Selection and Adaptation in the Human Genome

Wenqing Fu and Joshua M. Akey

Department of Genome Sciences, University of Washington, Seattle, Washington 98195-5065; email: akeyj@uw.edu

Annu. Rev. Genomics Hum. Genet. 2013. 14:467–89

First published online as a Review in Advance on July 3, 2013

The *Annual Review of Genomics and Human Genetics* is online at genom.annualreviews.org

This article's doi: 10.1146/annurev-genom-091212-153509

Keywords

adaptive evolution, purifying selection, hard sweep, soft sweep, polygenic selection, mutation load

Abstract

An enduring goal of evolutionary biology is to understand how natural selection has shaped patterns of polymorphism and divergence within and between species and to map the genetic basis of adaptations. The rapid maturation of next-generation sequencing technology has generated a deluge of genomics data from nonhuman primates, extinct hominins, and diverse human populations. These emerging genome data sets have simultaneously broadened our understanding of human evolution and sharply defined existing gaps in knowledge about the mechanistic basis of evolutionary change. In this review, we summarize recent insights into how natural selection has influenced the human genome across different timescales. Although the path to a more comprehensive understanding of selection and adaptation in humans remains arduous, some general insights are beginning to emerge, such as the importance of adaptive regulatory evolution, the absence of pervasive classic selective sweeps, and the potential roles that selection from standing variation and polygenic adaptation have likely played in recent human evolutionary history.

INTRODUCTION

Fitness: a central, albeit nebulous, concept in evolutionary theory that quantifies the ability of individuals to survive and contribute to the next generation

Positive selection: natural selection that acts to increase the prevalence of advantageous mutations

Purifying selection: natural selection that acts to eliminate deleterious mutations

The universe of human phenotypic variation is vast, and individuals vary in innumerable pathogenic and nonpathogenic ways (110). Although the environment plays an important role, genetic variation is also a major determinant of phenotypic variation. Thus, understanding the evolutionary forces that create, maintain, and shape patterns of human genetic variation is fundamentally important to explaining the spectrum of human diversity. Population genetics, which emerged from the early seminal contributions of Fisher, Wright, and Haldane (124), provides the necessary intellectual and quantitative framework to delineate the evolutionary determinants of polymorphism within and divergence between species.

A key insight from theoretical population genetics is that both random (genetic drift) and deterministic (natural selection) forces govern changes in allele frequency within and between populations. Natural selection (34, 35) occurs when there are differences in fitness among individuals and mechanistically involves the differential reproduction and survival of genotypes. However, even in the absence of natural selection, allele frequencies are expected to change from one generation to the next in populations of finite size because of the stochastic effects of genetic drift. Indeed, the neutral theory of molecular evolution posits that the vast majority of evolution at the molecular level is caused by random genetic drift (76, 77). Ohta (109) subsequently extended the neutral theory to allow for the fixation of mutations with small fitness effects (referred to as the nearly neutral theory). The neutral and nearly neutral theories are commonly misunderstood as denying the existence of adaptive mutations, but these theories posit simply that adaptive mutations constitute a minority of the polymorphisms present within a population or the fixations observed between species.

Despite the predominant role that genetic drift has likely played in shaping patterns of human genomic variation, there has been intense interest in identifying regions of the human genome that have been subject to natural selection (reviewed in 4, 9, 17, 73, 105) and, more specifically, the loci that harbor adaptive alleles. Adaptively evolving loci have been sought with such vigor because they may yield insights into the genetic basis of human-specific phenotypes and provide a window into human history. For example, strong signatures of adaptive evolution are present in genes that confer resistance to malaria and other infectious diseases (33, 55, 148), allow the digestion of milk and dairy products into adulthood (14, 121, 139, 147), influence skin pigmentation (82, 92), and enable individuals to live at high altitude (15, 115, 161, 163). These examples reveal the potent selective pressures exerted by pathogens and climatic conditions, and detailed analyses of genetic variation at the responsible loci provide insight into where and over what time frames selection has acted. Indeed, the human genome has been subjected to selective pressure over many different timescales, and the study of particular epochs can be informative about specific evolutionary processes or critical periods of innovation (**Figure 1**).

The search for signatures of natural selection in the human genome has entered a golden age (125) with the advent of next-generation sequencing technology. The emerging deluge of whole-genome sequences, including those from nonhuman primates and extinct archaic hominins (27, 51, 88, 97, 128), provides an extraordinary opportunity to develop a more comprehensive and mechanistic understanding of how natural selection has shaped human genomic variability. In this review, we discuss specific examples of loci that harbor signatures of natural selection, general themes that have emerged from genome-wide scans of selection, and outstanding gaps in knowledge. Natural selection operates in different guises, and we focus on the consequences of both selection for advantageous variants (positive or Darwinian selection) and selection against deleterious variants (purifying selection). Throughout, we highlight different mechanisms of natural

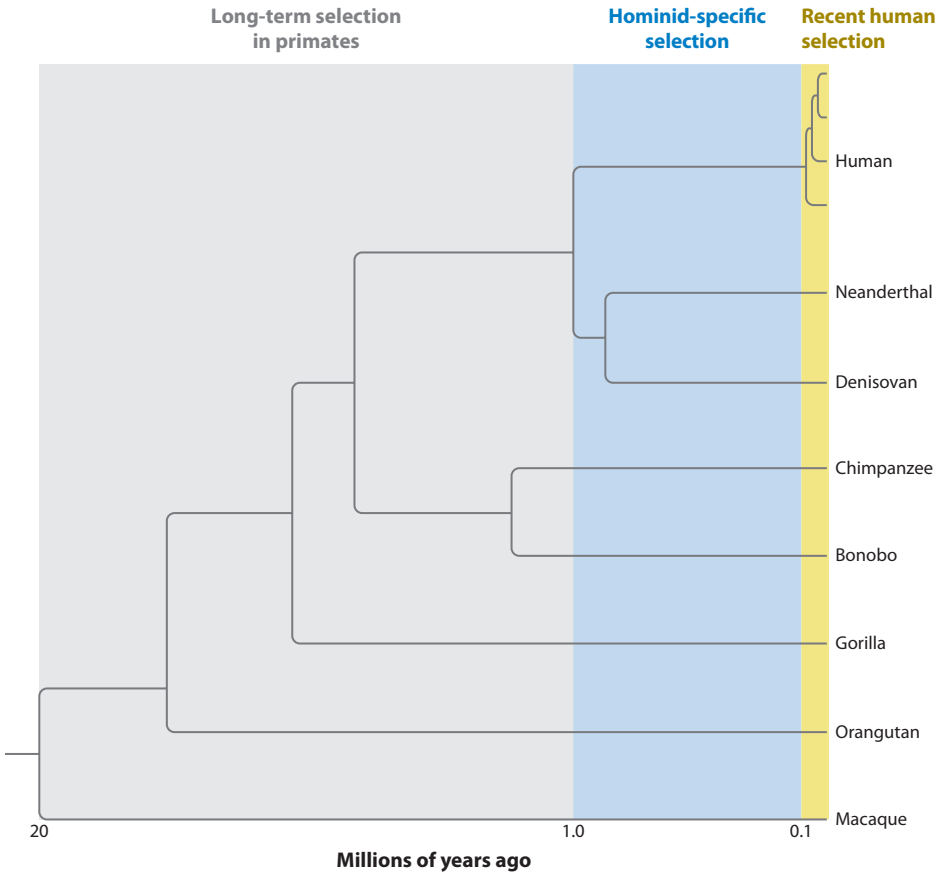

Figure 1

Natural selection acting on the human genome across many timescales. The study of particular epochs can be informative about specific evolutionary processes or periods of suspected evolutionary innovations. For example, long-term selective pressure occurring throughout primate evolution can provide insights into conserved sequences and loci subjected to balancing selection, and recent selection can pinpoint loci that have conferred selective advantages for novel environments as modern humans migrated out of Africa. The multiple lineages for humans denote polymorphism data.

selection (summarized in **Figure 2**) and the signatures they are expected to impart on patterns of DNA sequence variation.

DETECTING ADAPTIVE EVOLUTION FROM INTERSPECIFIC DIVERGENCE

Signatures of Adaptive Protein Divergence

Comparing orthologous DNA sequences from multiple species is a powerful way to detect selection acting over longer evolutionary timescales (**Figure 1**). A common approach using comparative sequence data is to contrast the rates of evolutionary divergence between two classes of sites. For example, the evolutionary forces acting on protein-coding sequences can be studied by looking at

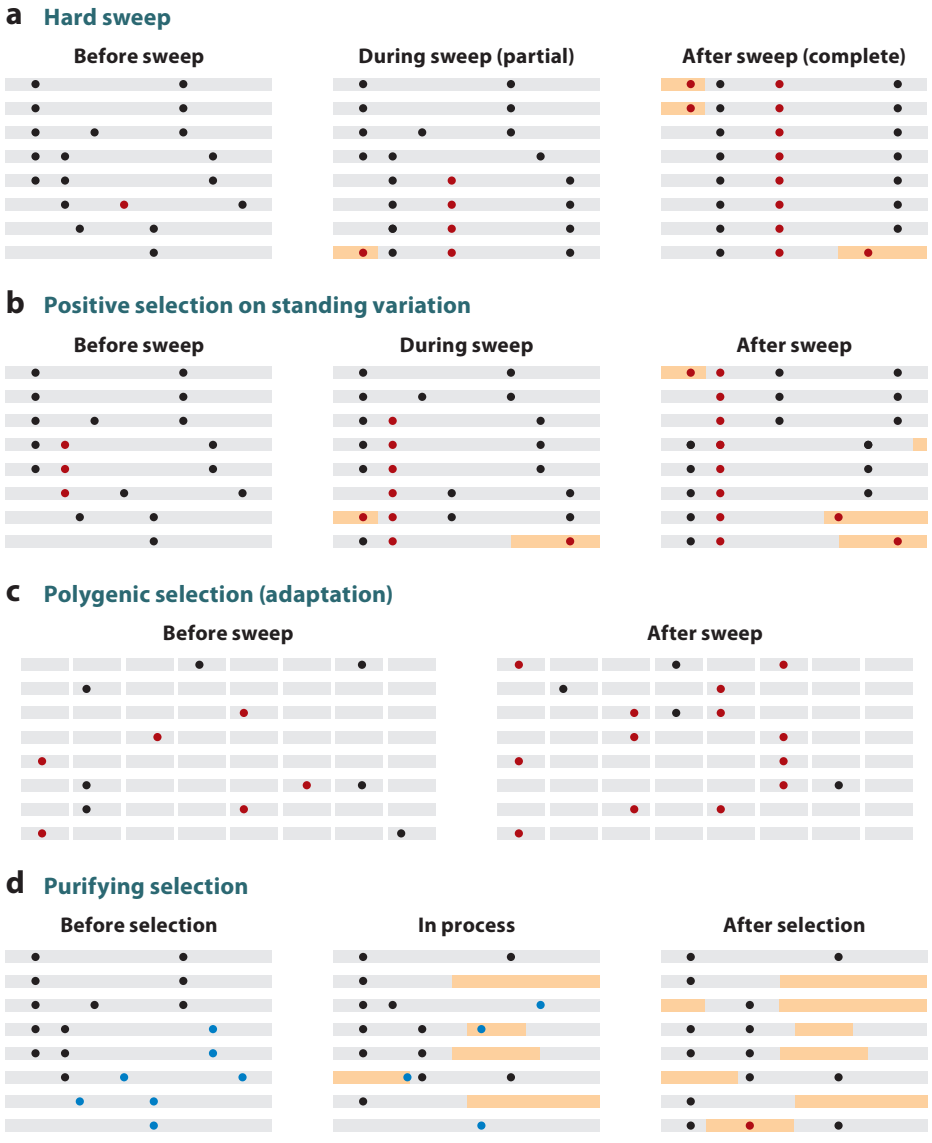

Figure 2

Different modes of natural selection. Lines represent chromosomes, and circles represent derived single-nucleotide polymorphism alleles. (*a*) A hard sweep, where a new advantageous mutation (*red*) appears initially on one chromosome and over time is driven to fixation by selection, creating unusually long haplotypes of low diversity. (*b*) Positive selection on standing variation, where selection acts on a preexisting variant (*red*) that before becoming advantageous is either neutral or mildly deleterious. A selective sweep from standing variation drags along more polymorphisms at linked sites than a sweep from a new mutation. (*c*) Polygenic selection (adaptation), where positive selection acts on multiple advantageous alleles (*red*) existing at low to moderate frequencies at various loci across the genome. After selection, the genome-wide abundance of advantageous alleles has increased. (*d*) Purifying selection, which eliminates deleterious mutations (*blue*), resulting in decreased genetic diversity. Note that, somewhat paradoxically, variants that are so deleterious that they are rapidly removed from the population have less of an effect on genetic diversity than mildly deleterious alleles that can persist for longer periods (and mark multiple chromosomes for selective purging).

ratios of nonsynonymous substitutions per nonsynonymous site (d_N) to synonymous substitutions per synonymous site (d_S). Under neutral evolution, synonymous and nonsynonymous substitutions should occur at the same rate, and the d_N/d_S ratio should be 1. If the nonsynonymous sites tend to be deleterious and eliminated by purifying selection, then d_N/d_S should be less than 1; if these sites are advantageous and have been driven to fixation, then d_N/d_S should be greater than 1 (48, 76). This general framework of comparing rates of divergence between two classes of sites can also be applied to noncoding regions (160).

Early studies using mouse protein-coding sequences as an outgroup identified hundreds of genes under positive selection in human and chimpanzee lineages (27, 28, 47) that are overrepresented in biological processes associated with immunity, olfaction, host defense, and reproduction (104). These enrichments are not particular to the human lineage, but rather represent adaptive changes occurring throughout the primates and more broadly in other mammalian lineages. Indeed, further studies have suggested that the chimpanzee lineage has an excess of positive selection compared with humans, and in contrast with earlier observations, these studies did not find an enrichment of positively selected genes related to brain development or size (8, 79).

Supplemental Table 1 (follow the Supplemental Material link from the Annual Reviews home page at **http://www.annualreviews.org**) summarizes genes potentially subject to positive selection in humans culled from 16 genome-wide analyses. This is an expanded list of loci relative to those we have summarized previously (4) and includes genes that harbor signals of selection based on patterns of interspecific divergence and/or intraspecific polymorphism (see below). It is important to stress the need for caution when interpreting sequences that appear to have elevated substitution rates in the human lineage. Although many of the results to date are likely enriched for functionally important sequences that have been subject to human-specific molecular adaptations, other evolutionary forces could potentially confound these observations. Perhaps the most insidious is GC-biased gene conversion, which creates signatures in sequence data that closely mimic those caused by natural selection (38).

Supplemental Material

Signatures of Adaptive Regulatory Divergence

The importance of gene regulation as a vehicle for evolutionary change was first suggested in the 1970s at a time when data were scarce (20, 78). These precocious observations have been the focus of considerable interest over the past decade, as technologies such as DNA microarrays have facilitated testing hypotheses of regulatory evolution. For example, significant gene expression differences between humans and chimpanzees have been described, particularly for genes expressed in the brain (22, 40, 107, 150), suggesting that regulatory evolution may have played an important role in human or primate adaptations (23, 78). Moreover, Haygood et al. (61) compared rates of DNA sequence evolution in the human lineage between promoter regions and nearby intronic sequences (an analog of the d_N/d_S test described above that is applicable to noncoding regions) and found an excess of human promoter divergence for genes related to neuronal development and nutrition compared with chimpanzee and rhesus macaque. Although these results are consistent with adaptive regulatory divergence in humans, differences in neutral mutation rate between promoters and introns could lead to spurious results (142), highlighting the challenges in distinguishing between neutral and nonneutral forces. Furthermore, adaptive regulatory divergence between humans and other primates has been observed in other classes of regulatory sequences, including alternative splicing (18), microRNA expression (13, 67), DNA methylation (111, 165), transcription factor binding sites (70), and DNase I–hypersensitive sites (133). We anticipate that ongoing efforts to comprehensively map human regulatory DNA (42) will provide a powerful resource for identifying adaptive variants that influence gene expression.

Insights from Archaic Hominins

One of the most exciting recent developments in studies of human evolutionary history was the sequencing of the genomes of Neanderthals and Denisovans, closely related hominin groups that coexisted with anatomically modern humans (51, 97, 128). These whole-genome sequences allowed new analyses to be pursued into the contentious topic of whether modern humans and archaic groups interbred, and led to the finding that ~1–4% of non-African genomes are derived from Neanderthals and ~4–6% of Melanesian genomes are derived from Denisovans (51, 128). Moreover, studies have also reported significant evidence of introgression in African populations from an unidentified archaic group distinct from Neanderthals and Denisovans (56, 81, 156).

A fundamentally important question is whether archaic sequences present in modern humans are neutral or have facilitated adaptive evolution. Abi-Rached et al. (3) recently found evidence for adaptive introgression of archaic immune system alleles. Specifically, they studied the highly polymorphic human leukocyte antigen (HLA) class I region, which is vital in immunity and subject to strong balancing selection, and suggested that modern humans acquired the *HLA-B*73* allele in western Asia through admixture with Denisovans. Further analyses also suggested that additional archaic HLA haplotypes from Neanderthal and Denisovan genomes have introgressed into modern Eurasian and Oceanian populations. Several alleles in these archaic HLA haplotypes, encoding unique or strong ligands for natural killer cell receptors, now represent more than half of the HLA alleles in modern Eurasians and appear to have been subsequently introduced into Africans (3). Although exciting, it is important to note that HLA loci have deep coalescent times that could potentially be mistaken for introgression and that next-generation sequencing is difficult to carry out for highly polymorphic loci like the HLA region, and we therefore suggest caution in interpreting these observations. More recently, adaptive introgression has also been proposed for the innate immune gene *STAT2* in Melanesian populations (96). The search for adaptive introgression is likely to be a major focus moving forward, and it is reassuring that the examples described to date involve immune-related genes, which are well-known substrates for adaptive evolution.

More generally, archaic genomes provide a potentially powerful resource to more clearly delineate human-specific substitutions, which likely underlie at least a subset of human-specific phenotypes. For example, by identifying regions where Neanderthals carried fewer derived alleles than expected relative to humans, Green et al. (51) identified a total of 212 regions that are likely enriched for substrates of positive selection, which occurred in conjunction with or shortly after the divergence of modern human lineages from Neanderthals ~825,000 years ago. Several disease genes that influence cognition, such as *DYPK1A*, *NRG3*, *CADPS2*, and *AUTS2*, are located in these candidate selected regions (**Supplemental Table 1**), suggesting that multiple genes involved in cognitive development were positively selected during the early history of modern humans. Overall, the sequencing of archaic genomes has provided new avenues of inquiry to identify the genetic substrates of human uniqueness.

Balancing selection: natural selection that maintains different alleles in a population, resulting in an increase of neutral polymorphism in linked genome regions

Selective sweep: the process by which a favorable mutation becomes fixed so rapidly that physically linked alleles also become either fixed or lost, depending on the linkage phase; also referred to as a hard sweep or genetic hitchhiking

Hard sweep: a classic selective sweep originating from a new advantageous mutation; selective sweep is sometimes used to refer to this mode of selection in a narrow sense

Supplemental Material

DETECTING ADAPTIVE EVOLUTION FROM INTRASPECIFIC POLYMORPHISM

Classic Selection Sweeps

Another major approach for detecting signatures of more recent selection is based on analysis of intraspecific polymorphism (**Figure 1**). The vast majority of statistical approaches to detect positive selection from polymorphism data are predicated on the classic selective sweep model (also referred to as a hard sweep), where a new advantageous mutation is rapidly driven to fixation

(136) (**Figure 2*a***). Because of a hitchhiking effect, the frequency of neutral variation closely linked to the advantageous allele also increases. Therefore, a classic selective sweep tends to eliminate nucleotide diversity in the region around the advantageous allele, create an excess of low-frequency variants, and increase linkage disequilibrium and haplotype homozygosity. Many neutrality tests based on the site frequency spectrum, such as Tajima's *D* (140), Fu & Li's *D* (46), and Fay & Wu's *H* (43), have been developed to detect the strong skew toward rare alleles in the swept region.

According to diffusion theory (138), a selective sweep in a randomly mating population of constant size takes approximately $2\ln(2N_e)/s$ generations to complete, where N_e is the effective population size and s is the selection coefficient. Assuming the long-term N_e of humans is 10,000 (130), even sweeps with very strong selection coefficients (e.g., $s = 5\%$) will take ~10,000 years to complete (assuming 25 years per generation), and those with weaker selective effects will take considerably longer. Thus, it is possible that many sweeps are ongoing (also called partial or incomplete sweeps) (**Figure 2*a***) and that the advantageous allele has not yet reached fixation. For partial selective sweeps, an excess of rare variants in the swept region may not be detectable by neutrality tests of the site frequency spectrum. However, partial sweeps typically exhibit long identical haplotypes when the selected allele has risen to moderate or high frequencies, whereas the remaining haplotypes show normal levels of variability. Taking advantage of this feature, methods based on haplotype homozygosity [such as extended haplotype homozygosity (EHH) (131), integrated haplotype score (iHS) (154), and integrated extended haplotype homozygosity of a single-nucleotide polymorphism (SNP) site (iES) (141)] have been developed to detect unusually long, frequent haplotypes expected for incomplete or ongoing sweeps.

Another signature that may arise in classic selective sweeps occurs when a locus is subject to selection in a geographically restricted population (85). In these cases, the allele frequencies around the selected locus are expected to change rapidly, resulting in an excess of SNPs with extreme levels of population structure in this region. Therefore, an unusually high level of population differentiation (most often measured by the fixation index F_{ST}) can be used to identify putative targets of selection (5, 85). It is important to note, however, that large allele frequency differences are not uniquely caused by positive selection, and other demographic processes can influence the variance in F_{ST} (101, 102, 129). Indeed, the confounding effects of natural selection and population history on patterns of polymorphism make the interpretation of all neutrality test statistics difficult, and this remains a formidable challenge even in the context of genome-wide data sets.

A number of celebrated and well-understood examples have been described of loci that exhibit features of classic selective sweeps. For example, the ability to digest lactose into adulthood is a human-specific trait that allows the consumption of dairy products and is correlated with the domestication of cattle (139). Lactase persistence is governed by regulatory variation that allows *LCT*, the gene encoding the enzyme lactase, to be expressed into adulthood. In northern Europeans, a common haplotype containing *cis*-regulatory polymorphisms that influence the expression of *LCT* has been identified (41, 80). This haplotype exhibits strong linkage disequilibrium that extends over 1 Mb (121). The most plausible causal allele is a C-to-T SNP located 13,910 base pairs (bp) upstream of *LCT*. The derived T allele, associated with lactase persistence, is found in approximately 77% of Europeans (and is nearly fixed in particular European populations) but is virtually absent in most non-European populations. The strong selection coefficient estimated for *LCT* ($s = \sim 1.4$–15%) makes it among the most intensely selected loci in the human genome (14). Interestingly, lactase persistence is also observed in pastoral African populations (147). However, the European variants that confer lactase persistence are either absent or at low frequency. Strikingly, other variants at sites very close to the European-associated polymorphisms (such as G/C-14010, T/G-13915, and C/G-13907) are significantly associated with lactase persistence and

Site frequency spectrum: the allele frequency distribution in a population; it can be "folded" or "unfolded" depending on whether ancestral and derived alleles are distinguished

Partial sweep: the process by which an advantageous mutation increases rapidly but has not yet reached fixation (perhaps because the sweep is still in progress or the allele's selective advantage has weakened)

contain signatures of recent selection, suggesting convergent adaptive evolution for this phenotype in Europeans and Africans (147).

Furthermore, a number of recent studies have identified genes subject to natural selection in highland populations (12, 15, 115, 135, 161, 163). High-altitude hypoxia is caused by decreased barometric pressure at high altitude and results in severe physiological stress. The Tibetan population, which has lived at >3,000 m for thousands of years, has been the subject of considerable interest. A study of 50 exomes of ethnic Tibetans identified a polymorphism in *EPAS1*, the gene encoding endothelial Per-Arnt-Sim domain protein 1, that exhibits a 78% difference in allele frequency between Tibetan and Han samples, representing one of the fastest allele frequency changes observed at any human gene to date (163). *EPAS1* encodes a transcription factor involved in the response to hypoxia, and SNPs across this gene are significantly correlated with hemoglobin concentration in Tibetan highlanders (12). The causal adaptive variant(s) in *EPAS1* remain to be discovered, however.

In addition to *EPAS1*, genome-wide scans for positive selection have implicated other genes (such as *EGLN1* and *PPARA*) as targets of natural selection for high-altitude adaptation in Tibetan populations (115, 161). Interestingly, *EGLN1* also shows evidence of positive selection in Andeans, another high-altitude population, although the patterns of variation for this gene differ between Tibetans and Andeans (15). Furthermore, beyond *EGLN1*, the apparent genetic targets of selection are different in the two populations. For example, *PRKAA1* and *NOS2A* exhibit strong signatures of recent positive selection in Andeans but not in Tibetans (15). The largely distinct evolutionary paths in altitude adaptation may be consistent with the physiological differences that exist between Tibetan and Andean populations or may suggest that there are parallel routes to the same physiological states (11).

Classic Selection Sweeps Are Rare in Humans

As alluded to above, it remains challenging to distinguish the footprints of a selective sweep from neutral deviations caused by population demographic history (mediated primarily through genetic drift) and mutation rate heterogeneity. For instance, in a summary of nine genome-wide scans from SNP data, ~14% of the genome containing ~23% of genes was identified as being under positive selection in at least one study (4). However, only ~20% of these regions were concordant among these studies, suggesting a high false-positive rate.

Furthermore, a systematic analysis of the HapMap phase II data showed that genic (especially nonsynonymous) SNPs are more likely than nongenic SNPs to exhibit high levels of population differentiation (30, 94). This observation is consistent with geographically restricted models of classic selective sweeps. However, more detailed analyses of high-F_{ST} SNPs revealed relatively few fixed or nearly fixed differences between populations from different continents. High-F_{ST} SNPs tend to lie in regions of reduced diversity, but the shift of cross-population extended haplotype homozygosity (XP-EHH) is relatively small. These observations imply that positive selection is common in the genome, whereas strong selection that rapidly drives new mutations to fixation is rare (30).

A comprehensive study by Hernandez et al. (64) using resequencing data from 179 human genomes provides the clearest evidence yet that classic selective sweeps in humans are rare. This study showed that reduced levels of genetic diversity exist around exons and conserved noncoding sequences, but the reduction of diversity for flanking human-specific amino acid substitutions is comparable to that for synonymous substitutions. Thus, although the reduction of diversity at coding and conserved noncoding sites is a clear signature of selection, it does not appear to have been driven by classic selective sweeps.

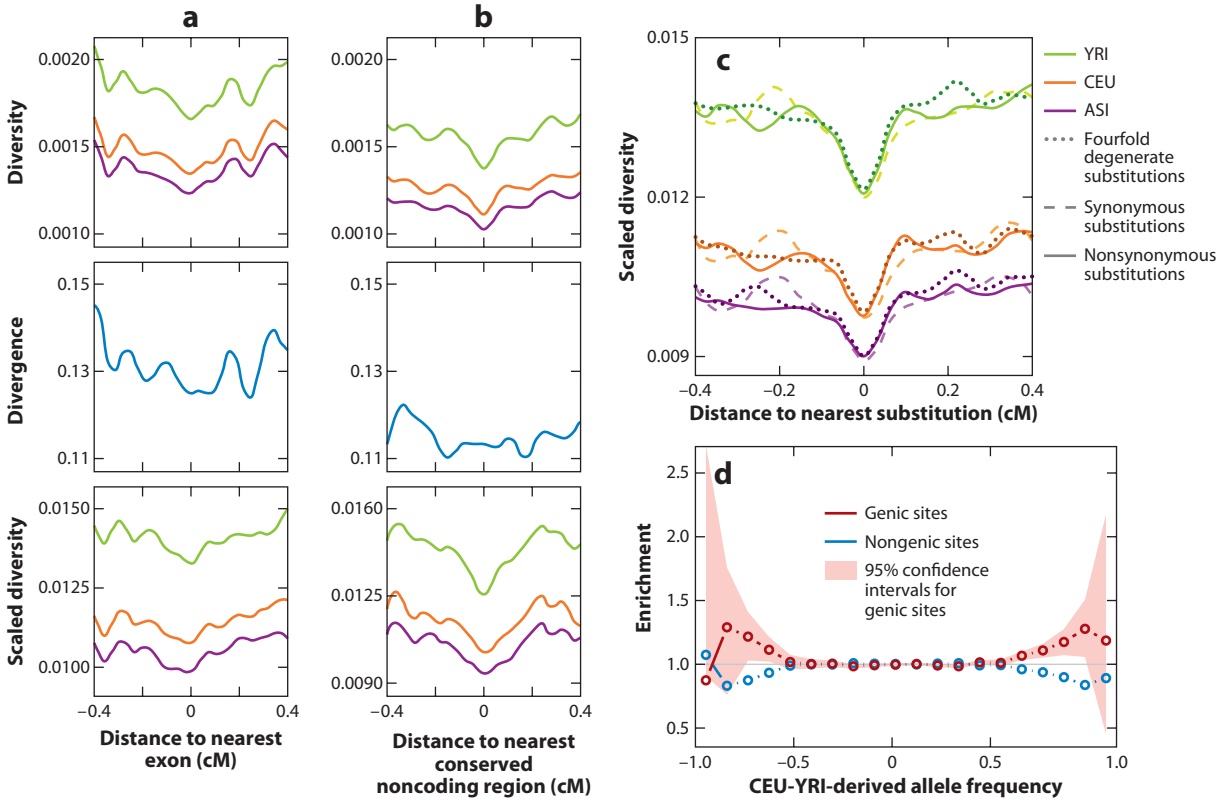

Figure 3

Analyses demonstrating the rarity of classic selective sweeps in human populations. To recapitulate and explore the observations of Hernandez et al. (64) in more detail, we analyzed the 1000 Genomes Project low-coverage data, including 59 individuals from Yoruba (YRI), 60 individuals with European ancestry (CEU), and 60 individuals from China or Japan (ASI). (*a–c*) LOESS curves for a span of 0.1 and a bin size of 1.2×10^{-5} centimorgans (cM). Panels *a* and *b* show diversity, human–rhesus macaque divergence, and scaled diversity (diversity divided by human–rhesus macaque divergence) at nonconserved, noncoding sites as a function of genetic distance from exons and conserved noncoding regions, respectively. Panel *c* shows scaled diversity around human-specific fourfold degenerate (*dotted lines*), synonymous (*dashed lines*), and nonsynonymous (*solid lines*) substitutions. (*d*) Enrichment analysis for highly differentiated single-nucleotide polymorphisms (the CEU-YRI comparison) in genic (*red*) and nongenic (*blue*) sites. The 95% confidence intervals for genic sites are shown in pink, and were obtained by bootstrapping 200-kb regions 1,000 times. Based on figures 1, 3, and 4 of Reference 64 with permission, using reanalyzed data.

To illustrate these important results and explore the observations of Hernandez et al. (64) in more detail, we repeated their analyses on the 1000 Genomes Project data as described in their study (**Figure 3**). We were able to recapitulate all of their observations; specifically, we found that genetic diversity scaled by human–rhesus macaque divergence is reduced for exons and conserved noncoding sequences (**Figure 3*a*,*b***). Interestingly, patterns of divergence vary, at least qualitatively, between exons and conserved noncoding sequences, with the former exhibiting more of a reduction in diversity compared with the latter (**Figure 3*a*,*b***), consistent with stronger selection acting on coding sequences. Moreover, divergence is uniformly lower for conserved noncoding sequences and their flanking regions compared with exons (**Figure 3*a*,*b***), as expected given that these sites were a priori selected based on their conservation. Overall, these patterns are consistent with some form of directional selection, as noted by Hernandez et al. (64). However, the

reduction in scaled diversity around human-specific amino acid substitutions is commensurate with that around synonymous substitutions (**Figure 3*c***), which Hernandez et al. (64) argued is inconsistent with the expectation of classic sweeps. This argument assumes that the signatures of a classic sweep should be observed more strongly on the changes most likely to have functional consequences (e.g., amino acid substitutions). Although this is a reasonable assumption, it is conceivable that a higher proportion of synonymous sites have fitness consequences than is currently appreciated (144). Moreover, SNPs in genic regions are not significantly enriched for highly differentiated alleles (**Figure 3*d***), which again is inconsistent with expectations of classic selective sweeps. Similarly, a recent study of allele frequency differentiation and haplotypes in African populations also did not show widespread evidence for classic selective sweeps (49). Thus, the emerging picture is that classic selective sweeps have been rare in recent human history, prompting the need to consider other modes of adaptation.

Soft sweep: a selective sweep model where standing variation becomes selectively favored and sweeps to fixation, or where multiple alleles at a locus are simultaneously favored and increase in frequency

ALTERNATIVE MODELS TO A CLASSIC SELECTIVE SWEEP: NONCLASSIC SWEEPS

Selection on Standing Variation

Recent empirical and theoretical studies have highlighted the potential importance of soft sweeps, i.e., selection on standing variation (63, 123) (**Figure 2*b***) or selection acting simultaneously on multiple independent loci (**Figure 2*c***). For example, Hermisson & Pennings (63) used diffusion theory to investigate the probability that selection acts on standing variation as opposed to new mutations. The key parameters in their model are the scaled mutation rate θ, measured by $4N_e\mu L$ (where N_e is the effective population size, μ is the mutation rate per generation per base pair, and L is the "mutational target size," or the number of base pairs in which mutations can lead to differences in fitness), and the relative selective advantage in a new environment compared with the ancestral one (63). They supposed that mutations at any of the L base pairs result in higher fitness for the organism in the new environment. The mutational target could be a regulatory region (e.g., a promoter or enhancer) or a set of amino acids that influence protein function. In general, very little is known about mutational target size for a typical locus or how it varies across different genes (90). Prior to the environmental change, variants in this mutational target are either neutral or mildly deleterious, which enables them to persist in the population at low or intermediate frequencies.

We used the theory of Hermisson & Pennings (63) to evaluate several characteristics of selection from standing variation. For example, **Figure 4*a*** shows the probability that a standing variant goes to fixation as a function of the strength of selection (s) and θ. Compared with a new advantageous mutation, standing variation is much more likely to be fixed by selection, provided that the mutational target size is sufficiently large—i.e., $\theta \geq 0.004$, which corresponds to ~10 bp, depending on assumptions of the mutation rate [e.g., 10^{-8}/(generation $\times$ number of base pairs)] and effective population size (e.g., $N_e = 10{,}000$). Furthermore, if a sweep has occurred, we can use the theory of Hermisson & Pennings (63) to calculate the probability that it did so from standing variation and not from a new mutation. When the effective population size is fixed at $N_e = 10{,}000$, a sweep is more probable from standing variation than from a new mutation for nearly all parameter values considered except for very strong selection acting on small mutational target sizes (e.g., $L = 1$ bp) (**Figure 4*b***). However, the waiting time for a new advantageous mutation to occur will be long when the mutational target size is small. Furthermore, the probability of a sweep from standing variation increases with larger effective population sizes (**Figure 4*c***). This observation has important implications for recent human evolutionary history, which has seen an

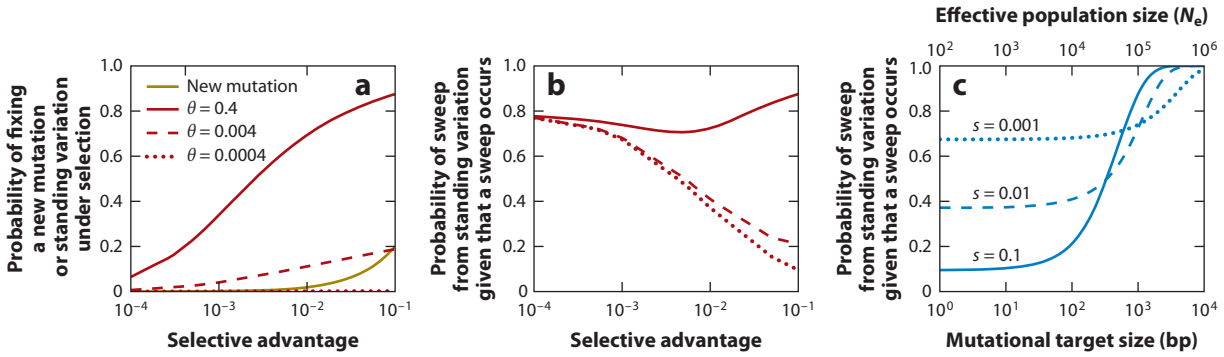

Figure 4

Characteristics of selective sweeps arising from standing variation. (*a*) The probability that a new mutation or standing variation is fixed under selection. (*b*,*c*) The probability that a sweep occurs from standing variation, conditional on a sweep occurring (either from standing variation or from a new advantageous mutation), as a function of the strength of selection (panel *b*) and the mutational target and effective population sizes (panel *c*). Where appropriate, we have assumed the effective population size is $N_e = 10{,}000$, the mutation rate is $\mu = 10^{-8}$ per generation, and the mutational target size is 100 bp. Before the environmental switch, variants are deleterious with a selection pressure $2N_e s = -10$. All selection is assumed to be additive. Adapted from Reference 63 with permission.

explosion in population size (45, 72, 103, 144), suggesting that selection occurring over the past 5,000–10,000 years did so primarily by acting on standing variation.

Signatures of Selection on Standing Variation

If a beneficial allele originates as standing variation, it will leave a different signature during the fixation process compared with that of a new mutation. As standing variation has persisted in the population, either neutrally or as a weakly deleterious variant, it has a different haplotype background than a new mutation does. In particular, recombination will have had more opportunities to move an existing variant onto different haplotype backgrounds. Therefore, a sweep from standing variation drags along more polymorphisms at linked sites than a sweep from a new mutation does, resulting (on average) in a narrower and shallower valley of low genetic diversity compared with a hard sweep (126). The difference in levels of genetic diversity expected from hard and soft sweeps depends primarily on the frequency f of a variant at the time selection begins. Specifically, if f is less than $1/(2N_e s)$ and selection is strong, then reductions in diversity as measured by π (140), θ_w (159), and θ_H (43) will be similar in both the hard and soft sweep models. However, as f increases, the effect of selection has a weaker effect on diversity at linked neutral sites (68).

To illustrate these effects, we simulated hard and soft sweeps using demographic models appropriate for African and European populations (details of the models are described in the caption of **Figure 5**). As expected, hard sweeps result in a greater reduction of genetic diversity compared to soft sweeps (**Figure 5*a–c***). Interestingly, as previously noted (126), for intermediate values of f, selection on standing variation leads to a much larger variance in the site frequency spectrum, as measured by Tajima's D (140), than a hard sweep does (**Figure 5*e***). Furthermore, it also results in weaker levels of linkage disequilibrium, as measured by haplotype homozygosity (**Figure 5*d***). Thus, many conventional neutrality test statistics used to detect a hard sweep will have decreased power to identify sweeps that arose from standing variation (146). As selection from standing variation is expected to be common in recent human history (**Figure 4*c***), these observations suggest that many of the substrates of positive selection in the human genome await discovery because they do not leave signatures in patterns of genetic variation that most neutrality test statistics are designed to detect.

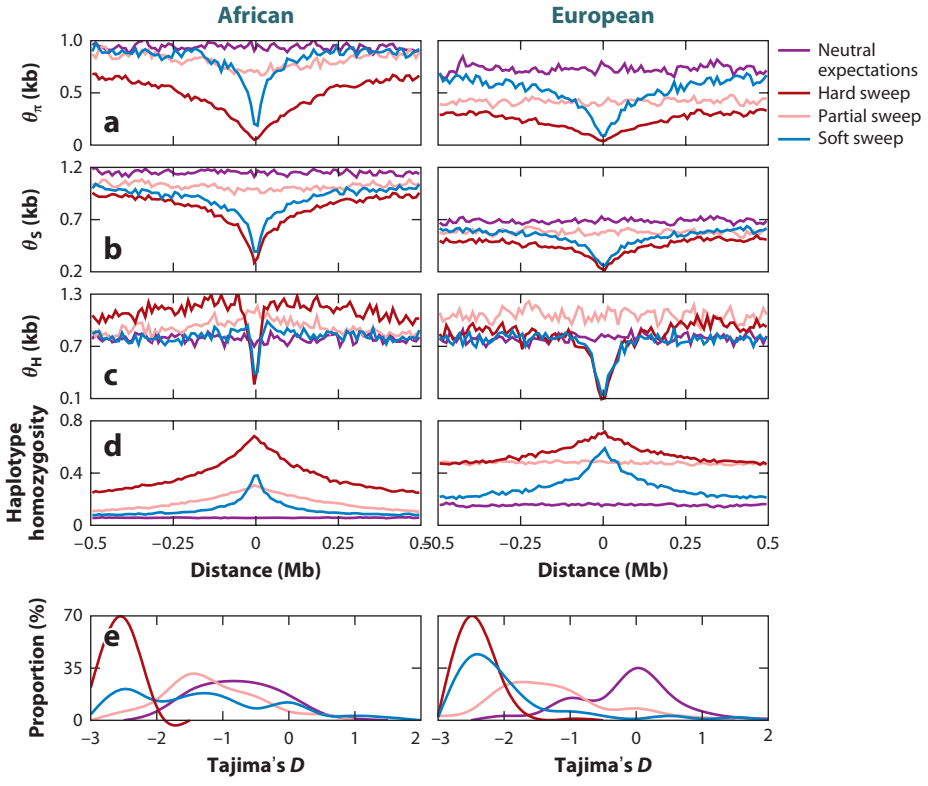

Figure 5

Signatures of different types of selective sweeps compared with neutral expectations. We conducted the analysis by simulating 100 chromosomes using previously inferred demographic models (132) for African and European populations. Models of selection include selection on a new mutation (hard sweep; *red*), on a standing variant with a frequency of 0.05 (soft sweep; *blue*) that began 1,200 generations ago, and on a new mutation but without fixation (partial sweep; *pink*) that began 400 generations ago. The additive selection pressure is assumed to be 0.02. In each scenario, 100 replications were performed. Different statistics were summarized based on a sliding window analysis of 10 kb. (*a–c*) The effects of selection on summary statistics of genetic diversity (θ_π, θ_S, and θ_H in panels *a*, *b*, and *c*, respectively). (*d*) The effects of selection on haplotype homozygosity. (*e*) The distribution of Tajima's *D* (140) in the 10-kb region surrounding the selected site.

Polygenic Selection on (Standing) Variation

The architecture of most traits is complex and governed by the action of many loci and environmental factors. More than 1,000 genome-wide association studies have been published to date (**http://www.genome.gov/gwastudies**) (66). A common observation of these studies is that the vast majority of associations have an odds ratio between 1.1 and 1.5, indicating relatively small phenotypic effects (91). Similarly, most quantitative traits are also architecturally complex and influenced by a large number of quantitative trait loci that individually account for only small amounts of phenotypic variation (10, 65). Thus, a typical trait has a large repository of standing variants for selection to act on, and selection acting simultaneously on many preexisting variants would be an efficient mechanism for phenotypic adaptation (21, 122).

To be concrete, consider a quantitative trait that is affected by a large number of loci (**Figure 6**), with individuals in the population distributed around an optimal phenotypic value (i.e., the

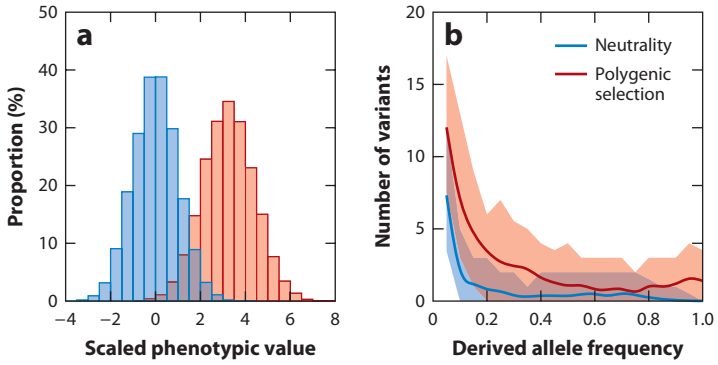

Figure 6

The consequences of polygenic selection from standing variation on a quantitative trait. We simulated 100 chromosomes from a population of constant size equal to 10,000 and assumed that 100 independent variants contributed to the phenotypic variation of the trait. The contribution of each variant to the trait was determined by drawing a value from the normal distribution *N*(mean, standard error), with *N*(0,1) for the ancestral allele and *N*(3,1) for the derived allele. Under neutrality, the fitness for individuals with different phenotypes is the same. Under the polygenic selection model, however, individuals with a larger value of the phenotype (i.e., >45) have 1% higher fitness compared with those with a smaller value of the phenotype. In total, we conducted 100 replications for each model. (*a*) Phenotypic distribution under the neutral and polygenic selection models after 500 generations. (*b*) Comparison of the site frequency spectrums under the polygenic selection and neutral models after 500 generations. Lines represent the average number of variants for a given derived allele frequency; shaded regions define the 95% confidence intervals obtained from the replications.

phenotypic value that confers the highest fitness). If the optimal phenotypic value changes, perhaps because of a change in the environment, then the population will adapt by allele frequency shifts at many loci. Thus, polygenic selection pulls alleles up and down in frequency but generally not to fixation (26, 30, 113) (**Figure 2*c***). A subtle allele frequency shift is expected for polygenic selection on standing variation (57), which can be used to detect such signatures across the genome (**Figure 6**). For example, by studying height, a classic polygenic trait, Turchin et al. (149) demonstrated that the frequencies of alleles associated with increased height are systematically elevated in northern Europeans compared with southern Europeans ($p < 4.3 \times 10^{-4}$), which is consistent with widespread weak selection with selection coefficients on the order of $\sim 10^{-3}$–10^{-5} per allele per generation compared with genetic drift alone.

Moreover, several genome-wide approaches have been developed to detect signatures expected from polygenic selection. For example, Tennessen & Akey (143) found significant evidence for parallel adaptive divergence in geographically diverse populations, consistent with modest shifts in allele frequencies at many loci. Another promising approach for detecting polygenic selection is to compare allele frequencies with environmental variables while controlling for population structure (31, 57–59). Collectively, these studies suggest that polygenic selection may be widespread in humans, although more theoretical and empirical work needs to be done to better understand models involving subtle shifts of allele frequencies at many loci and how best to detect this mechanism of adaptation.

Polygenic selection: a process in which selection occurs at many (standing) variants simultaneously; it is likely common for traits with a complex genetic architecture

DELETERIOUS VARIANTS AND BACKGROUND SELECTION

Relative Importance of Background Selection and Genetic Hitchhiking

Human diversity is positively correlated with recombination on a large scale (84, 99, 100) but negatively correlated with coding sequence density (114), consistent with the expectations of natural

selection. In theory, this correlation could arise either from the genetic hitchhiking effects associated with advantageous alleles or from background selection against deleterious mutations (24, 75). Hellmann et al. (62) investigated the relationship between recombination rate and diversity in human polymorphism data by establishing models based on these two types of selection and found that the hitchhiking model gave a slightly better fit to genome-wide estimates of diversity. However, they emphasized that their models are greatly simplified, and therefore their results were not conclusive.

Background selection: a process in which deleterious mutations drift up to low frequencies and are then purged from the population, with an occasional removal of linked variation resulting in a reduction in diversity

More recently, McVicker et al. (94) analyzed the genomic distribution of human polymorphisms and divergence among five primate species relative to the locations of conserved sequence features. They applied a theoretical model of background selection to explore the reduction of nucleotide diversity due to purifying selection by considering recombination rates, selected site locations, deleterious mutation rates, and the distribution of selection intensities. They found a surprisingly large reduction in average diversity across the genome attributable to selection of 19–26% on the autosomes and 12–40% on the X chromosome, indicating that background selection may have played a prominent role in shaping patterns of human genomic variation. However, because the signals of background selection can mimic patterns produced by positive selection, they noted that they were not able to formally exclude the possibility that their findings were the result of hitchhiking.

Thus, the relative contributions of genetic hitchhiking and background selection in shaping genomic patterns of variation remain open (137). Our intuition is that background selection provides a better first-order approximation for existing data than genetic hitchhiking does, simply because more mutations are expected to be deleterious than advantageous and because of how widespread the signal is across the genome. Definitive inferences are not possible at this point, but perhaps will be in the future with the development of new methods and data.

Purifying Selection Facilitates the Identification of Functionally Important Sequences

Despite the remarkable progress in molecular biology and biochemistry, interpreting the human genome remains a challenging endeavor. For example, over a decade has passed since the initial drafts of the human genome were published (83, 152), yet there is still considerable uncertainty about the fraction of the genome that encodes for functionally important information. The Encyclopedia of DNA Elements (ENCODE) project, the goal of which was to systematically identify functional elements encoded in the human genome, recently described their results (42). In this project, a functional element was defined as a discrete genomic segment that encodes a defined product (e.g., protein or noncoding RNA) or displays a reproducible biochemical signature (e.g., protein binding or a specific chromatin structure). This systematic analysis revealed that a large fraction of the human genome (80.4%) is "biochemically active" and covered by at least one ENCODE-identified element (42). However, it seems rather unlikely that ~80% of the genome is functionally important (37, 39, 50, 106).

An evolutionary perspective provides an important framework for identifying functionally important sequences through conservation-based metrics. In an evolutionary context, conservation arises through time in functional sequences because most mutations that occur are deleterious and purged from the population by purifying selection (**Figure 2*d***). An initial comparison of the mouse and human genomes estimated that up to ~5% of the mammalian genome is under purifying selection (158); subsequent estimates ranged from 2.6% to 20% (6, 16, 32, 36, 89, 95, 112, 119, 120, 134). For instance, a recent study comparing 29 eutherian genomes estimated that at least 5.5% of the human genome experiences purifying selection and located constrained

elements covering ~4.2% of the genome at a resolution of 12 bp (87). Furthermore, using the ENCODE data, Ward & Kellis (157) estimated that 5% of the human genome is conserved across mammals, with at least an additional 4% experiencing purifying selection in the human lineage alone. Thus, although evolutionary analyses lead to different estimates about the proportion of the human genome that is subject to functional constraint, they are uniformly inconsistent with the notion that ~80% of our DNA is functionally significant. Although there are inherent limitations to simple conservation-based metrics (e.g., identifying sequences that are necessary to maintain the proper spatial organization of functional elements but whose specific nucleotide composition is not important), it is a good approximation that is likely to identify the bulk of functionally, evolutionarily, and phenotypically significant DNA.

Mutation load: the loss of fitness resulting from deleterious alleles maintained by mutation-selection balance

The Burden of Deleterious Variants Among Individuals

The human population size has increased dramatically in recent history (45, 72, 103, 144). Given an estimated world population of more than 7 billion individuals (151), humanity now incurs on the order of 10^{11} de novo mutations per generation (assuming a mutation rate of 2.2×10^{-8} per generation). How many of these variants are deleterious, and how many such variants do individuals carry? Studies of exome and whole-genome sequencing performed using large cohorts of individuals ($n > 1{,}000$) have provided new insights into the mutation loads across individuals (45, 103, 144). For example, conservative estimates suggest that approximately 14–17% of all protein-coding variants are deleterious, with each individual possessing an average of 500 functional protein-coding single-nucleotide variants (144). Similarly, by comparing the ratios of nonsynonymous to synonymous variants in different frequency ranges, another study estimated that 25–50% of rare nonsynonymous variants are deleterious (2).

Beyond the exome, studies are also beginning to integrate whole-genome sequence and functional genomics data. For example, using the ENCODE data (42), Vernot et al. (153) investigated the population genetics characteristics of polymorphisms located in DNase I–hypersensitive sites and concluded that in regulatory DNA, individuals harbor many more deleterious variants than protein-coding variants, although the deleterious variants are likely to have smaller effect sizes on average. In addition to being important in personal and clinical genomics, delineating the mutational burden among individuals may facilitate more quantitative estimates of the deleterious mutation rate in humans (127) and provide insight into the implications of the increased mutation load for contemporary human populations (71, 90). Finally, it is interesting to speculate that a small fraction of the 10^{11} de novo mutations that occur in each generation are advantageous and therefore that the increased mutational capacity of recent human populations may have created a repository of advantageous alleles that adaptive evolution will act on in subsequent generations (60).

FUTURE DIRECTIONS AND CHALLENGES

Progress in understanding human evolutionary history has often been marked by technological advances. The maturation of next-generation sequencing technology is the latest advance, and is transforming the ability to access sequence variation on a scale that has never before been possible. Indeed, many human and nonhuman primate sequencing data sets have already been described (1, 2, 7, 45, 51, 97, 128, 144), and many more will certainly be generated in the near future. However, simply generating extensive catalogs of human variation will not in itself lead to a deeper understanding of human evolutionary history, and many challenges remain. An immediate pragmatic issue is simply how to store, visualize, and share such massively large data sets in a computationally efficient manner (145).

We are confident that the computational challenges, although formidable, will be addressed, and that genomics can leverage the expertise and knowledge of other data-intensive branches of science (86, 93). More important, however, is the acute need to develop the next generation of population genetics models and statistical tools tailored to the analysis of whole-genome data sets (155). Although the classic selective sweep model and outlier-based genome-wide scans for selection have served the field well, it is now clear that this mode of selection has not played a prominent role in recent human evolutionary history and that most of the dramatic signatures of a hard sweep have been identified. The analysis of additional, typically understudied populations that have experienced unique environments obviously may still reveal novel loci that have been subject to classic selective sweeps [such as *EPAS1* in Tibetans (163)], but such loci are likely to be the exception rather than the rule. Thus, theoretical studies that explore the dynamics of more realistic mechanisms of evolutionary change, such as polygenic and epistatic selection, will be important for developing novel statistical methods to detect signatures from different modes of adaptive evolution. Indeed, as Haldane (53) noted, "A satisfactory theory of natural selection must be quantitative. In order to establish the view that natural selection is capable of accounting for the known facts of evolution we must show not only that it can cause a species to change, but that it can cause it to change at a rate which will account for present and past transmutations" (p. 19). Furthermore, a renewed focus on background selection is warranted not only because it appears to have pervasive effects on human genomic diversity but also because it is a well-defined intellectual construct to assess a variety of key questions, ranging from the fraction of the genome that is functionally significant to the burden of deleterious variants among individuals.

Moreover, it is clear that a deeper understanding of how selection has influenced the human genome is impeded by the mostly statistical descriptions of putatively selected loci to date. Except for a handful of cases (14, 41, 44, 54, 69, 82, 121, 147, 148), the biological functions of adaptively evolving loci are ambiguous, and the causal adaptive alleles underlying signatures of selection remain elusive. Thus, a better understanding of the biology of the human genome would have profound implications for gaining a deeper understanding of its evolution (52). To this end, functional genomics data sets developed by the ENCODE and Roadmap Epigenomics projects using techniques such as RNA-Seq, ChIP-Seq, and DNase I–hypersensitivity mapping (42) have provided an expansive resource for the study of human biology that will facilitate the interpretation and generation of hypotheses in studies of human selection and adaptation.

Another area that we envision will be fruitful to pursue is the evolutionary history of structural variation (indels, duplications, deletions, inversions, and copy-number variations), which affects more base pairs in the genome than single-base changes do (19, 25, 29, 98, 117, 118, 164). Genome duplication is a dominant force in evolution, providing the raw material for genomic innovation that is then sculpted by mutation, drift, and selection (108). To date, however, the vast majority of human studies of adaptation have focused on sequence and not structural variation (with a few notable exceptions; see 74, 116, 162). With the improved resolution of structural variant identification afforded by next-generation sequencing technology, now is the time to begin integrating sequence and structural variation into a comprehensive theory of how natural selection has influenced the human genome.

Considerable progress has been made in understanding how natural selection has influenced human divergence and diversity, yet much remains to be discovered. Beyond sequencing and statistical methods, experimental analysis, and functional genomics profiling, a comprehensive view of natural selection and adaptation must synthesize knowledge from disparate disciplines, including ecology, anthropology, and linguistics, to name only a few. Only then will a coherent narrative emerge about where and (more interestingly) why natural selection has influenced the human genome.

DISCLOSURE STATEMENT

J.M.A. consults for Glenview Capital.

ACKNOWLEDGMENTS

This work was funded by National Institutes of Health grants R01GM098360 and HL106034 to J.M.A.

LITERATURE CITED

1. 1000 Genomes Proj. Consort. 2010. A map of human genome variation from population-scale sequencing. *Nature* 467:1061–73
2. Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, et al. 2012. An integrated map of genetic variation from 1,092 human genomes. *Nature* 491:56–65
3. Abi-Rached L, Jobin MJ, Kulkarni S, McWhinnie A, Dalva K, et al. 2011. The shaping of modern human immune systems by multiregional admixture with archaic humans. *Science* 334:89–94
4. Akey JM. 2009. Constructing genomic maps of positive selection in humans: Where do we go from here? *Genome Res.* 19:711–22
5. Akey JM, Zhang G, Zhang K, Jin L, Shriver MD. 2002. Interrogating a high-density SNP map for signatures of natural selection. *Genome Res.* 12:1805–14
6. Asthana S, Noble WS, Kryukov G, Grant CE, Sunyaev S, Stamatoyannopoulos JA. 2007. Widely distributed noncoding purifying selection in the human genome. *Proc. Natl. Acad. Sci. USA* 104:12410–15
7. Auton A, Fledel-Alon A, Pfeifer S, Venn O, Segurel L, et al. 2012. A fine-scale chimpanzee genetic map from population sequencing. *Science* 336:193–98
8. Bakewell MA, Shi P, Zhang J. 2007. More genes underwent positive selection in chimpanzee evolution than in human evolution. *Proc. Natl. Acad. Sci. USA* 104:7489–94
9. Bamshad M, Wooding SP. 2003. Signatures of natural selection in the human genome. *Nat. Rev. Genet.* 4:99–111
10. Barton NH, Keightley PD. 2002. Understanding quantitative genetic variation. *Nat. Rev. Genet.* 3:11–21
11. Beall CM. 2007. Two routes to functional adaptation: Tibetan and Andean high-altitude natives. *Proc. Natl. Acad. Sci. USA* 104(Suppl. 1):8655–60
12. Beall CM, Cavalleri GL, Deng L, Elston RC, Gao Y, et al. 2010. Natural selection on *EPAS1* (*HIF2α*) associated with low hemoglobin concentration in Tibetan highlanders. *Proc. Natl. Acad. Sci. USA* 107:11459–64
13. Berezikov E, Thuemmler F, van Laake LW, Kondova I, Bontrop R, et al. 2006. Diversity of microRNAs in human and chimpanzee brain. *Nat. Genet.* 38:1375–77
14. Bersaglieri T, Sabeti PC, Patterson N, Vanderploeg T, Schaffner SF, et al. 2004. Genetic signatures of strong recent positive selection at the lactase gene. *Am. J. Hum. Genet.* 74:1111–20
15. Bigham AW, Mao X, Mei R, Brutsaert T, Wilson MJ, et al. 2009. Identifying positive selection candidate loci for high-altitude adaptation in Andean populations. *Hum. Genomics* 4:79–90
16. Birney E, Stamatoyannopoulos JA, Dutta A, Guigo R, Gingeras TR, et al. 2007. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* 447:799–816
17. Biswas S, Akey JM. 2006. Genomic insights into positive selection. *Trends Genet.* 22:437–46
18. Blekhman R, Marioni JC, Zumbo P, Stephens M, Gilad Y. 2009. Sex-specific and lineage-specific alternative splicing in primates. *Genome Res.* 20:180–89
19. Britten RJ. 2002. Divergence between samples of chimpanzee and human DNA sequences is 5%, counting indels. *Proc. Natl. Acad. Sci. USA* 99:13633–35
20. Britten RJ, Davidson EH. 1971. Repetitive and non-repetitive DNA sequences and a speculation on the origins of evolutionary novelty. *Q. Rev. Biol.* 46:111–38
21. Burger R, Gimelfarb A. 1999. Genetic variation maintained in multilocus models of additive quantitative traits under stabilizing selection. *Genetics* 152:807–20

22. Cáceres M, Lachuer J, Zapala MA, Redmond JC, Kudo L, et al. 2003. Elevated gene expression levels distinguish human from non-human primate brains. *Proc. Natl. Acad. Sci. USA* 100:13030–35
23. Carroll SB. 2005. Evolution at two levels: on genes and form. *PLoS Biol.* 3:e245
24. Charlesworth B, Morgan MT, Charlesworth D. 1993. The effect of deleterious mutations on neutral molecular variation. *Genetics* 134:1289–303
25. Cheng Z, Ventura M, She X, Khaitovich P, Graves T, et al. 2005. A genome-wide comparison of recent chimpanzee and human segmental duplications. *Nature* 437:88–93
26. Chevin LM, Hospital F. 2008. Selective sweep at a quantitative trait locus in the presence of background genetic variation. *Genetics* 180:1645–60
27. Chimpanzee Seq. Anal. Consort. 2005. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* 437:69–87
28. Clark AG, Glanowski S, Nielsen R, Thomas PD, Kejariwal A, et al. 2003. Inferring nonneutral evolution from human-chimp-mouse orthologous gene trios. *Science* 302:1960–63
29. Conrad DF, Pinto D, Redon R, Feuk L, Gokcumen O, et al. 2009. Origins and functional impact of copy number variation in the human genome. *Nature* 464:704–12
30. Coop G, Pickrell JK, Novembre J, Kudaravalli S, Li J, et al. 2009. The role of geography in human adaptation. *PLoS Genet.* 5:e1000500
31. Coop G, Witonsky D, Di Rienzo A, Pritchard JK. 2010. Using environmental correlations to identify loci underlying local adaptation. *Genetics* 185:1411–23
32. Cooper GM, Stone EA, Asimenos G, Green ED, Batzoglou S, Sidow A. 2005. Distribution and intensity of constraint in mammalian genomic sequence. *Genome Res.* 15:901–13
33. Currat M, Trabuchet G, Rees D, Perrin P, Harding RM, et al. 2002. Molecular analysis of the β-globin gene cluster in the Niokholo Mandenka population reveals a recent origin of the β^S Senegal mutation. *Am. J. Hum. Genet.* 70:207–23
34. Darwin CR. 1859. *On the Origin of Species by Means of Nature Selection; or, The Preservation of Favoured Races in the Struggle for Life.* London: Murray
35. Darwin CR, Wallace AR. 1858. On the tendency of species to form varieties; and on the perpetuation of varieties and species by natural means of selection. *Proc. Linn. Soc. Lond. Zool.* 3:46–50
36. Davydov EV, Goode DL, Sirota M, Cooper GM, Sidow A, Batzoglou S. 2010. Identifying a high fraction of the human genome to be under selective constraint using GERP++. *PLoS Comput. Biol.* 6:e1001025
37. Doolittle WF. 2013. Is junk DNA bunk? A critique of ENCODE. *Proc. Natl. Acad. Sci. USA* 110:5294–300
38. Duret L, Galtier N. 2009. Biased gene conversion and the evolution of mammalian genomic landscapes. *Annu. Rev. Genomics Hum. Genet.* 10:285–311
39. Eddy SR. 2012. The C-value paradox, junk DNA and ENCODE. *Curr. Biol.* 22:R898–99
40. Enard W, Khaitovich P, Klose J, Zollner S, Heissig F, et al. 2002. Intra- and interspecific variation in primate gene expression patterns. *Science* 296:340–43
41. Enattah NS, Sahi T, Savilahti E, Terwilliger JD, Peltonen L, Jarvela I. 2002. Identification of a variant associated with adult-type hypolactasia. *Nat. Genet.* 30:233–37
42. ENCODE Proj. Consort. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature* 489:57–74
43. Fay JC, Wu CI. 2000. Hitchhiking under positive Darwinian selection. *Genetics* 155:1405–13
44. Friedman MJ. 1978. Erythrocytic mechanism of sickle cell resistance to malaria. *Proc. Natl. Acad. Sci. USA* 75:1994–97
45. Fu W, O'Connor TD, Jun G, Kang HM, Abecasis G, et al. 2013. Analysis of 6,515 exomes reveals the very recent origin of most human protein-coding variants. *Nature* 493:216–20
46. Fu YX, Li WH. 1993. Statistical tests of neutrality of mutations. *Genetics* 133:693–709
47. Gibbs RA, Rogers J, Katze MG, Bumgarner R, Weinstock GM, et al. 2007. Evolutionary and biomedical insights from the rhesus macaque genome. *Science* 316:222–34
48. Goldman N, Yang Z. 1994. A codon-based model of substitution for protein-coding DNA sequences. *Mol. Biol. Evol.* 11:725–36
49. Granka JM, Henn BM, Gignoux CR, Kidd JM, Bustamante CD, Feldman MW. 2012. Limited evidence for classic selective sweeps in African populations. *Genetics* 192:1049–64

50. Graur D, Zheng Y, Price N, Azevedo RB, Zufall RA, Elhaik E. 2013. On the immortality of television sets: "function" in the human genome according to the evolution-free gospel of ENCODE. *Genome Biol. Evol.* 5:578–90
51. Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, et al. 2010. A draft sequence of the Neandertal genome. *Science* 328:710–22
52. Grossman SR, Andersen KG, Shlyakhter I, Tabrizi S, Winnicki S, et al. 2013. Identifying recent adaptations in large-scale genomic data. *Cell* 152:703–13
53. Haldane JBS. 1924. A mathematical theory of natural and artificial selection. Part 1. *Trans. Camb. Philos. Soc.* 23:19–41
54. Hamblin MT, Di Rienzo A. 2000. Detection of the signature of natural selection in humans: evidence from the Duffy blood group locus. *Am. J. Hum. Genet.* 66:1669–79
55. Hamblin MT, Thompson EE, Di Rienzo A. 2002. Complex signatures of natural selection at the Duffy blood group locus. *Am. J. Hum. Genet.* 70:369–83
56. Hammer MF, Woerner AE, Mendez FL, Watkins JC, Wall JD. 2011. Genetic evidence for archaic admixture in Africa. *Proc. Natl. Acad. Sci. USA* 108:15123–28
57. Hancock AM, Alkorta-Aranburu G, Witonsky DB, Di Rienzo A. 2010. Adaptations to new environments in humans: the role of subtle allele frequency shifts. *Philos. Trans. R. Soc. B* 365:2459–68
58. Hancock AM, Witonsky DB, Alkorta-Aranburu G, Beall CM, Gebremedhin A, et al. 2011. Adaptations to climate-mediated selective pressures in humans. *PLoS Genet.* 7:e1001375
59. Hancock AM, Witonsky DB, Ehler E, Alkorta-Aranburu G, Beall C, et al. 2010. Colloquium paper: human adaptations to diet, subsistence, and ecoregion are due to subtle shifts in allele frequency. *Proc. Natl. Acad. Sci. USA* 107(Suppl. 2):8924–30
60. Hawks J, Wang ET, Cochran GM, Harpending HC, Moyzis RK. 2007. Recent acceleration of human adaptive evolution. *Proc. Natl. Acad. Sci. USA* 104:20753–58
61. 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.* 39:1140–44
62. Hellmann I, Mang Y, Gu Z, Li P, de la Vega FM, et al. 2008. Population genetic analysis of shotgun assemblies of genomic sequences from multiple individuals. *Genome Res.* 18:1020–29
63. Hermisson J, Pennings PS. 2005. Soft sweeps: molecular population genetics of adaptation from standing genetic variation. *Genetics* 169:2335–52
64. Hernandez RD, Kelley JL, Elyashiv E, Melton SC, Auton A, et al. 2011. Classic selective sweeps were rare in recent human evolution. *Science* 331:920–24
65. Hill WG. 2009. Understanding and using quantitative genetic variation. *Philos. Trans. R. Soc. B* 365:73–85
66. Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, et al. 2009. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl. Acad. Sci. USA* 106:9362–67
67. Hu HY, Guo S, Xi J, Yan Z, Fu N, et al. 2011. MicroRNA expression and regulation in human, chimpanzee, and macaque brains. *PLoS Genet.* 7:e1002327
68. Innan H, Kim Y. 2004. Pattern of polymorphism after strong artificial selection in a domestication event. *Proc. Natl. Acad. Sci. USA* 101:10667–72
69. Kamberov YG, Wang S, Tan J, Gerbault P, Wark A, et al. 2013. Modeling recent human evolution in mice by expression of a selected EDAR variant. *Cell* 152:691–702
70. Kasowski M, Grubert F, Heffelfinger C, Hariharan M, Asabere A, et al. 2010. Variation in transcription factor binding among humans. *Science* 328:232–35
71. Keightley PD, Lynch M. 2003. Toward a realistic model of mutations affecting fitness. *Evolution* 57:683–85
72. Keinan A, Clark AG. 2012. Recent explosive human population growth has resulted in an excess of rare genetic variants. *Science* 336:740–43
73. Kelley JL, Swanson WJ. 2008. Positive selection in the human genome: from genome scans to biological significance. *Annu. Rev. Genomics Hum. Genet.* 9:143–60
74. Kidd JM, Newman TL, Tuzun E, Kaul R, Eichler EE. 2007. Population stratification of a common *APOBEC* gene deletion polymorphism. *PLoS Genet.* 3:e63

75. Kim Y, Stephan W. 2002. Detecting a local signature of genetic hitchhiking along a recombining chromosome. *Genetics* 160:765–77
76. Kimura M. 1968. Evolutionary rate at the molecular level. *Nature* 217:624–26
77. King JL, Jukes TH. 1969. Non-Darwinian evolution. *Science* 164:788–98
78. King MC, Wilson AC. 1975. Evolution at two levels in humans and chimpanzees. *Science* 188:107–16
79. Kosiol C, Vinar T, da Fonseca RR, Hubisz MJ, Bustamante CD, et al. 2008. Patterns of positive selection in six mammalian genomes. *PLoS Genet.* 4:e1000144
80. Kuokkanen M, Enattah NS, Oksanen A, Savilahti E, Orpana A, Jarvela I. 2003. Transcriptional regulation of the lactase-phlorizin hydrolase gene by polymorphisms associated with adult-type hypolactasia. *Gut* 52:647–52
81. Lachance J, Vernot B, Elbers CC, Ferwerda B, Froment A, et al. 2012. Evolutionary history and adaptation from high-coverage whole-genome sequences of diverse African hunter-gatherers. *Cell* 150:457–69
82. Lamason RL, Mohideen MA, Mest JR, Wong AC, Norton HL, et al. 2005. SLC24A5, a putative cation exchanger, affects pigmentation in zebrafish and humans. *Science* 310:1782–86
83. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, et al. 2001. Initial sequencing and analysis of the human genome. *Nature* 409:860–921
84. Lercher MJ, Hurst LD. 2002. Human SNP variability and mutation rate are higher in regions of high recombination. *Trends Genet.* 18:337–40
85. Lewontin RC, Krakauer J. 1973. Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics* 74:175–95
86. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, et al. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078–79
87. Lindblad-Toh K, Garber M, Zuk O, Lin MF, Parker BJ, et al. 2011. A high-resolution map of human evolutionary constraint using 29 mammals. *Nature* 478:476–82
88. Locke DP, Hillier LW, Warren WC, Worley KC, Nazareth LV, et al. 2011. Comparative and demographic analysis of orang-utan genomes. *Nature* 469:529–33
89. Lunter G, Ponting CP, Hein J. 2006. Genome-wide identification of human functional DNA using a neutral indel model. *PLoS Comput. Biol.* 2:e5
90. Lynch M. 2009. Rate, molecular spectrum, and consequences of human mutation. *Proc. Natl. Acad. Sci. USA* 107:961–68
91. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, et al. 2009. Finding the missing heritability of complex diseases. *Nature* 461:747–53
92. McEvoy B, Beleza S, Shriver MD. 2006. The genetic architecture of normal variation in human pigmentation: an evolutionary perspective and model. *Hum. Mol. Genet.* 15(Suppl. 2):R176–81
93. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, et al. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20:1297–303
94. McVicker G, Gordon D, Davis C, Green P. 2009. Widespread genomic signatures of natural selection in hominid evolution. *PLoS Genet.* 5:e1000471
95. Meader S, Ponting CP, Lunter G. 2010. Massive turnover of functional sequence in human and other mammalian genomes. *Genome Res.* 20:1335–43
96. Mendez FL, Watkins JC, Hammer MF. 2012. A haplotype at *STAT2* introgressed from Neanderthals and serves as a candidate of positive selection in Papua New Guinea. *Am. J. Hum. Genet.* 91:265–74
97. Meyer M, Kircher M, Gansauge MT, Li H, Racimo F, et al. 2012. A high-coverage genome sequence from an archaic Denisovan individual. *Science* 338:222–26
98. Mills RE, Walter K, Stewart C, Handsaker RE, Chen K, et al. 2011. Mapping copy number variation by population-scale genome sequencing. *Nature* 470:59–65
99. Nachman MW. 2001. Single nucleotide polymorphisms and recombination rate in humans. *Trends Genet.* 17:481–85
100. Nachman MW, Bauer VL, Crowell SL, Aquadro CF. 1998. DNA variability and recombination rates at X-linked loci in humans. *Genetics* 150:1133–41
101. Nei M, Chakravarti A. 1977. Drift variances of F_{ST} and G_{ST} statistics obtained from a finite number of isolated populations. *Theor. Popul. Biol.* 11:307–25

102. Nei M, Maruyama T. 1975. Lewontin-Krakauer test for neutral genes. *Genetics* 80:395
103. Nelson MR, Wegmann D, Ehm MG, Kessner D, St. Jean P, et al. 2012. An abundance of rare functional variants in 202 drug target genes sequenced in 14,002 people. *Science* 337:100–4
104. Nielsen R, Bustamante C, Clark AG, Glanowski S, Sackton TB, et al. 2005. A scan for positively selected genes in the genomes of humans and chimpanzees. *PLoS Biol.* 3:e170
105. Nielsen R, Hellmann I, Hubisz M, Bustamante C, Clark AG. 2007. Recent and ongoing selection in the human genome. *Nat. Rev. Genet.* 8:857–68
106. Niu DK, Jiang L. 2012. Can ENCODE tell us how much junk DNA we carry in our genome? *Biochem. Biophys. Res. Commun.* 430:1340–43
107. Nowick K, Gernat T, Almaas E, Stubbs L. 2009. Differences in human and chimpanzee gene expression patterns define an evolving network of transcription factors in brain. *Proc. Natl. Acad. Sci. USA* 106:22358–63
108. Ohno S. 1970. *Evolution by Gene Duplication*. New York: Springer-Verlag
109. Ohta T. 1973. Slightly deleterious mutant substitutions in evolution. *Nature* 246:96–98
110. Olson MV. 2012. Human genetic individuality. *Annu. Rev. Genomics Hum. Genet.* 13:1–27
111. Pai AA, Bell JT, Marioni JC, Pritchard JK, Gilad Y. 2011. A genome-wide study of DNA methylation patterns and gene expression levels in multiple human and chimpanzee tissues. *PLoS Genet.* 7:e1001316
112. Parker SC, Hansen L, Abaan HO, Tullius TD, Margulies EH. 2009. Local DNA topography correlates with functional noncoding regions of the human genome. *Science* 324:389–92
113. Pavlidis P, Metzler D, Stephan W. 2012. Selective sweeps in multilocus models of quantitative traits. *Genetics* 192:225–39
114. Payseur BA, Nachman MW. 2002. Gene density and human nucleotide polymorphism. *Mol. Biol. Evol.* 19:336–40
115. Peng Y, Yang Z, Zhang H, Cui C, Qi X, et al. 2010. Genetic variations in Tibetan populations and high-altitude adaptation at the Himalayas. *Mol. Biol. Evol.* 28:1075–81
116. Perry GH, Dominy NJ, Claw KG, Lee AS, Fiegler H, et al. 2007. Diet and the evolution of human amylase gene copy number variation. *Nat. Genet.* 39:1256–60
117. Perry GH, Tchinda J, McGrath SD, Zhang J, Picker SR, et al. 2006. Hotspots for copy number variation in chimpanzees and humans. *Proc. Natl. Acad. Sci. USA* 103:8006–11
118. Perry GH, Yang F, Marques-Bonet T, Murphy C, Fitzgerald T, et al. 2008. Copy number variation and evolution in humans and chimpanzees. *Genome Res.* 18:1698–710
119. Pheasant M, Mattick JS. 2007. Raising the estimate of functional human sequences. *Genome Res.* 17:1245–53
120. Ponting CP, Hardison RC. 2011. What fraction of the human genome is functional? *Genome Res.* 21:1769–76
121. Poulter M, Hollox E, Harvey CB, Mulcare C, Peuhkuri K, et al. 2003. The causal element for the lactase persistence/non-persistence polymorphism is located in a 1 Mb region of linkage disequilibrium in Europeans. *Ann. Hum. Genet.* 67:298–311
122. Pritchard JK, Di Rienzo A. 2010. Adaptation—not by sweeps alone. *Nat. Rev. Genet.* 11:665–67
123. Pritchard JK, Pickrell JK, Coop G. 2010. The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. *Curr. Biol.* 20:R208–15
124. Provine WB. 1971. *The Origins of Theoretical Population Genetics*. Chicago: Univ. Chicago Press
125. Przeworski M. 2011. The golden age of human population genetics. *Science* 331:547
126. Przeworski M, Coop G, Wall JD. 2005. The signature of positive selection on standing genetic variation. *Evolution* 59:2312–23
127. Reed FA, Akey JM, Aquadro CF. 2005. Fitting background-selection predictions to levels of nucleotide variation and divergence along the human autosomes. *Genome Res.* 15:1211–21
128. Reich D, Green RE, Kircher M, Krause J, Patterson N, et al. 2010. Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* 468:1053–60
129. Robertson A. 1975. Remarks on the Lewontin-Krakauer test. *Genetics* 80:396
130. Sabeti PC, Reich DE, Higgins JM, Levine HZ, Richter DJ, et al. 2002. Detecting recent positive selection in the human genome from haplotype structure. *Nature* 419:832–37

131. Sabeti PC, Schaffner SF, Fry B, Lohmueller J, Varilly P, et al. 2006. Positive natural selection in the human lineage. *Science* 312:1614–20
132. Schaffner SF, Foo C, Gabriel S, Reich D, Daly MJ, Altshuler D. 2005. Calibrating a coalescent simulation of human genome sequence variation. *Genome Res.* 15:1576–83
133. Shibata Y, Sheffield NC, Fedrigo O, Babbitt CC, Wortham M, et al. 2012. Extensive evolutionary changes in regulatory element activity during human origins are associated with altered gene expression and positive selection. *PLoS Genet.* 8:e1002789
134. Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, et al. 2005. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* 15:1034–50
135. Simonson TS, Yang Y, Huff CD, Yun H, Qin G, et al. 2010. Genetic evidence for high-altitude adaptation in Tibet. *Science* 329:72–75
136. Smith JM, Haigh J. 1974. The hitch-hiking effect of a favourable gene. *Genet. Res.* 23:23–35
137. Stephan W. 2010. Genetic hitchhiking versus background selection: the controversy and its implications. *Philos. Trans. R. Soc. B* 365:1245–53
138. Stephan W, Wiehe THE, Lenz MW. 1992. The effect of strongly selected substitutions on neutral polymorphism: analytical results based on diffusion theory. *Theor. Popul. Biol.* 41:237–54
139. Swallow DM. 2003. Genetics of lactase persistence and lactose intolerance. *Annu. Rev. Genet.* 37:197–219
140. Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123:585–95
141. 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
142. Taylor MS, Massingham T, Hayashizaki Y, Carninci P, Goldman N, Semple CA. 2008. Rapidly evolving human promoter regions. *Nat. Genet.* 40:1262–63
143. Tennessen JA, Akey JM. 2011. Parallel adaptive divergence among geographically diverse human populations. *PLoS Genet.* 7:e1002127
144. Tennessen JA, Bigham AW, O'Connor TD, Fu W, Kenny EE, et al. 2012. Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science* 337:64–69
145. Tennessen JA, O'Connor TD, Bamshad MJ, Akey JM. 2011. The promise and limitations of population exomics for human evolution studies. *Genome Biol.* 12:127
146. Teshima KM, Coop G, Przeworski M. 2006. How reliable are empirical genomic scans for selective sweeps? *Genome Res.* 16:702–12
147. Tishkoff SA, Reed FA, Ranciaro A, Voight BF, Babbitt CC, et al. 2007. Convergent adaptation of human lactase persistence in Africa and Europe. *Nat. Genet.* 39:31–40
148. Tishkoff SA, Varkonyi R, Cahinhinan N, Abbes S, Argyropoulos G, et al. 2001. Haplotype diversity and linkage disequilibrium at human *G6PD*: recent origin of alleles that confer malarial resistance. *Science* 293:455–62
149. Turchin MC, Chiang CW, Palmer CD, Sankararaman S, Reich D, Hirschhorn JN. 2012. Evidence of widespread selection on standing variation in Europe at height-associated SNPs. *Nat. Genet.* 44:1015–19
150. Uddin M, Wildman DE, Liu G, Xu W, Johnson RM, et al. 2004. Sister grouping of chimpanzees and humans as revealed by genome-wide phylogenetic analysis of brain gene expression profiles. *Proc. Natl. Acad. Sci. USA* 101:2957–62
151. US Census Bur. 2013. *World POPClock Projection.* **http://www.census.gov/population/popclockworld.html**
152. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, et al. 2001. The sequence of the human genome. *Science* 291:1304–51
153. Vernot B, Stergachis AB, Maurano MT, Vierstra J, Neph S, et al. 2012. Personal and population genomics of human regulatory variation. *Genome Res.* 22:1689–97
154. Voight BF, Kudaravalli S, Wen X, Pritchard JK. 2006. A map of recent positive selection in the human genome. *PLoS Biol.* 4:e72
155. Wakeley J. 2004. Recent trends in population genetics: More data! More math! Simple models? *J. Hered.* 95:397–405
156. Wall JD, Lohmueller KE, Plagnol V. 2009. Detecting ancient admixture and estimating demographic parameters in multiple human populations. *Mol. Biol. Evol.* 26:1823–27

157. Ward LD, Kellis M. 2012. Evidence of abundant purifying selection in humans for recently acquired regulatory functions. *Science* 337:1675–78
158. Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, et al. 2002. Initial sequencing and comparative analysis of the mouse genome. *Nature* 420:520–62
159. Watterson GA. 1975. On the number of segregating sites in genetical models without recombination. *Theor. Popul. Biol.* 7:256–76
160. Wong WSW, Nielsen R. 2004. Detecting selection in noncoding regions of nucleotide sequences. *Genetics* 167:949–58
161. Xu S, Li S, Yang Y, Tan J, Lou H, et al. 2010. A genome-wide search for signals of high-altitude adaptation in Tibetans. *Mol. Biol. Evol.* 28:1003–11
162. Xue Y, Sun D, Daly A, Yang F, Zhou X, et al. 2008. Adaptive evolution of *UGT2B17* copy-number variation. *Am. J. Hum. Genet.* 83:337–46
163. Yi X, Liang Y, Huerta-Sanchez E, Jin X, Cuo ZX, et al. 2010. Sequencing of 50 human exomes reveals adaptation to high altitude. *Science* 329:75–78
164. Yohn CT, Jiang Z, McGrath SD, Hayden KE, Khaitovich P, et al. 2005. Lineage-specific expansions of retroviral insertions within the genomes of African great apes but not humans and orangutans. *PLoS Biol.* 3:e110
165. Zeng J, Konopka G, Hunt BG, Preuss TM, Geschwind D, Yi SV. 2012. Divergent whole-genome methylation maps of human and chimpanzee brains reveal epigenetic basis of human regulatory evolution. *Am. J. Hum. Genet.* 91:455–65

Annual Review of Genomics and Human Genetics

Volume 14, 2013

Contents

Errata

An online log of corrections to *Annual Review of Genomics and Human Genetics* articles may be found at http://genom.annualreviews.org