Gene Evolution and Human Adaptation

Eugene E Harris, *Queensborough Community College, City University of New York, Bayside, New York, USA and Profesor Visitante, Departmento de Genética e Biologia Evolutiva, Universidade de São Paulo, Brasil*

Regions of the human genome that have been subject to past positive selection contain patterns of genetic variation that are markedly different in specific ways from regions that have not experienced positive selection. By uncovering these so-called signatures of positive selection in the genome we can discover the unique ways in which humans have evolved.

Advanced article

Article Contents

Online posting date: 14th March 2008

Introduction

In the course of human evolution, genetic adaptation has occurred at two levels. Species-wide adaptations have occurred along the human lineage as humans diverged from their shared common ancestor with chimpanzees and prior to the differentiation of geographical human populations. Population-level adaptations have occurred during the process of differentiation among human populations as they adapted to different geographic environments. Genetic adaptations are produced through the evolutionary process of positive selection. In this process, an advantageous deoxyribonucleic acid (DNA) variant is driven to higher frequencies because of the increased reproductive fitness it confers to individuals bearing the variant over individuals who do not. We can learn about adaptation during human evolution by studying the effects that positive selection can leave in localized regions of the genome that experienced positive selection. Human lineage adaptations can be studied by comparing DNA differences between humans and chimpanzees (substitutions) and identifying regions having signatures of positive selection. Population-level adaptations can be studied by comparing DNA differences between and among individuals from different populations (polymorphisms) and identifying regions having signatures of positive selection.

Signatures of positive selection can be identified when patterns of DNA differences (either between species or between individuals) deviate in specific ways from patterns expected under an evolutionary history in which there was no positive selection (i.e. the null hypothesis). The null hypothesis is traditionally based on the standard neutral model stemming from Kimura's *Neutral Theory of Molecular Evolution* that proposes that most observed genetic differences between and within species have no effect on fitness (i.e. are selectively neutral). The model allows explicit predictions to be made concerning patterns of DNA variation against which actual observed data can be compared. However, as implemented in most tests, the model makes a number of simplifying assumptions about past populations including constancy in size, random mating and a single panmictic population (no population structure). Unfortunately, violations of some of these assumptions can produce signatures that look almost identical to signatures of positive selection, thereby making it difficult to determine whether a genomic region has actually experienced positive selection. In particular, the effects of demographic events such as bottlenecks, size-increases and population subdivisions can confound analyses of genes when comparisons are made against the neutral model.

ELS subject area: Evolution and Diversity of Life

How to cite:
Harris, Eugene E (March 2008) Gene Evolution and Human Adaptation. In: Encyclopedia of Life Sciences (ELS). John Wiley & Sons, Ltd: Chichester.
DOI: 10.1002/9780470015902.a0020773

Positive Selection and Species-wide Adaptations

As mentioned earlier, we can learn about positive selection along the human lineage ($\sim$6–7 million years) by comparing DNA differences (substitutions) between human and chimpanzee genomes (see **Figure 1**). Substitutions in protein-coding genes have traditionally been the target of such studies. About three-fourth of DNA mutations that arise in a gene in a population produce changes at the amino acid level that will likely negatively affect the proteins' function. Such mutations are called nonsynonymous mutations, and they are usually removed soon after they appear in the population by the process known as purifying selection. However, about one-quarter of mutations occurring in genes, do not result in amino acid changes (called synonymous mutations) and can potentially become fixed in the population by the stochastic process of genetic drift. These so-called neutral substitutions are usually assumed

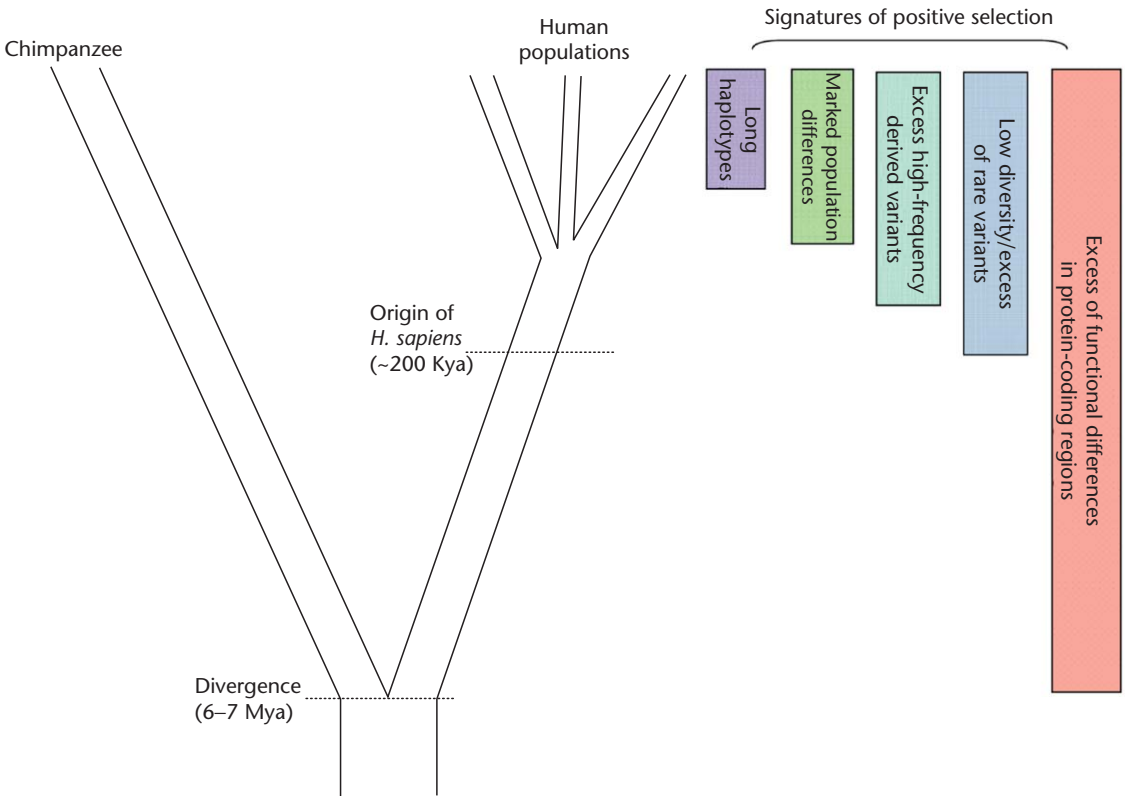

Figure 1 From bottom to top, this illustration shows the evolutionary divergence of chimpanzees and humans from their common ancestor. More recently along the human lineage is represented the origin of anatomically morden *Homo sapiens* and its subsequent differentiation into different regional populations. (Time scales along the lineages leading to humans and chimpanzees have been compressed). Human species-wide adaptations have evolved along the human lineage from the initial time of divergence (~ 6–7 million years ago) up until the time of the differentiation of modern human populations. However, population-level adaptations have evolved as human populations differentiated and adapted to different geographic regions. On the right are shown five signatures in the patterns of DNA variation that can indicate that a DNA region has experienced positive selection. These are arranged according to the time frame in human evolution in which they can reveal the process of positive selection.

under neutral theory to make up the majority of DNA differences between species. For example, in whole genome comparisons between chimpanzees and humans, there are approximately 1.02% synonymous differences but only about 0.242% nonsynonymous differences (Bustamante *et al*., 2005). Positive selection, however, can increase the number of nonsynonymous substitutions along a lineage above the number of synonymous substitutions when these substitutions have been advantageous. This can be detected through the d_N/d_S test (also referred to as the K_a/K_s test), where d_N is the number of nonsynonymous substitutions per site, and d_S is the number of synonymous substitutions per site. Another test, known as the McDonald-Kreitman (MK) test, compares the between-species ratio of the number of substitutions per gene at each class of site (denoted D_n/D_s) to the within-species ratio of the number of polymorphisms per gene at each of these classes (denoted P_n/P_s). In this comparison, an increase in D_n/D_s compared to P_n/P_s can be another signature that a gene has been positively selected during divergence between species. Since these methods are based directly on the functional differences themselves, it is often assumed that they can help reveal the actual site(s) that experienced positive selection.

The d_N/d_S ratio is believed to be robust to the confounding effects of demography. This is because both nonsynonymous and synonymous substitutions (derived from the same gene) are assumed to have been equally affected by the same demographic events, thus controlling for demography. The MK test, however, can potentially be confounded by demographic effects (Eyre-Walker, 2002) because it cannot distinguish between positive selection and factors that might cause purifying selection to become less efficient. For example, the *nearly neutral model* (a corollary of the *Neutral Model*; see Ohta, 1973) predicts that during periods of small population-size, purifying selection will become less efficient compared to genetic drift. Under these conditions, a significant fraction of slightly deleterious variants will drift to fixation. Thus far, few studies of positive selection along the human lineage have controlled for the effects of demographic history, even though both bottleneck events as well as population-size increases are believed to have occurred.

The between species d_N/d_S test is conservative because a large number of nonsynonymous substitutions are needed before the test yields significant results. For this reason, the method is best suited to detect positive selection when it has occurred recurrently over extended time periods at a gene,

producing numerous nonsynonymous substitutions. Furthermore, the excess of nonsynonymous substitutions may be restricted to relatively small domains where positive selection has been localized. The remainder of the protein, however, may show very few if any nonsynonymous substitutions due to the effects of strong purifying selection in these regions. Therefore, the signature of positive selection in a protein may be hidden because it has been overwhelmed by the pattern in the gene as a whole. Thus, the test appears to be most powerful when it can be applied to regions of a gene that are hypothesized to be under positive selection on the basis of a priori information regarding the protein's function. Following are several examples of genes for which the d_N/d_S method has been successful at revealing positive selection along the human lineage. In major histocompatibility genes (*MHC*), and several genes encoding sperm-specific proteins (e.g. *PRM1*, *PRM2*) large excesses of nonsynonymous substitutions over synonymous substitutions have led to well-supported inferences of positive selection (Hughes and Nei, 1988; Wykoff *et al*., 2000). In *MHC* genes, the excess is found in the binding domains that determine antigen specificity. Further examples include genes encoding the candidate sperm-receptor protein (*PKDREJ*) and the zonadhesion protein (*ZAN*) in which excesses of nonsynonymous substitutions are concentrated in functional domains where male/female gametes interact (Hamm *et al*., 2007; Gasper and Swanson, 2007).

The above examples of positively selected genes (PSGs) were found in 'candidate-gene studies' in which genes are studied for evidence of positive selection based on their known (or presumed) functional or phenotypic association. However, with the increasing availability of genome-wide data, studies (known as 'genome-scanning studies') can be designed to discover which genes within the protein-coding portion of the genome have experienced positive selection. Such studies have typically applied the d_N/d_S or MK methods to comparisons of extensive sets of orthologous protein-coding genes between humans and chimpanzees. Usually, these analyses have turned up hundreds of genes with signatures suggestive of positive selection, though this number decreases considerably after statistical corrections to determine significance are applied.

There is considerable overlap among studies in the biological categories that appear enriched for putative PSGs including genes involved in immunity and pathogen resistance, sensory reception (olfactory and auditory), reproduction (gametogenesis and fertilization), apoptosis and in nucleotide metabolism and repair. A 'transcription-factor' category was also enriched for PSGs in some studies (Bustamante *et al*., 2005; Mikkelsen *et al*., 2005) including homeotic, forkhead and other genes involved in early development, possibly indicating that changes in gene regulation have played important roles in human evolution. Several studies found significantly more PSGs along the chimpanzee lineage compared to the human lineage (Bakewell *et al*., 2007; Arbiza *et al*., 2006). This finding is probably related to the larger long-term effective population size estimated for chimpanzees compared to humans. Neutral theory predicts that positive selection will be more effective at fixing advantageous variants in larger populations than in smaller populations because the effects of random genetic drift are reduced in large populations. For the same reason, purifying selection is also expected to be more effective in larger populations, but in this case at removing mutations having slightly deleterious effects. The significantly lower d_N/d_S ratio in chimpanzees (.245) compared to humans (.259) found in genome-wide studies (Bakewell *et al*., 2007) appears consistent with this prediction as it indicates that a greater proportion of nonsynonymous mutations have been removed in chimpanzee genes than in human genes over evolutionary time. A further finding from scanning studies is that chimpanzee and human PSGs are mostly nonoverlapping and are distributed differently amongst biological categories, presumably pointing to unique sets of adaptations in the two species.

It is of little surprise that genome scans have detected a concentration of PSGs within the categories of reproduction, immune defence and apoptosis. The proteins encoded by these genes (similar to the cases mentioned earlier) are typically involved in coevolutionary arm's-race-type interactions (e.g. against pathogens, in sperm competition, in tumour suppression). Such interactions are characterized by recurrent events of positive selection that can produce multiple nonsynonymous substitutions in the functional domains where the interaction is localized. As described earlier, methods based on detecting excesses of amino acid changes (the d_N/d_S and MK tests) are best able to detect positive selection in which repeated amino acid substitutions occurred. However, it has been pointed out that this type of positive selection is rare, and that most phenotypic adaptations have not been produced through recurrent selection (see Hughes, 2007). Consequently, our current methods (that compare nonsynonymous to synonymous changes) will be of limited utility to us in identifying PSGs that underlie many important species-wide adaptations, and new approaches will be needed. Another limitation (particularly of the d_N/d_S test) is that relaxation of purifying selection on a gene can also produce elevated nonsynonymous substitutions, creating a signature similar to that left by positive selection. For example, the unusual finding of extensive positive selection in human olfactory receptor genes could instead be explained (at least partially) by relaxation of purifying selection. Studies of the olfactory receptor gene repertoire have revealed that many gene copies have become pseudogenes (see Gilad *et al*., 2005). Therefore, nonsynonymous variants could have become fixed through the process of genetic drift and not via positive selection.

Major Human Morphological and Behavioural Adaptations

How do we uncover the signatures of positive selection underlying major human morphological and behavioural

adaptations? Initial scanning studies on sets of brain-associated genes (Dorus *et al*., 2004; Khaitovitch *et al*., 2005) suggested that the rate of nonsynonymous substitution in these genes is accelerated along the human lineage relative to chimpanzees. However, more comprehensive analyses in which comparisons were made against average rates across the genome (Wang, 2007; Nielsen *et al*., 2005; Shi *et al*., 2006) showed that human brain-associated genes are not accelerated in their rate. In fact, rates appear to be slower in humans relative to chimpanzees. Besides underscoring the importance of placing findings in a genome-wide context, this finding suggests the possibility that morphological changes in the human brain evolved less through selection on favourable amino acid variants in proteins than through favourable mutations in regulatory regions that control the expression of genes. Observations of extensive gene-expression differences in human brains compared to chimpanzee brains (Enard *et al*., 2002; Caceres *et al*., 2003) may point to the importance of adaptation through gene regulation.

A candidate-gene study of the *FOXP2* gene (a forkhead transcription factor), associated with articulation of speech, shows an elevated rate of nonsynonymous substitutions along the human lineage, though the gene otherwise appears to have been under tight constraint in mammalian evolution (Enard *et al*., 2002). Such a signature could arise if the gene was under relaxed selective constraints. However, this does not appear likely because under relaxation of constraints an excess of within-population amino acid polymorphism would be expected, but no amino acid variants were in fact detected. It is tantalizing to surmise that the substitutions in *FOXP2* (estimated to have arisen within the past 100 000 years or so) might in part underlie the evolutionary acquisition of human speech. Yet before such a conclusion can be made, further research will be needed to identify if the amino acid differences between humans and chimpanzees can be functionally linked to differences in, for example, orofacial movements related to speech.

As with the human brain, speech and language, many unique human adaptations such as bipedalism, increased hand/finger dexterity, reduced prognathism and reduced canine-size involve a complex set of functionally integrated morphological features. Studies of complex traits in model organisms have usually indicated that such traits have a polygenic basis and that regulatory processes are often critical in their development (e.g. transcriptions-factors and signal-transduction pathways; see Carroll, 2003). Where the critical mutations for these traits have been identified, they are usually located in regulatory or other noncoding regions. For these reasons, it is likely to be a slow and complex enterprise to decipher the genetic basis of the major human morphological adaptations. Nevertheless, encouraging signs are emerging through the application of new methods that enable detection of functionally important noncoding regions. Initial results suggest that such regions are widespread in the genome. Moreover, some of these regions have undergone accelerated evolution along the human lineage (Prabhaker *et al*., 2006; Bird *et al*., 2007) and may therefore underlie human adaptations.

Selection and Population-level Adaptations

Positive selection occurring since the origins of anatomically modern *Homo sapiens* around 200 000 years ago can be studied by analysing patterns of DNA differences between and among populations (polymorphism). Earlier, I have described *direct* methods of detecting selection that compare ratios of functional to nonfunctional substitutions. In contrast, most methods based on polymorphism are *indirect* because they explore the effects that positive selection has on neutral sites linked to an advantageous mutation. These methods can therefore potentially detect selection in noncoding regions of the genome. A disadvantage of indirect methods is that the actual site (or sites) under selection is (are) usually difficult to determine. It is important to understand that the effects of positive selection on patterns of polymorphism only endure for limited spans of evolutionary time before they weaken and eventually disappear. The reason is that subsequent to the positive selection event, neutral genetic processes like mutation and genetic drift will restore patterns to neutral expectations. Consequently, signatures based on different aspects of polymorphism data can tell us about positive selection within particular windows of time in human evolution (**Figure 1**). The four main signatures of positive selection produced in patterns of DNA polymorphism are described later and listed in relation to the time depth within which they can reveal positive selection (from oldest to more recent).

Demography is extremely relevant to detecting positive selection using polymorphism data. Multilocus analyses have indicated that *H. sapiens* has had a complex demographic history since its origin, probably including population-bottlenecks, size-increases, range-expansions, fragmentations, as well as local extinctions and re-colonization events. The challenge therefore is to identify the signature of selection against a background pattern of genetic variation shaped by a complex demographic history. An important feature distinguishing the two processes, and that can be used to disentangle their effects, is that demographic events influence variation over the entire genome, whereas positive selection has effects localized to specific loci.

Signatures of Positive Selection in Patterns of Polymorphism

Reduction in genetic diversity and excess of rare variants

Positive selection drives advantageous variants to higher frequencies within populations. Simultaneously, neutral

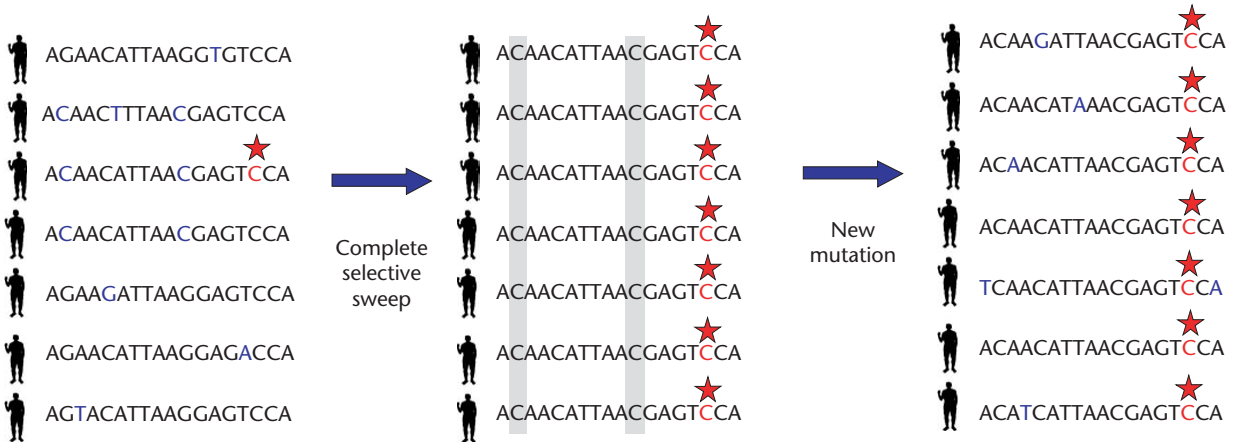

Figure 2 Proceeding from left to right, the figure illustrates the evolutionary process known as a positive selective sweep. Each of the seven DNA sequences represents a copy of DNA carried by a human individual in the population. In the left hand panel, the seven DNA copies show a pattern of variants (coloured in blue) that might be expected under neutrality, some variants are present in only a single individual in the sample (i.e. are rare variants) and some are present at higher frequencies. The third sequence from the top has recently acquired an advantageous mutation (coloured in red and denoted by a star). Under positive selection, this DNA copy will be driven to higher and higher frequencies in the population while all other copies become lost, so that in the middle panel, all DNA copies drawn from the population possess the advantageous mutation. According to the process known as 'hitchhiking', all sequence copies have also inherited the two linked neutral variants (shaded) that were present on the originally selected gene copy. The process has also 'swept' (or removed) all other variants originally present in the population, decreasing genetic diversity in the population in the region of DNA bearing the advantageous mutation. The panel on the right represents DNA copies drawn from the population just after the sweep while the population is in recovery phase. The neutral mutation rate has introduced new variants to the population (coloured in blue). Since the population has not reached equilibrium, all mutations still exist at low frequencies (i.e. they are rare variants).

variants linked to the advantageous variant are also driven to high frequency in a process called 'hitchhiking'. As the advantageous variant (and linked variants) replace other gene copies in the population, the overall genetic variation in the region surrounding the selected mutation is reduced in a process known as a 'selective sweep' (Maynard Smith and Haigh, 1974). A complete sweep will eliminate all variation from the region (**Figure 2**). Immediately after the sweep, however, new variants are introduced by mutation and will appear in the data as low-frequency mutations (rare variants). Thus, positive selection produces a signature (in the region near the selected mutation) of unusually low genetic variation and an excess of rare variants compared to regions unaffected by selection. As time continues, neutral processes will begin to restore levels of genetic variation, as well as the frequency of variants, to neutral expectations. This signature can reveal positive selection over a time frame encompassing the origin of anatomically modern *H. sapiens* until the present. Demographic events like rapid population expansion, however, can confound interpretation of this signature because expansions are also known to produce an excess of rare variants.

High-frequency derived variants

At polymorphic sites in humans, there are usually only two variants. One is the ancestral variant because it was inherited from the human–chimpanzee ancestor, while the other variant is derived because it is due to a mutation event in human evolutionary history. Because derived variants evolved more recently, they are expected to have lower frequencies compared to ancestral variants. However, positive selection may drive derived variants to unusually high frequencies by 'hitchhiking' when they are linked to the beneficial variant. Therefore, an excess of high-frequency derived variants in a genomic region can be a signature of recent positive selection. This signature persists, however, for a relatively shorter time (than the signature described earlier) because high-frequency derived variants have a high probability of rapidly going to fixation. Demographic events like population subdivision and bottlenecks are also known to produce high-frequency derived variants (see Przerworski, 2002) and so can confound interpretation of this signature.

Marked frequency differences between populations

When positive selection has occurred in one population but not another (perhaps because selective pressures were geographically restricted), the frequency of the beneficial variant is expected to be markedly higher in the population in which positive selection took place compared to populations in which it did not take place. This frequency difference between populations will also occur at neutral variants linked to the beneficial variant. Thus, when a genomic region contains variants showing marked frequency differences between populations, this can be a signature of positive selection. The time frame over which this signature is useful is restricted to the time since human populations first began to differentiate (roughly 100 000 to 50 000 years ago). Demographic events such as population subdivision and population-specific bottlenecks can also produce

marked frequency differences between populations, and therefore can confound interpretations of this signature.

Long haplotypes

DNA variants in strong association with one another on a single chromosome form what is called a haplotype. The process of recombination (due to crossovers of different haplotypes during meiosis), however, breaks down associations between variants over time according to the recombination rate. Under positive selection, however, an advantageous variant and its linked neutral variants can be driven to high frequency at a rate faster than the recombination rate can break down their association. This gives rise to an unusually long haplotype. Therefore, when a haplotype is found at high frequency in a population, and also shows unusually long-range association between its variants, this can be a signature of positive selection. Since recombination will quickly break down a long haplotype, the signature will be most useful for detecting recent positive selection, for example within the last several tens of thousands of years (see Przerworski, 2002). Furthermore, the signature can usually only be detected if the beneficial variant has not gone to fixation. This signature can be confounded by population bottlenecks, which also can produce unusually long haplotypes.

Candidate-gene Studies Using Polymorphism Data

The most successful studies of single candidate genes are those framed in light of an a priori hypothesis about the type and direction of selection. The studies of the genes encoding the Duffy red blood cell antigen (*FY*) and the enzyme lactase (*LCT*) are good examples (see Hamblin *et al*., 2002; Bersaglieri *et al*., 2004). Both genes showed variants with marked frequency differences between populations: in the Duffy gene, the *FY*O* variant is nearly fixed in sub-Saharan African populations, but is virtually absent elsewhere; at *LCT*, a variant in tight association with the lactase persistence trait was found at extremely high frequencies in Northern Europe but at lower frequencies (or absent) elsewhere. Both genes are clearly linked to phenotypes, and the underlying causal variants in each case are known: *FY*O* homozygotes lack expression of the antigen on red blood cells due to a mutation in the GATA promoter (GATA indicating the DNA recognition motif); lactase persistence is caused by a mutation in a transcription-site located kilobases upstream of *LCT*. Moreover, it is easy to see how these traits may have been selectively favoured: absence of the Duffy antigen prevents the invasion of *Plasmodium vivax* (thus providing resistance to vivax malaria); lactase persistence allows individuals to digest milk-sugar (lactase) into adulthood. Analyses of polymorphism have revealed signatures consistent with positive selection and these signatures are uniquely found in the same populations for which selection was initially hypothesized, providing strong additional support for the hypothesis. The Duffy gene in sub-Saharan Africans shows markedly reduced variation in and around the gene (i.e. $\sim$3-fold less than in Italians), displays marked population differentiation of linked neutral variants, and contains an excess of high-frequency derived variants. The *LCT* gene in Northern European populations shows the presence at very high frequencies ($\sim$80%) of an extremely long haplotype ($\sim$1 Mb in size) bearing the lactase persistence causal variant.

One of the most significant problems for studies of single candidate genes has been in separating the confounding effects of demography from the signature of positive selection. Comparing levels of variation at a gene against the standard neutral model can be problematic. Even if a gene is shown to deviate from the model it is still difficult to decide whether this is due to positive selection or if a demographic explanation is more likely. Alternatively, a gene found to be consistent with the neutral model could still have been positively selected. Thus, both false-positive and false-negative results are possible. The use of computer simulations to model the effects of demographic events (e.g. bottlenecks and population expansion) has allowed some researchers to construct new null models, but such methods cannot reflect the real complexity of human demographic history.

Our increasing knowledge of genome-wide patterns of polymorphism, through such efforts as the International HapMap and Perlegen BioSciences projects (see Web-links below), represents a potential way to control for the confounding effects of demography. Genome-wide data has been used in two basic ways in evaluating genes for evidence of positive selection. One approach constructs a new (human-specific) null model based on genome-wide patterns of polymorphism against which patterns found in specific regions can be compared. Another approach is purely empirical and orders genomic regions on the basis of summary statistics (e.g. the frequencies of variants and/or diversity levels in particular regions), and identifies regions falling within the tails of the genome-wide distribution as potential PSGs (i.e. so-called 'outlier-genes'). The usefulness of these approaches in helping to distinguish actual cases of positive selection can be seen by the fact that some previously claimed PSGs (e.g. *CCR5* and *ASPM*, a chemokine-receptor gene and a gene associated with cerebral cortex-size, respectively) do not appear unusual when compared with a genome-wide distribution (see Sabeti *et al*., 2005, 2006; Yu *et al*., 2007). Alternatively, it also is possible that some genes that appeared consistent with the neutral model may stand out when compared to a null model based on genome-wide data.

Genome-scanning Studies Using Polymorphism Data

With the availability of genome-wide polymorphism data, methods for detecting selection can be applied across

the genome in a discovery-driven effort to identify new potential candidate PSGs (e.g. Voight *et al*., 2006; Williamson *et al*., 2007). A number of common themes are emerging from studies thus far. Like in the between-species scans, scans based on polymorphism are also identifying categories such as immunity and pathogen defence, reproduction, olfactory and chemosensory as enriched for PSGs. This presumably indicates continuity over evolutionary time in selection pressures on genes in these categories. Categories specific to polymorphism scans, include metabolism-associated genes (e.g. of lipids and carbohydrates), and genes involved in coenzyme and vitamin transport. Such categories of PSGs might indicate adaptation to new diets. Genes related to pigmentation also show strong signatures of positive selection (particularly in European populations), and some have been previously identified in candidate studies (e.g. *SLC24A5* and *OCA2*). Several scanning studies (but not all) have found that non-African populations show more PSGs than African-derived populations. This might indicate that adaptation played a greater role in non-Africans as they colonized new environments outside Africa. However, such a discrepancy may be merely an artifact since there are reasons to believe that positive selection may be more difficult to detect in African populations. For example, several studies based their analyses on African-American samples. However, admixture between Africans and Europeans in these samples may have had the effect of weakening the signature of positive selection. Additionally, demographic differences between the two populations (a bottleneck inferred for non-Africans and larger long-term population size inferred for Africans) may make positive selection more difficult to detect in Africans. Resolution of this question will require studies of diverse African populations unaffected by admixture, and with better controls for demographic differences between the two populations. A further finding, common to many scanning studies, is that signatures of positive selection are not uncommon in regions in which there are no known genes. At present, we know little about the basis of these signatures, but they could potentially indicate regions containing important regulatory elements.

Until now there has been only moderate degree of overlap between scanning studies, with about one-third or less of PSGs identified in one study also being found in another study. Nevertheless, genes with extreme signatures in one study are commonly found in others. The considerable differences in results between studies could be due to a high rate of discovery of false-positives common to most studies. To some extent this is expected because of the stochastic nature of the neutral evolutionary process. However, discrepancies may also be due to the different methods used in separate studies, in which usually only a single summary statistic is applied to scan the genome for positive selection. Different methods are known to preferentially detect different types of positive selection and to preferentially detect selection that occurred within different time windows in the past. For example, scans for long haplotypes preferentially detect partial selective sweeps (where the advantageous variant has not been fixed) that occurred relatively recently in the past. In contrast, scans for reduced diversity (and excesses of rare alleles) preferentially detect completed sweeps that occurred more distantly in the past. To some extent, therefore, studies are complementary to each other and differences are to be expected. A considerable limitation of scan-studies is that signatures cover rather broad genomic regions (e.g. hundreds of thousands of base pairs) within which a number of genes are sometimes found. Therefore, a major challenge will lie in accurately identifying which gene was under positive selection. Furthermore, there is still the crucial task of identifying an adaptively favourable variant, and linking it to its function or phenotypic expression. **See also**: Molecular Evolution: Neutral Theory; Neutrality and Selection in Molecular Evolution: Statistical Tests; Selection Operating on Protein-Coding Regions in the Human Genome; Selective and Structural Constraints

Acknowledgements

I thank City University of New York (through a PSC CUNY Award) for providing funds that helped support me writing this article.

References

Arbiza L, Dopazo J and Dopazo H (2006) Positive selection, relaxation, and acceleration in the evolution of the human and chimp genome. *PLoS Computational Biology* **2**(4): e38.

Bakewell MA, Shi P and Zhang J (2007) More genes underwent positive selection in chimpanzee evolution than in human evolution. *Proceedings of the National Academy of the Sciences of USA* **104**(18): 7489–7494.

Bersaglieri T, Sabeti PC, Patterson N *et al*. (2004) Genetic signatures of strong recent positive selection at the lactase gene. *American Journal of Human Genetics* **74**: 1111–1120.

Bird CP, Stranger BE, Liu M *et al*. (2007) Fast-evolving non-coding sequences in the human genome. *Genome Biology* **8**: R118.

Bustamante CD, Fledel-Alon A, Williamson S *et al*. (2005) Natural selection on protein coding genes in the human genome. *Nature Genetics* **437**: 1153–1157.

Caceres M, Lachuer J, Zapala MA *et al*. (2003) Elevated gene expression levels distinguish human from non-human primate brains. *Proceedings of the National Academy of the Sciences of USA* **100**: 13030–13035.

Carroll SB (2003) Genetics and the making of *Homo sapiens*. *Nature* **22**: 849–857.

Dorus S, Vallender EJ, Evans PD *et al*. (2004) Accelerated evolution of nervous system genes in the origin of *Homo sapiens*. *Cell* **119**: 1027–1040.

Enard W, Przeworski M, Fisher SE *et al*. (2002) Intra- and interspecific variation in primate gene expression patterns. *Science* **296**: 340–343.

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

Gasper J and Swanson WJ (2007) Molecular population genetics of the gene encoding the human fertilization protein zonadhesion reveals rapid adaptive evolution. *American Journal of Human Genetics* **79**: 820–830.

Gilad Y, Man O and Glusman G (2005) A comparison of the human and chimpanzee olfactory gene repertoires. *Genome Research* **15**: 224–230.

Hamblin MT, Thompson EE and Di Rienzo A (2002) Complex signatures of natural selection at the Duffy blood group locus. *American Journal of Human Genetics* **70**: 369–383.

Hamm D, Mautz BS, Wolfner MF, Aquadro CF and Swanson WJ (2007) Evidence of amino-acid diversifyity-enhancing selection within humans and among primates at the candidate sperm-receptor gene *PKDREJ*. *American Journal of Human Genetics* **81**: 44–52.

Hughes AL (2007) Looking for Darwin in all the wrong places: the misguided quest for positive selection at the nucleotide level. *Heredity* **1**: 1–10.

Hughes AL and Nei M (1988) Pattern of nucleotide substitution at MHC class I loci reveals overdominant selection. *Nature* **335**: 167–170.

Khaitovich P, Hellmann I, Enard W *et al*. (2005) Parallel patterns of evolution in the genomes and transcriptones of humans and chimpanzees. *Science* **309**: 1850–1854.

Maynard Smith J and Haigh J (1974) The hitch-hiking effect of a favourable gene. *Genetics Research* **23**: 23–35.

Mikkelsen TS, LaDeana W Hiller, Eichler EE *et al*. (Chimpanzee Sequencing and Analysis Consortium) (2005) Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* **437**(7055): 69–87.

Nielsen R, Bustamante C, Clark AG *et al*. (2005) A scan for positively selected genes in the genomes of humans and chimpanzees. *Plos Biology* **3**: e170.

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

Prabhaker S, Noonan JP, Paabo S and Rubin EM (2006) Accelerated evolution of conserved noncoding sequences in humans. *Science* **314**: 786.

Przerworski M (2002) The signature of positive selection at randomly chosen loci. *Genetics* **160**: 1179–1189.

Sabeti PC, Walsh E, Schaffner SF *et al*. (2005) The case for selection at *CCR5Δ32*. *PLos Biology* **3**(11): e378.

Sabeti PC, Schaffner SF, Fry B *et al*. (2006) Positive natural selection in the human lineage. *Science* **312**: 1614–1620.

Shi P, Bakewell MA and Zhang J (2006) Did brain-specific genes evolve faster in humans than in chimpanzees? *Trends in Genetics* **22**(11): 608–613.

Voight BF, Kudaravalli S, Wen X and Pritchard JK (2006) A map of recent positive selection in the human genome. *PLoS Biology* **4**(3): e72.

Wang H-Y (2007) Rate of evolution in brain-expressed genes in humans and other primates. *PLos Biology* **5**(2): e13.

Williamson SH, Hubisz MJ, Clark AG *et al*. (2007) Localizing recent adaptive evolution in the human genome. *PLoS Genetics* **3**(6): e90.

Wykoff GJ, Wang W and Wu C-I (2000) Rapid evolution of male reproductive genes in the descent of man. *Nature* **403**: 304–309.

Yu F, Hill RS, Schaffner SF *et al*. (2007) Comment on ongoing adaptive evolution at *ASPM*, a brain size determinant in *Homo sapiens*. *Science* **316**: 370b.

Further Reading

Bamshad M and Wooding SP (2003) Signatures of natural selection in the human genome. *Nature Reviews. Genetics* **4**: 99–111.

Harris EE and Meyer D (2006) The molecular signature of selection underlying human adaptations. *Yearbook of Physical Anthropology* **49**: 89–130.

Kimura M (1983) *The Neutral Theory of Molecular Evolution*. Cambridge: Cambridge University Press.

Nielsen R (2005) Molecular signatures of natural selection. *Annual Review of Genetics* **39**: 197–218.

Web Links

International HapMap Project http://www.hapmap.org
Perlegen Biosciences http://www.perlegen.com