Journal of Molecular Microbiology and Biotechnology

J Mol Microbiol Biotechnol 2011;21:20–35
DOI: 10.1159/000332747

Published online: January 13, 2012

Fitness Effects of Mutations in Bacteria

Isabel Gordo[a] Lilia Perfeito[b] Ana Sousa[a]

[a]Instituto Gulbenkian de Ciência, Oeiras, Portugal; [b]Institut für Theoretische Physik, Universität zu Köln, Köln, Germany

Key Words
Evolution · Mutations · Population genetics

Abstract
Mutation is the primary source of variation in any organism. Without it, natural selection cannot operate and organisms cannot adapt to novel environments. Mutation is also generally a source of defect: many mutations are not neutral but cause fitness decreases in the organisms where they arise. In bacteria, another important source of variation is horizontal gene transfer. This source of variation can also cause beneficial or deleterious effects. Determining the distribution of fitness effects of mutations in different environments and genetic backgrounds is an active research field. In bacteria, knowledge of these distributions is key for understanding important traits. For example, for determining the dynamics of microorganisms with a high genomic mutation rate (mutators), and for understanding the evolution of antibiotic resistance, and the emergence of pathogenic traits. All of these characteristics are extremely relevant for human health both at the individual and population levels. Experimental evolution has been a valuable tool to address these questions. Here, we review some of the important findings of mutation effects in bacteria revealed through laboratory experiments.

Introduction

Evolution, the process by which the frequencies of alleles change generation after generation, is affected by different mechanisms. The primary is mutation upon which genetic drift and natural selection act. Genetic drift is the non-deterministic (stochastic) change in frequency of mutations in any natural population. Because the number of individuals in a population is limited, not all get to reproduce every generation, even if they are genetically identical. This leads to a stochastic sampling process from parents to offspring that continuously occurs. It can have a substantial effect when population bottlenecks occur, such as those characteristic of bacteria when colonizing new hosts. If a mutational change affects a phenotype that influences the survival (e.g. resistance to stress) or reproduction (e.g. increased growth rate) of its carrier, then that mutation is said to have a fitness effect. The fitness effect of mutations is key in determining their fate and their frequency distribution in natural populations, which greatly affects the level of genetic diversity observed. The extent to which an organism can maintain a phenotype (related to fitness) in the face of

All authors contributed equally to this study.

KARGER

Fax +41 61 306 12 34
E-Mail karger@karger.ch
www.karger.com

1464–1801/11/0212–0020$38.00/0

Accessible online at:
www.karger.com/mmb

Isabel Gordo
Instituto Gulbenkian de Ciência
Rua da Quinta Grande, 6
PT–2780-156 Oeiras (Portugal)
Tel. +351 21 440 7915, E-Mail igordo@igc.gulbenkian.pt

mutation has been called mutational robustness [Wagner, 2005]. If the mean fitness effect of random mutations is small, then the organism is said to be more robust than if it is large. In this review, we focus on the efforts that have been done, using laboratory experiments, to infer mutation effects in bacteria. It is also possible to infer effects of mutations segregating in natural populations, using population genetics theory and assuming different models [Ashley et al., 2010; Eyre-Walker and Keightley, 2007], but we will not review those here. Several studies performed in different laboratory conditions indicate that mutations present a wide variation in effects. These experiments show that the mean effect of deleterious mutations is around a few percent per genome per generation, and they also indicate that beneficial effects tend to be larger when populations evolve in stressful environments. The various studies typically employ different experimental fitness assays and different modeling frameworks to estimate fitness effects, which present difficulties in answering a fundamental question: how does the distribution of fitness effects depend on the environment and genetic background?

Molecular Basis of Mutations

All mutations comprise changes in the sequence of DNA. These changes can have different sizes and origins. In brief, mutations can range from the substitution of a single base pair to deletions and insertions of DNA fragments, inversions, translocations and duplications. Duplications are particularly important since they provide the raw material for the evolution of new genes by subsequent mutation [Hughes, 1994]. Transposable elements such as insertion sequences (IS) and transposons also play a very important role in bacterial mutagenesis since they probably lead to mutations at high rates [Papadopoulos et al., 1999]. Transposons can carry many types of information, like antibiotic resistance genes, whereas IS contain information that relates exclusively to their transposition and regulation. The latter elements are very abundant in bacterial genomes and contribute in high degree to spontaneous mutagenesis [Schneider and Lenski, 2004; Schneider et al., 2000]. Upon insertion, IS elements can abolish gene function, activate the expression of cryptic genes or modify the expression of active genes (they often carry promoter sequences). They can also act as substrates for recombination, leading to chromosomal rearrangements such as inversions or deletions. The rate at which different IS elements transpose in the genome has not been directly estimated; however, based on a number of studies it is thought to be high. For example, Papadopoulos et al. [1999] determined the genetic fingerprint for 8 IS elements in 12 lines of *Escherichia coli* evolved for 10,000 generations. They estimated an average of around 10^{-3} to 10^{-4} changes per generation. Importantly, without knowing the fitness effect of these IS movements, the rate of transposition is hard to estimate, so an accurate estimate of the spontaneous rate of transposition is still missing. A different approach to this problem was taken by Hall [1999], who selected for inactivation of the *ebg*R gene in *Escherichia coli* and then looked at the mutational spectra of the mutants. *ebg*R codes for the repressor of the *ebg* operon, so in a situation where it is essential for the cell to express the operon but no inhibition of the repressor occurs, only mutants that abolish *ebg*R function are viable. He estimated that inactivation of the target gene mediated by IS elements constituted 61% of all the events detected. In contrast, single point mutations, regularly considered as the most abundant type of mutations, constituted only 28%.

The classical way to determine the mutational spectrum, in any circumstance, is by reducing selection to a minimum allowing an unbiased detection of all arising mutations. A common way of reaching this goal is a protocol that has become known as 'mutation accumulation' (MA), where mutations fix at the same rate as they arise (see fig. 1). Andersson and Lind [2008] did such an experiment in *Salmonella typhimurium*. Distinct DNA repair mutant clones and a wild-type strain were subject to daily bottlenecks for 5,000 generations. Subsequently, whole-genome sequencing of two of the lines was performed. Their analysis, focused exclusively on base pair substitutions, allowed them to conclude that there was no mutational bias from GC to AT in the wild type, whereas in the two DNA repair mutants 98% of the mutations were of these types. These results suggest that the deletion of the repair systems and loss of these control mechanisms could be responsible for a rapid reduction in the GC content. An estimate of the genomic mutation rate in wild-type *Salmonella* was also obtained. Dividing the number of mutations detected by the number of elapsed generations yielded an estimate of 3.4×10^{-3} base pair substitutions per genome per generation, or 7×10^{-10} substitutions per base pair per generation. This estimate is consistent with what is known as Drake's rule. Drake [1991] estimated a mean value for the mutation rate per replication of 3.3×10^{-3} based on DNA phages, *E. coli*, *Saccharomyces cerevisiae* and *Neurospora crassa*, a rule that is also supported now by *Salmonella*.

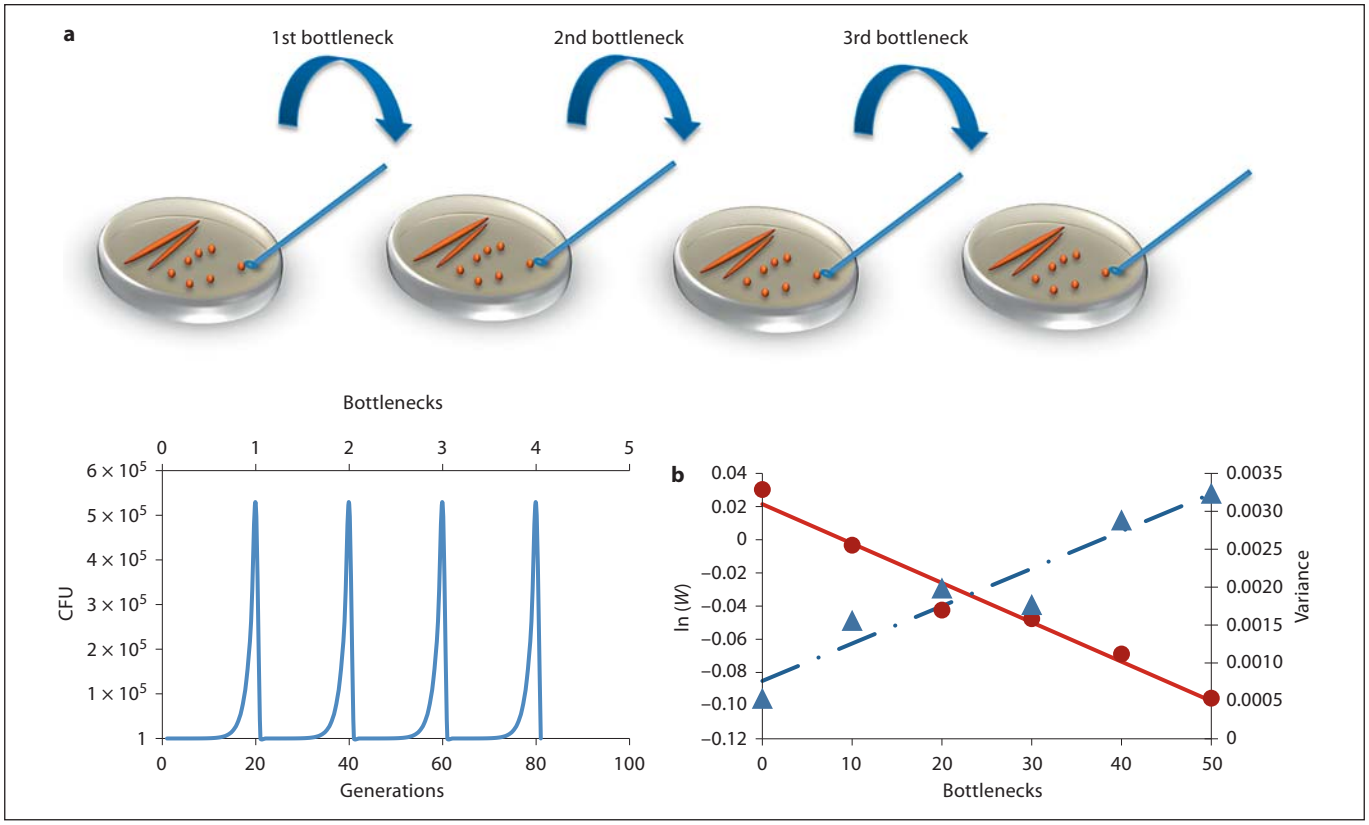

Fig. 1. MA in bacteria. In this kind of experiments, the evolving population is periodically reduced to one individual. In the case of bacteria, this is accomplished by propagating the cells in solid media where the development of isolated colonies (typically assumed to have originated from a single individual) is possible. **a** The typical MA experiment starts from a single colony, which is spread in the agar using an inoculation loop, followed by a period of cell growth (incubation), and then a new colony is randomly picked to continue the experiment. These are the steps that take place between any two bottlenecks. **b** Average fitness and variance of 50 lines of *E. coli* in a MA experiment performed by Trindade et al. [2010]. Typically, fitness (○) declines whereas variance (△) increases as the number of generations increases. Trindade et al. [2010] followed the model by Gordo and Dionisio [2005], which assumed that the number of newly arising deleterious mutations is Poisson distributed with mean U_d and that each each mutation causes a fitness decline of s_d, and estimated these parameters according to the following equations:

$$U_d = \frac{m_1}{\left(1-\left(1-s_d\right)^G\right)} \quad \text{and} \quad s_d = \frac{m_2}{m_1}$$

where m_1 is the slope of the natural logarithm of the mean fitness (of all 50 lines) with bottleneck number, and m_2 is the slope of the natural logarithm of

$$F = \frac{\overline{w_i}^2}{\overline{w_i^2}}$$

with bottleneck number i; w is the mean fitness of each line at bottleneck i.

Besides the endogenous origin of variability (all the above-described mechanisms), prokaryotes have other ways of increasing diversity, which comprise a variety of promiscuous gene transfer systems, such as conjugation, transformation and transduction. These processes allow bacteria to expand their genome by acquiring genes from other bacteria, a mechanism that became known as horizontal gene transfer (HGT) [Ochman et al., 2000]. These are highly pervasive and can involve distantly related species. Bacterial plasmids are very important players in HGT, specifically conjugative plasmids, which are those that code for the functions necessary for their own or other plasmids' transference from one cell to the other. Plasmids do not accommodate any of the core fundamental genes for growth or multiplication of bacterial cells. Yet sometimes they may be essential, since they can carry a wide range of genes such as antibiotic and toxic heavy metal resistance genes, genes that code

for extra nutritional abilities or virulence factors that allow the cell to colonize specific habitats. Plasmids can be as big as about half the size of the genome of their host. For example pSymB, which is a natural plasmid of *Sinorhizobium meliloti* strain 1021, has a size of 1.68 Mb, whereas the bacteria's chromosome is 3.65 Mb. Every time plasmids are transmitted by conjugation, both the recipient and the donor cell end up with a copy of the plasmid. Conjugation can also be mediated by transposons; in this case these are called conjugative transposons [for a detailed description, see Salyers et al. 1995]. These kinds of transposable elements, like conjugative plasmids, also code for the necessary functions that promote their transmission. First, they excise from the genome, circularize and then either integrate back into the genome, into a plasmid or are transmitted by conjugation to a new cell in a process similar to plasmid conjugation. In general, transposons can lead to gene movement between two plasmids, a plasmid and the chromosome or even between two chromosomes from different cells; most of the times without requiring DNA homology between the element and the site of insertion [Craig, 1997].

The other two mechanisms that promote the acquisition of exogenous DNA are transformation and transduction. Briefly, the first process relates to the ability of some bacteria, naturally competent, to incorporate DNA from the environment into their cytoplasm and integrate it into their own genome. The second process is due to specific phages that sometimes, instead of their own genome, encapsidate fragments of the host DNA and release it when they infect a new cell.

Recently, a new way of acquiring genetic elements was reported [Ben-Yehuda and Dubey, 2011]. The authors observed the formation of some structures that they coined nanotubes between cocultured bacteria. These nanotubes were able to bridge bacteria of the same species *(Bacillus subtilis)* and from different species with dissimilar cell wall composition *(B. subtilis* and *E. coli)*. They constitute a way of exchanging different cytoplasmatic contents ranging from proteins to mRNAs and non-conjugative plasmids. As long as the right selective pressure is applied, the plasmids acquired in this way can be maintained in the population, since they can be replicated and vertically inherited. The same does not apply to the functions coded by the transmitted mRNAs. Nevertheless, these can be actively translated for a period long enough to make the difference between life and death, for instance if they provide antibiotic resistance enzymes when the respective antibiotic is present. This mechanism does not account for long-lasting increased genetic diversity but may be important for the transient expression of essential functions for the cell.

Estimation of Fitness

Fitness is a measure of natural selection, and plays a central role in evolutionary theory. It represents an individual's ability to survive and reproduce in its environment [Orr, 2009] and it is quantified by the number of offspring an organism produces throughout its life. For organisms with binary fission such as many species of bacteria, it is not straightforward to apply this definition. The maximum number of offspring per individual is fixed (2) and natural selection acts either to increase survival and/or to reduce generation time. For bacteria with symmetrical division, it is also difficult to define what one individual is, since the individual ceases to exist as soon as it reproduces. It then becomes useful to define fitness for the genotype and estimate it from the genotype's growth rate. The effect of a mutation on fitness is the mutation's selection coefficient (typically called s in evolutionary genetics literature: s_d for deleterious and s_b for beneficial effects), which is a key parameter in determining its fate in a population. Mutations with very small s values are greatly influenced by the stochastic effect of genetic drift: small deleterious mutations can fix in populations, and small beneficial mutations are easily lost. Mutations with minute fitness effects (compared with the size of the population) accumulate at the rate at which they appear [Kimura, 1983]. In contrast, mutations with strong detrimental effects are quickly eliminated, and highly beneficial mutations enjoy large probabilities of fixation.

The key step in estimating the effect of mutations empirically is to construct an assay to measure bacterial fitness. Two main types of assays have been generally used: competitive fitness and growth curve assays. In the first case, an experiment is performed where wild-type and mutant bacteria compete in a given environment and the comparison of the frequency of each genotype before and after competition leads to a measure of the mutation's selection coefficient (fig. 2). In the second case, the wild-type and the mutant bacteria grow in isolation over a period of time, and the dynamics of growth is determined. Typically, from the maximum growth rates of both genotypes a selection coefficient can be estimated. The sensitivities of each assay are different. Competitive fitness is more robust to small changes in the experimental condi-

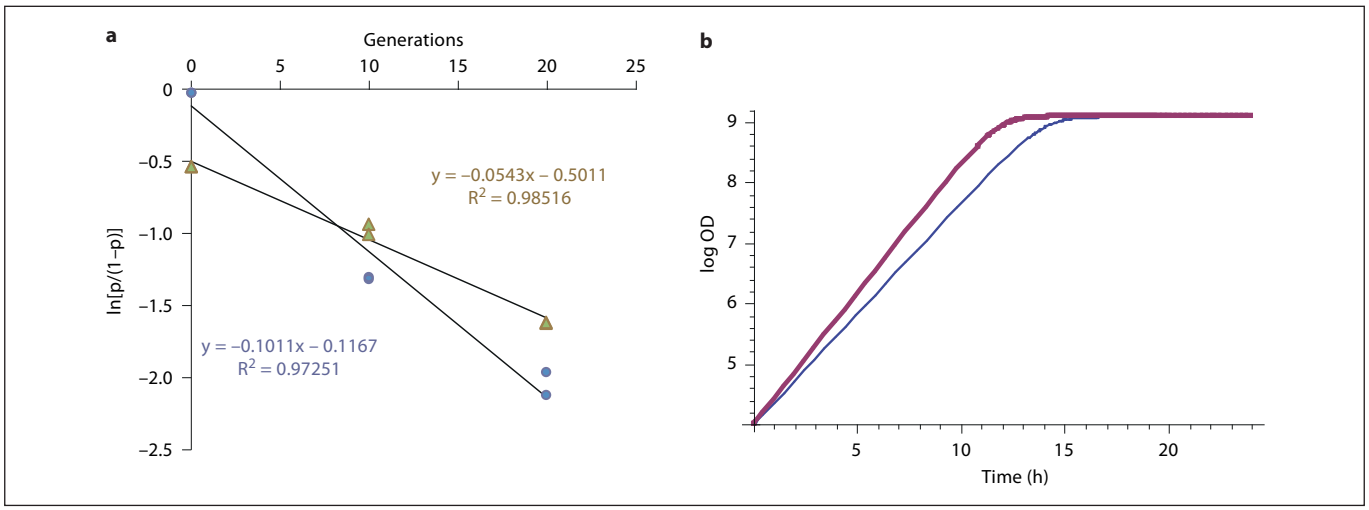

Fig. 2. Common fitness assays to estimate selection coefficients. **a** Competitive fitness assay: a wild-type strain competes with a mutant strain in a given environment. Typically, the mutant genotype has a starting frequency, p, close to 0.5, but the assay can be performed with different initial frequencies to test for frequency-dependent selection. Over a short period of competition, a few tens of generations, the numbers of both genotypes are determined and the change in frequency estimated. The expected change in frequency per generation is $\Delta p = sp(1 - p)$, which means that from the slope of $\ln[p(1 - p)]$ with time a value of s can be estimated. The figure is an example of two different mutants of *E. coli*, each carrying a distinct mutation that confers resistance to the antibiotic rifampicin; one of the mutants (△) has an s_d of 5%, the other mutation (○) is more costly with an s_d of 10%. Each mutant (carrying a YFP fluorescence marker) was put in competition with a sensitive *E. coli* (carrying a CFP fluorescence marker) in LB medium, and the results of two replicate competitions are shown. Another common way of determining s_d is from $(r_{mut} - r_{wt})/r_{wt}$, where $r_{mut} = \ln[N_{mut}(24)/N_{mut}(0)]$ and $r_{wt} = \ln[N_{wt}(24)/N_{wt}(0)]$ and $N_{mut}(t)$ or $N_{wt}(t)$ are the numbers of bacteria with the mutation, or wild type, at time t, respectively. **b** Growth curve assay. The wild-type (in pink) and mutant (in blue) genotypes are grown separately, and growth of each culture is estimated by taking periodic measurements of optical density (OD; e.g. every hour). OD is proportional to cell density, and the slope of OD against time provides an estimate of the r of each bacterial clone. If most of the differences come from the exponential phase, then s can be calculated as follows: $(r_{mut} - r_{wt})/r_{wt}$. If there are also differences in the other phases of growth, then the measure of the selection coefficient has to be adapted to incorporate them.

tions because they will affect both genotypes equally. More importantly, the two assays reflect different fitness components. Whereas there are conditions under which both assays can be theoretically expected to give similar results, generally they should not [Chevin, 2011]. One striking case where this will be evident is when interactions between different genotypes are an important form of selection, e.g. the fitness is frequency dependent [Levin, 1988]. Evidently, this can only be detected when the two genotypes are in competition. Imagine the case where a genotype produces some toxin to which it is immune. The advantageous effect of such mutation can only be seen when a competitive fitness assay is performed. On the other hand, only a growth assay can reveal if producing the toxin is costly. The growth curve assay is also useful to study the ability of a genotype to colonize an empty niche.

As implicit above, most measures of fitness involve one or two strains. However, in nature there can be any number of genotypes segregating in a population. In this case, the frequency distribution of multiple alleles can be analyzed.

Fitness Effects of Mutations at the Genome Level

It is intuitive to suppose that not all mutations affect individuals in the same way. Are they all of similar effect? Do they fall into distinct categories? The first way to classify mutations is according to their sign, whether they increase or decrease fitness. It was first proposed and later observed that most mutations are deleterious, and only a small fraction is beneficial. As far as deleterious mutations are concerned, they also fall into three main categories: nearly neutral, mildly deleterious and lethal. All of these classes are likely to be present for any given organism in any given environment. The question then is: what is the relative frequency of these classes? We must keep in

mind that these frequencies may change according to the environment and genetic background.

The first theoretical framework to make predictions on the fraction of deleterious to advantageous mutations was proposed by Fisher [1930]. In the 1930s, he proposed a model of adaptation nowadays known as Fisher's geometrical model (FGM). In FGM, an organism is described by several phenotypic traits (n), each having an optimum value in a given environment, and a population of such individuals, which is, at present, located away from that optimum. Random mutations continuously occur, the majority of which are deleterious, but those that move the population closer to the optimum are advantageous and can contribute to adaptation to that environment. This is much like a hill climbing process towards a single fitness peak, where every move up the hill is more likely to involve small steps than larger ones, because very large steps are deleterious. Fisher's model makes the prediction that the mean effect of beneficial mutations should decrease as the population gets closer to the fitness peak. Furthermore, at any given distance to the peak, the bigger the n, the smaller the fraction of beneficial mutations. Recently, Martin and Lenormand [2008] studied a generalized version of FGM and provided predictions for the distribution of fitness effects of both beneficial and deleterious mutations. In particular, this extended model has led to the prediction that, when populations are not too far from the fitness peak, arising beneficial mutations should follow a beta distribution with unit shape parameter.

Gillespie [1984] proposed another theoretical model that relies on the statistical properties of extreme values of distributions. Because beneficial mutations are assumed to be rare, they lie on the tails of the general distribution of fitness effects. In many general circumstances, those tails follow an exponential distribution, and therefore the distribution of beneficial mutations, under Gillespie's mutational landscape model, is expected to be exponential. None of the models makes strong assumptions about genetic architecture of the organisms, hence their generality and potential relevance for all species.

Direct measures of the effects of mutations and experimental evolution allow testing the predictions of these models. At the genome level, three main strategies are used to estimate the distribution of fitness effects of mutation (DFEM): MA, construction of mutants and evolution of large populations in controlled conditions.

MA in Bacteria

One of the classical experiments to estimate phenotypic effects of spontaneous mutations is a MA (fig. 1). An approach that has been applied to different organisms, including *Drosophila melanogaster* [Haag-Liautard et al., 2007, 2008], *Caenorhabditis elegans* [Davies et al., 1999; Denver et al., 2000; Estes et al., 2004; Keightley and Caballero, 1997; Vassilieva and Lynch, 1999] and *S. cerevisiae* [Dickinson, 2008; Joseph and Hall, 2004; Wloch et al., 2001; Zeyl and DeVisser, 2001], *Tetrahymena* [Brito et al., 2010], *E. coli* [Kibota and Lynch, 1996; Trindade et al., 2010] and some viruses [de la Iglesia and Elena, 2007; Elena and Moya, 1999; Lazaro et al., 2003; Li and Roossinck, 2004]. The assumption underlying an MA experiment is that, when the population size (N_e) is kept extremely low, genetic drift overwhelms natural selection and mutations accumulate at the rate at which they appear (i.e. neutrally). In higher eukaryotes, MA is done by propagating the populations under extreme inbreeding. In microorganisms, it is done by exposing the populations to periodic bottlenecks of a single individual. Since this individual is chosen randomly and independently of its fitness, selection will be very inefficient in purging deleterious mutations. At such low effective population sizes, the mean fitness of the population declines, reflecting the intuitive assumption that most mutations are deleterious. At the same time, the variance between different lines increases, because different mutations are fixed at different (stochastic) times. With the quantification of change in fitness along the bottlenecks and the increase in variance between the different lines that are propagated, it is possible to estimate both the genomic mutation rate and the mean fitness effects of mutations [Gordo and Dionisio, 2005; Kibota and Lynch, 1996]. Under the assumption that the DFEM follows a continuous distribution, maximum likelihood can also be used to estimate its parameters [Keightley, 1994]. A gamma distribution has been typically assumed and is predicted by FGM [Martin and Lenormand, 2006]. The recent advance in whole-genome sequencing technologies will complement the phenotypic data generated in a MA experiment with genetic data [Andersson and Lind, 2008], allowing better estimates of the rate and effects of mutations. This will be critical, since without knowledge of the number of mutations occurring in a MA line, the estimates of the mutation rate tend to be biased downwards from the measured decline in fitness. In addition, the deleterious effects of mutations tend to be overestimated [Trindade et al., 2010]. This is because mutations of small fitness effect are difficult to measure experimentally, even though they may be very numerous. One example of these very slightly deleterious mutations is changes in codon usage. In *E. coli*, the average strength of selection (s) versus drift ($1/N_e$) for optimal

codons is $N_e s \sim 1.7$, as estimated from population genetics theory that assumes that this sort of mutations are in an equilibrium between mutation, selection and drift [Sharp et al., 2010]. This implies that, in order to measure directly the effect of a mutation that changes one optimal codon to a suboptimal one, we would require an experimental setup with a measurement error of less than 10^{-7} to 10^{-8}, a scenario that we can only dream about with the current assays.

Three species of bacteria have been subjected to MA: *E. coli*, *S. typhimurium* and *Streptococcus pneumoniae*. In *E. coli*, two strains with different DNA repair abilities have been studied. Kibota and Lynch [1996] studied a mismatch-proficient *E. coli* and were the first to estimate the average fitness effects of deleterious mutations in a bacterium. After propagating 50 independent lines during 300 bottlenecks, they observed a modest fitness decline of 0.02 and an increase in fitness variance between the lines of 0.0003. This allowed an estimation of the mean fitness effect of deleterious mutations of 1.2% and of the genomic deleterious mutation rate of 0.0002. When a mutator strain of *E. coli* was subjected to the same type of experiment [Funchain et al., 2000; Trindade et al., 2010] even under a smaller number of bottlenecks, a much larger fitness decrease was observed (fig. 1). The mean fitness decrease of 50 lines of a *mutS*$^-$ strain of *E. coli* was 0.12 [Trindade et al., 2010]. In a similar experiment (involving 100 lines of a *mutS*$^-$ *E. coli*), some mutator lines were even doomed to extinction after 40 bottlenecks. Furthermore, the majority of the lines had defects in at least one metabolic pathway [Funchain et al., 2000]. The current estimate of the genomic deleterious mutation rate of the *mutS*$^-$ strain is 0.005 per generation with an effect per deleterious mutation of 3% [Trindade et al., 2010]. This value of the fitness effect may be an overestimation due to two main reasons: variation in the effects, which is disregarded in the estimation method and the possible occurrence of beneficial or compensatory mutations during the experiment.

Maisnier-Patin et al. [2005] performed an MA in *S. typhimurium* under conditions that resulted in different mutation rates. They propagated a strain that carried a plasmid with an inducible error-prone polymerase (DinB, that when overexpressed increases mutagenesis). As observed in *E. coli*, mean population fitness declined. Since the authors sequenced a substantial portion of the genome, determination of the number of mutations accumulated could be done and an important observation emerged: the decline in fitness was not linear with the number of mutations accumulated. A trend of a decrease less than linear was particularly evident after a large number of mutations had accumulated. These data provided evidence for antagonistic epistasis between deleterious mutations – whereby the fitness effect of two mutations combined is smaller than adding the effects of each separately. The mechanism of buffering was identified as the overproduction of chaperones. These assist in protein folding and are therefore capable of reducing the impact of deleterious mutations as they accumulate. The experiment also demonstrated that robustness can evolve and that the fitness effects of mutations depend on the genetic background. Another MA experiment in *Salmonella*, mentioned above [Andersson and Lind, 2008], provided an estimate of s_d around 0.15%. This estimate is an order of magnitude lower than that measured in *E. coli*. One of the reasons for this discrepancy may be due to the different methodologies employed in the estimation: in the MA of *E. coli*, fitness-related phenotypes were the bases for the estimation, whereas in this study the number and genetic basis of the mutations were the raw data. In contrast to the results obtained in *Salmonella* and *E. coli*, Stevens and Sebert [2011] observed an increase in fitness during an MA experiment in the pathogen *S. pneumoniae*. At first sight, it would seem that *S. pneumoniae* has a distribution of fitness effects totally different from that of *E. coli* and *Salmonella*. Although possible, we should first consider a key difference between the studies: the fitness assay used in *S. pneumoniae* (the colony growth dynamics in a Petri dish) was different from that in *E. coli* or *Salmonella* (competitive fitness assay or exponential growth rate measurement in liquid culture). Stevens and Sebert [2011] showed that the fitness increase was due to the occurrence of beneficial mutations that were amplified by selection during stationary phase. From the fitness increments observed in the different replicate lines, an extraordinarily high mutation rate (4.8×10^{-4} per genome per generation) towards beneficial alleles was inferred for *S. pneumoniae*, the highest estimate to date in a bacterium.

Direct Estimates of the Fitness Effects

It is possible to measure the fitness of a sample of the mutational neighborhood of a given genotype (wild type), by constructing mutants that differ from the wild type by a single mutation. The largest genomic studies that followed this approach surveyed the fitness effects of all possible knockout mutations [Baba et al., 2006; Nichols et al., 2011; Yamamoto et al., 2009]. These are probably the strongest effect mutations because they result in gene inactivation, though their effect may range from lethality

(revealing an essential gene) to no detectable change in phenotype. A few studies have tried to measure the frequency of knockout mutations that are lethal in different organisms and to relate these to the level of genome complexity. For example, in the bacterium *Mycoplasma genitalium*, which possesses the smallest gene complement of an independently replicating cell (517 genes), global random transposon mutagenesis revealed that about two thirds of the genes are essential [Hutchison et al., 1999]. In another bacterium, *E. coli* (~4,186 genes tested), this number was found to be about 8% [Baba et al., 2006; Yamamoto et al., 2009]. Although it is tempting to make the hypothesis that the number of essential genes could be around 350 or so, the approaches used to estimate the number of essential genes usually underestimate the actual number in the wild, where nutrients and environmental conditions are not always so benign as in the test tube. Nevertheless, one may still wonder what explains the survival of genes whose deletions produce no detectable phenotype. Part of the explanation for this may rely on the fact that mutations with undetectable effects in the laboratory may actually cause small fitness effects that, in large populations, are kept in check by selection.

Elena et al. [1998] used transposon mutagenesis to construct 226 clones of *E. coli*, which differed from the parental clone by a single random insertion in the genome. By performing a competitive fitness assay between each single step mutational neighbor and the parental strain, they estimated a mean deleterious effect per transposition of 3%. None of the mutants showed a beneficial effect, and 80% showed significant deleterious effects. This indicates that the fraction of beneficial mutations caused by random transposon insertions is below 0.4% and that 20% of these mutations are neutral or do not present measurable fitness effects. Although a gamma distribution, which is predicted theoretically [Martin and Lenormand, 2006], fitted reasonably well the data of fitness effects, the distribution that provided the best fit was a gamma combined with a uniform, which allowed the observed class of large deleterious effects to be probable.

Evolution in Large Populations: Effects of Beneficial Mutations

In contrast to MA experiments, in evolution experiments involving large population sizes, mean fitness increases as the population adapts. From the mean increase in fitness of a population, it is possible to get an estimate of the beneficial mutation rate (U_b) and the mean effect of mutations (s_b). With this (and other questions) in mind, Lenski et al. [1991] started in 1988 what is now a classical long-term experiment of evolution in *E. coli*. They followed the fitness increase in populations adapting to glucose minimal medium through tens of thousands of generations. They used a population genetics model to estimate the mutational parameters, and concluded that a U_b of 2×10^{-9} and an s_b of 3% [Gerrish and Lenski, 1998] provided the highest likelihood of explaining the data. However, in the method used for the likelihood calculation, they made a simple assumption: that U_b was not very large [Sniegowski and Gerrish, 2010]. At the time, beneficial mutations were thought to be very rare, and this simplification made sense. Once it became clear that U_b was probably much higher (see below), such that it was likely that many beneficial mutations were segregating in the population at the same time, the data were reanalyzed, yielding new estimates of $U_b \sim 5.7 \times 10^{-5}$ and $s_b \sim 0.3\%$ [Sniegowski and Gerrish, 2010]. In general, from the data of fitness increase along generations, model fitting can provide either a high U_b and a small s_b or a small U_b and high s_b [Hegreness et al., 2006; Perfeito et al., 2007; Rozen et al., 2002]. Which conditions are more likely and which model to use when analyzing data from adapting populations depends on which regime of adaptation the populations lay. There are three different regimes to consider when studying the process of adaptation: the periodic selection (PS), clonal interference (CI) and the multiple mutations (MM) regimes. The key issue in distinguishing between them is the rate at which beneficial mutations (U_b) occur and the number of individuals in the population (N) [for a recent review on this particular subject see Sniegowski and Gerrish, 2010].

In the PS regime, mutations are rare compared to the strength of selection. Once an individual acquires a beneficial mutation, it starts reproducing faster, and eventually its descendants dominate the population. This is called a fixation or a selective sweep because all diversity in the population is swept away, being replaced by a single clonal lineage carrying one beneficial mutation. In PS, fixations happen independently and are well separated in time. Most classical models of adaptation are based on PS because early population genetics models were dominated by the weak mutation-strong selection assumption. Adaptation was assumed to be a slow process, punctuated by short periods of fitness increase associated with fixation events. In contrast, the CI regime is characterized by a high beneficial mutation rate, and a significant number of mutations of large effect. Due to the large U_b in the time it takes for a mutation to fix, others appear in the population. These mutations will compete for fixation, and only the strongest will be successful. This type of in-

terference between mutations was first studied by Hill and Robertson [1966] in sexual species, and later extensively analyzed by Gerrish and Lenski [1998] in asexual populations. Two main predictions emerge from the theory of CI. One is that many small effect mutations are lost from the population, even though they provide a fitness advantage to the individuals. The second prediction is that the rate of adaptation increases slower than linearly with the population size (experimentally tested in de Visser et al. [1999]). Despite some unrealistic simplifications, such as not allowing for MM in the same genotype, this theory has proved to be very robust [Park et al., 2010]. The MM regime is closely related to CI but relieves the assumption of a single mutation in each genotype. However, in the first models considering this regime, all mutations were assumed to have the same effect. This means that no single mutation can outcompete all the others (as is the case in CI). Instead, clonal lineages only have a chance of increasing in frequency if they harbor more than one beneficial mutation. This regime has been extensively studied in a theoretical framework adapted from physics, called the traveling wave [e.g. Desai et al., 2007; Hallatschek, 2011; Rouzine et al., 2003]. In this approach, populations are divided into genotypes and ordered according to the number of beneficial mutations they carry. This forms a distribution, shaped like a wave, with most genotypes having an average number of mutations. New mutations are constantly added to all genotypes in the population, but they only survive if they fall on the classes with the highest number of mutations (the 'nose' of the wave). These genotypes then become the most frequent in the population, causing the wave to travel forward as time passes. Like CI, the MM predicts that the rate of fixation of beneficial mutations does not increase linearly with U_b or N. It also predicts that fixations will happen at a regular speed and that mutations do not fix alone. Instead, each fixation event is associated with a cluster of beneficial mutations.

The three regimes are not mutually exclusive. In fact, as populations become adapted to their environment, the different regimes are likely to follow each other in time. Picture a population that just colonized a new environment. It is intuitive to assume that there are many ways in which fitness can be improved, many of which involving mutations of large effect [Orr, 2005]. We then expect the regimes of CI and MM to be important. After the initial steps of adaptation and if the environment stays constant, the supply of beneficial mutations diminishes because there are fewer and fewer ways for further improvement in the population, we are likely to observe a PS regime of evolution. The transitions between the regimes are not expected to be sharp, with many intermediate phases. In general, the type of regime one may expect to see depends on the mutation rate for beneficial mutations and the distribution of their effects. It is therefore important to estimate these parameters in different environments and for different genetic backgrounds.

In several laboratory experiments where bacterial adaptation was studied, beneficial mutations were detected through their effects on neutral markers, from which mutation rates and effects can be estimated (fig. 2). The principle relies on a phenomenon called hitchhiking effect, which can be detected in a clonal population that contains some variability at a neutral locus. If a strong beneficial mutation arises in a clone carrying a particular neutral allele, it will rapidly increase in frequency and, due to complete linkage, the frequency of that neutral allele will also rapidly increase. This phenomenon was first observed in the 1950s in populations of *E. coli* [Atwood et al., 1951]. Over the years, people have developed different marker systems that produce a readily visible phenotype or can be quantified by some other procedure. Among the first type is the inactivation of genes that render bacteria unable to synthesize some compound essential for their survival, like histidine [Atwood et al., 1951], or unable to metabolize a given carbon source like arabinose [Lenski et al., 1991] or lactose [Nguyen et al., 1989] and others. In the first case [Atwood et al., 1951], bacteria can be distinguished based on the requirement of histidine supplementation in order to grow, and in the case of Ara^- and Lac^- mutants these can be distinguished from the wild type by their ability to generate red colonies when cultivated in tetrazolium arabinose or lactose agar, respectively. Among the second category of markers are those like microsatellites [Imhof and Schlotterer, 2001; Perfeito et al., 2007] and fluorescent proteins [Barrick et al., 2010; Hegreness et al., 2006]. Imhof and Schlotterer [2001] provided the first multiple marker system to track beneficial mutations. In this system, bacteria were labeled with a plasmid that contained a microsatellite sequence. These sequences are known to be unstable, so the number of the repeat unit varies with time making it possible to quantify the frequency of bacteria with a given repeat number.

The regime of adaptation where the population is expected to be greatly determines the importance of using more than two markers. In the PS regime, a two-marker system is assumed to be sufficient to infer the effect of mutations that sweep to fixation. Nevertheless, even in this regime such a system will, to some extent, underes-

Table 1. Rates and effects of beneficial mutations across environments in different bacteria

Organism	Environment	s_b	U_b	Reference
E. coli	Minimal medium	0.03	2×10^{-9} [a]	Lenski et al. [1991]
E. coli	Minimal medium	0.003	5.7×10^{-5} [c]	Sniegowski and Gerrish [2010]
E. coli	Minimal medium	0.022	5.9×10^{-8} [c]	Rozen et al. [2002]
E. coli	Rich medium Small N_e Large N_e	 0.012 0.023	2×10^{-5}	Perfeito et al. [2007]
P. fluorescens	Stressful nutritional environment	2.1	3.8×10^{-8}	Barret et al. [2006]
P. fluorescens	Colonizing ability air-liquid interface	2	Not estimated	McDonald et al. [2011]
E. coli	Rich medium	0.054[b] 0.01[c] (0.005–0.015)	$10^{-6.7}$ [b] 10^{-5} [c] (10^{-4} to 10^{-6})	Hegreness et al. [2006]
S. pneumoniae	MA, rich solid medium	0.025	4.8×10^{-4}	Stevens and Sebert [2011]
S. typhimurium (rpsL)	Compensation Small N_e Large N_e	 0.10 0.05	 2×10^{-7} $>10^{-7}$	Maisnier-Patin et al. [2002]
E. coli (rpoB)	Compensation Minimal media Genetic background (I572L, S574Y, D516Y, Q148L, D516G, S512F, Q513P, Δ532–535)	 (0.08, 0.09, 0.10, 0.19, 0.15, 0.15, 0.18, 0.33)	10^{-7} [b]	Barrick et al. [2010]
P. fluorescens (nalidixic acid resistance)	Rich medium Glucose Mannitol Sorbitol	0.087 (0.06–0.14) 0.024 (0.013–0.052) 0.044 (0.029–0.079) 0.031 (0.029–0.059)	5×10^{-7}	Kassen and Bataillon [2006]

[a] From dynamics of fitness increase. [b] Assuming a single value for the effect of each mutation. [c] Assuming an exponential distribution for beneficial effects.

timate the mutation effects. This is because the increase in frequency used to calculate the effect of the mutation responsible for that shift is the frequency of the whole population tagged with the marker, and not exclusively that of the beneficial mutation itself. As more complex scenarios take shape, as in the case of CI or MM, the error inherent to the two-marker system becomes more important since the increase in frequency of a marker would represent the selective effect of a number of different beneficial mutations. In this case, a larger number of markers or an analysis of single clones in the population (as was done in Desai et al. [2007]) is probably more appropriate.

Most studies to date have applied a system of two markers, and in table 1 we show several estimates of fitness effects of mutations in bacteria from the different studies. Rozen et al. [2002] determined the distribution of effects of beneficial mutations that fixed during adaptation of *E. coli* to minimal medium with glucose as the sole carbon source. Using a two-allele marker system and following the frequencies of the markers during 400 generations, they measured the fitness of clones sampled from the winning populations. Under the assumption that the increase in fitness was the result of a single mutation that was destined to be fixed, the distribution of fitness effects of fixed beneficial mutations was estimated. They found that mutations of intermediate effects contribute to adaptation in these populations. Furthermore, the mean effect of arising mutations that maximized the probability of obtaining the observed distribution of fixed beneficial mutations was 2.2%. Barret et al. [2006] estimated the distribution of effects of fixed beneficial mutations using a similar experimental system, but this time in *Pseudomonas fluorescens*, adapting to a more stressful environment. The value of the mean effect esti-

mated was as high as 2.1 (about two orders of magnitude larger than the estimates in *E. coli*). They concluded that large-effect beneficial mutations are not uncommon, at least if environments are stressful.

In these experiments, the genetic bases of the adaptation are not described, and it is assumed that a single beneficial mutation has fixed. If many mutations of smaller effect are more probable, then the mean effect per mutation can be much smaller and the mutation rate much higher. The issue can possibly be resolved in at least two ways. One is at the experimental level, by applying whole-genome sequencing of sampled clones, which allows determining the number of mutations. Another possible approach is to increase the number of neutral markers to detect small effect mutations that may be lost in competition. At the theoretical level, more robust methods should be constructed for analyzing the dynamics of two marker systems, which should be capable of determining the conditions under which a unique distribution of arising beneficial mutations can explain those data. This can be a difficult task though. Rozen et al. [2002] pointed out that in adapting populations experiencing intense CI, many different distributions of arising beneficial mutations can lead to the same distribution of fixed beneficial mutations. The latter are the ones that can be detected through the associated marker once they reach a high frequency. Later, the work of Hegreness et al. [2006] elegantly demonstrated that the complex dynamics of adaptation observed in a double-marker system can be described with only two parameters: an effective selection coefficient (s_e) and effective rate of beneficial mutations (U_e). This model assumes that all mutations have the same selective effect and all occur at the same rate. The relation between these effective parameters of evolutionary dynamics and the more biologically meaningful parameters U_b and s_b is difficult to determine exactly. When studying the dynamics of *E. coli* populations adapting to rich medium using two fluorescent markers, Hegreness et al. [2006] determined that a value of $s_e \sim 5.4\%$ and $U_e \sim 10^{-6.7}$ would explain the dynamics during the initial time periods of adaptation. However, a biologically more realistic assumption that selection coefficients follow a continuous distribution also explained the dynamics. When assuming an exponential distribution, the estimates obtained were $10^{-4} < U_b < 10^{-6}$ and $s_b \sim 1\%$. Although this issue has not been carefully addressed from a theoretical standpoint, it seems that the value of the effective parameter U_e will be an underestimate of the beneficial mutation rate and that the value of s_e will correspond to an overestimate of the mean effects of beneficial mutations.

An important component of adaptation in natural bacterial populations may involve the occurrence and fixation of compensatory mutations. These are beneficial mutations that ameliorate the effects of deleterious mutations that were previously present in the genome. One widely studied type of such deleterious mutations is that conferring antibiotic resistance. Maisnier-Patin et al. [2002] performed one of the first studies that assayed beneficial mutations, which compensate for the cost of streptomycin resistance, in *Salmonella* populations. They scored for mutants able to produce colonies of increased size in periods of about a hundred of generations. This method does not detect many compensatory mutations of small effect because they might not lead to distinct colony sizes, so it is biased to detect only mutations of strong effect (or several mutations of small effect). That is probably why adaptation was only detected in 2 out 10 populations in 500 generations. As a result, they found large effects for compensatory mutations (s_b values above 10%) and a rate of beneficial mutations of about 10^{-7} per genome per generation. More recently, Barrick et al. [2010] examined the adaptation of several strains of *E. coli*, each carrying a distinct allele at the *rpoB* gene and all resistant to rifampicin. They used the two-marker system method to estimate the effective parameters U_e and s_e mentioned above. When focusing on the initial phase of the marker frequency change that is likely to only capture the increase in frequency of the first beneficial mutation, they estimated a common value of U_e of about 10^{-7}, and s_e values ranging from 8 to 33% depending on the resistance mutation studied. Interestingly, the fitness effect estimated for the first step of the adaptation correlated very neatly with the fitness cost of the resistance mutation, whereas the estimated mutation rate was independent of the specific genetic background. This suggests that the distribution of fitness effects of beneficial mutations may depend on the genetic background in a predictable manner. FGM predicts exactly such a result. If we assume that a population is displaced from the optimum by the fixation of a strong (weak) effect deleterious resistance mutation, then, under FGM, we expect arising beneficial mutations that compensate for that defect of large (small) effect.

In another study, adaptation of *E. coli* to rich medium was followed, and a much higher estimate of the mutation rate towards newly arising beneficial mutations was obtained [Perfeito et al., 2007]. The authors took the approach of increasing the number of the neutral markers, and found values of $U_b \sim 2 \times 10^{-5}$ and $s_b \sim 1\%$. They followed the changes in allelic diversity at a variable micro-

satellite locus, and estimated the fitness effects of those beneficial mutations that were able to reach a detectable frequency. Importantly, they studied bacterial populations evolving at two different effective population sizes. In one case, the N_e was about 2×10^4, and in the other case 2×10^7. The first N_e is expected to minimize the effects of CI and still be large enough to reduce the accumulation of deleterious mutations. Increasing the N_e about 1,000 times should increase significantly the effect of CI. The estimated values of U_b and s_b were well supported both by the dynamics of variation in the microsatellite marker in the populations of small N_e, and by the observed increments in mean fitness of all the populations that were evolved. Consistently, a clear signal of intense CI could be detected in the distribution of microsatellite alleles of the populations evolving under the large N_e setting [Imhof and Schlotterer, 2001; Perfeito et al., 2007]. Furthermore, the estimated mean effect of segregation mutations in large populations was stronger than in the small populations.

Fitness Effects of Mutations at the Gene Level

As mentioned above, mutations can be targeted to particular genes or pathways. By observing the fitness effects of these mutations, we can shed light onto the evolution of the associated phenotypes. For example, this type of analysis can be used to address the evolution of medically relevant phenotypes, or the acquisition of new metabolic functions. We can also use the DFEM of single genes to compare and possibly extrapolate to the whole genome.

The first descriptions of fitness effects at the gene level focused on the fitness effects of mutations conferring antibiotic resistance. It was soon realized that resistance-conferring mutations were deleterious in the absence of the drug [e.g. Nguyen et al., 1989; Schrag and Perrot, 1987]. From an evolutionary point of view, it is useful to distinguish two ways in which bacteria can become resistant to antibiotics. They can mutate the drug target (often single point mutations) or acquire genes that lead either to the secretion or destruction of the drug. The first case is achieved by spontaneous mutation while the second requires HGT. In the case of spontaneous mutations, a comprehensive study in *P. fluorescens* was able to describe the distribution of effects of mutations conferring resistance to nalidixic acid [Kassen and Bataillon, 2006]. They measured the growth rate of 665 spontaneous mutants in rich medium in the presence and absence of the antibiotic. They found that the distribution of fitness in both environments was roughly normal, with a right skew (towards less harmful effects). Surprisingly, 28 of 665 (4%) mutations were found to be beneficial in the absence of the antibiotic. The distribution of these mutations was exponential and had a mean beneficial effect between 3.1 and 8.7%, depending on the richness of the medium and carbon source used. A comparison between these results and the genome-wide studies suggests that the effects of mutations in a single gene might not be very different from the distribution of effects in the whole genome.

A similar study was conducted in a resistance-conferring enzyme, beta-lactamase [Weinreich et al., 2006]. Here, fitness was defined as the level of resistance against an antibiotic to which the enzyme had low affinity. Although the authors only studied a small subset of mutations, they showed that the effect of beneficial mutations is highly dependent on the amino acid sequence at other sites of the protein. The same was observed for a protein involved in central metabolism, IMDH [Lunzer et al., 2010]. The authors recreated all 168 mutations that separate the IMDH of *E. coli* from the homologous enzyme in *Pseudomonas aeruginosa* in different clones. They then looked at the enzyme-specific activity in these mutants, which is related to fitness. They found a distribution skewed for lower values with 56 mutants with decreased activity and 5 with increased activity. This yields a fraction of 3% activity-enhancing mutations, similar to the fraction of beneficial mutations in the whole genome. However, the relationship between IMDH activity and fitness is not linear, at least for wild-type *E. coli* [Lunzer et al., 2010], and so it is difficult to extrapolate between activity and fitness. Furthermore, it is not clear whether the phenotype fitness map for this enzyme also changed since these mutations became fixed. This sample of mutations is biased, in the sense that these already survived natural selection. The fact that most of them are deleterious indicates that there are highly epistatic interactions between them, as was observed for beta-lactamase [Weinreich et al., 2006].

The distribution of effects of random mutations was measured for two ribosomal genes, *rpsT* and *rplA*, in *S. typhimurium* [Lind et al., 2010]. The mutations were generated randomly, and their effects on growth rate and competitive fitness were measured. In minimal medium with glucose as carbon source, the average deleterious selection coefficient was 5 and 1% for each assay, respectively. Interestingly, the effects of mutations differ between the assays, likely because the competitive fit-

ness assay is a composite measure of the entire growth curve and not just the exponential phase. The values of the selection coefficient are relatively low, considering that the knockout of either gene reduces fitness by more than 70%. No loss-of-function or advantageous mutations were detected. The mutations were further analyzed by category: synonymous and nonsynonymous. The former mutations do not change the protein sequence while the latter do. Surprisingly, the mean selection coefficient was very similar in both sets (0.96 and 1.31%, respectively). In general, synonymous mutations are thought to be much milder than nonsynonymous mutations because they do not affect protein sequence. It appears that, at least for these two genes, both are equally important. The question is then: what phenotypes are these mutations affecting? The authors ruled out codon bias effects and found only a weak correlation between mRNA free energy and the effects of the synonymous mutations. But it can still be that other properties of the mRNA molecule are important for growth. The fact that the two distributions are similar indicates that a common biological mechanism might be responsible for the effects. As the authors point out, if the protein is very robust to amino acid changes, then all fitness effects of both synonymous and nonsynonymous mutations should come from changes in the mRNA. As ribosomal proteins are the most abundant in the cell, it will be very important to study the effect of both synonymous and nonsynonymous mutations in other core genes. Furthermore, because nonsynonymous mutations can affect protein stability and both influence translation rate, factors that will cause changes in protein concentration, it would be important to measure these in the mutants.

Even though the effect of specific mutations in proteins is highly dependent on the background [Weinreich et al., 2006], the previous results suggest that the shape of the distribution may be invariant. If that is the case, it is plausible that there is a universal biological mechanism behind it. One way to test this is to look at the particular phenotypes that mutations are affecting and how they interact with each other. In the case of proteins, the most relevant biophysical phenotypes are protein stability and activity. As reviewed in DePristo et al. [2005], the interaction between these two can give rise to very complex adaptive pathways and a high degree of epistasis, similar to what was observed. This emerges from the fact that the optimal level of stability is intermediate and that mutations affecting activity also affect stability. As such, mutations, which are selected for their beneficial effects on activity, tend to be deleterious for the stability of the protein. These can then be compensated for by subsequent mutations that affect stability only. The fitness effects of mutations depend only on how far the proteins are from their optimal activity and stability and not on their specific amino acid sequence. It is likely that wild-type proteins are close to optimal levels of stability, which would explain their similar distributions of fitness effects, while at the same time explain that the effect of particular mutations changes with the background. A biophysical model incorporating these effects was able to accurately describe the fitness effects of mutations in viral proteins [Wylie and Shakhnovich, 2011], but it is likely to apply to bacterial proteins as well.

A similar approach can be used to direct selection to a whole pathway, rather than a single gene. Ibarra et al. [2002] showed that *E. coli* adapting to glycerol is able to optimize growth as predicted from a metabolic model. Although this study did not address the problem of the DFEM, they were later able to show that few mutations of high effect are able to confer a large advantage in this environment where the wild type grows very poorly [Herring et al., 2006]. More recently, McDonald et al. [2011] examined the DFEM for mutations that improve a particular phenotype in *P. fluorescens*: the ability to colonize the air-liquid interface in static broth. They picked random mutants in a medium that did not select for this phenotype by using a reporter construct instead. This meant that the probability of picking up a mutant was not affected by its fitness, as long as the reporter gene was activated. They estimated the distribution in two environments: in a static environment, where the phenotype is advantageous, and in a shaken environment, where this phenotype had no known fitness effect. In both cases, the distribution was well fitted by a normal distribution with similar variances but different means. In the first environment, where all mutations were beneficial, they were able to reject an exponential distribution of beneficial effects (as theory predicts). To our knowledge, there is no theoretical framework that predicts a normal distribution of beneficial mutations. It can still be that the distribution is exponential but that the low-effect mutations are not picked up by their reporter construct. If not, then we need a new theoretical framework to describe the fitness effects for mutations affecting this particular phenotype (see below). In the second environment, mutations were expected to be random and any residual selection due to the use of the reporter would not affect the distribution. However, there were not enough beneficial mutations to test a distribution.

Analyzing fitness as a function of phenotypes instead of the underlying mutations can be a powerful approach to determining the DFEM. It has been appreciated for a long time that natural selection acts on genotypes through the phenotypes encoded by them. But it has been a daunting task to uncover these maps at the genome level. However, the approach can be taken at a smaller scale for well-characterized genetic systems. For example, Perfeito et al. [2011] constructed a number of mutants with different levels of expression of the *lac* operon in *E. coli*. In this case, the genotype is known and the phenotypes are well defined and can be measured, namely the level of protein production and of protein activity of the *lac* genes. The fitness of these mutants was measured in an environment where only one of these phenotypes (level of protein production) affected fitness, and in one where both were important. This allowed the authors to disentangle the effects of the two phenotypes and to see how they interact. Using the experimental data and a biophysical model of protein expression and growth, they were able to predict the fitness of the whole parameter space, including mutants that were not observed. This type of approach enables us to understand and predict how changes in both genotype and environment affect the relative fitness of different genotypes.

Ideally, one would like to construct mutants in all the fitness-relevant phenotypes in a given environment. If we could do that, we could potentially get the whole fitness landscape of an organism and the DFEM would naturally come from it. However, that is a very laborious task to do at the organism level and requires a deep knowledge of both the genetics of the organism, as well as the details of its ecology. Recently, high-throughput measurements allowed the generation of unprecedented data on genotype, phenotype and fitness [Nichols et al., 2011]. If we can guide the analysis of such datasets by integrative biochemical and biophysical models of bacterial growth, we can finally begin to build whole-genome phenotype-based fitness landscapes. The importance of integrating this type of data and of using such bottom-up approaches in population genetics is currently a big interest focus of several groups, and has been reviewed in Dean and Thornton [2007] and Loewe [2009].

Another key question relates to determining the distribution of fitness effects of DNA elements that can be transferred among bacteria. Although there are some studies [e.g. Bouma and Lenski, 1988] where the fitness costs imposed on bacteria by plasmid carriage have been measured, a thorough study on such costs or benefits across environments has not been done. Recently, the costs of 5 natural conjugative plasmids (which ranged from 2.8 to 8% in *E. coli* strain MG1655) have been shown to depend on the genetic background of the bacteria to where they were transferred [Silva et al., 2011]. Since such elements are of extreme importance in the evolution of bacteria, in particular of antibiotic resistance, future studies with larger samples will allow a better understanding of their fitness effects.

In this paper, we reviewed some of the most relevant studies that directly address several fundamental questions in the field of evolutionary biology. Nevertheless, theoretical models of evolution make many more predictions that need to be experimentally tested. Experimental evolution in prokaryotes constitutes one of the best methods to answer both fundamental questions in evolution and test model predictions in controlled laboratory experiments.

References

Andersson DI, Lind PA: Whole-genome mutational biases in bacteria. Proc Natl Acad Sci U S A 2008;105:17878–17883.

Ashley RD, Feil EJ, Falush D: Bacterial Population Genetics in Infectious Disease. New York, John Wiley & Sons, 2010.

Atwood KC, Schneider LK, Ryan FJ: Periodic selection in *Escherichia coli*. Proc Natl Acad Sci U S A 1951;37:146–155.

Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, Datsenko KA, Tomita M, Wanner BL, Mori H: Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol Syst Biol 2006; 2:2006.0008.

Barrett RD, MacLean RC, Bell G: Mutations of intermediate effect are responsible for adaptation in evolving *Pseudomonas fluorescens* populations. Biol Lett 2006;2:236–238.

Barrick JE, Kauth MR, Strelioff CC, Lenski RE: *Escherichia coli rpoB* mutants have increased evolvability in proportion to their fitness defects. Mol Biol Evol 2010;27:1338–1347.

Ben-Yehuda S, Dubey GP: Intercellular nanotubes mediate bacterial communication. Cell 2011;144:590–600.

Bouma JE, Lenski RE: Evolution of a bacteria plasmid association. Nature 1988;335:351–352.

Brito PH, Guilherme E, Soares H, Gordo I: Mutation accumulation in Tetrahymena. BMC Evol Biol 2010;10:354.

Chevin LM: On measuring selection in experimental evolution. Biol Lett 2011;7:210–213.

Craig NL: Target site selection in transposition. Annu Rev Biochem 1997;66:437–474.

Davies EK, Peters AD, Keightley PD: High frequency of cryptic deleterious mutations in *Caenorhabditis elegans*. Science 1999;285: 1748–1751.

Dean AM, Thornton JW: Mechanistic approaches to the study of evolution: the functional synthesis. Nat Rev Genet 2007;8:675–688.

de la Iglesia F, Elena SF: Fitness declines in tobacco etch virus upon serial bottleneck transfers. J Virol 2007;81:4941–4947.
Denver DR, Morris K, Lynch M, Vassilieva LL, Thomas WK: High direct estimate of the mutation rate in the mitochondrial genome of *Caenorhabditis elegans*. Science 2000;289: 2342–2344.
DePristo MA, Weinreich DM, Hartl DL: Missense meanderings in sequence space: a biophysical view of protein evolution. Nat Rev Genet 2005;6:678–687.
Desai MM, Fisher DS, Murray AW: The speed of evolution and maintenance of variation in asexual populations. Curr Biol 2007;17:385–394.
de Visser JA, Zeyl CW, Gerrish PJ, Blanchard JE, Lenski RE: Diminishing returns from mutation supply rate in asexual populations. Science 1999;283:404–406.
Dickinson WJ: Synergistic fitness interactions and a high frequency of beneficial changes among mutations accumulated under relaxed selection in *Saccharomyces cerevisiae*. Genetics 2008;178:1571–1578.
Drake JW: A constant rate of spontaneous mutation in DNA-based microbes. Proc Natl Acad Sci USA 1991;88:7160–7164.
Elena SF, Ekunwe L, Hajela N, Oden SA, Lenski RE: Distribution of fitness effects caused by random insertion mutations in *Escherichia coli*. Genetica 1998;102–103:349–358.
Elena SF, Moya A: Rate of deleterious mutation and the distribution of its effects on fitness in vesicular stomatitis virus. J Evol Biol 1999; 12:1078–1088.
Estes S, Phillips PC, Denver DR, Thomas WK, Lynch M: Mutation accumulation in populations of varying size: the distribution of mutational effects for fitness correlates in *Caenorhabditis elegans*. Genetics 2004;166: 1269–1279.
Eyre-Walker A, Keightley PD: The distribution of fitness effects of new mutations. Nat Rev Genet 2007;8:610–618.
Fisher RA: The Genetical Theory of Natural Selection. Oxford, Clarendon Press, 1930.
Funchain P, Yeung A, Stewart JL, Lin R, Slupska MM, Miller JH: The consequences of growth of a mutator strain of *Escherichia coli* as measured by loss of function among multiple gene targets and loss of fitness. Genetics 2000;154:959–970.
Gerrish PJ, Lenski RE: The fate of competing beneficial mutations in an asexual population. Genetica 1998;102–103:127–144.
Gillespie JH: Molecular evolution over the mutational landscape. Evolution 1984;38:1116–1129.
Gordo I, Dionisio F: Nonequilibrium model for estimating parameters of deleterious mutations. Phys Rev E Stat Nonlin Soft Matter Phys 2005;71:031907.
Haag-Liautard C, Coffey N, Houle D, Lynch M, Charlesworth B, Keightley PD: Direct estimation of the mitochondrial DNA mutation rate in *Drosophila melanogaster*. PLoS Biol 2008;6:e204.
Haag-Liautard C, Dorris M, Maside X, Macaskill S, Halligan DL, Houle D, Charlesworth B, Keightley PD: Direct estimation of per nucleotide and genomic deleterious mutation rates in Drosophila. Nature 2007;445:82–85.
Hall BG: Spectra of spontaneous growth-dependent and adaptive mutations at ebgR. J Bacteriol 1999;181:1149–1155.
Hallatschek O: The noisy edge of traveling waves. Proc Natl Acad Sci USA 2011;108: 1783–1787.
Hegreness M, Shoresh N, Hartl D, Kishony R: An equivalence principle for the incorporation of favorable mutations in asexual populations. Science 2006;311:1615–1617.
Herring CD, Raghunathan A, Honisch C, et al: Comparative genome sequencing of *Escherichia coli* allows observation of bacterial evolution on a laboratory timescale. Nat Genet 2006;38:1406–1412.
Hill WG, Robertson A: The effect of linkage on limits to artificial selection. Genet Res 1966; 8:269–294.
Hughes AL: The evolution of functionally novel proteins after gene duplication. Proc Biol Sci 1994;256:119–124.
Hutchison CA, Peterson SN, Gill SR, Cline RT, White O, Fraser CM, Smith HO, Venter JC: Global transposon mutagenesis and a minimal Mycoplasma genome. Science 1999;286: 2165–2169.
Ibarra RU, Edwards JS, Palsson BO: *Escherichia coli* K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. Nature 2002;420:186–189.
Imhof M, Schlotterer C: Fitness effects of advantageous mutations in evolving *Escherichia coli* populations. Proc Natl Acad Sci USA 2001;98:1113–1117.
Joseph SB, Hall DW: Spontaneous mutations in diploid *Saccharomyces cerevisiae*: more beneficial than expected. Genetics 2004;168: 1817–1825.
Kassen R, Bataillon T: Distribution of fitness effects among beneficial mutations before selection in experimental populations of bacteria. Nat Genet 2006;38:484–488.
Keightley PD: The distribution of mutation effects on viability in *Drosophila melanogaster*. Genetics 1994;138:1315–1322.
Keightley PD, Caballero A: Genomic mutation rates for lifetime reproductive output and lifespan in *Caenorhabditis elegans*. Proc Natl Acad Sci USA 1997;94:3823–3827.
Kibota TT, Lynch M: Estimate of the genomic mutation rate deleterious to overall fitness in *E. coli*. Nature 1996;381:694–696.
Kimura M: The Neutral Theory of Molecular Evolution. Cambridge, Cambridge University Press, 1983.
Lazaro E, Escarmis C, Perez-Mercader J, Manrubia SC, Domingo E: Resistance of virus to extinction on bottleneck passages: study of a decaying and fluctuating pattern of fitness loss. Proc Natl Acad Sci USA 2003;100: 10830–10835.
Lenski RE, Rose MR, Simpson SC, Tadler SC: Long-term experimental evolution in *Escherichia coli*. 1. Adaptation and divergence during 2,000 generations. Am Nat 1991;138: 1315–1341.
Levin BR: Frequency-dependent selection in bacterial populations. Philos Trans R Soc Lond B Biol Sci 1988;319:459–472.
Li H, Roossinck MJ: Genetic bottlenecks reduce population variation in an experimental RNA virus population. J Virol 2004;78: 10582–10587.
Lind PA, Berg OG, Andersson DI: Mutational robustness of ribosomal protein genes. Science 2010;330:825–827.
Loewe L: A framework for evolutionary systems biology. BMC Syst Biol 2009;3:27.
Lunzer M, Golding GB, Dean AM: Pervasive cryptic epistasis in molecular evolution. PLoS Genet 2010;6:e1001162.
Maisnier-Patin S, Berg OG, Liljas L, Andersson DI: Compensatory adaptation to the deleterious effect of antibiotic resistance in *Salmonella typhimurium*. Mol Microbiol 2002;46: 355–366.
Maisnier-Patin S, Roth JR, Fredriksson A, Nystrom T, Berg OG, Andersson DI: Genomic buffering mitigates the effects of deleterious mutations in bacteria. Nat Genet 2005;37: 1376–1379.
Martin G, Lenormand T: The fitness effect of mutations across environments: a survey in light of fitness landscape models. Evolution 2006;60:2413–2427.
Martin G, Lenormand T: The distribution of beneficial and fixed mutation fitness effects close to an optimum. Genetics 2008;179: 907–916.
McDonald MJ, Cooper TF, Beaumont HJ, Rainey PB: The distribution of fitness effects of new beneficial mutations in *Pseudomonas fluorescens*. Biol Lett 2011;7:98–100.
Nguyen TN, Phan QG, Duong LP, Bertrand KP, Lenski RE: Effects of carriage and expression of the Tn10 tetracycline-resistance operon on the fitness of *Escherichia coli* K12. Mol Biol Evol 1989;6:213–225.
Nichols RJ, Sen S, Choo YJ, et al: Phenotypic landscape of a bacterial cell. Cell 2011;144: 143–156.
Ochman H, Lawrence JG, Groisman EA: Lateral gene transfer and the nature of bacterial innovation. Nature 2000;405:299–304.
Orr HA: Theories of adaptation: what they do and don't say. Genetica 2005;123:3–13.
Orr HA: Fitness and its role in evolutionary genetics. Nat Rev Genet 2009;10:531–539.

Papadopoulos D, Schneider D, Meier-Eiss J, Arber W, Lenski RE, Blot M: Genomic evolution during a 10,000-generation experiment with bacteria. Proc Natl Acad Sci USA 1999; 96:3807–3812.
Park SC, Simon D, Krug J: The speed of evolution in large asexual populations. J Stat Phys 2010; 381–410.
Perfeito L, Fernandes L, Mota C, Gordo I: Adaptive mutations in bacteria: high rate and small effects. Science 2007;317:813–815.
Perfeito L, Ghozzi S, Berg J, Schnetz K, Lässig M: Nonlinear fitness landscape of a molecular pathway. PLoS Genet 2011;7:e1002160.
Rouzine IM, Wakeley J, Coffin JM: The solitary wave of asexual evolution. Proc Natl Acad Sci USA 2003;100:587–592.
Rozen DE, de Visser JA, Gerrish PJ: Fitness effects of fixed beneficial mutations in microbial populations. Curr Biol 2002;12:1040–1045.
Salyers AA, Shoemaker NB, Stevens AM, Li LY: Conjugative transposons: an unusual and diverse set of integrated gene transfer elements. Microbiol Rev 1995;59:579–590.
Schneider D, Duperchy E, Coursange E, Lenski RE, Blot M: Long-term experimental evolution in *Escherichia coli*. IX. Characterization of insertion sequence-mediated mutations and rearrangements. Genetics 2000;156: 477–488.
Schneider D, Lenski RE: Dynamics of insertion sequence elements during experimental evolution of bacteria. Res Microbiol 2004;155: 319–327.
Schrag SJ, Perrot V: Reducing antibiotic resistance. Nature 1987;381:120–121.
Sharp PM, Emery LR, Zeng K: Forces that influence the evolution of codon bias. Philos Trans R Soc Lond B Biol Sci 2010;365:1203–1212.
Silva RF, Mendonça SCM, Carvalho LM, Reis AM, Gordo I, et al: Pervasive sign epistasis between conjugative plasmids and drug-resistance chromosomal mutations. PLoS Genet 2011;7:e1002181.
Sniegowski PD, Gerrish PJ: Beneficial mutations and the dynamics of adaptation in asexual populations. Philos Trans R Soc Lond B Biol Sci 2010;365:1255–1263.
Stevens K, Sebert M: Frequent beneficial mutations during single-colony serial transfer of *Streptococcus pneumoniae*. Plos Genet 2011; 7:e1002232.
Trindade S, Perfeito L, Gordo I: Rate and effects of spontaneous mutations that affect fitness in mutator *Escherichia coli*. Philos Trans R Soc Lond B Biol Sci 2010;365:1177–1186.
Vassilieva LL, Lynch M: The rate of spontaneous mutation for life-history traits in *Caenorhabditis elegans*. Genetics 1999;151:119–129.
Wagner A: Distributed robustness versus redundancy as causes of mutational robustness. Bioessays 2005;27:176–188.
Weinreich DM, Delaney NF, Depristo MA, Hartl DL: Darwinian evolution can follow only very few mutational paths to fitter proteins. Science 2006;312:111–114.
Wloch DM, Szafraniec K, Borts RH, Korona R: Direct estimate of the mutation rate and the distribution of fitness effects in the yeast *Saccharomyces cerevisiae*. Genetics 2001;159: 441–452.
Wylie CS, Shakhnovich EI: A biophysical protein folding model accounts for most mutational fitness effects in viruses. Proc Natl Acad Sci USA 2011;108:9916–9921.
Yamamoto N, Nakahigashi K, Nakamichi T, et al: Update on the Keio collection of *Escherichia coli* single-gene deletion mutants. Mol Syst Biol 2009;5:335.
Zeyl C, DeVisser JA: Estimates of the rate and distribution of fitness effects of spontaneous mutation in *Saccharomyces cerevisiae*. Genetics 2001;157:53–61.