strong parallelism at the level of which gene it is adaptive to mutate, but little or no parallelism for a particular mutation in a particular gene. The fact that multiple point mutations in each gene could serve an adaptive role—and that disruption by IS insertion was beneficial—suggests that the point mutations were decreasing or eliminating the protein's function.

In an investigation of global protein profiles of the evolved *E. coli*, Lenski's group discovered that the MalT protein of the maltose operon had suffered mutations in 8 out of 12 strains (Pelosi et al. 2006). Several mutations were small deletions while others were point mutations, thus suggesting that decreasing the activity of the MalT protein was adaptive in minimal glucose media. This was tested by moving the mutation into the ancestral strain, which subsequently gained in fitness.

Recently, Lenski's group reported the isolation of a mutant *E. coli* that had evolved a Cit^+ phenotype. That is, the strain could grow under aerobic conditions in a culture of citrate (Blount et al. 2008). Wild *E. coli* cannot grow under such conditions, as it lacks a citrate permease to import the metabolite under oxic conditions. (It should be noted that, once inside the cell, however, *E. coli* has the enzymatic capacity to metabolize citrate.) The phenotype, whose underlying molecular changes have not yet been reported, conferred an enormous growth advantage because the culture media contained excess citrate but only limited glucose, which the ancestral bacteria metabolized. Blount et al. (2008) marshaled evidence to show that multiple mutations were needed in the population before the final mutation conferred the ability to import citrate; the activating mutation did not appear until after the 30,000th generation.

As Blount et al. (2008) discussed, several other laboratories had, in the past, also identified mutant *E. coli* strains with such a phenotype. In one such case, the underlying mutation was not identified (Hall 1982); however, in another case, high-level constitutive expression on a multicopy plasmid of a citrate transporter gene, *citT*, which normally transports citrate in the absence of oxygen, was responsible for eliciting the phenotype (Pos et al. 1998). If the phenotype of the Lenski Cit^+ strain is caused by the loss of the activity of a normal genetic regulatory element, such as a repressor binding site or other FCT, it will, of course, be a loss-of-FCT mutation, despite its highly adaptive effects in the presence of citrate. If the phenotype is due to one or more mutations that result in, for example, the addition of a novel genetic regulatory element, gene-duplication with sequence divergence, or the gain of a new binding site, then it will be a noteworthy gain-of-FCT mutation.

The results of future work aside, so far, during the course of the longest, most open-ended, and most extensive laboratory investigation of bacterial evolution, a number of adaptive mutations have been identified that endow the bacterial strain with greater fitness compared to that of the ancestral strain in the particular growth medium. The goal of Lenski's research was not to analyze adaptive mutations in terms of gain or loss of function, as is the focus here, but rather to address other longstanding evolutionary questions. Nonetheless, all of the mutations identified to date can readily be classified as either modification-of-function or loss-of-FCT.

AN AMBIGUOUS CASE

A fascinating case of concurrent gain- and loss-of FCT can be seen in the work of Zinser et al. (2003). The authors examined a mutant of *E. coli* that thrived when the culture was starved, but that was outcompeted under conditions of growth. They demonstrated that the result depended on two mutational events: 1) an initial insertion of an IS element between the transcriptional promoter of the *cstA* gene and a nearby CRP box that regulated it; and 2) inversion of the sequence flanked by the new IS element and another IS element ~60 kb distant and upstream of the *ybeJ-gltJKL-ybeK* operon. The inversion brought the *ybeJ* gene of the mutant, which encodes a

portion of an amino acid transporter, under the influence of the CRP box that previously regulated the *cstA* gene and that encodes a starvation-inducible oligopeptide permease. In the new arrangement, the *ybeJ* gene was actively transcribed, while the *cstA* gene no longer was. Interestingly, the *ybeJ* gene product helped the bacterium grow under conditions of starvation as expected, but the authors demonstrated that expression of the *cstA* gene itself would also be beneficial under the same conditions. Nonetheless, the inverted arrangement was selected because it possessed the greatest *net* fitness.

This example points to unavoidable ambiguity in the classification of some adaptive evolutionary events. One can view the adaptation investigated by Zinser et al. (2003) in several ways: 1) as a modification-of-function event (i.e., the insertion of a duplicate IS element) followed, as a result of the inversion, by the loss of a functional coded element controlling the *cstA* gene (the CRP box) and the gain of a functional coded element controlling the *ybeJ* gene (the same CRP box); or 2) since no FCT was gained or lost in the inversion but rather was simply rearranged, the inversion sums to a second modification-of-function mutation, albeit one in which there is now a potential for a recombination switch under conditions of starvation vs. growth.

Evolution Experiments with Viruses

Because of their ability to rapidly reproduce to enormous population sizes, viruses, like bacteria, are well-suited to experimental evolutionary studies, and such studies have been well-reviewed in recent years (Elena and Lenski 2003; Elena and Sanjuan 2007; Bull and Molineux 2008; Elena et al. 2008). Several general features distinguish viruses from bacteria and alter evolutionary expectations. 1) Viral genomes are usually much smaller than bacterial genomes, often by a thousand-fold, and one consequence of this is that mutations are much easier to identify by sequencing for viruses than for bacteria. Another consequence, however, is that viruses do not have deep reserves of homologous genes or other genetic material to recruit as bacteria do, except that which comes from their cellular hosts. 2) Because of their small genome size, most of their genome is required for their life cycle; relatively few viral genes are dispensable. 3) For RNA viruses, the mutation rate is greatly increased, as much as a million-fold over that of cells. This allows various mutation "solutions" to a selective challenge to present themselves much more rapidly than for cells.

Table 4 summarizes evolution experiments with viruses discussed in the reviews above or in related papers. Most of the experiments discussed below disturbed virus growth in various ways and followed their evolutionary reaction to the disturbance. The papers are categorized by the general selective pressure that was brought to bear on the viruses.

Recovery from multiple bottleneck-induced deleterious mutations

Two laboratories (Burch and Chao 1999; Bull et al. 2003) subjected two different viruses—an RNA virus (φ6) and a DNA virus (T7)—to repeated population bottlenecks (i.e., severe reductions in population size that cause the loss of genetic variation) of just one or a few viruses. Because of the high mutation rate of RNA viruses, each virus has a high probability of having a mutation, most likely deleterious, and, through many bottlenecks, the researchers were able to accumulate many mutations in the virus. Because of the lower rate of mutation in DNA viruses, the workers using phage T7 added a mutagen to the growth medium in order to increase its mutation rate.

Burch and Chao (1999) showed that φ6 suffered a ten-fold loss of fitness after 20 bottleneck generations of repeatedly picking individual plaques grown on a plate of bacterial host, and then using the plaque to seed a new plate. When subsequently allowed to grow at larger population sizes, the phage was able to recover fitness by compensatory mutations; however, because the molecular natures of the mutations were not determined, they cannot be classified here.

Bull et al. (2003) were able to accumulate hundreds of mutations in bacteriophage T7, and to reduce its fitness to about 0.0001% of the wild type. During the recovery phase of the experiment, two separate populations of phage gained a factor of about 100-fold and 10,000-fold in fitness during 115 and 82 passages, respectively. Both populations were continuing to recover fitness when the experiment was halted. Among the mutations accumulated during the bottleneck phase of the experiment were several small and large deletions in nonessential genes, one small insertion, and about 400 point mutations, almost equally divided between missense mutations and silent mutations. There were also several nonsense mutations in nonessential genes.

In the recovery phase, there were also several large and small deletions, as well as a few dozen point mutations, most of which were missense mutations that were likely subject to positive selection. Recovery-phase mutations involving large or frame-shifting deletions in genes can, with reasonable confidence, be classified as loss-of-FCT. Selectable point mutations are more difficult to classify definitively. None were reported to have formed identifiable, functional coded elements, such as new promoter sites or protein processing sites. However, the possibility remains that some new functional coded element that is not readily identifiable from its nucleic acid sequence may have arisen in one or more cases. To rigorously test such a possibility, careful biochemical analysis would be required. In the absence of evidence for gain-of-FCT, selectable point mutations are tentatively classified as "modification-of-function" in Table 4.

adapting to a new environment

A second category of selective pressure is that of viruses placed in new environments. Cuevas et al. (2002) described the competition of two neutrally-marked strains of the single-stranded RNA vesicular stomatitis virus (VSV) when grown on the same medium and the same cells, but at varying population sizes. There was remarkable parallelism in the exact nucleotide changes for the two strains over multiple experiments, thus suggesting that the mutations compensate for growth in the particular environment. All mutations were nucleotide substitutions; no deletions or insertions were seen. Twelve substitutions were synonymous, fifteen were nonsynonymous, and three were in intergenic regions.

Bull et al. (1997) have studied the adaptation of replicate lineages of the single-stranded DNA virus φX174 to high temperature (43.5°C) on several hosts. In one study of 119 total substitutions at 68 sites in the virus, over half were identical with substitutions in other lineages, and no deletions or insertions were reported. Adaptation improved fitness in primary lines by about 4000-fold with *S. typhimurium* as host, and by about 100,000-fold with *E. coli* as host. In another study, Wichman et al. (1999) grew φX174 on *S. typhimurium* at elevated temperature (43.5°C). Half of the mutations in one line also appeared in a second line grown under identical conditions. Most changes were amino acid substitutions, but one common mutation was an intergenic deletion of 27 nucleotides. Since no discrete molecular features were attributed to the deleted intergenic region, that deletion can be considered a modification-of-function—no functional coded element was lost. No discrete new molecular features were identified within mutated coding regions, but further biochemical analysis would be required to determine if a functional coded feature—such as a new protein binding site—were gained. In the absence of such evidence, selectable point mutations are tentatively classified as "modification-of-function" in Table 4.

adapting to a new host

Another category of selective pressure is that of viruses forced to adapt to a novel host organism in the laboratory. The adaptation of a virus to a new host in the wild can, of course, be a major—even catastrophic—epidemiological event, as the examples of HIV and, potentially, H1N1 show. Nonetheless, as emphasized throughout this review, to understand how adaptation proceeds, the phenotypic and molecular aspects must be kept separate. The ability of a virus to grow on an alternate host is a phenotype that may have a variety of

TABLE 4
Summary of selected viral evolution experiments

Investigator action	Underlying mutation	Mode of adaptation	Reference
Viruses subject to drift			
Bacteriophage φ6 subject to drift in bottleneck populations	Recovery by growth in larger populations showed many small steps; no sequence data	—	Burch and Chao (1999)
Bacteriophage T7 repeatedly passed through single-virus population bottlenecks	Several dozen point mutations; several large and small deletions in recovery phase	L,M	Bull et al. (2003)
Viruses adapting to new environments			
Two neutrally-marked strains of VSV competed at different population ratios	Extensive parallel convergence at the nucleotide and amino acid level; some co-variations noted	M	Cuevas et al. (2002)
Bacteriophage φX174 adapted to high temperature growth	Multiple convergent substitution mutations across several lineages	M	Bull et al. (1997)
Two populations of bacteriophage φX174 adapted to high temperature and novel host	Multiple convergent mutations, mostly substitutions, in two populations accumulated in different orders	M	Wichman et al. (1999)
Viruses adapting to new hosts			
φX174 adapted to growth on *Eschericia* and *Salmonella* hosts	Amino acid substitutions at 5 sites in major capsid attachment protein affect host preference	M	Crill et al. (2000)
φX174 adapted to growth on *E. coli* strains differing in lipopolysaccharide attachment site	Amino acid substitutions at 10 sites in major capsid attachment protein affect host preference	M	Pepin et al. (2008)
φ6 adapted to growth on 15 *Pseudomonas syringae* strains	Amino acid substitutions at 9 sites in spike attachment protein affect host range	M	Duffy et al. (2006)
Broad-specificity φ6 adapted to growth on single *Pseudomonas syringae* strain	Single substitution in spike attachment protein gene narrows host range	M	Duffy et al. (2007)
Viruses manipulated to be defective			
Deletion of 19 intercistronic nucleotides from RNA virus MS2 containing Shine-Dalgarno sequence and two hairpins	One revertant deleted 6 nucleotides; another duplicated an adjoining 14-nucleotide sequence; missing functional coded elements substantially restored	G,G	Olsthoorn and van Duin (1996)
4 nucleotide deletion in lysis gene of MS2	Reading frame restored by deletions, insertions	G,G	Licis and van Duin (2006)
Randomized operator sequence of MS2	Modification-of-function mutations appeared, but operator hairpin not restored	M	Licis et al. (2000)
Deletion of T7 viral ligase gene	Loss-of-FCT and modification-of-function mutations in several genes; new ligase activity not obtained from host	L,M	Rokyta et al. (2002)
Bacteriophage T7 forced to replicate using T3 RNA polymerase	9 of 16 promoters altered sequence; several missense modification-of-function changes	M	Bull et al. (2007)
T7 RNA polymerase gene moved toward end of viral genome	Loss of one *E. coli* polymerase termination site; facilitated rearrangement of RNAP gene closer to the beginning of the viral genome	L,M	Springman et al. (2005)

continued

TABLE 4
Continued

Investigator action	Underlying mutation	Mode of adaptation	Reference
Viruses forced to be interdependent			
Separate viruses, f1 and IKe, engineered to carry distinct antibiotic resistance markers	In media containing both antibiotics, phages co-packaged into f1 protein coats; two-thirds of IKe genome deleted, second antibiotic gene captured by f1	L,M G	Sachs and Bull (2005)

Abbreviations:
G - adaptive gain of functional coded element
L - adaptive loss of functional coded element
M - adaptive modification of function

possible underlying molecular bases. As reviewed below and as observed in experimental evolutionary studies, viruses have presumably adapted to new hosts by modification-of-function mutations rather than by gain- or loss-of-FCT ones.

Many experimental studies have been performed investigating the adaptation of viruses to new hosts, and several recent papers illustrate the topic of gain- or loss-of-FCT that this review is concerned with. Commonly, although not exclusively (see Yin and Lomax 1983; Subbarao et al. 1993; Agudelo-Romero et al. 2008), mutations in coat proteins underlie the phenotypic variation. Crill et al. (2000) studied the adaptation of φX174 to *Eschericia* and *Salmonella* hosts. Over 11 days of selection on a host, up to 28 substitutions, as well as a 2-base insertion and 27-base deletion, were observed in all genes except one. The authors determined that nonsynonymous nucleotide substitutions at 5 sites in the major capsid protein gene F affected host preference. The altered amino acid residues lie on the virion surface and correlate strongly with attachment efficiency but, intriguingly, are at sites on the protein other than the putative carbohydrate binding site identified by X-ray crystallographic studies (McKenna et al. 1994). Pepin et al. (2008) revisited host adaptation of φX174, but in a more specific context. They examined the adaptation of the virus to three strains of *E. coli* that differed from each other only by a single sugar group in the lipopolysaccharide site of attachment. Of 21 mutations, all were nucleotide substitutions; seven of those were synonymous, and, of the remaining 14, 10 occurred in the major capsid protein gene F. Since the substitutions apparently did not form a new FCT or cause the loss of one, but simply modified the strength of attachment, they are modification-of-function mutations.

Duffy et al. (2006) examined the evolution of the RNA bacteriophage φ6 on 15 *Pseudomonas syringae* strains. Nine mutations were isolated that affected host range, and all were amino acid substitutions in the host-attachment spike-protein P3. One virus strain carrying a substitution of P3 that broadened the φ6 host range was used for further studies (Duffy et al. 2007). When the mutant strain was grown exclusively on an alternate permissive host, it acquired a further single nucleotide substitution in the gene for the spike protein. The substitution conferred increased growth on the alternate host and re-narrowed the host range.

The work reviewed above presents an opportunity to further clarify the classification categories introduced in this paper. Host adaptations could potentially be classified in several ways. Some investigators might categorize the adaptations reported by Duffy et al. (2007), for instance, as gains and losses of the function of binding very specific host ligands. As I discussed earlier in this paper, however, I classify shifts in ligand preference as "modification-of-function" events if the binding occurs at the same site on the protein, as in the examples discussed here. From

this view, the functional coded element is the protein binding site, which can be modified either by amino acid substitution directly at the site or elsewhere in the protein. Considered in this way, a functional site is not gained or lost as host preference is altered, but is instead modified.

RECOVERY FROM INTRODUCED DEFECTS: THE RNA BACTERIOPHAGE MS2

Another category of selective pressure is the intentional introduction by the investigator of defined defects in a virus. Olsthoorn and van Duin (1996) deleted a 19-nucleotide intercistronic segment of RNA phage MS2 between the stop codon for the maturation protein gene and the start codon for the coat protein gene; this segment contains the Shine-Dalgarno sequence and usually forms two hairpins. The synthesis of the coat protein is especially sensitive to the presence of a hairpin containing its initiator codon, and the investigators reported that the titer of the altered phage dropped ten orders of magnitude (Olsthoorn and van Duin 1996). Two separate kinds of revertants were isolated—one that had deleted 6 nucleotides that coded the final two residues of the maturation protein, and another that contained a duplication of an adjoining 14-nucleotide sequence. Remarkably, the deletion revertant substantially restored the coat protein initiator hairpin and the Shine-Dalgarno sequence and, subsequently, approximated them more closely by accumulating several point mutations. The duplication revertant partially restored both native hairpins and the Shine-Dalgarno sequence, and subsequently accumulated point mutations to more closely approximate the starting phage. The titers of the reconstructed phages were very similar to the starting phage.

Both the 6-nucleotide deletion and the 14-nucleotide duplication are gain-of-FCT mutations since they both produced new coded molecular features in the virus that did not exist in the immediate precursor; that is, the virus that sustained the deliberately-deleted 19 nucleotides. The 6-nucleotide deletion did not inactivate the maturation protein, so it is not a loss-of-FCT mutation. The 14-nucleotide duplication led to the formation of new functional coded elements (it did not simply repeat pre-existing elements), so it is not just a modification-of-function mutation. The subsequent point mutations that improve activity are modification-of-function.

Another paper from van Duin's laboratory investigated the evolution of phage MS2 in which a 4-nucleotide sequence was deleted from a 38-nucleotide intercistronic region between the coat protein gene and the replicase gene (Licis and van Duin 2006). The gene for the lysis protein spans the intercistronic region and overlaps into the coat protein and replicase genes, so that the deletion introduces a frameshift into it. Furthermore, the deletion destabilizes the operator loop, which is needed to bind a dimer of coat protein, initiating capsule formation, as well as regulating expression of the replicase gene. The mutant phage decreased in fitness by a factor of 10^6.

Most initial revertants at relatively low titers first repaired the frameshift by inserting a nucleotide in this region. One deleted two more nucleotides, which also restored the correct reading frame. Several further revertants were obtained by increasing the initial population size of mutant viruses and passaging the virus. In particular, insertion revertants were obtained, one of which (revIN4) had inserted four nucleotides exactly at the site where they initially had been deleted. These are all gain-of-FCT mutations. After further passages, most revertants acquired point mutations that repaired damaged features, including the operator loop. Only one revertant achieved a fitness equal to the wild type phage after multiple passages, and this was a derivative of revIN4 that went on to revert completely to the wild type.

A third paper from van Duin's lab concerning MS2 examined the evolutionary consequences of partially randomizing the sequence of the operator hairpin (Licis et al. 2000). The fitness of the initial mutants decreased anywhere from 10^3 to 10^7. Most revertants failed to reconstruct the operator. The two best revertants, with 2% and

20% of the fitness of the wild type, respectively, did not restore the operator hairpin. Thus, the mutations in this study are modification-of-function ones.

Recovery from introduced defects: the DNA bacteriophage T7

Bull, Wichman, Molineux, and colleagues (Bull et al. 1997, 2003, 2007; Bull and Molineux 2008) introduced large changes into the double stranded DNA bacteriophage T7 to observe how it might recover. The DNA polymerase of T7 is much more accurate than the RNA polymerase of MS2, the system used by van Duin. Nonetheless, they were able to show considerable evolution of the DNA phage.

In one experiment, the gene for DNA ligase from phage T7 was intentionally deleted and the phage's evolutionary recovery was observed (Rokyta et al. 2002). The authors noted that the deletion exerts little selective pressure if the phage is grown on a host that contains an active cellular ligase. They grew the phage in a cell line whose ligase had been previously mutated to decrease its activity. Initially, the mutant T7 had very low fitness, but it eventually recovered to approximately 20% of wild type. Of the ten mutations recovered, five were determined to compensate for the loss of ligase activity. Three mutations were in genes, like ligase, involved in DNA metabolism: single amino acid substitutions in phage DNA polymerase and primase-helicase, and formation of a stop codon at position 26 of phage endonuclease. One mutation was a deletion of 18 nucleotides in gene *1.5*—a gene of unknown function. The final mutation was the insertion of a nucleotide into gene *2.8*, causing a frameshift. The authors determined, however, that the likely beneficial effect was to increase the transcription of neighboring gene *3*. These can be classified as loss-of-FCT and modification-of-function mutations. Rokyta et al. (2002) remarked that they initially expected the phage to acquire a new ligase activity, either by recombination or by gene duplication and divergence by point mutation. If that had happened, it would have been a notable gain-of-FCT mutation; however, ligase activity was not recovered.

In another study, the gene for RNA polymerase (RNAP) was deleted from bacteriophage T7, and the modified phage was used to infect cells that harbored a plasmid carrying a gene for bacteriophage T3 RNAP (Bull et al. 2007). The bacteriophages T3 and T7 are closely related, and the two polymerases differ in sequence by only 18%. T3 RNAP transcribes from T7 promoters only at low levels, but a single nucleotide change in the T7 consensus promoter sequence allows significant activity by the T3 polymerase. Bull et al. (2007) used this system to follow the expected evolution in T7 promoters to accommodate T3 polymerase. The initial fitness of the phage was a factor of about 10^{10} less than wild type, but, after adaptation, the fitness approached that of wild type. Sequencing revealed that 9 of 16 T7 promoters had acquired mutations—mostly nucleotide substitutions and one deletion. Seven missense mutations accumulated in proteins that do not bind RNAP, and these were determined to be generally beneficial and not strictly compensatory for T3 RNAP. Several missense mutations also occurred in some genes to compensate for T3 RNAP. All of the mutations can be categorized as modification-of-function.

In a third study, the order of genes in bacteriophage T7 was rearranged, with the T7 RNAP gene, which normally enters the cell early, placed at an ectopic position near the opposite end of its genome, thus enabling it to enter the cell late. Ordinarily, the phage polymerase would assist in pulling the phage DNA into the cell by its transcriptional activity. By placing the gene toward the further end of the phage, that activity was precluded, and the fitness of the phage dropped nine orders of magnitude. Springman et al. (2005) studied three revertants. Two revertants recovered only a comparatively small portion of the fitness they lost, increasing by about 1000-fold. Both of these revertants abolished an *E. coli* polymerase termination site, thereby allowing the cellular polymerase, which usually only transcribes early genes of the

phage, to continue down to the ectopic site of the phage polymerase gene and transcribe it. These are loss-of-FCT mutants, as they have lost a coded regulatory site. A third revertant recovered a much larger amount of fitness, increasing by about a million-fold. The originally constructed mutant that gave rise to this revertant had had some flanking DNA sequence of the phage RNA polymerase gene purposely left behind in its original position, in order to allow for subsequent recombination of the ectopic gene at its original site. After about five passages, the expected recombination occurred with a concomitant large increase in fitness. This recombination can be categorized as a modification-of-function mutation, because coded features were rearranged, not gained or lost.

SIMILAR VIRUSES CONSTRAINED TO OCCUPY THE SAME CELL

Sachs and Bull (2005) examined the evolution of two filamentous single-stranded DNA bacteriophages of *E. coli*, f1 and IKe. The two phages share 55% nucleotide identity and contain ten similar genes. The investigators inserted different antibiotic resistance genes into each phage, so that cells growing in a culture containing both antibiotics would need to be infected by both phages in order to survive. Similarly, one kind of phage could only survive in a cell if the other kind of phage were there also. Over the course of 50 passages, the fitness of the phages grew as they increasingly came to be packaged together into single f1 coats. During the evolutionary process, f1 acquired eight point mutations and IKe acquired nine point mutations, as well as two large deletions. A deletion of 206 nucleotides of noncoding DNA around passage 15 gave the largest increase in fitness, and allowed virtually all IKe to be packaged into an f1 coat. Subsequently, around passage 40, IKe suffered a massive deletion, losing two-thirds of its genome and retaining only the antibiotic resistance gene, plus two genes controlling copy number and co-interference. This resulted in a small, additional increase in phage fitness.

One can view these adaptations from several points of view. When considering the mutations individually, the large deletion in IKe is a loss-of-FCT mutation, and the smaller deletion in IKe and the point mutations in both phages are likely simple modification-of-function. When considering the phages as competing organisms, however, IKe was essentially destroyed in the evolutionary process, and f1acquired the use of its critical antibiotic resistance gene. This could be considered a pseudo-horizontal-gene-transfer event from IKe to f1 and, therefore, a gain-of FCT event for f1.

MUTATIONAL SATURATION

The longer an evolution experiment is run, and the larger the population of microbes it harbors, the greater the chance for mutations to appear that are rare and particularly beneficial. Lenski's (2004) long-term experiment with *E. coli*, which is approaching 50,000 generations and a cumulative population size of about 10^{14} organisms, is the clear leader in such projects.

There are several other such experiments to note, which involve particularly large numbers of microbes or generations. Wichman et al. (2005) described the adaptive evolution of the bacteriophage φX174 over the course of 13,000 generations, which, for the rapidly reproducing microbe, required only 180 days. Although it was not reported, the total population size of phage over the course of the experiment likely reached 10^{13}-10^{14}. Couñago et al. (2006) replaced the essential gene for adenylate kinase in *Geobacillus stearothermophilus*—a moderate thermophile—with that of *Bacillus subtilis*—a mesophile—which they then grew in a turbidostat at increasing temperatures. Over the course of 1500 generations, they isolated six thermostable mutants of the enzyme—one single point mutant and five double point mutants derived from the single mutant. The cumulative number of bacteria over the term of the experiment was 10^{13}-10^{14}.

Table 5 shows the calculated saturation of genomes with mutations in these large-scale experiments, which can be estimated for *E. coli*, *G. stearothermophilus*, and φX174 by using the value for the mutation rate of

TABLE 5
Saturation coverage of mutations in three microbes

Microbe	Generations	Cumulative population size N	Estimated mutation rate per nucleotide	Average fold-saturation with mutations
E. coli	50,000	10^{13}–10^{14}	7×10^{-10}	7×10^{3}–7×10^{4}
G. stearothermophilus	1,500	10^{13}–10^{14}	5×10^{-10}	5×10^{3}–5×10^{4}
φX174	13,000	10^{13}–10^{14}	6×10^{-7}	6×10^{6}–6×10^{7}

Drake et al. (1998): approximately 0.003 per genome per generation for DNA-based microbes. On average in the population during the course of long-term experiments, each nucleotide in each genome is expected to be substituted from 7×10^3 to 6×10^7 times; of course, the exact rate of substitution could vary considerably from nucleotide to nucleotide. Thus these experiments plumb the depths of what adaptive mutation can accomplish in these systems in a single step, or in a series of single, related steps—such as Couñago et al.'s (2006) double mutants and Blount et al.'s (2008) Cit^+ phenotype.

Implications

circumstances of gain- and loss-of-fct mutations

Tables 2 through 4 summarize results from the past four decades of evolutionary experiments with microbes, categorizing the adaptive mutations as loss- or gain-of-FCT, or modification-of-function. As can be seen, only one of the adaptive mutations from bacteria (Tables 2 and 3) is gain-of-FCT, yet several adaptive mutations from experiments with viruses (Table 4) belong to that class as well. Why the difference? One reason may be that, except for the capture of an antibiotic resistance gene by phage f1, the viral gain-of-FCT mutations all reconstruct functional coded elements that had been deliberately removed from—or rendered inactive in—the ancestral virus, thus restoring pieces of a once-integrated system. That is, they began at a point that was known to be able to benefit from a gain-of-FCT mutation. The lysis gene of MS2 was rescued from a frameshift deletion mutation by adding or deleting additional nucleotides to restore the correct reading frame. The mutant that achieved the greatest fitness was the one that reverted completely to the wild type sequence. Similarly, the very deleterious effects of a 19-nucleotide deletion of MS2 containing important functional coded elements (several hairpins and the Shine-Dalgarno sequence) was overcome by gain-of-FCT mutations that restored those same elements to a greater or lesser degree.

Considering the time- and population-scale constraints of the experiments, it is not surprising that, when large experimental deletions were constructed that removed the coding sequences of whole genes (rather than just frame shift mutations or short control elements), the deleted genes were not restored. However, it was surprising that more modest adaptive gain-of-FCT mutations were not seen either. The removal of T7 ligase resulted in point mutations and deletions in other genes involved in DNA metabolism, which are loss-of-FCT and modification-of-function mutations. Intentional deletion of the gene for T7 RNA polymerase and infection of a cell harboring a T3 polymerase gene yielded mutations that apparently strengthened weak T3 promoters, which are modification-of-function changes. Rearrangement of the order of bacteriophage T7 genes, thereby decreasing its fitness, did not provoke the evolutionary construction of new coded control elements. Rather, one existing element was lost (an *E. coli* polymerase termination site) and the gene order reverted, guided by flanking DNA that Springman et al. (2005) intentionally left in the viral DNA sequence.

A second reason why several adaptive gain-of-FCT viral mutations but few bacte-

rial ones were identified might be that RNA viruses have increased mutation rates compared to those of bacteria. This factor may also partially explain the occurrence of gain-of-FCT mutations in experiments with the RNA virus MS2, but not in experiments with the DNA virus T7, which has a much lower mutation rate.

Tables 2–4 also show that adaptive loss-of-FCT mutations are less common for small viruses than for bacteria. No such mutations were reported for the small bacteriophages φ6, φX174, VSV, or MS2 in the reviewed laboratory experiments. The small DNA virus IKe suffered a massive loss when two-thirds of its genome was deleted. However, it had been co-infected into cells with the similar small DNA virus f1, which likely could have supplied all the functions that IKe lost. No adaptive loss-of-FCT mutations were sustained by f1 itself. The larger DNA virus T7 did acquire several loss-of-FCT mutations in various experiments. It seems likely that the general explanation for this pattern is that the smaller the virus, the smaller the percentage of its genome that is dispensable for its basic life cycle. The larger a viral genome, the greater the percentage that is dispensable. For larger viral genomes, and for cells, more coded elements are available for adaptation by loss-of-FCT mutations.

THE FIRST RULE OF ADAPTIVE EVOLUTION

As seen in Tables 2 through 4, the large majority of experimental adaptive mutations are loss-of-FCT or modification-of-function mutations. In fact, leaving out those experiments with viruses in which specific genetic elements were intentionally deleted and then restored by subsequent evolution, only two gain-of-FCT events have been reported: the development of the ability of a fucose regulatory protein to respond to D-arabinose (Lin and Wu 1984), and the antibiotic gene capture by f1 (Sachs and Bull 2005). Why is this the case? One important factor is undoubtedly that the rate of appearance of loss-of-FCT mutations is much greater than the rate of construction of new functional coded elements. Suppose an adaptive effect could be secured by diminishing or removing the activity of a certain protein. If the gene for the protein were, for instance, 1000 nucleotides in length, then there would be numerous targets of opportunity for a loss-of-FCT mutation. The deletion of any single nucleotide in the coding sequence would alter the reading frame and likely destroy or greatly diminish protein activity. The insertion of a nucleotide anywhere in the coding sequence would do the same. Longer insertions or deletions would commonly have the same effect, as would alteration of a codon from sense to nonsense. All these would fall into the category of loss-of-FCT mutations.

Nucleotide substitutions resulting in missense mutations, although not likely to completely eliminate protein activity, are very likely to diminish activity to a greater or lesser extent, as, in multiple experiments, the majority of amino acid substitutions have been found to decrease a protein's activity (Reidhaar-Olson and Sauer 1988; Bowie and Sauer 1989; Lim and Sauer 1989; Bowie et al. 1990; Reidhaar-Olson and Sauer 1990; Axe et al. 1996; Huang et al. 1996; Sauer et al. 1996; Suckow et al. 1996). Although, if residual protein activity remained, these would be categorized as modification-of-function mutations under the accounting system used here—that is, the partial diminishment of the function provides the adaptive effect. (A caveat: in the particular case of microbes that were first allowed to accumulate deleterious mutations and then recover, such as in Bull et al. (2003), it is likely that many adaptive point mutations are compensatory for initial deleterious mutations and therefore increase (mutated) protein function. Biochemical analysis would be needed in order to prove that protein function increased or decreased and was the basis of the adaptive effect.) If the basic point mutation rate per nucleotide per generation were 10^{-9}, then the rate of appearance of an adaptive loss-of-FCT mutation would likely be on the order of 10^{-6}, because of the many ways possible to decrease the activity of a protein. Indeed, an adaptive mutation rate of $\sim 10^{-5}$ was recently measured in *E. coli* (Perfeito et al. 2007).

Contrast this with a situation in which a particular nucleotide in a gene for a certain protein has to be mutated in order to gain an adaptive effect. (The particular mutation can be thought to help code for

a new binding site in the protein or to construct a new genetic control element from a sequence that was already a near-match, in addition to other possibilities. These would be gain-of-FCT mutations.) If the basic point mutation rate per nucleotide were 10^{-9}, then that would also be the rate of appearance of the beneficial mutation. Even if there were several possible pathways by which to construct a gain-of-FCT mutation, or several possible kinds of adaptive gain-of-FCT features, the rate of appearance of an adaptive mutation that would arise from the diminishment or elimination of the activity of a protein is expected to be 100–1000 times the rate of appearance of an adaptive mutation that requires specific changes to a gene.

This reasoning can be concisely stated as what I call "The First Rule of Adaptive Evolution":

> *Break or blunt any functional coded element whose loss would yield a net fitness gain.*

It is called a "rule" in the sense of being a rule of thumb. It is a heuristic, useful generalization, rather than a strict law; other circumstances being equal, this is what is usually to be expected in adaptive evolution. Since the rule depends on very general features of genetic systems (that is, the mutation rate and the probability of a loss-of-FCT versus a gain-of-FCT mutation), it is expected to hold for organisms as diverse as viruses, prokaryotes, and multicellular eukaryotes. It is called the "first" rule because the rate of mutations that diminish the function of a feature is expected to be much higher than the rate of appearance of a new feature, so adaptive loss-of-FCT or modification-of-function mutations that decrease activity are expected to appear first, by far, in a population under selective pressure.

Illustrations of the First Rule

The first rule, gleaned from laboratory evolution experiments, can be used to interpret data from evolution in nature, including human genetic mutations in response to selective pressure by malaria (Table 1). Hundreds of distinct mutations are known that diminish the activities of G6PD or the α- or β- chains of hemoglobin, leading to thalassemia. Yet it is estimated that the gain-of-FCT mutation leading to sickle hemoglobin has arisen independently only a few times, or perhaps just once, within the past 10,000 years (Cavalli-Sforza et al. 1994). Thus, loss-of-FCT adaptive mutations in this situation appeared several orders of magnitude more frequently than did a gain-of-FCT mutation. Nonetheless, the sickle hemoglobin mutation did arise and spread in a regional population. Therefore, if a gain-of-FCT mutation such as the sickle gene has a sufficiently large selection coefficient, then, even though adaptive loss-of-FCT mutations arrive more rapidly and in greater numbers, it is possible for the gain-of-FCT mutation to outcompete them.

Another illustration from nature of the first rule can be seen in the adaptive evolution of the plague bacterium *Yersinia pestis* over the past 1,500–20,000 years (Wren 2003). A likely evolutionary scenario for its great virulence is that the plague bacterium serially acquired several plasmids that conferred upon it the ability to be transferred by flea bite (Carniel 2003). The 101 kb pFra plasmid carries the *yplD* gene, which codes for a phospholipase D that is necessary for the survival of the bacterium in the flea proventriculus. The 9.6 kb pPla plasmid codes for a plasminogen activator, which allows the bacterium to move in its host unhindered by blood clotting. The acquisitions of these genes are gain-of-FCT events. *Y. pestis* has also subsequently lost a large number of chromosomal genes—a general estimate is that 150 genes have been lost (Carniel 2003; Chain et al. 2006). A number of discarded genes have activities in other *Yersinia* species that allow pathogen-host adhesion. The plague bacterium has also acquired hundreds of missense mutations (Carniel 2003). The discarded genes are of course loss-of-FCT mutations, and many missense mutations are likely to diminish activity. Thus, the organism adapted relatively quickly to its new lifestyle—first made possible by several gain-of-FCT events—through much more numerous loss-

of-FCT and modification-of-function mutations.

Two recent laboratory studies also illustrate the first rule. Ferenci (2008) discusses the adaptive value of mutations to the gene for a specialized *E. coli* RNA polymerase σ factor that contributes to the general stress response. The author observed that "*rpoS* mutations occurred, and indeed spread at rapid rates within a few generations of establishing glucose-limited chemostats", and also that "The majority of *rpoS* mutations accumulating in glucose-limited cultures are loss-of-function mutations with little or no residual RpoS protein. . . . The mutations include stop codons, deletions, insertions as well as point mutations" (Ferenci 2008:447). A second study showed that the loss of mating genes during asexual growth in *Saccharomyces cerevisiae* provided a 2% per-generation growth-rate advantage (Lang et al. 2009). The authors noted that "in bacteria, gratuitous gene expression reduces growth rate. . . . We suspect that the cost of gene expression is not specific to bacterial enzymes or genes in the yeast mating pathway, but rather reflects a universal cost of gene expression and that this cost must be borne in all environments where the gene is expressed" (Lang et al. 2009:5758). Thus in any environment in which a gene becomes superfluous or, more generally, in any environment where its loss would yield a net fitness gain, the frequent mutations occurring in the population that tend to eliminate the functional coded element will turn adaptive.

HOW FREQUENTLY ARE LOSS-OF-FCT AND GAIN-OF-FCT MUTATIONS ADAPTIVE?

Although the rates of appearance of loss-of-FCT and modification-of-function mutations that degrade protein activity are always expected to be much greater than the rate of appearance of gain-of-FCT mutations, a separate question concerns what fractions of the mutations in those categories are adaptive. That is, although loss-of-FCT mutations might appear rapidly, if they do not yield a selective benefit—i.e., if they are not adaptive—then they will not usually spread in a population. In the same vein, a gain-of-FCT mutation may eventually appear that builds some new genetic feature such as a transcription factor binding site, for instance; yet if the feature is not adaptive within the organism's genetic context, it will not be selected (Stone and Wray 2001).

These are empirical questions that are difficult to answer conclusively. However, the data and experiments discussed in this review offer some insights. In the most open-ended laboratory evolution experiment (Lenski 2004), in which no specific selection pressure was intentionally brought to bear, all of the adaptive mutations that have been so far identified have either been loss-of-FCT or modification-of-function mutations, and there is strong reason to believe that most of the modification-of-function mutations diminished protein activity. Except in cases where specific genetic features were first removed, as well as in the case of antibiotic gene capture by f1, all adaptive mutations in laboratory evolution experiments with viruses seem to be loss-of-FCT or modification-of-function mutations. Thus, in general laboratory evolutionary situations (that is, where a microorganism was under a general selective pressure rather than a specific one), adaptive loss-of-FCT or modification-of-function mutations were always available. This cannot be said for gain-of-FCT mutations.

One objection might be that the above examples are artificial. They concern laboratory evolution, and it may be that diminished expression of some pre-existing, commonly-held genes gives an organism an advantage over its conspecifics in such a constant environment, but not in the varied and changing environments of nature. It is true that the laboratory is an artificial environment, and the opportunities for some events that occur at irregular intervals in the wild—such as lateral gene transfer—are essentially nonexistent. This is clearly an area that needs to be addressed in more detail. Further laboratory evolution studies with more complex environments or cultures of mixed species would serve to shed light on the extent to which

such factors affect opportunities for adaptation by gain- or loss-of-FCT mutations. Nonetheless, results arguably similar to those that have been seen in laboratory evolution studies to date have also been seen in nature, such as the loss of many genes by *Yersinia pestis* (after, of course, the acquisition of new genetic material in the form of several plasmids), and the loss-of-FCT mutations that have spread in human populations in response to selective pressure from malaria. A tentative conclusion suggested by these results is that the complex genetic systems that are cells will often be able to adapt to selective pressure by effectively removing or diminishing one or more of their many functional coded elements.

A second possible objection is that many of the reviewed experiments were conducted on comparatively small populations of microbes for relatively short periods of time, so that although loss-of-FCT and modification-of-function mutations might be expected to occur, there simply was not much opportunity to observe gain-of-FCT mutations. After all, one certainly would not expect new genes with complex new properties to arise on such short time-scales. Although it is true that new complex gain-of-FCT mutations are not expected to occur on short time-scales, the importance of experimental studies to our understanding of adaptation lies elsewhere. Leaving aside gain-of-FCT for the moment, the work reviewed here shows that organisms do indeed adapt quickly in the laboratory—by loss-of-FCT and modification-of-function mutations. If such adaptive mutations also arrive first in the wild, as they of course would be expected to, then those will also be the kinds of mutations that are first available to selection in nature. This is a significant addition to our understanding of adaptation. That knowledge is also a necessary prerequisite for elucidating the nature of long-term adaptation, as consideration of how long-term adaptation proceeds must take into account how organisms adapt in the short-term.

Furthermore, although complex gain-of-FCT mutations likely would occur only on long time-scales unavailable to laboratory studies, simple gain-of-FCT mutations need not take nearly as long. As seen in Table 1, a gain-of-FCT mutation in sickle hemoglobin is triggered by a simple point mutation, which helps code for a new protein binding site. It has been estimated that new transcription-factor binding sites in higher eukaryotes can be formed relatively quickly by single point mutations in DNA sequences that are already near matches (Stone and Wray 2001). In general, if a sequence of genomic DNA is initially only one nucleotide removed from coding for an adaptive functional element, then a single simple point mutation could yield a gain-of-FCT. As seen in Table 5, several laboratory studies have achieved thousand- to million-fold saturations of their test organisms with point mutations, and most of the studies reviewed here have at least single-fold saturation. Thus, one would expect to have observed simple gain-of-FCT adaptive mutations that had sufficient selective value to outcompete more numerous loss-of-FCT or modification-of-function mutations in most experimental evolutionary studies, if they had indeed been available.

A third objection could be that the time and population scales of even the most ambitious laboratory evolution experiments are dwarfed when compared to those of nature. It is certainly true that, over the long course of history, many critical gain-of-FCT events occurred. However, that does not lessen our understanding, based upon work by many laboratories over the course of decades, of how evolution works in the short term, or of how the incessant background of loss-of-FCT mutations may influence adaptation.

Conclusion

Adaptive evolution can cause a species to gain, lose, or modify a function. Therefore, it is of basic interest to determine whether any of these modes dominates the evolutionary process under particular circumstances. The results of decades of experi-mental laboratory evolution studies strongly suggest that, at the molecular level, loss-of-FCT and diminishing modification-of-function adaptive mutations predominate. In retrospect, this

conclusion is readily understandable from our knowledge of the structure of genetic systems, and is concisely summarized by the first rule of adaptive evolution. Evolution has myriad facets, and this one is worthy of some notice.

REFERENCES

Agudelo-Romero P., Carbonell P., Perez-Amador M. A., Elena S. F. 2008. Virus adaptation by manipulation of host's gene expression. *PLoS ONE* 3:e2397.

Axe D. D., Foster N. W., Fersht A. R. 1996. Active barnase variants with completely random hydrophobic cores. *Proceedings of the National Academy of Sciences USA* 93:5590–5594.

Behe M. J., Englander S. W. 1979. Mixed gelation theory. Kinetics, equilibrium and gel incorporation in sickle hemoglobin mixtures. *Journal of Molecular Biology* 133:137–160.

Blount Z. D., Borland C. Z., Lenski R. E. 2008. Historical contingency and the evolution of a key innovation in an experimental population of *Escherichia coli*. *Proceedings of the National Academy of Sciences USA* 105:7899–7906.

Bowie J. U., Sauer R. T. 1989. Identifying determinants of folding and activity for a protein of unknown structure. *Proceedings of the National Academy of Sciences USA* 86:2152–2156.

Bowie J. U., Reidhaar-Olson J. F., Lim W. A., Sauer R. T. 1990. Deciphering the message in protein sequences: tolerance to amino acid substitutions. *Science* 247:1306–1310.

Bull J. J., Badgett M. R., Rokyta D., Molineux I. J. 2003. Experimental evolution yields hundreds of mutations in a functional viral genome. *Journal of Molecular Evolution* 57:241–248.

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.

Bull J. J., Molineux I. J. 2008. Predicting evolution from genomics: experimental evolution of bacteriophage T7. *Heredity* 100:453–463.

Bull J. J., Springman R., Molineux I. J. 2007. Compensatory evolution in response to a novel RNA polymerase: orthologous replacement of a central network gene. *Molecular Biology and Evolution* 24:900–908.

Burch C. L., Chao L. 1999. Evolution by small steps and rugged landscapes in the RNA virus φ6. *Genetics* 151:921–927.

Carniel E. 2003. Evolution of pathogenic *Yersinia*: some lights in the dark. Pages 3–12 in *The Genus* Yersinia*: Entering the Functional Genomic Era*, edited by M. Skurnik et al. New York: Kluwer Academic/Plenum.

Cavalli-Sforza L. L., Menozzi P., Piazza A. 1994. *The History and Geography of Human Genes*. Princeton (NJ): Princeton University Press.

Chain P. S., Hu P., Malfatti S. A., Radnedge L., Larimer F., Vergez L. M., Worsham P., Chu M. C., Andersen G. L. 2006. Complete genome sequence of *Yersinia pestis* strains Antiqua and Nepal516: evidence of gene reduction in an emerging pathogen. *Journal of Bacteriology* 188:4453–4463.

Clarke P. H. 1984. Amidases of *Pseudomonas aeruginosa*. Pages 187–232 in *Microorganisms as Model Systems for Studying Evolution*, edited by R. P. Mortlock. New York: Plenum Press.

Cooper T. F., Rozen D. E., Lenski R. E. 2003. Parallel changes in gene expression after 20,000 generations of evolution in *Escherichia coli*. *Proceedings of the National Academy of Sciences USA* 100:1072–1077.

Cooper V. S., Schneider D., Blot M., Lenski R. E. 2001. Mechanisms causing rapid and parallel losses of ribose catabolism in evolving populations of *Escherichia coli* B. *Journal of Bacteriology* 183:2834–2841.

Couñago R., Chen S., Shamoo Y. 2006. In vivo molecular evolution reveals biophysical origins of organismal fitness. *Molecular Cell* 22:441–449.

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

Crozat E., Philippe N., Lenski R. E., Geiselmann J., Schneider D. 2005. Long-term experimental evolution in *Escherichia coli*. XII. DNA topology as a key target of selection. *Genetics* 169:523–532.

Cuevas J. M., Elena S. F., Moya A. 2002. Molecular basis of adaptive convergence in experimental populations of RNA viruses. *Genetics* 162:533–542.

Darwin C. 1859. *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. London (UK): John Murray.

Drake J. W., Charlesworth B., Charlesworth D., Crow J. F. 1998. Rates of spontaneous mutation. *Genetics* 148:1667–1686.

Duffy S., Burch C. L., Turner P. E. 2007. Evolution of host specificity drives reproductive isolation among RNA viruses. *Evolution* 61:2614–2622.

Duffy S., Turner P. E., Burch C. L. 2006. Pleiotropic costs of niche expansion in the RNA bacteriophage φ6. *Genetics* 172:751–757.

Elena S. F., Agudelo-Romero P., Carrasco P., Codoner F. M., Martin S., Torres-Barcelo C., Sanjuan R. 2008. Experimental evolution of plant RNA viruses. *Heredity* 100:478–483.

Elena S. F., Lenski R. E. 2003. Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. *Nature Reviews Genetics* 4:457–469.

Elena S. F., Sanjuan R. 2007. Virus evolution: Insights from an experimental approach. *Annual Review of Ecology, Evolution, and Systematics* 38:27–52.

Ferenci T. 2008. The spread of a beneficial mutation in experimental bacterial populations: the influence of the environment and genotype on the fixation of *rpoS* mutations. *Heredity* 100:446–452.

Flint J., Hill A. V., Bowden D. K., Oppenheimer S. J., Sill P. R., Serjeantson S. W., Bana-Koiri J., Bhatia K., Alpers M. P., Boyce A. J. 1986. High frequencies of α-thalassaemia are the result of natural selection by malaria. *Nature* 321:744–750.

Forget B. G. 1998. Molecular basis of hereditary persistence of fetal hemoglobin. *Annals of the New York Academy of Sciences* 850:38–44.

Friedman M. J. 1978. Erythrocytic mechanism of sickle cell resistance to malaria. *Proceedings of the National Academy of Sciences USA* 75:1994–1997.

Hall B. G. 1982. Chromosomal mutation for citrate utilization by *Escherichia coli* K-12. *Journal of Bacteriology* 151:269–273.

Hall B. G. 1983. Evolution of new metabolic functions in laboratory organisms. Pages 234–257 in *Evolution of Genes and Proteins*, edited by M. Nei and R. K. Koehn. Sunderland (MA): Sinauer Associates.

Hall B. G. 1984. The evolved β-galactosidase system of *Escherichia coli*. Pages 165–186 in *Microorganisms as Model Systems for Studying Evolution*, edited by R. P. Mortlock. New York: Plenum Press.

Hall B. G. 1995. Evolutionary potential of the *ebgA* gene. *Molecular Biology and Evolution* 12:514–517.

Hall B. G. 2003. The EBG system of *E. coli*: origin and evolution of a novel β-galactosidase for the metabolism of lactose. *Genetica* 118:143–156.

Hartley B. S. 1984. Experimental evolution of ribitol dehydrogenase. Pages 23–54 in *Microorganisms as Model Systems for Studying Evolution*, edited by R. P. Mortlock. New York: Plenum Press.

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

Kemper J. 1984. Gene recruitment for a subunit of isopropylmalate isomerase. Pages 255–284 in *Microorganisms as Model Systems for Studying Evolution*, edited by R. P. Mortlock. New York: Plenum Press.

Kennedy J. R. 2002. Modulation of sickle cell crisis by naturally occurring band 3 specific antibodies—a malaria link. *Medical Science Monitor* 8:HY10–HY13.

Lang G. I., Murray A. W., Botstein D. 2009. The cost of gene expression underlies a fitness trade-off in yeast. *Proceedings of the National Academy of Sciences USA* 106:5755–5760.

Lenski R. E. 2004. Phenotypic and genomic evolution during a 20,000-generation experiment with the bacterium *Escherichia coli*. *Plant Breeding Reviews* 24:225–265.

Lenski R. E., Simpson S. C., Nguyen T. T. 1994. Genetic analysis of a plasmid-encoded, host genotype-specific enhancement of bacterial fitness. *Journal of Bacteriology* 176:3140–3147.

Lenski R. E., Travisano M. 1994. Dynamics of adaptation and diversification: a 10,000-generation experiment with bacterial populations. *Proceedings of the National Academy of Sciences USA* 91:6808–6814.

Lenski R. E., Winkworth C. L., Riley M. A. 2003. Rates of DNA sequence evolution in experimental populations of *Escherichia coli* during 20,000 generations. *Journal of Molecular Evolution* 56:498–508.

Licis N., Balklava Z., van Duin J. 2000. Forced retroevolution of an RNA bacteriophage. *Virology* 271: 298–306.

Licis N., van Duin J. 2006. Structural constraints and mutational bias in the evolutionary restoration of a severe deletion in RNA phage MS2. *Journal of Molecular Evolution* 63:314–329.

Lim W. A., Sauer R. T. 1989. Alternative packing arrangements in the hydrophobic core of lambda repressor. *Nature* 339:31–36.

Lin E. C. C., Wu T. T. 1984. Functional divergence of the L-fucose system in mutants of *Escherichia coli*. Pages 135–164 in *Microorganisms as Model Systems for Studying Evolution*, edited by R. P. Mortlock. New York: Plenum Press.

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. *Journal of Molecular Biology* 237:517–543.

Modiano D., Luoni G., Sirima B. S., Simpore J., Verra F., Konate A., Rastrelli E., Olivieri A., Calissano C., Paganotti G. M., D'Urbano L., Sanou I., Sawadogo A., Modiano G., Coluzzi M. 2001. Haemoglobin C protects against clinical *Plasmodium falciparum* malaria. *Nature* 414:305–308.

Mortlock R. P. 1984a. *Microorganisms as Model Systems for Studying Evolution*. New York: Plenum Press.

Mortlock R. P. 1984b. The development of catabolic pathways for the uncommon aldopentoses. Pages 109–134 in *Microorganisms as Model Systems for Studying Evolution*, edited by R. P. Mortlock. New York: Plenum Press.

Mortlock R. P. 1984c. The utilization of pentitols in studies of the evolution of enzyme pathways. Pages 1–21 in *Microorganisms as Model Systems for Studying Evolution*, edited by R. P. Mortlock. New York: Plenum Press.

Olsthoorn R. C., van Duin J. 1996. Evolutionary reconstruction of a hairpin deleted from the genome of an RNA virus. *Proceedings of the National Academy of Sciences USA* 93:12256–12261.

Patrick W. M., Quandt E. M., Swartzlander D. B.,

Matsumura I. 2007. Multicopy suppression underpins metabolic evolvability. *Molecular Biology and Evolution* 24:2716–2722.

Pelosi L., Kuhn L., Guetta D., Garin J., Geiselmann J., Lenski R. E., Schneider D. 2006. Parallel changes in global protein profiles during long-term experimental evolution in *Escherichia coli*. *Genetics* 173: 1851–1869.

Pepin K. M., Domsic J., McKenna R. 2008. Genomic evolution in a virus under specific selection for host recognition. *Infection, Genetics and Evolution* 8:825–834.

Perfeito L., Fernandes L., Mota C., Gordo I. 2007. Adaptive mutations in bacteria: high rate and small effects. *Science* 317:813–815.

Pogo A. O., Chaudhuri A. 2000. The Duffy protein: a malarial and chemokine receptor. *Seminars in Hematology* 37:122–129.

Pos K. M., Dimroth P., Bott M. 1998. The *Escherichia coli* citrate carrier CitT: A member of a novel eubacterial transporter family related to the 2-oxoglutarate-malate translocator from spinach chloroplasts. *Journal of Bacteriology* 180:4160–4165.

Reidhaar-Olson J. F., Sauer R. T. 1988. Combinatorial cassette mutagenesis as a probe of the informational content of protein sequences. *Science* 241: 53–57.

Reidhaar-Olson J. F., Sauer R. T. 1990. Functionally acceptable substitutions in two α-helical regions of lambda repressor. *Proteins* 7:306–316.

Rokyta D., Badgett M. R., Molineux I. J., Bull J. J. 2002. Experimental genomic evolution: extensive compensation for loss of DNA ligase activity in a virus. *Molecular Biology and Evolution* 19:230–238.

Rozen D. E., Lenski R. E. 2000. Long-term experimental evolution in *Escherichia coli*. VIII. Dynamics of a balanced polymorphism. *American Naturalist* 155:24–35.

Ruwende C., Khoo S. C., Snow R. W., Yates S. N., Kwiatkowski D., Gupta S., Warn P., Allsopp C. E., Gilbert S. C., Peschu N. 1995. Natural selection of hemi- and heterozygotes for G6PD deficiency in Africa by resistance to severe malaria. *Nature* 376:246–249.

Sachs J. L., Bull J. J. 2005. Experimental evolution of conflict mediation between genomes. *Proceedings of the National Academy of Sciences USA* 102: 390–395.

Sauer R. T., Milla M. E., Waldburger C. D., Brown B. M., Schildbach J. F. 1996. Sequence determinants of folding and stability for the P22 Arc repressor dimer. *FASEB Journal* 10:42–48.

Schneider D., Duperchy E., Coursange E., Lenski R. E., Blot M. 2000. Long-term experimental evolution in *Escherichia coli*. IX. Characterization of insertion sequence-mediated mutations and rearrangements. *Genetics* 156:477–488.

Sniegowski P. D., Gerrish P. J., Lenski R. E. 1997. Evolution of high mutation rates in experimental populations of *E. coli*. *Nature* 387:703–705.

Springman R., Badgett M. R., Molineux I. J., Bull J. J. 2005. Gene order constrains adaptation in bacteriophage T7. *Virology* 341:141–152.

Stone J. R., Wray G. A. 2001. Rapid evolution of *cis*-regulatory sequences via local point mutations. *Molecular Biology and Evolution* 18:1764–1770.

Stott R. 2003. *Darwin and the Barnacle*. New York: W. W. Norton.

Stover C. K., Kemper J., Marsh R. C. 1988. Molecular cloning and characterization of *supQ/newD*, a gene substitution system for the *leuD* gene of *Salmonella typhimurium*. *Journal of Bacteriology* 170:3115–3124.

Subbarao E. K., London W., Murphy B. R. 1993. A single amino acid in the PB2 gene of influenza A virus is a determinant of host range. *Journal of Virology* 67:1761–1764.

Suckow J., Markiewicz P., Kleina L. G., Miller J., Kisters-Woike B., Muller-Hill B. 1996. Genetic studies of the Lac repressor. XV: 4000 single amino acid substitutions and analysis of the resulting phenotypes on the basis of the protein structure. *Journal of Molecular Biology* 261:509–523.

Verra F., Simpore J., Warimwe G. M., Tetteh K. K., Howard T., Osier F. H., Bancone G., Avellino P., Blot I., Fegan G., Bull P. C., Williams T. N., Conway D. J., Marsh K., and Modiano D. 2007. Haemoglobin C and S role in acquired immunity against *Plasmodium falciparum* malaria. *PLoS ONE* 2:e978.

Weatherall D. J., Miller L. H., Baruch D. I., Marsh K., Doumbo O. K., Casals-Pascual C., Roberts D. J. 2002. Malaria and the red cell. *Hematology: American Society of Hematology Education Program* 2002:35–57.

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.

Wichman H. A., Millstein J., Bull J. J. 2005. Adaptive molecular evolution for 13,000 phage generations: a possible arms race. *Genetics* 170:19–31.

Wills C. 1984. Structural evolution of yeast alcohol dehydrogenase in the laboratory. Pages 233–254 in *Microorganisms as Model Systems for Studying Evolution*, edited by R. P. Mortlock. New York: Plenum Press.

Woods R., Schneider D., Winkworth C. L., Riley M. A., Lenski R. E. 2006. Tests of parallel molecular evolution in a long-term experiment with *Escherichia coli*. *Proceedings of the National Academy of Sciences USA* 103:9107–9112.

Wren B. W. 2003. The *Yersiniae*–a model genus to

study the rapid evolution of bacterial pathogens. *Nature Reviews Microbiology* 1:55–64.

Yin F. H., Lomax N. B. 1983. Host range mutants of human rhinovirus in which nonstructural proteins are altered. *Journal of Virology* 48:410–418.

Zinser E. R., Schneider D., Blot M., Kolter R. 2003. Bacterial evolution through the selective loss of beneficial genes. Trade-offs in expression involving two loci. *Genetics* 164:1271–1277.

Handling Editor: Daniel E. Dykhuizen