Evolutionary Anthropology 17:146–157 (2008)

ARTICLES

Searching the Genome for Our Adaptations

EUGENE E. HARRIS

The current genomic revolution represents a turning point in our understanding of human evolution. For the first time, we are able to begin to investigate human evolutionary adaptations by comparing our entire genome with the genomes of other animal species with which we are related by descent. We are also able to begin fully investigating the genetic differences within and among human populations to understand exactly how human populations have evolved and adapted over time.

One principal aim of biological anthropology and human evolutionary genetics is to develop a full understanding of human genetic adaptations. Our species has acquired genetic adaptations at two levels. At the first level are adaptations that are common to all human individuals, which we can call species-wide adaptations. Such adaptations evolved along our lineage after we diverged from the lineage that led to chimpanzees and before our species differentiated into geographic populations (Fig. 1). At the second level are adaptations that arose as our species differentiated into geographic populations and adapted to local environmental conditions. These are referred to as population-level adaptations. At both levels, the process of genetic adaptation happened via the same evolutionary mechanism of positive selection. In this process, an advantageous DNA variant is driven to higher frequencies in a population because of the increased reproductive fitness it confers on individuals bearing the variant compared to individuals that do not bear the variant. An important consequence of this process is the telltale footprint or signature left in the genomic region surrounding the selected variant.[1,2] For researchers, the accurate recognition of a signature of positive selection provides a crucial clue for understanding exactly what our adaptations are and their genetic basis. One aim of this paper is to describe methods employed to find evidence of positive selection in comparisons of DNA sequences between species and between individuals within species.

Until recently, our ability to dissect the nature of human adaptations has been limited because the needed level and magnitude of data have been wanting. However, within the past five years or so, this has all changed. A once rather dry terrain is now inundated with DNA data of two major types. The DNA sequence of the full genomes of humans, chimpanzees, and rhesus macaques, as well as the mouse and rat, are available to researchers and are ideal for examining species-wide adaptations. In addition, endeavors such as the International HapMap and Perlegen Sciences projects (see glossary) are genotyping single nucleotide polymorphisms (SNPs) within and among different human populations on a genome-wide scale, providing the data necessary to examine population-level adaptations. In the near future, we can look forward to comparative analyses of the complete genomes of different human individuals, an approach called population genomics. Already, analyses have compared complete coding regions obtained for 20 European Americans and 15 African Americans.[3,4]

Studies of human adaptation using these data have adopted one of two general approaches. In the first approach, called the candidate-gene approach, specific genes or sets of genes are scrutinized for evidence of positive selection based on their known or presumed function or association with a phenotype of medical or anthropological interest. Using this approach, several human traits such as lactase persistence, resistance to vivax malaria, and bitter-taste sensation have been found to show strong evidence of having been positively selected (reviewed in Harris and Meyer[49]). Another approach, known as the genome-scan approach, has become increasingly popular with the availability of large-scale datasets. This approach is aimed at identifying the subsets of positively selected genes (PSGs) within our genome. In only a short time, a relatively large number of genome-scan studies have been undertaken. A second aim of this paper, therefore, is to review and discuss

Eugene Harris is in the Department of Biological Sciences and Geology at Queensborough Community College, City University of New York, and a Visiting Professor at the University of São Paulo, Brazil. His research focuses on human and nonhuman primate evolutionary genetics.
Queensborough Community College, Department of Biological Sciences and Geology, City University of New York, 222-05 56th Avenue, Bayside, NY 11364. E-mail: Eharris@qcc.cuny.edu; harris_eugene@hotmail.com

Key words: positive selection; human evolution; natural selection; population adaptations; human adaptations

DOI 10.1002/evan20174
Published online in Wiley InterScience (www.interscience.wiley.com).

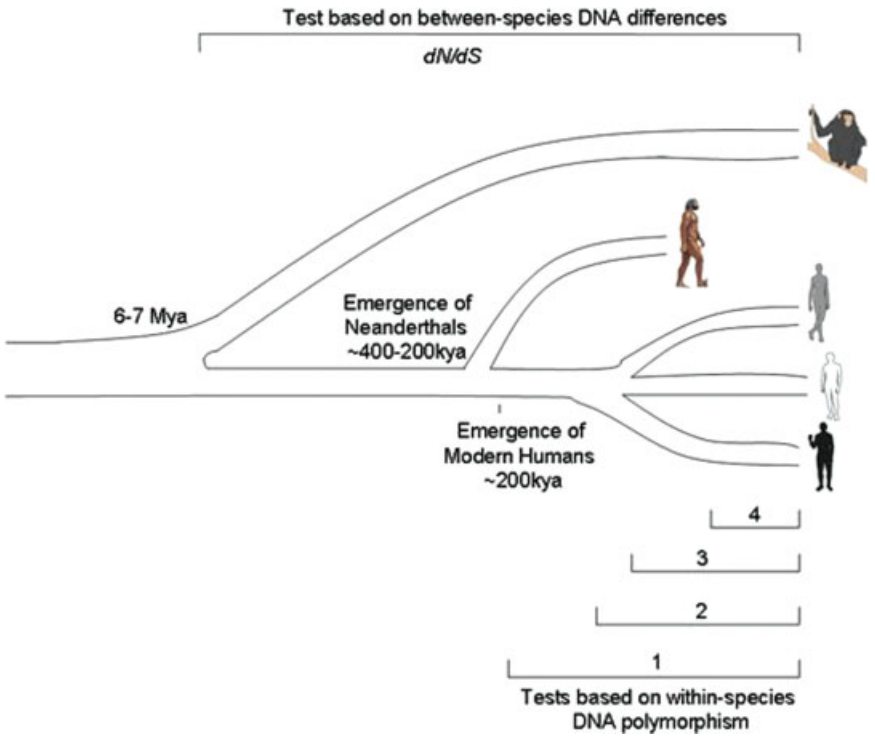

Figure 1. A tree representing the divergence of humans from chimpanzees and the differentiation of human populations. Signatures of positive selection (numbered 1–4 and described in Box 1) are shown according to the windows in the evolutionary past within which they can reveal positive selection. Also shown is the divergence of Neanderthals from the human lineage based on estimated dates.[45] At present, it is unclear how much gene flow, if any, occurred between early modern humans and Neanderthals.[45,47]

the many exciting discoveries that are emerging from these studies.

A central challenge faced by researchers in all such studies is to identify a signature of selection accurately when it is present. The problem stems from the fact that factors other than natural selection have also shaped our genome and its variation. These factors can make it difficult to detect a signature of selection. The persistent bugbear is the imprint that past demography has left on our genome. Demography can have either of two effects: it can make a signature of selection difficult to detect because it acts to conceal it or it can mimic the signature of selection itself and thereby lead to spurious inferences of positive selection. Several different ways have been devised to disentangle the effects of demography from positive selection. One important advance is made possible through the availability of genome-wide data. Because of the general problem that demography presents, a recurring theme in this paper involves a description of how demography can confound the detection of positive selection. In addition, I will discuss various approaches researchers have devised to help overcome demography's confounding effects.

POSITIVE SELECTION ALONG THE HUMAN LINEAGE

Analytical Methods for Detecting Selection

One way to discover adaptations along the human lineage is to analyze the DNA differences, or substitutions, between chimpanzees and humans. Such analyses typically focus on the amino acid coding regions of genes. DNA substitutions that accrue within coding regions will either produce amino-acid changes or will have no effect in changing amino acids. The first type of DNA substitution is called a nonsynonymous substitution; the second type is referred to as a synonymous substitution. About three-quarters of the mutations that arise within populations are nonsynonymous.[5] However, most of these mutations are deleterious and are usually quickly removed through purifying selection.[5] As a consequence of the degenerate nature of the genetic code, only approximately one-quarter of mutations arising in a population are of the synonymous type.[5] Due to genetic drift, these DNA mutations can eventually become fixed in the population. The expected rate at which this occurs can be estimated using the neutral theory of molecular evolution.[6] As a result, synonymous substitutions rather than nonsynonymous substitutions are expected to make up the majority of DNA differences between species. For example, when we compare the coding genomes of chimpanzees and humans, using the mouse or macaque as an outgroup, the rate of nonsynonymous changes that became fixed along the human lineage is estimated to be less than 25% of the rate of synonymous changes.[7,8] This implies that purifying selection has prevented more than 75% of the amino-acid altering mutations from becoming fixed along the human lineage. This is not a surprising result, however, because most proteins play vital roles in the biology of organisms and are therefore generally conserved through the action of purifying selection.

The Between-Species *dN/dS* Test

Positive selection may be inferred when the rate of nonsynonymous substitutions (dN) supersedes the rate of synonymous substitutions (dS). This signature points to the accelerated functional evolution of a gene and can be detected by employing the so-called dN/dS test.[9] This test estimates the ratio between dN and dS obtained by comparing the DNA sequences of two different species. (Note that the test makes a correction for the fact mentioned previously, that synonymous mutations arise at only a fraction of the rate at which nonsynonymous mutations arise.[5]) There are three possibilities: $dN/dS = 1$, indicating no selection or neutrality; $dN/dS < 1$, indicating purifying selection, or selection against deleterious amino acid changes (expected

Glossary

dN—the number of nonsynonymous substitutions per nonsynonymous site (also the nonsynonymous substitution rate).

dS—the number of synonymous substitutions per synonymous site (also the synonymous substitution rate).

***dN/dS* ratio**—the ratio of nonsynonymous substitutions per nonsynonymous site to synonymous substitutions per synonymous site (or the ratio of the nonsynonymous substitution rate to the synonymous substitution rate).

Haplotype—a copy of DNA that bears a unique set of linked DNA variants.

Hitchhiking—the phenomenon whereby neutral variants occurring on the same chromosome as advantageous variants are driven to higher frequency by positive selection along with the advantageous variant.

International HapMap Project (http://www.hapmap.org/)—an international effort to identify the genetic differences in the form of single nucleotide polymorphisms (SNPs) across the entire human genome by screening four different populations: Yoruban, Japanese, Chinese, and European. To date, 3.1 million SNPs have been genotyped.[96]

Perlegen Science SNP Project (http://www.perlegen.com)—a project that has genotyped more than 1.58 million SNPs in 23 African Americans, 24 Han Chinese, and 24 European Americans. The SNPs were initially discovered through a genome-chip array-based method on a subset of individuals from different geographic origins.[97]

Single nucleotide polymorphism (SNP)—a DNA difference between different individuals at a single nucleotide position.

McDonald-Kreitman (MK) test—comparison of the ratio between the number of nonsynonymous to synonymous substitutions between two species to the ratio between the number of nonsynonymous to synonymous polymorphisms within a species.

Positive selection—selection that drives an advantageous variant to higher frequencies within a population.[16]

Purifying selection—selection against deleterious mutations arising in a population (also called negative selection).

Selective sweep—the phenomenon whereby genetic variation is eliminated at neutral sites linked to a selectively favored variant as that variant is driven to higher frequencies by positive selection (Fig. 2).

for important functional genes); and $dN/dS > 1$, often interpreted as indicating positive selection.

The *dN/dS* test has been widely applied as a means to detect positive selection. An oft-cited advantage of the *dN/dS* test is that it is robust to past demographic fluctuations. The reason is that both nonsynonymous and synonymous substitutions are drawn from the same genomic region and both substitutions are assumed to have been equally affected by past demographic events. Therefore, demographic effects will be canceled out in the ratio.[10] Another commonly cited advantage of the *dN/dS* test is that, being based directly on the functional differences between DNA sequences, it has the potential to reveal the site or sites under positive selection. That said, the test is restrictive in a sense because it can only detect selection within amino-acid coding regions of the genome; any cases of positive selection in functionally important noncoding regions, such as gene regulatory regions, will go undetected.

The between-species *dN/dS* test is known to be conservative, meaning that multiple nonsynonymous substitutions are required within a coding region before the test can yield significant results. Therefore, when positive selection acts on a "one-time basis" in the past and produces a substitution at a single nonsynonymous site, it is probable that this change will go undetected by the test. For this reason, the method is best suited to detect positive selection that acted recurrently over extended periods of time, producing multiple nonsynonymous substitutions within a gene.[11] Furthermore, the test is most effective when it is applied to specific functional regions of a gene. This is because a gene as a whole is likely to be under purifying selection acting to conserve its function over evolutionary time. Yet, if positive selection has been confined to a particular functional domain of a protein, the signal may go undetected if the *dN/dS* test is applied over the entire gene. The test, therefore, is most powerful when detailed functional information about the protein is known and when the researcher has an *a priori* hypothesis concerning where the effect of positive selection was localized.[12]

The following are several examples of human genes studied through a candidate-gene approach where use of the *dN/dS* test has been successful at revealing positive selection. When the test is applied to genes in the major histocompatability (MHC) system and genes involved in male reproduction, large excesses of nonsynonymous substitutions over synonymous substitutions are revealed, undoubtedly pointing to the action of positive selection. In MHC genes, excesses are localized within binding domains that determine antigen specificity.[9] In some sperm-specific genes (*PRM1*, *PRM2*, and others), large excesses of nonsynonymous substitutions drive *dN/dS* ratios considerably over 1.0 (ranging from 1.35–2.89) and point to strong positive selection, probably due to sperm-competition.[13] Additional examples include the human genes encoding the candidate sperm-receptor protein on the surface of eggs in females (*PKDREJ*) and the zonadhesion protein (*ZAN*) on the heads of sperm in males. In these examples, excesses of nonsynonymous substitutions are localized within functional domains where

male and female gametes interact.[14,15] Later, I will further discuss selection on gametes.

Genome Scans for Positively Selected Genes (PSGs) Between Humans and Chimpanzees

Genome-scan approaches aim to discover genes within our genome that have experienced positive selection. Most studies have applied the *dN/dS* test or the related McDonald-Kreitman method,[16] which includes polymorphism data, over large sets of orthologous human and chimpanzee genes. Hundreds of PSGs are usually found, but the number often decreases considerably when statistical corrections for significance are applied. Since the evolutionary process is stochastic, a certain number of genes will show test results suggestive of positive selection merely by chance.[11] Such results are known as false positives. The rate of their occurrence can be quite high in genome-scanning studies.[17] For humans and chimpanzees, the problem partly stems from the fact that the overall number of nucleotide differences is exceedingly small (1.5%), which decreases the power of statistical tests.[11,18] For example, the median number of nonsynonymous differences per gene is two, whereas the median number of synonymous differences is three.[7]

The results of different scanning studies (about five studies so far)[3,7,10,19,20] have overlapped considerably in terms of the biological categories into which putative PSGs fall. PSGs tend to be involved in functions of immunity and pathogen-resistance, sensory perception (olfactory and chemosensation), reproduction (fertilization and gametogenesis), and apoptosis. Several studies also found a "transcription factor" category of genes (that is, genes coding for proteins involved in gene regulation) being enriched for PSGs.[3,7,10] These genes include homeotic, forkhead and other genes involved in early development, and could indicate that alterations in gene regulation played an important role in human adaptation.

When the exact sets of putative PSGs in humans are compared with PSGs in chimpanzees, they fall largely into separate functional categories. This seems to point to unique sets of adaptations in these two closely related species. Furthermore, comparing the overall number of putative PSGs in chimpanzees with the overall number in humans indicates that humans have about 50% fewer PSGs than do chimpanzees.[8] This finding is likely to be a consequence of the fact that the effective population size for humans has been estimated to be several fold smaller relative to the value estimated for chimpanzees and also relative to the effective population size estimated for their last common ancestor.[98] The result is not unexpected since the neutral theory predicts that natural selection will be relatively less efficient at fixing advantageous mutations in smaller rather than larger populations because the effects of random genetic drift are increasingly dominant in smaller populations.[6] Using the same reasoning, the neutral model predicts that purifying selection will be less efficient at removing deleterious mutations in populations of small size. Not unexpectedly, the *dN/dS* ratio averaged over the coding region of the genome is significantly greater in humans (0.259) than in chimpanzees (0.245), indicating that significantly fewer nonsynonymous mutations were removed in human evolution than in chimpanzee evolution.[8] I will return to this point in the coda section when I comment on the nonadaptive evolution of our genome.

Caveats of Human-Chimpanzee Genome-Scan Studies

It is not surprising that genome-scan studies have detected positive selection in genes involved in immunological defense, reproduction, and apoptosis. These genes encode proteins that are involved in coevolutionary interactions or instances of internal genomic conflict. For example, immune defense genes are involved in a molecular arms race between pathogens and host cells. For the spermatogenesis genes, the exact cause of positive selection is unknown, but could be related to several factors, including competition among sperm to be the first to fertilize the egg, pathogen-driven selection in the female reproductive tract, selection to distort segregation, or some other factor.[14,18] (For an in-depth discussion of rapid evolution in reproductive proteins, see Swanson and Vacquier.[21]) The category of apoptosis may also be partially related to spermatogenesis since there is naturally a high rate of apoptosis in sperm production and there would be intense competition to avoid destruction.[10] As a consequence of such interactions such as those described, above, genes in these categories are characterized by recurrent bouts of positive selection, producing multiple nonsynonymous substitutions and yielding relatively high *dN/dS* ratios. Thus, it is specifically *recurrent* positive selection that is being detected through this method.

While studies using the *dN/dS* method are best suited to detecting *recurrent* positive selection, we need to consider some important issues. For example, just how common is the mode of recurrent positive selection? Was recurrent positive selection responsible for producing human phenotypic adaptations? Recently, Hughes[11] has raised two quite relevant points of caution here. One is that recurrent natural selection is likely to be rare overall and that many phenotypic adaptations probably result from small numbers of nonsynonymous changes within coding regions. The other is the possibility that important adaptive substitutions have occurred in elements controlling gene expression (that is, regulatory elements) and that fewer have occurred within protein-coding regions, a point originally raised by King and Wilson.[22]

A second problem is that a high *dN/dS* ratio may point to a spurious case of positive selection. For example, a relative increase in the rate of nonsynonymous substitution compared to synonymous substitution can result from a scenario in which purifying selection has been relaxed. One example is the unusual finding in some scanning studies that a rela-

tively large number of human olfactory-receptor genes show evidence of positive selection.[19] An alternative explanation is that a large fraction of these genes experienced relaxed purifying selection. Interestingly, detailed studies of the olfactory gene repertoire revealed that many of these genes have become pseudogenes (that is, nonfunctional genes).[23] Under relaxed selective constraints, multiple nonsynonymous mutations can reach fixation within a species via the effects of random genetic drift, with no role played by positive selection. Another example may be seen in the 25 or more genes within the gene repertoire (*TAS2R*) involved in detecting bitter substances. Some of these genes have relatively high *dN/dS* ratios in primates, but there is disagreement about whether this is due to positive selection or general relaxation of purifying selection in these genes.[24–26] Nevertheless, it is clear that two bitter-taste genes, *TAS2R38* and *TAS2R16*, show relatively unambiguous signs of positive selection in humans, possibly due to selection favoring the ability to detect certain bitter-tasting foods so as to avoid ingesting them.[27,28]

INTEGRATED FUNCTIONAL MORPHOLOGICAL ADAPTATIONS: BIPEDALISM AND OTHER TRAITS

How do we find the sets of genes that underwent positive selection and produced the major morphological and behavioral adaptations of human evolution? How can we ascertain the types of genetic changes that were responsible for these adaptations?

Many unique human adaptations involve complex sets of functionally interrelated morphological features. These include bipedalism, increased hand and finger dexterity, reduced prognathism, reduced size of the dentition, and development of a nonhoning canine-premolar complex with reduced canines. Studies of complex traits in model organisms such as fruit flies, fish, and mice can provide a basis for understanding the evolution of complex features in humans. Findings from such studies suggest that complex traits are likely to be polygenic and that critical mutations underlying these traits will probably lie within elements regulating the expression of genes, such as transcription factors and signal transduction pathways, as well as promoters and enhancers.[29–32]

If the complex traits of humans are due to mutations in noncoding regions, this raises a problem because traditional methods for detecting selection between species, such as the *dN/dS* ratio and the McDonald-Kreitman test,[16] are based on properties of coding regions. These methods will therefore be unable to detect adaptive changes in noncoding regions. For this reason, new analytical techniques are being developed that are applicable to detecting adaptation in noncoding regions.[33–36]

How do we find the sets of genes that underwent positive selection and produced the major morphological and behavioral adaptations of human evolution? How can we ascertain the types of genetic changes that were responsible for these adaptations?

Brain Adaptations

Scanning studies using the *dN/dS* test have compared large sets of genes associated with brain and nervous system functions and compared rates of change between the chimpanzee and human lineages. Initial studies suggested that the rate of nonsynonymous change in these sets of genes was, on average, significantly greater along the human lineage than along the chimpanzee lineage.[37,38] However, one limitation of these studies was that the set of brain and nervous system genes was only compared against a limited set of genes across the genome, making it difficult to determine if the brain genes truly represent an unusually fast evolving set of genes.

In more recent analyses, in which brain genes have been compared against a more comprehensive set of genes across the genome,[39,40] these genes have been found to show no special acceleration along the human lineage. In fact, some studies show that rates may have been even slower along the human lineage than along the chimpanzee lineage. Such an approach highlights the necessity of placing findings in a genome-wide context. Furthermore, it suggests the possibility that adaptations in the human brain and nervous system may have evolved less through selection on favorable amino-acid variants and more through advantageous mutations in elements outside the coding portions of genes. Interestingly, comparisons of gene expression profiles between human and chimpanzee brains find extensive differences between the two species.[41,42] This suggests that changes in gene regulation may indeed have been significant in the evolution of the human brain.

Language

Language is a major human adaptation marking our species' ability to develop complex social relationships and culture. A study of the evolution of the *FOXP2* gene, encoding a forkhead transcription factor, was performed because mutations in this gene are associated with defects in the articulation of speech and cognitive deficiencies in processing language.[43,44] The gene shows very little change when compared across mammalian species, including primates, indicating it is under strong evolutionary constraint. Nevertheless, it shows two nonsynonymous substitutions on the human lineage that occurred after our divergence from chimpanzees, which produce two amino acid changes in exon seven of the gene (although a *dN/dS* test was not significant). This finding leads to the tantalizing hypothesis that the

Box 1. Signatures of Positive Selection in DNA Polymorphism Data

It may be inferred that a particular genomic region has experienced positive selection when that region shows any of the following patterns of polymorphism.

1. Reduced genetic diversity with an excess of rare variants
   - Signature can reveal positive selection since the emergence of modern humans about 200,000 years ago.[90] The signature will appear subsequent to a selective sweep (Fig. 2).
   - Signature can be confounded by rapid population growth.[91]
   - Detected using the test statistic known as Tajima's D.[92]
2. Excess of derived variants at high frequencies
   - Signature will persist for a shorter time than signature 1 because high-frequency variants will quickly rise to 100% frequency (Fig. 1).
   - Signature can be confounded by population subdivision and population bottlenecks.[91]
   - Detected using the Fay and Wu H test.[93]
3. Large differences in the frequencies of DNA variants between populations
   - Signature can indicate positive selection unique to different populations. Informative back to roughly 75,000 years ago.[2]
   - Detected using the F_{ST} coefficient applied between geographic populations.[94]
4. Unexpectedly long haplotypes unbroken by recombination
   - Signature will persist for only relatively recent evolutionary periods on the order of the past several tens of thousands of years.[2]
   - Signature can be confounded by population bottlenecks.[91]
   - Detected using the Long-Range Haplotype test[95] and related approaches (for example, Voight and coworkers[60] and Tang, Thornton, and Stoneking.[61]

amino acid changes in *FOXP2* are linked to the evolutionary innovation of human speech.

However, recent analyses have shown that the Neanderthal *FOXP2* gene also has these two amino acid changes, indicating that the origins of these substitutions likely predates the divergence between Neanderthals and modern humans, which is estimated to have occurred somewhere between 300,000 and 400,000 years ago.[45] This could indicate that Neanderthals were capable of a form of speech similar to that possessed by early modern humans. Such a finding would be consistent with the evidence that Neanderthals had a relatively complex culture, including intricate tool-making capacity (the Mousterian tradition), cultural traditions like ritual burials, and symbolic representation (body adornment and geometric representations).[46] However, before this conclusion can be made, and before we can say for certain whether *FOXP2* partly underlies human speech, more detailed research is needed to work out the precise functional significance of the amino acid substitutions in the *FOXP2* gene.

Interestingly, our increasing knowledge of the genome of Neanderthals,[45,47] the extinct hominid group most closely related to us, will allow us to be more precise in narrowing down the set of traits that uniquely evolved in anatomically modern humans, such as adaptations to climate and diet as well as details of brain function, cognition, and in behavior. Conversely, as the *FOXP2* gene potentially suggests, for some traits it will lead us to enlarge our concept of what it meant to be a Neanderthal.

. . . our increasing knowledge of the genome of Neanderthals, the extinct hominid group most closely related to us, will allow us to be more precise in narrowing down the set of traits that uniquely evolved in anatomically modern humans, such as adaptations to climate and diet as well as details of brain function, cognition, and in behavior.

SEARCHING FOR HUMAN ADAPTATIONS AT THE POPULATION LEVEL

Researchers can search for genetic adaptations that have occurred since the emergence of modern humans (*Homo sapiens sapiens*) around 200,000 years ago by analyzing DNA differences or *polymorphisms* among different human individuals and between different human populations. The various methods used for detecting selection using polymorphism data can be described as *indirect* methods because they explore the effect that positive selection has on neutral variants linked to an advantageous variant (that is, a neutral variant found nearby on the same chromosome). Therefore, these methods have the advantage that they can be applied to both coding and noncoding regions, potentially revealing selection in both regions. On the other hand, these methods have the disadvantage that the actual DNA variant

Figure 2. The effects of a positive selective sweep on the variants within a set of DNA sequences drawn from a population.

under selection is usually difficult to pinpoint since a signature may extend over a broad genomic region (for example, >100,000 base pairs).

Box 1 provides short descriptions of the four main signatures of positive selection in polymorphism data and gives the population genetic statistics commonly used to detect them. These signatures will be localized to the region of the genome surrounding the genetic change that has been positively selected. Several recent papers provide thorough and accessible descriptions of signatures of positive selection (Bamshad and Wooding[1]; Sabeti and coworkers[2]; Nielsen[48]; and Harris and Meyer[49]). One important property of these signatures is that they only endure for limited evolutionary time spans before they dissipate. This is because subsequent to positive selection the processes of mutation, recombination, and genetic drift will restore patterns of polymorphism to neutral levels. As a result, the various signatures can only tell us about selection within specific time frames of human evolution (Fig. 1).

Separating the Effects of Demography From Selection

As described previously, demographic events can potentially confound the signature of positive selection in polymorphism data because past population events can leave patterns in polymorphism data that are closely similar to patterns left by selection (Box 1). This is particularly important since it is likely that human evolutionary history has been demographically complex, with populations having experienced such varied events as bottlenecks, range expansions, population size increases, subdivisions, local extinctions and recolonization events.[50] Therefore, a major challenge has been to verify that positive selection, rather than some possible demographic process, accounts for a particular pattern of polymorphism.

Traditionally, positive selection is inferred for a gene when the observed level and pattern of polymorphism at that gene are significantly different from the pattern expected under an evolutionary model in which selection is not posited. Such an evolutionary model, which serves as the null model in tests of selection, is based on *The Neutral Model of Molecular Evolution,* first described by Motoo Kimura in 1963.[6] Using the neutral model, researchers can make explicit theoretical predictions about levels and patterns of between- and within-species variation, then compare empirical data with the expected pattern. A determination can then be made about whether the observed data are or are not consistent with neutrality and, if not, whether the deviation is in the direction expected for positive selection.

However, an inference of positive selection is not without complication. While deviations from neutrality may suggest positive selection, this interpretation is not always correct. To understand why, it is crucial to realize that the neutral model makes various simplifying assumptions about population dynamics. (Because of these assumptions, the neutral model is often referred to as the neutral-*equilibrium* model.) Significantly, violation of these assumptions can produce deviations from neutrality that can be in the same direction as deviations caused by positive selection. This problem can lead to spurious inferences of selection.

What are some of the assumptions made by the neutral model? One assumption is that populations have remained constant in size over evolutionary time. Other assumptions are that there has been no splitting or combining events among different regional subpopulations or groups. However, as mentioned earlier, it is undoubtedly true that human population history has been extremely complex that and such assumptions have been violated. Therefore, it is widely recognized that great care is needed to disentangle the effects of demography from those of selection.

Controlling for Demography Through Whole-Genome Analyses

At present, our knowledge of patterns of polymorphism across the genome is improving through projects such as the International HapMap

and Perlegen Sciences projects (see Glossary). Using genome-wide data, researchers have taken one of two basic approaches in attempting to deal with the problem of demography. One approach uses the genome pattern of DNA polymorphism to construct a new and human-specific null model against which patterns observed at specific genes can be compared.[51] The other approach is empirical. It orders specific regions across the genome according to certain summary statistics that describe patterns and levels of polymorphism in these regions.[52] For example, summary statistics might include the frequencies of derived variants within a specific region (signature 2, Box 1), or the length of haplotypes in a specific region (signature 4). The approach then identifies regions of the genome that fall within the tails of the genome-wide distribution as potential candidates for PSGs. These genes are referred to as outlier genes. Using this approach, researchers have shown that some previously claimed PSGs in the European population, such as *CCR5,* the chemokine receptor gene,[53] and *ASPM*, the gene associated with cerebral cortex size,[54] do, in fact, fit within the genome-distribution.[2,55,56] Thus, it appears that instead of having been subject to positive selection in recent human populations, the pattern at these genes can be explained either by neutrality or demography,[57] though debate exists. On the other hand, a relatively high *dN/dS* ratio for *ASPM* in comparisons between humans and chimpanzees does indicate that positive selection at this gene occurred earlier along the human lineage.[58]

Whole-Genome Scans for Adaptations at the Population Level

Considering the short time that the needed data have been available, a considerable number of studies have already scanned the genome in search of PSGs.[18,51,52,59–64] Several common themes emerge from these studies. For example, the functional categories of immunity and pathogen defense, olfaction, reproduction and chemosensation are found in many, but not all,[61] as being enriched for genes with signals of positive selection in human populations. Since these categories overlap with the categories found in genome comparisons with the chimpanzee, as described earlier, this finding seems to indicate that selection pressures were continuous over time, at least for genes with these functions. In fact, genes in these categories appear to have undergone positive selection throughout mammalian evolution.[65] New categories to emerge in scans of human populations include genes involved in vitamin and co-enzyme transport and in the metabolism of lipids and carbohydrates.[61] Selection on these genes may point to shifts to new diets as human populations moved into new environments.

Another category specific to population scans includes multiple genes associated with pigmentation that bear strong signatures of selection, especially in European populations. Some of these genes, including *SLC24A5, SLC45A2, MATP*, and *OCA2*, have previously been identified in candidate-gene studies.[66,67] This not only further supports the concept that skin color is a trait influenced by multiple genes, but also gives researchers confidence that scanning methods can identify signatures initially detected in single-gene studies. One newly detected PSG (*DCT*) shows a strong signature of positive selection that is not found in Europeans and is, in fact, restricted to the Chinese.[68] (Certain variants in *DCT* in mice are known to produce lightness of coat coloration.[68]) Interestingly, scans restricted to pigmentation genes have found little or no evidence of shared signatures of selection between Asians and Europeans.[68] This finding challenges the assumption that non-African populations, both Asian and European, owe their lighter skin color to a common genetic basis and instead appears to support the hypothesis that similarities in skin color between the populations evolved through convergence.[68,69]

Another finding in several scanning studies is that non-African populations have greater numbers of PSGs than do African populations.[51,59,70–72] This finding could indicate that positive selection played larger roles in human groups as they colonized new environments after leaving Africa. Conversely, some studies, for example, by Voight and coworkers[60] and Kelley and colleagues,[52] found equivalent or greater numbers of PSGs in Africans as compared to non-Africans. Therefore, it is possible that findings of fewer PSGs in Africans are due to some biasing factor. For example, it may be that signatures of selection are harder to detect in Africans than in non-Africans due to differences in their demographic histories. Carlson and coworkers[59] suggested that because non-Africans apparently experienced a recent bottleneck as they left Africa, their frequency spectrum of polymorphisms is shifted toward intermediate-frequency variants. This shift would make it easier to detect PSGs because such genes show shifts in the frequency spectrum toward low-frequency variants. Another bias may stem from the fact that several studies analyzed African-American populations rather than pure Africans.[73] Admixture may weaken the signature of positive selection. Interestingly, Voight and colleagues,[60] who found greater numbers of PSGs in Africans, based their analysis on the nonadmixed Yoruban samples of the International HapMap project. We can expect to better understand this issue in the future through increased sampling of nonadmixed Africans and increased control for differences in demography.

An additional finding in scanning studies is that signatures can cover large genomic regions (for example, >100,000 bases) within which multiple genes may be found. Identifying which genes have been positively selected is one of the challenges of these studies. Signatures of positive selection may also be encountered in regions in which no genes are known.[17,59,63] The functional basis for these signatures remains unclear. However, it is conceivable that some fraction of these signatures is due to positive selection on regulatory

regions positioned far away from the genes they influence (for example, *trans*-acting regulatory elements).

Currently, only a modest degree of overlap exists among different population-level scanning studies in terms of PSGs identified. Roughly, only about one-third or less of the genes identified in one study were also identified in another study.[17] Nevertheless, genes showing strong signatures in one study frequently are also identified in other studies. Also, genes shown in candidate studies to have experienced positive selection, such as Duffy (vivax malaria), *LCT* (lactase persistence), *HBB*, and MHC genes frequently stand out in genome scans and are clear outliers in an empirical genome distribution.[2] Nevertheless, the considerable differences in results could stem from several factors, such as low power of the particular tests employed,[2] the stochastic nature of the neutral evolutionary process,[17] or the fact that most studies scan the genome using only a single summary statistic. For example, studies might assess only the relative sizes of haplotypes (signature 4 above) or the frequency-spectrum of DNA variants (signature 1 above). It is known that different summary statistics are better at detecting positive selection of different types and at different times. For example, scans for reduced diversity and an excess of rare variants (signature 1) are better able to detect completed selective sweeps that occurred deeper in the evolutionary past. In contrast, scans for long haplotypes (signature 4) are better at detecting incomplete or partial selective sweeps occurring more recently in evolution. Therefore, to some degree, results from different studies are complementary to each other.

LOOKING AHEAD

There is concern about whether methodological approaches relying on multiple recurrent nonsynonymous changes within coding regions are able to detect the genetic basis of phenotypic adaptations. Several examples suffice to show that large effects on phenotype can be produced by one or a few amino acid changes. For example, marked pigmentation differences within the Rock pocket mouse (*Chaetodipus intermedius*)[74] and the Beach mouse (*Peromyscus polionatus*),[75] are produced by one or several amino acid mutations in the *M1CR* gene. For humans, examples include the thick or thin hair phenotype difference between Asians and non-Asians and the dry or wet earwax phenotype difference between Asians and non-Asians, which result from single coding polymorphisms in, respectively, the *EDAR76* and *ABCC11* genes.[76,77]

There is concern about whether methodological approaches relying on multiple recurrent nonsynonymous changes within coding regions are able to detect the genetic basis of phenotypic adaptations. Several examples suffice to show that large effects on phenotype can be produced by one or a few amino acid changes.

These examples support the idea that one or a few amino acid changes can have large effects (Nei's[78] major gene effect model).

There is also the general question of whether regulatory and coding changes are more important as the basis of phenotypic adaptations. In 1975, King and Wilson[22] proposed that regulatory changes were likely to be more important in explaining the adaptive differences between chimpanzees and humans. This issue, though not necessarily specific to humans and chimpanzees, has been resurrected in the genome age and is being debated in the context of new molecular data. Some argue, based on findings in model organisms, for the fundamental importance of *cis*-acting regulatory elements in producing morphological adaptations.[31,32,79] (*Cis*-acting elements are sequence motifs proximal to coding regions that serve as activators or repressors of gene expression). Others argue that while present evidence seems to favor a more important role for coding substitutions, the debate appears rather premature.[80] Even in the midst of increased data, there are many adaptations we still have little or no information about. Therefore, it is not possible to make definitive judgments on this issue. Nevertheless, as we come to understand better how genetic changes interface with phenotypic diversity, it is likely we will find that mutations of both types play significant roles.[78,80]

I have raised two issues regarding phenotypic adaptation: the importance of single or several amino acid substitutions and the importance of regulatory versus coding changes. To further address these issues, new tools and methods will be needed to discover adaptive substitutions or variants in either coding or noncoding parts of our genome. If single or several substitutions have large effects in producing adaptations, then current statistical methods that rely on multiple amino acid substitutions to reject neutrality (i.e., the *dN/dS* test) will be unable to detect these adaptations. For these adaptations, experimental studies that investigate the *functional* effects of particular nonsynonymous substitutions will become increasingly important. Also, if many important adaptations occur in regulatory elements, then increasing focus should be placed on developing methods to detect selection in these regions. In this regard, encouraging signs are emerging. For example, one new approach uses a method similar in design to methods comparing nonsyonymous to synonymous changes. However, instead of focusing on increased *nonsynonymous* change as an indicator of adaptive evolution, this method monitors increased change within regulatory elements of

genes, and compares these to synonymous changes.[81] Interestingly, when the method was applied to humans and chimpanzees, little or no evidence of adaptive change was found.[81] Nevertheless, other studies have identified sets of noncoding regions in the genome that are presumed to have important functions because they are conserved across phylogenetic groups. Significantly, some of these regions appear to have experienced accelerated change along the human lineage,[34,82] with one study finding accelerated regions to be positioned near genes associated with brain function and brain development.[34]

Differences in genomic architecture are also likely to have produced important adaptations. One area being investigated is differences in gene copy number between humans and chimpanzees and between different individuals and human populations. For example, it is estimated that over 6.0% of genes in our genome are not found in the chimpanzee genome.[83] These genes, which likely arose through gene duplication, have been found in gene families related to brain function, development, autoimmune disorders, and immunity.[84] Even within our species, substantial variation in gene copy number exists.[85] These differences might be associated with phenotypic differences among individuals and could underlie population-level adaptations. One example has already been found and appears to represent an adaptation to differences in diet. The gene encoding the salivary enzyme amylase, which hydrolyzes starch, shows increased copy number in populations that have historically relied on high-starch diets, such as European Americans, Japanese, and Hadza. In contrast, populations that have low-starch diets, such as Yakut and Biaki, have fewer copies of the gene.[86]

CODA

The emphasis on adaptive evolution in this review should not be taken to indicate that nonadaptive evolutionary processes such as mutation, genetic drift, and recombination have not played important roles in human evolution. (The process of *genomic* drift, which produces random fluctuations in the numbers of genes within multi-gene families by random duplication and deletion events,[87] can be added to these.) As we gain a better understanding of the details of the human genome and variation within it, it is becoming clear that many aspects can only be explained by nonadaptive evolutionary forces.[87] Humans have a long-term effective population size estimated to be relatively small (on the order of 10,000) and, as discussed earlier, such a small population size would have increased the relative influence that random genetic drift exerted over past genetic variation in our species. Thus, for much of our history, the effectiveness of natural selection (both purifying and positive selection) has been relatively diminished. As explained earlier, this phenomenon helps to explain why there are considerably fewer PSGs in humans than in chimpanzees, which, over evolutionary time, had larger effective populations. It also helps to explain why there exists a large class of low-frequency nonsynonymous variants within human populations, variants that likely contribute to genetic disease.[3,4,88]

Importantly, the effects of nonadaptive evolution have consequences with regard to detecting positive selection in humans. For example, based on principles of the *nearly* neutral theory (a subfeature of the neutral theory worked out by Ohta[89] and others), Hughes[11] has described how any reductions in population size during human evolutionary history (e.g., bottleneck events) since we diverged from chimpanzees would have led to increased fixation of nonsynonymous variants via random genetic drift. (The effects would be greatest for the class of nonsynonymous variants that are weakly deleterious, which, in large populations, would normally be removed by purifying selection.) The effect of this nonadaptive process operating during population size reductions would be to augment the ratio of nonsynonymous change relative to synonymous change. The effect would have serious consequences for tests that compare the ratio of nonsynonymous to synonymous substitutions, particularly the McDonald-Kreitman test. These consequences would occur because adaptive evolution would be inferred from test results even though, in actuality, no adaptive evolution took place. Despite this limitation, it is undoubtedly true that the process of adaptation has occurred in our divergence from chimpanzees. Furthermore, with the field of evolutionary genetics advancing so rapidly on all fronts, we can look forward optimistically to new methods that can better control for the effects of nonadaptive evolution and help reveal actual adaptations.

ACKNOWLEDGMENTS

I thank John Fleagle, Henry Harpending, Sean Myles, and one anonymous reviewer for providing helpful reviews of this manuscript and thank Diogo Meyer for thoughtful discussions.

REFERENCES

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

2 Sabeti PC, Schaffner SF, Fry B, Lohmueller J, Varilly P, Shamovsky O, Palma A, Mikkelsen TS, Altshuler D, Lander ES. 2006. Positive natural selection in the human lineage. Science 312:1614–1620.

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

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

5 Hurst L. 2002. The *Ka/Ks* ratio: diagnosing the form of sequence evolution. Trends Genet 18:486–487.

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

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

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 104:7489–7494.

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

10 Nielsen R, Bustamante C, Clark AG, Glanowski S, Sackton TB, Hubisz MJ, Fledel-Alon A, Tanenbaum DM, Civello D, White TJ, Sninsky JJ, Adams MD, Cargill M. 2005. A scan for positively selected genes in the genomes of humans and chimpanzees. PLoS Biol 3:e170.

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

12 Wagner A. 2007. Rapid detection of positive selection in genes and genomes through variation clusters. Genetics 176:2451–2463.

13 Wyckoff GJ, Wang W, Wu CI. 2000. Rapid evolution of male reproductive genes in the descent of man. Nature 403:304–309.

14 Gasper J, Swanson WJ. 2006. Molecular population genetics of the gene encoding the human fertilization protein *zonadhesin* reveals rapid adaptive evolution. Am J Hum Genet 79:820–830.

15 Hamm D, Mautz BS, Wolfner MF, Aquadro CF, Swanson WJ. 2007. Evidence of amino acid diversity-enhancing selection within humans and among primates at the candidate sperm-receptor gene *PKDREJ*. Am J Hum Genet 81:44–52.

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

17 Biswas S, Akey JM. 2006. Genomic insights into positive selection. Trends Genet 22:437–446.

18 Nielsen R, Williamson S, Kim Y, Hubisz MJ, Clark AG, Bustamante C. 2005. Genomic scans for selective sweeps using SNP data. Genome Res 15:1566–1575.

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

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

21 Swanson WJ, Vacquier VD. 2002. The rapid evolution of reproductive proteins. Nat Rev Genet 3:137–144.

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

23 Gilad Y, Man O, Glusman G. 2005. A comparison of the human and chimpanzee olfactory receptor gene repertoires. Genome Res 15:224–230.

24 Kim U, Wooding S, Ricci D, Jorde LB, Drayna D. 2005. Worldwide haplotype diversity and coding sequence variation at human bitter taste receptor loci. Hum Mutat 26:199–204.

25 Wang X, Thomas SD, Zhang J. 2004. Relaxation of selective constraint and loss of function in the evolution of human bitter taste receptor genes. Hum Mol Genet 13:2671–2678.

26 Fischer A, Gilad Y, Man O, Paabo S. 2005. Evolution of bitter taste receptors in humans and apes. Mol Biol Evol 22:432–436.

27 Wooding S, Kim UK, Bamshad MJ, Larsen J, Jorde LB, Drayna D. 2004. Natural selection and molecular evolution in *PTC*, a bitter-taste receptor gene. Am J Hum Genet 74:637–646.

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

29 Stam LF, Laurie CC. 1996. Molecular dissection of a major gene effect on a quantitative trait: the level of alcohol dehydrogenase expression in *Drosophila melanogaster*. Genetics 144:1559–1564.

30 Gompel N, Prud'homme B, Wittkopp PJ, Kassner VA, Carroll SB. 2005. Chance caught on the wing: *cis*-regulatory evolution and the origin of pigment patterns in *Drosophila*. Nature 433:481–487.

31 Prud'homme B, Gompel N, Carroll SB. 2007. Emerging principles of regulatory evolution. Proc Natl Acad Sci USA 104(Suppl 1):8605–8612.

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

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

34 Prabhakar S, Noonan JP, Paabo S, Rubin EM. 2006. Accelerated evolution of conserved noncoding sequences in humans. Science 314:786.

35 Ponting CP, Lunter G. 2006. Signatures of adaptive evolution within human non-coding sequence. Hum Mol Genet 15(Spec No 2):R170–175.

36 Haygood R, Fedrigo O, Hanson B, Yokoyama KD, Wray GA. 2007. Promoter regions of many neural- and nutrition-related genes have experienced positive selection during human evolution. Nat Genet 39:1140–1144.

37 Dorus S, Vallender EJ, Evans PD, Anderson JR, Gilbert SL, Mahowald M, Wyckoff GJ, Malcom CM, Lahn BT. 2004. Accelerated evolution of nervous system genes in the origin of *Homo sapiens*. Cell 119:1027–1040.

38 Khaitovich P, Hellmann I, Enard W, Nowick K, Leinweber M, Franz H, Weiss G, Lachmann M, Paabo S. 2005. Parallel patterns of evolution in the genomes and transcriptomes of humans and chimpanzees. Science 309:1850–1854.

39 Wang HY, Chien HC, Osada N, Hashimoto K, Sugano S, Gojobori T, Chou CK, Tsai SF, Wu CI, Shen CK. 2007. Rate of evolution in brain-expressed genes in humans and other primates. PLoS Biol 5:e13.

40 Shi P, Bakewell MA, Zhang J. 2006. Did brain-specific genes evolve faster in humans than in chimpanzees? Trends Genet 22:608–613.

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

42 Caceres M, Lachuer J, Zapala MA, Redmond JC, Kudo L, Geschwind DH, Lockhart DJ, Preuss TM, Barlow C. 2003. Elevated gene expression levels distinguish human from non-human primate brains. Proc Natl Acad Sci USA 100:13030–13035.

43 Lai CS, Fisher SE, Hurst JA, Vargha-Khadem F, Monaco AP. 2001. A forkhead-domain gene is mutated in a severe speech and language disorder. Nature 413:519–523.

44 Enard W, Przeworski M, Fisher SE, Lai CS, Wiebe V, Kitano T, Monaco AP, Paabo S. 2002. Molecular evolution of *FOXP2*, a gene involved in speech and language. Nature 418:869–872.

45 Noonan JP, Coop G, Kudaravalli S, Smith D, Krause J, Alessi J, Chen F, Platt D, Paabo S, Pritchard JK, Rubin EM. 2006. Sequencing and analysis of Neanderthal genomic DNA. Science 314:1113–1118.

46 Trinkaus E. 2007. Human evolution: Neandertal gene speaks out. Curr Biol 17:R917–R9.

47 Green RE, Krause J, Ptak SE, Briggs AW, Ronan MT, Simons JF, Du L, Egholm M, Rothberg JM, Paunovic M. 2006. Analysis of one million base pairs of Neanderthal DNA. Nature 444:330–336.

48 Nielsen R. 2005. Molecular signatures of natural selection. Annu Rev Genet 39:197–218.

49 Harris EE, Meyer D. 2006. The molecular signature of selection underlying human adaptations. Am J Phys Anthropol 43(Suppl):89–130.

50 Harding RM, McVean G. 2004. A structured ancestral population for the evolution of modern humans. Curr Opin Genet Dev 14:667–674.

51 Williamson SH, Hubisz MJ, Clark AG, Payseur BA, Bustamante CD, Nielsen R. 2007. Localizing recent adaptive evolution in the human genome. PLoS Genet 3:e90.

52 Kelley JL, Madeoy J, Calhoun JC, Swanson W, Akey JM. 2006. Genomic signatures of positive selection in humans and the limits of outlier approaches. Genome Res 16:980–989.

53 Stephens JC, Reich DE, Goldstein DB, Shin HD, Smith MW, Carrington M, Winkler C, Huttley GA, Allikmets R, Schriml L, and others. 1998. Dating the origin of the *CCR5-Delta32* AIDS-resistance allele by the coalescence of haplotypes. Am J Hum Genet 62:1507–1515.

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

55 Sabeti PC, Walsh E, Schaffner SF, Varilly P, Fry B, Hutcheson HB, Cullen M, Mikkelsen TS, Roy J, Patterson N. 2005. The case for selection at *CCR5-Delta32*. PLoS Biol 3:e378.

56 Yu F, Hill RS, Schaffner SF, Sabeti PC, Wang ET, Mignault AA, Ferland RJ, Moyzis RK, Walsh CA, Reich D. 2007. Comment on "Ongoing adaptive evolution of *ASPM*, a brain size determinant in *Homo sapiens*." Science 316:370.

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

58 Evans PD, Anderson JR, Vallender EJ, Gilbert SL, Malcom CM, Dorus S, Lahn BT. 2004. Adaptive evolution of *ASPM*, a major determinant of cerebral cortical size in humans. Hum Mol Genet 13:489–494.

59 Carlson CS, Thomas DJ, Eberle MA, Swanson JE, Livingston RJ, Rieder MJ, Nickerson DA. 2005. Genomic regions exhibiting positive selection identified from dense genotype data. Genome Res 15:1553–1565.

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

61 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.

62 Kimura R, Fujimoto A, Tokunaga K, Ohashi J. 2007. A practical genome scan for population-specific strong selective sweeps that have reached fixation. PLoS ONE 2:e286.

63 Wang ET, Kodama G, Baldi P, Moyzis RK. 2006. Global landscape of recent inferred Darwinian selection for *Homo sapiens*. Proc Natl Acad Sci USA 103:135–140.

64 Zhang C, Bailey DK, Awad T, Liu G, Xing G, Cao M, Valmeekam V, Retief J, Matsuzaki H, Taub M, Seielstad M, Kennedy GC. 2006. A whole genome long-range haplotype (WGLRH) test for

detecting imprints of positive selection in human populations. Bioinformatics 22:2122–2128.

65 Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P, Agarwala R, Ainscough R, Alexandersson M, An P and others. 2002. Initial sequencing and comparative analysis of the mouse genome. Nature 420:520–562.

66 Lamason RL, Mohideen MA, Mest JR, Wong AC, Norton HL, Aros MC, Jurynec MJ, Mao X, Humphreville VR, Humbert JE, Sinha S, Moore JL, Jagadeeswaran P, Zhao W, Ning G, Makalowska I, McKeigue PM, O'Donnell D, Kittles R, Parra EJ, Mangini NJ, Grunwald DJ, Shriver MD, Canfield VA, Cheng KC. 2005. *SLC24A5*, a putative cation exchanger, affects pigmentation in zebrafish and humans. Science 310:1782–1786.

67 Soejima M, Tachida H, Ishida T, Sano A, Koda Y. 2006. Evidence for recent positive selection at the human *AIM1* locus in a European population. Mol Biol Evol 23:179–188.

68 Myles S, Somel M, Tang K, Kelso J, Stoneking M. 2007. Identifying genes underlying skin pigmentation differences among human populations. Hum Genet 120:613–621.

69 Norton HL, Kittles RA, Parra E, McKeigue P, Mao X, Cheng K, Canfield VA, Bradley DG, McEvoy B, Shriver MD. 2007. Genetic evidence for the convergent evolution of light skin in Europeans and East Asians. Mol Biol Evol 24:710–722.

70 Akey JM, Eberle MA, Rieder MJ, Carlson CS, Shriver MD, Nickerson DA, Kruglyak L. 2004. Population history and natural selection shape patterns of genetic variation in 132 genes. PLoS Biol 2:e286.

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

72 Storz JF, Payseur BA, Nachman MW. 2004. Genome scans of DNA variability in humans reveal evidence for selective sweeps outside of Africa. Mol Biol Evol 21:1800–1811.

73 Ronald J, Akey JM. 2005. Genome-wide scans for loci under selection in humans. Hum Genomics 2:113–125.

74 Nachman MW, Hoekstra HE, D'Agostino SL. 2003. The genetic basis of adaptive melanism in pocket mice. Proc Natl Acad Sci USA 100:5268–5273.

75 Hoekstra HE, Hirschmann RJ, Bundey RA, Insel PA, Crossland JP. 2006. A single amino acid mutation contributes to adaptive beach mouse color pattern. Science 313:101–104.

76 Fujimoto A, Kimura R, Ohashi J, Omi K, Yuliwulandari R, Batubara L, Mustofa MS, Samakkarn U, Settheetham-Ishida W, Ishida T, Morishita Y, Furusawa T, Nakazawa M, Ohtsuka R, Tokunaga K. 2008. A scan for genetic determinants of human hair morphology: *EDAR* is associated with Asian hair thickness. Hum Mol Genet 17:835–843.

77 Yoshiura K, Kinoshita A, Ishida T, Ninokata A, Ishikawa T, Kaname T, Bannai M, Tokunaga K, Sonoda S, Komaki R and others. 2006. A SNP in the *ABCC11* gene is the determinant of human earwax type. Nat Genet 38:324–330.

78 Nei M. 2007. The new mutation theory of phenotypic evolution. Proc Natl Acad Sci USA 104:12235–12242.

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

80 Hoekstra HE, Coyne JA. 2007. The locus of evolution: evo devo and the genetics of adaptation. Evolution Int J Org Evolution 61:995–1016.

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

82 Bird CP, Stranger BE, Liu M, Thomas DJ, Ingle CE, Beazley C, Miller W, Hurles ME, Dermitzakis ET. 2007. Fast-evolving noncoding sequences in the human genome. Genome Biol 8:R118.

83 Demuth JP, De Bie T, Stajich JE, Cristianini N, Hahn MW. 2006. The evolution of mammalian gene families. PLoS ONE 1:e85.

84 Hahn MW, Demuth JP, Han SG. 2007. Accelerated rate of gene gain and loss in primates. Genetics 177:1941–1949.

85 Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P, Maner S, Massa H, Walker M, Chi M, Navin N, Lvcito R, Healy J, Hicks J, Ye K, Reiner A, Gilliam TC, Trask B, Patterson N, Zetterberg A, Wigler M. 2004. Large-scale copy number polymorphism in the human genome. Science 305:525–528.

86 Perry GH, Dominy NJ, Claw KG, Lee AS, Fiegler H, Redon R, Werner J, Villanea FA, Mountain JL, Misra R, Carter N, Lee C, Stone AC. 2007. Diet and the evolution of human amylase gene copy number variation. Nat Genet 39(10):1256–1260.

87 Lynch M. 2007. The orgins of genome architecture. Sunderland, MA: Sinauer Associates.

88 Gojobori J, Tang H, Akey JM, Wu CI. 2007. Adaptive evolution in humans revealed by the negative correlation between the polymorphism and fixation phases of evolution. Proc Natl Acad Sci USA 104:3907–3912.

89 Ohta T. 1973. Slightly deleterious mutant substitutions in evolution. Nature 246:96–98.

90 Tishkoff SA, Verrelli BC. 2003. Patterns of human genetic diversity: implications for human evolutionary history and disease. Annu Rev Genomics Hum Genet 4:293–340.

91 Przeworski M. 2002. The signature of positive selection at randomly chosen loci. Genetics 160:1179–1189.

92 Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585–595.

93 Fay JC, Wu CI. 2000. Hitchhiking under positive Darwinian selection. Genetics 155:1405–1413.

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

95 Sabeti PC, Reich DE, Higgins JM, Levine HZ, Richter DJ, Schaffner SF, Gabriel SB, Platko JV, Patterson NJ, McDonald GJ, Ackerman HC, Campbell SJ, Altshuler D, Cooper R, Kwiat Kowski D, Ward R, Lander ES. 2002. Detecting recent positive selection in the human genome from haplotype structure. Nature 419:832–837.

96 The International HapMap Consortium. 2007. A second generation human haplotype map of over 3.1 million SNPs. Nature 449:851–861.

97 Hinds DA, Stuve LL, Nilsen GB, Halperin E, Eskin E, Ballinger DG, Frazer KA, Cox DR. 2005. Whole-genome patterns of common DNA variation in three human populations. Science 307:1072–1079.

98 Chen FC, Li WH. 2001. Genomic divergences between humans and other hominoids and the effective population size of the common ancestor of humans and chimpanzees. Am J Hum Genet 68:444–456.