YEARBOOK OF PHYSICAL ANTHROPOLOGY 53:13–45 (2010)

Nonadaptive Processes in Primate and Human Evolution

Eugene E. Harris*

Department of Biological Sciences and Geology, Queensborough Community College, City University of New York, Bayside, NY 10364

KEY WORDS genetic drift; positive selection; purifying selection; effective population size; neutral theory; selective constraint; gene duplication; compensatory substitution

ABSTRACT Evolutionary biology has tended to focus on adaptive evolution by positive selection as the *primum mobile* of evolutionary trajectories in species while underestimating the importance of nonadaptive evolutionary processes. In this review, I describe evidence that suggests that primate and human evolution has been strongly influenced by nonadaptive processes, particularly random genetic drift and mutation. This is evidenced by three fundamental effects: a relative relaxation of selective constraints (i.e., purifying selection), a relative increase in the fixation of slightly deleterious mutations, and a general reduction in the efficacy of positive selection. These effects are observed in protein-coding, regulatory regions, and in gene expression data, as well as in an augmentation of fixation of large-scale mutations, including duplicated genes, mobile genetic elements, and nuclear mitochondrial DNA. The evidence suggests a general population-level explanation such as a reduction in effective population size (N_e). This would have tipped the balance between the evolutionary forces of natural selection and random genetic drift toward genetic drift for variants having small selective effects. After describing these proximate effects, I describe the potential consequences of these effects for primate and human evolution. For example, an increase in the fixation of slightly deleterious mutations could potentially have led to an increase in the fixation rate of compensatory mutations that act to suppress the effects of slightly deleterious substitutions. The potential consequences of compensatory evolution for the evolution of novel gene functions and in potentially confounding the detection of positively selected genes are explored. The consequences of the passive accumulation of large-scale genomic mutations by genetic drift are unclear, though evidence suggests that new gene copies as well as insertions of transposable elements into genes can potentially lead to adaptive phenotypes. Finally, because a decrease in selective constraint at the genetic level is expected to have effects at the morphological level, I review studies that compare rates of morphological change in various mammalian and island populations where N_e is reduced. Furthermore, I discuss evidence that suggests that craniofacial morphology in the *Homo* lineage has shifted from an evolutionary rate constrained by purifying selection toward a neutral evolutionary rate. Yrbk Phys Anthropol 53:13–45, 2010.

Although we are nearly 99% similar to chimpanzees in nucleotide sequence, phenotypically we appear to have diverged considerably farther from our common ancestor. Our unique human traits include bipedal walking, increased manual dexterity, increased brain size, reduced body hair, and our ability to make and use tools and complex language. Most efforts to understand the evolution of these features have emphasized adaptive evolution as the most important, if not singular, evolutionary force. Yet, adaptive evolution is only one of the forces of evolutionary change, others being mutation, random genetic drift, and recombination. These forces are described as nonadaptive because the evolutionary change they produce is due to factors unrelated to individual differences in relative fitness. The importance of nonadaptive forces in influencing evolutionary change is becoming increasingly clear (Stoltzfus, 1999, 2006; Koonin, 2004, 2009a,b; Hughes, 2007, 2008, 2009; Lynch, 2007a,b; Stoltzfus and Yampolsky, 2009; Ellegren, 2009). Within anthropological and human genetics, random genetic drift has most often been studied with respect to founder effects on isolate populations (e.g., Hutterites, Old Order Amish, Ashkenazi Jews, Yanomamo etc.; see Smouse et al., 1981; Sokal et al., 1986; Arcos-Burgos and Muenke, 2002; Risch et al., 2003; Slatkin, 2004) as well as in bottleneck events (e.g., first migration into the Americas; see Cavalli-Sforza et al., 1994). Much less explored has been the role of random genetic drift in influencing the evolution of primates, and humans in particular.

The well-known neutralist versus selectionist controversy in evolutionary genetics is essentially a debate about whether evolutionary change is mostly due to genetic drift acting on neutral variation, or whether it is due to selection acting on adaptive variation (Ford, 1964; Kimura, 1968, 1969; King and Jukes, 1969; Lewontin, 1974; Gillespie, 1991; Kreightman, 1996; see Nei, 2005). The controversy has subsided somewhat since its most vociferous days in the 1970s and 80s but has not reached resolution (see Hey, 1999; Crow, 2008; Hahn, 2008). Interest in this issue, however, has reignited within the last decade in the context of analyses of large amounts of genome-wide DNA data from humans and other

*Correspondence to: Eugene E. Harris, Department of Biological Sciences and Geology, Queensborough Community College, City University of New York, 222-05 56th Avenue Bayside, NY 11364. E-mail: eharris@qcc.cuny.edu

DOI 10.1002/ajpa.21439
Published online in Wiley Online Library (wileyonlinelibrary.com).

species. Through analyses of these data, it is becoming increasingly revealed that the population size of a species, specifically its long-term effective population size (N_e), is fundamental in determining the relative importance of genetic drift and selection. Furthermore, it is becoming clear that the relative power of these forces has varied in different animal lineages (for discussion see Ellegren, 2009). The findings are presumably explained by population genetics theory that posits a central role of N_e in influencing the relative importance of these two evolutionary processes. Explained briefly, when molecular variants have large selective coefficients, either because they are strongly advantageous or strongly deleterious, natural selection can act as a powerful force in shaping the direction of molecular evolution. However, if most variants have only slight effects on fitness, a finding that has become increasingly supported (Ohta, 1973, 1974, 2003; Hughes et al., 2003, 2005; Eyre-Walker et al., 2006a; Eyre-walker and Keightley, 2007), then the extent to which natural selection can discriminate these variants is a function of N_e. For these variants, N_e becomes the critical determining factor that tips the balance between either natural selection being the dominant evolutionary force acting on them, or random genetic drift becoming the dominant force. When N_e is sufficiently small, the effects of genetic drift dominate over natural selection and completely determine the fate of these variants. This theory is known as the Nearly Neutral Theory by Tomoko Ohta (1973, 1974), which is an extension of Motoo Kimura's Neutral Theory Of Molecular Evolution (Kimura, 1968, 1969, 1983). On the other hand, if N_e is sufficiently large, natural selection is better able to discriminate the selective effects of genetic variants, tending to promote advantageous variants while removing deleterious variants, even if these have only slightly negative selective coefficients.

Analyses of human genetic data have consistently estimated relatively small values of N_e approximating 10,000 (Burgess and Yang, 2008; Yu et al., 2001) or perhaps considerably smaller (see Tenesa et al., 2007). Note that N_e is approximately equal to the harmonic mean rather than the arithmetic mean of population sizes through evolutionary time and is, therefore, strongly influenced by historically smaller population sizes. Thus, even if there was a recent and profound demographic expansion, the much smaller population sizes that characterized most of human evolution results in an N_e that approaches the smaller values. Because low estimates of N_e for humans have been attained in two basic ways, through analyses of levels of human nucleotide polymorphism (Harris and Hey, 1999, 2001; Przeworski et al., 2000; Yu et al., 2001; Wall, 2003) and through phylogenetic comparisons of multiple genes among hominoid species (Chen and Li, 2001; Burgess and Yang, 2008), they indicate that N_e has been consistently small over the ~6 Myrs of time since human divergence from chimpanzees through the emergence of anatomically modern humans (~200 kya) until recently. Additionally, the N_e of the common ancestral population between humans and Neandertals is also estimated to be close to 10,000 (see Noonan et al., 2006; Premo and Hublin, 2009). Furthermore, a recent bottleneck in the evolution of *Homo sapiens* supports a greater reduction in the effective population size of non-African human populations (Gherman et al., 2007). In contrast, the human (H)/chimpanzee (C) ancestral population is estimated to have had a much larger N_e, with estimates ranging from 104K to 157K (Chen and Li, 2001; Burgess and Yang, 2008). Estimates of N_e in the human H/C/gorilla (G) ancestral population range from 55K to 83K, and estimates for the ancestral H/C/G/orangutan population range from 84K to 126K (Burgess and Yang, 2008). These estimates indicate an approximately 10-fold reduction in N_e in the course of human evolution, a reduction that seems to be extreme for a primate species.

Although few good estimates of N_e exist for nohuman primates, it is generally assumed that N_e is reduced generally in primates compared with other animal groups. Recent analyses of polymorphism in macaques estimate that N_e is approximately 70,000 (Hernandez et al., 2007). In contrast, estimates of N_e in murids (mice and rats) are placed at around 450K–820K (Eyre-Walker et al., 2002). Although more distantly related, *Drosophila* species, which have effective population sizes an order of magnitude larger than mammal species (1–2 million; see Eyre-Walker et al., 2002), they provide a useful contrast for studying the evolutionary effects of differences in population size.

A reduction in N_e in the evolution of primates and more so in the course of human evolution is expected to have had considerable effects on the dynamics of molecular evolution. The reduction is expected to have augmented the relative power of random genetic drift relative to natural selection. This, in turn, is expected to have its greatest effects on slightly deleterious variants, both increasing the fixation of these variants by genetic drift during divergence, and in permitting slightly deleterious variants to increase in frequency in recent human populations. These predictions are supported by studies making large-scale genomic comparisons between primates and other species (Eyre-Walker et al., 2002; Keightley et al., 2005a,b; Kryukov et al., 2005) and in genome-wide analyses of polymorphism within human populations (Fay et al., 2001; Hughes et al., 2003, 2005; Gojobori et al., 2007).

In this article, I explore the genetic and genomic evidence that indicates a relative augmentation in the power of random genetic drift in relation to natural selection in primate and human evolution. There are two central questions I explore. First, what is the evidence that genetic drift has played a relatively increased role in primates, and specifically in the human lineage, compared with other animal lineages? Second, how might an augmentation in random genetic drift have influenced the direction of human evolution at the phenotypic level?

To address these questions, I examine two different types of genetic evidence. The first evidence compares the rate and pattern of nucleotide divergence observed between humans and chimpanzees with that observed between other pairs of species (e.g. mice versus rats). As an extension, I discuss comparative data on the divergence of gene expression. Although I focus on describing nonadaptive processes acting over time periods associated with primate and human evolution, I also consider how nonadaptive processes may have affected genetic variation between populations of modern humans. The second evidence compares large-scale differences in genome architecture between the human genome and the genomes of other species in their numbers of gene duplications, transposable element (TE) insertions, and mitochondrial insertions into the nuclear genome (i.e., nuclear mitochondrial DNA, *numts*). I explore the evidence that indicates that there was an increase in the fixation of TEs, gene duplications, and *numts* in primates and human evolution and how this may be a consequence of a

reduction in N_e. I then discuss how nonadaptive genetic changes at the nucleotide and genomic levels can potentially influence evolution at the phenotypic level.

EFFECTIVE POPULATION SIZE AND NEARLY NEUTRAL VARIANTS

When genetic variants arise in a population, there are essentially two evolutionary forces that influence their fates: natural selection and genetic drift. If we consider natural selection first, there are two predominant forms: purifying selection and positive selection. Purifying selection (or negative selection) removes deleterious variants from a population. Positive selection acts to drive beneficial variants to higher frequencies and can potentially drive them to 100% frequency (i.e., fixation) in a population or species. Genetic drift, on the other hand, is a stochastic force because of the random sampling of gametes from generation to generation within populations. The process leads to the random fluctuation of frequencies of genetic variants over time.

Neutral theory (Kimura, 1968, 1969, 1983) has posited that most evolutionary change occurs by random genetic drift and not by positive selection. It describes that molecular variants continuously arise within a population by mutation and that these variants are either neutral or deleterious. Under neutral theory, deleterious variants are assumed to be continuously purged from populations by the ubiquitous process of purifying selection. On the other hand, neutral variants (with no effect on fitness) are fixed or lost within a population according to random chance effects. Although neutral theory acknowledges positive selection is an important evolutionary force, because advantageous variants are assumed to be extremely rare relative to neutral or deleterious variants, positive selection is also assumed to be rare.

The *nearly* neutral theory (Ohta, 1973, 1974) proposes that most variants are not strictly neutral but have slight selective effects with most believed to be slightly deleterious. A class of slightly advantageous variants is also assumed to exist, but is believed to be much smaller. According to nearly neutral theory, the fate of variants with slight selective effects is understood to be dependent on the joint effects of selection and genetic drift.

The relative influence of these forces on the fate of newly arisen genetic variants can be quantified by calculating the ratio of the fixation probability of a selected variant over the fixation probability of a hypothetical neutral variant. The ratio is expressed mathematically by $\theta_f = 4N_e s/1/e^{-4Nes}$ and is graphed in Figure 1, in which N_e represents the effective population size and s is the selective coefficient. On the x axis, negative values indicate deleterious variants and positive values indicate advantageous variants. $N_e s = 0$ represents a neutral variant in which case the ratio becomes 1.0. The fate of a new variant within a population is a function of two factors: the selective effect of the variant and the effective size of the population in which the variant arises.

Let us consider the effects of different values of N_e on the fixation probability of a slightly deleterious variant (e.g., $s = -10^{-5}$) by considering two different effective population sizes: 1) the case in which $N_e = 100{,}000$ and 2) the case in which $N_e = 10{,}000$. These values approximate the values of N_e empirically estimated for the common ancestor of chimpanzees and humans and for humans, respectively (Burgess and Yang, 2008). In the first case, $N_e s = -0.1$ and $\theta_f = 0.81$, and, in the second case, $N_e s = -1.0$ and $\theta_f = 0.07$. As can be observed, when N_e is large, the slightly deleterious variant has only an extremely slight chance of becoming fixed in the population (7%). On the other hand, when N_e is relatively small, its chance of fixation increases to as much as 81% of the probability of fixation of a neutral variant. Thus, even though the variant is selectively deleterious, the variant behaves very nearly like a neutral variant. Put another way, when the selection coefficient of a variant, s, is considerably greater than the reciprocal of the effective population size (i.e., $s \gg 1/4N_e$), selection is able to largely dominate over genetic drift in determining the fate of the variant. On the other hand, if the selection coefficient is much smaller than the reciprocal of the effective population size (i.e., $s \ll 1/4N_e$), then genetic drift dominates. Note that the relationship is the same for both slightly deleterious and slightly advantageous variants.

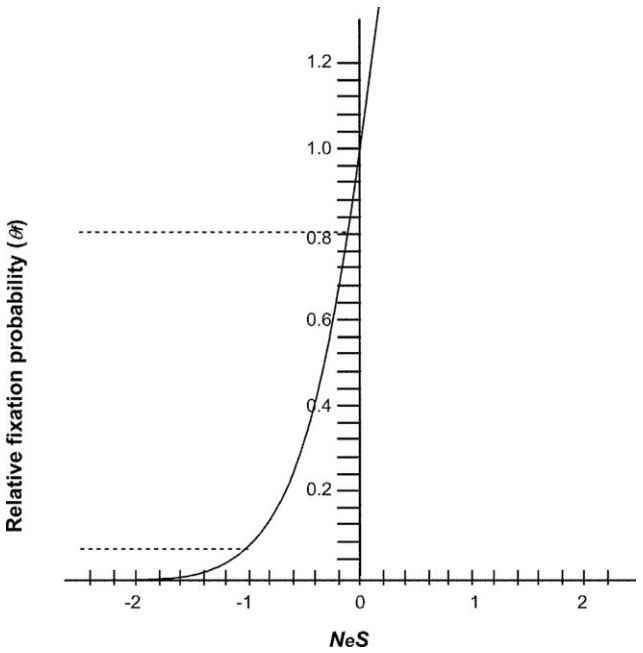

Fig. 1. The probability of fixation of a new variant with respect to the neutral expectation of 1/(2N) graphed as a function of the product of the effective population size and the selection coefficient of the variant ($N_e \times s$). The dashed lines represent cases in which $s = -10^{-5}$ but N_e takes on different values, either 10,000 (upper dashed line) or 100,000 (lower dashed line).

How does the nearly neutral theory impact our understanding of human evolution? Because N_e during human evolution is estimated to be 10-fold reduced with respect to the chimpanzee–human ancestor, and is markedly reduced compared with other mammals (e.g., murids), we can make several predictions that are testable through comparative genomic studies between humans and other animal species. Compared with species with larger estimated N_e, humans are expected to 1) show evidence of reduced purifying selection during their evolution; 2) have fixed a relatively large fraction of slightly deleterious variants during the course of their evolution; and 3) exhibit evidence of reduced levels of positive selection during their evolution. In addition, an examination of genetic variation within human populations should show a relatively large class of slightly deleterious variants maintained at low frequencies. This effect

should be exacerbated for human populations that have undergone reduction in the recent evolutionary past. In the following, I will explore the extent to which human genetic data of various types meet these predictions. In addition, because N_e is reduced in primates compared with other animal groups, we expect to see similar effects in primates generally.

EVOLUTION OF THE PROTEIN-CODING GENOME

Humans and chimpanzees are very similar when protein-coding portions of their genomes are compared. Mikkelsen et al. (2005) found that they are identical at 29% of proteins, and, at the remaining 71%, they differ to a very small degree, at only one or two amino acid sites. Using such observations, it is estimated that the two species differ over the genome at a relatively small number of amino acid sites—around 60,000 amino acid sites (Eyre-Walker, 2006a). The differences between the two species that we observe today originally arose as mutational variants within populations of either species, though represent the small subset of all variants that ultimately became fixed in the diverging species. The fixation of these variants was either the result of random genetic drift or positive selection. (It should be noted that a very large fraction of deleterious variants that arose in the two species was removed by purifying selection.) Overall, we can place the variants that arose along the human lineage into three categories based on the type of evolutionary forces acting on them: 1) those variants removed by purifying selection; 2) those fixed by genetic drift, either because they had no effect on fitness or because their selection coefficients were sufficiently small relative to N_e that they behaved as if they were neutral; and 3) those variants fixed by positive selection. Working within the framework of neutral molecular theory (Kimura, 1983), it is possible to quantify the relative proportions of these different types of variants, which can allow us to estimate the relative importance of these different modes of evolution in human–chimpanzee divergence. The results of these types of analyses can allow us to evaluate our predictions based on reduced N_e in humans.

Purifying selection on proteins

According to neutral theory (Kimura, 1968, 1969, 1983), most variants that alter amino acid sequences are assumed to be deleterious and are, therefore, expected to be removed from the population by purifying selection. Judging from the very small number of amino acid differences between human and chimpanzee proteins, we can infer that their similarity has been conserved by purifying selection. Can we quantify the degree to which purifying selection has conserved human and chimpanzee proteins? Was the strength of purifying selection that acted along human and chimpanzee lineages comparable with the strength of purifying selection that acted in other species lineages?

The method most often used to quantify purifying selection is the dN/dS statistic (also denoted by ω) (Hughes and Nei, 1988). The method compares the rates of substitution at two different classes of nucleotide sites within protein-coding regions, nonsynonymous and synonymous sites. Nonsynonymous substitutions cause a change in an amino acid, whereas synonymous substitutions do not. The method relies on the assumption that substitutions between species at synonymous sites occur at a neutral rate and then compares the rate of change at nonsynonymous sites against this rate. Thus, when dN is less than dS (i.e., $\omega < 1$), purifying selection is inferred, and this is the most common situation across genes. However, when the rate of change at nonsynonymous sites (dN) is equal to the rate at synonymous sites (dS) (i.e., when dN/dS or $\omega = 1$), a gene is inferred to be evolving neutrally. Alternatively, when dN exceeds dS (i.e., $\omega > 1$), a situation that occurs at only a minority of genes, positive selection is inferred.

To frame the discussion of how N_e is associated with values of ω, let us consider species with vastly different estimated values of N_e. As mentioned above, humans have an estimated N_e of approximately 10,000, whereas *Drosophila* species have an N_e near 1–2 million (Eyre-Walker et al., 2002). Consistent with this difference, estimates for ω estimated between *Drosophila* species-pairs (Hegger and Ponting, 2007) are very low (0.06–0.11), whereas the values estimated between humans and chimpanzees falls in the range of 0.169–0.259 (see references in Table 1). This is consistent with the theoretical prediction that purifying selection is less efficient at removing amino acid altering variants in smaller populations.

The dN/dS statistic has been applied to obtain genome-wide values of ω for different species lineages (Mikkelsen et al., 2005; Bakewell et al., 2007; Kosiol et al., 2008; Table 1). The most recent estimates are by Kosiol et al. (2008) who compared 16,000 orthologous genes between the eight different mammal species. The ω values for the human lineage (0.249) and chimpanzee lineage (0.245) are about twice as high as the values for the mouse (0.127) and the rat (0.121) lineages. Lower values of ω indicate greater degrees of purifying selection, and higher values indicate lesser degrees of purifying selection. One way to interpret ω is that it implies that considerably fewer amino acid altering variants were removed by purifying selection in human (75.1%) and chimpanzee (75.5%) evolution compared with the number removed during mouse (87.3%) and rat (87.9%) evolution (see Mikkelsen et al., 2005). In sum, it seems that humans and chimpanzees have experienced considerably reduced purifying selection on protein sequence compared with murids, and this difference is supported by all studies that have reported values of ω (Table 1).

It should be noted that values of ω vary (sometimes considerably) between studies even though the pattern of increasing values in species with larger effective population sizes remains consistent. This variation is probably due to differences in the exact set of genes used in the different studies and differences in methods used. For example, studies that focus on subsets of genes that are more conserved across species will estimate average values of ω that are lower than the values of ω estimated in studies that focus on less-conserved sets of genes. Differences, between studies, in the stringency of filters used to identify orthologous genes across species (e.g., in the *E*-value cutoff used in a BLAST search to identify genes across species having a specific level of sequence similarity) will lead to data sets of different average conservation and will yield different average values of ω. Also, if different studies use different methods in estimating dN/dS (ω) [e.g., counting methods versus maximum likelihood methods as implemented in Phylogenetic Analysis by Maximum Likelihood (PAML)], this can lead to different values of ω.

Within primates, we can compare ω between different species. For example, Kosiol et al. (2008) estimated ω to

TABLE 1. dN/dS compared between mammalian species having different estimates of effective population size (given in parentheses)

Human (10,000)[a]	Chimpanzee (21,000)[b]	Macaque (73,000)	Dog (–)	Mouse (450,000–810,000)[c]	Rat	Opossum	Platypus	*Drosophila* (~1–2 million)[d]	Reference
0.208	0.194	–	–	0.142	0.137	–		–	Mikkelsen et al. (2005)
0.259	0.245	0.226	–	–	–	–		–	Bakewell et al. (2007)
0.169	0.175	0.124	0.111	0.104	–	–		–	Gibbs et al. (2007)
0.249	0.245	0.191	0.140	0.127	0.121	–		–	Kosiol et al. (2008)
0.132	–	–	0.128	0.105	–	0.125	0.132	–	Warren et al. (2008)
0.112	–	–	0.095	0.088	–	–			Lindblad-Toh et al. (2005)
–	–	–	–	–	–	–		0.06–0.11	Heger and Ponting (2007)

[a] This is an approximate estimate of N_e based on many different studies.
[b] From Caswell et al. (2008); this is the N_e estimated for chimpanzee and bonobo. However, this article estimated much higher estimates for Central chimpanzees (118K).
[c] From Keightley et al. (2005b).
[d] From Eyre-Walker et al. (2002).

be 0.191 in the macaque lineage. This indicates that genetic constraints due to purifying selection were greater in the macaque lineage. Comparison of humans with macaques indicates that, although approximately 81% of mutations were purged by purifying selection in the macaque, only 75% were purged in humans. The increased constraint in macaques is observed in all three studies for which data are available (Table 1). Finally, comparing chimpanzees with humans, we see that ω is somewhat higher in humans compared with chimpanzees in three of the four studies, indicating reduced purifying selection in humans. The Kosiol et al. (2008) study suggests only a small difference. However, the degree of relaxed constraint in humans is higher in the Mikkelsen et al.'s (2005) and Bakewell et al.'s (2007) studies. Bakewell et al. (2007) determined that the difference between the two species was significant. (It is not clear why the macaque genome study (Gibbs et al., 2007) estimated smaller values of ω in humans compared with chimpanzees.) In sum, there seems to be a pattern in recent studies indicating slightly less purifying selection in humans compared with chimpanzees and considerably less purifying selection compared with macaques (Mikkelsen et al., 2005; Bakewell et al., 2007; Gibbs et al., 2007; Kosiol et al., 2008). Less purifying selection in humans compared with chimpanzees could be due to relatively smaller N_e in humans (see Bakewell et al., 2007).

In a more detailed study of the relationship between N_e and the magnitude of selective constraints, Hughes and Friedman (2009) explicitly compared the divergence of approximately 5,000 protein genes between two pairs of recently diverged species—between humans and rhesus monkey (*Macaca mulatta*), on the one hand, and between mouse (*Mus domesticus*) and Norway rat (*Rattus norvegicus*), on the other hand (Fig. 2 illustrates their most important findings). As in other studies, they found significantly increased mean ω (0.196) in the divergence between the primates compared with the murids (0.138). Their more detailed analysis categorized the qualitative types of amino acid substitutions that took place in the divergences between the pairs of species. They found that amino acid replacements in primates generally introduced more chemically dissimilar amino acids (based on the distance method of Miyata et al. 1979). The difference was most pronounced for the set of genes in primates that have ω values higher than in murids. The authors conclude: "this finding is consistent with the hypothesis that the elevated ω in primates is due mainly to the fixation of slightly deleterious mutations, since slightly deleterious mutations are likely to involve greater chemical dissimilarity of amino acid residues than those that are strictly neutral, yet are less likely to involve extremely radical changes (Hughes and Friedman, 2009, p 55)."

In sum, we can see that there is relatively strong evidence for the relaxation of selective constraints on protein evolution (vis-à-vis purifying selection) in primates compared with other animal species. Relaxed constraint along the hominid lineage was also supported in a recent genome-wide analysis of six mammalian genomes (human, chimpanzee, gorilla, orangutan, macaque, and dog) by McVicker et al. (2009). The relaxed constraint was found not only for protein-coding regions but also for conserved noncoding (CNC) regions (presumably regulatory in function).

Most authors have attributed the evidence for reduced selective constraints on protein evolution in primates to their generally lower effective population sizes. Are alternative interpretations possible? One potential explanation could be the possibility that positive selection within primate evolution has generally been more prevalent than in murids. Thus, evidence for relaxed selection in primates could instead be reinterpreted to be evidence of more widespread positive selection in which there is increased adaptive substitution. However, this explanation seems unlikely because it would be difficult to explain positive selection acting at amino acid sites across a very broad set of genes in the primate genome but not in the murid genome. One consequence of reduced purifying selection would be a relative increase in the fixation of less-conservative (slightly damaging) amino acids in primates by random genetic drift, with such substitutions having relatively slight selective coefficients. The Hughes and Friedman (2009) analysis supports the augmentation of slightly deleterious substitutions in primates. In addition, several studies (described above) show further relaxation of selective constraints along the human lineage after separation from the chimpanzee lineage. Thus, it can be further hypothesized that there was a relative increase in the fixation of slightly deleterious variants during human evolution, a topic I consider in the following section.

Slightly deleterious variants in proteins

The neutral theory originally considered that most nucleotide variants that cause amino acid alterations had

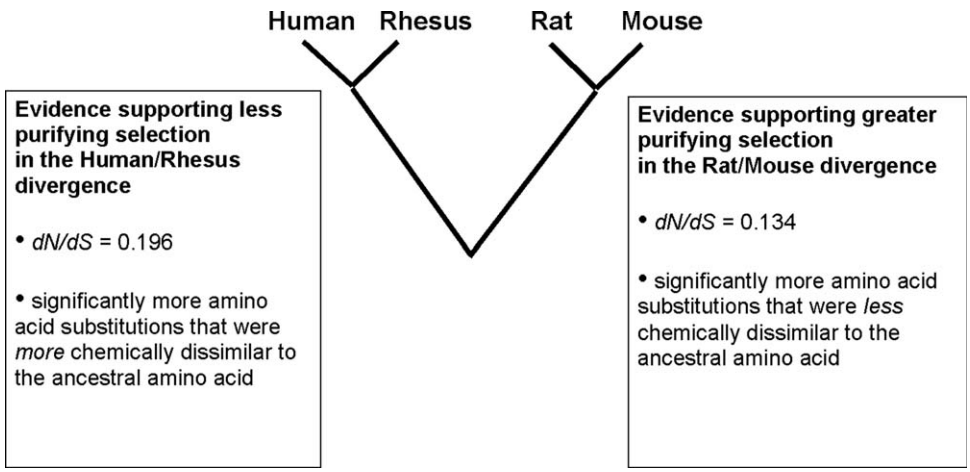

Fig. 2. Summarization of some of the findings of an analysis by Hughes and Friedman (2009), which compared 5,000 protein coding genes between two pairs of species (i.e., between human and rhesus macaque and between mouse and rat). The results indicate considerably less-effective purifying selection in the human divergence from rhesus macaques than in the divergence between the mouse and rat.

sufficiently strong negative effects that they would be quickly removed from a population. These variants were believed neither to contribute to divergence between species nor to polymorphism between individuals within a species. However, as described above, some amino acid-altering variants have small selection coefficients, and the fates of these variants in a population are determined largely by the balance between two evolutionary forces (i.e., selection versus genetic drift). When N_e is sufficiently small, genetic drift can overwhelm the power of selection. The low effective population size estimated for humans ($\sim$10,000) and the resulting relaxation in selective constraints due to purifying selection would have allowed an increased proportion of variants having slight selective coefficients to drift to relatively high frequencies, with some fraction being fixed in the species. (Because most variants that alter protein structure are deleterious, this class has become known as the slightly deleterious class of variants, even though a small fraction of variants having slight selective coefficients are slightly advantageous (Charlesworth and Eyre-Walker, 2007)). The net effect of fixation of these variants would be to increase rates of protein evolution in species with small N_e relative to closely related species with larger population sizes. Is it possible to quantify the amount of effectively neutral (slightly deleterious) substitutions that accumulated in the human lineage relative to a closely related species? Can we estimate the effect on fitness of these substitutions?

In the divergence between murids and primates, there exists a fraction of amino acid substitutions that fall into the effectively neutral category that were fixed in primates because of random genetic drift, but were removed in murids by purifying selection (see Kondrashov, 1995; Eyre-Walker et al., 2002). Let us consider two lineages, the murid lineage with $N_e(M)$ and the human lineage with $N_e(H)$, diverging from a common ancestor. The murid lineage has an effective population size greater than that for humans such that $N_e(M) > N_e(H)$, where M indicates murids and H indicates humans. The selective effects of those substitutions that are effectively neutral in humans but not in murids will lie between the reciprocal of the effective populations of the two species: $1/4N_e(M) < s < 1/4N_e(H)$. Let us assume that N_e in murids is roughly 100,000, whereas that in humans is roughly 10,000. Then, the selective effects of these substitutions fall in the range of 10^{-5} ($1/4 \times 100{,}000 = 0.25 \times 10^7$) and 10^{-4} ($1/4 \times 10{,}000 = 2.5 \times 10^{-5}$), and, therefore, these substitutions have very slight effects on fitness. Eyre-Walker et al. (2002) conducted an analysis of this sort using the differences in selective constraints in large sets of orthologous proteins compared between two species-pairs: mouse and rat, and human and chimpanzee. Assuming $N_e(M)$ and $N_e(H)$ are 220,000 and 15,000 respectively, they found that $\sim$15% of variants have $-1/4N_e(M) < s < -1/4N_e(P)$. This represents the subfraction of all substitutions between humans and murids that have gone to fixation in primates (but not in murids) because they were effectively neutral in primates even if many had slightly deleterious effects.

Evidence of a relatively large class of slightly deleterious variants in humans also comes from human polymorphism studies. Some time ago, Lewontin (1974) and others in electrophoresis experiments observed an excess of low-frequency allozyme (protein) variants. More recently, many studies of DNA polymorphism in human populations have detected a relatively large number of nonsynonymous (protein-altering) polymorphisms segregating in human populations at low frequencies (Cargill et al., 1999; Halushka et al., 1999; Fay et al., 2001; Bustamante et al., 2005; Gojobori et al., 2007). For example, Figure 3 shows ratios of numbers of nonsynonymous to synonymous polymorphisms in different frequency classes from which it can be observed that, at relatively low frequencies (usually below 20%), there exists a relatively large proportion of protein-altering variants segregating in human populations (Fay et al., 2001; Gojobori et al., 2007). Neutral theory allows us to quantify the excess of slightly deleterious polymorphisms. Under neutrality, it is expected that the ratio of the number of nonsynonymous (Dn) to synonymous substitutions (Ds) between species will be proportional to the ratio of number of nonsynonymous (Pn) to synonymous polymorphisms (Ps) (i.e., DNA variants) within populations. If these fractions are calculated across the coding genome, the relative proportion of protein-altering polymorphisms within human populations ($Pn/Ps = 38.42\%$) is found to

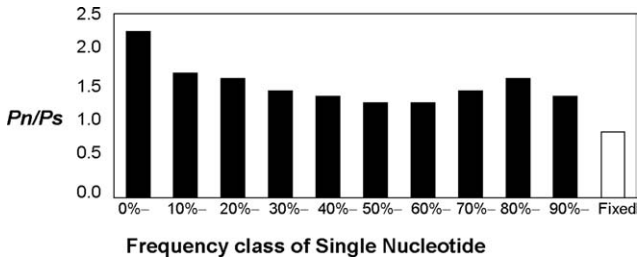

Fig. 3. The ratio of the numbers of nonsynonymous (amino acid-altering) and synonymous (non-amino acid-altering) polymorphisms are given for different frequency classes. As can be seen, the frequency class with the greatest proportion of nonsynonymous to synonymous polymorphisms lies at a frequency below 10%. For comparison, the proportion of nonsynonymous to synonymous fixed differences between chimpanzees and humans is shown in the right-most bar (Redrawn from Gojobori et al., 2007).

be considerably greater than the proportion of protein-altering substitutions between humans and chimpanzees (*Dn*/*Ds* = 23.76%) (Bustamante et al., 2005). The excess of slightly deleterious variants existing as polymorphism (38.42–23.76 = 14.66%) is similar to the estimate of ~15% by Eyre-Walker et al. (2002) (see above) and only slightly smaller than the estimate of ~20% by Fay et al. (2001). These polymorphisms rise to low frequencies within human populations via genetic drift though are ultimately prevented from rising to higher frequencies through the action of purifying selection. Further evidence that purifying selection is acting on these polymorphisms comes from studies by Hughes et al. (2003, 2005) in which large sets of polymorphisms are separated into two classes: those causing conservative amino acid changes and those causing radical amino acid changes. The more radical polymorphisms are found to segregate at significantly reduced frequencies compared with conservative polymorphisms, indicating that purifying selection acts more strongly on more radical polymorphisms (Hughes et al., 2003, 2005). Similar evidence derives from analyses by Lohmueller et al. (2008) of a large genome-wide database of single nucleotide polymorphisms (SNPs) (determined via the resequencing of 15 African-Americans and 20 European-Americans) in which SNPs inferred to be "probably damaging" were found to segregate at significantly lower frequencies than "benign" or "possibly damaging" SNPs.

More slightly deleterious variants in European than African populations: Effects of a population bottleneck and subsequent expansion in Europeans. Lohmueller et al. (2008) also found evidence of a significantly larger proportion of nonsynonymous SNPs in European populations compared with African populations, specifically when population-specific SNPs were considered (i.e., SNPs present only in Africans or only in Europeans). Their analyses of independent genome-wide data sets (e.g., the SeattleSNP data set and HapMap data) also supported the finding (see Lohmueller et al., 2008). Furthermore, of the nonsynonymous SNPs, Europeans were found to have significantly larger proportions, compared with Africans, of SNPs believed to be probably damaging. The researchers ran simulations under different demographic scenarios and found that an excess of damaging SNPs was generated under bottleneck scenarios (see Lohmueller et al., 2008). A likely explanation for the pattern is that the excess of slightly deleterious variants in the European population accumulated because of the reduced efficacy of purifying selection at the time of the bottleneck during the migration out of Africa (Lohmueller et al., 2008). Furthermore, the proportion of slightly deleterious alleles is expected to have further increased during the expansion subsequent to a bottleneck. One reason is due to the increased numbers of variants that accumulate in larger populations due to an increase in targets for mutations as well as the fact that most of these newly arising variants will be nonsynonymous because of the nature of the genetic code. Another process that can augment the proportion of deleterious variants in an expanding population is known as "gene surfing" (see Travis et al., 2007; Hallatschek and Nelson, 2009). This phenomenon is due to the effects of strong genetic drift acting on rare variants present within the small population at the leading edge of an expanding population (see Edmonds et al., 2004; Klopfstein et al., 2006). Travis et al. (2007) carried out simulations of a population expansion across Europe (modeling human expansions) and showed that slightly deleterious variants can indeed attain relatively high frequencies in the expanding population. In laboratory microbial experiments, Hallatscheck and Nelson (2009) found that deleterious mutations proliferated at expanding frontiers. In humans, this could have augmented the level of genetic load because of deleterious variants outside Africa.

Functional consequences of slightly deleterious variants. It is believed that slightly deleterious variants contribute to genetic diseases and to phenotypic differences between individuals within human populations (Fay et al., 2001; Bustamante et al., 2005). Because of their slight effects of fitness, one possibility is that some genetic diseases result from the combined effects of multiple very weakly selected polymorphisms (Fay et al., 2001). Although we see evidence of slightly deleterious variants segregating in human populations today, as discussed above, it seems likely that a relatively large fraction of slightly deleterious variants have been fixed by genetic drift in species that have relatively small effective sizes. Because N_e in humans is reduced compared with chimpanzees, it would be expected that there has been an augmentation in the fixation of slightly deleterious amino acid variants during human evolution. Furthermore, these slightly deleterious differences (between humans and chimpanzees) might very well have contributed to biological differences between humans and chimpanzees. The possible consequences of an increased fixation of slightly deleterious variants during primate evolution and more so in human evolution will be considered in the Discussion section. In the next section, however, I will discuss how the presence of a large class of slightly deleterious variants segregating within modern human populations might affect the accuracy of estimates of adaptive amino acid substitutions in the human lineage.

Positive selection on proteins

Quantifying levels of adaptive evolution. A number of studies have attempted to quantify levels of adaptive change in proteins as humans and chimpanzees diverged. One approach has used d*N*/d*S* comparisons between humans and an outgroup species (divergence

TABLE 2. Estimates of adaptive evolution during human evolution based on genome-wide studies of protein-coding genes

Research study	Total genes	Method	Species compared with humans	Adaptive evolution by positive selection (%)
Clark et al. (2003)	7,645	d*N*/d*S*	Chimpanzee, mouse	0.08[a]
Arbiza et al. (2006)	9,674	d*N*/d*S*	Chimpanzee, mouse/rat	1.12[a]
Mikkelsen et al. (2005)	13,454	MK	Chimpanzee, mouse	~0.0[b]
Nielsen et al. (2005)	8,079	d*N*/d*S*	Chimpanzee	0.4[a]
Gojobori et al. (2007)	5,008[c] and 5,535[d]	MK	Chimpanzee	10.4[b] and 12.8[b]
Fay et al. (2001)	182	MK	Old World monkeys	35
Boyko et al. (2008)	11,404	MK	Rhesus macaque	10–20
Bakewell et al. (2007)	13,888	d*N*/d*S*	Chimpanzee, rhesus macaque	1.1[a]
Zhang and Li (2005)	479	MK	Chimpanzee, Old World monkeys, mouse	~0.0[b]
Bustamante et al. (2005)	4,916	MK	Chimpanzee	~6.0[a]

[a] Percentage of genes that show evidence of positive selection.
[b] These values are the percentage of amino acid substitutions over all genes that show evidence of positive selection estimated using an McDonald–Kreitman (1991) approach applied over large sets of genes.
[c] Based on the SNP data set from Perlegen Biosciences (http://www.perlegen.com) used to obtain the first estimate of adaptive evolution in column five.
[d] Based on the SNP data set from the International HapMap Project (http://www.hapmap.org) used to obtain the second estimate of adaptive evolution in Column 5.

TABLE 3. Comparisons of estimates of the proportion of amino acid variants driven to fixation by positive selection in species with very large effective population sizes

Species	Total genes	Species used for comparison	Analysis type	Adaptive evolution (%)	Reference
Mus musculus castaneus	77	*M. famulus* and *Rattus norvegicus*	MK	57.0	Halligan et al. (2010)
Drosophila melanogaster	44	*D. simulans*	MK	45.0	Bierne and Eyre-Walker (2004)
D. simulans	115	*D. yakuba*	MK	41.0	Welch (2006)
D. melanogaster	91	*D. simulans*	MK	~95	Sawyer et al. (2007)
Escherichia coli	410	*S. enterica*	MK	>56.0	Charlesworth and Eyre-Walker (2006)
E. coli	410	*S. enterica*	MK	74	Charlesworth and Eyre-Walker (2008)
Salmonella enterica	410	*E. coli*	MK	65	Charlesworth and Eyre-Walker (2008)

data) to estimate the proportion of human genes that have experienced positive natural selection. As described above, d*N*/d*S* values significantly greater than 1.0 are considered to be evidence of positive selection. A second approach is based on the McDonald–Kreitman (MK) (1991) method that compares the amount of polymorphism within human species to the amount of divergence between species. Specifically, the method compares the ratio of the number of nonsynonymous to synonymous differences (*Dn*/*Ds*) between species to the ratio of the number of nonsynonymous to synonymous polymorphisms within species (*Pn*/*Ps*). When *Dn*/*Ds* is significantly greater than *Pn*/*Ps*, positive selection is inferred. Under neutrality, we expect *Dn*/*Ds* = *Pn*/*Ps*, whereas under purifying selection we expect *Pn*/*Ps* to be less than *Pn*/*Ps*. When the method is applied over the protein coding genome, it yields an estimate of the proportion of amino acids that have undergone adaptive evolution.

Results from both approaches (i.e. d*N*/d*S* and MK methods) are generally comparable and are given in Table 2. As can be observed, many estimates of positive selection in the divergence of humans from chimpanzees indicate very little evidence of adaptive evolution. However, some estimates range up to 10% or 20% (Boyko et al., 2008) or even up to 35% (Fay et al., 2001). There is reason, however, to believe the result of Fay et al. (2001) is an overestimate because divergence data and polymorphism data were drawn from different sets of genes and the polymorphism data came from a set of disease genes (see Eyre-Walker, 2006a). In majority, though, estimates of positive selection in the divergence of humans and chimpanzees seem to be relatively low. This is especially apparent compared with estimates of adaptive evolution in protein-coding genes in organisms with larger effective population sizes (Table 3). For example, estimates of adaptive evolution in various *Drosophila* species and in bacteria species are usually greater than 40% and in some studies range to as great as 95% (e.g., in a study of *D. simulans* by Sawyer et al., 2007). Aside from humans, the only other mammal for which an estimate of adaptive evolution in protein-coding genes is available is for the house mouse (*Mus musculus castaneus*), for which the estimate is 57% (Halligan et al., 2010). The N_e for wild populations of *M. musculus castaneus* is estimated to be about two orders of magnitude greater than N_e in humans, and more similar to that of *Drosophila*. Therefore, the higher estimate of adaptive protein evolution in this species is consistent with the idea that positive selection is more effective in larger populations.

Because estimates of adaptive evolution are relatively low in human evolution (Table 2), it seems that many of the protein-coding differences between humans and chimpanzees have resulted from random genetic drift acting on variants that were either neutral or slightly deleterious (but effectively neutral). Low estimates of adaptive evolution in human evolution may be a consequence of small effective population size in humans, and many studies have presented evidence that natural selection has been less effective in human evolution than in other mammal species (Eyre-Walker and Keightley, 1999; Keightley et al., 2005a,b; Bakewell et al., 2007).

Has adaptive change in proteins during human evolution been underestimated? It is possible that estimates of adaptive evolution have, in fact, been considerably underestimated (Charlesworth and Eyre-Walker, 2008; Eyre-Walker and Keightley, 2009). As described above, the most common method used to estimate the proportion of adaptive substitutions is based on the MK test (McDonald and Kreitman, 1991). However, the MK test can yield biased results when a class of slightly deleterious variants is segregating in the current population. This class of variants is affected by very weak purifying selection and, therefore, tends to segregate at relatively low frequencies in the population. Because these variants contribute more to polymorphism than to divergence in the MK test, they have the effect of making adaptive evolution harder to detect and can result in a considerable underestimate of adaptive evolution (Fay et al., 2001). As described in the previous section, analyses of human populations indicate that there is a relatively large class of slightly deleterious variants.

A commonly adopted approach to ameliorate the problem caused by slightly deleterious variants has been to exclude polymorphisms segregating at low frequencies because most of these are assumed to be slightly deleterious variants. For example, Fay et al. (2001) advocated removing polymorphisms below a frequency of 15%. However, Charlesworth and Eyre-Walker (2008) found that this method does not sufficiently reduce the underestimation of adaptive evolution. They pointed out at least two problems with the approach. First, the decision is arbitrary as to where to draw the frequency cutoff value. Second, the method fails to remove all slightly deleterious variants and, at the same time, removes some effectively neutral variants.

Two recent studies attempt to better account for the bias caused by slightly deleterious variants by using an approach in which the distribution of fitness effects (DFE) of new variants is estimated based on the full frequency spectrum of polymorphisms in the current human population (Boyko et al., 2008; Eyre-Walker and Keightley, 2009). Based on the frequencies of polymorphisms, estimates are made of the fraction of variants in classes with varying selection coefficients (i.e., neutral or nearly neutral variants, moderately deleterious variants, and highly deleterious variants). Because the frequency spectrum of polymorphisms can be skewed by recent demographic history (see Nielsen, 2001, 2005; Bamshad and Wooding, 2003; and Harris and Meyer, 2006 for a review), it presents a possible bias to estimates of the DFE. The methods used in Boyko et al. (2008) and Eyre-Walker and Keightley (2009) attempt to account for this problem by simultaneously using polymorphism data to infer recent human demographic history and then using the results to correct for demographic effects on the DFE. Next, the DFE is used to infer the expected numbers of fixed differences between humans and chimpanzees (i.e., substitutions) that originated as neutral or slightly deleterious variants (Boyko et al., 2008; Eyre-Walker and Keightley 2009). Expected values are then compared with the observed numbers of substitutions, and, if the observed number is greater, the difference is inferred to be the amount of substitutions driven to fixation by adaptive evolution. Using such a method, Boyko et al. (2008) inferred that 10–20% of amino acid differences between humans and chimpanzees were fixed by adaptive evolution (positive selection).

On the other hand, results from the recent study of Eyre-Walker and Keightley (2009) found little evidence of adaptive evolution in proteins in their comparison between humans and macaques, with values approximating zero. However, this result was quite possibly an artifact related to the effects of long-term changes in effective population size during human evolution and how these changes can bias the MK test (see Eyre-Walker, 2002; Eyre-Walker and Keightley, 2009; Halligan et al., 2010). For example, if the effective population size in the divergence phase of human evolution was larger than that in the polymorphism phase (e.g., in the case of a population contraction), then estimates of adaptive evolution are underestimated because slightly deleterious variants make a larger relative contribution to polymorphism than to divergence. On the other hand, if the effective population size in the divergence phase was smaller than in the polymorphism phase (e.g., the case of a population expansion), then estimates of adaptive evolution are overestimated because slightly deleterious variants become fixed in the divergence phase and make a larger relative contribution to divergence than to polymorphism.

Most demographic analyses have found that the effective population size of the human-chimpanzee common ancestor was considerably larger, possibly as much as 5–10 times larger than the current effective population size of humans (Burgess and Yang, 2008). Assuming that the effective population size during macaque and human evolution was of a similar large size and that the human effective population size became reduced only recently in human evolution, Eyre-Walker and Keightley (2009) calculated that the true value of adaptive amino acid substitutions could be as high as 31% or even 40% based on the assumption of a 5- or 10-fold population size reduction, respectively.

These recent estimates of adaptive evolution in human proteins by Boyko et al. (2008) and Eyre-Walker and Keightley (2009) could possibly bring values closer to the estimated values for mice or *Drosophila*. However, the higher values estimated by Eyre-Walker and Keightley (2009) depend on the accuracy of estimates of the long-term demographic changes along the human lineage and methods used to account for these changes.

Comparing numbers of positively selected genes. The overall numbers of positively selected genes (PSGs) estimated in humans can be compared with estimates of PSGs in other primates as well as in other mammal species (Table 4). Comparing humans with chimpanzees, several studies have found that humans have considerably fewer numbers of PSGs than in chimpanzees (Arbiza et al., 2006; Bakewell et al., 2007; Gibbs et al., 2007; Kosiol et al., 2008). The most extreme difference was estimated by Bakewell et al. (2007) who found that 51% more genes were positively selected along the chimpanzee lineage. Also, both humans and chimpanzees have been found to have reduced numbers of PSGs compared with rhesus macaques (*M. mulatta*) (Gibbs et al., 2007). These findings are most often attributed to population genetic theory that predicts positive selection to be less efficient at fixing beneficial variants in smaller populations compared with larger populations (Arbiza et al., 2006; Bakewell et al., 2007; Kosiol et al., 2008). However, such differences need to be verified for accuracy using higher quality genome sequences. For example, a recent analysis has suggested that larger estimates of PSGs in chimpanzees can simply result from base errors in the chimpanzee sequence, which is of lower quality

TABLE 4. Estimated numbers of positively selected genes in primate and nonprimate mammal species

	Study				
	Kosiol et al. (2008)	Kosiol et al. (2008)	Gibbs et al. (2007)[a]	Bakewell et al. (2007)	Arbiza et al. (2006)
Method	LRT[b]	Bayesian Inference[c]	LRT	LRT[d]	LRT
Humans	10.0	207.9	2	154 (2)	108
Chimpanzees	18.0	233.5	14	233 (59)	577
Macaques	16.0	340.9	131	–	–
Ancestral primate branch	21.0	360.5	–	–	–
Mouse	30.5	290.4	–	–	–
Rat	30.5	253.8	–	–	–
Ancestral murid branch	56.0	393.9	–	–	–
Dog	–	308.4	–	–	–

[a] Numbers of genes are those that met an false-discovery rate (FDR) <0.10.
[b] Likelihood ratio test as developed in Nielsen and Yang (1998) and Yang and Nielsen (2002). Numbers of genes are those that met an FDR <0.05.
[c] Bayesian inference model as described in Kosiol et al. (2008).
[d] Numbers in parentheses are numbers of PSGs after correction for multiple testing and that meet an FDR < 0.05.

than the human sequence (see Mallick et al., 2009). The d*N*/d*S*-based tests used to test for positive selection can be confounded by even small percentages of sequence errors, causing genes to artifactually seem to be PSGs (Mallick et al., 2009).

When numbers of PSGs in primates have been compared with numbers of PSGs in murids, considerably larger numbers of PSGs have been estimated in murids. This is true across studies that use the d*N*/d*S* test for each codon in a likelihood ratio test (LRT) for each branch (Nielsen and Yang, 1998) (Table 4). For example, compared with the ancestral primate branch estimated to have 21 PSGs, the ancestral murid branch was estimated to have 56 (Kosiol et al., 2008). Similarly, compared with the primate clade estimated to have 24 PSGs, the murid clade was estimated to have 61. On the basis of this method, murids seem to have two to three times greater numbers of PSGs, which again seems to be consistent with population genetics theory predicting increased efficacy of positive selection in species with larger N_e.

On the other hand, it is possible that the discrepancy in PSGs between primates and murids is an artifact. One of the problems is the lack of power due to the low interspecies divergence especially in the primate part of the tree. Power does not improve with the addition of multiple species' genomes, and it is possible, therefore, that the discrepancy between numbers of PSGs between primates and murids is due to differences in power to detect PSGs on different parts of the tree (see Kosiol et al., 2008). Kosiol et al. (2008) also used a Bayesian approach to estimate positive selection. The method pools information across genes with the consequence of combining weak information to potentially increase sensitivity and improve accuracy in detecting PSGs. Results showed considerably larger numbers of PSGs on all branches of the tree compared with the LRT method and yielded more equivalent numbers of PSGs along the different mammalian branches. Also, longer branches (e.g., the ancestral primate and murid branches as well as the macaque and dog branches) generally showed more PSGs than shorter branches (e.g., the terminal mouse, rat, human, or chimpanzee branches), indicating that the longer branches may provide greater power to detect PSGs than shorter branches. Despite these findings, the ancestral murid branch showed considerably higher gains of positive selective genes compared with the ancestral primate branch. As Kosiol et al. (2008) point out, this finding might indicate a real tendency toward greater positive selection in the ancestral murids and could be related to their larger effective population size compared with primates. Alternatively, it is possible that the difference relates to real differences in species-specific ecology whereby the specific ecology of murids necessitates increased biological adaptation.

When humans are compared with chimpanzees using the Bayesian approach (Kosiol et al., 2008), humans again showed fewer PSGs (~208) than chimpanzees (~234), although the 95% credible intervals for these estimates broadly overlapped, thus providing little statistical support for the difference. At present, it is unclear whether the marked differences found in LRT studies between these two species (and particularly in the analysis by Bakewell et al., 2007) reflect actual differences in numbers of PSGs. There is also the possibility that the discrepancy is due to difference in power between different methods or could be artifacts because of lower sequence quality in the chimpanzee genome sequence compared with the human sequence (see Mallick et al., 2009). If a real difference in numbers of PSGs exists between the two species, the future use of improved quality genome sequence, and the use of more sensitive methods to detect PSGs, should be able to detect this difference.

EVOLUTION OF GENE-REGULATORY REGIONS

It has been proposed that changes in gene expression are likely to play important roles in phenotypic evolution (King and Wilson, 1975; Carroll, 2003; Prud'homme et al., 2007; Wray, 2007). Although it is unknown exactly how much of the genome is dedicated to regulatory function, it is estimated that roughly 3.5% of the non-protein-coding portion of the genome, which is strongly conserved across species, is involved (Kim and Pritchard, 2007). Many regulatory regions lie in the sequence near genes they influence (Kim and Pritchard, 2007).

To investigate the extent to which purifying selection has constrained divergence within regulatory regions between humans and chimpanzees (hominids), Keightley et al. (2005b) compared the 6 kb of sequence lying upstream and downstream of genes, as well as sequence in the 5′ regions of first introns, for a large set of randomly selected orthologous genes between the two spe-

cies. The same comparisons were made between randomly selected genes in the mouse (*Mus domesticus*) and rat (*R. norvegicus*). The result of the comparison between the hominids and the murids were then compared. To test the null hypothesis of neutrality, they used modified d*N*/d*S* and MK methods that could be applied to noncoding regions. The methods were modified in the following ways: instead of using nonsynonymous sites as the class of putatively selected (as is done when the method is applied to protein-coding regions), they used presumed regulatory sites as the putatively selected sites; also, instead of using synonymous sites as the surrogate for neutrally evolving sites, they used sites within introns (though not those within the first intron). They made two important findings. First, levels of selective constraint (due to purifying selection) in the divergence of regulatory regions between the hominids were markedly reduced compared with the levels of constraint in the divergence between the two murid species. In fact, in their analysis, hominids (humans and chimpanzees) showed a near absence of selective constraints in the regions studied. Second, there was little evidence of positive selection having affected regulatory regions in the divergence between humans and chimpanzees. (However, see Haygood et al. (2007) who did detect signals of positive selection in regulatory regions in the divergence of humans and chimpanzees).

Bush and Lahn (2005) compared shorter segments of presumed regulatory sequence and also found reduced levels of constraint in hominids compared with murids, although the difference seemed to be less extreme (for further discussion see Keightley et al., 2006). Relaxed selective constraint in hominids compared with murids in presumed regulatory regions was also found in the study by Kim and Pritchard (2007) in their comparison of around 99,000 CNC regions. Taylor et al. (2006) studied numbers of substitutions and microindel events (<11 bp) in the regulatory sequence upstream of transcription start sites and found that rates of change were significantly higher within primates (humans, chimpanzees, and macaques), particularly in humans, compared with rates in mouse, rat, and dog, thus suggesting reduced selective constraint generally in primates and particularly in humans. Furthermore, Gaffney et al. (2008) studied the divergence of experimentally defined regulatory regions within three different primate species and found greatest levels of constraint in rhesus macaques, moderate levels in chimpanzees, and least levels of selective constraint in humans.

A common explanation for reduced selective constraint in primates, and hominids, relative to murids and other species is that primates have relatively smaller effective population sizes compared with other species (Keightley et al., 2005b; Kim and Pritchard, 2007; Gaffney et al., 2008). Keightley et al. (2005b) reasoned that the selective coefficients of most variants in regulatory regions must be very slightly negative. As discussed previously, the fate of such variants depends on the relative relationship between s, the selection coefficient, and N_e. If $s > 1/N_e$, then the variant will be removed by purifying selection. However, if $s < 1/N_e$, then the probability of fixation of the slightly deleterious variant increases. The selective coefficient of variants that were fixed in the divergence between humans and chimpanzees are assumed to fall within the range of $1/N_e$ (murids) $< s < 1/N_e$ (hominids), whereby they tend to be purged by purifying selection in murids, but can drift to fixation in chimpanzees and humans (see Keightley et al., 2005b). Thus, similar to protein-coding regions, within regulatory regions, there seems to be a considerable fraction of fixed differences between humans and chimpanzees that have slightly negative selective coefficients and that fall into an effectively neutral class because of reduced N_e in these species. Moreover, regulatory regions in primates generally seem to have been under less-selective constraint compared with other mammal species (mice, rats, and dogs). Although even within primates, evidence suggests reduced selective constraint in humans compared with chimpanzees and, furthermore, reduced constraint in chimpanzees compared with rhesus macaques. Such an effect within primates may be associated with differences in effective population sizes in different primate species. In the future, increased population genetic samples from a variety of primate species should allow more accurate estimates of species' effective population sizes. Then, comparisons between species should allow researchers to test how finely tuned the relationship is between effective population size, degree of selective constraint, and positive selection.

Evolution of gene expression

Because regulatory regions can modulate levels of expression of gene products, it seems reasonable for us to expect to see similar patterns when we examine gene expression divergence between humans and chimpanzees compared with that between mice and rats. That is, we might expect to see evidence of relatively reduced constraints on gene expression in the divergence of humans and chimpanzees. However, questions about the evolution of gene expression are not straightforward to answer. One reason is that the nature of expression data is qualitatively different from DNA sequence data. The evolutionary analysis of DNA sequences is facilitated by the fact that a well-developed body of theory exists, Kimura's Neutral Theory of Molecular Evolution (Kimura, 1968, 1969, 1983), which can be used as a basis for understanding sequence divergence between species and polymorphism within populations. No comparably well-developed theory exists for explaining gene expression evolution. However, recent work (Khaitovich et al., 2004, 2005, 2006) has found that much of the expression differences between species seems to be neutral or nearly neutral (i.e., has slight selective coefficients), but there is also considerable evidence for widespread purifying selection helping to constrain changes in gene expression. Although the application of a neutral model to expression data is not perfect, Khaitovich et al. (2006) argue that until adjustments are made a neutral model (based on Kimura's Neutral Theory) can serve as a useful null hypothesis against which observations of gene expression changes can be tested.

To assess levels of selective constraint, Keightley et al. (2005b) compared levels of gene expression differences between human and chimpanzee and between two different mouse species (*Mus musculus* and *M. spretus*) using data from Enard et al. (2002) on expression levels in liver and brain tissue. They found similar levels of expression divergence between each species pair. They then compared the level of gene expression divergence to the amount of nucleotide divergence between each species pair at presumably neutrally evolving sites and found that nucleotide divergence between humans and chimpanzees was about half that between the mice.

They found that, relative to nucleotide divergence, gene expression divergence is about 1.8-fold higher between the hominids than between the mice species. Thus, if it is true that gene expression evolution is generally under selective constraint (Khaitovich et al., 2004; Lemos et al., 2005; Gilad et al., 2006), they argued that an increase in expression divergence between humans and chimpanzees presumably indicates a relative relaxation of selective constraints on gene expression (Keightley et al., 2005b). It should be noted, however, that the phylogenetic distance between humans and chimpanzees (6–8 MY) (Patterson et al., 2006) is nearly four times that between the mice species compared (~1.8 MY; Khaitovich et al., 2004) and this difference could perhaps make direct comparisons between their expression divergences problematic. Keightley et al. (2005b) argued, however, that such a "saturation" effect would not affect the results because other primate species, at greater phylogenetic distances, showed relatively greater expression divergence.

Khaitovich et al. (2004) found that gene expression divergence between humans and chimpanzees for large sets of brain-associated genes is not significantly different from divergence between the two species in expressed pseudogenes (assumed to evolve neutrally). This was unexpected because selective constraint could easily be expected to be acting on the gene-expression levels since purifying selection is widely assumed under the standard neutral model (see Kimura, 1983). One possible explanation, and one that could indicate selective constraint on the gene-expression data, is that the set of pseudogenes might actually be functional. Thus, to some degree, it could be that the expressed pseudogenes themselves are under some level of selective constraint, a possibility that should be tested in comparisons among closely related species (Khaitovich et al., 2004).

The interpretation of neutral divergence in gene expression divergence should be considered tentative until further studies permit the development of more robust models. One prediction of the neutral model that does not seem to be consistent with expression data is that the level of divergence between two species will be proportional to the phylogenetic distance between them. However, one might reasonably assume that there is a limit to the degree to which gene expression can diverge between two species. After a certain degree of divergence, further increases in expression divergence become untenable. Thus, although gene expression differences among recently diverged species might generally evolve by neutral forces (i.e., genetic drift and purifying selection), as genetic distances become greater as two species diverge, expression divergence might become increasingly constrained because of purifying selection (see Khaitovich et al., 2006; Whitehead and Crawford, 2006).

In sum, there seems to be evidence of relatively relaxed constraint in gene expression divergence between humans and chimpanzees compared with that between pairs of mice species. This evidence is consistent with the reduced levels of constraint observed in protein-coding and regulatory regions of the genome. However, without a good model of neutral evolution of gene expression, it is difficult to accurately assess this hypothesis. Future research on gene expression divergence between and within species should allow the formulation of more accurate models of gene expression evolution that take into account the specific ways in which gene expression evolution is different from DNA evolution. One worthwhile area of research would be to extensively test levels of gene expression divergence in larger numbers of closely related species separated by similar phylogenetic distances (facilitating their direct comparison) and among species that have different effective population sizes.

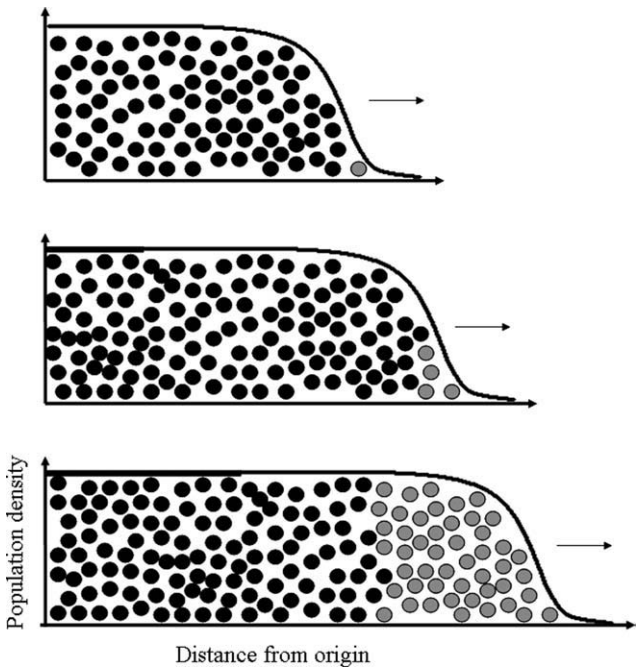

Fig. 4. The three illustrations from top to bottom represent the expansion of a population in the direction of the arrow. In the upper illustration, a new mutation (gray dots) is observed to arise at the wave front. This mutation has a greater chance to "surf" to higher frequencies and obtain relatively high frequencies in the descendent populations (as seen in the two lower illustrations) than does a mutation that occurred within populations in the occupied area (black dots). See text for details and supporting references.

Nonadaptive "surfing" in spatially expanding populations

Range expansions have occurred, sometimes recurrently, in the histories of most species. Despite this fact, the genetic consequences of spatial expansions has only recently come under study especially in relation to recent human expansions (Edmonds et al., 2004; Klopfstein et al., 2006; Travis et al., 2007). It has been shown in simulation studies that neutral variants that arise on the edge of a range expansion can potentially "surf" on the wave of advance, reaching large spatial distributions and much higher frequencies in newly colonized regions than would be expected under stationary conditions, and without positive selection (Edmonds et al., 2004; Klopfstein et al., 2006). Recently, Hallatschek and Nelson (2008) have provided experimental evidence of the surfing process in laboratory studies of expanding microbial populations. The surfing process is illustrated in Figure 4 in which a variant that initially arises on the wave front of spatial expansion ultimately goes to fixation in the newly colonized area.

The surfing phenomenon is due to the effects of strong genetic drift acting within the small populations located at the leading edge of the expansion. The genetic compo-

sition of the populations on this leading edge determines the genetic diversity that is propagated in a forward direction into new territories. Non-neutral variants, including both advantageous and deleterious variants, are also capable of surfing (Travis et al., 2007), though surfing in the case of disadvantageous variants occurs much less frequently than for the other kinds of variants. As described earlier, the surfing and promotion of deleterious variants is a result of reduced purifying selection in the population at the wave front followed by the propagation of the deleterious variant into the newly colonized geographic regions.

Recent large-scale studies of genetic variation have supported a model of serial founder effects on genetic diversity as humans colonized the world (Prugnolle et al., 2005; Ramachandran et al., 2005; Liu et al., 2006; Handley et al., 2007; Li et al., 2008; DeGiorgio et al., 2009; Deshpande et al., 2009; Hunley et al., 2009). Surfing is likely to have occurred during the repeated bottleneck events followed by spatial expansions and is likely to have had considerable effects on the genetic diversity of modern human populations worldwide. Recently, Excoffier and colleagues (Currat et al., 2006; Hofer et al., 2009) have shown that nonadaptive surfing can produce markedly high measures of genetic differentiation between different geographic populations, a pattern traditionally interpreted as evidence of local adaptation (Cavalli-Sforza, 1966; Lewontin and Krakauer, 1973; Beaumont and Balding, 2004). The F_{ST} statistic (Wright, 1951), and tests based on this statistic, are the most popular statistics used to measure population differentiation. These measures have featured prominently in both locus-specific studies and genome-wide scans for loci under positive selection (see Akey et al., 2002; Kayser et al., 2003; Evans et al., 2005; Mekel-Bobrov et al., 2005; Soranzo et al., 2005; Xue et al., 2006; Barreiro et al., 2008;Myles et al., 2008). Furthermore, many locus-specific studies that have detected large degrees of population differentiation have focused on adaptive explanations rather than the possibility of nonadaptive demographic processes (see Table 1 in Hofer et al., 2009 for a list of such studies). To examine the extent to which demographic processes explain cases of marked geographic differentiation, Hofer et al. (2009) analyzed genome-wide data sets in the 53 worldwide populations represented in the HGDP-CEPH Diversity Panel (Cann et al., 2002) consisting of three different classes of molecular markers: short tandem repeats (STRs), SNPs, and insertion–deletion polymorphism. For all three classes of markers, a very considerable fraction of these (about one-third of loci) were found to show strong differences in gene frequencies between continents with either very narrow or wider clines (i.e., gradients in gene frequencies). [Such clines in gene frequencies are usually attributed to environmentally varying selection pressures on populations (see Endler, 1973) or to demic diffusion (Cavalli-Sforza et al., 1993) but can arise via nonadaptive demographic surfing (Excoffier and Ray, 2008; Excoffier et al., 2009).]. Hofer et al. (2009) believe that the large numbers of loci that show marked geographic differentiation are unlikely to have resulted from the effects of positive selection but are better explained by demographic factors, particularly surfing. If surfing was a major driving force for these allele frequency differences, Hofer et al. (2009) reasoned that STR alleles (i.e., short DNA repeat sequences that are 2–16 bp in length and presumably neutral) should show increased rather than decreased frequencies outside Africa and into Eurasia, East Asia, and the Americas (assuming a major Out of Africa migration) because surfing promotes an increase in the frequency of initially low frequency variants. This pattern, of increasing frequency with distance from Africa, was found for the majority of STR loci and, thus, was interpreted as supporting the surfing explanation (Hofer et al., 2009). Chiaroni et al. (2009) have hypothesized that the surfing phenomenon also explains the spatial distribution of some Y chromosome variants.

Nevertheless, local adaptation must explain some fraction of loci showing extreme geographic differentiation. Some well-known examples of selected loci, for which a link to phenotypic variation is known, include the large frequency differences between populations for lactase persistence variants (Bersaglieri et al., 2004; Tishkoff et al., 2007), for the Duffy blood group negative variant (Hamblin and DiRienzo, 2000; Hamblin et al., 2002), for variants in skin pigmentation genes as well as in various malaria resistance variants (see Novembre and DiRienzo, 2009). Neutral allele surfing, also, will only occur at a subset of loci for the reason that it is governed by the stochastic effects of random genetic drift acting within the population leading the expansion. On the other hand, the effects due to general demographic phenomenon (e.g., expansions, bottlenecks, and inbreeding) will be evident at the majority of loci in the genome. Because positive selection and gene surfing affect a minority of loci genome-wide, it is a particularly troublesome problem, and it will be crucially important to learn how to successfully disentangle the two. One way to help distinguish the two is to use more complex and spatially explicit demographic models in simulation studies to determine accurate distributions of values of geographic differentiation that can be expected under neutral scenario. Using such simulations, Currat et al. (2006) showed that the high degree of differentiation at *ASPM* and *MCPH1* between Africa and Non-Africa (initially claimed to be a result of positive selection in non-Africa populations; see Evans et al., 2005; Mekel-Bobrov et al., 2005) could have been generated merely by surfing. Excoffier et al. (2009) has shown that when only simple geographic scenarios are simulated these will lead to a relatively high number of false positive loci (i.e., loci that show levels of population differentiation that fall outside the null distribution and that are, therefore, assumed to result from positive selection). Simulations derived using more complex models lead to a much smaller number of significantly outlying loci (Excoffier et al., 2009).

The problem presented by gene surfing is particularly pernicious because the signature and pattern of variation left at a locus after a surfing event is very closely similar to the signature left after a selective sweep (Nielsen et al., 2007; Excoffier and Ray, 2008). Beyond the marked degree of population differentiation (as measured by F_{ST}), surfing can also increase linkage disequilibrium producing unusually extended haplotypes (Nielsen et al., 2007; Hofer et al., 2009). The presence of extended haplotypes is a signature used often in scans for positive selection across populations (e.g., Sabeti et al., 2006; Voight et al., 2006; Pickrell et al., 2009). It is not clear whether recent tests that scan simultaneously for multiple different signatures of selection (i.e., highly differentiated alleles, extended haplotypes, and high frequency derived alleles), like the composite of multiple signals (Grossman et al., 2010), can be confounded by the nonadaptive gene surfing phenomenon.

Although gene surfing because of spatial expansions may complicate the detection of positive selection by genome researchers, the phenomenon may have contributed to the evolution of innovative genotypes. Research findings by Burton and Travis (2008) suggest that mutation surfing can lead to fitness peak shifts at the front of a spatially expanding population. Later in the Discussion section, these findings will be described and discussed in the context of human evolution.

GENOMIC MUTATIONS

Lynch (Lynch and Conery, 2003; Lynch, 2007a) showed that there are vast differences in genome size and architecture between prokaryotes and eukaryotes that can be generally correlated with their differences in N_e. Prokaryotes have population sizes that are orders of magnitude greater than that of eukaryotes, and prokaryotes tend to have much more smaller and more streamlined genomes than eukaryotes. Increase in genome size is attributable to overall greater numbers of genes within the genome, the presence or absence of introns within genes, and the accumulation of mobile genetic elements and segmental duplications, all of which eukaryotes show. Lynch (2007a) has argued that each of these genomic embellishments are slightly deleterious and that purifying selection has tended to remove them from the population, thus maintaining the generally streamlined genomes of prokaryotes. In contrast, Lynch (2007) argues that decreased purifying selection and increased influence of random genetic drift in eukaryotic species, with vastly smaller N_e, has allowed the passive accumulation of genomic embellishments.

Thus far, most research has focused on genome size and architectural differences between groups of organisms with vast differences in N_e. It is unclear, however, to what extent differences in genomic architecture occur between species that are closely related to each other and that have much smaller differences between them in effective population sizes. However, one recent study compared 1,043 ray-finned fish species (representing approximately 190 families) and found genome size to be negatively correlated with N_e (Yi and Streelman, 2005), supporting the relationship between N_e and genome size seen between prokaryotes and eukaryotes. However, other researchers have noted that at least some data may not fit the pattern. For example, Vinogradov (2004) has presented data indicating that genome size in herbivores is smaller than that of carnivores, even though herbivores are usually assumed to have relatively larger effective population sizes compared with carnivores because they are on a lower trophic level. However, Lynch and Conery (2004) argue that population dynamics are fundamentally stochastic, and, therefore, a significant scatter around the general trend is expected (Lynch and Conery, 2004). They also suggest that assumptions concerning N_e in many species may be inaccurate (Lynch and Conery, 2004), especially because our understanding of levels of genetic diversity in many species is limited.

Lynch (2007a) has raised the hypothesis that a relative reduction in N_e in the course of primate and particularly human evolution has led to the passive accumulation of genomic elements. Below, I review briefly the theory that explains how relaxed selective constraints (purifying selection) due to small N_e can lead to relative accumulation of genomic features, specifically segmental duplications, mobile genetic elements, and *numts* (mitochondrial inserts into the nuclear genome). I then review some of the recent findings that generally indicate relative increases in the proportions of these elements in the genomes of anthropoid primates, hominoids, and along the human lineage.

TABLE 5. Transposable elements in human genome (from Lynch, 2007)

Class	Number of copies	Fraction of genome (%)
LINEs	868,000	20.4
SINEs (including primate specific-*Alu* elements)	1,558,000	13.1
LTR elements	443,000	8.3
Transposons	294,000	2.8

Transposable genetic elements

Approximately 44–46% of the human genome is comprised of transposable genetic elements that have the capacity to self-proliferate throughout the genome through internal drive-like mechanisms (Mills et al., 2006; Lynch, 2007a). Lynch (2007a) has described the vast proliferation of these elements in eukaryotic organisms. The proportions of the major classes of these elements in the human genome are given in Table 5. The internal drive-like aspect of TEs is due to the fact that many contain genes that encode an enzyme for reverse transcription and integration into the genome. These include the long interspersed elements (LINEs), approximately 6 kb in length, and the LTR (long terminal repeat) elements. Other TEs are nonautonomous in that they do not contain their own machinery for self-replication. Instead, they proliferate by hijacking the replication machinery of other TEs (e.g., LINEs). These nonautonomous TEs include the short interspersed elements (SINEs), the most common being the *Alu*s, as well as the SVAs, named for their composite nature being comprised of a SINE-R element, a variable nucleotide repeat, and an *Alu*. Vastly fewer SVAs (~5,000; see Weber, 2006) are present in the human genome compared with LINEs (~868,000) and SINEs (~1,558,000; Table 5).

Lynch (2007a) provides a detailed description (briefly summarized below) of the population conditions that influence the capacity for accumulation of these elements within the genome. For persistence of a TE within a population, it is required that the TE be able to produce a descendent element in the genome before any of four possible events are able to eliminate it: 1) elimination of the TE through the acquisition of inactivating mutations; 2) elimination of the TE via genomic deletion; 3) elimination of the TE through the deleterious effects caused by TE induced-ectopic recombination (see Langley et al., 1988) or 4) elimination of the TE through the action of purifying selection. One critical factor is the effect the element's insertion has on fitness of the host. If an element is very deleterious, then it will be removed immediately by purifying selection. However, many TEs may have only slightly deleterious effects and will behave as effectively neutral variants in sufficiently small populations. Under these conditions, the probability increases that the TEs will become fixed in the population merely through the effects of random genetic drift. As discussed previously, for mutations to behave effectively neutral within a population then s needs to be less than $1/4N_e$. With respect to TE insertion, s represents the proportional decrease in fitness resulting from the

TABLE 6. Proportion of different types of transposable elements (TEs) in the human and chimpanzee genomes[a]

TE class	Human	Chimpanzee	Total (both species)
Overall number	7,786 (72.6%)	2,933 (27.4%)	10,719
Alu elements	5,530 (77.1%)	1,642 (22.9%)	7,172
LINEs (L1)	1,174 (60.8%)	758 (39.2%)	1,932
SVA	864 (68.6%)	396 (31.4%)	1,260
Others (e.g., LTRs and other elements)	219 (63.6%)	127 (36.9%)	344

[a] Data from Mills et al. (2006).

element's insertion. According to Lynch (2007a), the persistence of a mobile element within a population (R) is dependent on the expression: $R = \mu p(N_e)/v$, where μ is the rate of insertion per generation; v is the rate of removal by genomic causes (i.e., nonselective processes); and $p(N_e)$ is the fraction of insertions that are effectively neutral conditional on the N_e of the population. In the case where $R \geq 1.0$, TEs will persist in the genome because the rate of insertion of new mutations is greater than the rate of loss of old elements by genomic processes (e.g., deletion, or inactivation) or by purifying selection.

Several studies have detected increased numbers of TEs in humans compared with other primates (Liu et al., 2003; Hughes et al., 2005; Mills et al., 2006). By comparing the human and chimpanzee genomes, Mills et al. (2006) identified 10,719 mobile genetic insertions that were presumed to be specific to either lineage. A large proportion of these (N = 7,786 or 72.6%) were believed to have been inserted specifically during human evolution, whereas only 2,933 (27.4%) were believed to have been inserted in the evolution of chimpanzees. The disparity in proportion of lineage-specific TEs was found for each of the different classes (Table 6). For *Alus*, humans contain 3.4 times the number found in chimpanzees. For LINEs (L1s), humans show 1.5 times the amount compared with chimpanzees and for both SVAs and for other types of TEs (LTRs, etc.) humans showed an almost twofold greater numbers. Beyond the considerably larger numbers of TEs that seem to have been inserted during human evolution, there are considerable differences in the array of subtypes, at least for *Alu*s and LINEs between the two species (Mills et al., 2006).

Segmental duplications and gene duplication

Within the human genome, there are numerous blocks of highly homologous duplicated sequence that are very large in size, often over 20 kb in length. Duplicated segments can potentially have enormous impacts on the evolutionary process because they can often contain entire genes and can, therefore, increase the number of duplicated genes (i.e., copies of genes). Is there any evidence for an increase in rate of segmental duplication in nonhuman primate evolution and along the human lineage? Cheng et al. (2005) identified segmental duplications of sizes >20 kb in length in the human and chimpanzee genomes and found a twofold increase in humans. More recently, Marques-Bonet et al. (2009b) analyzed a greater number of primate species (i.e., macaque, orangutan, gorilla, and human) and found significantly increased numbers of duplication events as well as amounts of overall base pairs duplicated during hominoid evolution. Duplications specific to the human lineage were estimated to total ~10 Mb comprising 210 duplication intervals with an average length of 53.1 kb. In terms of megabases duplicated, this was several-fold larger than duplications believed to be specific to the chimpanzee, orangutan, or rhesus macaque lineages. They inferred that the most significant bursts of duplication activity (four to 10-fold increases) occurred in the ancestral lineages leading to the African Great Apes and humans with an initial burst occurring before the divergence of the gorilla and then another burst occurring in the ancestral lineage leading to chimpanzees and humans. Although other causal factors are likely at work, one possible explanation of the fixation of these segmental duplications, even if the duplications are slightly deleterious, is a reduction in the effective population size in primates, especially in apes and humans (see Marques-Bonet et al., 2009a,b).

Studies of gene duplication in mammals generally have found increased rates of gene duplication in primates. In a recent study, Hahn et al. (2007; see Demuth and Hahn, 2009) found that duplication rates in the primate lineage (0.0024/gains and losses/gene/year) is nearly twofold greater than the rate inferred for lineages leading to dogs, mice, or rat (0.0014). They inferred a further acceleration in the great ape lineage (0.0039), which is an almost threefold increase over the rate in nonprimate mammals (Demuth and Hahn, 2009).

Lynch and coworkers (Force et al., 1999; Lynch and Force, 2000; see Lynch, 2007) have developed a model that could explain the increased preservation of duplicated genes under conditions of reduced selective constraint and increased random genetic drift in populations with small effective sizes. Because their model does not invoke positive selection as the causal mechanism of fixation of duplicated genes, the model is a nonadaptive one. In their model called Duplication–Degeneration–Complementation (DDC), after the duplication of a gene that has multiple functions (a common aspect of genes), degenerative mutations accumulate in the regulatory regions of each gene copy that cause each copy to develop complementary expression patterns specific to different tissues. Because the ancestral functions are now partitioned between the two copies, and both functions are biologically required, both copies are fixed in the population. Force et al. (1999) describe several examples of gene duplications that are consistent with evolution via the DDC process, including duplicate engrailed genes in zebrafish, the *ZAG1* and *ZMM2* gene pair in maize, and the *Hoxa1* and *Hoxb1* genes in mice. A similar model of subfunctionalization (SF) was introduced earlier by Hughes (1994), but he emphasized the accumulation of mutations in the gene's protein-coding regions, resulting in the copies gaining different subfunctions. Lynch and Force (2000) have developed theory showing that duplicate preservation by the DDC process becomes more likely in populations smaller than about 10^5. In contrast, in very large populations (e.g., over 10^5 or 10^7), natural selection starts to play a more influential role, and the probability of fixation of a gene copy decreases, and the time to fixation is prolonged.

Mitochondrial insertions into the nuclear genome

numts (nuclear mitochondrial sequences) are partial copies of the mitochondrial genome ranging in length from >100 bp up to 16 kb that are found in abundance

in the nuclear genome. *numts* are common in mammalian genomes, and their evolution has been studied recently in primates (Gherman et al., 2007). Compared with mice and rats, which have roughly 636 and 529 *numts*, the human genome has been estimated to contain approximately twice this or ~1,200 *numt* copies (Gherman et al., 2007). When homologous *numt* sequences are compared across a phylogenetic sample of primate species, a nonuniform accumulation of *numts* is observed with a surge in accumulation of these elements between 45 and 55 million years ago (Gherman et al., 2007), a time associated with the evolution of and early diversification of anthropoid primates (Steiper and Young, 2006; Gherman et al., 2007). A previous study (Hazkani-Covo et al., 2007) also found a large relative increase in *numts* at a similar time in primate evolution, specifically in the ancestral lineage to Old World monkeys and apes (Catarrhines).

Interestingly, the temporal phase of *numt* accumulation in primate evolution is roughly coincident with *Alu*s subfamily accumulation (Bailey et al., 2003). Gherman et al. (2007) hypothesized that because the insertion mechanisms of *numts* and *Alu*s are different, their independent accumulation at roughly the same time may be accounted for by a general population-level phenomena. They suggested that reduced N_e during early anthropoid evolution could have provided the generally permissive conditions for the independent accumulation of these different elements by random genetic drift (Gherman et al., 2007). (However, it should be noted that at present no good evidence exists concerning the effective population size of these primates.) Two recent studies examined *numt* insertions along the human and chimpanzee lineages (Hazkani-Covo and Graur, 2007; Jensen-Seaman et al., 2009). Both studies found greater numbers of *numts* inferred to have inserted specifically in the chimpanzee lineage (46 and 66 *numts* in the two studies, respectively) compared with the human lineage (34 and 37 *numts*). This finding does not fit expectations of an increase in accumulation of *numts* with decreased long-term population size in humans compared with chimpanzees and may point to other factors playing important roles in *numt* insertion and fixation at least in these lineages.

Summary of genomic mutations

Reasonable evidence exists of an augmentation in genomic events in the evolution of nonhuman primates and humans, including segmental duplications, TE activity, and *numt* insertion. It is possible that reduced N_e in primates has generally allowed for the passive accumulation of these elements. Nevertheless, the association between N_e and the accumulation of elements does not seem to be perfect. For example, with respect to *Alu*s, Liu et al. (2003) have found that rates of *Alu* accumulation in different primate lineages are extremely variable, and that mechanistic influences such as insertion site availability, degree and efficiency of reverse transcriptase activity, and other factors play important roles in determining their accumulation (Liu et al., 2003). Also, Bailey et al. (2003) examined junctions between duplicated segments of the genome and the nearby sequence into which they were inserted and found a significant enrichment of *Alu* elements near or within junctions. They argued that a process known as *Alu–Alu*-mediated recombination is a possible mechanism producing segmental duplication within the genome. They suggest that increased *Alu*s subfamily accumulation between 35 and 40 million years ago could have seeded the genome landscape for segmental duplications whose rate increased after this time. Thus, it is possible that increased segmental duplication rates in the hominoid evolution (as described above) and *Alu* accumulation rates are not independent of each other and could, to a presently unknown degree, be causally connected.

These kinds of evidence likely indicate that multiple causal mechanisms are involved in the accumulation of genomic elements and architectural changes within primate genomes. Despite various causal mechanisms at work, reduced N_e and relaxation of selective constraints would seem to ultimately play an important role. Teasing out the exact and probably multiple causal factors explaining the relative concentrations of different genomic elements in different species lineages will be one of the major challenges for comparative genomic research in the future.

DISCUSSION

In the preceding pages, I have described the proximate effects that resulted presumably from the reduced efficacy of natural selection in primate and human evolution because of reductions in effective population size. For example, there is evidence that reduced purifying selection has led to an augmentation in the proportion of slightly deleterious substitutions in protein-coding and regulatory regions of the genome and has also probably enabled increased accumulations of segmental duplications, mobile genetic elements (*Alu*s, LINEs, etc.), and possibly *numts* in nonhuman primate and human evolution. On the other hand, reduced positive selection is evidenced in the generally smaller numbers of genes, and the fewer amino acid substitutions inferred to have been under positive selection in primate and human evolution compared with the evolution of other animal species. Even within primates, there is some evidence for reduced positive selection in species with relatively smaller effective population sizes. In the following, I would like to explore what may have been the evolutionary consequences of a relative reduction in the efficiency of natural selection on primate and human evolution.

Slightly deleterious substitutions and compensatory substitution

There is approximately a 1% difference in nucleotide composition between chimpanzees and humans. Assuming about 3 billion base pairs in the total genome, this means there are about 30 million fixed nucleotide differences between the two species. Many of the substitutions separating chimpanzees and humans are neutral with inconsequential effects on fitness, but reduced purifying selection has undoubtedly led to the fixation of a sizeable fraction of nearly neutral mutations having slight negative impacts on fitness. Eyre-Walker et al. (2002) estimated that the fraction of slightly deleterious substitutions fixed in the divergence of human and chimpanzees may be as large as 15% (i.e., 15% of all substitutions fixed between the two species) and that these substitutions have selection coefficients smaller than 10^{-5} (Eyre-Walker and Keightley, 2007). Because the effective population size of humans is estimated to be about 5- to 10-fold smaller than the common chimpanzee-human ancestral species (Burgess and Yang, 2008) and about twofold

smaller than the N_e estimated for chimpanzees (Caswell et al., 2008), we can assume that a larger percentage of slightly deleterious substitutions have accumulated along the human lineage than along the chimpanzee lineage.

What is the evolutionary significance of an increase in fixation of slightly deleterious substitutions in human evolution? Functional significance depends on where substitutions are located in the genome. About 1.5% of the human genome is comprised of protein-coding genes (Lander et al., 2001), whereas at least as much of the genome is likely to be involved in the regulation of these genes (Ponting and Lunter, 2006). It is expected that both coding and regulatory regions have accumulated slightly deleterious substitutions that would have affected phenotypic evolution. In protein regions, slightly deleterious substitutions would have caused slightly suboptimal amino acids within proteins. Few research projects have focused on analyzing the evolutionary and functional consequences of slightly suboptimal amino acids. A number of studies have, however, estimated a very high deleterious mutation rate in humans and chimpanzees (Kondrashov, 1995; Eyre-Walker and Keightley, 1999) prompting Kondrashov (1995) to discuss this in an article entitled: "Contamination of the genome by very slightly deleterious mutations: why have we not died 100 times over?" One possible explanation he gives that would allow survival despite a high genetic load is if multiple slightly deleterious mutations are able to combine their effects in epistatic ways so that although they may be separately deleterious their joint effects are much less deleterious, neutral, or even beneficial. Kimura (1991) has shown that such double mutants, which restore fitness through compensation, can fix relatively rapidly within populations by random genetic drift alone (although positive selection could also potentially operate in their fixation). Even under drift alone, Kimura's (1991) estimates put fixation times on the order of tens to hundreds of thousands of years, and, so, these times are relevant to the time frame of human evolution. He suggested that such a nonadaptive process may potentially be a way by which populations can quickly move from one adaptive peak to another equally adaptive peak by crossing a valley of low fitness merely by random genetic drift and mutation.

There are several experimental studies on bacteriophage (T7, Φ174, and Φ6), *Salmonella typhimurium*, and *Caenorhabditis elegans* (reviewed in Estes and Lynch, 2003; Davis et al., 2009) that show that the evolution of compensatory mutations can help populations regain fitness after they have accumulated deleterious mutations. Several examples of deleterious mutations that behave in such a way are given in Camps et al. (2007). Results from these studies suggest that as mean fitness declines because of the fixation of slightly deleterious mutations, the proportion and magnitude of beneficial mutations increases (see Lenski and Travisano, 1994; Burch and Chao, 1999; reviewed in Whitlock, 2000). Therefore, it is possible that, as populations decrease in size, there will be an increase in the rate of fixation of compensatory mutations. The rationale would be that when a biological molecule (e.g., an enzyme) is performing less than optimally, there are more opportunities to improve it from its current suboptimal performance than to improve on its optimal performance (see Whitlock, 2003). At the same time, it should be kept in mind that, in small populations, two phenomena occur that could constrain recovery via compensation: in small populations, there are fewer potential targets for beneficial mutations and positive selection becomes less effective. Thus, in very small populations (smaller than the N_e estimated for humans), it becomes difficult for populations to recover from the accumulation of deleterious mutations. Several researchers have estimated this threshold N_e to be around several hundred individuals and have pointed out the importance for conservation of determining the so-called critical N_e for an endangered species (Schultz and Lynch, 1997; Poon and Otto, 2000; Whitlock, 2000). Nevertheless, a return to larger population size seems to be important for recovery. Estes and Lynch (2003) found in studies of the nematode *Caenorhabditis elegans* (mutationally-degraded in the laboratory) that recovery by compensatory mutation to initial mean fitness levels occurs rapidly but only after the population returns to population genetic conditions that favor the efficacy of natural selection (i.e., larger sizes). Therefore, any fitness decline that the human species may have acquired as a result of relaxed purifying selection may only now (and in the future) become recoverable from in vastly larger population sizes. This is because population growth in recent human evolution (beginning roughly 50,000 years ago; see Hawks et al., 2007) has increased the number of targets that can acquire compensatory mutations. Increased population size has also augmented the effectiveness by which positive selection can fix compensatory mutations. [Further discussion of evolutionary processes acting in recent and future human evolution can be found in Hawks et al. (2007), Reed and Aquadro (2006), and Premo and Hublin, (2009).]

How common is compensatory evolution? Through experimental studies, Poon et al. (2005) estimated that for every fixed deleterious mutation there are on average approximately 12 possible yet different compensatory mutations that could recover fitness. Furthermore, the suppression of the negative effects of just one deleterious mutation may require multiple compensatory mutations, especially if deleterious effects are pleiotropic (DePristo et al., 2005). In another study of compensatory evolution, Kondrashov et al. (2002) compared 32 different proteins across nonhuman mammalian species and searched within them for any known pathogenic mutations in humans. In cases where a human pathogenic mutation was found in another species, these mutations were called compensated pathogenic deviations because a compensatory mutation must have allowed the deleterious mutation to have become fixed. In total, roughly 10% of deviations of a nonhuman protein from the human protein were found to be compensated pathogenic deviations, which also indicates that compensatory evolution is relatively common.

Given that humans have presumably accumulated a relatively large fraction of slightly deleterious mutations, it is possible that humans show increased fractions of compensating amino acid substitutions. However, it is unknown at present how changes in effective population size impacts the rate of compensatory evolution. Studying *Drosophila*, Kulathinal et al. (2004) also found a relatively large fraction of amino acid deviations to be compensated, the fraction being closely similar to the 10% found by Kondrashov et al. (2002) for mammals. This was unexpected, however, because increased effectiveness of purifying selection in *Drosophila*, as a result of large N_e, is expected to reduce the length of time that deleterious mutations will segregate in the population. Therefore, we might expect there to be less time for the

evolution of compensatory changes. Kulathinal et al. (2004) raised the hypothesis that compensatory evolution may be independent of population size, though the idea needs rigorous testing.

If, in human evolution, there has been an increase in compensatory evolution in response to a relative increase in the fixation of slightly deleterious mutations, there are implications for studies aimed at identifying genes that have experienced positive selection in human evolution. In recent years, many studies have scanned the human genome for putative PSGs turning up relatively large sets of candidate genes (Clark et al., 2003; Bustamante et al., 2005; Nielsen et al., 2005; Arbiza et al., 2006; Bakewell et al., 2007). In many of these studies, positive selection is inferred when there is evidence that the rate of substitution at nonsynonymous sites has been greater than the rate of substitution at synonymous sites (e.g., d*N*/d*S* tests). Hughes (2007) has pointed out, however, that such a signature may result through the fixation by genetic drift of one or several slightly deleterious mutations followed by the fixation of several compensatory mutations. The fact that compensatory mutations in proteins tend to be physically close to the deleterious mutation with which they interact (Kondrashov et al., 2002; Poon et al., 2005; Camps et al., 2007; Davis et al., 2009) makes encountering such a signature more probable. Therefore, it is possible that some genes detected in studies of positive selection (using the method described above) could be cases whereby a protein is merely evolving to attain functional equivalency to an initial condition, a situation Hartl and Taubes (1996) has described as "selection without adaptation." Note that compensatory substitutions in regulatory regions could also confound tests of selection that use methods similar to the d*N*/d*S* method but modified for use in regulatory regions (e.g., Wong and Nielsen, 2004; Haygood et al., 2007).

Furthermore, Hughes (2007, p 175) observed that "identifying such a mixture of deleterious and compensatory changes as positive selection contributes little to our understanding of the molecular basis of evolutionary novelties." An important question is: how many signatures of positive selection detected in human genes are actually signatures of compensatory evolution? A challenge for the future will be to tease apart the fractions of substitutions that are slightly deleterious, compensatory, and those that actually underlie new evolutionary phenotypes.

Slightly deleterious substitutions in regulatory regions: Less-selective constraint in regulatory regions compared with protein-coding regions. King and Wilson (1975) noted that the relatively small number of differences in the proteins between chimpanzees and humans seemed to be too small to account for the large phenotypic differences between the species. Therefore, they proposed that the phenotypic differences between the species are caused by evolutionary changes within regulatory regions of genes. Recently, increased attention has been given to evolution within regulatory regions for generating new phenotypes (Carroll, 2003; Prud'homme et al., 2007; Wray, 2007).

As described earlier, evidence suggests that selective constraint in both regulatory regions and protein-coding regions have been relaxed in hominids (humans and chimpanzees) compared with levels of constraint in murids. However, regulatory regions may have experienced considerably greater reduction in selective constraint compared with protein-coding regions (Bush and Lahn, 2005; Keightley et al., 2005b; Kryukov et al., 2005; Keightley et al., 2006; Eyre-Walker and Keightley, 2007). In a detailed comparison of selective constraints in protein-coding regions versus conserved non-coding (CNC) regions that are potentially regulatory in two species-pairs (chimpanzee versus humans and rat versus mouse), Kryukov et al. (2005) found that although protein regions showed 12% less constraint in the hominids compared with the murids, CNC regions in hominids showed around 50% less-selective constraint compared with murids. Why is there considerably less-selective constraint in CNC regions (presumably regulatory) compared with protein-coding regions in hominids? One explanation is that mutations in CNC regions have smaller negative effects on fitness compared with mutations in protein-coding regions (see Eyre-Walker and Keightley, 2007; Kryukov et al., 2005).

The role of compensatory mutations in regulatory regions

Ohta (2002) discussed the impact of nearly neutral substitutions (slightly deleterious substitutions) in gene regulatory regions and how these might affect morphological evolution. She considered the possibility that regulatory regions could accumulate slightly deleterious mutations that could then be compensated for by nearby regulatory mutations that were able to stabilize the gene's expression. Experiments by Ludwig et al. (2000) show this to be possible. Ludwig et al. (2000) found that although phenotypic expression of the gene *even-skipped stripe 2* (a gene contributing to proper segmentation during embryogenesis) is conserved across different *Drosophila* species, *cis* regulatory elements involved in regulating the gene's expression are not conserved across species but have accumulated numerous substitutions. This raises the question of whether *cis* regulatory substitutions are inconsequential. To examine this question, Ludwig et al. (2000) constructed a chimera between the *Stripe 2* regulatory elements from two different *Drosophila* species. They then compared the expression pattern of the chimeric element to the expression pattern of each of the two undisturbed or "native" elements. The two native elements gave normal expression patterns, whereas the chimeric element gave a defective expression. As an explanation, Ludwig et al. (2000) proposed that stabilizing selection has acted to maintain phenotypic conservation despite allowing mutations to accumulate at functionally important sites. Nucleotide substitutions between species in regulatory regions are thought to accumulate by random genetic drift because of their extremely weak selective effects. To preserve gene expression across species, Ludwig et al. (2000) hypothesized that different sets of compensatory mutations have evolved within the regulatory regions of each species, a phenomenon they hypothesize as being common in the evolution of regulatory regions.

One implication of the Ludwig et al. (2000) study is that there seems to be a considerable amount of cryptic variation that accumulates within regulatory regions because individual point mutations may have very small fitness effects and because compensatory mutations may hide any possible phenotypic effects. Of greater significance for evolution is that such cryptic variation may evolve by nonadaptive processes and may subsequently become advantageous in a changed environmental context and/or within a changed population genetic context (e.g., increased population size) when positive selection becomes more effective. On this point, Ludwig et al.

(2000) note that variation in regulatory regions may subsequently come under selection to produce important phenotypic differences between species. As examples, they note that for two traits that show phenotypic differences between *Drosophila* species, abdominal bristle-number and wing morphology, that quantitative genetics research has determined strong associations with nucleotide differences in noncoding presumably regulatory regions. Thus, the cumulative effect of slightly deleterious and compensatory changes in regulatory regions may be substantial and some phenotypes may have evolved because of the fixation of a large number of simultaneously acting noncoding variants (see Kryukov et al., 2005).

Effective population size and differences in adaptive evolution

Compared with other animal species, primates seem to have experienced less overall positive selection in driving amino acid change within genes. The difference is especially remarkable when comparisons are made with distant species that have much large effective population sizes. In bacteria, *Drosophila*, and mice, estimates of the fraction of amino acids driven to fixation by positive selection are 40%, or considerably greater (Table 3), compared with estimates of adaptive evolution in the divergence of humans and chimpanzees (between ~0.0% and 20%; Table 2). Smith and Eyre-Walker (2002) estimated that approximately 45% of the total number of amino acid substitutions between *Drosophila simulans* and *D. yakuba* were adaptive. Thus, they estimated that over the 6 Myr divergence of the two species, about 270,000 amino acids were driven to fixation by positively selection.

To compare, we can make a rough estimate of the overall number of amino acids that have been positively selected in the divergence of humans and chimpanzees. Between chimpanzee and humans, Mikkelsen et al. (2005) found 38,773 amino acid differences in 13,454 genes, indicating that there is an average of 2.88 amino acid substitutions per gene. Dividing 2.88 by 443 (the average number of codons per gene; see Lander, 2001), there are 0.0064 substitutions per codon. The total number of amino acids in the human genome is obtained by multiplying 443 × 25,000, the estimated total number of genes in the genome (see International Human Genome Consortium, 2004), to give 11,075,000. If we then divide it by 0.0064, we get the total number of amino acid substitutions between humans and chimpanzees (70,880). For an estimate of the percentage of amino acids positively selected between these two species, we can take an average of the estimated fraction of positively selected genes in Table 2 (6.0%). Six percent of 70,880 yields 4,252 amino acid substitutions, and this is the number of amino acids estimated to be adaptive in the divergence of humans and chimpanzees. Thus, the estimated number of adaptive substitutions between humans and chimpanzees is merely 2% of the number thought to be adaptive in the divergence of the two *Drosophila* species.

This large difference is likely to be attributable to their vast differences in population size. The N_e for *Drosophila* is near 1 million, and, for the hominids (including both humans and chimpanzees), it is roughly 100,000. In large populations, the efficacy of positive selection is augmented, therefore, allowing variants with a wider range of fitness effects to be positively selected. Also, as mentioned previously, in large populations the probability of advantageous variants arising via mutation is augmented because there are more targets on which variants can arise. On the other hand, the difference in estimates of N_e between chimpanzees (~21K) and humans (~10,000) is much smaller than between hominids and *Drosophila*. However, even for smaller differences in N_e, similar population genetics effects may be apparent. As described earlier, Bakewell et al. (2007) estimated that chimpanzees have 50% more PSGs compared with humans. If this is true, then it indicates how natural selection can be sensitive to even relatively small changes in population size.

What do these differences tell us about evolution by adaptation in populations of different sizes? The answers are unclear at present, but Eyre-Walker (2006a) suggests two possibilities. It might indicate that small populations are less able to adapt to their environments than are large populations. On the other hand, it could be that most of the important adaptations are underlain by mutations that have strong selective effects and these are positively selected across a very large range of population sizes. Thus, the increased efficacy of positive selection in large populations may be merely "fine tuning" in which variants that have extremely small benefits are fixed by positive selection.

What is the biological significance of the considerably larger fraction of adaptive amino acids in *Drosophila* compared with hominids? If adaptive amino acid differences give rise to adaptive differences at the morphological level, then there seems to be a disconnect because the magnitude of phenotypic differences between humans and chimpanzees on the face of it seems much greater than between *Drosophila* species (although morphological differences do exist between *Drosophila* species). Eyre-Walker (2006a) has suggested that the differences in *Drosophila* may relate more to the physiological and ecological behavior of these species than to morphological aspects. For example, a lot of adaptive divergence may be related to the physiological aspects of host–plant interactions, reproductive interactions, and other phenomena for which we know little about (and flies "know" a lot about!).

The fact that chimpanzees are hypothesized to have more PSGs than humans is also counterintuitive because in terms of morphology humans seemed to have diverged to a larger extent than have chimpanzees (e.g., in brain size and structure, in traits related to bipedalism, and in hand anatomy). However, it may be that these morphological traits derive less from substitutions coded within proteins and more from subtle changes within regulatory regions. The larger numbers of PSGs in chimpanzees compared with the number in humans may point to the existence of a relatively large set of physiological traits in which chimpanzees differ from us. In support of this idea, genome-wide scans of selection (Mikkelsen et al., 2005; Nielsen et al., 2005; Arbiza et al., 2006; Bakewell et al., 2007; Kosiol et al., 2008) have usually found that the biological categories most enriched for PSGs are related to physiological processes such as reproduction, immune defense, and chemosensory functions. Furthermore, Bakewell et al. (2007) found that humans and chimpanzees differ significantly in the exact sets of genes distributed to these biological categories, indicating substantial species-specific differences in adaptation.

Possible consequences of gene surfing in the human evolution

The spatial dynamics of populations and species can have consequences for evolutionary processes (Burton

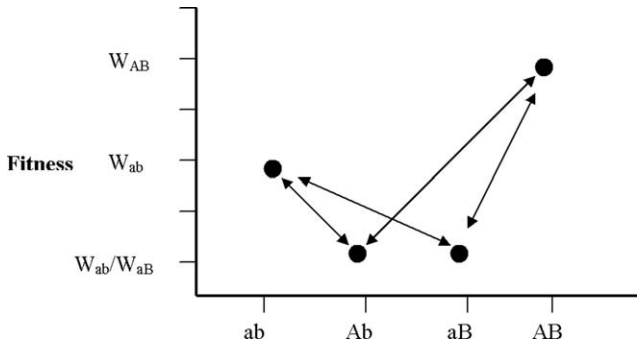

Fig. 5. Graph of the relative fitnesses for the four different genotypes for two biallelic loci (*A* and *B*). The *A* locus has alleles *a* and *A* and the *B* locus has alleles *b* and *B*. The *ab* genotype has intermediate fitness and *AB* has the highest fitness. However, genotypes *Ab* and *aB* have the lowest fitness and represent fitness valleys (Redrawn from Burton and Travis, 2008).

and Travis, 2008; Excoffier et al., 2009). With respect to human evolution, Eswaran and coworkers (Eswaran, 2002; Eswaran et al., 2005) showed that a coadapted combination of genes carrying a complex advantage could spread out of Africa on a wave front of diffusion. Recently, Burton and Travis (2008), using simulations, found that populations undergoing range expansions have an increased likelihood (compared with stationary populations) for fitness peak shifts, moving from one peak to a higher fitness peak by passing through a low-fitness valley. The model incorporates the idea of sign epistasis in which the fitness effects of a mutation at one locus is dependent on the genetic background provided by another locus. This is similar to the epistasis modeled by Kimura (1991), described above, but with the added dimension of how population expansion affects the shift between fitness peaks across a valley. Figure 5 shows a two-locus haploid system consisting of locus *A* (with alleles *a* and *A*) and locus *B* (with alleles *b* and *B*). Genotype *ab* has an intermediate fitness and *AB* has the highest fitness; however, the intermediate genotypes, *aB* and *Ab*, have the lowest fitnesses. Results of forward simulations of population growth, dispersal, and mutation showed an increase in the probability of fitness peak shifts during periods of range expansion (Burton and Travis, 2008). They showed that the lower fitness genotypes (*aB* and *Ab*) can accumulate on a wave front via mutation surfing, which leads to increased probability that a new mutation leads to a shift to a higher peak through epistasis. In the evolution of modern humans, there is genetic evidence of successive founder effects as populations spread out of Africa (Prugnolle et al., 2005; Ramachandran et al., 2005; Liu et al., 2006; Handley et al., 2007; Li et al., 2008; DeGiorgio et al., 2009; Deshpande et al., 2009; Hunley et al., 2009). The research by Burton and Travis (2008) suggests that frequent founder effects of this type could have facilitated the evolution and geographic spread of advantageous genotypes in humans.

Such dynamic demographic history including successive founder effects and spatial expansions are examples of nonadaptive processes that, in theory, are capable of impacting considerably the evolution of species and may have had a special impact on the evolution of modern humans. Such processes may have also acted at earlier times in human evolution. Recent studies of African lakes has indicated that climate in East Africa during the Plio-Pleistocene has been punctuated by periods of extreme climate variability particularly between 2.7 and 2.5 Ma, 1.9 and 1.7 Ma, and 1.1 and 0.9 Ma (Trauth et al., 2005, 2007, 2009; Maslin and Christensen, 2007). During these periods, stratigraphic transitions in East African paleo lakes suggest that they underwent rapid oscillations—appearing rapidly, persisting for thousands of years, and then rapidly disappearing (Maslin and Christiansen, 2007). Extreme climatic variability might have led to the expansion and contraction of hominin populations in East Africa. These demographic dynamics could have facilitated peak shifts in fitness by mutation surfing that enabled the evolution of novel hominin phenotypes.

Preservation of gene duplicates and potential functional effects

As described above, relaxed purifying selection in primates and humans concomitant with reductions in population size may have been responsible for the passive accumulation of various features of genomic architecture. However, it must be reiterated that although relaxed purifying selection because of reductions in N_e may have been an important contributing factor, it was probably not the only factor in allowing the accumulation of genomic features. Below, I consider the possible evolutionary and functional consequences that the accumulation of genomic features potentially had on primate and human evolution.

Population genetics theory predicts that a functionally redundant duplicate gene copy will not persist long in a population because deleterious mutations will accumulate and cause the gene to become nonfunctional (He and Zhang, 2005). Yet, gene duplication is believed to be the principal font of new genes (Ohno, 1970; see Taylor and Raes, 2004 for a historical treatment). Essentially two major hypotheses have been put forward to explain how duplicates become fixed within populations. The neofunctionalization (NF) hypothesis describes that, after duplication, one of the duplicates retains the ancestral function, and the other copy gains a new and different function and is subsequently fixed by positive natural selection (Ohno, 1970). An alternative hypothesis, known as SF (see Hughes, 1994; Force et al., 1999) proposes that duplicates are fixed in populations via the reciprocal loss of different subfunctions between the original gene and its descendent copy making each copy biologically requisite. Lynch and Force (2000) showed that the probability of fixation of duplicate genes by SF is thwarted by strong purifying selection in larger populations but enhanced in small populations with reduced levels of purifying selection (Lynch and Force, 2000). He and Zhang (2005) tested both models (i.e., NF versus SF models) using experimentally determined genome-wide patterns of protein interactions in yeast and human gene expression and found that a joint model is more likely whereby duplicates are initially fixed and stably preserved in a population by SF after which NF has ample time to occur.

More recently, it has been proposed that positive selection might act immediately after duplication to drive duplicates to fixation because of the general benefit offered by the dosage increase in gene product (see Kondrashov and Koonin, 2004; Kondrashov and Kondrashov, 2006). Qian and Zhang (2008) tested this idea by comparing sets of known haploinsufficient genes in the genome (i.e., genes for which the gene product is halved

when one copy of a homologous pair is defective) and found that the former set of genes do not duplicate more frequently than latter genes, as might be expected if the dosage-effect hypothesis were true. On this basis, they argued that the gene dosage hypothesis is unlikely to be an explanation for gene duplicate preservation. Furthermore, they point out that the initial assumption, that an increase in gene product is generally advantageous, is likely to be false for two reasons. First, increased gene product may generally be deleterious and second, increased production of gene product may have an energy cost associated with it that lowers fitness (Qian and Zhang, 2008).

The initial nonadapative preservation and fixation of duplicates does not exclude the possibility that positive selection acts subsequently on duplicates to promote new and divergent functions. The duplication of the *RNASE1* gene to produce the daughter copy *RNASE1B* in at least two Colobine primates (*Colobus guereza* and *Pygathrix nemaeus*) seems to be an example whereby SF was followed (probably relatively rapidly) by NF (see Zhang et al., 2002; Zhang, 2006). *RNASE1* is expressed in the pancreas and also in other widespread tissues of the body. It is known to have at least two different functions, the digestion of single-stranded RNA in the small intestine and a nondigestive function (of which little is known) in the degradation of double-stranded RNA, presumably of viral RNA (see Sorrentino and Libonati, 1997). After duplication, the two functions seem to have been partitioned between the duplicates whereby *RNASE1* retained the nondigestive function of degrading dsRNA and *RNASE1B* lost the ability to degrade dsRNA and became restricted to degrading ssRNA. Subsequently, because of the accumulation of multiple amino acid substitutions in *RNASE1B*, this gene gained a new function, the capacity to work in low pH conditions as exist in the small intestine of foregut-fermenters like the two Colobine species. In this example, the gain of a new function in *RNASE1B* would have been impossible without relaxation of selective constraints on the gene (Zhang et al., 2002), although it may have been due to functional redundancy rather than a decrease in population size. Otherwise, the amino acid alterations (that inhibit dsRNA degradation) would not have been accepted. Duplication of *RNASE1*, therefore, seems to be explained by an initial nonadaptive phase followed by an adaptive phase.

Are the initial phases of gene duplication generally governed by relaxed selective constraints? At present, we do not have the answer to this question. Too few duplicate genes have been studied at the level of detail necessary to make such inferences. Some studies have tested specifically for positive selection and hypothesize that a considerable fraction of duplicate genes in primates (10%) have experienced positive selection (Han et al., 2009). Others have presented suggestive evidence of adaptive duplicates in humans and other primates (e.g., Dumas et al., 2007). Yet, from the data, it is impossible to tell whether positive selection was indeed the evolutionary force responsible for the initial maintenance and/or fixation of the duplicated copy(ies). Another problem is that precise nucleotide sequence is unavailable at present for many individual duplicate copies of genes, making it impossible to tell whether some gene copies have been degraded by stop codons, frameshifts, or splice-site mutations. Even in the case of duplicated copies of the salivary-amylase gene, which seems to be a case in which increased dosage was positively selected in human populations with high starch diets (Perry et al., 2007), it is impossible to tell whether copies were initially maintained and fixed by nonadaptive processes. For example, it is known that a wide range of copies of the salivary amylase gene (from two to 13 copies) are segregating within populations with low-starch diets, making it possible that the gene's copy numbers fluctuate by random genetic drift (see Perry et al., 2007). In fact, Nei and colleagues have hypothesized that a large fraction of the copy number variation observed within populations (Redon et al., 2006; Nguyen et al., 2008; Perry et al., 2008) and between species (especially notable in the expansive differences within chemosensory gene families) is governed by a neutral process they call *genomic drift*, the random fluctuation in numbers of gene copies between different individuals (Nei and Rooney, 2005; Nozawa et al., 2007; Nozawa and Nei, 2008; Nei et al., 2008; also see Zhang, 2007). In the future, only with more comprehensive nucleotide sequence from each individual gene copy, as well as more detailed study of their molecular function and evolution, will the issue of the timing and significance of nonadaptive versus adaptive evolution in gene duplication be better understood.

Nonadaptive accumulation of TEs and evolutionary implication

About half of the human genome is comprised of mobile genetic elements (Lynch, 2007). As reviewed previously, evidence suggests that reduced selective constraints as a result of smaller population sizes may have been a contributing factor in permitting the accumulation of mobile genetic elements in primates and humans. What are the evolutionary and functional consequences of these TEs?

Some TEs are known to be associated with known genetic diseases such as autoimmune lymphoproliferative syndrome, glycerol kinase deficiency, Apert syndrome, neurofibromatosis Type 1, and hemophilia A (see Lev-Maor et al., 2008). Despite this, there is evidence that some mobile genetic elements can have important functional and even adaptive importance (Brosius, 1991; Brosius and Gould, 1992; Makalowski, 2003). Functional effects can result through several mechanisms: through TEs contributing regulatory sites to genes situated near their insertion points (Thornburg et al., 2006), TEs contributing to the alternative splicing of proteins (Lev-Maor et al., 2003), and TEs causing chromosomal inversions that become potentially important for speciation (Lee et al., 2008). In brief, mobile genetic elements that accumulated by nonadaptive processes can provide raw material for future functions that may subsequently be promoted by positive selection because they become beneficial.

With respect to a regulatory role, Thornburg et al. (2006) showed that TEs often carry functional regulatory motifs such as polyadenylation sites, promoters, enhancers, and silencers that, when inserted near genes, have the potential of giving rise to new species-specific expression patterns. They found that SINEs (e.g., *Alu*s) were found most commonly within regulatory regions of genes. Hamdi et al. (2000) gave examples of four genes (*PTH*, nicotinic acetylcholine receptor alpha 3 gene, *CD8* alpha gene, and hematopoietic cell-specific receptor gene) that were differentially regulated in higher primates because of the presence or absence of *Alu* elements

in their upstream regulatory regions. Clarimon et al. (2004) found that the *APP* gene, which is up-regulated in Alzheimer's disease, seems to be differentially affected in higher primates because of the presence/absence of regulatory sites from nearby *Alu* elements.

Lev-Maor et al. (2003) described one mechanism by which TEs (*Alu*s) inserted into genes can lead to the formation of new exons (a general process called exonization) that can lead to alternatively-spliced variants of proteins. They showed that more than 5% of human alternatively spliced proteins are due to *Alu*s and that most exons derived from *Alu*s are alternatively spliced. Alternative splicing can lead to expression complexity whereby different isoforms of the protein can be expressed in different tissues. Krull et al. (2005) reconstructed the evolutionary steps presumed to lead to the exonization or nonexonization of *Alu* elements in four primate genes. They showed that exonization is not instant after *Alu* insertion into a gene but proceeds by steps over evolutionary time whereby *Alu* elements acquire internal mutations (e.g., alternative splice sites and additional mutations) that eventually allow them to become functional.

Two genes that seem to have become exonized via TE incorporation during primate evolution and that seem to have become differently expressed in humans and chimpanzees are *TTLL6* (tubulin tyrosine ligase-like family, member 6) and *SEPN1* (selenoprotein N) (Lin et al., 2008, 2009). *TTLL6* is an apoptosis-related gene with preferential expression in testis (Chen et al., 2006). It gained a splice site in the human lineage that led to a new isoform that shows considerable expression in the testis. Interestingly, Chen et al. (2006) found evidence of positive selection on *TTLL6* in humans and postulate that it plays an important function in human male reproduction. The second gene, *SEPN1*, plays an important protective function against oxidative damage in skeletal muscle (Lin et al., 2008). The isoform of *SEPN1* that contains the *Alu*-derived exon shows considerable expression in skeletal muscle in humans and, therefore, would seem to contribute to species physiological differences. This difference may relate to Bramble and Lieberman's (2004) hypothesis that endurance running was an important adaptation in the genus *Homo* about 2 million years ago, although it is unclear whether *SEPN1* has a signature of positive selection. Evidence indicates that intense and prolonged exercise can produce oxidative damage to biological molecules and can negatively affect cell signaling pathways and the expression of genes (Powers and Jackson, 2008). Therefore, the expression of *SEPN1* in human skeletal muscle during human evolution may have been important for reducing oxidative damage resulting from long distance running.

Modrek and Lee (2003) proposed that relaxed purifying selection plays an important role in the evolution of alternative splicing. They made the observation that most newly created exons (whether *Alu*-derived or not) are alternatively spliced and that most new exons are included only as minor-form transcripts. Because new isoforms are only expressed at low levels whereas the original gene product is still expressed at high levels, the new isoform experiences reduced purifying selection (i.e., it is less likely to be selected out of the population) and can potentially lead to new gene functions. Smaller effective population size would also contribute to reduced selective constraints and enable new exon formation as well as allow the accumulation of nearly neutral variants within these new exons.

The accumulation of mobile genetic elements (because of either relaxed selective constraints or other factors) might also have played a role in speciation in primates by causing chromosomal rearrangements. Lee et al. (2008) found that *Alu* and *L1* elements are responsible for 44% of the inversion differences between humans and chimpanzees. Inversions were found to be associated with reduced recombination rates in surrounding chromosomal regions, and many inversions were found within genes and, thus, could have a variety of possible consequences on genes by either disrupting their function, causing alternative splicing, or modifying regulatory control. Beyond potential roles in the speciation process, these TE-derived inversions could potentially produce species phenotypic differences.

Nonadaptive evolution and morphology

It seems reasonable to assume that an increase in the rate of fixation of slightly deleterious mutations due to relaxed purifying selection would have recognizable effects on morphological evolution. However, few studies have explicitly tested whether populations or species for which there is evidence of widespread relaxed constraints display higher rates of morphological change. Nevertheless, based on the neutral molecular theory (Kimura, 1968, 1969), several methods were developed for studying morphological evolution. A popular approach has been the so-called "rate test" that compares the observed rate of phenotypic change between populations with the rate expected under neutrality (Lande, 1976, 1979; Chakroborty and Nei, 1982; Lynch and Hill, 1986; Lynch, 1990). More explicitly, if populations have differentiated largely through random genetic drift, population genetic theory predicts that levels of between-species divergence should be proportional to levels of within species levels of variation.

Lynch (1990) applied the rate test to data sets of metric variation for craniofacial traits in 10 different mammalian groups using both fossil and extant morphological specimens. Similar to molecular evolution, where widespread selective constraints dominate, the observed rate of morphological evolution in the majority of mammalian groups was found to be markedly below the rate expected under neutral evolution. A recent study of numerous well-documented and diverse fossil lineages also found that stabilizing selection (selective constraint) plays a dominant role (Hunt, 2007). Although both studies found that selective constraint was true for lineages over long-term evolution, more rapid divergence was detected in initial phases (Lynch, 1990; Hunt, 2007). Whether this is related to smaller N_e or increased adaptive evolution in populations in the early phases of lineages is unknown. An exception to the generally constrained rates of change among mammals was found when morphological variation between human populations was analyzed. The observed rate of differences was similar to the neutral rate (Lynch, 1990). Additionally, although intergeneric comparisons among the great ape species generally showed rates well below neutrality, humans were exceptional and showed rates 10 times higher. Lynch (1990) suggested that the higher rates of morphological evolution (i.e., close to neutral expectations) were generally due to relaxed purifying selection, with the suggestion that this was due to increased cultural evolution. More recently, Lynch (2007) has related

the increased rates to reduced effective population size in humans.

Ackermann and Cheverud (2004) applied similar population genetic methods to metric data from hominin craniofacial fossils, finding evidence for considerable positive selection driving the divergence of Pliocene hominins (particularly Paranthropines). In contrast, genetic drift was found to dominate in the *Homo* lineage. Variation in cranial morphology among human geographic populations, and between humans and Neanderthal fossils, is also found to be most consistent with models of genetic drift (Relethford, 1994, 2002; Roseman, 2004; Roseman and Weaver, 2007; Weaver et al., 2007; Betti et al., 2009; Weaver, 2009). In contrast, when Marroig and Cheverud (2004) applied similar methods as used by Ackermann and Cheverud (2004) to Neotropical monkeys, they found that positive selection accounts for much of the divergence in cranial morphology (especially among genera and higher taxa). They did, however, find that genetic drift may be more important in divergence between species. It is possible that increased evidence of drift in hominins is related to their reduced population sizes compared with nonhuman primates, although this was not specifically tested. Extensive morphological evolutionary rate tests between species with different effective population sizes and different quantified levels of molecular constraint (e.g., using the dN/dS method) have not yet been carried out but will address the issue of the relative importance of drift versus selection in morphological evolution.

Island species and their mainland ancestors (or other known founder events) represent special cases where hypotheses about the effects of reduced N_e can be tested using both molecular and morphological data sets (see Bromham and Woolfit, 2004). Woolfit and Bromham (2005) analyzed molecular data for 70 different island–mainland (descendent-ancestor) pairs using diverse vertebrate species (including lemurid primates), invertebrates, and plants and found that the rates of nonsynonymous to synonymous substitutions were significantly greater in island species compared with their mainland relatives. They concluded that the effect stemmed from an accumulation of slightly deleterious (nearly neutral) substitutions in the island forms and that, because the effect was observed over such a diverse group of organisms, it was likely due to the founder event and smaller sizes of populations on islands.

If genetic drift becomes a relatively more predominant force in the molecular divergence of island forms, we might then expect to observe patterns of drift in their morphological divergence. Only a few studies have explicitly studied whether drift or selection predominates in morphological divergence of island forms, and several have indicated that adaptation may have a more important role than drift, though clearly more studies are needed. Clegg et al. (2002) studied the avian species-complex *Zosterops lateralis*, which represents one of the most successful island-colonizing bird groups repeatedly colonizing islands in the southwestern Pacific from a mainland origin in Australia. Three pieces of evidence were interpreted to favor positive selection over drift as the primary explanation for island diversification (a persistent shift to larger body size in island forms, little association between morphological and presumed neutral molecular divergence, and a rate of morphological change that seemed too high to be accounted for by drift alone). Although a repeated directional size shift is taken as evidence of general positive selection in divergence, Hunt (2007), in a broad analysis of fossil lineages, found that body size may be more labile to positive selection than shape changes.

Morphological divergence between a large set of island and mainland mammal populations based on both fossil and extant material (mostly rodents, but also sloths and the fossil primate *Cantius*) were analyzed by Millien (2006). She concluded that morphological evolution was accelerated in island forms particularly over short time periods but not longer ones (though see Pérez-Claros and Aledo, 2007 for an alternative view). She surmised that this was the result of rapid adaptation of populations to new island environments. The effects of reduced N_e and increased rates of drift as drivers of the accelerated short-term rates, though possible, were not discussed as explanations.

In sum, there is evidence that purifying selection (i.e., stabilizing selection) has operated over the long-term evolution of many animal lineages. Evidence does support episodes of accelerated evolution in the initial phase of lineages that could be associated with divergence. However, long adaptive trends in lineages are not supported (see Lynch, 1990; Hunt, 2007). In the human lineage, there seems to be an increased rate of morphological change that is consistent with a neutral rate. This shift toward neutrality is especially prominent in more recent hominins within the *Homo* lineage. The few studies that have explicitly tested hypotheses of selection versus genetic drift have focused largely on craniodental material, and there is need to broaden studies to include general features of the postcranium. Genome-level effects resulting from relaxed selective constraints because of reduced N_e should have widespread effects on the skeleton. Additionally, although there are some studies on nonhuman primates, including apes (Lynch, 1990) and New World primates (Marroig and Cheverud, 2004), there is need to widen studies to include more diverse nonhuman primates, as well as closely related mammals. Cases of the evolution of populations on islands, where N_e is generally and considerably reduced, represent opportunities to test the effects of relaxed selection on animal species. To date, however, studies of island versus mainland relatives, although perhaps pointing to an important role for adaptive evolution in island population divergence, are too few in number and have not explicitly tested hypotheses of genetic drift versus positive selection.

Genetic draft: the effect of selection on linked variants

Thus far, our discussion has focused on genetic drift as one of the most important stochastic forces in populations. Yet, recently, the evolutionary force "genetic draft" has been suggested as perhaps a more important stochastic force in natural populations than is genetic drift (Gillespie, 2000a,b, 2001). According to the genetic draft model, selected substitutions at one locus can induce dynamics that resemble genetic drift at linked neutral loci even when populations are large. Linked loci are two loci that experience little or no recombination between them. This leads to some unusual consequences that are the opposite of the expectations under neutral theory. For example, with increasing population size, genetic variation at linked neutral loci will be insensitive to population size. Also, with increasing population size, the rate of substitu-

tion of deleterious variants is expected to increase; yet, at the same time, the rate of substitutions of selected advantageous variants is expected to decrease.

The curious dynamics of genetic draft are due to the effects of selection on linked variants. To understand these effects it is useful to think of three types of linked selection (for review see Williford and Comeron, 2010). The first, termed "hitchhiking," was first described by Maynard Smith and Haigh (1974). It describes how neutral variation linked to a strongly advantageous variant is reduced as the advantageous variant spreads and fixes within the population. The magnitude and size of the genomic region affected by the hitchhiking event depends on the strength of selection, the frequency at which advantageous mutations arise, and the rate of recombination. The second, termed "background selection," was first described by Charlesworth and colleagues (Charlesworth et al., 1993; Charlesworth, 1994). Background selection can cause a reduction in neutral variation linked to a strongly deleterious variant. That is, when a neutral variant arises in a low recombination region of a chromosome containing a deleterious variant, it has an increased chance of being eliminated by purifying selection relative to a neutral mutation arising in a region of high recombination. Neutral variants arising on a chromosome not containing a deleterious variant have an increased chance of persisting in the population. The effects of hitchhiking and background selection in knocking down neutral variation linked to selected sites are very similar. However, it should be pointed out that hitchhiking is due to positive selection, whereas background selection is due to purifying (negative) selection.

Another type of selection of importance for linked variants may be called "interference selection." With interference selection, there is a reduction in the effectiveness of selection because of the linkage of two variants that are both under natural selection (Hill and Robertson, 1966; Birky and Walsh, 1988). Theoretical studies have shown that the fixation probability of a beneficial variant is reduced in the presence of a second beneficial variant at a linked locus. Consider a pair of linked beneficial variants. When these variants arise on different chromosomes, and provided they cannot recombine onto the same chromosome, they compete or "interfere" with each other so that only one of them can become fixed. However, if recombination was able to bring both variants onto the same background, then both could be brought to fixation in the population. Thus, recombination can decrease the interference between linked beneficial variants and lead to an overall increase in adaptation. Now, consider a slightly deleterious variant that arises near a beneficial variant within a genomic region of reduced or no recombination. In this case, the unit of selection is the entire genomic region of strong linkage. The overall positive selection coefficient of the region will increase the chance that both the beneficial variant and slightly deleterious variant will go to fixation (Birky and Walsh, 1988). Of course, if the slightly deleterious variant were able to recombine onto a chromosome without the beneficial variant, then there would be an increased chance it would be eliminated by purifying selection. Thus, when linked to a beneficial variant, a slightly deleterious variant has an increased probability of becoming fixed (Felsenstein, 1974; Birky and Walsh, 1988).

Interestingly, the ultimate effects of linked selection (encapsulated in Gillespie's draft model) are very similar to the effects of genetic drift; both will lead to the loss of variation in a population either by fixation or loss of variants. It is important to note that both forces are stochastic rather than deterministic forces and, therefore, are nonadaptive processes. One important difference is the influence of population size. The influence of genetic drift is directly dependent on population size. It is a powerful force in small populations and a weaker force in larger populations. As a result, under neutral theory, the amount of genetic variation within a species should be directly proportional to its effective population size. For example, the expected variation (or heterozygosity) in a population is given by

$$\hat{H} = 4N_e\mu/1 + 4N_e\mu$$

where N_e is the effective population size and μ is the mutation rate. Observe that $\hat{H}$ is extremely sensitive to population size. If $4N_e\mu$ is small, then one expects little heterozygosity, but if $4N_e\mu$ is large, then one expects very high heterozygosity. However, and unlike with drift, the effects of genetic draft are only weakly dependent on population size. In fact, with increasing population size, the effects of genetic draft become more influential, exactly the reverse of the case with drift. One way to explain this is that as population size increases, draft causes a decrease in the effective population size (N_e), especially for genomic regions that show stronger linkage (i.e., in regions with reduced recombination rates). In these regions, the effects of draft such as a decrease in neutral variation, reduction in fixation of advantageous variants, and an increased fixation of deleterious variants, are expected to become more pronounced.

Gillespie (2000a,b, 2001) has proposed that genetic draft can resolve a long standing paradox first pointed out by Lewontin (1974). This paradox lies in the observation that levels of variation within species are remarkably similar even when compared among species having vastly different population numbers (e.g., between bacteria and humans). The explanation is that draft becomes more effective at knocking down genetic variation as species' population sizes become larger (because of an increase in the effectiveness of selection in large populations), and this process helps to maintain homogeneity in variation even between species with very different population sizes. Recent studies of mitochondrial DNA variation across a large array of species showed a remarkable homogeneity in levels of genetic variation (Bazin et al., 2006). Thus, mtDNA variation in invertebrates (7.67%) was found to be closely similar to that in vertebrates (7.99%), even though they have vastly different population sizes. This result is expected under the draft model because the mitochondrial genome experiences essentially no recombination and is compact and gene-dense, all factors that contribute to linked selection. Interestingly, nuclear variation conforms closer to expectations based on neutral theory (i.e., increased diversity in species with presumed larger population sizes); however, the differences are surprisingly small given that population sizes are vastly different (Bazin et al., 2006). The lack of a strong correlation between nuclear diversity and population size could also be due to the effects of genetic hitchhiking and the draft model (see Eyre-Walker, 2006b).

Empirical evidence for draft in human evolution. Gillespie (2000a,b, 2001) believes that genetic draft is more important than drift, especially for genomes and regions of genomes that have reduced recombination

rates. What is the empirical evidence? There seems to be strong supporting evidence. There are a number of studies that have shown that levels of variation are significantly reduced in areas of low recombination in *Drosophila* (Aguade et al., 1989; Begun and Aquadro, 1992; Begun et al., 2007). The positive correlation between levels of variation and rates of recombination is now known to characterize a variety of organisms, including humans (Nachman et al., 1998; Przeworski et al., 2000; Cai et al., 2009; McVicker et al., 2009). Although these regions show reduced diversity, they show normal rates of interspecies divergence (Hellmann et al., 2008), which is not expected under the neutral theory. Neutral theory predicts a strong positive correlation between polymorphism and divergence. A strong possibility is that the pattern is a result of natural selection reducing levels of neutral variation linked to selected sites (Begun et al., 2007; Cai et al., 2009; McVicker et al., 2009). Further evidence that selection is the cause of the correlation is that variation within regions of lower recombination tends to be more reduced in functional regions as well as in regions where there is increased nonsynonymous divergence between species (Cai et al., 2009). Functional regions might be expected to have experienced greater amounts of selection and regions of increased nonsynonymous divergence may have experienced recurrent selective sweeps through evolutionary time (Cai et al., 2009). Indeed, Hellmann et al. (2008) found that regions of low diversity were strongly correlated with gene regions identified in previous studies as candidates for recent positive selection (e.g., Bustamante et al., 2005; Mikkelsen et al., 2005; Nielsen et al., 2005; Gibbs et al., 2007; Tang et al., 2007).

What type of selection is responsible for reducing levels of variation in regions of low recombination? Two major types of selection (described previously) may be involved: 1) hitchhiking because of recurrent positive selective sweeps or 2) background selection because of the repeated removal of deleterious variants. The major problem in discerning between these two explanations is that both leave a similar signature of variation in the genome: a reduction in variation and a skew toward rare alleles, although the skew may not be as marked for background selection as for hitchhiking (see Charlesworth, 1994; Fu, 1997; Nachman et al., 1998; Sella et al., 2009). Recent studies on genomic variation in humans have not been able to discern between the two forms of selection (Cai et al., 2009; McVicker et al., 2009). Nevertheless, it is possible that both types of selection operate simultaneously or operate to different extents in various regions of the genome. Indeed, a challenge for the future will be to discern the relative importance of adaptive or deleterious mutations in causing the correlation between reduction in variation and decreased recombination. In addition, the relative importance of background selection and hitchhiking will also likely vary between different animal lineages. For example, the much higher estimated fraction of adaptive substitutions in *Drosophila* compared with the estimate of adaptive substitutions in humans could suggest that background selection has been more prevalent than hitchhiking in human evolution. Indeed, some researchers (Hellmann et al., 2005; Reed et al., 2005) have found that models of background selection (negative selection) explain human genetic variation data slightly better than models of hitchhiking (positive selection), although a study by Hellmann et al. (2008) found the reverse to be the case. In *D. simulans*, Begun et al. (2007) found an excess of high-frequency variants, which could indicate the prevalence of hitchhiking due to positive selection in this species.

Lastly, in terms of my focus on the influence of nonadaptive forces in human evolution, it should be noted that studies of linkage and selection on variation demonstrate that much of human neutral variation across the genome is affected by stochastic forces induced by selection and linkage (i.e., Gillespie's genetic draft). Indeed, Cai et al. (2009) estimated that genome-wide levels of neutral variation have been reduced by 6% (and up to 11% in gene dense areas) because of the effects of linked selection, and estimates by McVicker et al. (2009) suggest that levels are even more reduced (>25%) by linked selection. However, the full details of draft's effects on human genomic variation are still largely unknown. Nevertheless, the assumption that neutral variation anywhere in the genome can be used to accurately infer human demographic history is becoming increasingly problematic. For example, estimates of effective population size in species and even rates of divergence should be carried out in regions of the genome far away from selected sites and in regions where recombination rates are relatively high. It will also be interesting in the future to examine how and where interference selection has either limited adaptive substitution or increased the fixation of deleterious variants. Furthermore, many methods in population genetics based on neutral theory as well as estimates that result from the implementation of these methods (e.g., proportion of adaptive change, etc.) may need to be adjusted to take into consideration the effects of linked selection and draft (see Hahn, 2008; Williford and Comeron, 2010).

SUMMARY AND FUTURE PROSPECTS

In this article, I review comparative genomic evidence that indicates that there was an increase in nonadaptive evolutionary forces (specifically, genetic drift) in primate and human evolution. In particular, the primate lineage leading to humans shows significantly relaxed selective constraint (i.e., reduced purifying selection) and decreased adaptive evolution compared with organisms having much larger effective population sizes, like mice and *Drosophila*. Almost all estimates of effective population size over human evolution center about 10,000. This is a 10- to 100-fold smaller effective population size compared with the effective population sizes of murids or *Drosophila*. Thus, in a general comparison between mice and humans, humans seem to have experienced more than a 10% reduction in selective constraint (i.e., purifying selection) in protein-coding regions (Table 1; Mikkelsen et al., 2005; Bakewell et al., 2007; Kosiol et al., 2008). The relaxation in selective constraint in humans seems to have been even greater in regions of the noncoding genome that regulate genes (Keightley et al., 2005b; Kryukov et al., 2005). A recent finer-scaled analysis (Eőry et al., 2010) has confirmed that selective constraints are generally reduced in humans, especially in 5′ untranslated regions (presumably regulatory). However, the study also found some very surprising features namely that fourfold degenerate sites in humans (i.e., sites at which any nucleotide at that site codes for the same amino acid) experience two times greater constraint in humans compared with murids. This could suggest more complex protein-coding gene structure in humans compared with murids, and might also indicate that fourfold degenerate sites play a role in alternative splicing because alternatively spliced genes showed sig-

nificantly greater constraint at these sites than did single transcript genes.

At the same time that we see evidence of generally reduced purifying selection in humans, adaptive evolution in human evolution (measured in terms of proportion of amino acids fixed by positive selection) also seems to be reduced. Estimates of adaptive evolution in human proteins range from ~0% to 20%, whereas estimates for other organisms are quite considerably higher (e.g., mice ~57% Halligan et al., 2010; *Drosophila* >45%; see Table 3). There is also some evidence suggesting relatively reduced adaptive evolution in humans compared with chimpanzees (Bakewell et al., 2007), perhaps because of reduced population size in humans, (although see Mallick et al. (2009) for further discussion). The differences in selective constraint and in adaptive evolution between humans compared with *Drosophila* and mice are commonly interpreted as consistent with neutral and nearly neutral predictions of increased genetic drift and the simultaneous decrease in efficacy of positive selection in smaller populations compared with larger populations. On the other hand, it is unclear whether the much smaller differences in population size between humans (~10,000) and chimpanzees (~20,000, estimated for the chimpanzee-bonobo ancestor; Caswell et al., 2008) would have had a significant impact on selection.

Relaxed selective constraint in human evolution would have had the proximate effect of increasing the fixation of effectively neutral substitutions (either slightly deleterious or slightly advantageous) throughout the genome. To compensate for slightly deleterious substitutions, I have suggested the possibility that humans have experienced an increase in nearby substitutions that partially compensate for (or suppress) deleterious substitutions via epistatic interactions between the substitutions. Although there are experimental studies of bacteria and viruses, as well as comparative gene studies between mammal species, that suggest that compensatory substitutions occur, the possible role of compensatory substitution in human evolution remains to be investigated.

Recent population growth in humans is expected to have led to concomitant increases in the effectiveness of selection (see Hawks et al., 2007). Substitutions that were effectively neutral when human populations were relatively small would have come under the increasing scrutiny of natural selection (both purifying and positive selection). Thus, slightly deleterious variants would have an increased chance of being purged by purifying selection, increasing the overall fitness in the population. On the other hand, it is possible that some slightly deleterious substitutions, either through epistatic interaction with nearby sites or through changed environments, have become adaptive. In this view, the nonadaptive accumulation of effectively neutral substitutions in past human evolution may have provided the raw material for recent human adaptations.

It is also expected that positive selection has been more effective due to recent growth in human populations. Thus, slightly advantageous variants that were effectively neutral in the past may have increased in frequency or have been fixed by positive selection in recent human evolution. Along with the expectation of a greater efficacy of positive selection in larger human populations is the expectation for an increase in the overall number of adaptive variants in the population. This is because, in larger populations, there is an increased number of targets on which beneficial variants can arise through the mutational process. Therefore, there is the theoretical expectation that as population size has increased, the rate of adaptive evolution has also increased. Indeed, genome scans have found considerable evidence of partial positive selective sweeps that are presumably ongoing in human populations (Tang et al., 2007), and there exists evidence for an acceleration of adaptive change that is correlated with increasing population size over the past 50,000 years (Hawks et al., 2007).

One important link that has not been adequately explored is that between evolution at the molecular level and evolution at the morphological level with respect to changes in population size. Thus, the effects of reduced selective constraint (with concomitant accumulation of slightly deleterious substitutions) and reduced adaptive evolution during past human evolution should have had measurable consequences on our morphological evolution. One way to possibly gain insight into this phenomenon is to study morphological variation in island species (or populations of small population size) and compare their levels of morphological constraint with that in their mainland ancestors (presumably of larger population size) (see Clegg et al., 2002; Bromham and Woolfit, 2004, for examples). Within primates, many island populations exist for which studies could be conducted. Also, explicit tests of selective constraints (e.g., Lande, 1976; Lynch, 1990) can be made on morphological features in species with lineages having different effective population sizes. It is possible that species of smaller effective population sizes show reduced selective constraints and less adaptive change compared with species with larger effective population sizes. It will be interesting to see in the future to what extent human morphological features (e.g., cranial shape, postcranial dimensions) have been influenced by nonadaptive processes. Furthermore, we might ask: to what extent have the unique features of human evolution (i.e., in brain size and organization, bipedal features, hand morphology, etc.) as well as differences in morphology among human population been influenced by nonadaptive processes?

In the 20th Century, population genetics theory outpaced the actual data that could affirm or refute it. In the beginning of the 21st Century, the tables have turned; data collection now outpaces theory. Population genomic data will soon be providing us with detailed information concerning variation within populations (both in humans and other species) and divergence between species. These data will enable researchers to estimate effective population size in a variety of nonhuman primate lineages and will permit the finer investigation of the nature of evolutionary forces acting across the genome and in different lineages. It will also allow us to examine genomic architectural differences between individuals and between species. These kinds of data will drive the development of new theory (for informative discussions see Hahn, 2009; Siepel, 2009; Pool et al., 2010).

This review has largely focused on neutral and nearly neutral theory. In part, recent data are consistent with neutral theory in showing relaxed constraint and relatively diminished adaptive evolution in species like humans with evolutionarily small population sizes. However, some predictions based on neutral theory seem to be inconsistent with new data (see Hahn, 2009; Sella et al., 2009). For example, it is becoming clear that much of human genetic variation, though nonfunctional in itself and, therefore, not directly under selection, has been influenced by selection acting on nearby linked variants

(see preceding section). The theory of genetic draft (Gillespie, 2000a,b, 2001) is one explanation that can be applied to interpret such variation (see Cai et al., 2009; McVicker et al., 2009). Although draft is dependent on selection, its consequences are largely stochastic, and, therefore, we may conclude that a considerable portion of human genetic variation is nonadaptive. Once more, neutral genetic variation effected by draft may prove to be of limited value for demographic inference, and previous estimates may need to be revised taking draft into account. At present, however, the relative roles of natural selection and neutral forces in shaping genetic diversity (i.e., the classic selectionist versus neutralist debate dating from the late 1960s) is still not resolved, though still of central importance (see Pool et al., 2010; Hahn, 2008).

In addition, new theory is being developed to try to disentangle the effects that demographic history can have on human genetic variation from those that selection can have. An exciting new area of research is on the phenomenon of gene surfing (discussed in relation to human colonization of Europe) whereby neutral or even slightly deleterious variants can sweep to fixation over large geographic areas simply because they were located on the leading edge of advance of a population (Edmonds et al., 2004; Klopfstein et al., 2006; Travis et al., 2007; Excoffier et al., 2009; François et al., 2010). Therefore, gene surfing is another nonadaptive evolutionary force that could have impacted considerably the genetic diversity of human populations, though at present we cannot tell to what degree.

In the future, deeper insights into human evolution will rest both on the collection of new genomic data and through the development of new theory permitting us to glean more accurate historical information. New data and theory will inevitably help us to discern the adaptive changes in human evolutionary history (for reviews see Harris and Meyer, 2006; Nielsen et al., 2007; Akey, 2009). However, our eagerness to identify adaptations should be tempered by the lessons of Gould and Lewontin (1979) and, more recently, Lynch (2007a), which encourage the prior rejection of nonadaptive explanations (however, for a contrary view see Hahn, 2009). Fortunately, new data and theory should enable us to more clearly appreciate the roles that nonadaptive processes (and adaptive processes) played in our evolution.

ACKNOWLEDGMENTS

The following colleagues graciously exchanged ideas with the author regarding their research: Mark Batzer, Jesse Bloom, Stephane Bossinot, Evan Eichler, Amanda Larracuente, Michael Lynch, Tomas Marques-Bonet, Art Poon, George Perry, Adam Siepel, James Sikela, Olin Silander, Arlin Stoltzfus, Martin Trauth, Andreas Wagner, Meg Woolfit, Soojin Yi, and Jianzhi Zhang. Any errors in interpreting their research findings or ideas are strictly mine. I also thank the following people: Haishen Yao, who helped with Figure 1; Diogo Meyer, who made valuable comments and suggestions on versions of this manuscript; Terry Harrison, who kindly arranged for access to research materials; and three anonymous reviewers for their thoughtful suggestions that improved the manuscript. I also thank Robert Sussman who guided this article through the review process to publication. A fellowship leave helped support this work.

LITERATURE CITED

Ackermann RR, Cheverud JM. 2004. Detecting genetic drift versus selection in human evolution. Proc Natl Acad Sci USA 101:17946–17951.

Aguade M, Miyashita N, Langley CH. 1989. Reduced variation in the Yellow-Achaete-Scute Region in natural populations of *Drosophila melanogaster*. Genetics 122:607–615.

Akey JM. 2009. Constructing genomic maps of positive selection in humans: where do we go from here? Genome Res 19:711–722.

Akey JM, Zhang G, Zhang K, Jin L, Shriver MD. 2002. Interrogating a high-density SNP map for signatures of natural selection. Genome Res 12:1805–1814.

Arbiza L, Dopazo J, Dopazo H. 2006. Positive selection, relaxation, and acceleration in the evolution of the human and chimp genome. PLoS Comput Biol 2:e38.

Arcos-Burgos M, Muenke M. 2002. Genetics of population isolates. Clin Genet 61:233–247.

Bailey JA, Liu G, Eichler EE. 2003. An *Alu* transposition model for the origin and expansion of human segmental duplications. Am J Hum Genet 73:823–834.

Bakewell MA, Shi P, Zhang J. 2007. More genes underwent positive selection in chimpanzee evolution than in human evolution. Proc Natl Acad Sci USA 104:7489–7494.

Bamshad M, Wooding SP. 2003. Signatures of natural selection in the human genome. Nat Rev Genet 4:99–111.

Barreiro LB, Laval G, Quach H, Patin E, Quintana-Murci L. 2008. Natural selection has driven population differentiation in modern humans. Nat Genet 40:340–345.

Bazin E, Glémin S, Galtier N. 2006. Population size does not influence mitochondrial genetic diversity in animals. Science 312:570–572.

Beaumont MA, Balding DJ. 2004. Identifying adaptive genetic divergence among populations from genome scans. Mol Ecol 13:969–980.

Begun DJ, Aquadro CF. 1992. Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster*. Nature 356:519–520.

Begun DJ, Holloway AK, Stevens K, Hillier LW, Poh YP, Hahn MW, Nista PM, Jones CD, Kern AD, Dewey CN, Pachter L, Myers E, Langley CH. 2007. Population genomics: whole-genome analysis of polymorphism and divergence in *Drosophila simulans*. PLoS Biol 5:e310.

Bersaglieri T, Sabeti PC, Patterson N, Vanderploeg T, Schaffner SF, Drake JA, Rhodes M, Reich DE, Hirschhorn JN. 2004. Genetic signatures of strong recent positive selection at the lactase gene. Am J Hum Genet 74:1111–1120.

Betti L, Balloux F, Amos W, Hanihara T, Manica A. 2009. Distance from Africa, not climate, explains within-population phenotypic diversity in humans. Proc Biol Sci 276:809–814.

Bierne N, Eyre-Walker A. 2004. The genomic rate of adaptive amino acid substitution in *Drosophila*. Mol Biol Evol 21:1350–1360.

Birky CW Jr, Walsh JB. 1988. Effects of linkage on rates of molecular evolution. Proc Natl Acad Sci USA 85:6414–6418.

Boyko AR, Williamson SH, Indap AR, Degenhardt JD, Hernandez RD, Lohmueller KE, Adams MD, Schmidt S, Sninsky JJ, Sunyaev SR, White TJ, Nielsen R, Clark AG, Bustamante CD. 2008. Assessing the evolutionary impact of amino acid mutations in the human genome. PLoS Genet 4:e1000083.

Bramble DM, Lieberman DE. 2004. Endurance running and the evolution of *Homo*. Nature 432:345–352.

Bromham L, Woolfit M. 2004. Explosive radiations and the reliability of molecular clocks: island endemic radiations as a test case. Syst Biol 53:758–766.

Brosius J. 1991. Retroposons-seeds of evolution. Science 251:753.

Brosius J, Gould SJ. 1992. On "genomenclature": a comprehensive (and respectful) taxonomy for pseudogenes and other "junk DNA." Proc Natl Acad Sci USA 89:10706–10710.

Burch CL, Chao L. 1999. Evolution by small steps and rugged landscapes in the RNA virus Φ6. Genetics 151:921–927.

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

Burton OJ, Travis JM. 2008. The frequency of fitness peak shifts is increased at expanding range margins due to mutation surfing. Genetics 179:941–950.

Bush EC, Lahn BT. 2005. Selective constraint on noncoding regions of hominid genomes. PLoS Comput Biol 1:e73.

Bustamante CD, Fledel-Alon A, Williamson S, Nielsen R, Hubisz MT, Glanowski S, Tanenbaum DM, White TJ, Sninsky JJ, Hernandez RD, Civello D, Adams MD, Cargill M, Clark AG. 2005. Natural selection on protein-coding genes in the human genome. Nature 437:1153–1157.

Cai JJ, Macpherson JM, Sella G, Petrov DA. 2009. Pervasive hitchhiking at coding and regulatory sites in humans. PLoS Genet 5:e1000336.

Camps M, Herman A, Loh E, Loeb LA. 2007. Genetic constraints on protein evolution. Crit Rev Biochem Mol Biol 42:313–326.

Cann HM, de Toma C, Cazes L, Legrand MF, Morel V, Piouffre L, Bodmer J, Bodmer WF, Bonne-Tamir B, Cambon-Thomsen A, Chen Z, Chu J, Carcassi C, Contu L, Du R, Excoffier L, Ferrara GB, Friedlaender JS, Groot H, Gurwitz D, Jenkins T, Herrera RJ, Huang X, Kidd J, Kidd KK, Langaney A, Lin AA, Mehdi SQ, Parham P, Piazza A, Pistillo MP, Qian Y, Shu Q, Xu J, Zhu S, Weber JL, Greely HT, Feldman MW, Thomas G, Dausset J, Cavalli-Sforza LL. 2002. A human genome diversity cell line panel. Science 296:261–262.

Cargill M, Altshuler D, Ireland J, Sklar P, Ardlie K, Patil N, Shaw N, Lane CR, Lim EP, Kalyanaraman N, Nemesh J, Ziaugra L, Friedland L, Rolfe A, Warrington J, Lipshutz R, Daley GQ, Lander ES. 1999. Characterization of single-nucleotide polymorphisms in coding regions of human genes. Nat Genet 22:231–238.

Carroll SB. 2003. Genetics and the making of *Homo sapiens*. Nature 422:849–857.

Caswell JL, Mallick S, Richter DJ, Neubauer J, Schirmer C, Gnerre S, Reich D. 2008. Analysis of chimpanzee history based on genome sequence alignments. PLoS Genet 4:e1000057.

Cavalli-Sforza LL. 1966. Population structure and human evolution. Proc R Soc Lond B Biol Sci 164:362–379.

Cavalli-Sforza LL. 1994. The history and geography of human genes. Princeton, NJ: Princeton University Press.

Cavalli-Sforza LL, Menozzi P, Piazza A. 1993. Demic expansions and human evolution. Science 259:639–646.

Chakroborty R, Nei M. 1982. Genetic differentiation of quantitative characters between populations or species. I. Mutation and random genetic drift. Genet Res 39:303–214.

Charlesworth B. 1994. The effect of background selection against deleterious mutations on weakly selected, linked variants. Genet Res 63:213–227.

Charlesworth B, Morgan MT, Charlesworth D. 1993. The effect of deleterious mutations on neutral molecular variation. Genetics 134:1289–1303.

Charlesworth J, Eyre-Walker A. 2006. The rate of adaptive evolution in enteric bacteria. Mol Biol Evol 23:1348–1356.

Charlesworth J, Eyre-Walker A. 2007. The other side of the nearly neutral theory, evidence of slightly advantageous backmutations. Proc Natl Acad Sci USA 104:16992–16997.

Charlesworth J, Eyre-Walker A. 2008. The McDonald-Kreitman test and slightly deleterious mutations. Mol Biol Evol 25:1007–1015.

Chen FC, Li WH. 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.

Chen XH, Shi H, Liu XL, Su B. 2006. The testis-specific apoptosis related gene TTL.6 underwent adaptive evolution in the lineage leading to humans. Gene 370:58–63.

Cheng Z, Ventura M, She X, Khaitovich P, Graves T, Osoegawa K, Church D, DeJong P, Wilson RK, Paabo S, Rocchi M, Eichler EE. 2005. A genome-wide comparison of recent chimpanzee and human segmental duplications. Nature 437:88–93.

Chiaroni J, Underhill PA, Cavalli-Sforza LL. 2009. Y chromosome diversity, human expansion, drift, and cultural evolution. Proc Natl Acad Sci USA 106:20174–20179.

Clarimon J, Andres AM, Bertranpetit J, Comas D. 2004. Comparative analysis of *Alu* insertion sequences in the APP 5′ flanking region in humans and other primates. J Mol Evol 58:722–731.

Clark AG, Glanowski S, Nielsen R, Thomas PD, Kejariwal A, Todd MA, Tanenbaum DM, Civello D, Lu F, Murphy B, Ferriera S, Wang G, Zheng X, White TJ, Sninsky JJ, Adams MD, Cargill M. 2003. Inferring nonneutral evolution from human-chimp-mouse orthologous gene trios. Science 302:1960–1963.

Clegg SM, Degnan SM, Moritz C, Estoup A, Kikkawa J, Owens IP. 2002. Microevolution in island forms: the roles of drift and directional selection in morphological divergence of a passerine bird. Evolution 56:2090–2099.

Crow J. 2008. Mid-century controversies in population genetics. Ann Rev Genet 42:1–16.

Currat M, Excoffier L, Maddison W, Otto SP, Ray N, Whitlock MC, Yeaman S. 2006. Comment on "Ongoing adaptive evolution of ASPM, a brain size determinant in *Homo sapiens*" and "Microcephalin, a gene regulating brain size, continues to evolve adaptively in humans." Science 313:172; author reply 172.

Davis BH, Poon AF, Whitlock MC. 2009. Compensatory mutations are repeatable and clustered within proteins. Proc Biol Sci 276:1823–1827.

DeGiorgio M, Jakobsson M, Rosenberg NA. 2009. Out of Africa: modern human origins special feature: explaining worldwide patterns of human genetic variation using a coalescent-based serial founder model of migration outward from Africa. Proc Natl Acad Sci USA 106:16057–16062.

Deshpande O, Batzoglou S, Feldman MW, Cavalli-Sforza LL. 2009. A serial founder effect model for human settlement out of Africa. Proc Biol Sci 276:291–300.

Demuth JP, Hahn MW. 2009. The life and death of gene families. Bioessays 31:29–39.

DePristo MA, Weinreich DM, Hartl DL. 2005. Missense meanderings in sequence space: a biophysical view of protein evolution. Nat Rev Genet 6:678–687.

Dumas L, Kim YH, Karimpour-Fard A, Cox M, Hopkins J, Pollack JR, Sikela JM. 2007. Gene copy number variation spanning 60 million years of human and primate evolution. Genome Res 17:1266–1277.

Edmonds CA, Lillie AS, Cavalli-Sforza LL. 2004. Mutations arising in the wave front of an expanding population. Proc Natl Acad Sci USA 101:975–979.

Ellegren H. 2009. A selection model of molecular evolution incorporating the effective population size. Evolution 63:301–305.

Enard W, Khaitovich P, Klose J, Zollner S, Heissig F, Giavalisco P, Nieselt-Struwe K, Muchmore E, Varki A, Ravid R, Doxiadis GM, Bontrop RE, Paabo S. 2002. Intra- and interspecific variation in primate gene expression patterns. Science 296:340–343.

Endler JA. 1973. Gene flow and population differentiation. Science 179:243–250.

Előry L, Halligan DL, Keightley PD. 2010. Distributions of selectively constrained sites and deleterious mutation rates in the hominid and murid genomes. Mol Biol Evol 1:177–192.

Estes S, Lynch M. 2003. Rapid fitness recovery in mutationally degraded lines of *Caenorhabditis elegans*. Evolution 57:1022–1030.

Eswaran V. 2002. A diffusion wave out of Africa: the mechanism of the modern human revolution? Curr Anthropol 43:749–774.

Eswaran V, Harpending H, Rogers AR. 2005. Genomics refutes an exclusively African origin of humans. J Hum Evol 49:1–18.

Evans PD, Gilbert SL, Mekel-Bobrov N, Vallender EJ, Anderson JR, Vaez-Azizi LM, Tishkoff SA, Hudson RR, Lahn BT. 2005. Microcephalin, a gene regulating brain size, continues to evolve adaptively in humans. Science 309:1717–1720.

Excoffier L, Hofer T, Foll M. 2009. Detecting loci under selection in a hierarchically structured population. Heredity 103:285–298.

Excoffier L, Ray N. 2008. Surfing during population expansions promotes genetic revolutions and structuration. Trends Ecol Evol 23:347–351.

Eyre-Walker A. 2002. Changing effective population size and the McDonald-Kreitman test. Genetics 162:2017–2024.

Eyre-Walker A. 2006a. The genomic rate of adaptive evolution. Trends Ecol Evol 21:569–575.

Eyre-Walker A. 2006b. Size does not matter for mitochondrial DNA. Science 312:537–538.
Eyre-Walker A, Keightley PD. 1999. High genomic deleterious mutation rates in hominids. Nature 397:344–347.
Eyre-Walker A, Keightley PD. 2007. The distribution of fitness effects of new mutations. Nat Rev Genet 8:610–618.
Eyre-Walker A, Keightley PD. 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, Keightley PD, Smith NG, Gaffney D. 2002. Quantifying the slightly deleterious mutation model of molecular evolution. Mol Biol Evol 19:2142–2149.
Eyre-Walker A, Woolfit M, Phelps T. 2006. The distribution of fitness effects of new deleterious amino acid mutations in humans. Genetics 173:891–900.
Fay JC, Wyckoff GJ, Wu CI. 2001. Positive and negative selection on the human genome. Genetics 158:1227–1234.
Felsenstein J. 1974. The evolutionary advantage of recombination. Genetics 78:737–756.
Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J. 1999. Preservation of duplicate genes by complementary, degenerative mutations. Genetics 151:1531–1545.
Ford EB. 1964. Ecological genetics. New York: John Wiley.
François O, Currat M, Ray N, Han E, Excoffier L, Novembre J. 2010. Principal component analysis under population genetic models of range expansion and admixture. Mol Biol Evol 27:1257–1268.
Fu YX. 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genetics 147:915–925.
Gaffney DJ, Blekhman R, Majewski J. 2008. Selective constraints in experimentally defined primate regulatory regions. PLoS Genet 4:e1000157.
Gherman A, Chen PE, Teslovich TM, Stankiewicz P, Withers M, Kashuk CS, Chakravarti A, Lupski JR, Cutler DJ, Katsanis N. 2007. Population bottlenecks as a potential major shaping force of human genome architecture. PLoS Genet 3:e119.
Gibbs RA, Rogers J, Katze MG, Bumgarner R, Weinstock GM, Mardis ER, Remington KA, Strausberg RL, Venter JC, Wilson RK, Batzer MA, Bustamante CD, Eichler EE, Hahn MW, Hardison RC, Makova KD, Miller W, Milosavljevic A, Palermo RE, Siepel A, Sikela JM, Attaway T, Bell S, Bernard KE, Buhay CJ, Chandrabose MN, Dao M, Davis C, Delehaunty KD, Ding Y, Dinh HH, Dugan-Rocha S, Fulton LA, Gabisi RA, Garner TT, Godfrey J, Hawes AC, Hernandez J, Hines S, Holder M, Hume J, Jhangiani SN, Joshi V, Khan ZM, Kirkness EF, Cree A, Fowler RG, Lee S, Lewis LR, Li Z, Liu YS, Moore SM, Muzny D, Nazareth LV, Ngo DN, Okwuonu GO, Pai G, Parker D, Paul HA, Pfannkoch C, Pohl CS, Rogers YH, Ruiz SJ, Sabo A, Santibanez J, Schneider BW, Smith SM, Sodergren E, Svatek AF, Utterback TR, Vattathil S, Warren W, White CS, Chinwalla AT, Feng Y, Halpern AL, Hillier LW, Huang X, Minx P, Nelson JO, Pepin KH, Qin X, Sutton GG, Venter E, Walenz BP, Wallis JW, Worley KC, Yang SP, Jones SM, Marra MA, Rocchi M, Schein JE, Baertsch R, Clarke L, Csürös M, Glasscock J, Harris RA, Havlak P, Jackson AR, Jiang H, Liu Y, Messina DN, Shen Y, Song HX, Wylie T, Zhang L, Birney E, Han K, Konkel MK, Lee J, Smit AF, Ullmer B, Wang H, Xing J, Burhans R, Cheng Z, Karro JE, Ma J, Raney B, She X, Cox MJ, Demuth JP, Dumas LJ, Han SG, Hopkins J, Karimpour-Fard A, Kim YH, Pollack JR, Vinar T, Addo-Quaye C, Degenhardt J, Denby A, Hubisz MJ, Indap A, Kosiol C, Lahn BT, Lawson HA, Marklein A, Nielsen R, Vallender EJ, Clark AG, Ferguson B, Hernandez RD, Hirani K, Kehrer-Sawatzki H, Kolb J, Patil S, Pu LL, Ren Y, Smith DG, Wheeler DA, Schenck I, Ball EV, Chen R, Cooper DN, Giardine B, Hsu F, Kent WJ, Lesk A, Nelson DL, O'brien WE, Prüfer K, Stenson PD, Wallace JC, Ke H, Liu XM, Wang P, Xiang AP, Yang F, Barber GP, Haussler D, Karolchik D, Kern AD, Kuhn RM, Smith KE, Zwieg AS. 2007. Evolutionary and biomedical insights from the rhesus macaque genome. Science 316:222–234.
Gilad Y, Oshlack A, Smyth GK, Speed TP, White KP. 2006. Expression profiling in primates reveals a rapid evolution of human transcription factors. Nature 440:242–245.
Gillespie JH. 1991. The causes of molecular evolution. Oxford: Oxford University Press.
Gillespie JH. 2000a. The neutral theory in an infinite population. Gene 261:11–18.
Gillespie JH. 2000b. Genetic drift in an infinite population. The pseudohitchhiking model. Genetics 155:909–919.
Gillespie JH. 2001. Is the population size of a species relevant to its evolution? Evolution 55:2161–2169.
Gojobori J, Tang H, Akey JM, Wu CI. 2007. Adaptive evolution in humans revealed by the negative correlation between the polymorphism and fixation phases of evolution. Proc Natl Acad Sci USA 104:3907–3912.
Gould SJ, Lewontin RC. 1979. The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. Proc R Soc Lond B Biol Sci 205:581–598.
Grossman SR, Shylakhter I, Karlsson EK, Byrne EH, Morales S, Frieden G, Hostetter E, Angelino E, Garber M, Zuk O, Lander ES, Schaffner SF, Sabeti PC. 2010. A composite of multiple signals distinguishes causal variants in regions of positive selection. Science 327:883–886.
Hahn MW. 2008. Toward a selection theory of molecular evolution. Evolution 62:255–265.
Hahn MW, Demuth JP, Han SG. 2007. Accelerated rate of gene gain and loss in primates. Genetics 177:1941–1949.
Hallatschek O, Nelson DR. 2008. Gene surfing in expanding populations. Theor Popul Biol 73:158–170.
Hallatschek O, Nelson DR. 2009. Life at the front of an expanding population. Evolution 64:193–206.
Halligan DL, Oliver F, Eyre-Walker A, Harr B, Keightley PD. 2010. Evidence for pervasive adaptive protein evolution in wild mice. PLoS Genet 6:1–9.
Halushka MK, Fan JB, Bentley K, Hsie L, Shen N, Weder A, Cooper R, Lipshutz R, Chakravarti A. 1999. Patterns of single-nucleotide polymorphisms in candidate genes for blood-pressure homeostasis. Nat Genet 22:239–247.
Hamblin MT, DiRienzo A. 2000. Detection of the signature of natural selection in humans: evidence from the Duffy blood group locus. Am J Hum Genet 66:1669–1679.
Hamblin MT, Thompson EE, DiRienzo A. 2002. Complex signatures of natural selection at the Duffy blood group locus. Am J Hum Genet 70:369–383.
Hamdi HK, Nishio H, Tavis J, Zielinski R, Dugaiczyk A. 2000. *Alu*-mediated phylogenetic novelties in gene regulation and development. J Mol Biol 299:931–939.
Han MV, Demuth JP, McGrath CL, Casola C, Hahn MW. 2009. Adaptive evolution of young gene duplicates in mammals. Genome Res 19:859–867.
Handley LJ, Manica A, Goudet J, Balloux F. 2007. Going the distance: human population genetics in a clinal world. Trends Genet 23:432–439.
Harris EE, Hey J. 1999. X chromosome evidence for ancient human histories. Proc Natl Acad Sci USA 96:3320–3324.
Harris EE, Hey J. 2001. Human populations show reduced DNA sequence variation at the factor IX locus. Curr Biol 11:774–748.
Harris EE, Meyer D. 2006. The molecular signature of selection underlying human adaptations. Am J Phys Anthropol Suppl 43:89–130.
Hartl DL, Taubes CH. 1996. Compensatory nearly neutral mutations: selection without adaptation. J Theor Biol 182:303–309.
Hawks J, Wang ET, Cochran GM, Harpending HC, Moyzis RK. 2007. Recent acceleration of human adaptive evolution. Proc Natl Acad Sci USA 104:20753–20758.
Haygood R, Fedrigo O, Hanson B, Yokoyama KD, Wray GA. 2007. Promoter regions of many neural- and nutrition-related genes have experienced positive selection during human evolution. Nat Genet 39:1140–1144.
Hazkani-Covo E, Graur D. 2007. A comparative analysis of numt evolution in human and chimpanzee. Mol Biol Evol 24:13–18.
He X, Zhang J. 2005. Rapid subfunctionalization accompanied by prolonged and substantial neofunctionalization in duplicate gene evolution. Genetics 169:1157–1164.

Heger A, Ponting CP. 2007. Evolutionary rate analyses of orthologs and paralogs from 12 *Drosophila* genomes. Genome Res 17:1837–1849.

Hellmann I, Mang Y, Gu Z, Li P, de la Vega FM, Clark AG, Nielsen R. 2008. Population genetic analysis of shotgun assemblies of genomic sequences from multiple individuals. Genome Res 18:1020–1029.

Hellmann I, Prüfer K, Ji H, Zody MC, Pääbo S, Ptak SE. 2005. Why do human diversity levels vary at a megabase scale? Genome Res 15:1222–1231.

Hernandez RD, Hubisz MJ, Wheeler DA, Smith DG, Ferguson B, Rogers J, Nazareth L, Indap A, Bourquin T, McPherson J, Muzny D, Gibbs R, Nielsen R, Bustamante CD. 2007. Demographic histories and patterns of linkage disequilibrium in Chinese and Indian rhesus macaques. Science 316:240–243.

Hey J. 1999. The neutralist, the fly and the selectionist. Trends Ecol Evol 14:35–38.

Hill WG, Robertson A. 1966. The effect of linkage on limits to artificial selection. Genet Res 8:269–294.

Hofer T, Ray N, Wegmann D, Excoffier L. 2009. Large allele frequency differences between human continental groups are more likely to have occurred by drift during range expansions than by selection. Ann Hum Genet 73:95–108.

Hughes AL. 1994. The evolution of functionally novel proteins after gene duplication. Proc Biol Sci 256:119–124.

Hughes AL. 2007. Looking for Darwin in all the wrong places: the misguided quest for positive selection at the nucleotide sequence level. Heredity 99:364–373.

Hughes AL. 2008. Near neutrality: leading edge of the neutral theory of molecular evolution. Ann N Y Acad Sci 1133:162–179.

Hughes AL. 2009. Evolution in the post-genome era. Perspect Biol Med 52:332–337.

Hughes AL, Friedman R. 2009. More radical amino acid replacements in primates than in rodents: support for the evolutionary role of effective population size. Gene 440:50–56.

Hughes AL, Nei M. 1988. Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. Nature 335:167–170.

Hughes AL, Packer B, Welch R, Bergen AW, Chanock SJ, Yeager M. 2003. Widespread purifying selection at polymorphic sites in human protein-coding loci. Proc Natl Acad Sci USA 100:15754–15757.

Hughes AL, Packer B, Welch R, Bergen AW, Chanock SJ, Yeager M. 2005. Effects of natural selection on interpopulation divergence at polymorphic sites in human protein-coding loci. Genetics 170:1181–1187.

Hunley KL, Healy ME, Long JC. 2009. The global pattern of gene identity variation reveals a history of long-range migrations, bottlenecks, and local mate exchange: implications for biological race. Am J Phys Anthropol 139:35–46.

Hunt G. 2007. The relative importance of directional change, random walks, and stasis in the evolution of fossil lineages. Proc Natl Acad Sci USA 104:18404–18408.

International Human Genome Sequencing Consortium. 2004 Finishing the euchromatic sequence of the human genome. Nature. 431(7011):931–945.

Jensen-Seaman MI, Wildschutte JH, Soto-Calderon ID, Anthony NM. 2009. A comparative approach shows differences in patterns of numt insertion during hominoid evolution. J Mol Evol 68:688–699.

Kayser M, Brauer S, Stoneking M. 2003. A genome scan to detect candidate regions influenced by local natural selection in human populations. Mol Biol Evol 20:893–900.

Keightley PD, Kryukov GV, Sunyaev S, Halligan DL, Gaffney DJ. 2005a. Evolutionary constraints in conserved nongenic sequences of mammals. Genome Res 15:1373–1378.

Keightley PD, Lercher MJ, Eyre-Walker A. 2005b. Evidence for widespread degradation of gene control regions in hominid genomes. PLoS Biol 3:e42.

Keightley PD, Lercher MJ, Eyre-Walker A. 2006. Understanding the degradation of hominid gene control. PLoS Comput Biol 2:e19; author reply e26.

Khaitovich P, Enard W, Lachmann M, Paabo S. 2006. Evolution of primate gene expression. Nat Rev Genet 7:693–702.

Khaitovich P, Paabo S, Weiss G. 2005. Toward a neutral evolutionary model of gene expression. Genetics 170:929–939.

Khaitovich P, Weiss G, Lachmann M, Hellmann I, Enard W, Muetzel B, Wirkner U, Ansorge W, Paabo S. 2004. A neutral model of transcriptome evolution. PLoS Biol 2:e132.

Kim SY, Pritchard JK. 2007. Adaptive evolution of conserved noncoding elements in mammals. PLoS Genet 3:1572–1586.

Kimura M. 1968. Evolutionary rate at the molecular level. Nature 217:624–626.

Kimura M. 1969. The rate of molecular evolution considered from the standpoint of population genetics. Proc Natl Acad Sci USA 63:1181–1188.

Kimura M. 1983. The neutral theory of molecular evolution. Cambridge: Cambridge University Press.

Kimura M. 1991. Recent development of the neutral theory viewed from the Wrightian tradition of theoretical population genetics. Proc Natl Acad Sci USA 88:5969–5973.

King JL, Jukes TH. 1969. Non-Darwinian evolution. Science 164:788–798.

King MC, Wilson AC. 1975. Evolution at two levels in humans and chimpanzees. Science 188:107–116.

Klopfstein S, Currat M, Excoffier L. 2006. The fate of mutations surfing on the wave of a range expansion. Mol Biol Evol 23:482–490.

Kondrashov AS. 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 AS, Sunyaev S, Kondrashov FA. 2002. Dobzhansky-Muller incompatibilities in protein evolution. Proc Natl Acad Sci USA 99:14878–14883.

Kondrashov FA, Kondrashov AS. 2006. Role of selection in fixation of gene duplications. J Theor Biol 239:141–151.

Kondrashov FA, Koonin EV. 2004. A common framework for understanding the origin of genetic dominance and evolutionary fates of gene duplications. Trends Genet 20:287–290.

Koonin EV. 2004. A non-adaptationist perspective on evolution of genomic complexity or the continued dethroning of man. Cell Cycle 3:280–285.

Koonin EV. 2009a. Darwinian evolution in the light of genomics. Nucleic Acids Res 37:1011–1034.

Koonin EV. 2009b. Evolution of genome architecture. Int J Biochem Cell Biol 41:298–306.

Kosiol C, Vinar T, da Fonseca RR, Hubisz MJ, Bustamante CD, Nielsen R, Siepel A. 2008. Patterns of positive selection in six mammalian genomes. PLoS Genet 4:e1000144.

Kreitman M. 1996. The neutral theory is dead. Long live the neutral theory. Bioessays 18:678–683; discussion 683.

Krull M, Brosius J, Schmitz J. 2005. *Alu*-SINE exonization: en route to protein-coding function. Mol Biol Evol 22:1702–1711.

Kryukov GV, Schmidt S, Sunyaev S. 2005. Small fitness effect of mutations in highly conserved non-coding regions. Hum Mol Genet 14:2221–2229.

Kulathinal RJ, Bettencourt BR, Hartl DL. 2004. Compensated deleterious mutations in insect genomes. Science 306:1553–1554.

Lande R. 1976. Natural selection and random genetic drift in phenotypic evolution. Evolution 30:314–334.

Lande R. 1979. Quantitative genetic analysis of multivariate evolution, applied to brain: body size, allometry. Evolution 33:402–416.

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann N, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Shownkeen R, Sims S, Waterston RH, Wilson RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT, Pepin KH, Gish WR, Chissoe SL, Wendl MC, Delehaunty

KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng JF, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M, et al. 2001. Initial sequencing and analysis of the human genome. Nature 409:860–921.

Langley CH, Montgomery E, Hudson R, Kaplan N, Charlesworth B. 1988. On the role of unequal exchange in the containment of transposable element copy number. Genet Res 52:223–235.

Lee J, Han K, Meyer TJ, Kim HS, Batzer MA. 2008. Chromosomal inversions between human and chimpanzee lineages caused by retrotransposons. PLoS One 3:e4047.

Lemos B, Meiklejohn CD, Caceres M, Hartl DL. 2005. Rates of divergence in gene expression profiles of primates, mice, and flies: stabilizing selection and variability among functional categories. Evolution 59:126–137.

Lenski RE, Travisano M. 1994. Dynamics of adaptation and diversification: a 10,000-generation experiment with bacterial populations. Proc Natl Acad Sci USA 91:6808–6814.

Lev-Maor G, Ram O, Kim E, Sela N, Goren A, Levanon EY, Ast G. 2008. Intronic *Alu*s influence alternative splicing. PLoS Genet 4:e1000204.

Lev-Maor G, Sorek R, Shomron N, Ast G. 2003. The birth of an alternatively spliced exon: 3′ splice-site selection in *Alu* exons. Science 300:1288–1291.

Lewontin RC. 1974. The genetic basis of evolutionary change. New York: Columbia University Press.

Lewontin RC, Krakauer J. 1973. Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. Genetics 74:175–195.

Li JZ, Absher DM, Tang H, Southwick AM, Casto AM, Ramachandran S, Cann HM, Barsh GS, Feldman M, Cavalli-Sforza LL, Myers RM. 2008. Worldwide human relationships inferred from genome-wide patterns of variation. Science 319:1100–1104.

Lin L, Jiang P, Shen S, Sato S, Davidson BL, Xing Y. 2009. Large-scale analysis of exonized mammalian-wide interspersed repeats in primate genomes. Hum Mol Genet 18:2204–2214.

Lin L, Shen S, Tye A, Cai JJ, Jiang P, Davidson BL, Xing Y. 2008. Diverse splicing patterns of exonized *Alu* elements in human tissues. PLoS Genet 4:e1000225.

Lindblad-Toh K, Wade CM, Mikkelsen TS, Karlsson EK, Jaffe DB, Kamal M, Clamp M, Chang JL, Kulbokas EJ 3rd, Zody MC, Mauceli E, Xie X, Breen M, Wayne RK, Ostrander EA, Ponting CP, Galibert F, Smith DR, DeJong PJ, Kirkness E, Alvarez P, Biagi T, Brockman W, Butler J, Chin CW, Cook A, Cuff J, Daly MJ, DeCaprio D, Gnerre S, Grabherr M, Kellis M, Kleber M, Bardeleben C, Goodstadt L, Heger A, Hitte C, Kim L, Koepfli KP, Parker HG, Pollinger JP, Searle SM, Sutter NB, Thomas R, Webber C, Baldwin J, Abebe A, Abouelleil A, Aftuck L, Ait-Zahra M, Aldredge T, Allen N, An P, Anderson S, Antoine C, Arachchi H, Aslam A, Ayotte L, Bachantsang P, Barry A, Bayul T, Benamara M, Berlin A, Bessette D, Blitshteyn B, Bloom T, Blye J, Boguslavskiy L, Bonnet C, Boukhgalter B, Brown A, Cahill P, Calixte N, Camarata J, Cheshatsang Y, Chu J, Citroen M, Collymore A, Cooke P, Dawoe T, Daza R, Decktor K, DeGray S, Dhargay N, Dooley K, Dorje P, Dorjee K, Dorris L, Duffey N, Dupes A, Egbiremolen O, Elong R, Falk J, Farina A, Faro S, Ferguson D, Ferreira P, Fisher S, FitzGerald M, Foley K, et al. 2005. Genome sequence, comparative analysis and haplotype structure of the domestic dog. Nature 438:803–819.

Liu G, Zhao S, Bailey JA, Sahinalp SC, Alkan C, Tuzun E, Green ED, Eichler EE. 2003. Analysis of primate genomic variation reveals a repeat-driven expansion of the human genome. Genome Res 13:358–368.

Liu H, Prugnolle F, Manica A, Balloux F. 2006. A geographically explicit genetic model of worldwide human-settlement history. Am J Hum Genet 79:230–237.

Lohmueller KE, Indap AR, Schmidt S, Boyko AR, Hernandez RD, Hubisz MJ, Sninsky JJ, White TJ, Sunyaev SR, Nielsen R, Clark AG, Bustamante CD. 2008. Proportionally more deleterious genetic variation in European than in African populations. Nature 451:994–997.

Ludwig MZ, Bergman C, Patel NH, Kreitman M. 2000. Evidence for stabilizing selection in a eukaryotic enhancer element. Nature 403:564–567.

Lynch M. 1990. The rate of morphological evolution in mammals from the standpoint of the neutral expectation. Am Nat 136:727–741.

Lynch M. 2007a. The origins of genome architecture. Sunderland, MA: Sinauer Associates, Inc.

Lynch M. 2007b. The frailty of adaptive hypotheses for the origins of organismal complexity. Proc Natl Acad Sci USA 104(Suppl 1):8597–8604.

Lynch M, Conery JS. 2003. The origins of genome complexity. Science 302:1401–1404.

Lynch M, Conery JS. 2004. Testing genome complexity. Science 304:389–390.

Lynch M, Force A. 2000. The probability of duplicate gene preservation by subfunctionalization. Genetics 154:459–473.

Lynch M, Hill W. 1986. Phenotypic evolution by neutral mutation. Evolution 40:915–935.

Makalowski W. 2003. Genomics. Not junk after all. Science 300:1246–1247.

Mallick S, Gnerre S, Muller P, Reich D. 2009. The difficulty of avoiding false positives in genome scans for natural selection. Genome Res 19:922–933.

Marques-Bonet T, Girirajan S, Eichler EE. 2009a. The origins and impact of primate segmental duplications. Trends Genet 25:443–454.

Marques-Bonet T, Kidd JM, Ventura M, Graves TA, Cheng Z, Hillier LW, Jiang Z, Baker C, Malfavon-Borja R, Fulton LA, Alkan C, Aksay G, Girirajan S, Siswara P, Chen L, Cardone MF, Navarro A, Mardis ER, Wilson RK, Eichler EE. 2009b. A burst of segmental duplications in the genome of the African great ape ancestor. Nature 457:877–881.

Marroig G, Cheverud JM. 2004. Did natural selection or genetic drift produce the cranial diversification of neotropical monkeys? Am Nat 163:417–428.

Maslin MA, Christensen B. 2007. Tectonics, orbital forcing, global climate change, and human evolution in Africa: introduction to the African paleoclimate special volume. J Hum Evol 53:443–464.

Maynard Smith J, Haigh J. 1974. The hitch-hiking effect of a favourable gene. Genet Res 23:23–35.

McDonald JH, Kreitman M. 1991. Adaptive protein evolution at the *Adh* locus in *Drosophila*. Nature 351:652–654.

McVicker G, Gordon D, Davis C, Green P. 2009. Widespread genomic signatures of natural selection in hominid evolution. PLoS Genet 5:e1000471.

Mekel-Bobrov N, Gilbert SL, Evans PD, Vallender EJ, Anderson JR, Hudson RR, Tishkoff SA, Lahn BT. 2005. Ongoing adaptive evolution of ASPM, a brain size determinant in *Homo sapiens*. Science 309:1720–1722.

Mikkelsen TS, Hillier LW, Eichler EE, et al. 2005. Initial sequence of the chimpanzee genome and comparison with the human genome. Nature 437:69–87.

Millien V. 2006. Morphological evolution is accelerated among island mammals. PLoS Biol 4:e321.

Mills RE, Bennett EA, Iskow RC, Luttig CT, Tsui C, Pittard WS, Devine SE. 2006. Recently mobilized transposons in the human and chimpanzee genomes. Am J Hum Genet 78:671–679.

Miyata T, Miyazawa S, Yasunaga T. 1979. Two types of amino acid substitutions in protein evolution. J Mol Evol 12:219–236.

Modrek B, Lee CJ. 2003. Alternative splicing in the human, mouse and rat genomes is associated with an increased frequency of exon creation and/or loss. Nat Genet 34:177–180.

Myles S, Tang K, Somel M, Green RE, Kelso J, Stoneking M. 2008. Identification and analysis of genomic regions with large between-population differentiation in humans. Ann Hum Genet 72:99–110.

Nachman MW, Bauer VL, Crowell SL, Aquadro CF. 1998. DNA variability and recombination rates at X-linked loci in humans. Genetics 150:1133–1141.

Nei M. 2005. Selectionism and neutralism in molecular evolution. Mol Biol Evol 22:2318–2342.

Nei M, Niimura Y, Nozawa M. 2008. The evolution of animal chemosensory receptor gene repertoires: roles of chance and necessity. Nat Rev Genet 9:951–963.
Nei M, Rooney AP. 2005. Concerted and birth-and-death evolution of multigene families. Annu Rev Genet 39:121–152.
Nguyen DQ, Webber C, Hehir-Kwa J, Pfundt R, Veltman J, Ponting CP. 2008. Reduced purifying selection prevails over positive selection in human copy number variant evolution. Genome Res 18:1711–1723.
Nielsen R. 2001. Statistical tests of selective neutrality in the age of genomics. Heredity 86(Pt 6):641–647.
Nielsen R. 2005. Molecular signatures of natural selection. Annu Rev Genet 39:197–218.
Nielsen R, Yang Z. 1998. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. Genetics 148:929–936.
Nielsen R, Bustamante C, Clark AG, Glanowski S, Sackton TB, Hubisz MJ, Fledel-Alon A, Tanenbaum DM, Civello D, White TJ, Sninsky JJ, Adams MD, Cargill M. 2005. A scan for positively selected genes in the genomes of humans and chimpanzees. PLoS Biol 3:e170.
Nielsen R, Hellmann I, Hubisz M, Bustamante C, Clark AG. 2007. Recent and ongoing selection in the human genome. Nat Rev Genet 8:857–868.
Noonan JP, Coop G, Kudaravalli S, Smith D, Krause J, Alessi J, Chen F, Platt D, Pääbo S, Pritchard JK, Rubin EM. 2006. Sequencing and analysis of Neanderthal genomic DNA. Science 314:1113–1118.
Novembre J, DiRienzo A. 2009. Spatial patterns of variation due to natural selection in humans. Nat Rev Genet 10:745–755.
Nozawa M, Kawahara Y, Nei M. 2007. Genomic drift and copy number variation of sensory receptor genes in humans. Proc Natl Acad Sci USA 140:20421–20426.
Nozawa M, Nei M. 2008. Genomic drift and copy number variation of chemosensory receptor genes in humans and mice. Cytogenet Genome Res 123:263–269.
Ohno S. 1970. Evolution by gene duplication. Berlin: Springer-Verlag.
Ohta T. 1973. Slightly deleterious mutant substitutions in evolution. Nature 246:96–98.
Ohta T. 1974. Mutational pressure as the main cause of molecular evolution and polymorphism. Nature 246:96–98.
Ohta T. 2002. Near-neutrality in evolution of genes and gene regulation. Proc Natl Acad Sci USA 99:16134–16137.
Ohta T. 2003. Origin of the neutral and nearly neutral theories of evolution. J Biosci 28:371–377.
Patterson N, Richter DJ, Gnerre S, Lander ES, Reich D. 2006. Genetic evidence for complex speciation of humans and chimpanzees. Nature 441:1103–1108.
Pérez-Claros JA, Aledo JC. 2007. Comment on "Morphological evolution is accelerated among island mammals." PLoS Biol 5:e180; author reply e176.
Perry GH, Dominy NJ, Claw KG, Lee AS, Fiegler H, Redon R, Werner J, Villanea FA, Mountain JL, Misra R, Carter NP, Lee C, Stone AC. 2007. Diet and the evolution of human amylase gene copy number variation. Nat Genet 39:1256–1260.
Perry GH, Yang F, Marques-Bonet T, Murphy C, Fitzgerald T, Lee AS, Hyland C, Stone AC, Hurles ME, Tyler-Smith C, Eichler EE, Carter NP, Lee C, Redon R. 2008. Copy number variation and evolution in humans and chimpanzees. Genome Res 18:1698–1710.
Pickrell JK, Coop G, Novembre J, Kudaravalli S, Li JZ, Absher D, Srinivasan BS, Barsh GS, Myers RM, Feldman MW, Pritchard JK. 2009. Signals of recent positive selection in a worldwide sample of human populations. Genome Res. 19:826–837.
Ponting CP, Lunter G. 2006. Signatures of adaptive evolution within human non-coding sequence. Hum Mol Genet 15(Spec No 2):R170–175.
Pool JE, Hellmann I, Jensen JD, Nielsen R. 2010. Population genetic inference from genomic sequence variation. Genome Res 20:291–300.
Poon A, Davis BH, Chao L. 2005. The coupon collector and the suppressor mutation: estimating the number of compensatory mutations by maximum likelihood. Genetics 170:1323–1332.
Poon A, Otto SP. 2000. Compensating for our load of mutations: freezing the meltdown of small populations. Evolution 54:1467–1479.
Powers SK, Jackson MJ. 2008. Exercise-induced oxidative stress: cellular mechanisms and impact on muscle force production. Physiol Rev 88:1243–1276.
Premo LS, Hublin JJ. 2009. Culture, population structure, and low genetic diversity in Pleistocene hominins. Proc Natl Acad Sci USA 106:33–37.
Prud'homme B, Gompel N, Carroll SB. 2007. Emerging principles of regulatory evolution. Proc Natl Acad Sci USA 104(Suppl 1):8605–8612.
Prugnolle F, Manica A, Balloux F. 2005. Geography predicts neutral genetic diversity of human populations. Curr Biol 15:R159–R160.
Przeworski M, Hudson RR, DiRienzo A. 2000. Adjusting the focus on human variation. Trends Genet 16:296–302.
Qian W, Zhang J. 2008. Gene dosage and gene duplicability. Genetics 179:2319–2324.
Ramachandran S, Deshpande O, Roseman CC, Rosenberg NA, Feldman MW, Cavalli-Sforza LL. 2005. Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. Proc Natl Acad Sci USA 102:15942–15947.
Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, Fiegler H, Shapero MH, Carson AR, Chen W, Cho EK, Dallaire S, Freeman JL, Gonzalez JR, Gratacos M, Huang J, Kalaitzopoulos D, Komura D, MacDonald JR, Marshall CR, Mei R, Montgomery L, Nishimura K, Okamura K, Shen F, Somerville MJ, Tchinda J, Valsesia A, Woodwark C, Yang F, Zhang J, Zerjal T, Armengol L, Conrad DF, Estivill X, Tyler-Smith C, Carter NP, Aburatani H, Lee C, Jones KW, Scherer SW, Hurles ME. 2006. Global variation in copy number in the human genome. Nature 444:444–454.
Reed FA, Akey JM, Aquadro CF. 2005. Fitting background-selection predictions to levels of nucleotide variation and divergence along the human autosomes. Genome Res 15:1211–1221.
Reed FA, Aquadro CF. 2006. Mutation, selection and the future of human evolution. Trends Genet 22:479–484.
Relethford JH. 1994. Craniometric variation among modern human populations. Am J Phys Anthropol 95:53–62.
Relethford JH. 2002. Apportionment of global human genetic diversity based on craniometrics and skin color. Am J Phys Anthropol 118:393–398.
Risch N, Tang H, Katzenstein H, Ekstein J. 2003. Geographic distribution of disease mutations in the Ashkenazi Jewish population supports genetic drift over selection. Am J Hum Genet 72:812–822.
Roseman CC. 2004. Detecting interregionally diversifying natural selection on modern human cranial form by using matched molecular and morphometric data. Proc Natl Acad Sci USA 101:12824–12829.
Roseman CC, Weaver TD. 2007. Molecules versus morphology? Not for the human cranium. Bioessays 29:1185–1188.
Sabeti PC, Schaffner SF, Fry B, Lohmueller J, Varilly P, Shamovsky O, Palma A, Mikkelsen TS, Altshuler D, Lander ES. 2006. Positive natural selection in the human lineage. Science 312:1614–1620.
Sawyer SA, Parsch J, Zhang Z, Hartl DL. 2007. Prevalence of positive selection among nearly neutral amino acid replacements in *Drosophila*. Proc Natl Acad Sci USA 104:6504–6510.
Schultz ST, Lynch M. 1997. Deleterious mutation and extinction: effects of variable mutational effects, synergistic epistasis, beneficial mutations, and degree of outcrossing. Evolution 51:1363–1371.
Sella G, Petrov DA, Przeworski M, Andolfatto P. 2009. Pervasive natural selection in the *Drosophila* genome? PLoS Genet 5:e1000495.
Siepel A. 2009. Phylogenomics of primates and their ancestral populations. Genome Res 19:1929–1941.
Slatkin M. 2004. A population-genetic test of founder effects and implications for Ashkenazi Jewish diseases. Am J Hum Genet 75:282–293.
Smith NG, Eyre-Walker A. 2002. Adaptive protein evolution in *Drosophila*. Nature 415:1022–1024.

Smouse PE, Vitzthum VJ, Neel JV. 1981. The impact of random and lineal fission of the genetic divergence of small human groups: a case study among the Yanomama. Genetics 98: 179–197.

Sokal RR, Smouse PE, Neel JV. 1986. The genetic structure of a tribal population, the Yanomama Indians. XV. Patterns inferred by autocorrelation analysis. Genetics 114:259–287.

Soranzo N, Bufe B, Sabeti PC, Wilson JF, Weale ME, Marguerie R, Meyerhof W, Goldstein DB. 2005. Positive selection on a high-sensitivity allele of the human bitter-taste receptor TAS2R16. Curr Biol 15:1257–1265.

Sorrentino S, Libonati M. 1997. Structure-function relationships in human ribonucleases: main distinctive features of the major RNase types. FEBS Lett 404:1–5.

Steiper ME, Young NM. 2006. Primate molecular divergence dates. Mol Phylogenet Evol 41:384–394.

Stoltzfus A. 1999. On the possibility of constructive neutral evolution. J Mol Evol 49:169–181.

Stoltzfus A. 2006. Mutationism and the dual causation of evolutionary change. Evol Dev 8:304–317.

Stoltzfus A, Yampolsky LY. 2009. Climbing mount probable: mutation as a cause of nonrandomness in evolution. J Hered 100:637–647.

Tang K, Thornton KR, Stoneking M. 2007. A new approach for using genome scans to detect recent positive selection in the human genome. PLoS Biol 5:e171.

Taylor JS, Raes J. 2004. Duplication and divergence: the evolution of new genes and old ideas. Annu Rev Genet 38:615–643.

Taylor MS, Kai C, Kawai J, Carninci P, Hayashizaki Y, Semple CA. 2006. Heterotachy in mammalian promoter evolution. PLoS Genet 2:e30.

Tenesa A, Navarro P, Hayes BJ, Duffy DL, Clarke GM, Goddard ME, Visscher PM. 2007. Recent human effective population size estimated from linkage disequilibrium. Genome Res 17:520–526.

Thornburg BG, Gotea V, Makalowski W. 2006. Transposable elements as a significant source of transcription regulating signals. Gene 365:104–110.

Tishkoff SA, Reed FA, Ranciaro A, Voight BF, Babbitt CC, Silverman JS, Powell K, Mortensen HM, Hirbo JB, Osman M, Ibrahim M, Omar SA, Lema G, Nyambo TB, Ghori J, Bumpstead S, Pritchard JK, Wray GA, Deloukas P. 2007. Convergent adaptation of human lactase persistence in Africa and Europe. Nat Genet 39:31–40.

Trauth MH, Larrasoana JC, Mudelsee M. 2009. Trends, rhythms and events in Plio-Pleistocene African climate. Quart Sci Rev 28:399–411.

Trauth MH, Maslin MA, Deino A, Strecker MR. 2005. Late Cenozoic moisture history of East Africa. Science 309:2051–2053.

Trauth MH, Maslin MA, Deino AL, Strecker MR, Bergner AG, Duhnforth M. 2007. High- and low-latitude forcing of Plio-Pleistocene East African climate and human evolution. J Hum Evol 53:475–486.

Travis JM, Munkemuller T, Burton OJ, Best A, Dytham C, Johst K. 2007. Deleterious mutations can surf to high densities on the wave front of an expanding population. Mol Biol Evol 24:2334–2343.

Vinogradov AE. 2004. Testing genome complexity. Science 304:389–390; author reply 389–390.

Voight BF, Kudaravalli S, Wen X, Pritchard JK. 2006. A map of recent positive selection in the human genome. PLoS Biol 4:e72.

Wall JD. 2003. Estimating ancestral population sizes and divergence times. Genetics 163:395–404.

Warren WC, Hillier LW, Marshall Graves JA, Birney E, Ponting CP, Grutzner F, Belov K, Miller W, Clarke L, Chinwalla AT, Yang SP, Heger A, Locke DP, Miethke P, Waters PD, Veyrunes F, Fulton L, Fulton B, Graves T, Wallis J, Puente XS, Lopez-Otin C, Ordonez GR, Eichler EE, Chen L, Cheng Z, Deakin JE, Alsop A, Thompson K, Kirby P, Papenfuss AT, Wakefield MJ, Olender T, Lancet D, Huttley GA, Smit AF, Pask A, Temple-Smith P, Batzer MA, Walker JA, Konkel MK, Harris RS, Whittington CM, Wong ES, Gemmell NJ, Buschiazzo E, Vargas Jentzsch IM, Merkel A, Schmitz J, Zemann A, Churakov G, Kriegs JO, Brosius J, Murchison EP, Sachidanandam R, Smith C, Hannon GJ, Tsend-Ayush E, McMillan D, Attenborough R, Rens W, Ferguson-Smith M, Lefevre CM, Sharp JA, Nicholas KR, Ray DA, Kube M, Reinhardt R, Pringle TH, Taylor J, Jones RC, Nixon B, Dacheux JL, Niwa H, Sekita Y, Huang X, Stark A, Kheradpour P, Kellis M, Flicek P, Chen Y, Webber C, Hardison R, Nelson J, Hallsworth-Pepin K, Delehaunty K, Markovic C, Minx P, Feng Y, Kremitzki C, Mitreva M, Glasscock J, Wylie T, Wohldmann P, Thiru P, Nhan MN, Pohl CS, Smith SM, Hou S, Nefedov M, et al. 2008. Genome analysis of the platypus reveals unique signatures of evolution. Nature 453:175–183.

Weaver TD. 2009. Out of Africa: modern human origins special feature: the meaning of Neandertal skeletal morphology. Proc Natl Acad Sci USA 106:16028–16033.

Weaver TD, Roseman CC, Stringer CB. 2007. Were Neandertal and modern human cranial differences produced by natural selection or genetic drift? J Hum Evol 53:135–145.

Weber MJ. 2006. Mammalian small nucleolar RNAs are mobile genetic elements. PLoS Genet 2:e205.

Welch JJ. 2006. Estimating the genomewide rate of adaptive protein evolution in *Drosophila*. Genetics 173:821–837.

Whitehead A, Crawford DL. 2006. Variation within and among species in gene expression: raw material for evolution. Mol Ecol 15:1197–1211.

Whitlock MC. 2000. Fixation of new alleles and the extinction of small populations: drift load, beneficial alleles, and sexual selection. Evolution 54:1855–1861.

Whitlock MC. 2003. Fixation probability and time in subdivided populations. Genetics 164:767–779.

Williford A, Comeron JM. 2010. Local effects of limited recombination: historical perspective and consequences for population estimates of adaptive evolution. J Hered 101(Suppl 1):S127–S34.

Wong WS, Nielsen R. 2004. Detecting selection in noncoding regions of nucleotide sequences. Genetics 167:949–958.

Woolfit M, Bromham L. 2005. Population size and molecular evolution on islands. Proc Biol Sci 272:2277–2282.

Wray GA. 2007. The evolutionary significance of cis-regulatory mutations. Nat Rev Genet 8:206–216.

Wright S. 1951. The genetical structure of populations. Ann Eugen 15:323–354.

Xue Y, Daly A, Yngvadottir B, Liu M, Coop G, Kim Y, Sabeti P, Chen Y, Stalker J, Huckle E, Burton J, Leonard S, Rogers J, Tyler-Smith C. 2006. Spread of an inactive form of caspase-12 in humans is due to recent positive selection. Am J Hum Genet 78:659–670.

Yang Z, Nielsen R. 2002. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. Mol Biol Evol 19:908–917.

Yi S, Streelman JT. 2005. Genome size is negatively correlated with effective population size in ray-finned fish. Trends Genet 21:643–646.

Yu N, Zhao Z, Fu YX, Sambuughin N, Ramsay M, Jenkins T, Leskinen E, Patthy L, Jorde LB, Kuromori T, Li WH. 2001. Global patterns of human DNA sequence variation in a 10-kb region on chromosome 1. Mol Biol Evol 18:214–222.

Zhang J. 2006. Parallel adaptive origins of digestive RNases in Asian and African leaf monkeys. Nat Genet 38:819–823.

Zhang J. 2007. The drifting human genome. Proc Natl Acad Sci USA 104:20147–20148.

Zhang J, Zhang YP, Rosenberg HF. 2002. Adaptive evolution of a duplicated pancreatic ribonuclease gene in a leaf-eating monkey. Nat Genet 30:411–415.

Zhang L, Li WH. 2005. Human SNPs reveal no evidence of frequent positive selection. Mol Biol Evol 22:2504–2507.