

# Rates and Fitness Consequences of New Mutations in Humans

Peter D. Keightley<sup>1</sup>

Institute of Evolutionary Biology, School of Biological Sciences, University of Edinburgh,  
Edinburgh EH9 3JT, United Kingdom

**ABSTRACT** The human mutation rate per nucleotide site per generation ( $\mu$ ) can be estimated from data on mutation rates at loci causing Mendelian genetic disease, by comparing putatively neutrally evolving nucleotide sequences between humans and chimpanzees and by comparing the genome sequences of relatives. Direct estimates from genome sequencing of relatives suggest that  $\mu$  is about  $1.1 \times 10^{-8}$ , which is about twofold lower than estimates based on the human–chimp divergence. This implies that an average of  $\sim 70$  new mutations arise in the human diploid genome per generation. Most of these mutations are paternal in origin, but the male:female mutation rate ratio is currently uncertain and might vary even among individuals within a population. On the basis of a method proposed by Kondrashov and Crow, the genome-wide deleterious mutation rate ( $U$ ) can be estimated from the product of the number of nucleotide sites in the genome,  $\mu$ , and the mean selective constraint per site. Although the presence of many weakly selected mutations in human noncoding DNA makes this approach somewhat problematic, estimates are  $U \approx 2.2$  for the whole diploid genome per generation and  $\sim 0.35$  for mutations that change an amino acid of a protein-coding gene. A genome-wide deleterious mutation rate of 2.2 seems higher than humans could tolerate if natural selection is “hard,” but could be tolerated if selection acts on relative fitness differences between individuals or if there is synergistic epistasis. I argue that in the foreseeable future, an accumulation of new deleterious mutations is unlikely to lead to a detectable decline in fitness of human populations.

FOR some time, it has been thought that each newborn human has many tens of new mutations that appeared in its mother’s or father’s germline. This extraordinarily high genomic rate of mutation includes a small fraction of advantageous mutations that have fueled the evolution of our species and that are the basis of ongoing adaptive evolution. The input of new variation brings along with it mutations that cause Mendelian genetic disease, kept at low frequencies by natural selection, and a burden of less harmful mutations that presumably maintain genetic variation in susceptibility to complex diseases. Timofeeff-Ressovsky (1940) and Muller (1950) argued that the rate for mutations with mildly deleterious effects exceeds that for mutations causing visible or lethal phenotypes, and Muller (1950) argued that these mutations cause a substantial load of genetic deaths. Haldane (1937) had shown that the reduction in mean fitness caused by a new deleterious mutation in a large population is largely

independent of its selective value, because strongly deleterious mutations reduce fitness more than mildly deleterious mutations, but have shorter persistence times. These insights have fueled a great deal of interest in the total mutation rate in humans, particularly the genomic deleterious mutation rate. For example, a high genomic rate of deleterious mutations can make it difficult to explain how humans, a species with a relatively low reproductive potential, are able to persist (Crow 1970).

For anthropocentric reasons, we are fundamentally interested in the mutation rate in our own species, and a number of questions are unresolved or still only partially answered. A central parameter is the average number of new genetic variants each of us has that our parents did not possess. How many of these mutations came from our father and our mother, how much does the mutation rate change with parental age, and is there significant variation in the mutation rate in the population as a consequence of other environmental or genetic factors? A more difficult question to answer concerns the frequency of mildly deleterious mutations and the distribution of their fitness effects. Along with the nature of selection on fitness in human populations, these parameters hold the key to understanding how a high genomic rate of

Copyright © 2012 by the Genetics Society of America  
doi: 10.1534/genetics.111.134668

Manuscript received September 13, 2011; accepted for publication November 13, 2011  
This article is dedicated to the memory of James F. Crow, whose writing and lectures inspired me and many others.

<sup>1</sup>Address for correspondence: University of Edinburgh, West Mains Road, Edinburgh EH9 3JT, UK. E-mail: keightley.genetics2011@gmail.com

deleterious mutation can be tolerated without causing an implausibly high rate of genetic death. Finally, if natural selection has been relaxed in current populations, what are the plausible consequences of mutation accumulation in our species?

In this article, I review evidence on all of these issues. Technological developments in molecular genetics have opened up many previously inaccessible questions, especially via DNA sequencing of samples of genes from humans and our nearest relatives, and more recently by whole-genome sequencing. We are now starting to see the results of sequencing of relatives, including parents and their offspring.

## The Mutation Rate in Humans

Three methods have been used for estimating the germline spontaneous mutation rate per nucleotide site ( $\mu$ ) in the nuclear genome of humans (Kondrashov and Kondrashov 2010) and have produced remarkably consistent results.

### Human genetic disease phenotype frequencies

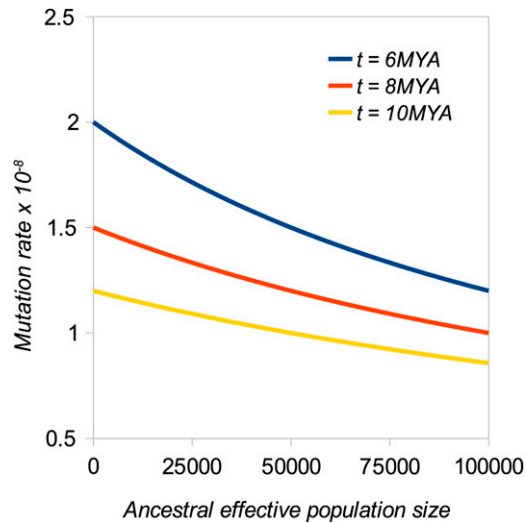
The incidence rate for a genetic disease can be used to estimate the gene mutation rate, under the assumption of mutation–selection balance, if the mean relative fitnesses of the mutant (or heterozygote) and wild type can be estimated (Haldane 1935). For example, the equilibrium frequency of an autosomal partially recessive mutation in an infinite population is approximately  $u/hs$  (Crow and Kimura 1970), where  $hs$  is the fitness disadvantage of the heterozygote and  $u$  is the mutation rate to the mutant allele. An estimate of  $\mu$  can then be obtained by dividing  $u$  by the number of sites in the gene that produce the mutant phenotype, if mutated, which can be estimated from the nucleotide sequence, under certain assumptions. The two most detailed studies have analyzed frequencies of loss-of-function mutations at loci causing human dominant autosomal or X-linked Mendelian disease (Kondrashov 2002; Lynch 2010). Kondrashov (2002) focused on loss-of-function generated by nonsense and indel mutations, arguing that this gives the best estimate of the number of sensitive sites. Lynch (2010) analyzed missense mutations, arguing that focusing on nonsense mutations will overestimate the mutation rate, since termination codons are A + T rich, and there is a mutational bias toward A + T. A number of other factors, however, will cause the mutation rate to be underestimated. These include underdiagnosis of the mutant phenotype caused by incomplete penetrance, the tendency for some truncated proteins to retain residual activity, and failure to detect the causal molecular event. On the other hand, the mutation rate might be overestimated if the most intensively studied loci have atypically high mutation rates and because coding sequences have a higher frequency of hypermutable CpG dinucleotides than the genome as a whole.

Kondrashov (2002) and Lynch (2010) obtained mean estimates for  $\mu$  of  $1.8 \times 10^{-8}$  and  $1.3 \times 10^{-8}$ , respectively, both agreeing fairly well with estimates based on the nucleotide divergence between humans and chimps and with recent estimates based on direct genome sequencing of parent–

offspring trios (see below). Kondrashov's (2002) study suggests that small insertion–deletion events (indels) are comparatively uncommon among human spontaneous mutations, comprising only ~4% of all mutation events, and deletions are about three times more common than insertions. Sequence divergence of processed pseudogenes between humans and chimpanzees suggests that the rate of indel mutation is about one-tenth that of single nucleotide events (Nachman and Crowell 2000). Humans, therefore, seem to be markedly less susceptible to small indels than the invertebrates *Caenorhabditis elegans* (Denver *et al.* 2004) and *Drosophila* (Petrov 2002; Haag-Liautard *et al.* 2007) and the flowering plant *Arabidopsis thaliana* (Ossowski *et al.* 2010).

### Between-species nucleotide divergence at putatively neutral sites

In a sequence evolving free from natural selection, the rate of accumulation of nucleotide substitutions is proportional to  $\mu$  (Wright 1938), and the expected divergence is  $k = 2\mu t/g$ , where  $t$  is the time since speciation and  $g$  is the generation time. Additionally, if the time since the speciation event is relatively short, polymorphism in the ancestral population contributes significantly to divergence; the frequency of observing a difference at an autosomal locus if one allele is randomly sampled from each species is the nucleotide diversity  $\theta = 4N_e\mu$ , where  $N_e$  is the ancestral effective population size. Uncertainty about four parameters,  $k$ ,  $t$ ,  $g$ , and  $N_e$ , therefore contributes to uncertainty affecting estimates of the mutation rate. By sequencing human- and chimp-processed pseudogene orthologs, which are among the best candidates for loci that evolve strictly neutrally, Nachman and Crowell (2000) carried out the most rigorous analysis to estimate the mutation rate by these means. They observed a mean autosomal single nucleotide divergence per site (*i.e.*, excluding indels) of 1.2%. If we assume values for  $g$  and  $t$  of 20 years and 5 MYA, respectively, and  $N_e = 10^4$ , on the basis of current nucleotide diversity in humans (Li and Sadler 1991), the autosomal single nucleotide mutation rate in hominids is therefore  $2.2 \times 10^{-8}$ . Two factors suggest, however, that this might be an overestimate. First, although controversial, new fossil finds have pushed back the human–chimp divergence date (Wood 2002; Benton and Donoghue 2007). For example, Benton and Donoghue (2007) suggest lower and upper limits of 6.5 and 10 MYA, respectively. However, speciation may not be an instantaneous event, and the time at which gene flow ceases may vary across the genome, leading to variation in lineage sorting among sites (Patterson *et al.* 2006). The importance of this effect concerning speciation in hominids has been debated (Barton 2006; Wakeley 2008; Presgraves and Yi 2009). Second, evidence from the pattern of substitutions among hominid lineages suggests that the effective population size of the human–chimp common ancestor (HCCA) could have been much larger than the recent effective sizes for humans or individual chimpanzee subspecies (Takahata *et al.* 1997; Chen and Li 2001; Burgess and Yang 2008). Information comes mainly from the variance of neutral



**Figure 1** Human mutation rate estimate plotted against HCCA effective population size, assuming that the neutral site divergence  $k = 0.012$ , the speciation date ( $t$ ) is varied between  $6 \times 10^6$  and  $10 \times 10^6$  MYA, and the generation interval  $g = 20$  years.

allele coalescence times, which is proportional to the ancestral effective population size. This leads to variation in divergence among loci, leading in turn to incongruence between the species tree and gene tree. Takahata *et al.* (1997) compared the frequency distribution of the number of substitutions in human, chimp, and gorilla sequences with its theoretical expectation and concluded that the HCCA  $N_e$  was  $\sim 10$  times higher than recent human  $N_e$ . Chen and Li (2001) examined incongruence between the species and gene trees in 53 hominid orthologous genomic segments, and observed that 22 segments showed evidence of incomplete lineage sorting. On the basis of Hudson's (1983) formula for the probability of incongruence between the gene and species trees, they obtained an estimate of the HCCA  $N_e$  of 52,000–96,000, depending on the generation time assumed. Burgess and Yang (2008) attempted to include the effects of variation in the mutation rate across the genome and sequencing errors (which have been a serious problem for the initial release of the chimp genome) and also concluded that HCCA  $N_e$  is 5- to 10-fold higher than recent human  $N_e$ . The effect of changing the ancestral population size on the estimate of  $\mu$  is shown in Figure 1. Assuming  $N_e = 10,000$  has little effect compared to ignoring ancestral polymorphism, whereas assuming  $N_e = 100,000$  reduces the estimated mutation rate by about one-third, yielding an estimate of  $1 \times 10^{-8}$ , if a divergence date of  $8 \times 10^6$  MYA is assumed. It is difficult to provide a confidence interval for this estimate, because the uncertainty affecting the divergence date and generation time cannot easily be quantified.

### Genome sequencing of relatives

High-throughput sequencing of related individuals, particularly one or more offspring and their two parents, offers the opportunity to directly estimate the mutation rate in current human populations, thereby avoiding most of the assumptions associated with approaches 1 and 2. This method has so far

been attempted four times (Table 1). Xue *et al.* (2009) employed Illumina sequencing of flow-sorted Y chromosomes of two individuals separated by 13 generations and detected four mutations, which were then verified by Sanger sequencing. This gives an estimate of  $\mu = 3.0 \times 10^{-8}$  per generation for males, which has wide confidence limits and is expected to be higher than the sex-averaged rate for autosomes (see below). Roach *et al.* (2010) employed sequencing array methodology to sequence two offspring and their parents at a mean depth of coverage of  $>50\times$ . Even at this high sequencing depth, there were many false positives (outnumbering confirmed mutations by more than 1000 to 1), which had to be excluded by further sequencing. There were also many false negatives, whose rate had to be estimated separately and a correction applied. The 28 confirmed mutations led to a mutation rate estimate of  $1.1 \times 10^{-8}$ . Awadalla *et al.* (2010) sequenced 294 Mb of putatively nonfunctional DNA as a control in a study of new mutations producing human genetic disease and found four new mutations, giving an estimate for  $\mu$  of  $1.4 \times 10^{-8}$ . The 1000 Genomes Project Consortium (2010) employed Illumina technology to sequence one offspring and two parents each from two trios (one European and one Yoruba from West Africa) at a mean depth of coverage of  $42\times$ . They employed a coarse filter to identify candidate *de novo* mutations, resequenced these, and in the case of the European trio, checked for transmission of confirmed mutations in grand offspring. There were 49 and 35 confirmed mutations in the European and Yoruba trios, respectively. After correcting for false negatives (which were relatively infrequent; Conrad *et al.* 2011), these numbers yielded estimates for  $\mu$  of  $1.2 \times 10^{-8}$  and  $1.0 \times 10^{-8}$  in the European and Yoruba trios, respectively. Although these estimates are very close to one another, further analysis of the parental origin of the mutations (Conrad *et al.* 2011) suggests that there could be substantial mutation rate variation among individuals. Excluding the Y chromosome study, the weighted average mutation rate is  $\mu = 1.1 \times 10^{-8}$  (Table 1), which is close to an estimate from the chimp–human nucleotide divergence, under the assumption of a larger ancestral population size ( $10^5$ ) and a more ancient speciation date than have typically been assumed (Figure 1). The findings in Table 1 should be viewed as work in progress, however, since current high-throughput sequencing is relatively error prone, produces many false positives, and potentially fails to detect some mutations; but this should improve in the future.

One implication of a lower estimate of the mutation rate is that the recent human effective population size may have been about twice as high as previous estimates (e.g., Li and Sadler 1991). Silent site variation varies among human populations, but taking a figure of 0.1% (Cargill *et al.* 1999), and equating autosomal silent nucleotide diversity to  $4N_e\mu$ , yields an estimate for  $N_e$  of 23,000.

### Sex Differences in the Human Mutation Rate

At loci causing human genetic disease, mutations originating in males greatly outnumber mutations originating in females

**Table 1** Estimates of the human single nucleotide mutation rate ( $\mu$ ) obtained by large-scale sequencing of relatives

Reference	Sequencing method	No. mutations detected	Mutation rate, $\mu$ ( $\times 10^{-8}$ )
Xue <i>et al.</i> (2009)	Illumina	4	3.0 (Chr Y)
Roach <i>et al.</i> (2010)	Array based	28	1.1
Awadalla <i>et al.</i> (2010)	Sanger	4	1.4
1000 Genomes Project Consortium (2010)	Illumina	49 (European)	1.2
1000 Genomes Project Consortium (2010)	Illumina	35 (Yoruba)	1.0
Weighted average			1.1 (excluding Xue <i>et al.</i> 2009)

(Crow 1997, 2000). This is believed to be a consequence of a higher number of cell divisions in male gametogenesis. The ratio of the male to female mutation rate,  $\alpha$ , can be estimated by comparing evolutionary divergence between the sex chromosomes and between the sex chromosomes and the autosomes (Miyata *et al.* 1987). In mammals, autosomes spend equal amounts of time in males and females, the X chromosome spends two-thirds of its time in females, and the Y chromosome spends all of its time in males. Comparisons of sex chromosome and autosomal sequences of primates typically suggest values for  $\alpha$  in the range 3–6 (*e.g.*, Huang *et al.* 1997; Ebersberger *et al.* 2002; Makova and Li 2002; Chimpanzee Sequencing and Analysis Consortium 2005; Rhesus Macaque Genome Sequencing and Analysis Consortium 2007). Estimates of  $\alpha$  from closely related species such as humans and chimps could be sensitive to ancestral polymorphism (Ebersberger *et al.* 2002; Makova and Li 2002; Chimpanzee Sequencing and Analysis Consortium 2005; Taylor *et al.* 2006). Consider a comparison of autosomal ( $k_A$ ) and X chromosome ( $k_X$ ) nucleotide divergences. For neutrally evolving sequences, these are expected to be

$$k_A = 2 t \mu_A / g + 4 N_e \mu_A, \quad (1)$$

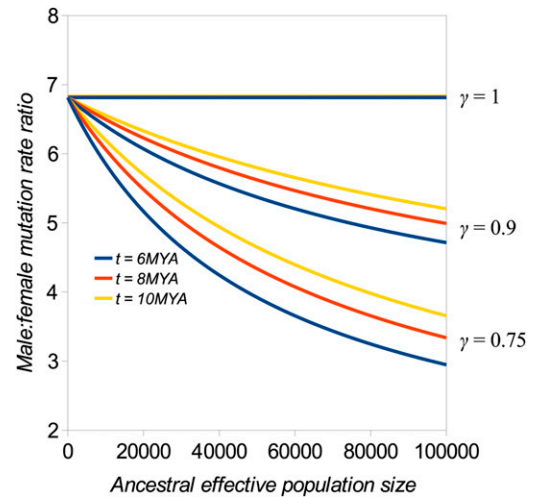
$$k_X = 2 t \mu_X / g + 4 \gamma N_e \mu_X, \quad (2)$$

where  $t$  is the time since speciation,  $g$  is the generation interval,  $\mu_A$  and  $\mu_X$  are sex-averaged mutation rates for the autosomes and the X chromosome, respectively,  $N_e$  is the ancestral effective population size for the autosomes, and  $\gamma$  is the ratio of X chromosome to autosomal effective population sizes in the ancestral population. The terms in  $N_e$  in (1) and (2) represent the contributions to nucleotide divergence from nucleotide diversity in the ancestral population, which might have been substantially higher than predicted by present-day diversity in humans (*e.g.*, Burgess and Yang 2008). Assuming that  $\mu_A = (\mu_M + \mu_F)/2$ , where  $\mu_M$  and  $\mu_F$  are male and female mutation rates, respectively, and that  $\mu_X = (\mu_M + 2 \mu_F)/3$ , we obtain the following estimator for  $\alpha$ ,

$$\alpha = \frac{\mu_M}{\mu_F} = \frac{4(\zeta + \gamma) - 3(K_X/K_A)(\zeta + 1)}{3(K_X/K_A)(\zeta + 1) - 2(\zeta + \gamma)}, \quad (3)$$

where  $\zeta = t/(2N_e g)$ . Figure 2 plots  $\alpha$  calculated from Equation 3 against ancestral  $N_e$  for three values of  $\gamma$ , assuming autosomal and X chromosome divergences of 1.25% and 0.94%, respectively, which are genome-wide average

estimates from the human–chimp genome comparison (Chimpanzee Sequencing and Analysis Consortium 2005). Estimates are therefore quite sensitive to the values assumed for ancestral  $N_e$  and  $\gamma$ . Under a neutral equilibrium model,  $\gamma = 3/4$ , but several factors, including a difference in reproductive success between males and females and differential effects of selective sweeps and background selection, are expected to cause departure from this expectation (Vicoso and Charlesworth 2009). In present-day human populations of European descent, for example, the genome-wide average for  $\gamma = 0.71$ , but  $\gamma$  exceeds 1 in regions distant from protein-coding genes, suggesting an influence of some or all of these processes (Hammer *et al.* 2010). Differences between assumed values of ancestral  $N_e$  and ancestral  $\gamma$  and variation in the contribution to divergence from ancestral polymorphism among sites in the genome could partly explain variation among estimates of  $\alpha$  based on comparing humans with chimps (Ebersberger *et al.* 2002; Chimpanzee Sequencing and Analysis Consortium 2005). The mutation rate for hypermutable CpG dinucleotides seems to depend on absolute time rather than the number of cell divisions, leading to CpGs showing substantially weaker male mutation bias than non-CpG sites (Taylor *et al.* 2006). This may explain the higher relative rate of



**Figure 2** The male:female mutation rate ratio ( $\alpha$ ) calculated by Equation 3 plotted against ancestral effective population size, assuming autosomal and X chromosome nucleotide divergences mentioned in the text, a generation interval ( $g$ ) of 20 years, for three speciation dates ( $t$ ), and three values of the ratio of X:autosomal effective population sizes in the ancestral population ( $\gamma$ ).

mutation at CpG sites compared to non-CpG sites in primates than murids and carnivores (Keightley *et al.* 2011).

A more direct estimate of  $\alpha = 6.5$  was obtained by Lynch (2010), on the basis of frequencies of human genetic disease originating from new parental and maternal mutations. This estimate is therefore at the upper end of the range obtained from nucleotide divergence between human and other apes (Huang *et al.* 1997; Ebersberger *et al.* 2002; Makova and Li 2002; Chimpanzee Sequencing and Analysis Consortium 2005). However, it is higher than an estimate obtained from a comparison of the human and macaque genomes ( $\alpha = 2.9$ ; Rhesus Macaque Genome Sequencing and Analysis Consortium 2007), which should not be substantially affected by ancestral polymorphism.

The advent of accurate, whole-genome sequencing of trios of offspring and their parents promises to resolve outstanding questions concerning the mutation rate sex bias, at least with reference to current human populations. It also promises to shed light on the causes of nonlinearity of the male mutation rate with age at some loci (Crow 2000, 2006; Goriely *et al.* 2003; Choi *et al.* 2008). Recently, the first direct information on the male:female mutation rate ratio has been published (Conrad *et al.* 2011). The results are surprising, since in one trio of European ancestry, 92% of mutations were paternal ( $\alpha = 12$ ), whereas in a second trio (Yoruba), only 36% of mutations were paternal ( $\alpha = 0.6$ ). There was significant heterogeneity in the maternal and paternal mutation rates between these trios, which could be caused by differences in paternal and/or maternal ages at conception (which are unknown for the individuals in question). The mutation rate differences may also reflect environmental and/or genetic variation in the mutation rate that is not associated with age of reproduction.

### The Human Genomic Deleterious Mutation Rate

The mean number of single-nucleotide mutation events per diploid human genome per generation ( $M$ ) is the product of the mutation rate per site and twice the number of base pairs in the genome, *i.e.*,  $M \approx 1.1 \times 10^{-8} \times 6 \times 10^9 \approx 70$ . How many of these mutations are subject to sufficiently strong natural selection as to overcome genetic drift (*i.e.*,  $|2N_e s| > 1$ )? The number of nonneutral mutations, and the distribution of their fitness effects, particularly those that are deleterious, is important for knowing the burden of “genetic deaths” sustained by the human population. Kondrashov and Crow (1993) proposed a method for estimating the genome-wide deleterious mutation rate ( $U$ ) for a model in which mutations are assumed to be either strongly deleterious ( $|2N_e s| \gg 1$ ) or selectively neutral. The between-species nucleotide divergence for sites at which new mutations are assumed to be exclusively neutral,  $k_S$  (for example, pseudogenes or transposable element remnants), is compared with the divergence for the genome as a whole,  $k_G$ . The genomic deleterious mutation rate is estimated from the number of “missing” substitutions in the genome as a whole, *i.e.*,

$$U = M(1 - k_G/k_S). \quad (4)$$

Consider the rate for deleterious mutations caused by amino acid change in protein-coding genes ( $U_p$ ). The human gene count is believed to be 20,000–25,000 (International Human Genome Sequencing Consortium 2004), the average length of concatenated coding exons per gene is  $\sim 1340$  bases (International Human Genome Sequencing Consortium 2001), about three-quarters of exonic mutations lead to an amino acid change, and  $\sim 70\%$  of these are strongly selected against (Eöry *et al.* 2010). The rate of deleterious amino acid mutation per diploid genome is therefore  $U_p \approx 2 \times 1.1 \times 10^{-8} \times 22,500 \times 1,340 \times 0.75 \times 0.7 \approx 0.35$  (assuming the median of the range for gene number and disregarding a contribution of indels). This is 2–10 times lower than previous estimates (Eyre-Walker and Keightley 1999; Nachman and Crowell 2000; Eöry *et al.* 2010), because the per-nucleotide mutation rate seems to be about twofold lower (see above), and gene number is severalfold lower than estimates made prior to the sequencing of the human genome.

Estimating  $U$  for the complete human genome is more problematic, because the assumption that new mutations in noncoding DNA are either neutral or strongly deleterious is almost certainly seriously violated. For example, many new mutations in regions of the mammalian genome regulating gene expression seem to be weakly selected (Eyre-Walker and Keightley 2009; Torgerson *et al.* 2009; Kousathanas *et al.* 2011) and can therefore drift to higher frequencies. Applying Kondrashov and Crow’s (1993) method will therefore underestimate the number of such mutations. By comparing the human and mouse genomes, the Mouse Genome Sequencing Consortium (2002) estimated that  $\sim 5\%$  of the mammalian genome is under selection, but this is subject to some uncertainty because mouse and human noncoding sequences are difficult to align accurately. Undeterred by the difficulty of inferring  $U$  in the presence of weakly selected mutations in noncoding DNA, we have attempted to apply Kondrashov and Crow’s (1993) approach to the human–chimpanzee genome comparison, using transposable element remnants as a neutral reference to estimate  $k_S$  (Eöry *et al.* 2010). We estimated that the nucleotide divergence for the fraction of the genome that excludes transposable elements,  $k_G^* \approx 0.95k_S$ . Assuming that the single nucleotide mutation rate is  $1.1 \times 10^{-8}$ , and allowing for indel mutations occurring at 5% the rate of single nucleotide events (Kondrashov 2002), this leads to an estimate for  $U$  of  $\sim 2.2$ . Let us now evaluate the potential consequences of this mutation rate for human fitness.

### The Mutation Load in Human Populations

Haldane (1937) noted that in an equilibrium population, new deleterious mutations are wiped out at the same rate as they appear. Furthermore, he showed that a deleterious

mutation reduces population mean fitness by an amount that is independent of its fitness effect. The term mutation “load,” first used by Muller (1950), is defined as the overall reduction in mean fitness relative to the mutation-free genotype brought about by recurrent deleterious mutation (Crow 1958). Muller also introduced the idea that the selective removal of each deleterious mutation is accompanied by one genetic death in an infinite population (see, e.g., Crow 1970, 1997; Kimura 1983; Kondrashov 1988; Eyre-Walker and Keightley 1999; Nachman and Crowell 2000; Reed and Aquadro 2006; Barton *et al.* 2007). However, the fitness consequences of deleterious mutations, and the mutation rate that the human population could tolerate (Kondrashov and Crow 1993; Crow 1999), depend on whether selection operates on absolute or relative fitness differences between individuals (Wallace 1970, 1975). Sved *et al.* (1967) had made a similar point in the context of the number of balanced polymorphisms that can be maintained. “One deleterious mutation one genetic death” applies to a model of “hard” selection (Wallace 1975), implying that selection is density independent and/or frequency independent. This would apply, for example, if a lethal mutation killed its carrier under all conditions, independent of the genotype of any other individual in the population. Consider a multiplicative model in which individuals survive to reproductive age with probability  $W = (1 - s)^n$ , where  $s$  is the selective disadvantage of the heterozygote and  $n$  is the number of deleterious mutations carried. If mutations are eliminated deterministically, the proportion of surviving individuals, equivalent to the population mean fitness relative to a population free from deleterious mutations, is approximately  $\bar{W} = e^{-U}$  (Kimura and Maruyama 1966). We can now evaluate  $\bar{W}$  assuming the values previously calculated for  $U$  in humans. Considering only mutations that change an amino acid of a protein-coding gene ( $U_p = 0.35$ ) predicted  $\bar{W} = 0.7$ . For deleterious mutations occurring anywhere in the genome ( $U = 2.2$ ),  $\bar{W} = 0.11$ . This latter figure seems implausibly low. For example, if each female were capable of producing 20 offspring, 18 of these progeny, on average, would need to undergo genetic death. With the caveat that some selection may occur prior to fertilization (Otto and Hastings 1998) or early in development (Edmonds *et al.* 1982), this level of selective mortality seems too high, given that there are presumably many nongenetic causes of death. High genomic deleterious mutation rates in humans and *Drosophila* (Mukai 1964; Mukai *et al.* 1972), predicting high mutation loads, led to the hypothesis that a load-reducing mechanism, such as quasi-truncation selection, speeds up the elimination of deleterious mutations and maintains higher mean fitness (Crow and Kimura 1979). Truncation selection generates synergistic epistasis, which is also an important component of the “mutational deterministic” hypothesis for the maintenance of obligate sexual reproduction (Kondrashov 1982, 1988): if  $U$  is sufficiently high, sexuals can outcompete asexuals in the presence of a twofold cost, because asexuals do not benefit from syner-

gistic epistasis (Kimura and Maruyama 1966). However, there is little empirical evidence that net synergistic epistasis for fitness is common (Kouyos *et al.* 2007; Halligan and Keightley 2009), although Peck and Waxman (2000) show that such epistasis can emerge in models involving competition in small groups.

So far, we have considered the load generated by strongly deleterious mutations eliminated deterministically. If mutations are slightly deleterious (*i.e.*, have fitness effects close to the reciprocal of the effective population size), the mutation load can be higher, because mutant alleles can drift to higher frequencies (Kimura and Crow 1963; Crow 1970). Furthermore, slightly deleterious mutations may fix, and, in a population close to equilibrium, their fixations would presumably be balanced by fixations of advantageous mutations. Even a modest proportion of fixating deleterious mutations can lead to implausibly low mean fitness (Kondrashov 1995), at least under a model in which fitness is expressed relative to a mutation-free genotype.

It is unrealistic to compare the fitnesses of mutationally loaded individuals to that of a mutation-free genotype, which has a vanishingly small probability of existing (Wallace 1970; Ewens 1979). If selection acts on relative fitness differences between extant individuals in the population, as under density- or frequency-dependent selection, the mutation load paradox can disappear entirely, and instead, load is manifest as genetic variance for fitness (Ewens 1979; Y. Lecesque, P. Keightley, and A. Eyre-Walker, unpublished data). Relative rather than absolute fitness differences are important under a range of scenarios, including competition between individuals and sexual selection. Analysis of a multiplicative model in which individuals mate with probability  $w = W/\bar{W}$  suggests that 10s or even 100s of deleterious mutations can be eliminated from a population each generation if there is a plausible input of new mutational variation for fitness. In reality, pure hard and soft selection is unlikely, and fitness effects of new mutations may be manifest via both relative and absolute fitness differences. However, a high genomic deleterious mutation rate, together with the absence of evidence for widespread synergistic epistasis, suggests that much natural selection in humans occurs via relative fitness differences between individuals.

### Consequences of Deleterious Mutation Accumulation for Human Population Fitness

Timofeoff-Ressovsky (1940) and Muller (1950) hypothesized that mutations with inconspicuous effects are more frequent than lethal mutations. Muller (1950) warned that mildly deleterious mutations may be accumulating in modern human populations, because natural selection has been relaxed by better environmental conditions. He also warned that mutation accumulation will be paid for by genetic deaths in future generations, unless improved environmental conditions can be sustained. This argument has been reiterated, with attempts to predict the rate of mutational

**Table 2** Estimates of the rate of change of fitness per generation in nematodes and *Drosophila* under whole-genome mutation accumulation ( $\Delta_{MA}$ ) and predicted rates of fitness loss per generation for humans ( $\Delta_R$ ), if selection is completely relaxed, calculated using Equation 5

Species	No. MA experiments	Mean $\Delta_{MA}$ under MA	$U$	Predicted $\Delta_R$ in humans
Nematodes	7	0.15%	2.5	0.13%
<i>Drosophila</i>	6	1.2%	1.4	1.8%

degradation, assuming that natural selection in current human populations is substantially reduced (Crow 1997, 2000; Lynch *et al.* 1999; Eyre-Walker *et al.* 2006; Lynch 2010). The maximal rate of loss of fitness, if selection is completely removed, is the “mutation pressure”  $U\bar{h}_s$  (Shabalina *et al.* 1997), where  $U$  is the deleterious mutation rate per diploid genome, and  $\bar{h}_s$  is the mean heterozygous fitness effect of a new mutation. The mutation pressure can be estimated by fitting models of the distribution of fitness effects (DFE) of new mutations to human polymorphism data, or by estimating the rate of decline of fitness in mutation accumulation experiments in model organisms and extrapolating to humans.

Unfortunately, parameter estimates based on the DFE inferred from polymorphism data are highly model dependent, making prediction of the rate of change of mean fitness under relaxed selection problematic. Most information currently comes from the analysis of polymorphism data for protein-coding genes. Boyko *et al.* (2008) fitted several models of the DFE to data for over 11,000 genes from individuals of African and European ancestry. The best-fitting DFE was a mixture distribution containing a normally distributed class of mutational effects and a point mass of strongly deleterious mutational effects. Estimates of the mean fitness effect of an amino acid mutation drawn from the normally distributed class were  $10^{-4}$  and  $10^{-3}$  for Africans and Europeans, respectively. Assuming that  $U_p = 0.35$  (see above), maximum rates of fitness loss per generation if selection were completely relaxed on these mutations are only 0.0015% and 0.02% for Africans and Europeans, respectively. The class of strongly deleterious mutations represents mutations (including some lethals) with effects  $N_e h_s > \sim 10$  that are kept so rare by purifying selection as to be essentially absent from polymorphism data. A meaningful estimate of their mean effect cannot be obtained, at least based on numbers of alleles sequenced in currently available data sets (Keightley and Eyre-Walker 2010). This difficulty in interpreting estimates of the mean of the DFE is a general problem of similar models fitted to polymorphism data (Eyre-Walker *et al.* 2006; Keightley and Eyre-Walker 2007; Boyko *et al.* 2008; Yampolsky *et al.* 2005).

### Inferences from Mutation Accumulation in Invertebrates

In a mutation accumulation (MA) experiment, new spontaneous mutations accumulate almost free from effective natural selection in inbred sublines or on replicated chromosomes protected by a balancer chromosome. There are

currently no MA experiments in model mammals, such as mice or rats, and data on the fitness consequences of spontaneous mutation accumulation in animals principally come from experiments in *Drosophila* and nematodes. Prediction of the rate of loss of fitness in humans under relaxed selection ( $\Delta_R$ ) can be made, albeit speculatively, by calculating the product of the rate of change of fitness,  $\Delta_{MA}$ , in *D. melanogaster* or nematodes under MA and the ratio of estimates of the genomic deleterious mutation rate for humans ( $U_h$ ) to *Drosophila* or nematodes ( $U_x$ ), *i.e.*,

$$\Delta_R = \Delta_{MA} U_h / U_x, \quad (5)$$

implicitly assuming that the mean selection coefficient is the same in the different species. We also assume, conservatively, that  $h = 0.5$  for all loci, and  $\Delta_{MA}$  can therefore be equated to the predicted rate of fitness loss in an outbred population evolving free from selection. Taking estimates of  $\Delta_{MA}$  from a recent review (Halligan and Keightley 2009), results of this calculation are shown in Table 2. In nematodes, there are no direct estimates of  $U$ , but an estimate of the total mutation rate per diploid genome in MA lines of *C. elegans* of  $M = 4.2$  (Denver *et al.* 2004) yields an estimate of  $U = 2.5$ , under the assumption that the fraction of selectively constrained sites in the *Caenorhabditis* and *Drosophila* genomes are similar (*i.e.*,  $\sim 60\%$ ; Halligan and Keightley 2006). For *Drosophila*, a direct estimate of  $U$  is obtained from the average mutation rate in *D. melanogaster* MA lines of three genotypes (Haag-Liautard *et al.* 2007; Keightley *et al.* 2009) multiplied by genome-wide selective constraint between closely related *Drosophila* species (Halligan and Keightley 2006).

At an arbitrary time point 200 years in the future (about eight human generations), the nematode and *Drosophila* parameters predict maximum fitness losses of 1 and 14%, respectively. However, several factors suggest that the reduction in fitness could be considerably smaller. First, calculations have assumed additive gene action, but theory suggests that large-effect mutations tend to be partially recessive (Wright 1934; Kacser and Burns 1981). Fitness reduction caused by an accumulation of recessive deleterious mutations is delayed, and predicted  $\Delta_R$  values could be 50% or more lower (García-Dorado and Caballero 2000; Vassilieva *et al.* 2000). Second, doubts have been raised about *D. melanogaster* MA experiments, particularly the validity of controls, and the possibility of nonmutational changes in fitness or unusual transposable element activity, and these doubts have been debated (Keightley 1996; García-Dorado 1997; Crow 1997; Keightley and Eyre-Walker 1999; Lynch *et al.* 1999;

Fry 2004; Halligan and Keightley 2009). Third, although modern medicine and lifestyle changes have undoubtedly reduced natural selection in some human populations, natural selection still occurs in all human populations. There is scope for selection in the germline cell lineages (Reed and Aquadro 2006), and many pregnancies spontaneously abort (Edmonds *et al.* 1982). Sexual selection still operates in human societies (Perrett *et al.* 1999), and this and other factors generate family size variation, allowing opportunities for natural selection. For example, selection associated with variation in male wealth in contemporary populations is at least as strong as selection measured in field studies of natural populations of other species (Nettle and Pollet 2008). Finally, a change in mean fitness could be inconsequential if selection is soft (for example, it might not matter if everyone becomes 5% less sexually attractive). The above considerations lead to doubts about whether deleterious mutation accumulation will produce a detectable fitness loss in humans in the foreseeable future. Less speculative, perhaps, is the existence of finite global energy, food, and water resources. Coupled with expanding human populations, these factors may intensify competition and lead to stronger natural selection in years to come.

## Acknowledgments

I thank Adam Eyre-Walker for helpful discussions and Brian Charlesworth, Bill Hill, Dan Halligan, Kateryna Makova, Alex Kondrashov, Michael Turelli, and three reviewers for comments on the manuscript. I am grateful for grant funding from the UK Biotechnology and Biological Sciences Research Council and the Wellcome Trust.

## Literature Cited

1000 Genomes Project Consortium, 2010 A map of human genome variation from population-scale sequencing. *Nature* 467: 1061–1073.

Awadalla, P., J. Gauthier, R. Myers, F. Casals, F. F. Hamdan *et al.*, 2010 Direct estimates of *de novo* mutation rates and constraints in autism and schizophrenia. *Am. J. Hum. Genet.* 87: 316–324.

Barton, N. H., 2006 How did the human species form? *Curr. Biol.* 16: R647–R650.

Barton, N. H., D. E. G. Briggs, J. A. Eisen, D. B. Goldstein, and N. H. Patel, 2007 *Evolution*, Cold Spring Harbor Laboratory Press, New York.

Benton, M. J., and P. C. J. Donoghue, 2007 Paleontological evidence to date the tree of life. *Mol. Biol. Evol.* 24: 26–53.

Boyko, A. R., S. H. Williamson, A. R. Indap, J. D. Degenhardt, R. D. Hernandez *et al.*, 2008 Assessing the evolutionary impact of amino acid mutations in the human genome. *PLoS Genet.* 4: e1000083.

Burgess, R., and Z. Yang, 2008 Estimation of hominoid ancestral population sizes under Bayesian coalescent models incorporating mutation rate variation and sequencing errors. *Mol. Biol. Evol.* 25: 1979–1994.

Cargill, M., D. Altshuler, J. Ireland, P. Sklar, K. Ardlie *et al.*, 1999 Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nat. Genet.* 22: 231–238.

Chen, F.-C., and W.-H. Li, 2001 Genomic divergences between humans and other hominoids and the effective population size

of the common ancestor of humans and chimpanzees. *Am. J. Hum. Genet.* 68: 444–456.

Chimpanzee Sequencing and Analysis Consortium, 2005 Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* 437: 69–87.

Choi, S. K., S. R. Yoon, P. Calabrese, and N. Arnheim, 2008 A germ-line-selective advantage rather than an increased mutation rate can explain some unexpectedly common human disease mutations. *Proc. Natl. Acad. Sci. USA* 105: 10143–10148.

Conrad, D. F., J. E. M. Keebler, M. A. DePristo, S. J. Lindsay, Y. Zhang *et al.*, 2011 Variation in genome-wide mutation rates within and between human families. *Nat. Genet.* 43: 712–714.

Crow, J. F., 1958 Some possibilities for measuring selection intensities in man. *Hum. Biol.* 30: 1–13.

Crow, J. F., 1970 Genetic loads and the cost of natural selection, pp. 128–177 in *Mathematical Topics in Population Genetics*, edited by Kojima, K. Springer-Verlag, New York.

Crow, J. F., 1997 The high spontaneous mutation rate: is it a health risk? *Proc. Natl. Acad. Sci. USA* 94: 8380–8386.

Crow, J. F., 1999 The odds of losing at genetic roulette. *Nature* 397: 293–294.

Crow, J. F., 2000 The origins patterns and implications of human spontaneous mutation. *Nat. Rev. Genet.* 1: 40–47.

Crow, J. F., 2006 Age and sex effects on human mutation rates: An old problem with new complexities. *J. Radiat. Res. (Tokyo)* 47: B75–B82.

Crow, J. F., and M. Kimura, 1970 *An Introduction to Population Genetics Theory*, Harper and Row, New York.

Crow, J. F., and M. Kimura, 1979 Efficiency of truncation selection. *Proc. Natl. Acad. Sci. USA* 76: 396–399.

Denver, D. R., K. Morris, M. Lynch, and W. K. Thomas, 2004 High mutation rate and predominance of insertions in the *Caenorhabditis elegans* nuclear genome. *Nature* 430: 679–682.

Ebersberger, I., D. Metzler, C. Schwarz, and S. Pääbo, 2002 Genome-wide comparison of DNA Sequences between humans and chimpanzees. *Am. J. Hum. Genet.* 70: 1490–1497.

Edmonds, D. K., K. S. Lindsay, J. F. Miller, E. Williamson, and P. J. Wood, 1982 Early embryonic mortality in women. *Fertil. Steril.* 38: 447–453.

Eöry, L., D. L. Halligan, and P. D. Keightley, 2010 Distribution of selectively constrained sites and the deleterious mutation rate in the hominid and murid genomes. *Mol. Biol. Evol.* 27: 177–192.

Ewens, W. J., 1979 *Mathematical Population Genetics*, Springer-Verlag, Berlin.

Eyre-Walker, A., and P. D. Keightley, 1999 High genomic deleterious mutation rates in hominids. *Nature* 397: 344–347.

Eyre-Walker, A., and P. D. Keightley, 2009 Estimating the rate of adaptive molecular evolution in the presence of slightly deleterious mutations and population size change. *Mol. Biol. Evol.* 26: 2097–2108.

Eyre-Walker, A., M. Woolfit, and T. Phelps, 2006 The distribution of fitness of new deleterious amino acid mutations in humans. *Genetics* 173: 891–900.

Fry, J. D., 2004 On the rate and linearity of viability declines in *Drosophila* mutation-accumulation experiments: Genomic mutation rates and synergistic epistasis revisited. *Genetics* 166: 797–806.

García-Dorado, A., 1997 The rate and effects distribution of viability mutation in *Drosophila*: Minimum distance estimation. *Evolution* 51: 1130–1139.

García-Dorado, A., and A. Caballero, 2000 On the average coefficient of dominance of deleterious spontaneous mutations. *Genetics* 155: 1991–2001.

Goriely, A., G. A. T. McVean, M. Rojmyr, B. Ingemarsson, and A. O. M. Wilkie, 2003 Evidence for selective advantage of pathogenic FGFR2 mutations in the male germ line. *Science* 301: 643–646.



- Haag-Liautard, C., M. Dorris, X. Maside, S. Macaskill, D. L. Halligan *et al.*, 2007 Direct estimation of per nucleotide and genomic deleterious mutation rates in *Drosophila*. *Nature* 445: 82–85.
- Haldane, J. B. S., 1935 The rate of spontaneous mutation of a human gene. *J. Genet.* 31: 317–326.
- Haldane, J. B. S., 1937 The effect of variation on fitness. *Am. Nat.* 71: 337–349.
- Halligan, D. L., and P. D. Keightley, 2006 Ubiquitous selective constraints in the *Drosophila* genome revealed by a genome-wide interspecies comparison. *Genome Res.* 16: 875–884.
- Halligan, D. L., and P. D. Keightley, 2009 Spontaneous mutation accumulation studies in evolutionary genetics. *Annu. Rev. Ecol. Evol. Syst.* 40: 151–172.
- Hammer, M. F., A. E. Woerner, F. L. Mendez, J. C. Watkins, M. P. Cox *et al.*, 2010 The ratio of human X chromosome to autosome diversity is positively correlated with genetic distance from genes. *Nat. Genet.* 42: 830–831.
- Huang, W., B. H.-J. Chang, X. Gu, D. Hewett-Emmett, and W.-H. Li, 1997 Sex differences in mutation rate in higher primates estimated from AMG intron sequences. *J. Mol. Evol.* 44: 463–465.
- Hudson, R. R., 1983 Testing the constant-rate neutral allele model with protein sequence data. *Evolution* 37: 203–217.
- International Human Genome Sequencing Consortium, 2001 Initial sequencing and analysis of the human genome. *Nature* 409: 860–921.
- International Human Genome Sequencing Consortium, 2004 Finishing the euchromatic sequence of the human genome. *Nature* 431: 931–945.
- Kacser, H., and J. A. Burns, 1981 The molecular basis of dominance. *Genetics* 97: 639–666.
- Keightley, P. D., 1996 Nature of deleterious mutation load in *Drosophila*. *Genetics* 144: 1993–1999.
- Keightley, P. D., and A. Eyre-Walker, 1999 Terumi Mukai and the riddle of deleterious mutation rates. *Genetics* 153: 515–523.
- Keightley, P. D., and A. Eyre-Walker, 2007 Joint inference of the distribution of fitness effects of deleterious mutations and population demography based on nucleotide polymorphism frequencies. *Genetics* 177: 2251–2261.
- Keightley, P. D., and A. Eyre-Walker, 2010 What can we learn about the distribution of fitness effects of new mutations from DNA sequence data? *Phil. Trans. R. Soc. B* 365: 1187–1193.
- Keightley, P. D., U. Trivedi, M. Thomson, F. Oliver, S. Kumar *et al.*, 2009 Analysis of the genome sequences of three *Drosophila melanogaster* spontaneous mutation accumulation lines. *Genome Res.* 19: 1195–1201.
- Keightley, P. D., L. Eöry, D. L. Halligan, and M. Kirkpatrick, 2011 Inference of mutation parameters and selective constraint in mammalian coding sequences by approximate Bayesian computation. *Genetics* 187: 1153–1161.
- Kimura, M., 1983 *The Neutral Theory of Molecular Evolution*, Cambridge University Press, Cambridge, UK.
- Kimura, M., and J. F. Crow, 1963 The mutation load in small populations. *Genetics* 48: 1301–1312.
- Kimura, M., and T. Maruyama, 1966 The mutational load with epistatic gene interactions in fitness. *Genetics* 54: 1337–1351.
- Kondrashov, A. S., 1982 Selection against harmful mutations in large sexual and asexual populations. *Genet. Res.* 40: 325–332.
- Kondrashov, A. S., 1988 Deleterious mutations and the evolution of sexual reproduction. *Nature* 336: 435–440.
- Kondrashov, A. S., 1995 Contamination of the genome by very slightly deleterious mutations. Why have we not died 100 times over? *J. Theor. Biol.* 175: 583–594.
- Kondrashov, A. S., 2002 Direct estimates of human per nucleotide mutation rates at 20 loci causing Mendelian disease. *Hum. Mutat.* 21: 12–27.
- Kondrashov, A. S., and J. F. Crow, 1993 A molecular approach to estimating the human deleterious mutation rate. *Hum. Mutat.* 2: 229–234.
- Kondrashov, F. A., and A. S. Kondrashov, 2010 Measurements of spontaneous rates of mutations in the recent past and the near future. *Phil. Trans. R. Soc. B* 365: 1169–1176.
- Kousathanas, A., F. Oliver, D. L. Halligan, and P. D. Keightley, 2011 Positive and negative selection on noncoding DNA close to protein-coding genes in wild house mice. *Mol. Biol. Evol.* 28: 1183–1191.
- Kouyos, R. D., O. K. Silander, and S. Bonhoeffer, 2007 Epistasis between deleterious mutations and the evolution of recombination. *Trends Ecol. Evol.* 22: 308–315.
- Li, W.-H., and L. A. Sadler, 1991 Low nucleotide diversity in man. *Genetics* 129: 513–523.
- Lynch, M., 2010 Rate, molecular spectrum, and consequences of human mutation. *Proc. Natl. Acad. Sci. USA* 107: 961–968.
- Lynch, M., J. Blanchard, D. Houle, T. Kibota, S. Schultz *et al.*, 1999 Perspective: Spontaneous deleterious mutation. *Evolution* 53: 645–663.
- Makova, K. D., and W.-H. Li, 2002 Strong male-driven evolution of DNA sequences in humans and apes. *Nature* 416: 624–626.
- Miyata, T., H. Hayashida, K. Kuma, K. Mitsuyasu, and T. Yasunaga, 1987 Male-driven molecular evolution: a model and nucleotide sequence analysis. *Cold Spring Harb. Symp. Quant. Biol.* 52: 863–867.
- Mouse Genome Sequencing Consortium, 2002 Initial sequencing and comparative analysis of the mouse genome. *Nature* 420: 520–562.
- Mukai, T., 1964 The genetic structure of natural populations of *Drosophila melanogaster*. I. Spontaneous mutation rate of polygenes controlling viability. *Genetics* 50: 1–19.
- Mukai, T., S. I. Chigusa, L. E. Mettler, and J. F. Crow, 1972 Mutation rate and dominance of genes affecting viability in *Drosophila melanogaster*. *Genetics* 72: 333–355.
- Muller, H. J., 1950 Our load of mutations. *Am. J. Hum. Genet.* 2: 111–176.
- Nachman, M. W., and S. L. Crowell, 2000 Estimate of the mutation rate per nucleotide in humans. *Genetics* 156: 297–304.
- Nettle, D., and T. V. Pollet, 2008 Natural selection on male wealth in humans. *Am. Nat.* 172: 658–666.
- Ossowski, S., K. Schneeberger, J. I. Lucas-Lledo, N. Warthmann, R. M. Clark *et al.*, 2010 The rate and molecular spectrum of spontaneous mutations in *Arabidopsis thaliana*. *Science* 327: 92–94.
- Otto, S. P., and I. M. Hastings, 1998 Mutation and selection within the individual. *Genetica* 102–3: 507–524.
- Patterson, N., D. J. Richter, S. Gnerre, E. S. Lander, and D. Reich, 2006 Genetic evidence for complex speciation of humans and chimpanzees. *Nature* 441: 1103–1108.
- Peck, J. R., and D. Waxman, 2000 Mutation and sex in a competitive world. *Nature* 406: 399–403.
- Perrett, D. I., D. M. Burt, I. S. Penton-Voak, K. J. Lee, D. A. Rowland *et al.*, 1999 Symmetry and human facial attractiveness. *Evol. Hum. Behav.* 20: 295–307.
- Petrov, D. A., 2002 DNA loss and evolution of genome size in *Drosophila*. *Genetica* 115: 81–91.
- Presgraves, D. C., and S. Yi, 2009 Doubts about complex speciation between humans and chimpanzees. *Trends Ecol. Evol.* 24: 533–540.
- Reed, F. R., and C. F. Aquadro, 2006 Mutation, selection and the future of human evolution. *Trends Genet.* 22: 479–484.
- Rhesus Macaque Genome Sequencing and Analysis Consortium, 2007 Evolutionary and biomedical insights from the rhesus macaque genome. *Science* 316: 222–234.
- Roach, J. C., G. Glusman, A. F. A. Smit, C. D. Huff, R. Hubley *et al.*, 2010 Analysis of genetic inheritance in a family quartet by whole-genome sequencing. *Science* 328: 636–639.

- Shabalina, S. A., L. Y. Yampolsky, and A. S. Kondrashov, 1997 Rapid decline of fitness in panmictic populations of *Drosophila melanogaster* maintained under relaxed natural selection. *Proc. Natl. Acad. Sci. USA* 94: 13034–13039.
- Sved, J. A., A. Reed, and W. F. Bodmer, 1967 The number of balanced polymorphisms that can be maintained in a natural population. *Genetics* 55: 469–481.
- Takahata, N., Y. Satta, and J. Klein, 1997 Divergence time and population size in the lineage leading to modern humans. *Theor. Popul. Biol.* 48: 198–221.
- Taylor, J., S. Tyekucheva, M. Zody, F. Chiaromonte, and K. D. Makova, 2006 Strong and weak male mutation bias at different sites in the primate genomes: insights from the human-chimpanzee comparison. *Mol. Biol. Evol.* 23: 565–573.
- Timofeeff-Ressovsky, N. W., 1940 Mutations and geographical variation, *The New Systematics*, edited by Huxley, J. Clarendon Press, Oxford.
- Torgerson, D. G., A. R. Boyko, R. D. Hernandez, A. Indap, X. Hu *et al.*, 2009 Evolutionary processes acting on candidate cis-regulatory regions in humans inferred from patterns of polymorphism and divergence. *PLoS Genet.* 5: e1000592.
- Vassilieva, L. L., A. M. Hook, and M. Lynch, 2000 The fitness effects of spontaneous mutations in *Caenorhabditis elegans*. *Evolution* 54: 1234–1246.
- Wakeley, J., 2008 Complex speciation of humans and chimpanzees. *Nature* 452: e3–e4.
- Wallace, B., 1970 *Genetic Load. Its Biological and Conceptual Aspects*, Prentice-Hall, Englewood Cliffs, New Jersey.
- Wallace, B., 1975 Hard and soft selection revisited. *Evolution* 29: 465–473.
- Wood, B., 2002 Palaeoanthropology: Hominid revelations from Chad. *Nature* 418: 133–135.
- Wright, S., 1934 Molecular and evolutionary theories of dominance. *Am. Nat.* 68: 24–53.
- Wright, S., 1938 The distribution of gene frequencies under irreversible mutation. *Proc. Natl. Acad. Sci. USA* 24: 253–259.
- Vicoso, B., and B. Charlesworth, 2009 Recombination may affect polymorphism levels on the X and the autosomes in African populations of *Drosophila*. *Genetics* 181: 1699–1701.
- Xue, Y., Q. Wang, Q. Long, B. L. Ng, H. Swerdlow *et al.*, 2009 Human Y chromosome base-substitution mutation rate measured by direct sequencing in a deep-rooting pedigree. *Curr. Biol.* 19: 1453–1457.
- Yampolsky, L. Y., F. A. Kondrashov, and A. S. Kondrashov, 2005 Distribution of the strength of selection against amino acid replacements in human proteins. *Hum. Mol. Genet.* 14: 3191–3201.

Communicating editor: M. Turelli