Review

Population size and the rate of evolution

Robert Lanfear[1,2], Hanna Kokko[1], and Adam Eyre-Walker[3]

[1] Ecology Evolution and Genetics, Research School of Biology, Australian National University, Canberra, ACT, Australia
[2] National Evolutionary Synthesis Center, Durham, NC, USA
[3] School of Life Sciences, University of Sussex, Brighton, UK

Does evolution proceed faster in larger or smaller populations? The relationship between effective population size (N_e) and the rate of evolution has consequences for our ability to understand and interpret genomic variation, and is central to many aspects of evolution and ecology. Many factors affect the relationship between N_e and the rate of evolution, and recent theoretical and empirical studies have shown some surprising and sometimes counterintuitive results. Some mechanisms tend to make the relationship positive, others negative, and they can act simultaneously. The relationship also depends on whether one is interested in the rate of neutral, adaptive, or deleterious evolution. Here, we synthesize theoretical and empirical approaches to understanding the relationship and highlight areas that remain poorly understood.

Does population size limit evolution?

Do small populations evolve faster or slower than large populations? When and why does population size limit adaptation? Answering these questions is important for understanding present-day diversity and the evolutionary past and future of life on Earth [1–3]. A quantitative answer requires one to measure the 'rate of evolution', and link it to population size.

Although there are many ways to measure the rate of evolution, in this review we specifically focus on the relationship between effective population size and the rate of molecular evolution, defined as the rate at which new substitutions accumulate in the genome over time. Substitution rates can be measured from DNA sequence data for almost any lineage [4]. They provide a convenient way to understand the relationship between effective population size and the rate of evolution, and additional data can be used to distinguish between rates of adaptive, deleterious, and neutral evolution. A substitution occurs when a new mutation spreads to fixation in a population; so the substitution rate depends on both the rate at which new mutations occur and the chance that each mutation spreads to fixation.

The purpose of this review is to synthesize theoretical and empirical knowledge of the relationship between effective population size (N_e, Box 1) and the substitution rate, which we term the N_e–rate relationship (N_eRR). A positive N_eRR implies faster evolution in larger populations relative to smaller ones, and a negative N_eRR implies the opposite (Figure 1A,B). Although N_e has long been known to be one of the most important factors determining the substitution rate [5–8], several novel predictions and observations have emerged in recent years, causing some reassessment of earlier theory and highlighting some gaps in our understanding.

Theory: the N_eRR of neutral and nearly neutral mutations ($s \approx 0$)

The neutral substitution rate reflects the mutation rate

Neutral and effectively neutral mutations have fitness effects at or very close to zero ($s \approx 0$, $N_e|s| << 1$, Box 2); therefore, their fate is dominated by genetic drift and largely unaffected by selection. Genetic drift is the stochastic fluctuation in allele frequencies caused by random differences in the fecundity and survival of individuals. As N_e increases, genetic drift becomes weaker because the larger the population, the smaller the proportional impact of each random event that concerns just one individual. Theory predicts that the increased production of neutral and effectively neutral mutations in larger populations is exactly balanced by the decreased probability that each mutation will fix through genetic drift ([9]; Box 3 provides a mathematical description, and describes how census and effective population size combine to determine rates of evolution). As a result, the neutral substitution rate equals the mutation rate, no matter what the value of N_e. Thus, if the mutation rate does not change, then the N_eRR for neutral mutations will be flat (as shown in Figure 1A,B, lines for $s = 0$).

The observation that the neutral substitution rate equals the mutation rate is surprisingly robust, despite the simplicity of the initial theory from which it was derived (Box 3). For example, factors such as interference among mutations [which can have important effects on the N_eRR for mutations under selection (i.e., with $s \neq 0$), see below] do not affect the rate of neutral substitution [10,11], and so will not change the N_eRR for neutral and effectively neutral mutations. A recent study suggested that the

Corresponding author: Lanfear, R. (rob.lanfear@anu.edu.au).
Keywords: population size; molecular evolution; mutation rate; substitution rate; genetic drift; natural selection.

0169-5347/$ – see front matter
 http://dx.doi.org/10.1016/j.tree.2013.09.009

Box 1. Estimating effective population size

The effective population size (N_e) provides a measure of the power of genetic drift, such that increasing N_e is associated with decreasing rates of genetic drift [81,83]. In an 'ideal' population – defined as one in which population size is constant, and in which offsprings' genes are randomly sampled from the parental generation – N_e will be equal to the census population size, N_C [83]. However, N_e is usually a lot lower than N_C because natural populations have many characteristics that reduce N_e [84]. Different parts of the genome can also differ in N_e (Box 4).

Broadly speaking, there are four methods that can be used to estimate N_e: (i) the reduction in N_e relative to N_C can be estimated from the species life history; (ii) N_e can be estimated from the variance in allele frequencies between generations; (iii) N_e can be estimated from genetic polymorphism data; and (iv) N_e can be estimated from its correlates, such as body size [67–71,81,84,85]. The last of these methods has been used in the vast majority of empirical studies because estimating N_e using the other methods requires substantially more information. However, no systematic analysis has been conducted into whether factors such as body size correlate with N_e, in large part because there are so few estimates of N_e available. Surprisingly estimates of N_e from the variation in allele frequencies and neutral DNA sequence diversity disagree, although they have rarely been applied to the same taxa [86]. The reasons for this discrepancy are unknown.

Because mutations of different selective effects spend different amounts of time in a population, it is important for studies of the N_eRR to pay close attention to the timescale over which N_e should be calculated. If N_e is constant over time and there is no linkage, then all mutations will experience the same N_e and any accurate estimate of N_e will be appropriate for any type of mutation. However, if N_e changes over time or there is linkage, then different mutations will experience different N_e because strongly selected mutations are either fixed or lost much more rapidly than neutral mutations [33]. This can lead to a mismatch between the N_e that can be estimated and the N_e of interest. This kind of mismatch might, for example, explain why two *Drosophila* species with different estimates of N_e appear to have similar rates of adaptive evolution [22].

neutral substitution rate is not equal to the mutation rate in situations where there are overlapping generations and fluctuating population size [12]. However, it seems likely that the reported effect is in fact due to changes in the mutation rate induced by changes in the generation time when populations expand or contract, rather than a true difference between the neutral substitution rate and the mutation rate.

Accounting for variation in mutation rates

Of course, we rarely expect the mutation rate per individual per year to be equal in different populations or species, which means that empirical estimates of the N_eRR for neutral mutations may not be flat. For example, mammals with larger N_e also tend to have higher mutation rates per year because they tend to have shorter generation times [13], creating a positive N_eRR for neutral mutations.

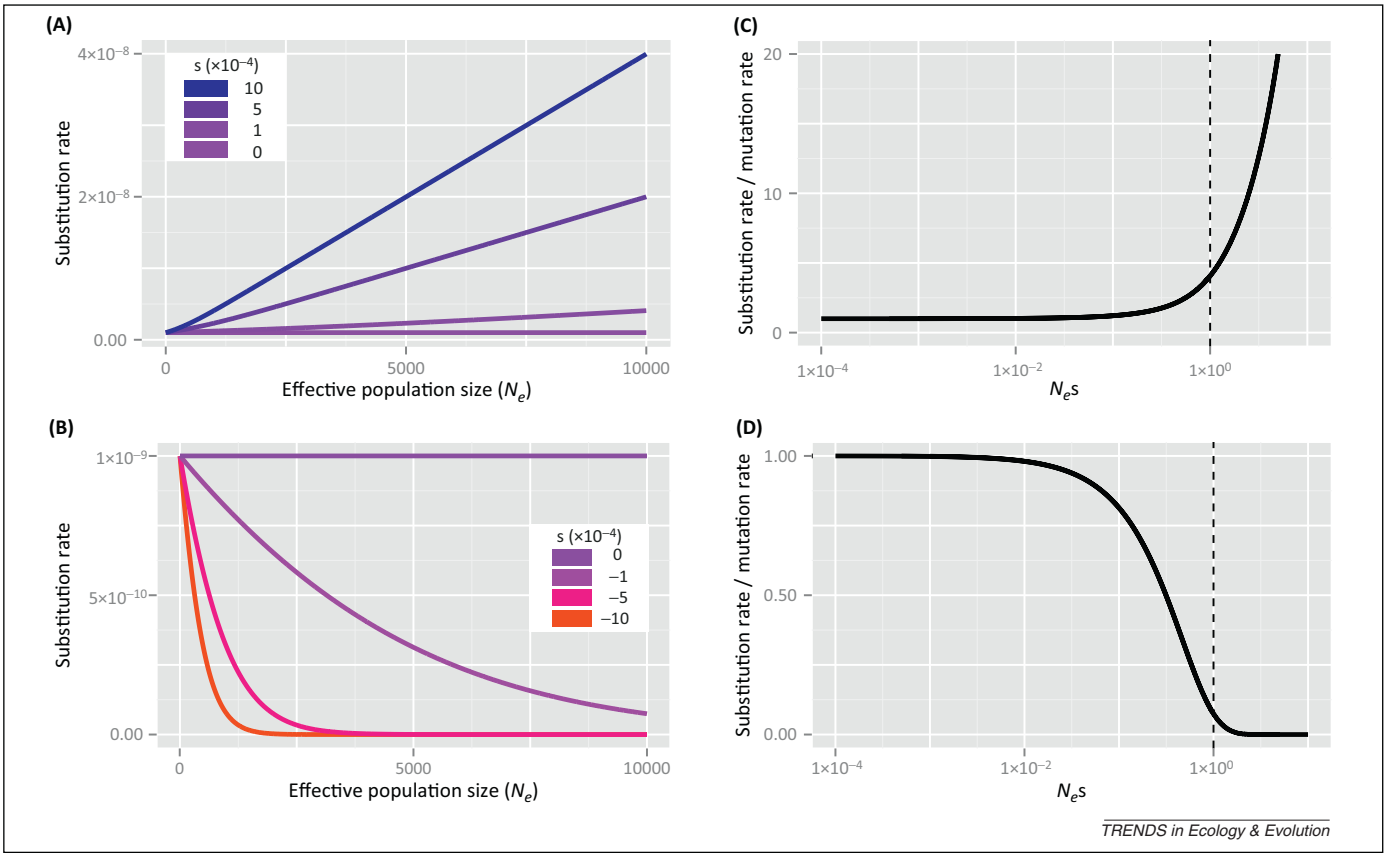

Figure 1. The relationship between substitution rate (in substitutions per site per year) and effective population size (N_e) under genetic drift and natural selection (the N_eRR) [8]. These relationships were calculated assuming a mutation rate of 1×10^{-9} mutations per site per year, approximately that found in humans. **(A,B)** show the substitution rate of mutations for a range of positive (A) and negative (B) selection coefficients (denoted 's'). **(C,D)** show the same data, but in this case the y-axis shows the substitution rate relative to the mutation rate, and the x-axis shows the product of N_e and the selection coefficient for positive (C) and negative (D) mutations respectively. A dashed line highlights where N_es = 1, below which mutations are often considered 'effectively neutral'. Note that genetic drift predicts a flat N_eRR for neutral mutations, where s = 0.00 in (A,B). In (C,D), this is reflected by the substitution rate equaling the mutation rate, giving a value of 1 on the y-axis, when N_es = 0.

Box 2. Five categories of mutation, and the distribution of fitness effects (DFE)

Typically, we categorize mutations based on the relative importance of genetic drift and natural selection in determining their fates.

Neutral and effectively neutral mutations
These mutations have fitness effects that are either zero or much smaller in magnitude than $1/N_e$; therefore, their fate is dominated by genetic drift rather than by selection. These might be the most common type of mutations in organisms with large genomes, but might be rare in species with small genomes.

Slightly deleterious and slightly advantageous mutations
These mutations, collectively known as nearly neutral mutations, have small effects on fitness, defined as an effect that is close to the reciprocal of the effective number of alleles in the population ($1/N_e$ for haploid populations, and $1/2N_e$ for diploid populations). Their fate depends on a combination of natural selection and genetic drift; as N_e decreases, a mutation that was previously slightly deleterious can become effectively neutral, increasing the proportion of mutations that behave as neutral; and as N_e increases, the same mutation can become subject to strong selection. Biologically, this reflects the fact that genetic drift becomes less important in determining the fate of mutations as N_e increases. There is evidence that both slightly deleterious and slightly advantageous mutations exist and contribute substantially to evolutionary dynamics [44,48,87].

Deleterious and advantageous mutations
These mutations have fitness effects that are much larger in magnitude than $1/N_e$. Their fates are determined primarily by natural selection, not genetic drift. Both of these types of mutation have been demonstrated to occur in natural and experimental populations [48].

Although mutations are often divided into these five categories for convenience, in reality there is likely to be a continuum of selective effects. Various methods have been devised to estimate this DFE, although this is challenging [48–51]. There is also growing interest in theoretical approaches that do not assume the DFE *a priori*, but allow it to change as populations evolve [39,52–54,88].

Another mechanism that might create an association between the mutation rate and N_e is selection to reduce the mutation rate. For example, one recent theory suggests that smaller populations will have higher mutation rates per generation than larger populations, because natural selection should be less effective at reducing the mutation rate in smaller populations [14,15]. All else being equal, this effect would produce a negative N_eRR for neutral mutations because the mutation rate per year will decline with increasing N_e. However, the importance of this theory for understanding the N_eRR is questionable: although there is some evidence that N_e is negatively associated with mutation rates per generation [16–18], the overall deleterious mutation rate is much higher than this theory predicts [19], and the modest effects predicted by this theory could be swamped by the other causes of variation in mutation rates per year such as differences in the generation time itself [16].

Theoretical studies of the N_eRR often account for mutation rate variation by reporting substitution rates relative to the underlying mutation rate. When the substitution rate is corrected in this way, we always expect a flat N_eRR for neutral mutations. However, this is difficult to do in empirical studies because it is tricky to obtain independent estimates of mutation and substitution rates; indeed, the neutral substitution rate is often the best estimate we have of the underlying mutation rate. Because of this, it is important to bear in mind that variation in both mutation and substitution rates may affect empirical estimates of the N_eRR.

Theory: the N_eRR for mutations under selection ($s \neq 0$)

The shape of the N_eRR depends on the fitness effects of mutations

For mutations on which natural selection can act (i.e., those with $s \neq 0$, Box 2), the N_eRR depends on the fitness effects of mutations (s, Figure 1). As N_e increases, natural selection becomes more effective at fixing advantageous mutations and removing deleterious mutations, but larger populations also produce more of both types of mutation. Theory suggests that as N_e increases the power of natural selection increases faster than the production of new mutations (see [5] for a recent review). This results in lower deleterious substitution rates as N_e increases (a negative N_eRR, Figure 1B,D), and higher advantageous substitution rates as N_e increases (a positive N_eRR, Figure 1A,C). However, these predictions can sometimes be altered when the simplifying assumptions of the underlying theory are not met.

The N_eRR for advantageous mutations when evolution is not mutation-rate limited

The prediction of a positive N_eRR for advantageous mutations (Figure 1A,C) depends on the advantageous substitution rate being limited by the mutation rate. This limitation occurs when the supply of advantageous mutations is low and there are periods in which the population contains no advantageous mutations. Species with large N_e or high mutation rates may not suffer from this limit, because they tend to produce more mutations [17,20]. If the rate of advantageous substitution is not limited by the supply of mutations, then the N_eRR for advantageous mutations will depend on how the level of genetic variation scales with N_e. However, surprisingly, the level of molecular variation seems to be largely independent of N_e [18,21], suggesting that the N_eRR for advantageous mutations could be relatively flat in many cases [20,22], although there are some notable exceptions [23].

The N_eRR for selected mutations in the presence of linkage

Simple models of molecular evolution assume that there is no linkage among loci, such that selection at one locus has no effect on the fixation of mutations at other loci. Linkage between loci can have knock-on effects on the N_eRR because selection at one locus can affect fixation at other loci through a set of processes collectively known as interference [24]. Interference is more prevalent when linkage is strong [25–31], but even modest amounts of recombination can be sufficient to alleviate the effects of interference on the N_eRR [28,30,32,33]. Little is known about the scaling of interference and recombination with N_e; so it is still too early to make general predictions about the effects of interference on the N_eRR. Nevertheless, theoretical studies have revealed certain situations in which interference can affect the N_eRR, in particular through clonal interference and selective sweeps.

Box 3. The rate of evolution under simple models of molecular evolution

For neutral and effectively neutral mutations, the expected rate of evolution depends simply on the mutation rate [5,7,9]. This elegant result arises because at any one time, only one chromosome at a site is destined to spread to fixation. The chance that a particular chromosome contains this mutation is $1/K$, where K is the number of chromosomes at a site. If N_c is the census population size then $K = N_c$ (haploid population) or $K = 2N_c$ (diploid population). Assuming mutations at a site are rare and arise at a rate of u, the rate at which mutations enter the population is Ku. Hence, the rate of neutral evolution at a site is $Ku \times 1/K = u$. This expectation holds irrespective of selection at other linked loci, changes in population size, or almost any other conceivable complication.

The rate of evolution at sites subject to selection, on the other hand, depends upon the mutation rate and the probability P that a mutation spreads and becomes fixed (Equation I):

$$R = 2N_c uP \quad [I]$$

The probability of fixation of a new mutation [8], present initially as a single copy, depends upon the effective and census population sizes and the strength of selection s. Assuming heterozygous individuals have an advantage $+s$ and homozygous mutant individuals an advantage $+2s$ (Equation II):

$$P = \frac{1 - e^{-2N_e s/N_c}}{1 - e^{-4N_e s}} \quad [II]$$

Hence, if s is small, as we expect for most mutations (Equation III):

$$P = \frac{2N_e s/N_c}{1 - e^{-4N_e s}} \quad [III]$$

For advantageous mutations in which $N_e s > 1$, this can be further simplified to Equation IV:

$$P = 2\frac{N_e}{N_c}s \quad [IV]$$

If we assume that $N_e = N_c$, we obtain the classic result of Haldane [89], $P = 2s$. The fixation probability does not depend on N_e in this case because an advantageous mutation is only vulnerable to loss by genetic drift when it is rare, and at low frequency the strength of genetic drift is independent of population size (this is because the variance in the number of copies of an allele, from generation to generation, is equal to the number of copies in the population, when it is rare, because it is Poisson distributed). More generally, the probability of fixation of an advantageous mutation depends upon N_e/N_c because only this fraction of the copies of the advantageous allele contribute to future generations.

For deleterious or weakly selected advantageous mutations, Equation III applies. The fixation probability depends upon both $N_e s/N_c$ and $N_e s$, reflecting the fates of mutations when they are rare and common, respectively; the effect of drift depends upon N_e. Using Equations I and III, the rate of evolution reduces to Equation V:

$$R = \frac{4N_e su}{1 - e^{-4N_e s}} \quad [V]$$

Hence, the rate relative to the rate of neutral evolution only depends on the compound parameter $4N_e s$ (Equation VI):

$$R' = \frac{4N_e s}{1 - e^{-4N_e s}} \quad [VI]$$

Clonal interference occurs when two or more adaptive mutations that originated in different individuals compete for fixation in a population [34]. Clonal interference can limit rates of adaptation in large populations and lead to a plateau in the N_eRR for adaptive substitutions, although this plateau can be somewhat overcome if additional mutations appear on the same genetic background as the original adaptive mutation [35–37]. Selective sweeps occur when adaptive mutations lead to the fixation of mutations at linked loci. Selective sweeps can increase the rate of slightly deleterious substitution and decrease the rate of slightly advantageous substitution [38]; so the effect of selective sweeps on the N_eRR depends on how their frequency scales with N_e. Some population genetic models suggest that the rate of selective sweeps increases with N_e until it reaches a plateau caused by clonal interference [38]. Others suggest that the rate of adaptation, and hence selective sweeps, is essentially independent of N_e, and is instead determined by the rate of environmental change and number of traits upon which selection acts [39]. If the former is correct, then at small values of N_e selective sweeps might produce a positive N_eRR for slightly deleterious mutations and a negative N_eRR for slightly advantageous mutations [38]; the opposite patterns to those predicted by genetic drift (Figure 1). However, both cases suggest that the N_eRR should be relatively flat at large values of N_e, a prediction that is simple to examine empirically (see below).

The N_eRR for selected mutations when there is variation in selection and N_e

The patterns we have described so far assume that the fitness effect of a given mutation remains constant over space and time. However, fitness effects may vary across a specie's range or over time as the environment or the genetic background changes, with knock-on effects on substitution rates [40–43]. Unfortunately, little is known about how spatiotemporal variation in fitness effects is linked to N_e, which limits understanding of how it affects the N_eRR. One intriguing but untested suggestion is that larger populations will occupy larger ranges than smaller populations, and therefore experience more spatial variation in fitness effects, which could lead to a plateau in the N_eRR [41]. Developing this aspect of population-genetic theory is an important avenue of future research into the N_eRR [27].

Variation in N_e can also affect the N_eRR. For example populations that have recently expanded are expected to fix a bout of advantageous mutations [44], and population structure can have complex effects on rates of adaptive substitution [45,46].

Theory: the N_eRR for all mutations

The distribution of fitness effects

The N_eRR of all mutations (i.e., combining substitutions from mutations of all fitness effects) is of particular interest because it describes how the overall substitution rate is related to N_e. Because mutations of different fitness effects have different N_eRRs (Figure 1), the distribution of fitness effects of new mutations (DFE) is important for predicting the N_eRR for all mutations. In general, we expect the N_eRR for all mutations to have a 'U' shape: resembling the N_eRR for deleterious mutations when N_e is small, and the N_eRR for advantageous mutations when N_e is large. This occurs because as N_e increases the proportion of deleterious substitutions declines towards zero (Figure 1) and the proportion of advantageous substitutions increases (Figure 1). The exact shape of the N_eRR for all mutations, and the

point at which the inflection occurs, depends on the shape of the DFE [47], which is typically not known for molecular datasets [48]. Further refinements to our prediction for the total N_eRR rely on methods to estimate the DFE from empirical data, which is fortunately a very active area of research [49–51] (Box 2).

Fitness landscapes

Many theoretical approaches to the N_eRR consider either one category of mutation (Box 2), or mutations drawn from a fixed DFE. A complementary approach is to study evolution on fitness landscapes. In this approach, the DFE is not assumed *a priori* but derived from the population's position on the fitness landscape (Figure 2). This requires a different type of distribution to be specified: the distance that genotypes move on the fitness landscape when they mutate. The fitness-landscape approach naturally accounts for the fact that advantageous mutations are less likely to occur when a population is already near a fitness peak (Figure 2). The downside is that the distribution of mutation sizes is probably even harder to estimate for natural populations than the DFE. A vast range of fitness landscapes can be defined [52,53]. The most commonly studied is Fisher's Geometric Model (e.g., [39,54]), which describes a smooth fitness landscape with one dimension for each trait of an organism, and a single adaptive peak.

Fitness landscapes provide some important insights into the N_eRR for all mutations. First, when populations are at equilibrium (i.e., at a fitness peak on the landscape) we expect smaller populations to fix more slightly deleterious mutations than larger ones, and so provide more opportunities for slightly advantageous compensatory mutations [55], a phenomenon known as 'selection without adaptation' [56]. This leads to smaller populations having faster rates of both deleterious and advantageous substitution than larger populations (i.e., a positive N_eRR for

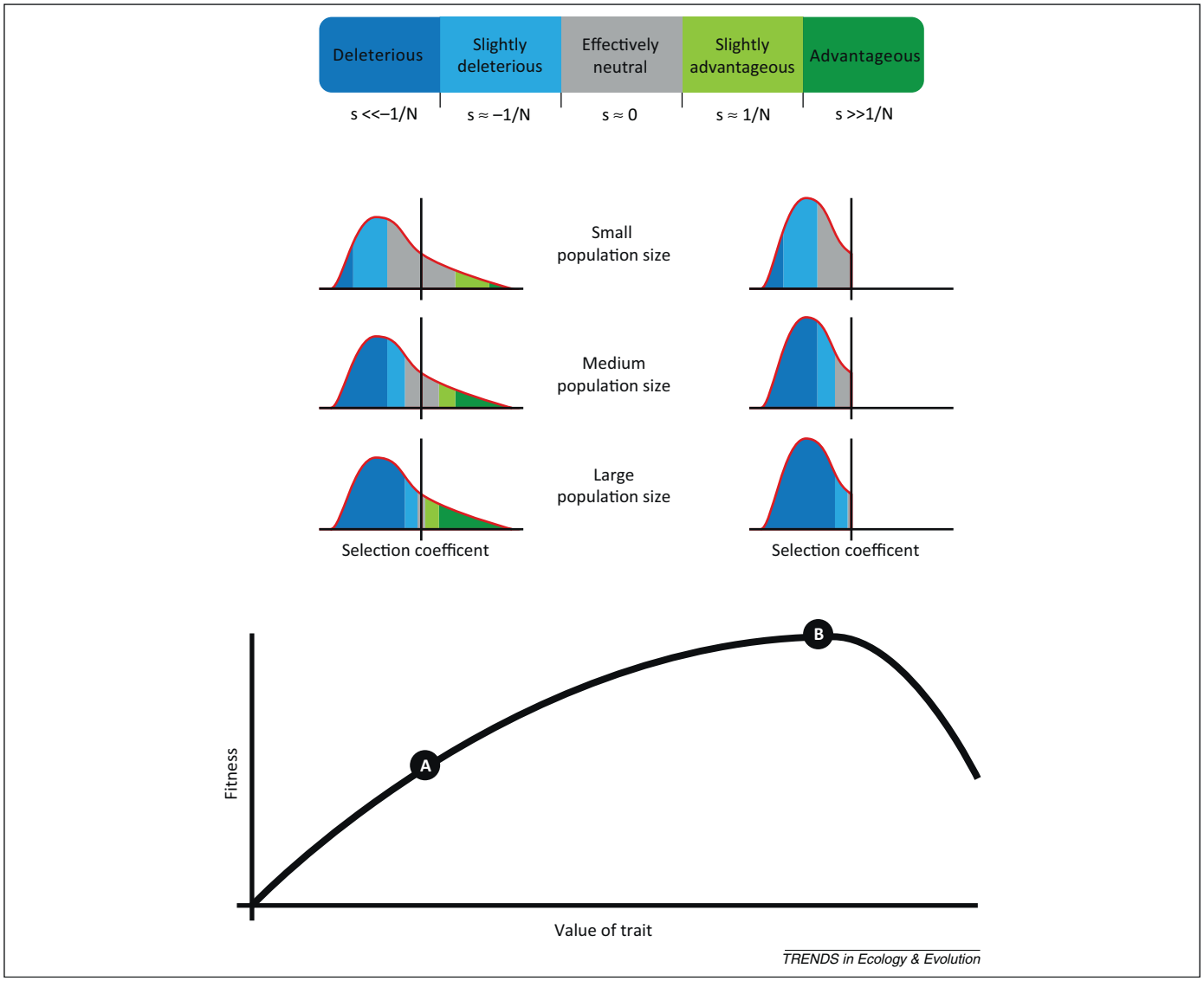

Figure 2. Fitness landscapes and the distribution of fitness effects. Hypothetical distributions of fitness effects for populations of different sizes at different positions on a simple fitness landscape. A population far from the optimum **(A)** has a certain proportion of mutations that confer increases in fitness. However, a population at the hypothetical optimum of the landscape **(B)** cannot increase it fitness, so all mutations are deleterious. In both cases, the proportion of mutations that fall into different categories (Box 2, main text) changes depending on the effective population size. Note that, for simplicity, we have drawn a fitness landscape that varies along a single dimension, but the distributions we have drawn are more similar to those that would come from higher-dimensional fitness landscapes. Furthermore, it is unlikely that any natural population sits at the precise optimum of any fitness landscape. The selection coefficients are shown on natural scales, not log-transformed scales.

both, the latter being the opposite of the predictions under simple models of drift), although this effect depends on the shape of the fitness landscape [57]. Second, if the optimal phenotype is moving because the environment is changing, this can have complex effects on the N_eRR, because species with larger N_e are expected to approach the optimum more rapidly and hence run out of advantageous mutations. Simulations suggest that the N_eRR for the overall substitution rate can be positive, negative, or constant in these cases, depending on the distribution of mutation sizes, the number of traits under selection, and the rate at which the environment changes [39,42,54].

Summary

In summary, we expect the rate of neutral and effectively neutral substitution to depend solely on the underlying mutation rate per individual per year. If the mutation rate per individual per year is not related to N_e then we expect a flat N_eRR for neutral and effectively neutral mutations (i.e., those with $s \approx 0$, Box 2). On the other hand, any process that leads to an association between N_e and mutation rates will cause a similar association between N_e and neutral and effectively neutral substitution rates. These processes could include effects such as the evolution of mutation rates, and the co-variation of N_e with life-history traits such as generation time.

For selected mutations (i.e., those with $s \neq 0$), the N_eRR depends on a variety of factors. Putting aside variation in the mutation rate, we largely expect the total rate of evolution to be negatively correlated with N_e if slightly deleterious mutations dominate evolution, and to be positively correlated with N_e if advantageous mutations dominate evolution. However, these simple predictions can change in complex ways if mutation rates are linked to N_e, or if linkage is tight and adaptive evolution is frequent (see above).

Empirical studies of the N_eRR

Most empirical studies of the N_eRR use a comparative approach in which substitution rates are estimated for two or more species [4] or regions of the genome (Box 4), and compared to estimates of N_e. The biggest challenge for these studies is obtaining estimates of N_e (Box 1); most comparative studies rely on crude proxies of N_e, which can limit the inferences that can be made about the N_eRR. It is also possible to use experimental approaches to study the N_eRR [58], although relatively few such studies have been performed to date [59]. Nevertheless, recent reductions in DNA sequencing costs hold great promise for experimental approaches, because it is now affordable to repeatedly sequence many genomes as populations evolve, and thus to observe mutation and substitution in action as well as estimate the fitness effects of the observed changes [58].

In empirical studies of the N_eRR it is helpful to disentangle the effects of mutation and selection, because variation in mutation rates can swamp other effects [60]. In comparative studies, this is typically done by assuming that synonymous mutations are neutral and nonsynonymous mutations are subject to selection [61–63]. If this is true, then the rate of synonymous substitution provides an estimate of the mutation rate, and the ratio (ω) of the nonsynonymous (dN) to the synonymous (dS) substitution rates gives an estimate of the rate of substitution at selected sites relative to the rate of mutation [63]. The interpretation and comparison of the results of comparative studies would be improved if researchers routinely reported dN, dS, and ω for each of the taxa in their study, which is currently rather rare.

Box 4. N_e and rates of evolution within a genome

Effective population size (N_e) is known to vary across a genome as well as between species. This was initially demonstrated by showing that levels of neutral nucleotide diversity were correlated to rates of recombination in *Drosophila* [28,90], humans [91], and some plant species [92,93]. This could have been due to either variation in N_e or the mutation rate, but levels of neutral divergence between species, a measure of the mutation rate, are not correlated to rates of recombination in *Drosophila* [28,90] or the plant species [92]. There is a correlation between divergence and the rate of recombination in humans, but this is not sufficient to explain the correlation between diversity and recombination [91]. Further analyses have shown that variation in N_e is present in the genomes of all species that have been examined [94]. This variation is thought to be consequence of genetic hitchhiking, background selection, and the fact that there are fewer sex chromosomes than autosomes in a population. However, the variation is modest, with most genomic regions having N_e values well within an order of magnitude of each other [94].

Does intragenomic variation in N_e affect rates of evolution? There is variation in the mutation rate across the genome both between the sex chromosomes and the autosomes [95] and across the autosomes [96]; however, there is no evidence that this is correlated with variation in N_e. The sex chromosomes have different mutation rates to the autosomes in some species, but this is thought to be because males can have higher mutation rates than females and sex chromosomes spend longer in one sex than the other [95]. However, intragenomic variation in N_e leads to variation in the rate of evolution at selected sites; it has been shown that ω is positively correlated with rates of recombination, and hence N_e, along the autosomes of *Drosophila* for genes that appear to be undergoing high rates of adaptive evolution [28,30,97], and negatively correlated for genes undergoing low rates of adaptive evolution [30]; ω is also significantly higher on the nonrecombining Y in humans [98], the W chromosome in birds [29], and the fourth chromosome in *Drosophila* [99]. By contrast, there appears to be no correlation between ω and rates of recombination in primates [100].

Empirical studies of the N_eRR for the overall substitution rate

Empirical studies have revealed negative N_eRRs for ω in bacteria, plants, animals, and fungi using various proxies of N_e [64–72]. These patterns are consistent with the view that most mutations are deleterious, adaptive substitutions are rare, and that smaller populations experience elevated rates of slightly deleterious substitution than larger populations as a result of genetic drift.

Some empirical studies of the N_eRR hint at more complex patterns. For example, a recent study showed that in island–mainland comparisons, island-to-mainland colonisations (a rarer occurrence than the converse) showed consistently faster substitution rates on the mainland, in line with the hypothesis that recent population size expansions cause bursts of adaptive evolution [44]. Another study showed that island birds have consistently lower total substitution rates than their mainland relatives [73]. This counterintuitive pattern lacks a definitive explanation, and is difficult to interpret with respect to the N_eRR

because of the confounding effect of variation in mutation rates; indeed, after accounting for this variation no consistent pattern was seen [73].

In addition to variation in mutation rates, comparisons between species suffer from a range of confounding factors that can be challenging to control. However, N_e is also expected to vary across the genome of a single species (Box 4). This variation can be used to examine the N_eRR without having to worry about many of the confounding factors associated with between-species comparisons. Comparisons within genomes have shown that N_e affects the efficiency of selection, and that regions with high N_e typically have lower rates of evolution than those with low N_e (Box 4).

Empirical studies of the N_eRR for adaptive substitutions

Positive N_eRRs for adaptive substitutions have been revealed in a range of eukaryotes [74–78], although in some of the studies the positive relationship might have been an artifact of using a common outgroup. By contrast, a recent study found no obvious association between DNA sequence diversity and the rate of adaptive substitution across a range of animals, and suggested that the N_eRR for adaptive substitutions might be flat [72]. Why these studies reach different conclusions is not well understood, and could result from a range of biological or methodological factors. This highlights the complexity of measuring and understanding the N_eRR for adaptive substitutions, and we suggest that future efforts should focus on better accounting for factors that can affect the N_eRR, such as variation in linkage and mutation rates among taxa (see above). Experimental studies have demonstrated that interference can dominate the N_eRR when N_e is large and/or linkage is strong. Experiments on viruses and prokaryotes have shown that the N_eRR is dominated by selective sweeps if the supply of new mutations is small. However, if the supply of new mutations is large, then clonal interference becomes the dominating factor [25,52]. Whether the genomic rate of adaptive evolution plateaus at large N_e in natural populations remains an open question, for which the study of 'hyperdiverse' species (those with large N_e and/or high mutation rates) will be particularly informative [17,23].

Concluding remarks and future perspectives

Existing studies of the N_eRR tend to fit our expectations that most mutations are deleterious, and that drift and selection are the most important forces determining the N_eRR. Some more recent studies hint at more complex effects, but in order to make progress we need to focus on obtaining more accurate estimates of N_e. We currently rely heavily on proxies of N_e that are highly imperfect, and tell us little or nothing about population structure or historical variation in N_e [45,46]. Luckily, the continued reductions in DNA sequencing costs promise to vastly increase the amount of available empirical data for estimating both N_e and substitution rates. It is now affordable to sequence hundreds or thousands of loci from many individuals of many species [72,79,80]. Sequencing at this scale not only enables better estimates of substitution rates, it can also be used to estimate N_e [81], the distribution of fitness effects [48–50], mutation rates [82], and variation in each of these parameters across the genome (Box 4). By integrating data and analyses across micro- and macro-evolutionary scales, we will be able to unravel more of the complexity of the N_eRR, and reveal more of the processes that drive and limit evolution.

References

1 Jamieson, I.G. and Allendorf, F.W. (2012) How does the 50/500 rule apply to MVPs? *Trends Ecol. Evol.* 27, 578–584
2 Gonzalez, A. *et al.* (2013) Evolutionary rescue: an emerging focus at the intersection between ecology and evolution. *Philos. Trans. R. Soc. Lond. B: Biol. Sci.* 368, 20120404
3 Frankham, R. *et al.* (2013) 50/500 rule and minimum viable populations: response to Jamieson and Allendorf. *Trends Ecol. Evol.* 28, 187–188
4 Lanfear, R. *et al.* (2010) Watching the clock: studying variation in rates of molecular evolution between species. *Trends Ecol. Evol.* 25, 495–503
5 Akashi, H. *et al.* (2012) Weak selection and protein evolution. *Genetics* 192, 15–31
6 Lynch, M. (2007) *The Origins of Genome Architecture*. (1st edn), Sinauer Associates
7 Ohta, T. (1992) The nearly neutral theory of molecular evolution. *Annu. Rev. Ecol. Syst.* 23, 263–286
8 Kimura, M. (1957) Some problems of stochastic processes in genetics. *Ann. Math. Stat.* 4, 882–901
9 Kimura, M. (1983) *The Neutral Theory of Molecular Evolution*, Cambridge University Press
10 Birky, C.W. and Walsh, J.B. (1988) Effects of linkage on rates of molecular evolution. *Proc. Natl. Acad. Sci. U.S.A.* 85, 6414–6418
11 Gillespie, J.H. (1999) The role of population size in molecular evolution. *Theor. Popul. Biol.* 55, 145–156
12 Balloux, F. and Lehmann, L. (2012) Substitution rates at neutral genes depend on population size under fluctuating demography and overlapping generations. *Evolution* 66, 605–611
13 Bromham, L. (2011) The genome as a life-history character: why rate of molecular evolution varies between mammal species. *Philos. Trans. R. Soc. Lond. B: Biol. Sci.* 366, 2503–2513
14 Lynch, M. (2011) The lower bound to the evolution of mutation rates. *Genome Biol. Evol.* 3, 1107–1118
15 Lynch, M. (2010) Evolution of the mutation rate. *Trends Genet.* 26, 345–352
16 Piganeau, G. and Eyre-Walker, A. (2009) Evidence for variation in the effective population size of animal mitochondrial DNA. *PLoS ONE* 4, e4396
17 Cutter, A.D. *et al.* (2013) Molecular hyperdiversity and evolution in very large populations. *Mol. Ecol.* 22, 2074–2095
18 Bazin, E. *et al.* (2006) Population size does not influence mitochondrial genetic diversity in animals. *Science* 312, 570–572
19 Martincorena, I. and Luscombe, N.M. (2012) Non-random mutation: the evolution of targeted hypermutation and hypomutation. *Bioessays* 35, 123–130
20 Karasov, T.T. *et al.* (2010) Evidence that adaptation in *Drosophila* is not limited by mutation at single sites. *PLoS Genet.* 6, e1000924
21 Nabholz, B. *et al.* (2008) Determination of mitochondrial genetic diversity in mammals. *Genetics* 178, 351–361
22 Bachtrog, D. (2008) Similar rates of protein adaptation in *Drosophila miranda* and *D. melanogaster*, two species with different current effective population sizes. *BMC Evol. Biol.* 8, 334
23 Dey, A. *et al.* (2013) Molecular hyperdiversity defines populations of the nematode *Caenorhabditis brenneri*. *Proc. Natl. Acad. Sci. U.S.A.* http://dx.doi.org/10.1073/pnas.1303057110
24 Charlesworth, B. *et al.* (2010) Genetic recombination and molecular evolution. *Cold Spring Harb. Symp. Quant. Biol.* 74, 177–186
25 de Visser, J.A.G.M. *et al.* (1999) Diminishing returns from mutation supply rate in asexual populations. *Science* 283, 404–406
26 Strelkowa, N. and Lassig, M. (2012) Clonal interference in the evolution of influenza. *Genetics* 192, 671–682
27 Barton, N.H. (2010) Genetic linkage and natural selection. *Philos. Trans. R. Soc. Lond. B: Biol. Sci.* 365, 2559–2569
28 Presgraves, D.C. (2005) Recombination enhances protein adaptation in *Drosophila melanogaster*. *Curr. Biol.* 15, 1651–1656

29 Berlin, S. and Ellegren, H. (2006) Fast accumulation of nonsynonymous mutations on the female-specific W chromosome in birds. *J. Mol. Evol.* 62, 66–72
30 Larracuente, A.M. *et al.* (2008) Evolution of protein-coding genes in *Drosophila*. *Trends Genet.* 24, 114–123
31 Paland, S. and Lynch, M. (2006) Transitions to asexuality result in excess amino acid substitutions. *Science* 311, 990–992
32 Colegrave, N. (2002) Sex releases the speed limit on evolution. *Nature* 420, 664–666
33 Weissman, D.B. and Barton, N.H. (2012) Limits to the rate of adaptive substitution in sexual populations. *PLoS Genet.* 8, e1002740
34 Orr, H.A. (2010) The population genetics of beneficial mutations. *Philos. Trans. R. Soc. Lond. B: Biol. Sci.* 365, 1195–1201
35 Desai, M.M. and Fisher, D.S. (2007) Beneficial mutation selection balance and the effect of linkage on positive selection. *Genetics* 176, 1759–1798
36 Lang, G.I. *et al.* (2011) Genetic variation and the fate of beneficial mutations in asexual populations. *Genetics* 188, 647–661
37 Good, B.H. *et al.* (2012) Distribution of fixed beneficial mutations and the rate of adaptation in asexual populations. *Proc. Natl. Acad. Sci. U.S.A.* 109, 4950–4955
38 Gillespie, J.H. (2001) Is the population size of a species relevant to its evolution? *Evolution* 55, 2161–2169
39 Lourenço, J.M. *et al.* (2013) The rate of molecular adaptation in a changing environment. *Mol. Biol. Evol.* http://dx.doi.org/10.1093/molbev/mst026
40 Bell, G. (2010) Fluctuating selection: the perpetual renewal of adaptation in variable environments. *Philos. Trans. R. Soc. Lond. B: Biol. Sci.* 365, 87–97
41 Ohta, T. (1972) Population size and rate of evolution. *J. Mol. Evol.* 1, 305–314
42 Mustonen, V. and Lässig, M. (2009) From fitness landscapes to seascapes: non-equilibrium dynamics of selection and adaptation. *Trends Genet.* 25, 111–119
43 Waxman, D. (2011) A unified treatment of the probability of fixation when population size and the strength of selection change over time. *Genetics* 188, 907–913
44 Charlesworth, J. and Eyre-Walker, A. (2007) The other side of the nearly neutral theory: evidence of slightly advantageous back-mutations. *Proc. Natl. Acad. Sci. U.S.A.* 104, 16992–16997
45 Frean, M. *et al.* (2013) The effect of population structure on the rate of evolution. *Proc. R. Soc. B* 280, 20130211
46 Lieberman, E. *et al.* (2005) Evolutionary dynamics on graphs. *Nature* 433, 312–316
47 Welch, J.J. *et al.* (2008) Divergence and polymorphism under the nearly neutral theory of molecular evolution. *J. Mol. Evol.* 67, 418–426
48 Eyre-Walker, A. and Keightley, P.D. (2007) The distribution of fitness effects of new mutations. *Nat. Rev. Genet.* 8, 610–618
49 Kousathanas, A. and Keightley, P.D. (2013) A comparison of models to infer the distribution of fitness effects of new mutations. *Genetics* 193, 1197–1208
50 Tamuri, A.U. *et al.* (2012) Estimating the distribution of selection coefficients from phylogenetic data using sitewise mutation-selection models. *Genetics* 190, 1101–1115
51 Peel, D. *et al.* (2012) Accounting for missing data in the estimation of contemporary genetic effective population size (Ne). *Mol. Ecol. Resour.* 13, 243–253
52 Miller, C.R. *et al.* (2011) Mutational effects and population dynamics during viral adaptation challenge current models. *Genetics* 187, 185–202
53 Nagel, A.C. *et al.* (2012) Stickbreaking: a novel fitness landscape model that harbors epistasis and is consistent with commonly observed patterns of adaptive evolution. *Genetics* 190, 655–667
54 Razeto-Barry, P. *et al.* (2012) The nearly neutral and selection theories of molecular evolution under the fisher geometrical framework: substitution rate, population size, and complexity. *Genetics* 191, 523–534
55 Poon, A. and Otto, S.P. (2000) Compensating for our load of mutations: freezing the meltdown of small populations. *Evolution* 54, 1467–1479
56 Hartl, D.L. and Taubes, C.H. (1996) Compensatory nearly neutral mutations: selection without adaptation. *J. Theor. Biol.* 182, 303–309
57 Cherry, J.L. (1998) Should we expect substitution rate to depend on population size? *Genetics* 150, 911–919
58 Kawecki, T.J. *et al.* (2012) Experimental evolution. *Trends Ecol. Evol.* 27, 547–560
59 Burke, M.K. (2012) How does adaptation sweep through the genome? Insights from long-term selection experiments. *Proc. R. Soc. B* 279, 5029–5038
60 Nabholz, B. *et al.* (2009) The erratic mitochondrial clock: variations of mutation rate, not population size, affect mtDNA diversity across birds and mammals. *BMC Evol. Biol.* 9, 54
61 Lanfear, R. *et al.* (2010) Mutation rate is linked to diversification in birds. *Proc. Natl. Acad. Sci. U.S.A.* 107, 20423–20428
62 Nabholz, B. *et al.* (2008) Strong variations of mitochondrial mutation rate across mammals: the longevity hypothesis. *Mol. Biol. Evol.* 25, 120–130
63 Woolfit, M. (2009) Effective population size and the rate and pattern of nucleotide substitutions. *Biol. Lett.* 5, 417–420
64 Moran, N.A. (1996) Accelerated evolution and Muller's rachet in endosymbiotic bacteria. *Proc. Natl. Acad. Sci. U.S.A.* 93, 2873–2878
65 Woolfit, M. and Bromham, L. (2003) Increased rates of sequence evolution in endosymbiotic bacteria and fungi with small effective population sizes. *Mol. Biol. Evol.* 20, 1545–1555
66 Woolfit, M. and Bromham, L. (2005) Population size and molecular evolution on islands. *Proc. Biol. Sci.* 272, 2277–2282
67 Popadin, K. *et al.* (2007) Accumulation of slightly deleterious mutations in mitochondrial protein-coding genes of large versus small mammals. *Proc. Natl. Acad. Sci. U.S.A.* 104, 13390–13395
68 Nikolaev, S.I. *et al.* (2007) Life-history traits drive the evolutionary rates of mammalian coding and noncoding genomic elements. *Proc. Natl. Acad. Sci. U.S.A.* 104, 20443–20448
69 Lartillot, N. and Poujol, R. (2011) A phylogenetic model for investigating correlated evolution of substitution rates and continuous phenotypic characters. *Mol. Biol. Evol.* 28, 729–744
70 Lartillot, N. and Delsuc, F. (2012) Joint reconstruction of divergence times and life-history evolution in placental mammals using a phylogenetic covariance model. *Evolution* 66, 1773–1787
71 Romiguier, J. *et al.* (2013) Genomic evidence for large, long-lived ancestors to placental mammals. *Mol. Biol. Evol.* 30, 5–13
72 Gayral, P. *et al.* (2013) Reference-free population genomics from next-generation transcriptome data and the vertebrate-invertebrate gap. *PLoS Genet.* 9, e1003457
73 Wright, S.D. *et al.* (2009) Slower tempo of microevolution in island birds: implications for conservation biology. *Evolution* 63, 2275–2287
74 Gossmann, T.I. *et al.* (2012) The effect of variation in the effective population size on the rate of adaptive molecular evolution in eukaryotes. *Genome Biol. Evol.* 4, 658–667
75 Strasburg, J.L. *et al.* (2011) Effective population size is positively correlated with levels of adaptive divergence among annual sunflowers. *Mol. Biol. Evol.* 28, 1569–1580
76 Phifer-Rixey, M. *et al.* (2012) Adaptive evolution and effective population size in wild house mice. *Mol. Biol. Evol.* 29, 2949–2955
77 Petit, N. and Barbadilla, A. (2009) Selection efficiency and effective population size in *Drosophila* species. *J. Evol. Biol.* 22, 515–526
78 Siol, M. *et al.* (2010) The population genomics of plant adaptation. *New Phytol.* 188, 313–332
79 Lemmon, A.R. *et al.* (2012) Anchored hybrid enrichment for massively high-throughput phylogenomics. *Syst. Biol.* 61, 727–744
80 Faircloth, B.C. *et al.* (2012) Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. *Syst. Biol.* 61, 717–726
81 Luikart, G. *et al.* (2010) Estimation of census and effective population sizes: the increasing usefulness of DNA-based approaches. *Conserv.Genet.* 11, 355–373
82 Millar, C.D. *et al.* (2008) Mutation and evolutionary rates in Adélie penguins from the Antarctic. *PLoS Genet.* 4, e1000209
83 Charlesworth, B. (2002) Effective population size. *Curr. Biol.* 12, R716–R717
84 Frankham, R. (1995) Effective population size/adult population size ratios in wildlife: a review. *Genet. Res.* 66, 95–107
85 Charlesworth, B. (1994) *Evolution in Age-Structured Populations*, Cambridge University Press
86 Palstra, F.P. and Ruzzante, D.E. (2008) Genetic estimates of contemporary effective population size: what can they tell us about

the importance of genetic stochasticity for wild population persistence? *Mol. Ecol.* 17, 3428–3447
87 Nielsen, R. and Yang, Z. (2003) Estimating the distribution of selection coefficients from phylogenetic data with applications to mitochondrial and viral DNA. *Mol. Biol. Evol.* 20, 1231–1239
88 Martin, G. and Lenormand, T. (2006) A general multivariate extension of Fisher's geometrical model and the distribution of mutation fitness effects across species. *Evolution* 60, 893–907
89 Haldane, J.B.S. (1927) A mathematical theory of natural and artificial selection, part V: selection and mutation. *Math. Proc. Camb. Philos. Soc.* 23, 838–844
90 Begun, D.J. and Aquadro, C.F. (1992) Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster*. *Nature* 356, 519–520
91 Hellmann, I. *et al.* (2003) A neutral explanation for the correlation of diversity with recombination rates in humans. *Am. J. Hum. Genet.* 72, 1527–1535
92 Roselius, K. *et al.* (2005) The relationship of nucleotide polymorphism, recombination rate and selection in wild tomato species. *Genetics* 171, 753–763
93 Tenaillon, M.I. *et al.* (2001) Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* ssp. *mays* L.). *Proc. Natl. Acad. Sci. U.S.A.* 98, 9161–9166
94 Gossmann, T.I. *et al.* (2011) Quantifying the variation in the effective population size within a genome. *Genetics* 189, 1389–1402
95 Ellegren, H. (2007) Characteristics, causes and evolutionary consequences of male-biased mutation. *Proc. Biol. Sci.* 274, 1–10
96 Hodgkinson, A. and Eyre-Walker, A. (2011) Variation in the mutation rate across mammalian genomes. *Nat. Rev. Genet.* 12, 756–766
97 Betancourt, A.J. and Presgraves, D.C. (2002) Linkage limits the power of natural selection in *Drosophila*. *Proc. Natl. Acad. Sci. U.S.A.* 99, 13616–13620
98 Wyckoff, G.J. *et al.* (2002) Molecular evolution of functional genes on the mammalian Y chromosome. *Mol. Biol. Evol.* 19, 1633–1636
99 Arguello, J.R. *et al.* (2010) Recombination yet inefficient selection along the *Drosophila melanogaster* subgroup's fourth chromosome. *Mol. Biol. Evol.* 27, 848–861
100 Bullaughey, K. *et al.* (2008) No effect of recombination on the efficacy of natural selection in primates. *Genome Res.* 18, 544–554