The distribution of fitness effects of new mutations

Adam Eyre-Walker*[‡] and Peter D. Keightley[§]

Abstract | The distribution of fitness effects (DFE) of new mutations is a fundamental entity in genetics that has implications ranging from the genetic basis of complex disease to the stability of the molecular clock. It has been studied by two different approaches: mutation accumulation and mutagenesis experiments, and the analysis of DNA sequence data. The proportion of mutations that are advantageous, effectively neutral and deleterious varies between species, and the DFE differs between coding and non-coding DNA. Despite these differences between species and genomic regions, some general principles have emerged: advantageous mutations are rare, and those that are strongly selected are exponentially distributed; and the DFE of deleterious mutations is complex and multi-modal.

Muller's ratchet
The process by which a genome with little or no recombination degenerates owing to the stochastic loss of the allelic class with fewest deleterious mutations.

All organisms undergo mutation, the effects of which can be broadly divided into three categories. First, there are mutations that are harmful to the fitness of their host; these mutations generally either reduce survival or fertility. Second, there are 'neutral' mutations, which have little or no effect on fitness. Finally, there are advantageous mutations, which increase fitness by allowing organisms to adapt to their environment. Although we can divide mutations into these three categories, there is, in reality, a continuum of selective effects, stretching from those that are strongly deleterious, through weakly deleterious mutations, to neutral mutations and then on to mutations that are mildly or highly adaptive. The relative frequencies of these types of mutation — the distribution of fitness effects (DFE) — is the subject of this Review (see BOX 1 for a description of some distributions).

The DFE is important for several reasons. First, it is of some intrinsic interest, particularly as far as we humans are concerned. It has been estimated that each of us receives more than 100 new mutations from our parents[1,2]. What effects do these mutations have? Are they good, bad or irrelevant to our well-being? Second, the DFE is central to many questions in evolutionary biology, including the molecular clock[3], the rate of genomic decay due to Muller's ratchet[4], the maintenance of genetic variation at the molecular level[5], and the evolution of sex and recombination[6].

The DFE is possibly of greatest practical importance in relation to two other problems: understanding the nature of quantitative genetic variation and hence complex human disease[7–9], and predicting the consequences of maintaining animals or plants at low population size, as in captive breeding programmes[10]. Many of the characters that are of most interest to geneticists are quantitative in nature; these include traits as diverse as milk yield in dairy cows and the probability of developing heart disease in humans. One of the central aims of quantitative genetics, and in particular medical genetics, is to map the alleles that cause variation in these traits. However, the ease with which this can be done depends on the genetic architecture of the trait. If most variation is contributed by mutations of large effect segregating at intermediate frequencies, as is hoped under the common disease–common variant model[11], then it should be possible to locate the mutations that affect the trait. Conversely, if most of the variation is contributed by mutations of large effect segregating at low frequencies, or mutations of small effect segregating at intermediate frequencies, then identifying the causal mutations will be difficult. The genetic architecture of the trait depends on both the nature of selection and the DFE[12].

Although the DFE has been the subject of much research, there has been no attempt to review the topic as a whole. This might be because a wide variety of approaches have been used to address the question, including theoretical, experimental and analytical methods. Here we attempt to fill this gap. We begin by briefly reviewing the approaches that have been used to investigate the DFE, before discussing what we have learnt about this important but elusive entity.

Experimental techniques
The most direct method for investigating the DFE is to induce or collect spontaneous mutations and then assay

*Centre for the Study of Evolution, University of Sussex, Brighton, BN1 9QG, UK.
‡National Evolutionary Synthesis Center, Durham, North Carolina 27705, USA.
*§Institute of Evolutionary Biology, University of Edinburgh, Edinburgh, EH9 3JT, UK.*
Correspondence to A.E.W.
e-mail:
a.c.eyre-walker@sussex.ac.uk
doi:10.1038/nrg2146

Box 1 | Mathematical distributions

A number of mathematical distributions have been used to model the distribution of fitness effects. The most common among these is the gamma distribution. This is a flexible distribution that has two parameters: a shape parameter, k, and a parameter that governs the mean of the distribution. Alteration of the shape parameter changes the distribution from an L-shaped distribution when $k < 1$, to an exponential distribution when $k = 1$, to a distribution that resembles a skewed normal distribution or a normal distribution when $k > 1$. The figure shows the probability density of the gamma distribution with a mean of 1 and varying shape parameters. If the shape parameter is large then the distribution becomes a spike, with all mutations having the same strength of selection.

The kurtosis of a distribution is the degree to which it is peaked; the greater the kurtosis, the more variance is contributed by infrequent observations of large effect. In this Review, the term leptokurtic is used to refer to a gamma distribution that is more leptokurtic than an exponential distribution — that is, it has a shape parameter of less than one. This is slightly different to the usual definition by which kurtosis is defined relative to that of the normal distribution.
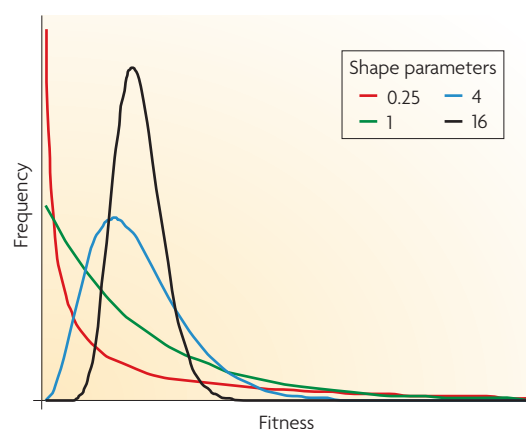
their effects on fitness in the laboratory (see FIG. 1 for an example). Unfortunately, accurate measurement of the effects of single mutations is possible only when they have fairly large effects on fitness (say >1%; that is, a mutation that increases or decreases viability or fertility by more than 1%)[13–17]. Furthermore, such experiments are difficult and time consuming, so they have been done in relatively few species, mostly microorganisms.

An alternative approach that has been extensively used to investigate the rates and effects of spontaneous mutations is the mutation accumulation experiment. In a mutation accumulation experiment, sublines (mutation accumulation lines) derived from a genetically uniform ancestral line are allowed to independently accumulate spontaneous mutations for many generations[18], or sublines are independently subjected to a mutagenesis treatment[19]. Each mutation accumulation line is kept at a small population size and under benign conditions during the period of mutation accumulation or after mutagenesis. In this way, the effects of natural selection are minimized, allowing all but the most deleterious mutations to accumulate. After the period of mutation accumulation, the fitnesses of the lines are measured in parallel with controls. The controls are often derivatives of the ancestral line, maintained in such a way that they can be assumed to be essentially mutation-free. These can be, for example, cryopreserved cells, embryos, eggs or dried seeds.

In general, mutation accumulation lines decrease in fitness as the experiment progresses and variance between lines increases[20] (but see REFS 21,22). This pattern is consistent with a net accumulation of deleterious mutations, some of which are strongly deleterious: these generate most of the variance between lines. It is straightforward to calculate the minimum rate and the maximum mean effect of deleterious mutations that explain the data from a mutation accumulation experiment, assuming independence of mutational effects[18,23], and several methods have also been developed to estimate the DFE from the fitnesses of the mutation accumulation lines[22,24,25]. Unfortunately, with realistic amounts of data, the amount of additional information that can be extracted about the DFE is limited, and parameter

estimates have large confidence intervals[26]. This is because the mutation rate and DFE are confounded with one another: the distribution of fitnesses of the mutation accumulation lines can usually be explained by a low mutation rate and high variance in fitness effects, or a high mutation rate and a low variance in fitness effects[26]. Some experiments have therefore sought to circumvent this problem by estimating the mutation rate directly in the mutation accumulation experiment[27,28]. These experiments, which are discussed below, have generally been much more informative about the DFE.

A potential problem with many mutagenesis and mutation accumulation experiments is the method that is used to generate the mutations. Ideally, we would like to measure the fitness effects of spontaneous mutations but, often, this is not possible, and the mutations are induced by either transposable element insertion or a chemical mutagen. As such, the DFE that is inferred from mutagenesis and mutation accumulation experiments might be different to the distribution of spontaneously occurring mutations; there is indeed evidence of this in yeast[17] (FIG. 2).

Mutagenesis and mutation accumulation experiments can give us detailed information about the DFE of mutations only if they have a moderately large effect, as these are the mutations that have detectable effects in laboratory assays. However, it seems likely that many and possibly the majority of mutations have effects that are too small to be detected in the laboratory[27,28]. For these mutations, DNA sequence data can provide valuable information. By examining DNA differences between and within species it is possible to infer various characteristics of the DFE for neutral, deleterious and slightly advantageous mutations.

Various methods have been developed to infer the DFE from DNA sequence data[9,29–34], all of which rely on two population genetic properties of mutations. First, they tap into the fact that the probability of a mutation spreading to a certain frequency in a population, and ultimately to fixation, depends on the strength of positive or negative selection acting on it. The more deleterious a mutation, the less likely it is to spread to high frequency in the population and to become fixed. The

Kurtosis
The peakedness of the distribution.
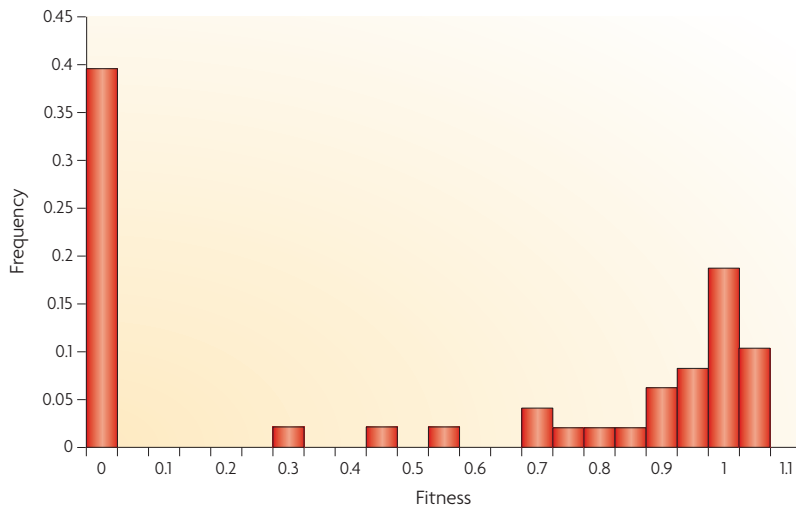
Figure 1 | **The distribution of fitness effects of random mutations in vesicular stomatitis virus.** In this experiment, random mutations were introduced into the virus, and the fitnesses of the mutants were compared against the unmutated wild type. A fitness of less than one indicates that the mutant was less fit than the wild type, so the mutation was deleterious. A fitness of zero indicates that no mutated progeny were recovered, and that the mutation was therefore lethal. Data from REF. 15.

second source of information comes from the fact that the efficiency of selection acting on a mutation depends on the effective population size: if the product of the effective population size, N_e, and the strength of selection, s, is much greater than one, then selection is effective; but, if $N_e s$ is much less than one, then the fate of the mutation is determined largely by random genetic drift. Hence, by comparing species, or genomic regions with different effective population sizes, it is possible to derive information about the DFE.

Although some of the methods that use DNA sequence data are quite powerful, it should be appreciated that they share several limitations. First, all methods require a set of sites at which most mutations are neutral, against which the evolution at selected sites can be compared. If mutations at these putatively neutral sites are in fact subject to selection, the inferences of the DFE could be seriously affected in ways that have not yet been determined. Second, in those cases in which a continuous distribution is fitted, one can be misled into believing that there is more information than actually exists. This is because there are horizons above and below which all mutations effectively act in the same manner.

Consider the following example. Recently, the DFE was inferred for new amino-acid-changing mutations in humans, under the assumption that the DFE is a gamma distribution (BOX 1), by using the frequencies of SNPs[9] (FIG. 3). About 43% of mutations are estimated to have effects such that $N_e s > 100$ and, of those, 30% are expected to have effects such that $N_e s > 1,000$. However, mutations for which $N_e s > 100$ have essentially no chance of appearing as polymorphisms in a reasonable sample size; so, in this analysis, all mutations with $N_e s > 100$ act in the same manner. Hence, the inference that 30% of mutations with $N_e s > 100$ have $N_e s > 1,000$ is almost entirely based on the assumption that the DFE follows

a gamma distribution, rather than on the actual experimental data. Third, mutations are generally assumed to segregate independently of one another. Analysis has suggested that this assumption of free recombination does not greatly affect the point estimates of parameters, but it can lead to a gross underestimation of the true confidence intervals on the estimates[35]. Fourth, although we are interested in the fitness effects of all mutations, almost all analyses have concentrated on a subset of mutations that exert effects that are independent of their frequency in the population and that stay approximately constant during the time they segregate in the population. Mutations that are subject to heterozygote advantage or frequency-dependent selection have not been studied in any detail with regard to the DFE. Unfortunately, there is little information on the proportion of mutations that are subject to various types of balancing selection. Recent evidence from a scan of human SNP variation supports the view that this phenomenon is unusual[36].
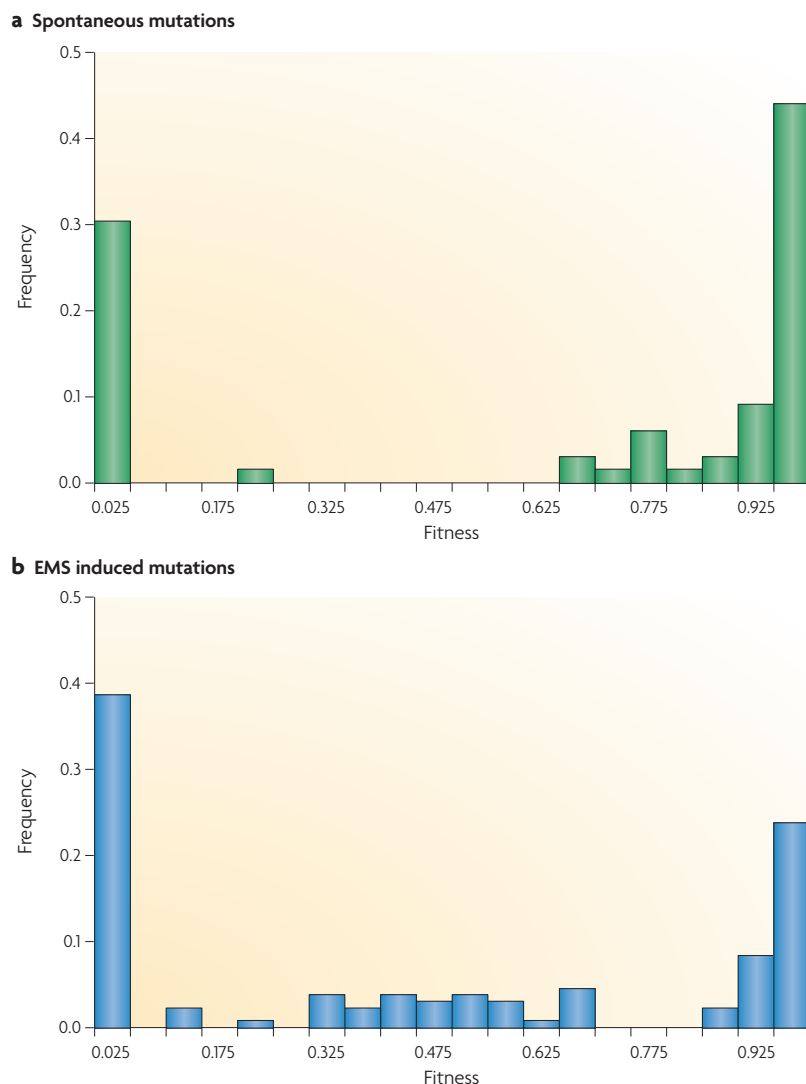
Neutral mutations

The first question one might ask about the DFE is: what proportion of mutations are neutral? As with many questions pertaining to the DFE, this has no easy answer. The first point to make is one of definition; it seems unlikely that any mutation is truly neutral in the sense that it has no effect on fitness. All mutations must have some effect, even if that effect is vanishingly small. However, there is a class of mutations that we can term effectively neutral. These are mutations for which $N_e s$ is much less than 1, the fate of which is largely determined by random genetic drift[3,37]. As such, the definition of neutrality is operational rather than functional; it depends on whether natural selection is effective on the mutation in the population or the genomic context in which it segregates, not solely on the effect of the mutation on fitness.

This definition of neutrality implies that the proportion of neutral mutations is expected to vary between species for two reasons. First, for a given DFE, a smaller proportion of mutations will be effectively neutral in species with large effective population sizes, a prediction that has been used to test for the presence of weakly selected deleterious mutations by, for example, comparing rates of non-synonymous and synonymous substitution in island and mainland species[38,39]. Second, because a smaller proportion of mutations are expected to be effectively neutral in species with large effective population sizes, one might expect such species to be better adapted. This is because fewer advantageous mutations will be effectively neutral and, as a consequence, more advantageous mutations will fixed. This is itself likely to change the DFE because, as a species adapts, its fitness is expected to move closer to an optimum[40].

There might also be second-order effects that cause the proportion of effectively neutral mutations to vary between species. For example, it has been suggested that species with small effective population sizes are susceptible to the accumulation of various kinds of non-protein-coding DNA, especially transposable element insertions, because the effectiveness of selection is reduced in

Effective population size
The population size of randomly mating individuals that would behave, in a population genetic sense, as the population being studied. For example, the genetic diversity in human populations is the same as one would find in 10,000 randomly mating individuals.

a Spontaneous mutations

b EMS induced mutations

Figure 2 | **The distribution of fitnesses among yeast lines.** Diploid yeast lines were either allowed to accumulate spontaneous mutations (panel **a**) or were subject to chemical mutagenesis using ethylmethane sulphonate (EMS) (panel **b**). After a period of inbreeding, the cells were made to undergo meiosis and the growth of the meiotic products was measured. Data from REF. 17.

such populations[41]. Mutations in these sequences are themselves likely to be neutral.

Although the proportion of effectively neutral mutations is likely to vary between species, there are a few general conclusions that can be made. First, analyses of protein-coding sequences suggest that only a small proportion of amino-acid-changing (non-synonymous) mutations are neutral. By considering the number of substitutions that have occurred between species at non-silent sites (d_n) and silent (d_s) sites, it is possible to estimate the proportion of non-silent mutations that are effectively neutral (BOX 2). In hominids, which seem to have effective population sizes in the range of 10,000 to 30,000 (REF. 29), the ratio d_n/d_s is less than 0.3 (REFS 29,42), and this suggests that fewer than 30% of amino-acid-changing mutations are effectively neutral. Given that mutations with a selection strength of much less than

$1/N_e$ are effectively neutral, this means that about 30% of amino-acid mutations in humans have effects of less than about 10^{-5}. Similar calculations in *Drosophila* species and enteric bacteria, which probably have effective population sizes in the millions and tens of millions, respectively, suggest that at most 16% (REF. 43) and 2.8% (REF. 44), respectively, of mutations are effectively neutral. The figures for *Drosophila* and bacteria might be overestimated by more than 50% because of adaptive substitutions (BOX 2). These observations might also indicate that most amino-acid-changing mutations are deleterious; for example, if we infer that at most 30% of non-synonymous mutations are neutral in humans, this implies that at least 70% are deleterious. Similarly, in *Drosophila* and enteric bacteria, the proportions are at least 84% and 97.2%, respectively.

The proportion of mutations that behave as effectively neutral occurring outside protein-coding sequences is much less clear. It is probably fair to say that until recently the majority of evolutionary biologists regarded most non-coding DNA as evolving neutrally, a view that led Orgel and Crick to term it 'junk' DNA[45]. However, this perspective has started to shift. In yeast[46], nematodes[47,48], *Drosophila melanogaster*[49–51] and mammals[52–54] a certain proportion of non-coding DNA seems to be more conserved than would be expected if all mutations were neutral. In yeast and nematodes, the proportion of non-coding nucleotides that is conserved by natural selection has been estimated to be 10–20% (REFS 46,48). By contrast, in *D. melanogaster*, the available evidence suggests that at least 50% of sites in non-coding DNA are constrained by natural selection[49,51]. In mammals, the proportion of the genome that is subject to natural selection is much lower, around 5% (REFS 55–57). It therefore seems likely that as much as 95% and as little as 50% of mutations in non-coding DNA are effectively neutral; therefore, correspondingly, as little as 5% and as much as 50% of mutations are deleterious. The differences between the estimates from different species might partly reflect differences in methodology.

Advantageous mutations

As expected, relatively few of the mutations that are not effectively neutral are advantageous. In three mutagenesis experiments, the proportion of advantageous mutations was 4% in the RNA virus vesicular stomatitis virus (VSV)[15] (FIG. 1), 0% in *Escherichia coli*[14], 0–15% in the bacteriophage ϕX174 (REF. 40), 0% in ϕ6 (REF. 13) and 6% in *Saccharomyces cerevisiae*[16]. However, although advantageous mutations are rare, they can contribute substantially to evolutionary change[58]. For example, in *D. melanogaster*, it has been estimated that more than 15% of all substitutions are due to advantageous mutations[49]. However, such analyses measure substitution rates rather than mutation rates, and do not tell us directly about the frequency of advantageous mutation. A certain amount of substitution could be due to a few strongly selected mutations, or many weakly selected mutations — most mutations, even those that are advantageous, are lost by random genetic drift; but the more strongly selected an advantageous mutation is, the less likely it is to be lost.
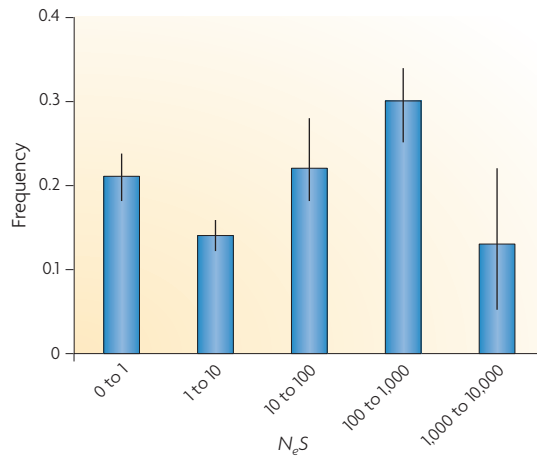
Figure 3 | **The distribution of fitness effects of new amino-acid-changing mutations in humans.** $N_e s$ denotes the product of the effective population size, N_e, and the strength of selection, s. The distribution was inferred from nearly 1,000 non-synonymous and more than 30,000 intron SNPs from 230 autosomal genes that were resequenced in 90 individuals. The proportion of mutations with strongly deleterious effects has probably been overestimated because of demographic effects[9]. The absolute strengths of selection can be inferred from this analysis because the effective population size of humans is approximately 10,000. Hence, mutations with $N_e s > 1,000$ have effects of greater than 10%. Data from REF. 9.

Although it is difficult to disentangle the rate and effects of advantageous mutations, the available data suggest that some aspects of the DFE of advantageous mutations are likely to differ between species. This has been most elegantly demonstrated in the bacteriophage φX174. Silander *et al.*[40] conducted a mutation accumulation experiment in which populations were passaged through a small population size of three individuals, or larger population sizes of up to 250 individuals. As expected, they found that the small population size lines had lower fitness because they had accumulated more deleterious mutations. However, they also found that 15% of the mutations in these lines were adaptive, in contrast to almost none in the large population size lines. Interestingly, the mean effect of mutations did not seem to differ between lines. Although this experiment was conducted in only one species, the results clearly suggest that the DFE is likely to differ between species that have different effective population sizes.

Other data corroborate this finding. For protein-coding sequences, the proportion of adaptively driven non-synonymous substitutions is estimated to be close to zero between humans and chimpanzees[59,60], about 50% between *Drosophila* species[61–63] (but see REF. 64), and might be even higher between species of enteric bacteria[44] and some viruses[32,65]. In hominid non-coding DNA, there is again little evidence of adaptive evolution[66], but in *Drosophila* it is estimated that ~15% of non-coding substitutions have been a consequence of positive selection[49]. The differences in the rates of adaptation

between species are either due to a difference in the rate of advantageous mutation or to a difference in the average effects of those mutations. The results from φX174 suggest that the first possibility is true but, clearly, more data are needed. Unfortunately, the rates and effects of mutations are generally difficult to disentangle.

Although the frequency of advantageous mutation remains an elusive quantity, progress has been made in understanding the shape of the DFE of new advantageous mutations. Theoretical work by Gillespie[67] and Orr[68], derived from a branch of mathematics called extreme value theory (EVT), predicts that the distribution should be exponential (BOX 1; FIG. 4). The only conditions are that the distribution of absolute fitness values of mutations should be invariant and that the population is reasonably well adapted to start with.

Several groups have sought to test the Gillespie–Orr theory by examining the fitness effects of new mutations. Generally, the prediction that the DFE of advantageous mutations is exponential has been supported[69–71] (but see REF. 15). However, the number of mutations that have been assessed is small, because advantageous mutations are rare. Hence, there is little power in these analyses to distinguish an exponential distribution from other similar distributions. Furthermore, the mutations that have been assayed are strongly advantageous, as these are the only mutations that are discernibly advantageous in laboratory assays.

In an attempt to circumvent the central problems of these tests — the rarity of advantageous mutations and size of the effect — Cowperthwaite *et al.*[72] carried out an *in silico* experiment in which they simulated the evolution of an RNA molecule to bind a ligand. They found that the DFE of advantageous mutations was poorly modelled by an exponential distribution because there was an excess of weakly advantageous mutations; only when 90% of the advantageous mutations were removed did the distribution become exponential.

This raises the question of whether these departures from the predictions of Gillespie–Orr theory are specific to this *in silico* system, or whether it uncovers a basic problem with the theory. Cowperthwaite *et al.*[72] suggest that the departure from the exponential distribution arises because one of the central assumptions of EVT is violated in this system. EVT assumes that mutations are drawn from some never-changing distribution; however, as an organism adapts, the DFE is likely to change (FIG. 4). In the *in silico* system studied by Cowperthwaite *et al.*, the DFE changes in such a way that there is always a category of advantageous mutations of small effect. Whether this is likely to be the case in real life is currently unclear and difficult to ascertain, but the data from the virus VSV are also consistent with a distribution that is more leptokurtic than exponential (R. Sanjuan, personal communication).

Although the Gillespie–Orr prediction might not be correct for all advantageous mutations, it may apply to mutations of large effect. Such large-effect mutations seem to be those that contribute most to adaptation. The evidence for this comes from several QTL-mapping experiments to determine the differences between species for traits of adaptive importance. In most cases, the differences

Leptokurtic
Used here to refer to distributions that are more peaked than an exponential distribution.

The simplest method by which we can obtain information about the distribution of fitness effects (DFE) using DNA sequence data is by comparing the number of substitutions at non-silent sites (d_n) to the number of substitutions at silent sites (d_s). Let us assume that all mutations at silent sites are neutral and that mutations at non-silent sites are either deleterious or 'effectively' neutral. By effectively neutral, we mean mutations that are either truly neutral, or deleterious mutations that behave as if they were neutral; these are mutations on which the strength of selection is much less than $1/N_e$ where N_e is the effective population size. Under this simple model, $d_s = 2ut$ where u is the nucleotide mutation rate, and hence the substitution rate, and t is the time of divergence of the two species that are being compared. $d_n = 2utf$ where f is the proportion of mutations that are effectively neutral. The ratio d_n/d_s, which is equal to f, can be used to estimate the proportion of non-silent mutations that are effectively neutral; in other words, the proportion of mutations for which $s \ll 1/N_e$. One can also estimate the proportion of mutations that are deleterious and on which selection is effective, since this is equal to $1-f$.

If there are advantageous mutations, then $d_n = 2utf/(1-\alpha)$ where α is the proportion of substitutions that are due to adaptive mutations[62]. From this it can be seen that the proportion of mutations that are effectively neutral tends to be overestimated by the ratio d_n/d_s. For example, if 50% of substitutions were due to adaptive evolution, then the proportion of effectively neutral mutations would be overestimated by twofold.

By comparing the d_n/d_s in species with different N_e, one can potentially gain further insight into the DFE[29]; however, this method makes the assumption that the DFE is relatively constant across species with different N_e.

between species in traits can be explained by a few QTLs of large effect[73–75], suggesting that just a few advantageous mutations were involved in the adaptations.

Deleterious mutations

The distribution may be complex. Although the DFE for advantageous mutations may have a relatively simple form, the DFE for deleterious mutations seems to be complex, in the sense that it does not appear to be described by a distribution with a single mode (that is, a single maximum). This is most readily apparent from experiments in which the fitness effects of single mutations can be assessed; in both RNA viruses and yeast, there is a distinct class of mutations that are lethal (FIGS 1,2), generating a bimodal DFE[15,17]. In VSV, nearly 40% of all randomly induced point mutations are lethal[15] (FIG. 1); similarly, in yeast, about 30–40% of mutations with fitness effects that are detectable in the laboratory are lethal[17] (FIG. 2). Mutation accumulation experiments in *D. melanogaster* that used balancer chromosomes[18,76] also identified a substantial class of lethal mutations.

There is also some evidence that the DFE of homozygous non-lethal mutations is multi-modal. Evidence for this comes from two experiments in which the nematode *Caenorhabditis elegans* was either mutagenized[27] or allowed to accumulate spontaneous mutations[77]. In each case, the phenotypic data were consistent with a model in which each line had a relatively small number of mutations with fairly similar effects. For example, Davies *et al.*[27] inferred that their lines had an average of 2.5 deleterious mutations, and the best fitting DFE was one in which all mutations had the same effect. Importantly, they could reject a model in which the mode was close to zero. However, in both experiments there was an estimate of the actual number of mutations in each line, and

Multi-modal
A distribution with more than one peak or mode.

from this minimum number of deleterious mutations could be inferred. Davies *et al.*[27] used a standardized dose of ethylmethane sulphonate (EMS), and inferred that each line had received, on average, approximately 200 new mutations (primarily G–C → A–T transitions). Furthermore, because protein-coding sequences are highly conserved in *Caenorhabditis* species, they estimated that at least 50 of these would be deleterious in the wild. Thus, the true rate of harmful mutation was at least 20-fold higher than the inferred mutation rate based on the analysis of fitnesses of the lines. Similar inferences were made by Denver *et al.*[28], on the basis of sequencing of the mutation accumulation lines of Vassilieva *et al.*[77] There are two explanations for these results: either the laboratory environment is so benign that most mutations that would be strongly deleterious in the wild have little or no effect in the laboratory, or there is a large category of deleterious mutations that have small effects on fitnesses. If we accept this second explanation, then we can also infer that the DFE is bimodal, with one mode near the effect sizes that are inferred from the experiment, and the other including the vast majority of deleterious mutations, which have small effects.

Results from other mutation accumulation experiments in which a direct estimate of the mutation rate is not available are generally consistent with the results reported by Davies *et al.*[27] and Denver *et al.*[28]: the DFE that fits the mutation accumulation line data is usually consistent with a distribution that has a mode above zero. Such experiments have been done in VSV[78] and φ6 (REF. 13), yeast[21,79], *Drosophila*[80–82] and plants from the genera *Amsinckia*[83] and *Arabidopsis*[22]. However, in many cases, the confidence intervals are large.

The fact that the DFE for deleterious mutations is a complex, multi-modal distribution is perhaps not surprising, given that there are several classes of sites and mutations that are likely to have different DFEs; for example, it seems likely that the DFE for non-synonymous mutations is different to that of non-coding mutations, and that point mutations will have smaller effects than, say, transposable element insertions. The DFE might therefore be the sum of several component distributions that make it multi-modal.

Although little is known about the DFE for heterozygous effects of mutations, the distribution might be less leptokurtic than that for homozygous mutations, as most evidence suggests that mutations with small fitness effects are partially recessive, whereas mutations with larger effects tend to be more recessive[84–86].

The distribution of mild-effect mutations. Although most deleterious mutations might have effects that are undetectable in an mutation accumulation experiment, analyses of DNA sequence can inform us about the DFE of mutations that have mild effects. Several methods have been developed and applied to various data from animals. In general, they yield a fairly consistent picture. The DFE of amino-acid-changing mutations is strongly leptokurtic; that is, the data are generally consistent with a gamma distribution that has a shape parameter of much less than one[9,30,31,33], with the one exception of
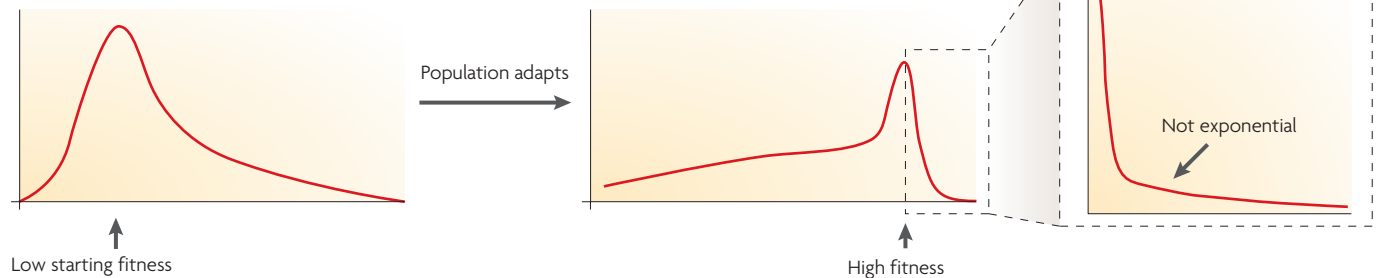
a Evolution under the extreme value theory

Population adapts

Low starting fitness

High fitness

Exponential
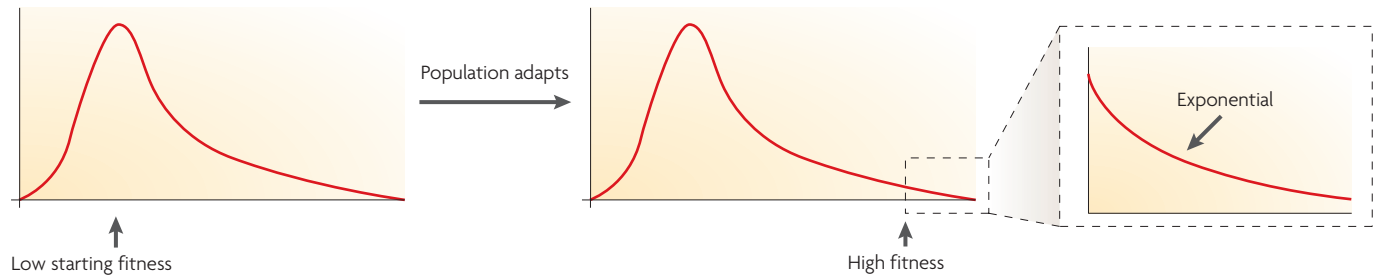
b Evolution in *in silico* experiments

Population adapts

Low starting fitness

High fitness

Not exponential

Figure 4 | **The evolution of the distribution of fitness effects (DFE) of advantageous mutations. a** | As observed under the assumptions of the extreme value theory. **b** | As observed during computer simulations. In both cases, the population evolves from a state of low fitness to a state of high fitness. However, under extreme value theory, the distribution of mutant effects is assumed to remain constant as the population evolves; this implies that the DFE for advantageous mutations becomes increasingly like an exponential distribution as the population evolves towards the right-hand tail of the distribution. By contrast, in computer simulations, the underlying distribution of mutant effects seems to change as the population evolves, such that there is always an excess of slightly advantageous mutations. Arrows pointing upwards indicate the mean fitness of the population.

the analysis of primate mitochondrial DNA by Nielsen and Yang[32].

For the most part, the estimates of the DFE from DNA sequence data have large confidence intervals. However, in humans, there is a fairly good estimate of the DFE for non-synonymous mutations (FIG. 3). This estimate has small confidence intervals because the method uses almost all the available information, and a large amount of data is available, both in terms of the number of genes (230 in this analysis) and the number of alleles sampled (180). In particular, the depth of sampling allows the method to accurately infer the shape of the distribution, because even mutations of strong effect are expected to be segregating in the sample of DNA sequences and hence to yield information about their effects. The available data suggest that relatively few amino-acid-changing mutations have effects of greater than 10% in humans, and that most have effects in the range of 10^{-3} and 10^{-1} (FIG. 3). However, two points of caution should be made. First, the method is ultimately limited in what it can infer about the DFE for mutations of very small and large effects (see above). To obtain more information about mutations with large effect, we would need to sequence many more alleles. To obtain information about mutations with weak effect, we would need to do the analysis in a species with a larger effective population size than we humans have. Second, the analysis

was carried out on data for which the demography is unknown; correcting for this factor is important in these kinds of analyses[9,87].

The DFE for mutations that occur outside protein-coding sequences is less well characterized. Most information comes from estimates of the fraction of mutations that are removed by natural selection (selective constraint), obtained by comparing the divergence of a sequence to that of some putatively neutrally evolving sequence. Two analyses that compared levels of selective constraint between hominids and murids suggest that many of the mutations that are selected might be weakly selected. Although there is a significant difference in the level of constraint in protein-coding sequences between hominids and murids[29,88,89], the difference seems to be much more dramatic for non-coding DNA, suggesting that a greater proportion of mutations in non-coding DNA are weakly deleterious. For example, the level of constraint upstream and downstream of genes is much lower in hominids than in rodents[66,90]. This effect probably arises because selection has been less effective in hominids because of their much smaller long-term effective population size. Similarly, Keightley *et al.*[91] and Kryukov *et al.*[92] found significantly lower conservation among blocks of non-coding DNA in hominids than in murids, blocks that had previously been identified as being significantly conserved between humans and mice.

The mean strength of selection. The mean strength of selection acting against deleterious mutations is a quantity that is important in, for example, the prediction of the likely consequences of maintaining a species as a small population, as one might do in a captive breeding programme. However, the mean effect of a deleterious mutation is not easy to measure. The only direct estimate we have is from mutagenesis experiments in VSV, in which the average effect of a new mutation has been estimated at 47% (REF. 15).

Mutation accumulation experiments can also give some information about the effects of mutations; in such an experiment, it is possible to estimate the maximum mean effect of mutations. This value typically comes out at a few percent[93]. However, the true mean might be much lower than this upper limit, because the upper limit is inferred under the assumption that all mutations have the same effect on fitness — this is clearly not the case. It should also be appreciated that lethal mutations are not assayed in most mutation accumulation experiments.

DNA sequence data can also yield estimates of the mean strength of selection acting against deleterious mutations. These estimates are for a particular category of mutations, usually non-synonymous mutations, rather than for all mutations, and the estimates typically have large standard errors. Furthermore, the mean of the DFE largely depends on the distribution for mutations of large effect, and it is these mutations that DNA-based methods have little direct information about (see above). In humans, the only species for which we have an estimate without large standard errors, the mean effect of a new non-synonymous mutation is estimated to be a few percent[9].

Conclusions

A number of general conclusions can be drawn about the DFE. First, the DFE seems to vary between species. This is evident in at least two different facets of the distribution: the proportion of mutations that are effectively neutral varies between species because it depends on the effective population size; and, the average effect of deleterious mutations varies dramatically. Second, the DFE of advantageous mutations seems to be exponential in character, at least for strongly advantageous mutations. Third, the DFE for deleterious mutations is probably both complex in character and variable between species. Fourth, focusing our attention on particular types of mutation, it seems that the DFE of non-synonymous mutations is leptokurtic and that non-coding mutations have a different DFE from coding mutations, with the DFE of non-coding DNA containing many more weakly selected mutations.

Although we have made some progress in understanding the properties of the DFE, there is still much to be learned. This raises a crucial question: can we ever really know what the DFE is? For a simple organism like a virus, this does seem to be possible, because most mutations have large effects that can be assayed in the laboratory. For most other organisms, and particularly for multicellular organisms, quite the opposite is the case; most mutations, even if they are deleterious, have such small effects that one cannot measure their fitness consequences. Furthermore, the environment in which most organisms live is probably sufficiently complex that laboratory assays give only the crudest measure of fitness. However, comparative methods using DNA sequence data potentially allow us to circumvent both of these problems. By examining the pattern of polymorphism in a species, we can estimate the effects of mutations with very small effects[9] and we can infer the overall fitness consequences of mutations in the environment in which the species evolved. Furthermore, if we are prepared to sequence hundreds, if not thousands, of alleles then we can, in principle, measure the DFE for both strongly and weakly selected mutations. So, the DFE is knowable, but uncovering it might require a large amount of sequencing and effort. Additionally, and possibly more crucially, we must be able to differentiate sites that are subject to selection from those that are not. This might prove to be especially difficult for sites that are subject to weak selection.

1. Crow, J. F. How much do we know about spontaneous human mutation rates? *Environ. Mol. Mutagen.* **21**, 122–129 (1993).
2. Kondrashov, A. S. & Crow, J. F. A molecular approach to estimating the human deleterious mutation rate. *Hum. Mutat.* **2**, 229–234 (1993).
3. Ohta, T. The nearly neutral theory of molecular evolution. *Annu. Rev. Ecol. Syst.* **23**, 263–286 (1992).
4. Loewe, L. Quantifying the genomic decay paradox due to Muller's ratchet in human mitochondrial DNA. *Genet. Res.* **87**, 133–159 (2006).
5. Charlesworth, D., Charlesworth, B. & Morgan, M. T. The pattern of neutral molecular variation under the background selection model. *Genetics* **141**, 1619–1632 (1995).
6. Peck, J. R., Barreau, G. & Heath, S. C. Imperfect genes, Fisherian mutation and the evolution of sex. *Genetics* **145**, 1171–1199 (1997).
7. Caballero, A. & Keightley, P. D. A pleiotropic nonadditive model of variation in quantitative traits. *Genetics* **138**, 883–900 (1994).
8. Zhang, X. S. & Hill, W. G. Genetic variability under mutation selection balance. *Trends Ecol. Evol.* **20**, 468–470 (2005).
9. Eyre-Walker, A., Woolfit, M. & Phelps, T. The distribution of fitness effects of new deleterious amino acid mutations in humans. *Genetics* **173**, 891–900 (2006).
10. Schultz, S. T. & Lynch, M. Mutation and extinction: the role of variable mutational effects, synergistic epistasis, beneficial mutations and degree of outcrossing. *Evolution* **51**, 1363–1371 (1997).
11. Reich, D. E. & Lander, E. S. On the allelic spectrum of human disease. *Trends Genet.* **17**, 502–510 (2001).
12. Johnson, T. & Barton, N. H. Theoretical models of selection and mutation on quantitative traits. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **360**, 1411–1425 (2005).
13. Burch, C. L., Guyader, C., Samarov, D. & Shen, H. Experimental estimate of the abundance and effects of nearly neutral mutations in the RNA virus φ6. *Genetics* **176**, 467–476 (2007).
14. Elena, S. F., Ekunwe, L., Hajela, N., Oden, S. A. & Lenski, R. E. Distribution of fitness effects caused by random insertion mutations in *Escherichia coli*. *Genetica* **102–103**, 349–358 (1998).
15. Sanjuan, R., Moya, A. & Elena, S. F. The distribution of fitness effects caused by single-nucleotide substitutions in an RNA virus. *Proc. Natl Acad. Sci. USA* **101**, 8396–8401 (2004).
    Estimates the DFE for an RNA virus by measuring the fitness consequences of single mutations.
16. Thatcher, J. W., Shaw, J. M. & Dickinson, W. J. Marginal fitness contributions of nonessential genes in yeast. *Proc. Natl Acad. Sci. USA* **95**, 253–257 (1998).
17. Wloch, D. M., Szafraniec, K., Borts, R. H. & Korona, R. Direct estimate of the mutation rate and the distribution of fitness effects in the yeast *Saccharomyces cerevisiae*. *Genetics* **159**, 441–452 (2001).
18. Mukai, T. The genetic structure of natural populations of *Drosophila melanogaster*. I. Spontaneous mutation rate of polygenes controlling viability. *Genetics* **50**, 1–19 (1964).
    The first mutation accumulation experiment.
19. Ohnishi, O. Spontaneous and ethyl methanesulfate-induced mutations controlling viability in *Drosophila melanogaster*. II. Homozygous effect of polygenic mutations. *Genetics* **87**, 529–545 (1977).
20. Bataillon, T. Estimation of spontaneous genome-wide mutation rate parameters: whither beneficial mutations? *Heredity* **84**, 497–501 (2000).
21. Joseph, S. B. & Hall, D. W. Spontaneous mutations in diploid *Saccharomyces cerevisiae*: more beneficial than expected. *Genetics* **168**, 1817–1825 (2004).
22. Shaw, F. H., Geyer, C. J. & Shaw, R. G. A comprehensive model of mutations affecting fitness and inferences for *Arabidopsis thaliana*. *Evolution* **56**, 453–463 (2002).
23. Bateman, A. J. The viability of near-normal irradiated chromosomes. *Int. J. Radiat. Biol.* **1**, 170–180 (1959).

REVIEWS

24. Garcia-Dorado, A. The rate and effects distribution of viable mutation in *Drosophila*: minimum distance estimation. *Evolution* **51**, 1130–1139 (1997).
25. Keightley, P. D. The distribution of mutation effects on viability in *Drosophila melanogaster*. *Genetics* **138**, 1315–1322 (1994).
26. Keightley, P. D. Inference of genome wide mutation rates and distributions of mutations effects for fitness traits: a simulation study. *Genetics* **150**, 1283–1293 (1998).
27. Davies, E. K., Peters, A. D. & Keightley, P. D. High frequency of cryptic deleterious mutations in *Caenorhabditis elegans*. *Science* **285**, 1748–1751 (1999).
   Shows that most mutations are undetectable in a mutation accumulation study and that the DFE of deleterious mutations must be complex and multi-modal.
28. Denver, D. R., Morris, K., Lynch, M. & Thomas, W. K. High mutation rate and predominance of insertions in the *Caenorhabditis elegans* nuclear genome. *Nature* **430**, 679–682 (2004).
29. Eyre-Walker, A., Keightley, P. D., Smith, N. G. C. & Gaffney, D. Quantifying the slightly deleterious model of molecular evolution. *Mol. Biol. Evol.* **19**, 2142–2149 (2002).
30. Loewe, L. & Charlesworth, B. Inferring the distribution of mutational effects on fitness in *Drosophila*. *Biol. Lett.* **2**, 426–430 (2006).
31. Loewe, L., Charlesworth, B., Bartolome, C. & Noel, V. Estimating selection on nonsynonymous mutations. *Genetics* **172**, 1079–1092 (2006).
32. Nielsen, R. & Yang, Z. Estimating the distribution of selection coefficients from phylogenetic data with applications to mitochondrial and viral DNA. *Mol. Biol. Evol.* **20**, 1231–1239 (2003).
33. Piganeau, G. & Eyre-Walker, A. Estimating the distribution of fitness effects from DNA sequence data: implications for the molecular clock. *Proc. Natl Acad. Sci. USA* **100**, 10335–10340 (2003).
34. Sawyer, S., Kulathinal, R. J., Bustamante, C. D. & Hartl, D. L. Bayesian analysis suggests that most amino acid replacements in *Drosophila* are driven by positive selection. *J. Mol. Evol.* **57**, S154–S164 (2003).
35. Bustamante, C. D., Wakeley, J., Sawyer, S. & Hartl, D. L. Directional selection and the site-frequency spectrum. *Genetics* **159**, 1779–1788 (2001).
36. Bubb, K. L. *et al.* Scan of human genome reveals no new loci under ancient balancing selection. *Genetics* **173**, 2165–2177 (2006).
37. Kimura, M. Genetic variability maintained in a finite population due to the mutational production of neutral and nearly neutral isoalleles. *Genet. Res.* **11**, 247–269 (1968).
38. Johnson, K. P. & Seger, J. Elevated rates of nonsynonymous substitution in island birds. *Mol. Biol. Evol.* **18**, 874–881 (2001).
39. Woolfit, M. & Bromham, L. Population size and molecular evolution on islands. *Proc. Biol. Sci.* **272**, 2277–2282 (2005).
40. Silander, O. K., Tenaillon, O. & Chao, L. Understanding the evolutionary fate of finite populations: the dynamics of mutational effects. *PLoS Biol.* **5**, e94 (2007).
   Shows that the DFE is highly dependent on the fitness of the population that is being considered.
41. Lynch, M. & Conery, J. S. The origins of genome complexity. *Science* **302**, 1401–1404 (2003).
42. Subramanian, S. & Kumar, S. Higher intensity of purifying selection on > 90% of the human genes revealed by the intrinsic replacement mutation rates. *Mol. Biol. Evol.* **23**, 2283–2287 (2006).
43. Eyre-Walker, A. Changing effective population size and the McDonald–Kreitman test. *Genetics* **162**, 2017–2024 (2002).
44. Charlesworth, J. & Eyre-Walker, A. The rate of adaptive evolution in enteric bacteria. *Mol. Biol. Evol.* **23**, 1348–1356 (2006).
45. Orgel, L. E. & Crick, F. H. C. Selfish DNA: the ultimate parasite. *Nature* **284**, 604–607 (1980).
46. Cliften, P. *et al.* Finding functional features in *Saccharomyces* genomes by phylogenetic footprinting. *Science* **301**, 71–76 (2003).
47. Shabalina, S. A. & Kondrashov, A. S. Pattern of selective constraint in *C. elegans* and *C. briggsae* genomes. *Genet. Res.* **74**, 23–30 (1999).
48. Webb, C. T., Shabalina, S. A., Ogurtsov, A. Y. & Kondrashov, A. S. Analysis of similarity within 142 pairs of orthologous intergenic regions of *Caenorhabditis elegans* and *Caenorhabditis briggsae*. *Nucleic. Acids Res.* **30**, 1233–1239 (2002).

49. Andolfatto, P. Adaptive evolution of non-coding DNA in *Drosophila*. *Nature* **437**, 1149–1152 (2005).
   Provided the first evidence that adaptive evolution is widespread in *Drosophila* non-coding DNA.
50. Bergman, C. M. & Kreitman, M. Analysis of conserved noncoding DNA in *Drosophila* reveals similar constraints in intergenic and intronic sequences. *Genome Res.* **11**, 1335–1345 (2001).
   Provided the first indication that extensive amounts of *Drosophila* non-coding DNA is subject to selection.
51. Halligan, D. L. & Keightley, P. D. Ubiquitous selective constraints in the *Drosophila* genome revealed by a genome-wide interspecies comparison. *Genome Res.* **16**, 875–884 (2006).
52. Dermitzakis, E. T. *et al.* Numerous potentially functional but non-genic conserved sequences on human chromosome 21. *Nature* **420**, 578–582 (2002).
   Provided the clearest evidence that substantial amounts of mammalian non-coding DNA is subject to selective constraint.
53. Koop, B. F. Human and rodent DNA sequence comparisons: a mosaic model of genomic evolution. *Trends Genet.* **11**, 367–371 (1995).
54. Shabalina, S. A., Ogurtsov, A. Y., Kondrashov, V. A. & Kondrashov, A. S. Selective constraint in intergenic regions of human and mouse genomes. *Trends Genet.* **17**, 373–376 (2001).
55. Mouse Genome Sequencing Consortium. Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520–562 (2002).
56. Dermitzakis, E. T., Reymond, A. & Antonarakis, S. E. Conserved non-genic sequences — an unexpected feature of mammalian genomes. *Nature Rev. Genet.* **6**, 151–157 (2005).
57. Gaffney, D. & Keightley, P. D. Genomic selective constraints in murid noncoding DNA. *PLoS Genet.* **2**, e204 (2006).
58. Eyre-Walker, A. The genomic rate of adaptive evolution. *Trends Ecol. Evol.* **21**, 569–575 (2006).
59. Chimpanzee Sequencing and Analysis Consortium. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* **437**, 69–87 (2005).
60. Zhang, L & Li, W.-H. Human SNPs reveal no evidence of frequent positive selection. *Mol. Biol. Evol.* **22**, 2504–2507 (2005).
61. Bierne, N. & Eyre-Walker, A. Genomic rate of adaptive amino acid substitution in *Drosophila*. *Mol. Biol. Evol.* **21**, 1350–1360 (2004).
62. Smith, N. G. C. & Eyre-Walker, A. Adaptive protein evolution in *Drosophila*. *Nature* **415**, 1022–1024 (2002).
63. Welch, J. J. Estimating the genome-wide rate of adaptive protein evolution in *Drosophila*. *Genetics* **173**, 821–837 (2006).
64. Bachtrog, D. & Andolfatto, P. Selection, recombination and demographic history in *Drosophila miranda*. *Genetics* **174**, 2045–2059 (2006).
65. Williamson, S. H. Adaptation in the *env* gene of HIV-1 and evolutionary theories of disease progression. *Mol. Biol. Evol.* **20**, 1318–1325 (2003).
66. Keightley, P. D., Lercher, M. J. & Eyre-Walker, A. Evidence for widespread degradation of gene control regions in hominid genomes. *PLoS Biol.* **3**, e42 (2005).
67. Gillespie, J. H. Molecular evolution over the mutational landscape. *Evolution* **38**, 1116–1129 (1984).
68. Orr, H. A. The distribution of fitness effects among beneficial mutations. *Genetics* **163**, 1519–1526 (2003).
   An extension of the work of Gillespie showing that the DFE of advantageous mutations should be an exponential distribution under certain simplifying assumptions.
69. Imhof, M. & Schlotterer, C. Fitness effects of advantageous mutations in evolving *Escherichia coli* populations. *Proc. Natl Acad. Sci. USA* **98**, 1113–1117 (2001).
70. Kassen, R. & Bataillon, T. Distribution of fitness effects among beneficial mutations before selection in experimental populations of bacteria. *Nature Genet.* **38**, 484–488 (2006).
71. Rokyta, D. R., Joyce, P., Caudle, S. B. & Wichman, H. A. An empirical test of the mutational landscape model of adaptation using a single-stranded DNA virus. *Nature Genet.* **37**, 441–444 (2005).
72. Cowperthwaite, M. C., Bull, J. J. & Meyers, L. A. Distributions of beneficial fitness effects in RNA. *Genetics* **170**, 1449–1457 (2005).
   Provided evidence that the DFE of advantageous mutations is not an exponential distribution.

73. Bratteler, M., Lexer, C. & Widmer, A. Genetic architecture of traits associated with serpentine adaptation of *Silene vulgaris*. *J. Evol. Biol.* **19**, 1149–1156 (2006).
74. Lexer, C., Rosenthal, D. M., Raymond, O., Donovan, L. A. & Rieseberg, L. H. Genetics of species differences in the wild annual sunflowers, *Helianthus annuus* and *H. petiolaris*. *Genetics* **169**, 2225–2239 (2005).
75. Schemske, D. W. & Bradshaw, H. D. Jr. Pollinator preference and the evolution of floral traits in monkeyflowers (*Mimulus*). *Proc. Natl Acad. Sci. USA* **96**, 11910–11915 (1999).
76. Mukai, T., Chigusa, S. I., Mettler, L. E. & Crow, J. F. Mutation rate and dominance of genes affecting viability in *Drosophila melanogaster*. *Genetics* **2**, 333–355 (1972).
77. Vassilieva, L., Hook, A. M. & Lynch, M. The fitness effects of spontaneous mutations in *Caenorhabditis elegans*. *Evolution* **54**, 1234–1246 (2000).
78. Elena, S. F. & Moya, A. Rate of deleterious mutation and the distribution of its effects on fitness in vesicular stomatitis virus. *J. Evol. Biol.* **12**, 1078–1088 (1999).
79. Zeyl, C. & DeVisser, J. A. Estimates of the rate and distribution of fitness effects of spontaneous mutation in *Saccharomyces cerevisiae*. *Genetics* **157**, 53–61 (2001).
80. Avila, V. *et al.* Increase of the spontaneous mutation rate in a long-term experiment with *Drosophila melanogaster*. *Genetics* **173**, 267–277 (2006).
81. Garcia-Dorado, A., Monedero, J. L. & Lopez-Fanjul, C. The mutation rate and the distribution of mutational effects of viability and fitness in *Drosophila melanogaster*. *Genetica* **102–103**, 255–256 (1998).
82. Keightley, P. D. Nature of deleterious mutation load in *Drosophila*. *Genetics* **144**, 1993–1999 (1996).
83. Schoen, D. J. Deleterious mutation in related species of the plant genus *Amsinckia* with contrasting mating systems. *Evolution* **59**, 2370–2377 (2005).
84. Garcia-Dorado, A. & Caballero, A. On the average coefficient of dominance of deleterious spontaneous mutations. *Genetics* **155**, 1991–2001 (2000).
85. Peters, A. D., Halligan, D. L., Whitlock, M. C. & Keightley, P. D. Dominance and overdominance of mildly deleterious induced mutations for fitness traits in *Caenorhabditis elegans*. *Genetics* **165**, 589–599 (2003).
86. Shaw, R. G. & Chang, S. M. Gene action of new mutations in *Arabidopsis thaliana*. *Genetics* **172**, 1855–1865 (2006).
87. Williamson, S. H. *et al.* Simultaneous inference of selection and population growth from patterns of variation in the human genome. *Proc. Natl Acad. Sci. USA* **102**, 7882–7887 (2005).
88. Li, W.-H., Tanimura, M. & Sharp, P. M. An evaluation of the molecular clock hypothesis using mammalian DNA sequences. *J. Mol. Evol.* **25**, 330–342 (1987).
89. Ohta, T. Synonymous and nonsynonymous substitutions in mammalian genes and the nearly neutral theory. *J. Mol. Evol.* **40**, 56–63 (1995).
90. Bush, E. C. & Lahn, B. T. Selective constraint on noncoding regions of hominid genomes. *PLoS Comput. Biol.* **1**, e73 (2005).
91. Keightley, P. D., Kryukov, G. V., Sunyaev, S., Halligan, D. L. & Gaffney, D. J. Evolutionary constraints in conserved nongenic sequences of mammals. *Genome Res.* **15**, 1373–1378 (2005).
92. Kryukov, G. V., Schmidt, S. & Sunyaev, S. Small fitness effect of mutations in highly conserved non-coding regions. *Hum. Mol. Genet.* **14**, 2221–2229 (2005).
93. Lynch, M. *et al.* Spontaneous deleterious mutation. *Evolution* **53**, 645–663 (1999).

FURTHER INFORMATION
Adam Eyre-Walker's homepage: http://www.lifesci.sussex.ac.uk/CSE/members/aeyrewalker/aeyrewalker.htm
Peter Keightley's homepage:
http://homepages.ed.ac.uk/eang33
Access to this links box is available online.