Published in final edited form as:

Annu Rev Anthropol. 2008 ; 37: 197–217. doi:10.1146/annurev.anthro.37.081407.085141.

Detecting the Genetic Signature of Natural Selection in Human Populations: Models, Methods, and Data

Angela M. Hancock and Anna Di Rienzo

Department of Human Genetics, University of Chicago, Chicago, Illinois 60637

Abstract

Patterns of DNA sequence variation in the genome contain a record of past selective events. The ability to collect increasingly large data sets of polymorphisms has allowed investigators to perform hypothesis-driven studies of candidate genes as well as genome-wide scans for signatures of adaptations. This genetic approach to the study of natural selection has identified many signals consistent with predictions from anthropological studies. Selective pressures related to variation in climate, diet, and pathogen exposure have left strong marks on patterns of human variation. Additional signals of adaptations are observed in genes involved in chemosensory perception and reproduction. Several ongoing projects aim to sequence the complete genome of 1000 individuals from different human populations. These large-scale projects will provide data for more complete genome scans of selection, but more focused studies aimed at testing specific hypotheses will continue to hold an important place in elucidating the history of adaptations in humans.

Keywords

genetic adaptations; selective sweeps; neutrality tests

INTRODUCTION

From a genetic perspective, evolution can be defined as changes in allele frequencies over time due to mutation, genetic drift, migration, and natural selection. The investigation of natural selection and genetic adaptations in humans has held a central place in biological anthropology as well as in disciplines such as human genetics and evolutionary biology. Although these disciplines differ slightly in the questions they address and the approaches they use, they intersect in the field of human population genetics, which uses genetic variation data to learn about the past demographic events and the history of natural selection in human populations.

Traditional approaches to identifying adaptations to local environments have often relied on comparing the distribution of a phenotype to an environmental variable hypothesized to mirror a selective pressure. This was, for example, pertinent to the malaria hypothesis, which explains the geographic distribution of the β -thalassemia (Haldane 1949) and HbS (Allison 1954) alleles in terms of the resistance they confer to malaria. Additional notable examples of this approach focused on quantitative phenotypes, such as the relationship between body mass and temperature (Katzmarzyk & Leonard 1998, Roberts 1978), skin

Copyright © 2008 by Annual Reviews. All rights reserved

ahancock@uchicago.edu, dirienzo@uchicago.edu.

DISCLOSURE STATEMENT The authors are not aware of any biases that might be perceived as affecting the objectivity of this review.

pigmentation and solar radiation (Jablonski & Chaplin 2000), and oxygen saturation of arterial hemoglobin and altitude (Beall et al. 1997). If the phenotypes are heritable, the correlation between the distribution of phenotype and environmental variable constitutes evidence for an adaptation at the genetic level (as opposed, for example, to acclimatization). The argument for positive selection is especially strong if a correlation can also be shown between the phenotype of interest (e.g., high oxygen saturation at high altitude) and a measure of fitness (e.g., number of surviving offspring) (e.g., Beall et al. 2004).

Although these traditional approaches directly test the hypothesis that natural selection acted on a given genetic variant, their application on a large scale is hampered by several limitations. They require investigators to collect phenotype data, which is expensive and time consuming, and analyze large samples to achieve adequate statistical power. More recently, an approach based on the genetic signature of natural selection on patterns of variation has found wide applicability in studies of adaptations in humans and other species. This approach is based on the idea that natural selection introduces a local perturbation in the patterns of neutral genetic variation surrounding an advantageous allele relative to regions of the genome where variation is shaped only by genetic drift. This approach has two major advantages: First, it does not require investigators to collect phenotype information, and second, it can detect adaptive changes resulting from selection coefficients, e.g., 1%, that would be very hard to detect by traditional phenotype-based approaches (Gillespie 1991). The recent development of technologies and resources for studying human genetic variation on an unprecedented scale has allowed investigators to scan the entire human genome for signals of positive natural selection. Moreover, as the same technologies are being applied to the study of the genetic bases of common diseases, signals of natural selection can be connected to genotype-phenotype association signals, where both signals arise from unbiased genome-scale analyses.

Despite the somewhat separate trajectories of anthropological, human genetic, and evolutionary genetic studies of human adaptations, by weaving together strands from these disciplines, it may soon be possible to reconstruct the history of selective events that occurred during human evolution and to infer the role of human adaptations in health and disease. Here, we briefly review models of selection and methods for detecting its signature. In describing the main signals of selection reported so far, we focus on those that pertain to variation that is still segregating in humans, rather than to mutations that became fixed between humans and closely related primates. Our goal is to provide a starting point for the synthesis of findings from the different disciplines that are contributing to our understanding of human adaptations.

MODELS OF POSITIVE NATURAL SELECTION

Natural selection can be defined as the process by which beneficial heritable traits increase in frequency over time and unfavorable heritable traits become less common. This process can occur according to a variety of models; the resulting signatures vary considerably across selection models and for different values of the relevant parameters (e.g., strength of selection, mode of inheritance, age of the selective pressure). Therefore, consideration of these models is crucial for interpreting empirical patterns of variation in regions of the genome that experienced positive selection.

Standard selection models involve loci that have two alleles, and selection occurs if the fitness differs across the three genotypes. Directional selection occurs when one of the two alleles is favored over the other so that its frequency may increase until it reaches fixation. The time needed for a new beneficial allele to become fixed is a function of the selection coefficient and mode of inheritance, with favored dominant and additive alleles increasing in

frequency much faster than recessive ones (see Figure 1a). Under balancing selection, the heterozygote has fitness greater than that of both homozygotes; in this case, a new allele quickly reaches a stable equilibrium frequency and fluctuates around that frequency for as long as the selective pressure is present (see Figure 1b). Therefore, although both directional and balancing selection will result in the rapid increase in frequency of a new beneficial allele, the former ultimately eliminates variation from a population while the latter tends to maintain it. If balancing selection is due to a long-standing selective pressure, a polymorphism may be maintained in the population well beyond the expected lifetime of a neutral polymorphism (i.e., $4N_e$ generations for an autosomal locus, where N_e is the effective population size). Another classical model is diversifying selection in which two or more phenotypes are simultaneously favored. This model generally predicts an increase in levels of genetic variation.

In addition to the models above, scenarios with temporally and spatially varying selection may be particularly relevant to humans (Gillespie 1991). Temporal variation in selective pressures probably resulted from the climatic changes of the Ice Ages, the diversity of habitats to which humans became exposed during their dispersal, and the environmental changes arising from the impact of human activities, e.g., the spread of agriculture. The new selective pressures may favor new alleles—as described in the models above—or standing (previously neutral) alleles, which may afford a faster adaptive response to environmental shifts. These two models make different predictions about the population dynamics of the favored allele and the expected signature of selection. The difference in predictions is mainly because a beneficial allele that was previously neutral underwent an initial phase of random drift before being driven to high frequency by selection; this initial neutral phase has important consequences for the expected patterns of linked variation. In addition to environmental and selective changes over the time span of human evolution, there is also ample evidence for spatial variation in selective pressures; this may be the consequence of variation in features of the physical environment, such as climate, or of the different subsistence strategies adopted by humans, which in turn influence other aspects of the environment, e.g., diet or pathogen exposure. In some cases, selective pressures may vary spatially in a graded fashion, e.g., climate or ultraviolet (UV) exposure, with important implications for the expected geographic distribution of favored alleles. Finally, differences across environments may lead to spatial variation in the intensity of purifying selection (i.e., selection removing deleterious mutations); for example, a gene that was under strong purifying selection in a given environment may undergo relaxation of selective constraints when the population migrates to a different location or adopts a different lifestyle.

METHODS FOR DETECTING THE SIGNATURE OF NATURAL SELECTION

As shown in Figure 1, natural selection acting on a beneficial variant has a dramatic impact on its rate of frequency change over time, resulting in a trajectory that is distinct from that expected for neutral variants. Under directional selection or during the initial phase of a balanced polymorphism, beneficial alleles are characterized by a rapid rise in frequency. Because of the low recombination distance, the histories of the beneficial allele and the nearby neutral alleles are strongly correlated. Because of this correlation, natural selection generates a local perturbation in the pattern of variation tightly linked to the selected site. In contrast, patterns of variation in neutrally evolving genomic regions are influenced only by genetic drift and, therefore, by properties of the population, including the history of population size and population structure. Therefore, detecting the signature of natural selection essentially consists in distinguishing the patterns of variation that are shaped solely by demographic history from those that are influenced by natural selection, in addition to demography.

Briefly, as shown in Figure 2, when a new advantageous allele is introduced into the population by the mutation process, it will be associated with a particular haplotype background. As this allele is driven quickly to an intermediate or high frequency, the neutral alleles that define the haplotype and that are tightly linked to the selected site will also tend to increase in frequency. Owing to the rapid increase in frequency of the advantageous allele, recombination does not erode the association between the selected site and the surrounding neutral sites, thereby generating a local pattern of extended identical haplotypes [referred to as extended haplotype homozygosity (EHH)] that occur at intermediate to high frequencies. This process is often referred to as a partial or incomplete selective sweep. A number of statistical tests have been developed to detect this haplotype pattern (Hanchard et al. 2006, Hudson et al. 1994, Innan et al. 2005, Sabeti et al. 2002, Voight et al. 2006). If selection is directional, the advantageous allele may go to fixation. In this case, which is referred to as a complete selective sweep, all variation near the selected site will also be fixed and only new mutations arising during the sweep will segregate in the population at low frequencies. Therefore, the expected pattern near a fixed advantageous allele consists of a reduction in polymorphism levels and a relative abundance of rare alleles. At a greater distance from the selected site, recombination events will tend to uncouple the advantageous allele from the neutral alleles. This may result in a pattern characterized by high frequency derived alleles (i.e., new mutations); because such alleles tend to be rare under neutrality, the occurrence of multiple high frequency derived alleles within a small region constitutes a strong signal of selection. Therefore, different aspects of the data (i.e., polymorphism levels, allele frequency spectrum, haplotype structure) will be informative to detect complete versus incomplete selective sweeps. A number of statistical tests have been developed to capture the impact of positive natural selection under different models (see Appendix); we refer the reader to several recent reviews for details (Biswas & Akey 2006, Nielsen 2005, Nielsen et al. 2007, Sabeti et al. 2006).

When selection acts on a variant that is advantageous only in a subset of populations, the frequency of that variant may differ across populations to a greater extent than predicted for variants evolving neutrally in all populations. Several approaches have been devised to detect such adaptations to local environments. Historically, the most widely used is based on the statistic F_{ST} and its modifications (Beaumont & Balding 2004, Consortium 2005), which simply summarize allele frequency differences between pairs of populations or among multiple populations (Weir 1996). Variants with unusually large F_{ST} values are typically interpreted as being the targets of local selective pressures (Lewontin & Krakauer 1973).

If the intensity of selection varies spatially in a graded fashion, advantageous allele frequencies may vary across populations following the geographic distribution of the selective pressure. Therefore, if the selective pressure—or a good proxy for it—is known, a test of the correlation between the advantageous allele frequency and the value of the environmental variable may provide evidence for spatially varying selection. This approach is particularly appropriate when the advantageous phenotype is known to have a clinal distribution; for example, this is true for human skin pigmentation as well as body mass and proportionality, which are correlated with UV radiation and temperature, respectively (Jablonski & Chaplin 2000, Katzmarzyk & Leonard 1998).

Additional information regarding the action of natural selection on genetic variation may be extracted by comparing the pattern of variation within species and among species (i.e., humans and a close outgroup) for synonymous and nonsynonymous variants within a coding region. If a gene is evolving neutrally, the ratios of nonsynonymous to synonymous mutations within species and among species are expected to be the same. However, if natural selection drives multiple advantageous mutations to fixation within the same gene, this action may generate an excess of nonsynonymous relative to synonymous changes

among species compared with the ratio observed within species. Alternatively, an excess of nonsynonymous mutations may be observed in the variation within species relative to among species. One possible interpretation for this pattern is that diversifying selection maintains a higher number of nonsynonymous variants; this explanation likely applies to the HLA genes (see below). However, a more common interpretation of this pattern, which is often observed in human genes, is that weak purifying (rather than positive) selection acts on nonsynonymous polymorphisms. Under this scenario, the nonsynonymous variants in excess are slightly deleterious mutations that reach nontrivial frequency in populations but are unlikely to become fixed.

Assessing Statistical Significance of Selection Signals

From the statistical standpoint, two main approaches have been used to detect the signature of natural selection: model-based and empirical approaches. In the model-based approach, a theoretical model of population history in which all variation is neutral is used to develop quantitative expectations about patterns of variation (quantitative expectations for different models of demography under neutrality can be readily obtained using simulations). If a test locus exhibits variation patterns that are significantly different from those expected under the model, the null hypothesis of neutrality is rejected. A variant of this approach is to compare the likelihood of the data under neutrality with that under a specified model of selection, using the statistical framework of the likelihood ratio test (Kim & Stephan 2002, Nielsen et al. 2005, Przeworski 2003). The drawback of model-based methods is that, if the assumptions about population history are violated, one may falsely reject neutrality. Although much progress has been made in reconstructing the broad picture of human population history (Schaffner et al. 2005, Stajich & Hahn 2005, Voight et al. 2005), it is widely recognized that the details of the true history are too complex to be captured by simple models. Unlike the model-based method, the empirical method is agnostic about population history, and, under the assumption that most loci in the human genome evolve neutrally, it aims at identifying the loci with the most unusual patterns compared with large-scale data sets of genetic variation. For example, empirical patterns of variation may be summarized by means of one or more test statistics, which are used to rank loci in the genome; those loci falling above an arbitrary cut-off (typically the top 5%) are identified as unusual and are often referred to as outliers. A recently developed composite likelihood method attempts to integrate the advantages of both approaches by comparing a model without selection that is fitted to genome-wide variation data with a model in which selection acted in a particular region of the genome (Nielsen et al. 2005). This method, which is tailored to the detection of complete selective sweeps, is remarkably robust to deviations from the demographic assumptions (Williamson et al. 2007).

Because it relies on large-scale variation data, the outlier approach has gained wider applicability since the completion of genome-wide projects, such as the International HapMap Project (Consortium 2005, Frazer et al. 2007) and the Perlegen Project (Hinds et al. 2005). This approach already generated many interesting candidate targets of selection. However, to reconstruct accurately the history of selective events, all or the vast majority of adaptations must be detected, and this is unlikely to occur with outlier approaches alone. This is because there is a trade-off between false positives and false negatives, the extent of which is sensitive to a number of factors that affect the statistical power of neutrality tests (see below) (Teshima et al. 2006). To complement outlier approaches, investigators can compare the strength of the evidence for positive selection across groups of genes in different biological processes; a significant excess in one group of genes suggests that the biological process as a whole evolved genetic adaptations, even though the signals at individual genes may not reach genome-wide significance (Akey et al. 2002, Tang et al. 2007, Voight et al. 2006, Wang et al. 2006).

Considerations about Statistical Power

The power of neutrality tests based on polymorphism levels, the spectrum of allele frequency, and haplotype structure is affected by a number of factors, most notably the age of the selective event and the strength of selection, whereby younger and more strongly advantageous alleles are more easily detected. In addition, it is important to note that the process illustrated in Figure 2 applies most directly to mutations that are immediately advantageous and that are dominant or codominant. If selection acts on standing—previously neutral—mutations or on recessive new mutations, the expected pattern of variation may not be distinguishable from that typical of neutrally evolving regions of the genome, leading to a marked reduction in power (Hermisson & Pennings 2005; Pennings & Hermisson 2006a,b; Przeworski et al. 2005). The similarity in expected variation patterns under neutrality and under the above selection models is due to the fact that the frequency trajectory of previously neutral or recessive new mutations is characterized by a longer initial phase, prior to the rapid rise in frequency, during which mutation and recombination events can generate substantial diversity in the chromosomal background of the mutation. For the case of selection on standing variation, the signature of selection is unlikely to be detectable if the frequency of the variant at the onset of selection is higher than 5% (Przeworski et al. 2005).

The above factors affect the power of different tests to a varying degree. Therefore, the results of genome-wide selection scans are inherently biased toward the type of signals most powerfully detected by the test used in the scan. This bias complicates the analysis of the overlap of signals across studies and potentially distorts the reconstruction of the overall history of selective pressures. Approaches that combine different summaries of the data have also been implemented in hopes of achieving greater power and accuracy (Tang et al. 2007, Zeng et al. 2006).

SIGNALS OF GENETIC ADAPTATIONS IN HUMAN POPULATIONS

Here, we review briefly the main signals of selection observed in studies of genetic variation within and among human populations. Some of these studies tested specific hypotheses arising from the anthropological literature on phenotypic variation and on its adaptive significance (see, for example, Biasutti 1959; Roberts 1953, 1978; Roberts & Kahlon 1976). With the advent of genome-wide scans for selection signals, investigators can evaluate the evidence for positive selection in an unbiased manner. Genome-wide scans have identified many selection targets that are consistent with previous anthropological studies. In addition, these scans implicated other, perhaps less expected signals of selection.

Climate and the Physical Environment

That climate and the physical environment played an important role in shaping human phenotypic variation is evident from the strong correlations between traits such as body mass, basal metabolic rate, or skin reflectance and variables such as temperature, latitude, and UV radiation (Henry & Rees 1991, Jablonski & Chaplin 2000, Leonard et al. 2005, Roberts & Kahlon 1976). Other phenotypes likely to have similar geographic distributions include those related to sodium homeostasis, salt and water retention, and thermogenesis. Overall, studies of phenotypic variation have provided strong support for the notion that heat and cold stress as well as UV radiation have exerted strong selective pressures on humans; genetic studies of selection have largely supported these predictions.

UV radiation—Among the phenotypes with complex patterns of inheritance, skin pigmentation is one of the more completely characterized phenotypes at the genetic level (Rees 2003 and references therein), with contributions from studies of animal models [e.g.,

Kelsh et al. 1996, Mammalian phenotype ontology database (see Related Resources)], Mendelian pigmentation disorders, and genotype-phenotype association (Bonilla et al. 2005; Duffy et al. 2004, 2007; Flanagan et al. 2000; Graf et al. 2005, 2007; Kanetsky et al. 2002; Lamason et al. 2005; Palmer et al. 2000; Rebbeck et al. 2002; Smith et al. 1998; Stokowski et al. 2007; Sulem et al. 2007).

An early study of the genetic signature of selection on skin pigmentation examined variation at the melanocortin 1 receptor (*MC1R*) locus, which influences variation in skin and hair color. The authors found a larger proportion of nonsynonymous variants in non-African compared with African samples, which was interpreted as evidence for strong purifying selection against nonsynonymous mutations in Africa and relaxation of constraints outside Africa (Harding et al. 2000). A recent study focused on a strong candidate gene for skin pigmentation, *SLC24A5*, which, when mutated, causes a pigmentation defect in zebrafish that can be rescued by the human gene (Lamason et al. 2005). In humans, a nonsynonymous alanine to threonine polymorphism at position 111 is associated with variation in skin pigmentation among African American and African Caribbean individuals. The light pigmentation allele is fixed or nearly fixed in Europeans, and, consistent with the expectations for a complete selective sweep, the genomic region spanning *SLC24A5* exhibits low levels of polymorphism.

The availability of extensive variation data allowed investigators to search for selection signatures in candidate genes for skin pigmentation on the basis of strong population differentiation (i.e., high F_{ST}) and EHH (Izagirre et al. 2006, Lao et al. 2007, Myles et al. 2007, Norton et al. 2007). The geographic distribution of some of these candidate single nucleotide polymorphisms (SNPs) was characterized in further detail; these analyses found evidence for variation leading to lighter pigmentation at *OCA2* in both Europe and Asia and at *ASIP* worldwide, whereas the *SLC24A5*, *MATP* and *TYR* genes showed evidence for selection only in Europeans. Because signatures of selection for lighter pigmentation are not found at the same genes in Europe and Asia, convergent evolution for reduced pigmentation appears to be relatively common for this trait (Norton et al. 2007).

Unbiased scans for signals of adaptations using genome-wide data have also provided strong evidence for selection on skin pigmentation. Strong signals were detected using haplotype structure [*OCA2*, *TYRP1*, *DNTPB1*, *SLC24A5* and *MYO5A* by Voight et al. (2006) and Tang et al. (2007)], EHH and strong population differentiation [*SLC24A5*, *SLC45A2* by Sabeti et al. (2007)], strong population differentiation alone [*ABCC11*, *SLC45A2*, *SLC24A5* by Barreiro et al. (2008)], and frequency spectrum [*KITLG*, *MATP*, *SILV*, *OCA2*, *TRPM1*, *SLC24A5*, *RAB27A*, *MC2R*, *ATRN* by Williamson et al. (2007)].

Heat and cold stress—The role of heat stress as a selective pressure in humans has been summarized in the sodium-retention hypothesis, which posits that, in hot, humid environments, selection for high sodium retention was strong because salts are lost quickly through sweat but are important for maintaining temperature homeostasis (Denton 1982, Gleibermann 1973). One prediction of this hypothesis is that genetic variation underlying adaptations to heat stress is correlated with climatic variables (e.g., temperature and potential evaporation) or latitude as a proxy for climate. In addition, signals consistent with complete or partial selective sweeps may be observed within individual populations. Consistent with these predictions, the frequency of genetic variants implicated in interindividual variation in sodium retention and risk to hypertension was significantly correlated with latitude. Researchers observed this pattern in a number of genes including those coding for angiotensinogen (*AGT*), which plays a vital role in the renin-angiotensin pathway, cytochrome P450 3A5 (*CYP3A5*), which activates cortisol in the kidney, and the G-protein beta-3 subunit, which is involved in signal transduction in a number of tissues

(Thompson et al. 2004, Young et al. 2005). In the case of the *AGT* and the *CYP3A5* genes, investigators also observed signals of natural selection using tests of the frequency spectrum or haplotype structure (Nakajima et al. 2004, Thompson et al. 2004).

In addition to heat tolerance, selective pressures related to climate probably acted to increase cold tolerance by shaping genetic variation in energy metabolism. Consistent with a role for positive selection on cold tolerance-enhancing variation, Mishmar et al. (2003) and Ruiz-Pesini et al. (2004) showed that the number and frequencies of nonsynonymous mutations in mitochondrial genes increase with distance from Africa. Furthermore, we recently investigated a large number of candidate susceptibility genes for common metabolic disorders (e.g., type 2 diabetes, obesity) to determine whether there was widespread evidence for spatially varying selection in energy metabolic genes (Hancock et al. 2008). As with the sodium-retention genes, we found that variation in these genes was significantly correlated with climate variables in worldwide population samples. Several of the strongest signals included variants previously associated with disease, and several of these fell in thermogenesis pathway genes.

The results of genome-wide selection scans have generally been consistent with those from hypothesis-driven studies. Carlson et al. (2005) scanned the human genome for signals of selection using the frequency spectrum and found evidence for selection at *CYP3A5*. Voight et al. (2006) used the genotype data from the HapMap project to test for evidence of an incomplete selective sweep by detecting regions of EHH across the genome (Voight et al. 2006); this study found significant signals at the *CYP3A5* and the leptin receptor (*LEPR*) genes, both showing strong correlations between allele frequency and climate variables. In addition to individual signals, they found a significant excess of signals in genes involved in biological processes related to energy metabolism, such as carbohydrate metabolism and electron transport (in Asians and Europeans, respectively). Tang et al. (2007) used a different measure of EHH that preferentially targeted signals in high frequency alleles; this study found signals for the *CYP3A5* region and for a group of genes with similar function (i.e., oxidoreductase activity). Williamson et al. (2007) used a composite likelihood test to scan the genome for signals of complete selective sweeps (Williamson et al. 2007); this study identified several strong signals in genes that may be important for heat or cold stress adaptations, including those coding for sterol carrier protein 2 (*SCP2*), farnesoid X receptor (*NR1H4*), and activator of heat shock 90kDa protein ATPase homolog 1 (*AHSA1*).

Diet and Food Availability

Specific dietary components—Lactase persistence is the production, after infancy, of the enzyme lactase, which breaks down the milk sugar lactose into glucose and galactose so that it can be further processed in the intestines. Although uncommon in nonhuman species, in some human populations lactase persistence is relatively common. Prevalence of lactase persistence varies greatly among human populations and has been shown to cooccur with cultural traits that involve the inclusion of dairy products in the diet. A cline in the lactase persistence phenotype exists within Europe such that persistence is highest among populations in the Northwest and lowest among those in the Southeast (Swallow 2003 and references therein). But the observed pattern is not restricted to Europe: Pastoral populations in the Middle East (e.g., Bedouin) and sub-Saharan Africa (e.g., Fulani and Tutsi) also have high prevalence of lactase persistence. These findings led to the hypothesis that alleles causing lactase persistence were advantageous when populations adopted dairy farming and shifted to a diet in which milk is a major adult staple (Kretchmer 1972, McCracken 1971, Simoons 1970).

In Europeans, lactase persistence was shown to be due to a polymorphism about 14 kb upstream of the gene coding for the lactase enzyme (*LCT*). This polymorphism, denoted C/

T₋₁₃₉₁₀, is likely to affect lactase production via changes in transcription levels. The genomic region containing this polymorphism was investigated for signatures of positive natural selection in ethnically diverse population samples, with a special focus on the haplotype associated with persistence in Europeans (Bersaglieri et al. 2004). Consistent with the idea that lactase persistence was advantageous in European populations, multiple polymorphisms displayed unusually large differences in allele frequencies between European and non-European populations, and the haplotype structure exhibited extremely high EHH in European populations. The time since expansion of the C/T₋₁₃₉₁₀ variant in Europeans was estimated to be 2188–20650 years, roughly consistent with the likely time of onset of dairy farming. The above findings were bolstered by genome-wide scans for selection, which showed that the *LCT* gene in Europeans contains the strongest signal of a partial selective sweep in the human genome (Consortium 2005, Sabeti et al. 2007, Voight et al. 2006, Williamson et al. 2007).

Although it is common in Europeans, the C/T₋₁₃₉₁₀ variant is absent or very rare in some African and Middle Eastern populations where lactase persistence is high, which led to the hypothesis that different variants underlie the lactase persistence phenotype across populations. Indeed, genotype-phenotype association studies in African pastoral populations found that several polymorphisms near the C/T₋₁₃₉₁₀ (G/T₋₁₄₀₁₀, T/G₋₁₃₉₁₅, and C/G₋₁₃₉₀₇) may cause lactase persistence. Furthermore, *in vitro* studies showed that these alleles affected *LCT* transcription levels. These findings suggested that convergent evolution occurred for lactase persistence in Europeans and in African pastoralist populations that regularly drink milk (Tishkoff et al. 2007). The high frequency of the lactase persistence allele(s) in African pastoralists compared with nonpastoralists and the high EHH among pastoralists are consistent with the notion that these alleles were driven quickly to high frequency by positive natural selection. As with European populations, the time estimated since the expansion of the lactase persistence allele in Africa, i.e., 3000–7000 years ago, is consistent with selection beginning around the time when a pastoral subsistence strategy was adopted in Africa. Enattah et al. (2008) described a high frequency haplotype in Saudi Arabians that is defined by two polymorphisms (G₋₁₃₉₁₅ and C₋₃₇₁₂) upstream to the *LCT* gene and is correlated with lactase persistence; *in vitro* assays showed that the variants defining this haplotype influenced *LCT* transcription (Enattah et al. 2008). As with the European and African lactase persistence haplotypes, they found evidence that this haplotype had been driven quickly to high frequency by positive selection. They estimated the age for the haplotype to be ~4000 years, broadly consistent with the timing of camel domestication in the region.

Starch consumption, like adult milk consumption, varies greatly across populations with different subsistence, most notably across agricultural versus nonagricultural populations. Therefore, the ability to digest and absorb starch is likely to have become advantageous since the spread of agriculture. Perry et al. (2007) recently presented evidence supporting this hypothesis. They examined variation in the number of copies of the amylase gene (*AMY1*), which encodes an enzyme important in starch hydrolysis, between agricultural and nonagricultural populations. They found that the agricultural populations had, on average, more copies of *AMY1* than did the combined hunter-gatherer and pastoralist populations, whose copy numbers were intermediate between those of the agricultural populations and chimpanzees and bonobos.

Alcohol dehydrogenase genes in the *ADH1* cluster, which play an important role in ethanol metabolism, harbor variation associated with susceptibility to alcoholism. This includes a nonsynonymous polymorphism (R47H) in the *ADH1B* gene, which has near-fixation frequency in Asian populations and is rare elsewhere (Osier et al. 2002a,b). Indeed, a genome-wide selection scan based on population differentiation found that the value for

ADH1B R47H was exceptionally high (Consortium 2005). Osier et al. (2002b) reported that several variants on the same haplotype showed high population differentiation, such that the haplotype was common in East Asian populations and was rare or absent elsewhere. Additionally, in their genome-wide scan, Voight et al. (2006) found evidence for selection at the *ADH1* gene cluster on the basis of strong EHH in East Asian populations (Voight et al. 2006). These results strongly argue for a selective advantage conferred by the *ADH1* polymorphism(s). However, whereas *ADH1* polymorphisms clearly affect variation in the processing of a common dietary component, i.e., ethanol, it is unclear whether ethanol consumption itself was the selective pressure underlying the observed population genetic pattern. Because the high-activity *ADH1* alleles result in an accumulation of acetaldehyde in response to alcohol load and because acetaldehyde has antiprotozoal activity, Goldman & Enoch (1990) proposed that these alleles are selectively advantageous because they protect against severe infectious diseases by protozoans.

Other specific dietary components may also be important. For example, two genome scans found evidence for selection from EHH for genes important in vitamin transporter activity and cofactor transporter activity (Tang et al. 2007, Voight et al. 2006). In addition, Wang et al. (2006) found signatures of positive selection for several genes involved in protein metabolism (*ADAMTS19–20*, *APEH*, *PLAU*, *HDAC8*, *UBR1*, and *USP26*) and a significant excess of signals in the group of genes in this biological category compared with other groups.

Food availability—In addition to variation in dietary components, fluctuation in food availability likely exerted strong selective pressures on the human genome. More than 40 years ago, Neel proposed the thrifty genotype hypothesis to explain why variants that increase risk to type 2 diabetes and obesity may be at high frequency in human populations (Neel 1962). He reasoned that because ancestral populations underwent seasonal cycles of feast and famine, they would have benefited from having extremely efficient fat and carbohydrate storage. When food production and storage resulted in more reliable food availability, this ancestral thriftiness became detrimental, and in contemporary populations, it contributes to the increased prevalence of diabetes and obesity. Consistent with the notion that thriftiness is an ancestral state, many alleles that increase risk to type 2 diabetes and other metabolic disorders are ancestral (i.e., shared with chimpanzee), whereas the alternative alleles at those polymorphic sites protect against the disease (*Chimpanzee Seq. Anal. Consort.* 2005, Di Rienzo & Hudson 2005).

Although researchers have well documented that the transition to a Western lifestyle and diet results in major prevalence increases of type 2 diabetes (O'dea 1991; Szathmary 1986, 1990; Weiss et al. 1984), it is unclear whether this is also true for the transition from hunting-gathering to agriculture. In a revision of the original thrifty genotype hypothesis, Neel proposed that the changes that likely accompanied the spread of agriculture, including the reduction in dietary diversity to a diet composed mainly of carbohydrates, represented an important step in the shift to overall environmental conditions that favor the development of type 2 diabetes in individuals carrying the thrifty genotype (Neel 1999). However, whether food availability and reliability were indeed higher in agriculturalists compared with foragers is arguable (Cohen & Armelagos 1984, Larsen 2003). For example, Benyshek & Watson (2006) compared the quantity of available food as well as the frequency and intensity of food shortages between contemporary agricultural and hunter-gatherer populations and found no significant differences. However, some interesting trends were observed: Occasional and mild-to-moderate food shortages were more common among agricultural populations, whereas frequent and severe shortages were more common among hunter-gatherers.

A corollary of the thrifty genotype hypothesis is that populations with reliable, steady access to food resources may have evolved adaptations that slowed the insulin response and decreased the storage of energy as fat, thus resulting in a decrease in type 2 diabetes prevalence over time. In this context, Diamond (2003) proposed that the low prevalence of type 2 diabetes in Europeans reflects a longer history of stable food supply in these populations. Under this scenario, alleles that protect against type 2 diabetes are expected to have increased in frequency and to carry a signature of positive selection (e.g., be associated with high EHH). It is important to note that changes of allele frequencies, even if driven by strong positive selection, take at least hundreds of generations. Therefore, the above scenario applies only if the change in diet and life style and the resulting shift in selective pressures occurred long before Westernization, possibly with the transition to agriculture as suggested by Neel (1999).

Until now a comprehensive test of the thrifty genotype hypothesis has not been possible because only few susceptibility variants are known; however, studies of individual genes have suggested some signals of selection. Fullerton et al. (2002) found that susceptibility variants in the calpain 10 (*CAPN10*) gene had large differences in allele frequencies between African and non-African populations and suggested that natural selection had favored different alleles across populations (Fullerton et al. 2002). Another study of the same gene found both a significant reduction in variability for the alleles carrying the putative protective allele, consistent with positive selection acting on this subset of alleles, and a region of unusually high polymorphism and decay of linkage disequilibrium (LD), consistent with balancing selection in the region (Vander Molen et al. 2005). More recently, Grant et al. (2006) showed that variation in the *TCF7L2* gene affects risk to type 2 diabetes. The haplotype carrying the protective allele at *TCF7L2* is associated with a signature of positive selection on the basis of tests of population differentiation and haplotype homozygosity (Helgason et al. 2007).

Response to Pathogens

Response to pathogens was likely an important selective pressure during human evolutionary history. Although selection in response to pathogen attack has probably been important throughout human evolutionary history, selection likely intensified over the past 10–20 kya as a result of a warming climate, more recent localized changes in environments, and the shifts to agriculture and animal husbandry.

Malaria—Whereas selection for resistance to several pathogens has been studied, response to malaria pathogens has received an exceptional amount of attention owing to the large number of people affected and the high resultant mortality. The most deadly strain by far, *Plasmodium falciparum*, is likely to have gained prominence rather recently; its prevalence is thought to have increased as a consequence of the transition to agriculture. A less deadly but widespread strain, *Plasmodium vivax*, is associated with high levels of morbidity; it is unclear if the spread of *P. vivax* occurred after the spread of agriculture, as for *P. falciparum*. A number of common polymorphisms confer resistance to malaria infection, including a number of hemoglobin variants such as Hb S, Hb C, Hb E, β -thalassemia, α -thalassemia, G6PD deficiency, and the null allele of the Duffy blood group (Kwiatkowski 2005). Most of these variants carry a genetic signature of selection.

Deficiency of the enzyme G6PD is correlated with reduced risk of malaria infection, and signatures of positive selection have been observed for alleles that confer this deficiency (Sabeti et al. 2002, Saunders et al. 2005, Tishkoff et al. 2001, Verrelli et al. 2002). Tishkoff et al. (2001) and Saunders et al. (2005) found reduced variability in the subset of haplotypes carrying the deficiency alleles, evidence for EHH associated with the haplotypes carrying

these alleles, and long branches separating the deficiency alleles from other alleles. The ages of the G6PD deficiency alleles were estimated using different methods and yielded values ranging from 2.5 to 45 kya (Coop & Griffiths 2004, Saunders et al. 2005, Tishkoff et al. 2001, Verrelli et al. 2002); the younger estimates are consistent with an increase in malaria infection with the transition to agriculture, but the older ones are not.

Carriers of the Duffy null (*FY*0*) mutation do not express a protein on the red blood cells to which *P. vivax* binds to invade the cells. As a result, *FY*0* homozygotes are resistant to this variety of malaria (Livingstone 1984); Kasehagen et al. (2007) recently showed that *FY*0* heterozygosity may also afford some degree of protection against vivax malaria. This allele has a striking geographic distribution in that it is fixed or nearly fixed in most Sub-Saharan African populations and is virtually absent elsewhere. This degree of differentiation in the frequency of the *FY*0* allele is highly unusual relative to genome-wide patterns (Consortium 2005) and is consistent with the notion that this allele underwent a complete selective sweep in Sub-Saharan African populations. Consistent with this hypothesis, Hamblin and colleagues showed that polymorphism levels in sub-Saharan Africans are low compared with divergence and that there is an excess of high frequency derived alleles near the *FY* gene (Hamblin & Di Rienzo 2000, Hamblin et al. 2002). Interestingly, the *FY*0* allele is found on two major haplotypes, which suggests that this is a case of selection on standing neutral variation.

Several hemoglobin variants, the most common of which include Hb S, Hb C, and Hb E, represent well-documented cases of balanced polymorphism whereby heterozygotes are resistant to *P. falciparum*, whereas homozygotes have reduced fitness. Accordingly, signatures of positive selection have been observed at the *HBB* locus on the basis of extensive haplotype homozygosity and strong population differentiation for HbS (Hanchard et al. 2006, Sabeti et al. 2007) and strong haplotype structure for HbC (Wood et al. 2005) and HbE (Ohashi et al. 2004). Consistent with the idea that the onset of malaria selection coincided with the spread of agriculture, the ages of the HbS, HbE, and HbC variants have been estimated to be 1350–2100 years (Currat et al. 2002), 1240–4440 years (Ohashi et al. 2004), and 1875–3750 years (Wood et al. 2005), respectively.

HLA region—Genes in the HLA region have received a great deal of attention in studies of mechanisms and selection for pathogen defense as a result of their roles in response to a broad range of pathogens. Balancing selection appears to have been pervasive in this region owing to the need for different HLA peptides to confer resistance to disease, but there is also evidence for strong directional selection on specific alleles. Although the HLA region contains at least 120 genes, many of which are involved in immune function, most studies have focused on nine, which are considered to be the classical HLA genes. Several lines of evidence suggest a history of natural selection at the classical HLA genes. As a group, these genes show exceptionally high levels of polymorphism especially in functional (peptide binding) regions (Meyer & Thomson 2001, Satta et al. 1994), and many genes have an excess of nonsynonymous compared with synonymous mutations (Hughes & Nei 1988). In addition to the high levels of polymorphism and nonsynonymous variation, the frequency spectrum tends to be skewed toward intermediate frequency variants relative to neutral expectation. Large sequence differences are observed among the HLA haplotypes, suggesting that the alleles are old, consistent with a model of long-term balancing selection (Meyer & Thomson 2001).

In addition to studies directed at single genes and small groups of related genes, multigene and genome-wide studies of selection have detected significant signals in immune response genes, either individually [*CD226*, *IGJ* (Akey et al. 2002, Carlson et al. 2005, Williamson et al. 2007); *ABO*, *IL1A*, *IL1RN*, *KEL* (Akey et al. 2002), *ABO*, *KEL* (Carlson et al. 2005)] or

as a group. Tang et al. (2007) found signals of positive selection for interleukin-1 receptor antagonist activity and cytokine activity. Using a test based on haplotype structure, Voight et al. (2006) found an enrichment for MHC-I-mediated immunity genes, and Wang et al. (2006) found signals in several genes (*CSF2*, *CCNT2*, *DEFB118*, *STAB1*, *SP*, *Zap70*) and an enrichment for pathogen response genes. Sabeti et al. (2007) found signals of positive selection based on haplotype structure and population differentiation at *LARGE* and *DMD*.

Other Signals of Selection

Chemosensory perception—The ability to sense aspects of the environment is necessary for survival and reproduction, and the array of cues encountered in different environments may be especially diverse for some senses, such as olfaction and taste. Olfaction may be important to detect and identify food and mates, whereas taste is likely crucial to identify suitable foods. Evidence that genes responsible for olfaction and taste evolved under positive selection is beginning to emerge.

Gilad et al. (2003) found evidence of positive selection on the human lineage compared with chimpanzees on the basis of the ratio of nonsynonymous to synonymous substitutions, low levels of polymorphism, and an excess of rare variants. Additional evidence for adaptation in olfactory receptor genes was detected in genome-wide scans. One study identified a signal for a complete selective sweep in a region within a cluster of olfactory receptor genes (Williamson et al. 2007); other studies found evidence for selection based on the frequency spectrum in several olfactory receptor genes (Carlson et al. 2005) and an enrichment of signals based on haplotype structure in genes involved in olfaction (Voight et al. 2006).

Inter-individual variation in the ability to taste PTC has been known for more than 70 years, but the genetic basis for this trait was only recently elucidated (Drayna et al. 2003 and references therein). Identification of the region that influences the PTC phenotype allowed testing for evidence of positive selection at the genetic level (Wooding et al. 2004). Two high-frequency haplotypes, defined by three nonsynonymous variants, were observed, consistent with the action of balancing selection at this locus. An additional 21 bitter taste receptor genes have been analyzed for evidence of selection: the average ratio of nonsynonymous to synonymous substitutions was high compared with that observed in 151 other genes; in addition, polymorphism levels tended to be high in the bitter taste receptor genes (Kim et al. 2005).

Fertility and reproduction—Genes that affect fertility and reproduction are expected to be subject to especially strong selection owing to their direct effects on fitness-related traits. This class of proteins evolves rapidly both within and among species (Swanson & Vacquier 2002). Significant variation in fertility exists among human populations with different subsistence patterns (Bentley et al. 1993). To the extent that variation in fertility within populations is heritable (Pettay et al. 2005), it is possible that natural selection has acted on this trait to influence variation among populations.

Stefansson et al. (Stefansson et al. 2005) showed that female carriers of a large inversion polymorphism, which shows evidence of selection based on population differentiation and long-range haplotype structure, have higher fertility than do noncarriers. Follicle stimulating hormone is important for fertility in both males (for Sertoli cell proliferation and in maintaining sperm quality in the testes) and females (for stimulation of ovarian follicles). Grigorova et al. (2007) examined sequence patterns for the follicle stimulating hormone receptor-binding beta-subunit (*FSHB*) in several human populations and found low overall diversity combined with an excess of intermediate frequency variants, consistent with balancing selection at this locus. The authors also found that one of the two high-frequency

haplotypes was associated with increased fertility and hypothesized that balancing selection at this locus may act on birth intervals.

Although relatively few single gene studies have been conducted for genes involved in fertility and reproduction, several signatures of selection have been found for genes in this class from genome-wide scans for selection. Williamson et al. (2007) found evidence for a complete sweep in European Americans and Han Chinese for *SPAG6*, a gene involved in sperm motility, and others found evidence for an excess of signals of partial selective sweeps in genes related to fertility, gametogenesis, spermatogenesis, sperm motility, and fertilization (Voight et al. 2006, Wang et al. 2006).

CONCLUSIONS

Early studies examined evidence for positive selection from the distribution of phenotypic traits and their cooccurrence with variables that were expected to represent underlying selective pressures. With the advent of DNA sequencing and genotyping technologies and the development of methods to detect evidence of selection from sequence variation data, testing for evidence of genetic adaptations in single genes became feasible. More recently, the availability of dense, genome-wide genotype data for multiple populations and the development of methods for detecting selection using SNP data have elicited many genome-wide scans for evidence of positive selection in human populations. Now, the HapMap project is expanding to include genotype data for additional populations as well as genome-wide resequencing data for some HapMap individuals. In addition, the 1000 Genomes Project was recently launched, which aims to resequence completely the genomes of 1000 individuals from diverse worldwide populations. The continued collection of data from large-scale projects will be useful for conducting more complete genome scans to detect evidence for positive selection. At the same time, more focused studies to test specific hypotheses or to follow up on results from genome-wide scans will continue to have an important place in reconstructing the overall history of selective pressures among human populations.

Glossary

Genetic drift	the change in frequency of a mutation due to the random sampling of alleles from one generation to the next
Fitness	a measure of the contribution of an organism to the next generation, related to the capacity to survive and reproduce
Fixation	the process by which a mutation reaches 100% frequency in the population
EHH	extended haplotype homozygosity
Selective sweep	the process by which an advantageous mutation removes linked neutral variation while it is driven to high frequency by selection
Allele frequency spectrum	the distribution of allele frequencies at polymorphic sites, specified by the proportion of alleles in different frequency ranges
Nonsynonymous mutations	mutations that change the amino acid sequence of a protein
Synonymous mutations	mutations that do not change the amino acid sequence of a protein

HLA	human leukocyte antigen
International HapMap Project	a project to characterize the haplotype structure of the human genome by genotyping more than 3.1 million SNPs in diverse populations
Perlegen Project	a project to characterize patterns of genetic variation 1.6 million SNPs in three diverse populations
SNP	single nucleotide polymorphism
kya	thousand years ago
G6PD	glucose-6-phosphate dehydrogenase
PTC	phenylthiocarbamide
Effective population size (N_e)	the size of an ideal population with the same rate of genetic drift of gene frequencies as the actual population

Acknowledgments

We thank Cynthia Beall and William Leonard for helpful comments on the manuscript. We gratefully acknowledge funding from National Institutes of Health grants DK56670 and GM79558. A.M.H. was partially supported by a predoctoral fellowship from the American Heart Association.

APPENDIX: STATISTICAL TESTS OF NEUTRALITY

Several methods have been developed to evaluate statistically whether empirical patterns of variation are consistent with the expectations of the null hypothesis of evolutionary neutrality. These tests, which are briefly described here, can be grouped on the basis of the aspect of variation that they use.

Amount of Variation Within and Between Species

The most widely used test to detect departures from neutrality based on polymorphism levels is the so-called HKA (Hudson-Kreitman-Aguade) test (Hudson et al. 1987), which is based on the notion that polymorphism levels within species and sequence divergence between species are proportional to the same underlying mutation rate. Therefore, by comparing polymorphism and divergence at two or more loci, one can test for departures from neutrality due to a deficit or an excess of polymorphism.

Spectrum of Allele Frequencies

The standard neutral model provides quantitative expectations for the spectrum of allele frequencies, whereby the expected fraction of alleles that occur i times in a sample is proportional to $1/i$. The most widely used test statistic summarizing information about the frequency spectrum is Tajima's D (TD) (Tajima 1989), which is based on the standardized difference between two estimators of the same parameter, $4N_e\mu$ (where N_e is the effective population size and μ is the mutation rate per site per generation). Under the standard neutral model, TD is expected to be near zero; an excess of intermediate frequency variants will generate a positive TD value, whereas an excess of rare variants will give rise to a negative TD value (because TD does not use information about the ancestral state of the polymorphisms, it does not distinguish between rare and high-frequency new mutations). A different test, referred to as H (Fay & Wu 2000), uses information from an outgroup

sequence to infer the ancestral allele at each polymorphic site and tests specifically for an excess of high-frequency nonancestral alleles.

Haplotype Structure

This family of tests relies on the property of partial sweeps to generate an extended region of haplotype homozygosity surrounding the selected site (Figure 2) (Hudson et al. 1994). In general, the extent of haplotype homozygosity is a function of the age of the mutation, with younger mutations exhibiting homozygosity over larger regions than would older ones. Under neutrality, young mutations are expected to be rare, whereas equally young advantageous mutations tend to occur at higher frequency. Therefore, tests based on the haplotype structure aim to capture the discrepancy between the extent of haplotype homozygosity in a region and the frequency of that haplotype in the population. The EHH in a region can be calculated for a core haplotype or a core SNP allele and then compared with that for other core haplotypes or the alternative allele at the core SNP. To perform a statistical test of neutrality, the relative EHH at a candidate region is compared with that for other core haplotypes or core SNPs within the same frequency range, obtained either by simulations or from empirical patterns (Sabeti et al. 2002, Voight et al. 2006).

Spatial Distribution of Allele Frequencies

Tests to detect the impact of selection on the extent of allele frequency differentiation have historically relied on the summary statistic F_{ST} , which is the proportion of the total genetic variance that occurs among populations. The spatial distribution of allele frequencies can also be tested for a correlation with an environmental variable, e.g., latitude or temperature, which is likely to reflect a selective pressure. For both F_{ST} and the correlation between allele frequency and an environmental variable, a statistical test of neutrality can be constructed by comparing the value of the test statistic at a candidate locus with the distribution of values expected for a null neutral model of subdivided populations. Because the models of human population structure investigated so far are too simplistic to provide an appropriate null model, an empirical approach is usually taken in which the value of the test statistic at the candidate locus is compared with the distribution of the statistic observed for a large set of independent loci.

LITERATURE CITED

- Akey JM, Zhang G, Zhang K, Jin L, Shriver MD. Interrogating a high-density SNP map for signatures of natural selection. *Genome Res* 2002;12:1805–14. [PubMed: 12466284]
- Allison AC. Protection afforded by sickle-cell trait against subtertian malarial infection. *Br. Med. J* 1954;1:290–94. [PubMed: 13115700]
- Barreiro LB, Laval G, Quach H, Patin E, Quintana-Murci L. Natural selection has driven population differentiation in modern humans. *Nat. Genet* 2008;40:340–45. [PubMed: 18246066]
- Beall CM, Song K, Elston RC, Goldstein MC. Higher offspring survival among Tibetan women with high oxygen saturation genotypes residing at 4,000 m. *Proc. Natl. Acad. Sci. USA* 2004;101:14300–4. [PubMed: 15353580]
- Beall CM, Strohl KP, Blangero J, Williams-Blangero S, Decker MJ, et al. Quantitative genetic analysis of arterial oxygen saturation in Tibetan highlanders. *Hum. Biol* 1997;69:597–604. [PubMed: 9299881]
- Beaumont MA, Balding DJ. Identifying adaptive genetic divergence among populations from genome scans. *Mol. Ecol* 2004;13:969–80. [PubMed: 15012769]
- Bentley GR, Goldberg T, Jasienska G. The fertility of agricultural and nonagricultural traditional societies. *Popul. Stud. A J. Demogr* 1993;47:269–81.

- Benyshek DC, Watson JT. Exploring the thrifty genotype's food-shortage assumptions: a cross-cultural comparison of ethnographic accounts of food security among foraging and agricultural societies. *Am. J. Phys. Anthropol* 2006;131:120–26. [PubMed: 16485298]
- Bersaglieri T, Sabeti PC, Patterson N, Vanderploeg T, Schaffner SF, et al. Genetic signatures of strong recent positive selection at the lactase gene. *Am. J. Hum. Genet* 2004;74:1111–20. [PubMed: 15114531]
- Biasutti, R. *Le Razze e i Popoli Della Terra*. Unione Tipografico/Editrice Torinese; Torino: 1959.
- Biswas S, Akey JM. Genomic insights into positive selection. *Trends Genet* 2006;22:437–46. [PubMed: 16808986]
- Bonilla C, Boxill LA, Donald SA, Williams T, Sylvester N, et al. The 8818G allele of the agouti signaling protein (*ASIP*) gene is ancestral and is associated with darker skin color in African Americans. *Hum. Genet* 2005;116:402–6. [PubMed: 15726415]
- Carlson CS, Thomas DJ, Eberle MA, Swanson JE, Livingston RJ, et al. Genomic regions exhibiting positive selection identified from dense genotype data. *Genome Res* 2005;15:1553–65. [PubMed: 16251465]
- Chimpanzee Seq. Anal. Consort. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* 2005;437:69–87. [PubMed: 16136131]
- Cohen, MN.; Armelagos, GJ. *Paleopathology at the Origins of Agriculture*. Academic; Orlando, FL: 1984. p. 635
- Consortium IH. A haplotype map of the human genome. *Nature* 2005;437:1299–320. [PubMed: 16255080]
- Coop G, Griffiths RC. Ancestral inference on gene trees under selection. *Theor. Popul. Biol* 2004;66:219–32. [PubMed: 15465123]
- Curat M, Trabuchet G, Rees D, Perrin P, Harding RM, et al. Molecular analysis of the beta-globin gene cluster in the Niokholo Mandenka population reveals a recent origin of the beta(S) Senegal mutation. *Am. J. Hum. Genet* 2002;70:207–23. [PubMed: 11741197]
- Denton, D. *The Hunger for Salt: An Anthropological, Physiological and Medical Analysis*. Springer-Verlag; New York: 1982.
- Diamond J. The double puzzle of diabetes. *Nature* 2003;423:599–602. [PubMed: 12789325]
- Di Rienzo A, Hudson RR. An evolutionary framework for common diseases: the ancestral-susceptibility model. *Trends Genet* 2005;21:596–601. [PubMed: 16153740]
- Drayna D, Coon H, Kim UK, Elsner T, Cromer K, et al. Genetic analysis of a complex trait in the Utah Genetic Reference Project: a major locus for *PTC* taste ability on chromosome 7q and a secondary locus on chromosome 16p. *Hum. Genet* 2003;112:567–72. [PubMed: 12624758]
- Duffy DL, Box NF, Chen W, Palmer JS, Montgomery GW, et al. Interactive effects of *MC1R* and *OCA2* on melanoma risk phenotypes. *Hum. Mol. Genet* 2004;13:447–61. [PubMed: 14709592]
- Duffy DL, Montgomery GW, Chen W, Zhao ZZ, Le L, et al. A three-single-nucleotide polymorphism haplotype in intron 1 of *OCA2* explains most human eye-color variation. *Am. J. Hum. Genet* 2007;80:241–52. [PubMed: 17236130]
- Enattah NS, Jensen TG, Nielsen M, Lewinski R, Kuokkanen M, et al. Independent introduction of two lactase-persistence alleles into human populations reflects different history of adaptation to milk culture. *Am. J. Hum. Genet* 2008;82:57–72. [PubMed: 18179885]
- Fay JC, Wu CI. Hitchhiking under positive Darwinian selection. *Genetics* 2000;155:1405–13. [PubMed: 10880498]
- Flanagan N, Healy E, Ray A, Philips S, Todd C, et al. Pleiotropic effects of the melanocortin 1 receptor (*MC1R*) gene on human pigmentation. *Hum. Mol. Genet* 2000;9:2531–37. [PubMed: 11030758]
- Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, et al. A second generation human haplotype map of over 3.1 million SNPs. *Nature* 2007;449:851–61. [PubMed: 17943122]
- Fullerton SM, Bartoszewicz A, Ybazeta G, Horikawa Y, Bell GI, et al. Geographic and haplotype structure of candidate type 2 diabetes susceptibility variants at the calpain-10 locus. *Am. J. Hum. Genet* 2002;70:1096–106. [PubMed: 11891618]

- Gilad Y, Bustamante CD, Lancet D, Paabo S. Natural selection on the olfactory receptor gene family in humans and chimpanzees. *Am. J. Hum. Genet* 2003;73:489–501. [PubMed: 12908129]
- Gillespie, JH. *The Causes of Molecular Evolution*. Oxford Univ. Press; New York: 1991. p. 336
- Gleibermann L. Blood pressure and dietary salt in human populations. *Ecol. Food Nutr* 1973;2:143–56.
- Goldman D, Enoch MA. Genetic epidemiology of ethanol metabolic enzymes: a role for selection. *World Rev. Nutr. Diet* 1990;63:143–60. [PubMed: 2197794]
- Graf J, Hodgson R, van Daal A. Single nucleotide polymorphisms in the *MATP* gene are associated with normal human pigmentation variation. *Hum. Mutat* 2005;25:278–84. [PubMed: 15714523]
- Graf J, Voisey J, Hughes I, van Daal A. Promoter polymorphisms in the *MATP* (*SLC45A2*) gene are associated with normal human skin color variation. *Hum. Mutat* 2007;28:710–17. [PubMed: 17358008]
- Grant SF, Thorleifsson G, Reynisdottir I, Benediktsson R, Manolescu A, et al. Variant of transcription factor 7-like 2 (*TCF7L2*) gene confers risk of type 2 diabetes. *Nat. Genet* 2006;38:320–23. [PubMed: 16415884]
- Grigorova M, Rull K, Laan M. Haplotype structure of *FSHB*, the beta-subunit gene for fertility-associated follicle-stimulating hormone: possible influence of balancing selection. *Ann. Hum. Genet* 2007;71:18–28. [PubMed: 17227474]
- Haldane, JBS. The rate of mutation of human genes. *Int. Congr. Genet.*, 8th; Stockholm, Sweden. 1949.
- Hamblin MT, Di Rienzo A. Detecting the signature of natural selection in humans: evidence from the Duffy blood group locus. *Am. J. Hum. Genet* 2000;66:1669–79. [PubMed: 10762551]
- Hamblin MT, Thompson EE, Di Rienzo A. Complex signatures of natural selection at the Duffy blood group locus. *Am. J. Hum. Genet* 2002;70:369–83. [PubMed: 11753822]
- Hanchard NA, Rockett KA, Spencer C, Coop G, Pinder M, et al. Screening for recently selected alleles by analysis of human haplotype similarity. *Am. J. Hum. Genet* 2006;78:153–59. [PubMed: 16385459]
- Hancock AM, Witonsky DB, Gordon AS, Eshel G, Pritchard JK, et al. Adaptations to climate in candidate genes for common metabolic disorders. *PLoS Genet* 2008;4(2):e32. [PubMed: 18282109]
- Harding RM, Healy E, Ray AJ, Ellis NS, Flanagan N, et al. Evidence for variable selective pressures at MC1R. *Am. J. Hum. Genet* 2000;66:1351–61. [PubMed: 10733465]
- Helgason A, Palsson S, Thorleifsson G, Grant SF, Emilsson V, et al. Refining the impact of *TCF7L2* gene variants on type 2 diabetes and adaptive evolution. *Nat. Genet* 2007;39:218–25. [PubMed: 17206141]
- Henry CJ, Rees DG. New predictive equations for the estimation of basal metabolic rate in tropical peoples. *Eur. J. Clin. Nutr* 1991;45:177–85. [PubMed: 1824042]
- Hermisson J, Pennings PS. Soft sweeps: molecular population genetics of adaptation from standing genetic variation. *Genetics* 2005;169:2335–52. [PubMed: 15716498]
- Hinds DA, Stuve LL, Nilsen GB, Halperin E, Eskin E, et al. Whole-genome patterns of common DNA variation in three human populations. *Science* 2005;307:1072–79. [PubMed: 15718463]
- Hudson RR, Bailey K, Skarecky D, Kwiatowski J, Ayala FJ. Evidence for positive selection in the superoxide dismutase (Sod) region of *Drosophila melanogaster*. *Genetics* 1994;136:1329–40. [PubMed: 8013910]
- Hudson RR, Kreitman M, Aguade M. A test of neutral molecular evolution based on nucleotide data. *Genetics* 1987;116:153–59. [PubMed: 3110004]
- Hughes AL, Nei M. Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature* 1988;335:167–70. [PubMed: 3412472]
- Innan H, Zhang K, Marjoram P, Tavare S, Rosenberg NA. Statistical tests of the coalescent model based on the haplotype frequency distribution and the number of segregating sites. *Genetics* 2005;169:1763–77. [PubMed: 15654103]

- Izagirre N, Garcia I, Junquera C, de la Rua C, Alonso S. A scan for signatures of positive selection in candidate loci for skin pigmentation in humans. *Mol. Biol. Evol* 2006;23:1697–706. [PubMed: 16757656]
- Jablonski NG, Chaplin G. The evolution of human skin coloration. *J. Hum. Evol* 2000;39:57–106. [PubMed: 10896812]
- Kanetsky PA, Swoyer J, Panossian S, Holmes R, Guerry D, Rebbeck TR. A polymorphism in the agouti signaling protein gene is associated with human pigmentation. *Am. J. Hum. Genet* 2002;70:770–75. [PubMed: 11833005]
- Kasehagen LJ, Mueller I, Kiniboro B, Bockarie MJ, Reeder JC, et al. Reduced *Plasmodium vivax* erythrocyte infection in PNG Duffy-negative heterozygotes. *PLoS ONE* 2007;2:e336. [PubMed: 17389925]
- Katzmarzyk PT, Leonard WR. Climatic influences on human body size and proportions: ecological adaptations and secular trends. *Am. J. Phys. Anthropol* 1998;106:483–503. [PubMed: 9712477]
- Kelsh RN, Brand M, Jiang YJ, Heisenberg CP, Lin S, et al. Zebrafish pigmentation mutations and the processes of neural crest development. *Development* 1996;123:369–89. [PubMed: 9007256]
- Kim U, Wooding S, Ricci D, Jorde LB, Drayna D. Worldwide haplotype diversity and coding sequence variation at human bitter taste receptor loci. *Hum. Mutat* 2005;26:199–204. [PubMed: 16086309]
- Kim Y, Stephan W. Detecting a local signature of genetic hitchhiking along a recombining chromosome. *Genetics* 2002;160:765–77. [PubMed: 11861577]
- Kretchmer N. Lactose and lactase. *Sci. Am* 1972;227:71–78. [PubMed: 4672311]
- Kwiatkowski DP. How malaria has affected the human genome and what human genetics can teach us about malaria. *Am. J. Hum. Genet* 2005;77:171–92. [PubMed: 16001361]
- Lamason RL, Mohideen MA, Mest JR, Wong AC, Norton HL, et al. *SLC24A5*, a putative cation exchanger, affects pigmentation in zebrafish and humans. *Science* 2005;310:1782–86. [PubMed: 16357253]
- Lao O, de Gruijter JM, van Duijn K, Navarro A, Kayser M. Signatures of positive selection in genes associated with human skin pigmentation as revealed from analyses of single nucleotide polymorphisms. *Ann. Hum. Genet* 2007;71:354–69. [PubMed: 17233754]
- Larsen CS. Animal source foods and human health during evolution. *J. Nutr* 2003;133:3893S–97S. [PubMed: 14672287]
- Leonard WR, Snodgrass JJ, Sorensen MV. Metabolic adaptation in indigenous Siberian populations. *Annu. Rev. Anthropol* 2005;34:451–71.
- Lewontin RC, Krakauer J. Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics* 1973;74:175–95. [PubMed: 4711903]
- Livingstone FB. The Duffy blood groups, *Vivax malaria*, and malaria selection in human populations: a review. *Hum. Biol* 1984;56:413–25. [PubMed: 6386656]
- McCracken RD. Origins and implications of the distribution of adult lactase deficiency in human populations. *J. Trop. Pediatr. Environ. Child Health* 1971;17:7–10. [PubMed: 5209856]
- Meyer D, Thomson G. How selection shapes variation of the human major histocompatibility complex: a review. *Ann. Hum. Genet* 2001;65:1–26. [PubMed: 11415519]
- Mishmar D, Ruiz-Pesini E, Golik P, Macaulay V, Clark AG, et al. Natural selection shaped regional mtDNA variation in humans. *Proc. Natl. Acad. Sci. USA* 2003;100:171–76. [PubMed: 12509511]
- Myles S, Somel M, Tang K, Kelso J, Stoneking M. Identifying genes underlying skin pigmentation differences among human populations. *Hum. Genet* 2007;120:613–21. [PubMed: 16977434]
- Nakajima T, Wooding S, Sakagami T, Emi M, Tokunaga K, et al. Natural selection and population history in the human angiotensinogen gene (*AGT*): 736 complete *AGT* sequences in chromosomes from around the world. *Am. J. Hum. Genet* 2004;74:898–916. [PubMed: 15077204]
- Neel JV. Diabetes mellitus: a “thrifty” genotype rendered detrimental by “progress”? *Am. J. Hum. Genet* 1962;14:353–62. [PubMed: 13937884]
- Neel JV. Diabetes mellitus: a “thrifty” genotype rendered detrimental by “progress”? 1962. *Bull. World Health Organ* 1999;77:694–703. discussion 692–93. [PubMed: 10516792]

- Nielsen R. Molecular signatures of natural selection. *Annu. Rev. Genet* 2005;39:197–218. [PubMed: 16285858]
- Nielsen R, Hellmann I, Hubisz M, Bustamante C, Clark AG. Recent and ongoing selection in the human genome. *Nat. Rev. Genet* 2007;8:857–68. [PubMed: 17943193]
- Nielsen R, Williamson S, Kim Y, Hubisz MJ, Clark AG, Bustamante C. Genomic scans for selective sweeps using SNP data. *Genome Res* 2005;15:1566–75. [PubMed: 16251466]
- Norton HL, Kittles RA, Parra E, McKeigue P, Mao X, et al. Genetic evidence for the convergent evolution of light skin in Europeans and East Asians. *Mol. Biol. Evol* 2007;24:710–22. [PubMed: 17182896]
- O'dea, D. Traditional diet of Australian Aboriginal hunter-gatherers. In: Whiten, A.; Widdowson, EM., editors. *Foraging Strategies and Natural Diet of Monkeys, Apes, and Humans*. Oxford Univ. Press; Oxford, UK: 1991. p. 73–81.
- Ohashi J, Naka I, Patarapotikul J, Hananantachai H, Brittenham G, et al. Extended linkage disequilibrium surrounding the hemoglobin E variant due to malarial selection. *Am. J. Hum. Genet* 2004;74:1198–208. [PubMed: 15114532]
- Osier MV, Pakstis AJ, Goldman D, Edenberg HJ, Kidd JR, Kidd KK. A proline-threonine substitution in codon 351 of *ADH1C* is common in Native Americans. *Alcohol Clin. Exp. Res* 2002a;26:1759–63. [PubMed: 12500098]
- Osier MV, Pakstis AJ, Soodyall H, Comas D, Goldman D, et al. A global perspective on genetic variation at the *ADH* genes reveals unusual patterns of linkage disequilibrium and diversity. *Am. J. Hum. Genet* 2002b;71:84–99. [PubMed: 12050823]
- Palmer JS, Duffy DL, Box NF, Aitken JF, O'Gorman LE, et al. Melanocortin-1 receptor polymorphisms and risk of melanoma: Is the association explained solely by pigmentation phenotype? *Am. J. Hum. Genet* 2000;66:176–86. [PubMed: 10631149]
- Pennings PS, Hermisson J. Soft sweeps II—molecular population genetics of adaptation from recurrent mutation or migration. *Mol. Biol. Evol* 2006a;23:1076–84. [PubMed: 16520336]
- Pennings PS, Hermisson J. Soft sweeps III: the signature of positive selection from recurrent mutation. *PLoS Genet* 2006b;2:e186. [PubMed: 17173482]
- Perry GH, Dominy NJ, Claw KG, Lee AS, Fiegler H, et al. Diet and the evolution of human amylase gene copy number variation. *Nat. Genet* 2007;39:1256–60. [PubMed: 17828263]
- Pettay JE, Kruuk LE, Jokela J, Lummaa V. Heritability and genetic constraints of life-history trait evolution in preindustrial humans. *Proc. Natl. Acad. Sci. USA* 2005;102:2838–43. [PubMed: 15701704]
- Przeworski M. Estimating the time since the fixation of a beneficial allele. *Genetics* 2003;164:1667–76. [PubMed: 12930770]
- Przeworski M, Coop G, Wall JD. The signature of positive selection on standing genetic variation. *Evol. Int. J. Org. Evol* 2005;59:2312–23.
- Rebeck TR, Kanetsky PA, Walker AH, Holmes R, Halpern AC, et al. P gene as an inherited biomarker of human eye color. *Cancer Epidemiol. Biomarkers Prev* 2002;11:782–84. [PubMed: 12163334]
- Rees JL. Genetics of hair and skin color. *Annu. Rev. Genet* 2003;37:67–90. [PubMed: 14616056]
- Roberts DF. Body weight, race and climate. *Am. J. Phys. Anthropol* 1953;11:533–58. [PubMed: 13124471]
- Roberts, DF. *Climate and Human Variability*. Cummings; Menlo Park, CA: 1978.
- Roberts DF, Kahlon DP. Environmental correlations of skin colour. *Ann. Hum. Biol* 1976;3:11–22. [PubMed: 1275430]
- Ruiz-Pesini E, Mishmar D, Brandon M, Procaccio V, Wallace DC. Effects of purifying and adaptive selection on regional variation in human mtDNA. *Science* 2004;303:223–26. [PubMed: 14716012]
- Sabeti PC, Reich DE, Higgins JM, Levine HZ, Richter DJ, et al. Detecting recent positive selection in the human genome from haplotype structure. *Nature* 2002;419:832–37. [PubMed: 12397357]
- Sabeti PC, Schaffner SF, Fry B, Lohmueller J, Varrilly P, et al. Positive natural selection in the human lineage. *Science* 2006;312:1614–20. [PubMed: 16778047]

- Sabeti PC, Varilly P, Fry B, Lohmueller J, Hostetter E, et al. Genome-wide detection and characterization of positive selection in human populations. *Nature* 2007;449:913–18. [PubMed: 17943131]
- Satta Y, O'Huigin C, Takahata N, Klein J. Intensity of natural selection at the major histocompatibility complex loci. *Proc. Natl. Acad. Sci. USA* 1994;91:7184–88. [PubMed: 8041766]
- Saunders MA, Slatkin M, Garner C, Hammer MF, Nachman MW. The extent of linkage disequilibrium caused by selection on *G6PD* in humans. *Genetics* 2005;171:1219–29. [PubMed: 16020776]
- Schaffner SF, Foo C, Gabriel S, Reich D, Daly MJ, Altshuler D. Calibrating a coalescent simulation of human genome sequence variation. *Genome Res* 2005;15:1576–83. [PubMed: 16251467]
- Simoons FJ. Primary adult lactose intolerance and the milking habit: a problem in biologic and cultural interrelations. II. A culture historical hypothesis. *Am. J. Dig. Dis* 1970;15:695–710. [PubMed: 5468838]
- Smith R, Healy E, Siddiqui S, Flanagan N, Steijlen PM, et al. Melanocortin 1 receptor variants in an Irish population. *J. Invest. Dermatol* 1998;111:119–22. [PubMed: 9665397]
- Stajich JE, Hahn MW. Disentangling the effects of demography and selection in human history. *Mol. Biol. Evol* 2005;22:63–73. [PubMed: 15356276]
- Stefansson H, Helgason A, Thorleifsson G, Steinthorsdottir V, Masson G, et al. A common inversion under selection in Europeans. *Nat. Genet* 2005;37:129–37. [PubMed: 15654335]
- Stokowski RP, Pant PV, Dadd T, Fereday A, Hinds DA, et al. A genomewide association study of skin pigmentation in a South Asian population. *Am. J. Hum. Genet* 2007;81:1119–32. [PubMed: 17999355]
- Sulem P, Gudbjartsson DF, Stacey SN, Helgason A, Rafnar T, et al. Genetic determinants of hair, eye and skin pigmentation in Europeans. *Nat. Genet* 2007;39:1443–52. [PubMed: 17952075]
- Swallow DM. Genetics of lactase persistence and lactose intolerance. *Annu. Rev. Genet* 2003;37:197–219. [PubMed: 14616060]
- Swanson WJ, Vacquier VD. Reproductive protein evolution. *Annu. Rev. Ecol. Syst* 2002;33:161–79.
- Szathmary EJE. Diabetes in arctic and sub-arctic populations undergoing acculturation. *Coll. Antropol* 1986;10:145–58.
- Szathmary, EJE. Diabetes in Amerindian populations: the Dogrib studies. In: Swedlund, AC.; Armelagos, GJ., editors. *Diseases in Populations in Transition: Anthropological and Epidemiological Perspectives*. Bergin and Garvey; New York: 1990. p. 75-103.
- Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 1989;123:585–95. [PubMed: 2513255]
- Tang K, Thornton KR, Stoneking M. A new approach for using genome scans to detect recent positive selection in the human genome. *PLoS Biol* 2007;5:e171. [PubMed: 17579516]
- Teshima KM, Coop G, Przeworski M. How reliable are empirical genomic scans for selective sweeps? *Genome Res* 2006;16:702–12. [PubMed: 16687733]
- Thompson EE, Kuttub-Boulos H, Witonsky D, Yang L, Roe BA, Di Rienzo A. *CYP3A* variation and the evolution of salt-sensitivity variants. *Am. J. Hum. Genet* 2004;75:1059–69. [PubMed: 15492926]
- Tishkoff SA, Reed FA, Ranciaro A, Voight BF, Babbitt CC, et al. Convergent adaptation of human lactase persistence in Africa and Europe. *Nat. Genet* 2007;39:31–40. [PubMed: 17159977]
- Tishkoff SA, Varkonyi R, Cahinhinan N, Abbes S, Argyropoulos G, et al. Haplotype diversity and linkage disequilibrium at human *G6PD*: recent origin of alleles that confer malarial resistance. *Science* 2001;293:455–62. [PubMed: 11423617]
- Vander Molen J, Frisse LM, Fullerton SM, Qian Y, del Bosque-Plata L, et al. Population genetics of *CAPN10* and *GPR35*: implications for the evolution of type 2 diabetes variants. *Am. J. Hum. Genet* 2005;76:548–60. [PubMed: 15696418]
- Verrelli BC, McDonald JH, Argyropoulos G, Destro-Bisol G, Froment A, et al. Evidence for balancing selection from nucleotide sequence analyses of human *G6PD*. *Am. J. Hum. Genet* 2002;71:1112–28. [PubMed: 12378426]

- Voight BF, Adams AM, Frisse LA, Qian Y, Hudson RR, Di Rienzo A. Interrogating multiple aspects of variation in a full resequencing data set to infer human population size changes. *Proc. Natl. Acad. Sci. USA* 2005;102:18508–13. [PubMed: 16352722]
- Voight BF, Kudaravalli S, Wen X, Pritchard JK. A map of recent positive selection in the human genome. *PLoS Biol* 2006;4:e72. [PubMed: 16494531]
- Wang ET, Kodama G, Baldi P, Moyzis RK. Global landscape of recent inferred Darwinian selection for *Homo sapiens*. *Proc. Natl. Acad. Sci. USA* 2006;103:135–40. [PubMed: 16371466]
- Weir, BS. *Genetic Data Analysis II: Methods for Discrete Population Genetic Data*. Sinauer; Sunderland, MA: 1996.
- Weiss KM, Ferrell RF, Hanis CL. A new world syndrome of metabolic diseases with a genetic and evolutionary basis. *Yearb. Phys. Anthropol* 1984;27:153–78.
- Williamson SH, Hubisz MJ, Clark AG, Payseur BA, Bustamante CD, Nielsen R. Localizing recent adaptive evolution in the human genome. *PLoS Genet* 2007;3:e90. [PubMed: 17542651]
- Wood ET, Stover DA, Slatkin M, Nachman MW, Hammer MF. The beta-globin recombinational hotspot reduces the effects of strong selection around HbC, a recently arisen mutation providing resistance to malaria. *Am. J. Hum. Genet* 2005;77:637–42. [PubMed: 16175509]
- Wooding S, Kim UK, Bamshad MJ, Larsen J, Jorde LB, Drayna D. Natural selection and molecular evolution in *PTC*, a bitter-taste receptor gene. *Am. J. Hum. Genet* 2004;74:637–46. [PubMed: 14997422]
- Young JH, Chang YP, Kim JD, Chretien JP, Klag MJ, et al. Differential susceptibility to hypertension is due to selection during the out-of-Africa expansion. *PLoS Genet* 2005;1:e82. [PubMed: 16429165]
- Zeng K, Fu YX, Shi S, Wu CI. Statistical tests for detecting positive selection by utilizing high-frequency variants. *Genetics* 2006;174:1431–39. [PubMed: 16951063]

Figure 1.

Allele frequency change over time under a directional selection model (*a*) and a balancing selection model (*b*). In the case of directional selection, different modes of inheritance of the advantageous allele are considered.

Figure 2.

The impact of selective sweeps on patterns of neutral variation tightly linked to an advantageous mutation. Neutral mutations are shown as blue circles, whereas advantageous mutations are shown as red circles. Gray bars indicate the results of recombination events involving a chromosome not carrying the advantageous mutation. Black line boxes mark mutations segregating at a high frequency; a large number of such mutations is a strong signal of selection.