



## **Determinants of genetic diversity**

Hans Ellegren, Nicolas Galtier

### **► To cite this version:**

Hans Ellegren, Nicolas Galtier. Determinants of genetic diversity. *Nature Reviews Genetics*, 2016, 17 (7), pp.422 - 433. <10.1038/nrg.2016.58>. <hal-01900639>

**HAL Id: hal-01900639**

**<https://hal.science/hal-01900639v1>**

Submitted on 30 Oct 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

## Determinants of genetic diversity

Hans Ellegren<sup>1</sup> and Nicolas Galtier<sup>2</sup>

**Abstract** | Genetic polymorphism varies among species and within genomes, and has important implications for the evolution and conservation of species. The determinants of this variation have been poorly understood, but population genomic data from a wide range of organisms now make it possible to delineate the underlying evolutionary processes, notably how variation in the effective population size ( $N_e$ ) governs genetic diversity. Comparative population genomics is on its way to providing a solution to ‘Lewontin’s paradox’ — the discrepancy between the many orders of magnitude of variation in population size and the much narrower distribution of diversity levels. It seems that linked selection plays an important part both in the overall genetic diversity of a species and in the variation in diversity within the genome. Genetic diversity also seems to be predictable from the life history of a species.

### Genetic diversity

(Also known as genetic polymorphism). Variation in a DNA sequence between distinct individuals (or chromosomes) of a given species (or population).

### Allozymes

Allelic variants of proteins that can be separated by electrophoresis based on differences in charge or structure.

### Fixation

The complete spread of a mutation in the population such that it replaces all other alleles at a site.

The distinct individuals of a given species are not genetically identical. Their DNA sequences differ to some extent, and these differences form the genetic diversity, known as polymorphism, of a species. Ever since the earliest empirical population genetic studies were able to demonstrate a wealth of polymorphism in many species<sup>1,2</sup>, thus settling a mid-twentieth century debate about the extent of genetic variation in natural populations (BOX 1), the variable nature of gene and genome sequences has been identified as a key aspect of biological evolution. In particular, genetic diversity contributes to the ability of a species to respond to environmental changes, with implications in terms of, for example, human health<sup>3,4</sup>, breeding strategies in crops and farm animals<sup>5</sup>, management of infectious diseases<sup>6</sup> and conservation of endangered species<sup>7,8</sup>.

It was recognized early from data on allozymes that genetic diversity differs considerably among species, and this observation was subsequently confirmed by DNA sequence data (BOX 2). For instance, approximately 3% of the genome of a fruitfly, *Drosophila simulans*, is variable<sup>9,10</sup>, whereas only 0.1% of a human genome shows allelic variation<sup>11,12</sup>. Remarkably, the reasons for such a variation in diversity have in large part been mysterious (BOX 2). Polymorphism also varies substantially across loci and chromosomes, as demonstrated in various species of plants<sup>13,14</sup>, fungi<sup>15</sup> and animals<sup>9,16,17</sup>.

From a theoretical viewpoint, genetic diversity can be thought of as reflecting the balance between appearance and disappearance of genetic variants (alleles). New alleles appear at each generation by spontaneous mutation due to DNA replication errors or mutagen-induced DNA damage. The rate of mutation is not constant across the genome<sup>18</sup> and among species<sup>19</sup>, which

might explain part of the observed variation in genetic diversity. Genetic diversity is also governed by the rate of allele loss and fixation. Loci with neutral alleles are largely influenced by genetic drift, which is the random fluctuation of allele frequency across generations<sup>20</sup>. In an idealized, panmictic (so-called Wright–Fisher) population with an equal expected contribution of individuals to reproduction, the strength of genetic drift is inversely proportional to the size of the population. However, real populations typically depart from Wright–Fisher assumptions in several respects, hence the concept of effective population size ( $N_e$ ), which is the size of an idealized population that would show the same amount of genetic diversity as the population of interest (BOX 3). The neutral theory of molecular evolution<sup>21</sup> therefore predicts that in a population of constant size the genetic diversity should be proportional to  $N_e$ . Further outlined in BOX 3, we here immediately make an important distinction:  $N_e$  varies over time, with long-term  $N_e$  explaining current levels of genetic diversity in populations but contemporary  $N_e$  explaining how strong drift currently is.

The census population size ( $N_c$ ), and therefore presumably  $N_e$ , varies by many orders of magnitude among taxa, with small organisms typically reaching densities immensely higher than larger ones. Very strong contrasts in the levels of genetic diversity should therefore be expected. Four decades ago, however, it was observed that the across-species variation in genetic diversity is much narrower than the variation in abundance<sup>22</sup>. This long-standing issue in population genetics is known as Lewontin’s paradox. Several potential solutions to this paradox have been considered, including: the influence of demographic fluctuations, which tend to level out the

<sup>1</sup>Department of Evolutionary Biology, Evolutionary Biology Centre, Uppsala University, Norbyvägen 18D, Uppsala SE-753 36, Sweden.

<sup>2</sup>Institute of Evolutionary Sciences, French National Centre for Scientific Research (CNRS), University of Montpellier 2, Place E. Bataillon, Montpellier 34095, France.

Hans.Ellegren@ebc.uu.se;  
Nicolas.Galtier@univ-montp2.fr

doi:10.1038/nrg.2016.58  
Published online 6 Jun 2016

## Box 1 | How genetic diversity was perceived before it could be measured

The ancient Greek philosophers' long-dominating metaphysical view of nature saw variation among material bodies as imperfections of the ideal forms. This typological doctrine was not seriously challenged until Charles Darwin postulated that varieties are essential facets of nature. He concluded that the variation that exists within species is directly connected to variation between species through evolution by natural selection (and isolation). Gregor Mendel's laws of heredity introduced the genetic component of variation, especially by formulating the expectations for variation among offspring.

Mid-twentieth century evolutionary biologists had been equipped with the theory of population genetics that was developed primarily by Ronald A. Fisher, J. B. S. Haldane and Sewall Wright. Integration of this theory with Mendelian genetics and Darwinian principles led to the formation of the modern evolutionary synthesis, in which gradual change in characters could be understood from changes in gene frequencies: that is, within the realm of genetic diversity. However, in the absence of means for investigating genetic diversity other than indirectly at the phenotypic level, the extent and character of this diversity remained elusive.

Two, sometimes described as polarized, views on genetic diversity held by evolutionary biologists at the time of the modern synthesis were named the 'classical' and the 'balance' theories by Theodosius Dobzhansky<sup>149</sup>. The classical school considered that genes usually have a 'wild-type allele' for which most individuals are homozygous. The prominent role of natural selection would function as an efficient 'sieve' to eliminate unfit alleles. On rare occasions, individuals might be heterozygous for recessive, deleterious alleles and inbreeding would enhance the rate by which these became exposed to selection. Hermann Müller and others attempted to estimate the fraction of heterozygous loci and arrived at very low estimates. When an advantageous allele sometimes arose, it would quickly sweep to fixation by natural selection. According to this view, genetic diversity in natural populations would thus be low.

By contrast, the balance school thought of most loci as heterozygous, containing many alleles without one of them necessarily corresponding to a wild-type gene. Selection could act to maintain such high genetic diversity by some form of balancing selection, such as heterozygote advantage (over-dominance) or frequency-dependent selection. An important distinction between the classical and the balance schools of thought was that the classical theory predicted more variation between than within species, whereas this was not the case for the balance theory.

When technology for the first time allowed genetic diversity to be quantified in populations in the 1960s, it was quickly realized that there is typically a high degree of variation. This could potentially have been seen as support for the balance theory; however, at the same time, the idea that many alleles would be selectively neutral was formalized under the neutral theory of molecular evolution by Motoo Kimura<sup>21</sup>, which in several respects revolutionized the field. Later, the neutral theory of molecular evolution was further developed into the nearly neutral theory by Tomoko Ohta<sup>150</sup>, in which slightly deleterious mutations are also thought to segregate in the population at a frequency inversely proportional to the effective population size. This theory largely represents our current view of genetic diversity at functional loci. Of course, the realization from molecular work in the 1970s and onwards that substantial parts of the genome are probably non-functional, and hence neutral, means that direct selection (but not linked selection; see the main text) would not be an issue for genetic diversity in such regions of the genome.

### Genetic drift

Fluctuation of allele frequency among generations in a population owing to the randomness of survival and reproduction of individuals, irrespective of selective pressures.

### Effective population size

( $N_e$ ). The number of breeding individuals in an idealized population that would show the same amount of genetic drift (or inbreeding, or any other variable of interest) as the population under consideration.

### Census population size

( $N$ ). The number of individuals in a population.

### Frequency-dependent selection

A form of selection in which the selective advantage or disadvantage of a genotype is dependent on its frequency relative to other genotypes.

### Bottleneck

A sharp and rapid reduction in the size of a population.

### Heterozygosity

The probability that two randomly sampled gene copies in a population carry distinct alleles; a measure of the genetic diversity.

long-term average  $N_e$ ; natural selection, which affects the diversity of neutral sequences through linkage; and mutation rate variation. Moreover, both demography and selection generate differences in  $N_e$  among genomic regions, and mutation rate variation may contribute to differences in local levels of genetic diversity. There are thus two important axes of variation in diversity: among species and within genomes.

For a long time, the lack of genome-wide data in a sufficiently large number of species meant that the literature on the variation in polymorphism and Lewontin's paradox remained mainly theoretical or speculative. Despite several attempts to relate genetic polymorphism to species traits and ecology (BOX 2), a 2012 review article on the subject, based on data from mostly just a small number of loci per species, concluded that diversity varies "mostly in ways that we still do not understand" (REF. 23). But with the advent of high-throughput sequencing technologies, it is now possible to approach genome-wide patterns of genetic diversity in non-model organisms at an affordable cost. Seizing this opportunity, several recent studies have revived the 'long riddle' of the determinants of genetic diversity, especially the influence of life-history traits versus the population history of species, the impact of mating systems and the prevalence of linkage effects. In this Review, we summarize

and synthesize recent findings pertinent to explaining genetic diversity. We focus on how and why the degree of genetic variability varies among species (populations) as well as the extent and character of variation in diversity within genomes. Our focus is on neutral diversity, which is easiest to define, measure and interpret. There is strong evidence that neutral genetic diversity correlates with phenotypic diversity as well as with the fitness of individuals and populations<sup>24,25</sup>. Given the many differences in demography, reproduction and genome evolution between eukaryotes and prokaryotes, we concentrate on the former category of organisms.

### Variation in diversity among species

**Life history, but not population history, predicts genetic diversity.** It is widely acknowledged that the genetic polymorphism currently carried by any particular species has been shaped by its demographic history. Demographic fluctuations due to biotic or abiotic environmental changes presumably affect most, if not all, living species<sup>26,27</sup>. Ecological drivers of such fluctuations include climate change, pest outbreaks and, more recently, human activities<sup>28</sup>. Changes in  $N_e$  over time are expected to deeply affect genetic diversity<sup>29</sup>. A strong population bottleneck, for instance, results in a rapid decay of heterozygosity due to enhanced genetic drift.

A similar effect is expected in the case of a founder event, when a new population is created by the immigration of a small number of individuals into a yet unoccupied area (such as an island)<sup>30</sup>. A large number of studies have therefore attempted to interpret the current amount and structure of genetic diversity in light of past climatic, biogeographical and ecological events, such as Pleistocene glaciations<sup>31</sup>, insularity<sup>32</sup>, habitat fragmentation<sup>33</sup>, ecological disturbance<sup>34</sup> and human intervention<sup>35</sup>. However, it is currently unclear how historical, contingent demographic effects compare to intrinsic determinants of species abundance (such as the size of organisms), which have been used as proxies for  $N_e$  in the molecular evolutionary literature<sup>36</sup>.

Romiguier and colleagues<sup>37</sup> recently investigated genome-wide patterns of genetic diversity across 90 distinct species of animals from various taxonomic

groups. The data set, which was made of thousands of polymorphic coding sequences per species, surprisingly revealed no major influence of latitude, endemism, conservation status or any variable potentially related to the demographic history of the species on current levels of genetic diversity. There was, however, a highly significant taxonomic effect: species from a given family tend to share a similar level of genetic diversity, irrespective of their specific history, indicating that recent fluctuations in  $N_e$  have not been strong enough to erase the dominant, long-term pattern.

By contrast, a strong correlation was reported between genetic diversity and life-history traits of the species, such as body mass, longevity and, surprisingly, reproductive strategy. Highly fecund species that release high numbers of small eggs into the environment (for example, mussels, urchins and earthworms; the so-called

#### Box 2 | Meta-analyses in population genetics — a historical perspective

The first comparative analyses of species-level genetic diversity appeared in the early 1970s. At this time, allelic variation was detected on the basis of the distinctive electrophoretic properties of proteins — the so-called allozyme technique<sup>2,151</sup>. Allozyme diversity has been estimated in hundreds of species, typically using 10–30 loci. This revealed substantial between-species variation in heterozygosity but no linear scaling with census population size<sup>22,152</sup>. Nevo and colleagues<sup>153</sup> correlated allozyme heterozygosity to various taxonomic, biogeographical, ecological and life-history variables in 1,111 species of plants and animals. They detected a higher amount of genetic polymorphism in invertebrates than in vertebrates and a weak but significant effect of body size and geographical range. Focusing on plants, Hamrick and Godt<sup>154</sup> reported a significant effect of life history and mating system on allozyme diversity, with self-fertilizing species being less polymorphic than outcrossing ones, whereas Cole<sup>155</sup> reported reduced diversity in rare compared with common species of plants. Allozyme data, however, are affected by several limitations, such as potential non-neutrality, gene duplication and the confounding effects of post-translational modifications of proteins.

The development of molecular biology in the 1980s, particularly PCR and Sanger sequencing, established the maternally inherited mitochondrial DNA (mtDNA) as the most popular population genetic marker, particularly in animals<sup>156</sup>. A large number of phylogeographical studies were published, in which one or several mtDNA fragments were sequenced in typically 10–100 individuals per species. In a meta-analysis of 1,683 animal species, Bazin and colleagues<sup>157</sup> observed no detectable influence of life-history traits of species or taxonomy on mtDNA diversity. Similarly low levels of correlation between mtDNA diversity and various indicators of species abundance were reported when specifically analysing mammals<sup>44</sup> or birds<sup>158</sup>, whereas a significantly positive relationship between some measures of mtDNA diversity and population density was reported in fish<sup>45</sup>.

A characteristic of the mitochondrial genome is the absence of recombination. For this reason, the distinct polymorphic positions observable in a mtDNA data set are part of a single locus: that is, share a common genealogy<sup>159</sup>. Across-species patterns of mtDNA diversity are therefore expected to be strongly affected by natural selection at linked loci, either within the mitochondrial genome or in other maternally inherited genetic units, such as the W chromosome in female heterogametic organisms<sup>160</sup> and intracellular symbionts in arthropods<sup>161</sup>. An erratic mtDNA mutation rate<sup>162</sup> is also a confounding factor that might partly explain the weak relationships between species abundance and mtDNA diversity<sup>163</sup>.

Besides mtDNA, microsatellites (also called short tandem repeats) have been heavily used to characterize within-species patterns of genetic diversity<sup>164</sup>. These loci, which are characterized by an extremely high mutation rate, correspond to short, repeated motifs of typically 1–6 bp each — the allelic state being determined by the number of repeats. Meta-analyses of microsatellite diversity have often been related to conservation issues. Spielman *et al.*<sup>43</sup>, for instance, found that endangered species of plants and animals tend to be less polymorphic than their non-endangered relatives, implying that recent population declines have had a genetic impact. McCusker and Bentzen<sup>45</sup> and Pinsky and Palumbi<sup>47</sup> also reported reduced diversity at microsatellite loci in overfished species, compared with other fish. Microsatellite analyses typically evaluate ~10 loci per species, a limited sample that is not always representative of the whole genome<sup>165</sup>. Importantly, the mutation rate can vary by orders of magnitude among microsatellite loci<sup>166</sup>, questioning the reliability of comparative analyses across distantly related species — in which distinct loci are to be genotyped.

With the advent of next-generation sequencing technologies, genome-wide data sets and single-nucleotide polymorphisms have greatly expanded the population genetics literature. Compared with Sanger sequencing<sup>23</sup>, the thousands of loci provided by these techniques offer the opportunity to investigate patterns of genetic variation among species with unprecedented accuracy<sup>167</sup>. The first meta-analyses of diversity genome wide, which were conducted at a very broad taxonomic scale<sup>168</sup>, yielded stimulating hypotheses regarding the impact of effective population size on molecular evolution<sup>36</sup>. More recently, taxon-specific analyses have provided interesting insight into the intricate relationship between species ecology, population history and genetic diversity<sup>37,46</sup>. It is likely that the forthcoming accumulation of genome-scale polymorphism data sets in large numbers of species will help to clarify these complex issues and perhaps eventually elucidate Lewontin's old paradox<sup>22</sup>.

Box 3 | The effective population size and its estimation

If populations were infinitely large and biological processes such as selection, mutation and migration were ignored, then the frequency of genetic variants should stay the same over generations. This scenario obviously does not apply in practice, and the fact that population sizes are finite means that stochastic fluctuations in allele frequencies are unavoidable. The sampling process of alleles is referred to as genetic drift, and the extent of genetic drift is given by the effective population size ( $N_e$ )<sup>169</sup>. The concept of  $N_e$  rather than simply population size ( $N$ ; or census population size ( $N_c$ )) is motivated by several very probable violations from the idealized scenario of equal reproduction and survival probabilities among all individuals in the population. A common verbal definition of  $N_e$  is that it corresponds to the number of individuals of an idealized population that would show the same amount of genetic diversity as an actual population. Another way of putting it is to refer to the number of individuals contributing to reproduction, but this is a gross over-simplification and would only be valid in an idealized Wright–Fisher population of constant size with random mating.

Directly estimating  $N_e$  is typically impracticable as it would require knowledge about, among other things, the total number of individuals of a species and the variance in lifetime reproductive success among individuals. Indirect estimates based on genetic parameters are therefore often used. The population genetic parameter  $\Theta$  (also known as the population mutation rate) predicts the amount of genetic diversity in a randomly mating population of constant size and equals  $4N_e\mu$ , where  $\mu$  is the rate of mutation per site per generation. The scaling factor is determined by the number of chromosomes per breeding pair, with ‘4’ referring to the autosomal sequence of a diploid and sexually reproducing species. Under neutrality and constant  $N_e$  over time, the expected value of  $\Theta$  is equal to the observed nucleotide diversity ( $\pi$ ; which is, essentially, heterozygosity), and if representative estimates of  $\pi$  and  $\mu$  are obtained,  $N_e$  can be solved from  $\Theta = 4N_e\mu$ . This approach provides a long-term estimate of  $N_e$  and is considered to reflect the harmonic mean of  $N_e$  over time. Recall that contemporary genetic diversity is the result of mutations and genetic drift that occurred in the past, with a particularly large influence of periods of low  $N_e$ . For example, it explains why the human long-term  $N_e$  is only a few tens of thousands ( $\pi \approx 1 \times 10^{-3}$  and  $\mu \approx 1 \times 10^{-8}$ ) despite the contemporary population being several orders of magnitude larger than that.

In conservation biology, it is often relevant to consider current  $N_e$  as a measure of how fast a population loses genetic diversity owing to genetic drift. Contemporary  $N_e$  can be seen in different ways — for example, the loss of heterozygosity due to inbreeding or a change in allele frequencies over time — and also on different spatial scales (ranging from local to global)<sup>170,171</sup>. Up until recently, two genetic methods have dominated in the estimation of contemporary  $N_e$ : temporal comparison of allele frequencies from two sampled cohorts of a population and approaches based on estimation of linkage disequilibrium in a population sample<sup>172</sup>. Both are sensitive to several confounding factors, such as migration, population structure and overlapping generations<sup>170</sup>. Recently, some new and exciting methods for estimating  $N_e$  in the recent past (~5–200 generations) based on analyses of chromosomal segments that are identical by descent have been presented<sup>173–175</sup>. In humans, such estimates are substantially larger than those of the long-term  $N_e$ , reflecting the contrast in density between modern and ancient human populations, and the effect of past population bottlenecks on current genetic diversity.

*r*-strategists) are much more polymorphic than species that produce a small number of relatively large offspring and provide parental care (for example, penguins, termites and seahorses; called *K*-strategists) (FIG. 1). Such a strong influence of life-history traits on genetic diversity had been suspected based on allozyme data (BOX 2) but was largely missed in previous comparative analyses of genomic sequences, probably because of lack of power due to limited sampling<sup>23</sup>.

A confounding effect of the mutation rate seems to be excluded. Being typically higher (generating more diversity) in long-lived than in short-lived eukaryotes<sup>38</sup>, the mutation rate is actually expected to act against the observed relationship, if anything. However, variation in mutation rate across organisms may help to explain

Lewontin’s paradox. According to the drift-barrier hypothesis, selection should only be able to reduce the rate of germline mutation (to protect gametes from deleterious mutations) to a point at which the selective advantage of further incremental reduction is not higher than the power of genetic drift<sup>19</sup>. As selection is more efficient in large populations, a negative correlation between mutation rate and  $N_e$  can be predicted. Although the number of species from which direct estimates of the germline mutation rate is available is still limited, existing data do seem to support this prediction<sup>19,38</sup>. Whole-genome re-sequencing of pedigrees provides a new and exciting approach for estimating mutation rate, and data from various organisms should soon become available<sup>39</sup>.

The analysis suggests that the effect of life-history traits on genetic diversity is mediated by  $N_e$  (REF. 37) (FIG. 2). A theory suggests that the minimal population size is what matters most as far as the long-term average  $N_e$  is concerned<sup>40</sup>. It would seem plausible that species traits, and reproductive strategy in particular, determine the minimal sustainable density of a species. This might be pertaining to the apparent ability of *K*-strategists to survive even severe population bottlenecks<sup>41,42</sup>: that is, experience substantial losses of diversity while escaping extinction. This is a speculative explanation that deserves additional investigation. According to this hypothesis, the intensity of past fluctuations would be better captured by life-history traits of a species than by the current range or density of a species. This study<sup>37</sup> therefore highlights the complex interplay between demographic dynamics and genetic diversity, which is a controversial issue<sup>43–47</sup> that has crucial relevance to conservation biology and the management of biodiversity.

**Estimating variation in  $N_e$  over time.** Given the likely importance of the demographic history to current levels of genetic diversity (although noting the caveats described above), knowledge about past demographic processes is key to understanding the relationship between population size and diversity. It has previously been difficult to assess temporal fluctuations in  $N_e$  using genetic data other than testing for simple demographic scenarios, such as exponential or logarithmic growth. The introduction of skyline-plot methods<sup>48,49</sup> based on the coalescent theory allowed for estimation of more complex demographic histories, but these methods are limited to analyses of non-recombining loci. Recently, a large number of studies have applied the pairwise sequentially Markovian coalescent (PSMC) approach<sup>50</sup> to estimate trajectories of changes in  $N_e$  over considerable time periods based on deep whole-genome sequencing data. This method analyses the length distribution and the time to the most recent common ancestor of identity-by-descent segments, identifies historical recombination events and infers past population sizes from the distribution of coalescence times. For instance, a bottleneck that occurred  $t_1$  generations ago is expected to result in a substantial proportion of the genome coalescing (that is, having a common ancestor that lived) at time  $t_1$ . Conversely, if a

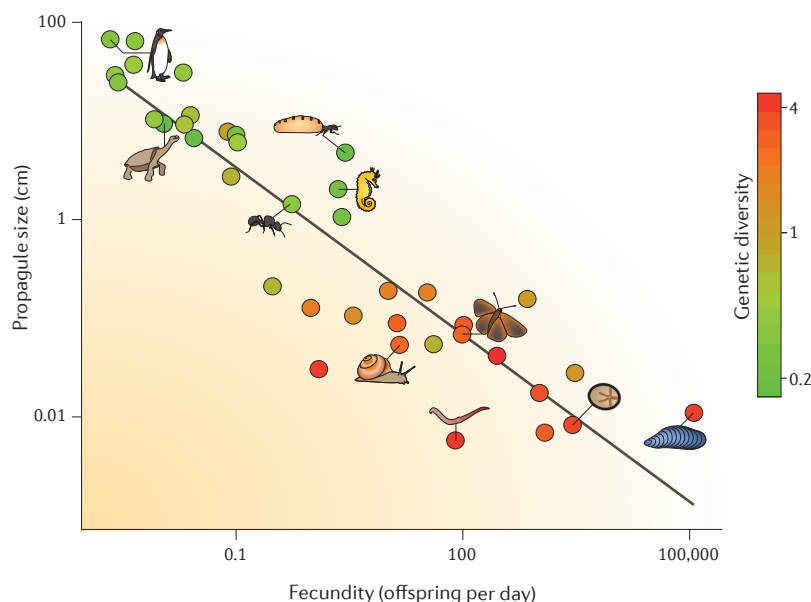
Drift-barrier hypothesis

The idea, based on the concept of diminishing returns, that selection can only improve a trait up to a point at which the next incremental improvement will be overwhelmed by the power of genetic drift.

Coalescent theory

A retrospective model of the distribution of gene divergence in a genealogy.





**Figure 1 | Genetic diversity and the r/K gradient in animals.** The average per-day fecundity is on the x axis and the average size of eggs or juveniles is on the y axis; each dot is for a family (one to four species each). The colour scale indicates the average nucleotide diversity at synonymous positions, expressed in per cent. The negative correlation reflects a trade-off between quantity and size of offspring. *r*-strategists (bottom right; for example, blue mussels, heart urchins and lumbricid earthworms) are more polymorphic than *K*-strategists (top-left; for example, penguins, Galapagos tortoises and subterranean termites). Figure from REF. 37, Nature Publishing Group.

#### Identity-by-descent segments

Chromosomal segments carried by two or more individuals that are identical because they have been inherited from a common ancestor, without recombination.

#### Polyploidization

A form of genome evolution in which the number of sets of chromosomes increases.

#### Linkage disequilibrium

The non-random association of alleles at two loci, often but not always due to physical linkage on the same chromosome.

#### Selective sweep

Elimination or reduction of genetic diversity in the neighbourhood of a beneficial allele that increases in frequency in the population, typically after an environmental change.

#### Hard sweeps

Selective sweeps in which the beneficial allele corresponds to a single, new mutation appearing after an environmental change.

relatively small number of genomic segments coalesce at time  $t_2$ , this is indicative of a large number of ancestors, hence a large  $N_e$  at  $t_2$ .

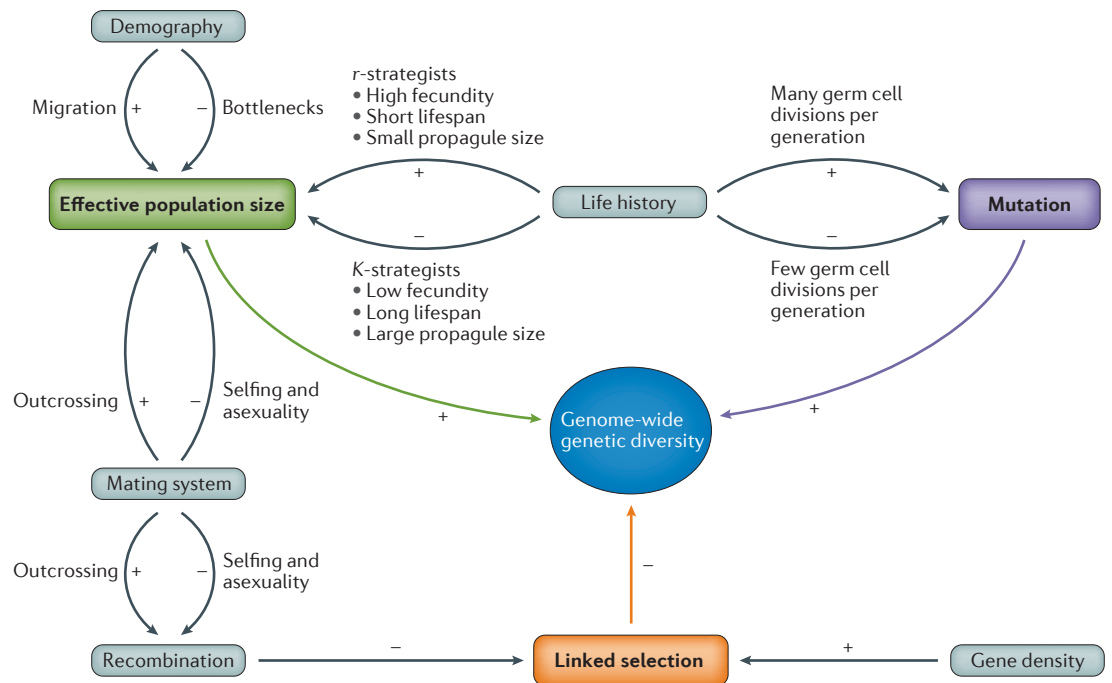
There is continued development of new methods for model-flexible inference of demographic history, including improved resolution in the detection of recent changes in population size<sup>51,52</sup>. An important conclusion from recent studies inferring the temporal dynamics of populations is that many extant species have gone through one or more periods of expansion, contraction or both, with significant effects on  $N_e$  (REF. 53).

**Influence of mating systems.** Comparative analyses of the distribution of genetic diversity across species have typically focused on sexual, outcrossing species, particularly of animals. Mating systems, however, vary considerably in nature and should influence genetic diversity. Self-fertilization (selfing) is expected to reduce  $N_e$  by a factor of two — and even more if, as suspected, selfing was associated with a higher propensity to experience bottlenecks through founding effects and a more pronounced effect of linked selection<sup>54,55</sup> (see the next section). Meta-analyses in plants have empirically confirmed this prediction<sup>56–58</sup>, as have genome-wide analyses of specific selfing versus outcrossing pairs of diploid plant<sup>59</sup> and animal<sup>60–63</sup> species. It has been suggested that the reduced efficacy of natural selection due to increased drift and linkage could increase extinction risks in selfers (self-fertilizing species), but the empirical evidence in favour of this hypothesis is not compelling so far<sup>64</sup>.

Asexual reproduction, by contrast, is expected to lead to increased heterozygosity because in the absence of chromosome segregation new mutations remain in a heterozygous state — the so-called Meselson effect<sup>65,66</sup>. This theoretical expectation has only met equivocal empirical corroboration so far. The Meselson effect was first reported in the asexual bdelloid rotifers<sup>66</sup>, but subsequent analyses in this taxon failed to confirm this process. In several other groups of asexual animals, the confounding effects of polyploidization, hybridization and/or young asexual lineages have hampered a convincing demonstration of accumulating heterozygosity<sup>67–69</sup>. A recent genome-wide analysis of asexual versus sexual species pairs in the plant genus *Oenothera*, however, revealed a significant increase in genetic diversity in asexual species that was associated with a reduced efficacy of purifying selection, which is in agreement with theoretical predictions<sup>70</sup>. This and other recent studies<sup>59–61</sup> demonstrate the power of high-throughput sequencing technologies in characterizing the effect of mating systems on genetic variation patterns. Significant advances in this topic are to be expected in forthcoming years.

Of note, despite the associated increase in heterozygosity, asexual populations are not expected to respond to environmental changes more efficiently than sexual populations. Quite the contrary, asexuality is predicted to increase the risk of population extinction because asexual populations lose the potential to create new combinations of alleles by recombination, which greatly reduces their adaptive potential<sup>71,72</sup>. This prediction is also valid for selfers<sup>73</sup>, and it has been empirically validated, with few exceptions<sup>66,74</sup>, that existing asexual or selfing lineages are of relatively recent origin<sup>75,76</sup>, which indicates a rapid turnover of these atypical mating systems.

**Linked selection strongly affects neutral diversity in large populations.** The fate of a mutation is determined not only by its own effect on fitness but also by selection applying to linked loci<sup>77,78</sup>. When selection (or any genetic process for that matter) changes the frequency of an allele in the population, it also changes the frequency of the haplotype in which the selected allele resides. From one generation to the next, haplotypes can be defined as chromosomal segments flanked by two adjacent recombination events (or between a recombination event and a chromosome end). In subsequent generations, further recombination will cause these segments to become progressively shorter, with the haplotypes that contain the selected allele varying in length depending on the particular history of recombination events along each coalescence lineage. As long as the selected allele is in linkage disequilibrium with segregating variants at nearby loci, neutral diversity is reduced by linked selection. For example, rapid fixation of a beneficial mutation (that is, a selective sweep) erases pre-existing polymorphisms at both the selected locus and nearby loci<sup>79,80</sup>. This is especially the case for hard sweeps. In the case of soft sweeps based on standing genetic variation, the selected allele is initially likely to be associated with several haplotypes, and the reduction in diversity resulting from the sweep will not be as pronounced. If pervasive, linked positive selection



**Figure 2 | Overview of determinants of genetic diversity.** Effective population size, mutation rate and linked selection are the main factors affecting diversity. These factors are in turn governed by several other parameters. The direction of correlation is indicated by the + and – symbols. Selfing, self-fertilization.

might affect the genome-wide average polymorphism level, which is known as the *genetic draft hypothesis*<sup>81,82</sup>. Purifying selection, by which deleterious alleles are lost, has a similar influence on local levels of diversity, which are reduced in the vicinity of the selective targets. This process is referred to as *background selection*<sup>83,84</sup>.

Selection might counteract the expected correlation between  $N_e$  and neutral diversity by virtue of linked selection being more prevalent in large populations, potentially making  $N_e$  a poor predictor of diversity<sup>82</sup>. Note that the fixation probability of a non-neutral allele is determined by both its selective advantage or disadvantage (the selection coefficient,  $s$ ) and  $N_e$ . For negative selection, this dual dependence is valid throughout a wide range of  $s$  and  $N_e$ , whereas for beneficial mutations the probability of fixation is largely governed by  $s$  when  $N_e$  increases. Conversely, there is more opportunity for beneficial mutations to arise in large populations.

Critically testing the effect of linked selection in relation to  $N_e$  on genetic diversity in different species has previously not been possible in the absence of genome-wide polymorphism data for most organisms. However, advances in next-generation sequencing technology have opened up new avenues for genomic investigation of non-model organisms, including population genomic studies. Corbett-Detig and colleagues<sup>85</sup> recently compiled genomic re-sequencing data from 40 sexually reproducing species of animals and plants, and used these data to test the relationship between proxies of  $N_e$  and genetic diversity. They took advantage of an expected correlation between recombination rate and diversity in different regions of the genome (see below), and reasoned that if linked selection shapes genome-wide diversity to

an extent that is dependent on  $N_e$ , then the correlation should be strongest in species with large populations. This correlation was found, and the authors concluded that “natural selection truncates the distribution of neutral genetic diversity” (REF. 85) among species.

This analysis demonstrates that linked selection is pervasive and results in a reduction of the genome-wide average diversity by a factor of up to three in large- $N$  species, which is considerable. It should be noted, however, that diversity varies by several orders of magnitude among species<sup>23,37</sup>. The evidence to date thus suggests that linked selection may not be a sufficient explanation for Lewontin’s paradox<sup>86</sup>, but further work is necessary to fully understand the effect of linked selection on diversity<sup>87</sup>; previous work in *Drosophila melanogaster* suggested that the effect might be stronger than predicted by simple models<sup>87–89</sup>. More generally, quantification of the relative effect of linked selection, life-history and ecological factors, through  $N_e$ , is an important challenge for future research in this field (FIG. 2).

### Variation in diversity within genomes

**Recombination rate and the density of targets for selection.** So far, we have considered  $N_e$  as a characteristic of a population. This indeed makes intuitive sense given its definition as the number of individuals in an idealized population expected to show the same amount of genetic diversity as the population under consideration (BOX 3). However, as touched upon above, neutral genetic diversity is not a constant entity across the genome, and this means that either or both of the factors that mediate this diversity — mutational input and  $N_e$  (BOX 3) — must vary across the genome. There is local

#### Soft sweeps

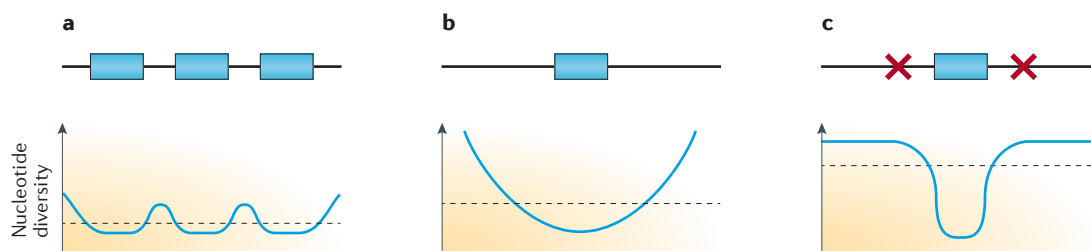
Selective sweeps in which the beneficial allele exists before an environmental change (thus representing standing variation) and is initially neutral or even slightly deleterious, or appears several times independently.

#### Genetic draft

Pervasive reduction of genetic diversity owing to recurrent selective sweeps.

#### Background selection

Reduction of genetic diversity owing to selection against deleterious mutations at linked loci.



**Figure 3 | Genetic diversity affected by the density of targets for selection and by recombination rate.** A schematic illustration of the effects of linked selection on genetic (nucleotide) diversity around genes or other functional elements (boxes; upper panels). In the lower panels, solid lines indicate the local variation in diversity level and dashed lines indicate the average diversity in the whole region in question. In regions with a high density of targets of selection (part **a**), linked selection is pervasive and significantly reduces diversity compared with regions with a lower density of selection targets (part **b**). When the recombination rate is high (part **c**), the effect of linked selection becomes less prevalent, allowing maintenance of high diversity levels.

(from adjacent nucleotides up to kilobase level) as well as regional (around megabase level) variation in the rate of mutation<sup>18</sup> in the genome, but there is little evidence that this would be a strong driver of variation in diversity. This is surprising, and more work should be devoted to understanding the relationship between mutation rate and genetic diversity. For now, these findings leave  $N_e$  as the primary determinant<sup>20,90</sup>. Diversity-reducing linked selection is a main cause of intragenomic variation in  $N_e$ , but additional factors could contribute to variation in  $N_e$ ; for example, parts of the genome may be less permeable to gene flow between diverging lineages by harbouring loci involved in reproductive incompatibility<sup>91</sup>. This keeps diversity lower around incompatibility loci than in genomic regions where there is introgression.

When it comes to linked selection, two aspects dominate as to what extent it causes variation in diversity across the genome: the density of targets of selection and the rate of recombination<sup>78</sup> (FIG. 3). The more targets for selection there are in a genomic region, the more reduction in diversity could, in principle, be expected in the surrounding sequence (FIG. 3a,b). The more recombination there is around selected loci, the smaller the genomic window that will be affected by linked selection (FIG. 3c). Both aspects contribute to within-genome heterogeneity by generating valleys of reduced diversity in the vicinity of the selected loci<sup>78</sup>.

A positive correlation between the rate of recombination and nucleotide diversity was first described in *D. melanogaster*<sup>92</sup> and has been confirmed in many organisms since then<sup>13,93–97</sup>. The correlation was initially understood as an effect of genetic hitch-hiking<sup>92,98</sup>, but background selection has been increasingly appreciated as a contributing factor<sup>78,85,88,99–103</sup>, perhaps in many cases the dominating one (FIG. 2). One reason for this shift might be that the early work was concentrated on fruit-flies, which have a large  $N_e$  and in which adaptive evolution is more prevalent than in organisms with smaller populations<sup>101</sup>. Forward-time simulations as well as coalescent methods will be important to elucidate the relative role of selective sweeps and background selection in this context, and considerable progress has recently been made using approaches that can accommodate complex demographic and selection models<sup>104–107</sup>.

The above-mentioned finding that the strength of correlation between recombination rate and nucleotide diversity seems to be dependent on  $N_e$  (REF. 85) is also important when it comes to the range of recombination rates for which a correlation with nucleotide diversity might be seen. In a comparison between a species with high  $N_e$  (*D. melanogaster*) and a species with low  $N_e$  (wild horses), Corbett-Detig and colleagues<sup>85</sup> showed that diversity is reduced even in regions of relatively high recombination in the former species, whereas diversity is reduced only in regions with low recombination in the latter species. There are also examples of species with low-to-medium  $N_e$  in which reduced diversity is only seen in genomic regions with the lowest recombination rates<sup>108</sup>.

Although it is a well-established theory that the density of targets of selection should be important to the role of linked selection<sup>109–112</sup>, especially when it comes to background selection, it was not until recently that annotated genome sequences of non-model organisms have allowed incorporation of the frequency of selection targets as an explanatory variable in analyses of variation in diversity<sup>90,113–115</sup>. Most of this work has been based on gene (exon) density, and although this may very well be an appropriate (relative) proxy for the total density of sites under selection, more detailed investigation will become possible when genomes are properly annotated for sites evolving under purifying selection outside coding regions. The fact that both the recombination rate and the density of targets of selection affect genetic diversity implies that spurious results can be obtained if the two explanatory variables co-vary but are analysed separately for possible correlation with diversity<sup>14</sup>. This is especially the case when regions of high recombination rates (promoting diversity) coincide with regions of high density of selected sites (reducing diversity)<sup>102</sup>.

There are further twists to the relationship between recombination rate and diversity that can potentially confuse or even revert the expected positive correlation between the two parameters. One twist is that even in the absence of linked selection, the recombination rate would scale positively with diversity if recombination is mutagenic (or if, through some other causal link, recombination and mutation are both associated with genomic regions of high diversity). Although the idea

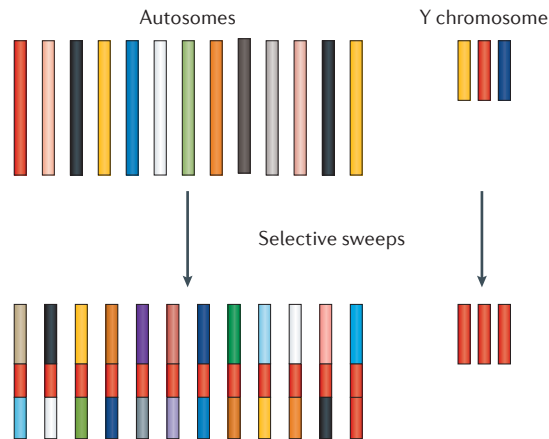
#### Introgression

New alleles entering the population by hybridization with members of a differentiated population or even a different species.

#### Hitch-hiking

The change in allele frequency at a locus that itself is not necessarily affected by selection but is genetically linked to a locus that is.





**Figure 4 | Genetic diversity in autosomes and the Y chromosome.** A schematic illustration of autosomes and the Y chromosome, with a range of colours symbolizing the different alleles present in the population and hence the degree of genetic diversity. The larger effective population size of the autosomes than of the Y chromosome means that the autosomes can sustain more variation than the Y chromosome. Moreover, in the absence of recombination on the Y chromosome (or on the W chromosome in organisms with female heterogamety), a strong selective sweep would effectively erase diversity on this chromosome. On autosomes, recombination around a locus that is subject to a selective sweep will allow diversity to be maintained elsewhere on the chromosome.

is not new<sup>94,116</sup>, it was not until recently that the direct detection of *de novo* mutations has been associated with crossovers<sup>117–119</sup>. It remains to be seen whether this association is quantitatively strong enough to have an impact on the correlation between recombination rate and diversity. However, many studies have failed to find a correlation between divergence and recombination rate, arguing against a strong role of recombination in governing variation in the rate of mutation<sup>78</sup>.

Another confounding factor may be GC-biased gene conversion (gBGC). This process, which is driven by recombination, increases the fixation probability of GC alleles<sup>120</sup> such that in regions of high recombination the allele frequency spectrum of AT → GC mutations is shifted towards high frequencies (counteracting stochastic loss of such alleles) and the spectrum of GC → AT mutations is shifted towards low frequencies<sup>121</sup>. gBGC can lead to a positive correlation between recombination rate and diversity if estimated based on AT → GC mutations, but not based on GC → AT mutations<sup>108</sup>. However, the predicted net effect of gBGC on genetic diversity is complex; it depends on genomic base composition and the proportion of AT → GC versus GC → AT mutations. Interestingly, as gBGC-like selection implies different fixation probabilities of different alleles, it (just as selection) becomes more pronounced with increasing  $N_e$ .

**Variation in diversity between autosomes and sex chromosomes.** An obvious heterogeneity in  $N_e$  within genomes is that expected between autosomes (A) and sex chromosomes (X and Y). Specifically, in a simple scenario

of a randomly mating diploid organism with X and Y (or Z and W) sex chromosomes and an even sex ratio, the  $N_e$  of A, X and Y should scale 4:3:1. However, deviations from the '4:3:1 rule' are likely to occur for several reasons. For example,  $N_e$  of the two sexes might differ, either because more individuals of one sex than of the other contribute to reproduction (that is, polygyny or polyandry) or because there are sex differences in migration. This has been assessed in several recent studies comparing the ratio of X-linked (or Z-linked) and autosomal diversity in humans<sup>122–124</sup> and other organisms<sup>125,126</sup>, with an X:A ratio of 0.75 forming a null hypothesis. When the variance in reproductive success among males is higher than among females, as is often the case with sexual selection, organisms with male heterogamety should show  $X:A > 0.75$  and organisms with female heterogamety should show  $Z:A < 0.75$ , which is generally observed<sup>126,127</sup>. A comparison of polygamous and monogamous bird species found lower Z:A ratios among polygamous species, which is consistent with sexual selection<sup>128</sup>, whereas another study found higher Z:A ratios among sexually dichromatic species than among sexually monochromatic species, which is seemingly inconsistent with the theory if dichromatism were a signature of sexual selection<sup>129</sup>.

Dominance and recombination are two other aspects of relevance when considering diversity of autosomes and sex chromosomes. The fact that the X (or Z) chromosome is hemizygous in the heterogametic sex means that recessive adaptive mutations are directly exposed to selection in that sex, making selection more efficient for X-linked loci and contributing to diversity reduction. The effect is reinforced if the sex-average rate of recombination on the X (or Z) chromosome is lower than on autosomes, boosting linked selection, given that full recombination only occurs in the homogametic sex. This probably explains why the signature of reduced genetic diversity near human genes is more pronounced on the X chromosome than on autosomes<sup>122</sup>. Interestingly, reduced diversity of sex chromosomes comes with one important exception. In organisms with highly differentiated sex chromosomes, recombination is often restricted to a small pseudoautosomal region (PAR) in which an obligate crossing over occurs between homologous sequences of the X and Y chromosomes (and of the Z and W chromosomes). When recombination is concentrated to a PAR that is only a few megabases or less in size, extremely high recombination rates (>50 cM per Mb) can be attained in such regions<sup>130,131</sup> and promote the maintenance of high diversity levels<sup>130,132</sup>.

The sex-limited Y and W chromosomes form a special case in many evolutionary genetic respects by suffering from the deleterious consequences of absence of recombination<sup>133–135</sup>. Relevant here is the sensitivity to selection of non-recombining chromosomes, which causes a reduction in diversity (FIG. 4). Studies of both animal<sup>136–138</sup> and plant<sup>139,140</sup> Y chromosomes have revealed much lower diversity than that given by the 4:1 relationship in  $N_e$  between autosomes due to the influence of the mating system. One potential contributing factor to low Y-linked diversity is sexual selection, as the Y chromosome often contains genes involved in male

#### Allele frequency spectrum

The distribution of the frequency of variants across biallelic loci in a population sample.

#### Polygyny

A mating system in which males mate with more than one female.

#### Polyandry

A mating system in which females mate with more than one male.

#### Heterogamety

When an organism of a particular sex carries two different types of sex chromosomes: that is, males of many animals and plants and females of birds, some fish and lizards, butterflies, and others.

#### Hemizygous

The situation when there is only one chromosome copy in an individual of a diploid species, as for the X chromosome in males of many species.

reproduction<sup>135</sup>. However, a recent study that quantified nucleotide diversity of flycatcher W chromosomes found the diversity to be approximately tenfold lower than expected from the rate of mutation (see below) and the mating system<sup>141</sup>, which is similar to the situation for the human Y chromosome<sup>142</sup>.

Finally, autosomes and sex chromosomes are likely to differ in the rate at which new mutations occur, affecting diversity (BOX 3). In many organisms, the number of mitotic cell divisions is higher in the male than in the female germ line; consequently, the number of sperm produced typically exceeds the number of egg cells produced. Assuming that the per-cell generation rate of mutation is similar in spermatogenesis and oogenesis, this should generate more replication-associated mutations in males than in females<sup>143</sup>. The most distinct effect of such male-biased mutations is elevated mutation rates of the Y chromosome in organisms with male heterogamety (counteracting diversity reduction) and reduced rates of the W chromosome in organisms with female heterogamety (reinforcing diversity reduction). This will contribute to differences in diversity between sex (and sex-limited in particular) chromosomes and autosomes. Among the species so far investigated, male-biased mutation is most pronounced in humans and chimpanzees, with four- to eightfold higher mutation rates in males than in females<sup>144,145</sup>.

## Conclusions

Lewontin's observation that genetic diversity varies among species much less than  $N_e$  contradicts the prediction of the neutral theory, which states that polymorphism should be proportional to population size. Natural selection and its effects through linkage were therefore invoked as a potential explanation for the paradox, but quantitative assessment of this hypothesis has long been impeded by the scarcity of population genomic data sets. We now have an answer as far as animals and plants are concerned: selection at linked sites does affect levels of genome-average diversity but to an extent that is probably insufficient to explain Lewontin's paradox by itself<sup>85,86</sup>. Another classical explanation for the narrow range of diversity across species invokes demographic fluctuations, which are expected to level out the long-term average  $N_e$ . The report in animals of a major effect of reproductive strategy of species on the genome-average heterozygosity<sup>37</sup> is indeed suggestive of a connection between population dynamic processes and the long-term  $N_e$  — even though the underlying ecological mechanisms remain to be clarified. These explanations account for Lewontin's paradox by invoking a departure from the assumption of mutation–drift equilibrium, through changes in either global (demographic

fluctuations) or local (linked selection)  $N_e$ . Besides, other factors not directly related to  $N_e$  deserve to be considered. As mentioned above, the per-generation mutation rate is negatively correlated to  $N_e$  across all domains of life. If confirmed at more restricted taxonomic scales, this effect could indeed partly explain the narrow range of observed genetic diversity among species. Finally, there might be some sort of physical limit to the genetic diversity a species can carry, due to constraints on the level of dissimilarity between two recombining DNA sequences<sup>146</sup>. It is noteworthy that recombination in yeast is impaired when heterozygosity exceeds a few per cent<sup>147,148</sup>, which is close to the maximal diversity observed in natural populations of animals<sup>37,62</sup>.

Conceptually, the forces that influence  $N_e$  are to a large extent similar on global (among species) and local (within genomes) scales. Just as variation in  $N_e$  among species is associated with variation in genetic diversity, it also causes variation in genetic diversity within genomes. There is ample evidence that the density of targets of selection and the rate of recombination determines to what extent linked selection affects neutral diversity. Both positive selection (manifested in selective sweeps) and purifying selection (through the process of background selection) reduce diversity at linked loci, which can be separated by large physical distances when the rate of recombination is low. The most pronounced effect is seen in non-recombining chromosomes: for example, the sex-limited Y and W chromosomes.

We are closer than ever to elucidating Lewontin's 40-year-old riddle, which is probably explained by a combination of demographic fluctuations, genetic hitch-hiking, mutation rate variation and molecular constraints on heterozygosity. The accumulation of genome-wide data from natural populations in a large number of species will soon, we believe, offer a quantification of the relative influence of these factors, which might differ between taxa. Further research should focus on accurate modelling and prediction of across-genome variation in selective pressure, both adaptive and purifying, and to move from large-scale to taxon-specific comparisons, thus hopefully clarifying the ecological determinants underlying the relationship between life-history traits and diversity. This would be an achievement of importance, owing to the many aspects of genome evolution involved in the determination of genetic diversity and the implications in terms of conservation and management policies. Assessment of the long-term sustainability and adaptive potential of endangered populations will indeed benefit from an understanding of the factors causing variation in genetic diversity within genomes as well as among species.

- Lewontin, R. C. & Hubby, J. L. A molecular approach to the study of genic heterozygosity in natural populations. II. Amount of variation and degree of heterozygosity in natural populations of *Drosophila pseudoobscura*. *Genetics* **54**, 595–609 (1966).
- Harris, H. Enzyme polymorphisms in man. *Proc. R. Soc. Lond. B* **164**, 298–310 (1966).
- Quintana-Murci, L. & Clark, A. G. Population genetic tools for dissecting innate immunity in humans. *Nat. Rev. Immunol.* **13**, 280–293 (2013).
- Bodmer, W. Genetic characterization of human populations: from ABO to a genetic map of the British people. *Genetics* **199**, 267–279 (2015).
- Hake, S. & Ross-Ibarra, J. Genetic, evolutionary and plant breeding insights from the domestication of maize. *eLife* **4**, e05861 (2015).
- Soares, M. P. & Weiss, G. The Iron Age of host–microbe interactions. *EMBO Rep.* **16**, 1482–1500 (2015).
- Vander Wal, E., Garant, D., Festa-Bianchet, M. & Pelletier, F. Evolutionary rescue in vertebrates: evidence, applications and uncertainty. *Phil. Trans. R. Soc. B* **368**, 20120090 (2012).
- Forcada, J. & Hoffman, J. I. Climate change selects for heterozygosity in a declining fur seal population. *Nature* **511**, 462–465 (2014).
- Begun, D. J. *et al.* Population genomics: whole-genome analysis of polymorphism and divergence in *Drosophila simulans*. *PLoS Biol.* **5**, e310 (2007).

10. Lack, J. B. *et al.* The *Drosophila* genome nexus: a population genomic resource of 623 *Drosophila melanogaster* genomes, including 197 from a single ancestral range population. *Genetics* **199**, 1229–1241 (2015).
11. McVean, G., Spencer, C. C. A. & Chaix, R. Perspectives on human genetic variation from the HapMap project. *PLoS Genet.* **1**, e54 (2005).
12. The 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
13. Tenaillon, M. I. *et al.* Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* ssp. *mays* L.). *Proc. Natl Acad. Sci. USA* **98**, 9161–9166 (2001).
14. Nordborg, M. *et al.* The pattern of polymorphism in *Arabidopsis thaliana*. *PLoS Biol.* **3**, e196 (2005).
15. Doniger, S. W. *et al.* A catalog of neutral and deleterious polymorphism in yeast. *PLoS Genet.* **4**, e1000183 (2008).
16. Wong, G. K. S. *et al.* A genetic variation map for chicken with 2.8 million single-nucleotide polymorphisms. *Nature* **432**, 717–722 (2004).
17. Sachidanandam, R. *et al.* A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* **409**, 928–933 (2001).
18. Hodgkinson, A. & Eyre-Walker, A. Variation in the mutation rate across mammalian genomes. *Nat. Rev. Genet.* **12**, 756–766 (2011).
19. Lynch, M. Evolution of the mutation rate. *Trends Genet.* **26**, 345–352 (2010).
20. Charlesworth, B. Effective population size and patterns of molecular evolution and variation. *Nat. Rev. Genet.* **10**, 195–205 (2009).
21. Kimura, M. *The Neutral Theory of Molecular Evolution* (Cambridge Univ. Press, 1983).
22. Lewontin, R. *The Genetic Basis of Evolutionary Change* (Columbia Univ. Press, 1974). **This book is a remarkably clear and early introduction to the problem of variation in genetic diversity and the first statement of the so-called Lewontin's paradox.**
23. Leffler, E. M. *et al.* Revisiting an old riddle: what determines genetic diversity levels within species? *PLoS Biol.* **10**, e1001388 (2012). **This article contains a thorough review of the distribution of DNA sequence diversity across hundreds of eukaryotic species.**
24. Reed, D. H. & Frankham, R. Correlation between fitness and genetic diversity. *Conserv. Biol.* **17**, 230–237 (2003).
25. Reed, D. H. & Frankham, R. How closely correlated are molecular and quantitative measures of genetic variation? A meta-analysis. *Evolution* **55**, 1095–1103 (2001).
26. Bjørnstad, O. N. & Grenfell, B. T. Noisy clockwork: time series analysis of population fluctuations in animals. *Science* **293**, 638–643 (2001).
27. Sun, J., Cornelius, S. P., Janssen, J., Gray, K. A. & Motter, A. E. Regularity underlies erratic population abundances in marine ecosystems. *J. R. Soc. Interface* **12**, 20150235 (2015).
28. Banks, S. C. *et al.* How does ecological disturbance influence genetic diversity? *Trends Ecol. Evol.* **28**, 670–679 (2013).
29. Alcalá, N. & Vuilleumier, S. Turnover and accumulation of genetic diversity across large time-scale cycles of isolation and connection of populations. *Proc. R. Soc. B* **281**, 20141369 (2014).
30. Mayr, E. *Animal Species and Evolution* (Harvard Univ. Press, 1963).
31. Hewitt, G. The genetic legacy of the Quaternary ice ages. *Nature* **405**, 907–913 (2000).
32. Stuessy, T. F., Takayama, K., López-Sepúlveda, P. & Crawford, D. J. Interpretation of patterns of genetic variation in endemic plant species of oceanic islands. *Bot. J. Linnean Soc.* **174**, 276–288 (2014).
33. Aguilar, R., Quesada, M., Ashworth, L., Herrerías-Diego, Y. & Lobo, J. Genetic consequences of habitat fragmentation in plant populations: susceptible signals in plant traits and methodological approaches. *Mol. Ecol.* **17**, 5177–5188 (2008).
34. Caplins, S. A. *et al.* Landscape structure and the genetic effects of a population collapse. *Proc. R. Soc. B* **281**, 20141798 (2014).
35. Coltman, D. W. Molecular ecological approaches to studying the evolutionary impact of selective harvesting in wildlife. *Mol. Ecol.* **17**, 221–235 (2008).
36. Lynch, M. *The Origins of Genome Architecture* (Sinauer Associates, 2007).
37. Romiguier, J. *et al.* Comparative population genomics in animals uncovers the determinants of genetic diversity. *Nature* **515**, 261–263 (2014). **This study shows a comparative analysis of patterns of diversity across animals revealing a strong influence of the life-history traits of species.**
38. Sung, W., Ackerman, M. S., Miller, S. F., Doak, T. G. & Lynch, M. Drift-barrier hypothesis and mutation-rate evolution. *Proc. Natl Acad. Sci. USA* **109**, 18488–18492 (2012).
39. Ness, R. W., Morgan, A. D., Vasanthakrishnan, R. B., Colegrave, N. & Keightley, P. D. Extensive *de novo* mutation rate variation between individuals and across the genome of *Chlamydomonas reinhardtii*. *Genome Res.* **25**, 1739–1749 (2015).
40. Wright, S. Size of population and breeding structure in relation to evolution. *Science* **87**, 430–431 (1938).
41. Weber, D., Stewart, B. S., Garza, J. C. & Lehman, N. An empirical genetic assessment of the severity of the northern elephant seal population bottleneck. *Curr. Biol.* **10**, 1287–1290 (2000).
42. Hedrick, P. W. Conservation genetics and North American bison (*Bison bison*). *J. Hered.* **100**, 411–420 (2009).
43. Spielman, D., Brook, B. W. & Frankham, R. Most species are not driven to extinction before genetic factors impact them. *Proc. Natl Acad. Sci. USA* **101**, 15261–15264 (2004).
44. Nabelholz, B., Mauffrey, J.-F., Bazin, E., Galtier, N. & Glemin, S. Determination of mitochondrial genetic diversity in mammals. *Genetics* **178**, 351–361 (2008).
45. McCusker, M. R. & Bentzen, P. Positive relationships between genetic diversity and abundance in fishes. *Mol. Ecol.* **19**, 4852–4862 (2010).
46. Perry, G. H. *et al.* Comparative RNA sequencing reveals substantial genetic variation in endangered primates. *Genome Res.* **22**, 602–610 (2012).
47. Pinsky, M. L. & Palumbi, S. R. Meta-analysis reveals lower genetic diversity in overfished populations. *Mol. Ecol.* **23**, 29–39 (2014).
48. Ho, S. Y. W. & Shapiro, B. Skyline-plot methods for estimating demographic history from nucleotide sequences. *Mol. Ecol. Resour.* **11**, 423–434 (2011).
49. Drummond, A. J., Rambaut, A., Shapiro, B. & Pybus, O. G. Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol. Biol. Evol.* **22**, 1185–1192 (2005).
50. Li, H. & Durbin, R. Inference of human population history from individual whole-genome sequences. *Nature* **475**, 493–496 (2011).
51. Liu, X. & Fu, Y.-X. Exploring population size changes using SNP frequency spectra. *Nat. Genet.* **47**, 555–559 (2015).
52. Schiffels, S. & Durbin, R. Inferring human population size and separation history from multiple genome sequences. *Nat. Genet.* **46**, 919–925 (2014).
53. Nadachowska-Brzyska, K., Li, C., Smeds, L., Zhang, G. & Ellegren, H. Temporal dynamics of avian populations during Pleistocene revealed by whole-genome sequences. *Curr. Biol.* **25**, 1375–1380 (2015).
54. Jarne, P. Mating system, bottlenecks and genetic polymorphism in hermaphroditic animals. *Genet. Res.* **65**, 193–207 (1995).
55. Charlesworth, D. & Wright, S. Breeding systems and genome evolution. *Curr. Opin. Genet. Dev.* **11**, 685–690 (2001).
56. Glémin, S., Bazin, E. & Charlesworth, D. Impact of mating systems on patterns of sequence polymorphism in flowering plants. *Proc. R. Soc. B* **273**, 3011–3019 (2006).
57. Glémin, S. & Muyle, A. Mating systems and selection efficacy: a test using chloroplast sequence data in angiosperms. *J. Evol. Biol.* **27**, 1386–1399 (2014).
58. Hartfield, M. Evolutionary genetic consequences of facultative sex and outcrossing. *J. Evol. Biol.* **29**, 5–22 (2016). **This review discusses the theoretical predictions and empirical evidence regarding genome evolution in asexual versus sexual contexts.**
59. Slotte, T. *et al.* The *Capsella rubella* genome and the genomic consequences of rapid mating system evolution. *Nat. Genet.* **45**, 831–835 (2013).
60. Burgarella, C. *et al.* Molecular evolution of freshwater snails with contrasting mating systems. *Mol. Biol. Evol.* **32**, 2403–2416 (2015).
61. Thomas, C. G. *et al.* Full-genome evolutionary histories of selfing, splitting, and selection in *Caenorhabditis*. *Genome Res.* **25**, 667–678 (2015).
62. Dey, A., Chan, C. K. W., Thomas, C. G. & Cutter, A. D. Molecular hyperdiversity defines populations of the nematode *Caenorhabditis brenneri*. *Proc. Natl Acad. Sci. USA* **110**, 11056–11060 (2013).
63. Dolgin, E. S., Charlesworth, B. & Cutter, A. D. Population frequencies of transposable elements in selfing and outcrossing *Caenorhabditis* nematodes. *Genet. Res.* **90**, 317–329 (2008).
64. Wright, S. I., Kalisz, S. & Slotte, T. Evolutionary consequences of self-fertilization in plants. *Proc. R. Soc. B* **280**, 20130133 (2013).
65. Balloux, F., Lehmann, L. & de Meeus, T. The population genetics of clonal and partially clonal diploids. *Genetics* **164**, 1635–1644 (2003).
66. Mark Welch, D. B. & Meselson, M. Evidence for the evolution of Bdelloid rotifers without sexual reproduction or genetic exchange. *Science* **288**, 1211–1215 (2000).
67. Delmotte, F. *et al.* Phylogenetic evidence for hybrid origins of asexual lineages in an aphid species. *Evolution* **57**, 1291–1303 (2003).
68. Schaefer, I. *et al.* No evidence for the ‘Meselson effect’ in parthenogenetic oribatid mites (Oribatida, Acari). *J. Evol. Biol.* **19**, 184–193 (2006).
69. Schwander, T., Henry, L. & Crespi, Bernard, J. Molecular evidence for ancient asexuality in *Timema* stick insects. *Curr. Biol.* **21**, 1129–1134 (2011).
70. Hollister, J. D. *et al.* Recurrent loss of sex is associated with accumulation of deleterious mutations in *Oenothera*. *Mol. Biol. Evol.* **32**, 896–905 (2015).
71. Maynard Smith, J. *The Evolution of Sex* (Cambridge Univ. Press, 1978).
72. McDonald, M. J., Rice, D. P. & Desai, M. M. Sex speeds adaptation by altering the dynamics of molecular evolution. *Nature* **531**, 233–236 (2016).
73. Stebbings, G. L. Self fertilization and population variability in the higher plants. *Am. Naturalist* **91**, 41–46 (1957).
74. Judson, O. P. & Normark, B. B. Ancient asexual scandals. *Trends Ecol. Evol.* **11**, 41–46 (1996).
75. Simon, J. C., Delmotte, F., Rispe, C. & Crease, T. Phylogenetic evidence for hybrid origins of asexual lineages in an aphid species. *Evolution* **57**, 1291–1303 (2003).
76. Igic, B. & Busch, J. W. Is self-fertilization an evolutionary dead end? *New Phytol.* **198**, 386–397 (2013).
77. Tajima, F. Relationship between DNA polymorphism and fixation time. *Genetics* **125**, 447–454 (1990).
78. Cutter, A. D. & Payseur, B. A. Genomic signatures of selection at linked sites: unifying the disparity among species. *Nat. Rev. Genet.* **14**, 262–274 (2013).
79. Maynard Smith, J. & Haigh, J. The hitch-hiking effect of a favourable gene. *Genet. Res.* **23**, 23–35 (1974).
80. Kaplan, N. L., Hudson, R. R. & Langley, C. H. The “hitchhiking effect” revisited. *Genetics* **123**, 887–899 (1989).
81. Gillespie, J. H. Genetic drift in an infinite population: the pseudohitchhiking model. *Genetics* **155**, 909–919 (2000).
82. Gillespie, J. H. Is the population size of a species relevant to its evolution? *Evolution* **55**, 2161–2169 (2001). **This paper shows a theoretical examination of the effects of recurrent adaptive substitutions on linked loci and their relationship with  $N_e$ .**
83. Charlesworth, B., Morgan, M. T. & Charlesworth, D. The effect of deleterious mutations on neutral molecular variation. *Genetics* **134**, 1289–1303 (1993). **This study shows a theoretical examination of the effects of recurrent deleterious substitutions on linked loci and the background selection model.**
84. Charlesworth, B. The effect of background selection against deleterious mutations on weakly selected, linked variants. *Genet. Res.* **63**, 213–227 (1994).
85. Corbett-Detig, R. B., Hartl, D. L. & Sackton, T. B. Natural selection constrains neutral diversity across a wide range of species. *PLoS Biol.* **13**, e1002112 (2015). **This article demonstrates the role of linked selection in shaping the within-genome variation in polymorphism and its relationship with  $N_e$ .**
86. Coop, G. Does linked selection explain the narrow range of genetic diversity across species? *bioRxiv* <http://dx.doi.org/10.1101/042598> (2016).
87. Elyashiv, E. *et al.* A genomic map of the effects of linked selection in *Drosophila*. *arXiv* <http://arXiv.org/abs/1408.5461v1> (2014).
88. Comeron, J. M. Background selection as baseline for nucleotide variation across the *Drosophila* genome. *PLoS Genet.* **10**, e1004434 (2014).



89. Enard, D., Messer, P. W. & Petrov, D. A. Genome-wide signals of positive selection in human evolution. *Genome Res.* **24**, 885–895 (2014).
90. Gossmann, T. I., Woolfit, M. & Eyre-Walker, A. Quantifying the variation in the effective population size within a genome. *Genetics* **189**, 1389–1402 (2011).
91. Wu, C.-I. The genic view of the process of speciation. *J. Evol. Biol.* **14**, 851–865 (2001).
92. Begun, D. J. & Aquadro, C. F. Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster*. *Nature* **356**, 519–520 (1992).
93. Nachman, M. W. Single nucleotide polymorphisms and recombination rate in humans. *Trends Genet.* **17**, 481–485 (2001).
94. Lercher, M. J. & Hurst, L. D. Human SNP variability and mutation rate are higher in regions of high recombination. *Trends Genet.* **18**, 337–340 (2002).
95. Dvorak, J., Luo, M. C. & Yang, Z. L. Restriction fragment length polymorphism and divergence in the genomic regions of high and low recombination in self-fertilizing and cross-fertilizing *Aegilops* species. *Genetics* **148**, 423–434 (1998).
96. Stephan, W. & Langley, C. H. DNA polymorphism in *Lycopersicon* and crossing-over per physical length. *Genetics* **150**, 1585–1593 (1998).
97. Cutter, A. D. & Choi, J. Y. Natural selection shapes nucleotide polymorphism across the genome of the nematode *Caenorhabditis briggsae*. *Genome Res.* **20**, 1103–1111 (2010).
98. Fay, J. C. & Wu, C. I. Hitchhiking under positive Darwinian selection. *Genetics* **155**, 1405–1413 (2000).
99. Campos, J. L., Halligan, D. L., Haddrill, P. R. & Charlesworth, B. The relation between recombination rate and patterns of molecular evolution and variation in *Drosophila melanogaster*. *Mol. Biol. Evol.* **31**, 1010–1028 (2014).
100. Messer, P. W. & Petrov, D. A. Frequent adaptation and the McDonald–Kreitman test. *Proc. Natl Acad. Sci. USA* **110**, 8615–8620 (2013).
101. Sella, G., Petrov, D. A., Przeworski, M. & Andolfatto, P. Pervasive natural selection in the *Drosophila* genome? *PLoS Genet.* **5**, e1000495 (2009).
- This article reviews the evidence for a pervasive role of linked selection on patterns of genetic variation in *Drosophila* species.**
102. Slotte, T. The impact of linked selection on plant genomic variation. *Brief. Funct. Genomics* **13**, 268–275 (2014).
103. Lohmueller, K. E. *et al.* Natural selection affects multiple aspects of genetic variation at putatively neutral sites across the human genome. *PLoS Genet.* **7**, e1002326 (2011).
104. Messer, P. W. SLiM: simulating evolution with selection and linkage. *Genetics* **194**, 1037–1039 (2013).
105. Hernandez, R. D. A flexible forward simulator for populations subject to selection and demography. *Bioinformatics* **24**, 2786–2787 (2008).
106. Bank, C., Ewing, G. B., Ferrer-Admetlla, A., Foll, M. & Jensen, J. D. Thinking too positive? Revisiting current methods of population genetic selection inference. *Trends Genet.* **30**, 540–546 (2014).
107. Coop, G. & Ralph, P. Patterns of neutral diversity under general models of selective sweeps. *Genetics* **192**, 205–224 (2012).
108. Bolivar, P., Mugal, C. F., Nater, A. & Ellegren, H. Recombination rate variation modulates gene sequence evolution mainly via GC-biased gene conversion, not Hill–Robertson interference, in an avian system. *Mol. Biol. Evol.* **33**, 216–227 (2016).
109. Payseur, B. A. & Nachman, M. W. Gene density and human nucleotide polymorphism. *Mol. Biol. Evol.* **19**, 336–340 (2002).
110. Charlesworth, B. Background selection and patterns of genetic diversity in *Drosophila melanogaster*. *Genet. Res.* **68**, 131–149 (1996).
111. Hudson, R. R. & Kaplan, N. L. Deleterious background selection with recombination. *Genetics* **141**, 1605–1617 (1995).
112. Nordborg, M., Charlesworth, B. & Charlesworth, D. The effect of recombination on background selection. *Genet. Res.* **67**, 159–174 (1996).
113. Flowers, J. M. *et al.* Natural selection in gene-dense regions shapes the genomic pattern of polymorphism in wild and domesticated rice. *Mol. Biol. Evol.* **29**, 675–687 (2012).
114. Burri, R. *et al.* Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of *Ficedula* flycatchers. *Genome Res.* **25**, 1656–1665 (2015).
- This study is a high-resolution examination of genome-wide patterns of diversity and the role of recombination and linked selection in several species of flycatcher.**
115. Nabholz, B. *et al.* Transcriptome population genomics reveals severe bottleneck and domestication cost in the African rice (*Oryza glaberrima*). *Mol. Ecol.* **23**, 2210–2227 (2014).
116. Hellmann, I., Ebersberger, I., Ptak, S. E., Pääbo, S. & Przeworski, M. A neutral explanation for the correlation of diversity with recombination rates in humans. *Am. J. Hum. Genet.* **72**, 1527–1535 (2003).
117. Yang, S. *et al.* Parent-progeny sequencing indicates higher mutation rates in heterozygotes. *Nature* **523**, 463–467 (2015).
118. Arbeitshuber, B., Betancourt, A. J., Ebner, T. & Tiemann-Boege, I. Crossovers are associated with mutation and biased gene conversion at recombination hotspots. *Proc. Natl Acad. Sci. USA* **112**, 2109–2114 (2015).
119. Rattray, A., Santoyo, G., Shafer, B. & Strathern, J. N. Elevated mutation rate during meiosis in *Saccharomyces cerevisiae*. *PLoS Genet.* **11**, e1004910 (2015).
120. Duret, L. & Galtier, N. Biased gene conversion and the evolution of mammalian genomic landscapes. *Annu. Rev. Genom. Hum. Genet.* **10**, 285–311 (2009).
121. Wallberg, A., Glémin, S. & Webster, M. T. Extreme recombination frequencies shape genome variation and evolution in the honeybee, *Apis mellifera*. *PLoS Genet.* **11**, e1005189 (2015).
122. Hammer, M. F. *et al.* The ratio of human X chromosome to autosome diversity is positively correlated with genetic distance from genes. *Nat. Genet.* **42**, 830–831 (2010).
123. Arbiza, L., Gottipati, S., Siepel, A. & Keinan, A. Contrasting X-linked and autosomal diversity across 14 human populations. *Am. J. Hum. Genet.* **94**, 827–844 (2014).
124. Gottipati, S., Arbiza, L., Siepel, A., Clark, A. G. & Keinan, A. Analyses of X-linked and autosomal genetic variation in population-scale whole genome sequencing. *Nat. Genet.* **43**, 741–743 (2011).
125. Charlesworth, B. The role of background selection in shaping patterns of molecular evolution and variation: evidence from variability on the *Drosophila* X chromosome. *Genetics* **191**, 233–246 (2012).
126. Frankham, R. How closely does genetic diversity in finite populations conform to predictions of neutral theory? Large deficits in regions of low recombination. *Heredity* **108**, 167–178 (2012).
- This paper reviews and demonstrates the reduction in genetic diversity in low-recombining genomic regions, including sex chromosomes, in plants and animals.**
127. Mank, J. E., Vicoso, B., Berlin, S. & Charlesworth, B. Effective population size and the faster-X effect: empirical results and their interpretation. *Evolution* **64**, 663–674 (2010).
128. Cori, A. & Ellegren, H. The genomic signature of sexual selection in the genetic diversity of the sex chromosomes and autosomes. *Evolution* **66**, 2138–2149 (2012).
129. Huang, H. & Rabosky, D. L. Sex-linked genomic variation and its relationship to avian plumage dichromatism and sexual selection. *BMC Evol. Biol.* **15**, 199 (2015).
130. Smeds, L. *et al.* Genomic identification and characterization of the pseudoautosomal region in highly differentiated avian sex chromosomes. *Nat. Commun.* **5**, 5448 (2014).
131. Lien, S., Szyda, J., Schechinger, B., Rappold, G. & Arnheim, N. Evidence for heterogeneity in recombination in the human pseudoautosomal region: high resolution analysis by sperm typing and radiation-hybrid mapping. *Am. J. Hum. Genet.* **66**, 557–566 (2000).
132. Bussell, J. J., Pearson, N. M., Kanda, R., Filatov, D. A. & Lahn, B. T. Human polymorphism and human–chimpanzee divergence in pseudoautosomal region correlate with local recombination rate. *Gene* **368**, 94–100 (2006).
133. Charlesworth, B. & Charlesworth, D. The degeneration of Y chromosomes. *Phil. Trans. R. Soc. Lond. B* **355**, 1563–1572 (2000).
134. Bachrog, D. Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. *Nat. Rev. Genet.* **14**, 113–124 (2013).
135. Mank, J. E. Small but mighty: the evolutionary dynamics of W and Y sex chromosomes. *Chromosome Res.* **20**, 21–33 (2011).
136. Hellborg, L. & Ellegren, H. Low levels of nucleotide diversity in mammalian Y chromosomes. *Mol. Biol. Evol.* **21**, 158–163 (2004).
137. Bachtrog, D., Thornton, K., Clark, A., Andolfatto, P. & Harrison, R. Extensive introgression of mitochondrial DNA relative to nuclear genes in the *Drosophila yakuba* species group. *Evolution* **60**, 292–302 (2006).
138. Shen, P. *et al.* Population genetic implications from sequence variation in four Y chromosome genes. *Proc. Natl Acad. Sci. USA* **97**, 7354–7359 (2000).
139. Qiu, S., Bergero, R., Forrest, A., Kaiser, V. B. & Charlesworth, D. Nucleotide diversity in *Silene latifolia* autosomal and sex-linked genes. *Proc. R. Soc. B* **277**, 3283–3290 (2010).
140. Filatov, D. A., Laporte, V., Vitte, C. & Charlesworth, D. DNA diversity in sex-linked and autosomal genes of the plant species *Silene latifolia* and *Silene dioica*. *Mol. Biol. Evol.* **18**, 1442–1454 (2001).
141. Smeds, L. *et al.* Evolutionary analysis of the female-specific avian W chromosome. *Nat. Commun.* **6**, 7330 (2015).
142. Wilson Sayres, M. A., Lohmueller, K. E. & Nielsen, R. Natural selection reduced diversity on human Y chromosomes. *PLoS Genet.* **10**, e1004064 (2014).
143. Ellegren, H. Characteristics, causes and evolutionary consequences of male-biased mutation. *Proc. R. Soc. B* **274**, 1–10 (2007).
144. Kong, A. *et al.* Fine-scale recombination rate differences between sexes, populations and individuals. *Nature* **467**, 1099–1103 (2010).
145. Venn, O. *et al.* Strong male bias drives germline mutation in chimpanzees. *Science* **344**, 1272–1275 (2014).
146. Cutter, A. D., Jovel, R. & Dey, A. Molecular hyperdiversity and evolution in very large populations. *Mol. Ecol.* **22**, 2074–2095 (2013).
- This article discusses the specificities and challenges associated with very highly polymorphic species, with a focus on *Caenorhabditis nematodes*.**
147. Drouin, G. Characterization of the gene conversions between the multigene family members of the yeast genome. *J. Mol. Evol.* **55**, 14–23 (2002).
148. Borts, R. H. & Haber, J. E. Meiotic recombination in yeast: alteration by multiple heterozygosities. *Science* **237**, 1459–1465 (1987).
149. Dobzhansky, T. *Evolution, Genetics, and Man* (Wiley, 1955).
150. Ohta, T. Slightly deleterious mutant substitutions in evolution. *Nature* **246**, 96–98 (1973).
151. Hubby, J. L. & Lewontin, R. C. A molecular approach to the study of genetic heterozygosity in natural populations. I. The number of alleles at different loci in *Drosophila pseudoobscura*. *Genetics* **54**, 577–594 (1966).
152. Soule, M. in *Molecular Evolution* (ed. Ayala, F.) 60–77 (Sinauer Associates, 1976).
153. Nevo, E., Beiles, A. & Ben-Shlomo, R. in *Evolutionary Dynamics of Genetic Diversity: Proceedings of a Symposium held in Manchester, England, March 29–30, 1983* (ed. Mani, G. S.) (Springer, 1984).
154. Hamrick, J. L. & Godt, M. J. W. Effects of life history traits on genetic diversity in plant species. *Phil. Trans. R. Soc. Lond. B* **351**, 1291–1298 (1996).
155. Cole, C. T. Genetic variation in rare and common plants. *Annu. Rev. Ecol. Evol. Systemat.* **34**, 213–237 (2003).
156. Avise, J. C. *et al.* Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annu. Rev. Ecol. Systemat.* **18**, 489–522 (1987).
157. Bazin, E., Glémin, S. & Galtier, N. Population size does not influence mitochondrial genetic diversity in animals. *Science* **312**, 570–572 (2006).
158. Nabholz, B., Glémin, S. & Galtier, N. The erratic mitochondrial clock: variations of mutation rate, not population size, affect mtDNA diversity across birds and mammals. *BMC Evol. Biol.* **9**, 1–13 (2009).
159. Ballard, J. W. O. & Whitlock, M. C. The incomplete natural history of mitochondria. *Mol. Ecol.* **13**, 729–744 (2004).
160. Berlin, S., Tomaras, D. & Charlesworth, B. Low mitochondrial variability in birds may indicate Hill–Robertson effects on the W chromosome. *Heredity* **99**, 389–396 (2007).



161. Hurst, G. D. D. & Jiggins, F. M. Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: the effects of inherited symbionts. *Proc. R. Soc. B* **272**, 1525–1534 (2005).
162. Galtier, N., Nabholz, B., Glémin, S. & Hurst, G. D. D. Mitochondrial DNA as a marker of molecular diversity: a reappraisal. *Mol. Ecol.* **18**, 4541–4550 (2009).
163. Piganeau, G. & Eyre-Walker, A. Evidence for variation in the effective population size of animal mitochondrial DNA. *PLoS ONE* **4**, e4396 (2009).
164. Jarne, P. & Lagoda, P. J. L. Microsatellites, from molecules to populations and back. *Trends Ecol. Evol.* **11**, 424–429 (1996).
165. Väli, Ü., Einarsson, A., Waits, L. & Ellegren, H. To what extent do microsatellite markers reflect genome-wide genetic diversity in natural populations? *Mol. Ecol.* **17**, 3808–3817 (2008).
166. Fungtammasan, A. *et al.* Accurate typing of short tandem repeats from genome-wide sequencing data and its applications. *Genome Res.* **25**, 736–749 (2015).
167. Ellegren, H. Genome sequencing and population genomics in non-model organisms. *Trends Ecol. Evol.* **29**, 51–63 (2014).
168. Lynch, M. & Conery, J. S. The origins of genome complexity. *Science* **302**, 1401–1404 (2003).
169. Wright, S. Evolution in Mendelian populations. *Genetics* **16**, 97–159 (1931).
170. Luikart, G., Ryman, N., Tallmon, D., Schwartz, M. & Allendorf, F. Estimation of census and effective population sizes: the increasing usefulness of DNA-based approaches. *Conserv. Genet.* **11**, 355–373 (2010).
171. Palstra, F. P. & Fraser, D. J. Effective/census population size ratio estimation: a compendium and appraisal. *Ecol. Evol.* **2**, 2357–2365 (2012).
172. Gilbert, K. J. & Whitlock, M. C. Evaluating methods for estimating local effective population size with and without migration. *Evolution* **69**, 2154–2166 (2015).
173. Browning, S. R. & Browning, B. L. Accurate non-parametric estimation of recent effective population size from segments of identity by descent. *Am. J. Hum. Genet.* **97**, 404–418 (2015).
174. Kirin, M. *et al.* Genomic runs of homozygosity record population history and consanguinity. *PLoS ONE* **5**, e13996 (2010).
175. Palamara, P. F., Lencz, T., Darvasi, A. & Pe'er, I. Length distributions of identity by descent reveal fine-scale demographic history. *Am. J. Hum. Genet.* **91**, 809–822 (2012).

## Acknowledgements

This work was supported by Swedish Research Council grants (2010–5650 and 2013–8271), a European Research Council grant (AdG 249976) and the Knut and Alice Wallenberg Foundation to H.E., and by a European Research Council grant (AdG 232971) and a French National Research Agency grant (ANR-10-BINF-01-01) to N.G. The authors thank N. Bierne, S. Glémin and M. Lascoux for comments on the manuscript.

## Competing interests statement

The authors declare no competing interests.