

PERSPECTIVE

Evolutionary genetics as a tool to target genes involved in phenotypes of medical relevance

Evelyne Heyer¹ and Lluís Quintana-Murci²

1 MNHN, Eco-Anthropologie, UMR 5145, CNRS-MNHN-P7, Musée de l'Homme, Paris, France

2 Institut Pasteur, Unit of Human Evolutionary Genetics, CNRS URA3012, Paris Cedex, France

Keywords

epidemiology, evolution, human adaptation, human diversity, human genome, infectious disease, natural selection, nutrition.

Correspondence

Evelyne Heyer, MNHN, Eco-Anthropologie, UMR 5145, CNRS-MNHN-P7, Musée de l'Homme, Paris, 75116 France.

Tel.: +33 1440 5725; fax: +33 1440 57241;

e-mail: heyer@mnhn.fr

Received: 24 October 2008

Accepted: 4 December 2008

doi:10.1111/j.1752-4571.2008.00061.x

Abstract

There is an increasing interest in detecting genes, or genomic regions, that have been targeted by natural selection. Indeed, the evolutionary approach for inferring the action of natural selection in the human genome represents a powerful tool for predicting regions of the genome potentially associated with disease and of interest in epidemiological genetic studies. Here, we review several examples going from candidate gene studies associated with specific phenotypes, including nutrition, infectious disease and climate adaptation, to whole genome scans for natural selection. All these studies illustrate the power of the evolutionary approach in identifying regions of the genome having played a major role in human survival and adaptation.

Clues from evolutionary genetics: natural selection in humans

As modern humans migrated from Africa to colonize new lands 60–80 000 Ka, they were exposed to different climatic, nutritional and pathogenic environments. In such a situation, it is expected that human migrations were accompanied by genetic adaptations to emergent selective forces (Harpending and Rogers 2000; Cavalli-Sforza and Feldman 2003). This represents the basis of *natural selection*, which is formally defined as the differential reproduction of genotypes in succeeding generations. Genotypic variation produces individuals with varying capacities to survive and reproduce in different environments (Nielsen 2005). Natural selection can alter the levels of variability in several ways; genetic variants that increase the fitness of an individual in his environment tend to increase in frequency as a result of *positive or directional selection*, whereas deleterious mutations tend to be eliminated by *negative or purifying selection* (Bamshad and Wooding 2003; Nielsen 2005; Sabeti et al. 2006). Furthermore, positive selection for two alternative proteins coded by the same gene can result in *balancing selection* at the DNA level (i.e. heterozygote advantage or

frequency-dependent selection) (Bamshad and Wooding 2003; Nielsen 2005; Sabeti et al. 2006). Positive selection is a fundamental process in speciation and adaptation, and the identification of targets of natural selection in humans is critical to understand the evolution of our species as a whole (Sabeti et al. 2006). In contrast, local adaptation due to spatial and temporal variation in selective pressures, such as those imposed by pathogens, climate or diet, may have been restricted to particular populations or environments (Sabeti et al. 2006).

There is an increasing interest in detecting genes, or genomic regions, that have been targeted by natural selection (Nielsen 2005; Sabeti et al. 2006). Indeed, inferences concerning the action of natural selection in the human genome provide a powerful tool for predicting regions of the genome potentially associated with disease (Nielsen 2005; Quintana-Murci et al. 2007; Blekhman et al. 2008). Genetic variants influencing human susceptibility to disease are likely to affect the fitness of the organism. There is therefore an intimate relationship between disease and selection that can be exploited for the identification of candidate disease loci. The recent genome-wide surveys of genetic variation provided by the International HapMap Consortium (Frazer et al. 2007) or other genotyping

efforts to catalogue genetic variation at the population level such as those provided by Perlegen (Hinds et al. 2005) and HGDP (Jakobsson et al. 2008; Li et al. 2008), represent a turning point in the study of natural selection in humans. These large public datasets of population genetic variation in 'normal' individuals represent a powerful tool that helps investigators to both identify where in the genome and how natural selection has acted and unmask genetic factors contributing to medical traits, such as susceptibility to disease, protection against illness or variation in drug response.

Today, a plethora of statistical tests are proposed to identify the multiple signatures of natural selection. These tests can be largely subdivided into those that take into account the divergence between species (i.e. inter-species neutrality tests) and those that only consider polymorphic variation within humans (i.e. intra-species neutrality tests). The inter-species neutrality tests are able to detect selective events that took place up to a few million years ago (i.e. adaptive events that participated in the emergence of modern human) and include the traditional Ka/Ks , dN/dS and MacDonal-Kreitman (MK) tests (for individual references and detailed reviews of all these tests, see (Bamshad and Wooding 2003; Nielsen 2005; Sabeti et al. 2006)). In contrast, intra-species neutrality tests will detect more recent selective events (i.e. those occurring after the emergence of modern humans ~ 200 Ka) and can be largely subdivided into (i) those evaluating the frequency spectrum of mutations (e.g. Tajima's D , Fu and Li's D , F and Fay and Wu's H); (ii) those measuring the levels of population differentiation between the different studied populations (e.g. F_{ST} statistic) and (iii) those based on the levels of LD associated with particular haplotypes and/or alleles [e.g. long-range haplotype, the haplotype similarity (HS) or the iHS tests (for individual references and detailed reviews of all these tests, see (Bamshad and Wooding 2003; Nielsen 2005; Sabeti et al. 2006)]. These latter tests have been shown to be the most powerful to detect events of recent positive selection (i.e. $<30,000$ years).

From genotypes to phenotypes: case studies

Identifying and understanding the traits or phenotypes that have been targets of selection is of major relevance to understand how humans have adapted to varying environmental conditions over time and can provide new insights into genes having played a major role in human survival. This, in turn, can provide insights into the location of functionally important polymorphisms and may help prioritize targets for association mapping [for a review, see (Nielsen et al. 2007)]. Below, we present a number of selected examples for which there is com-

pellent evidence of natural selection targeting genes involved in three major phenotypes, namely nutrition, host-pathogen interactions and climate adaptation.

Natural selection and nutritional resources

The history of humans is punctuated with several well-identified shifts where cultural innovations played a major role. The emergence of agriculture in human societies, for the first time 10 000 years ago, involved a number of cultural and demographic changes. In the context of the 100 000 years of modern human evolution, the shift from a nomadic hunter-gatherer to an agriculture-based lifestyle was extremely rapid. It was certainly made possible by a major cultural transition sometimes referred to as the Neolithic *revolution*. Some populations specialised in agriculture, while others became herders and some remained hunter-gatherers. Among other factors, the diet dramatically changed during this period, with increased intake of cereals for agriculturist populations, and a meat-rich and high dairy input diet for herder populations. Many genes involved in the processing of food are likely to have responded to strong selection pressures at that time, and the nutritional and metabolic adaptations resulting from that, are believed to be important.

A recent study has provided evidence supporting the hypothesis that selective pressure can be sufficiently intense to modify the expression profiles of digestive enzymes (Perry et al. 2007). They assessed the extent to which the consumption of starch, a prominent component of the diet of farmers, has exerted selective pressures detectable in our genome. They characterised patterns of genetic variation for salivary amylase, the enzyme responsible for starch digestion in humans. The amylase-encoding *AMY1* gene is present in multiple copies on chromosome 1p21.1. The number of copies varies between individuals. A significant correlation between *AMY1* genotype (copy number polymorphism) and phenotype (amylase concentration in saliva) was observed. In addition, populations with traditional starch-dense diets had a higher *AMY1* copy number than populations with low-starch diets, such as rainforest and arctic hunter-gatherers.

The ability to digest milk in adulthood is another clear example of natural selection related to nutritional resources. Like in most mammals, the ability to digest milk rapidly declines after weaning in most humans, owing a decrease in lactase levels in the small intestine. However, many human populations worldwide, particularly those who have traditionally based their mode of subsistence on pastoralism, can continue to drink milk into adulthood without problems, because lactase expression persists in the gut. Twin studies first indicated a

genetic basis for lactase persistence, a trait that seems to be inherited as a single dominant mutation. Genotype–phenotype studies subsequently demonstrated that lactase persistence in Europeans was due to a single mutation in the promoter of *LCT* – the gene encoding lactase – extending *LCT* expression into adulthood (Enattah et al. 2002). In Europeans, this lactase-persistence allele, T-13910, is part of a haplotype (a combination of alleles) largely exceeding its expected length under neutrality (Bersaglieri et al. 2004). This genomic feature is a clear signature of recent positive selection, which has allowed the *LCT* allele to reach very high frequencies in European populations, particularly in northern Europe. The geographic distribution of the lactase-persistence allele is strongly correlated with historical milk consumption. This observation, together with the estimated age of expansion of the T-13910 mutation at 5000–12 000 years ago (Beja-Pereira et al. 2003), suggests that lactase persistence emerged in response to the cultural innovation of dairying associated with an agricultural lifestyle. Intriguingly some East African populations with a high prevalence of lactose tolerance in adulthood do not present the European T-13910 allele (Ingram et al. 2007). A recent study has identified a different mutation in the *LCT* promoter region – C-14010 – that also increases *LCT* expression and has been positively selected in East African pastoralists (Tishkoff et al. 2007). This provides a fascinating example of convergent adaptation within our species, and shows that the cultural trait of milk consumption confers such a strong selective advantage in terms of human survival that lactase persistence alleles have increased in frequency independently in several parts of the world.

Indirect evidence witnessing past adaptations to low carbohydrate food before the Neolithic transition is provided by the increased prevalence of diabetes, obesity, hypertension, etc. associated with the recent shift in diet and lifestyle in many countries. This last major transition concerns mainly western countries and is associated with a drastic increase of glucides intake. This so-called ‘high carbohydrate diet’ can lead to metabolic disorders such as diabetes, obesity and hypertension. For example, the World Health Organization considers that more than 180 million people in the world suffer from diabetes (type-2 diabetes in 90% of the cases) and this number is likely to double before 2030. It is indeed admitted that these nutritional disorders may reflect a process of *maladaptation*, i.e. a poor capacity of humans to process the present-day diet in modern societies. Two alternative hypotheses have been put forward to explain the selective pressure in the past. The first is known as the ‘carnivore connection’ and postulates that during the Palaeolithic ‘Ice-Age’, meat was the main nutrient and glucose intake was very low (Colagiuri and Brand Miller 2002). To maintain sufficient

level of glucose in the blood (glycaemia of 1 g/L) and ‘feed’ the brain, the efficiency of insulin, responsible of glucose consumption and storage, had to be reduced. The resulting phenotype is referred to as *Insulin Resistance*. On the other side, the pathway which increases hepatic glucose production from protides (gluconeogenesis) had to be positively selected for. The other hypothesis, ‘the thrifty genotype hypothesis’ (Neel 1962, 1999), postulates that cycles of food scarcity and abundance selected for an increased capacity to store food energy as body fat during periods of plenty for subsequent use during periods of food scarcity. Recent studies however challenge this view by showing that foragers may have had no more scarcity in food production than agriculturists (Benyshek and Watson 2006). Both hypotheses have nevertheless the same consequences: to select for insulin resistance and for gluconeogenesis during the Palaeolithic period, so as to constantly maintain sufficient levels of glucose in the blood. The selective pressures on these genes declined strongly for agriculturist populations because the agricultural revolution brought a sharp increase in the quantity of carbohydrate consumed through cereals intake. For herder populations, where the intake of meat remained high, or for hunter–gatherer populations, the selective pressures for insulin resistance and gluconeogenesis were maintained strongly. Another consequence from this recent maladaptation hypothesis is that variants associated with these metabolic disorders should be frequent (Di Rienzo and Hudson 2005).

Today, approximately 10 genes have been associated with type-2 diabetes using linkage analyses or association studies (Freeman and Cox 2006). Recent studies using high-density microarrays have confirmed some of these associations and identified new candidate loci (Sladek et al. 2007; Zeggini et al. 2008). In only a small proportion of the clinical cases, a single variant has been identified as the main cause for the disease (Malecki 2005). In all other cases (referred to as complex, polygenic or multifactorial type-2 diabetes) several genes contribute to the disease jointly.

Natural selection and host–pathogen interactions

The evolutionary dynamics of host–pathogen interactions lead to constant natural selection for adaptation and counter-adaptation in the two competing species (Cooke and Hill 2001). Pathogens continually develop new ways to avoid host recognition or elimination, and the host, in turn, must evolve to keep pace with this increasingly sophisticated evasion by pathogens. The first evidence of selection acting on a human gene involved in host–pathogen interactions was obtained for the sickle cell anaemia Hb^S allele and malarial resistance (Haldane 1949).

Similarities in the geographic distribution of hemoglobinopathies and *Plasmodium falciparum* infection led Haldane to propose that red blood cell disorders, such as thalassemia, might protect the host against malarial infection (Haldane 1949). Individuals homozygous for the Hb^S variant suffer sickle-cell anaemia, but heterozygous individuals are better protected against *falciparum* malaria (Hill et al. 1991; Ackerman et al. 2005). Another classical example of intense natural selection is the major histocompatibility complex (*MHC*), a group of related proteins involved in antigen presentation (Ohta 1991; Hughes et al. 1994). The most striking selective manifestation in humans is the sharing of ancestral polymorphisms with other hominoid taxa. Although the exact nature of the selective pressures remains unclear, selection seems to favour heterozygotes and low-frequency alleles, leading ultimately to high levels of allelic diversity at the *MHC* locus. Selection pressure on the human genes involved in immune-related processes or, more generally, in host–pathogen interactions, is not limited to these instances. Other examples include β -defensins, interleukins, immunoglobulins, killer cell inhibitory receptors and cell-surface molecules [for a review, see Vallender and Lahn (2004)].

Because malaria is a major killer of children worldwide, there has been intense research focused on host genetic variability and susceptibility to malarial infection (Kwiatkowski 2005). Most of the human genes thought to decrease the risk of malarial infection are expressed in red blood cells or encode proteins involved in the immune system. For example, G6PD enzyme deficiency has been shown to be associated with a lower risk of *falciparum* malaria (Beutler 1994). Specifically, the A⁻ allele leads to a reduction in enzyme activity of ~12% and is found in African populations at frequencies of up to 25% (Tishkoff et al. 2001). Several studies seem to converge toward a similar evolutionary pattern: G6PD alleles associated with low-level enzymatic activity have been subject to positive selection (Tishkoff et al. 2001; Sabeti et al. 2002; Verrelli et al. 2002). For example, the rapid expansion of the A⁻ allele in Africa accounts for the low levels of microsatellite diversity observed within this lineage (Tishkoff et al. 2001). This low level of internal diversity, in conjunction with the high frequency of the A⁻ allele, indicates that the G6PD A⁻ allele may have increased in frequency so rapidly that there was no time to accumulate new variation in nearby polymorphisms. In addition, long-range LD is observed around haplotypes bearing the G6PD A⁻ variant, whereas this is not the case for other G6PD variants of similar frequency (Sabeti et al. 2002). The date of origin of the G6PD A⁻ allele has been estimated to 2500–6500 years ago (Tishkoff et al. 2001; Sabeti et al. 2002). These dates are consistent with archaeological data

showing that malaria has had a major impact only in the last 10 000 years, coinciding with the expansion of *Plasmodium falciparum* populations (Hughes and Verra 2001).

Another interesting example of natural selection and host–pathogen interactions concerns CCR5 and HIV. The role of CCR5 in the pathogenesis of AIDS was highlighted by the observation that individuals homozygous for a 32-bp deletion in CCR5 (i.e. CCR5- Δ 32) were more resistant to HIV-1 infection (Dean et al. 1996). The high proportion of nonsynonymous mutations observed at CCR5 first suggested the occurrence of selective pressure for amino-acid divergence targeting this gene (Carrington et al. 1997). More compelling evidence came from the observation that CCR5- Δ 32 occurs at relatively high frequency (up to 16%) in Europeans, but is absent elsewhere (Stephens et al. 1998). Haplotype and coalescent analyses showed that the CCR5- Δ 32 ancestral haplotype probably emerged ~700 years ago. This observation was taken as evidence of a past strong selective event (e.g. an epidemic of a pathogen that, like HIV-1, utilizes CCR5), driving CCR5- Δ 32 frequencies upward in European populations (Stephens et al. 1998). However, other studies seem to contradict these interpretations, suggesting that the patterns of variation observed at the CCR5- Δ 32 are largely consistent with neutrality (Austerlitz et al. 2003; Sabeti et al. 2005). Other studies have suggested the action of balancing selection targeting the 5' *cis*-regulatory region of CCR5 (Bamshad et al. 2002), which has also been shown to be associated with variation in HIV-1 disease progression (Gonzalez et al. 2001). However, the introduction of HIV-1 into human populations occurred too recently to account for the observed selection signatures in CCR5, indicating that the HIV-1 resistance afforded by CCR5 *cis*-regulatory variants results from adaptive changes to older pathogens (Lalani et al. 1999; Bamshad et al. 2002).

Another infectious disease that may have imposed strong selective pressure in the human genome is tuberculosis. This disease tops the WHO list for numbers of deaths due to a single infectious agent, killing between 2 and 3 million people per year (Lienhardt 2001). There is increasing evidence to suggest that host genetic factors determine differences in host susceptibility to mycobacterial infection and might contribute to the pattern of clinical disease (Casanova and Abel 2002; Bellamy 2003). The C-type lectin DC-SIGN has been shown to be the major receptor for *Mycobacterium tuberculosis* on human dendritic cells (Tailleux et al. 2003). A recent association case–control study showed that two DC-SIGN promoter region variants are associated with a reduced risk of developing tuberculosis in a South African cohort (Barreiro et al. 2006). Interestingly, the much higher prevalence in

non-African populations of the two 'protective' DC-SIGN variants may thus be seen as the result of genetic adaptation to a longer period of TB exposure. However, DC-SIGN is known to interact with a plethora of other bacteria, viruses and parasites (Cambi et al. 2005), making difficult to correlate a particular pathogen with the geographic distribution of the two promoter variants. An independent study has investigated the issue of how natural selection has shaped the patterns of variability of DC-SIGN (Barreiro et al. 2005). Analyses of the sequence of the whole gene in a multi-ethnic panel of individuals have shown that DC-SIGN has been under strong selective constraints, preventing the accumulation of aminoacid changes over time. The strong pattern of conservation of DC-SIGN and the selective constraints acting on it (Barreiro et al. 2005) indicate that this lectin may play a crucial role in host human defenses (Neyrolles et al. 2006).

It becomes clear that searching for evidence of selective pressures on genes involved in immunity-related processes or host-pathogen interactions represents a complementary strategy for identifying functionally important genes and variants involved in host immunity to infection and disease outcome (Olson 2002; Bamshad and Wooding 2003; Bradbury 2004). As most genomic regions evolve under neutrality (Kimura 1968), the specific regions targeted by natural selection reflect the importance of genes residing there for host defence (Bradbury 2004; Quintana-Murci et al. 2007). The evolutionary approach also makes it possible to define the redundant and nonredundant functions of individual immunity-related genes *in natura* (Sabeti et al. 2006; Quintana-Murci et al. 2007). For example, mannose-binding lectin (MBL) has long been considered a key player in protective immunity, because MBL deficiency has been associated with susceptibility to infectious diseases. However, the high worldwide prevalence of *MBL2* deficiency variants, associated with the production of little or no MBL, challenges this view. This observation has been traditionally interpreted as indicating a selective advantage for low MBL levels. Using an evolutionary approach, it has been shown that the pattern of *MBL2* variation is consistent with neutral evolution (Verdu et al. 2006). These results, based on sequence-based neutrality tests, LD-based tests and inter-species comparisons, therefore suggest that the high worldwide frequencies of *MBL2* deficiency alleles result from a relaxation of the selective constraints on the human lineage. These observations indicate that the high worldwide prevalence of deleterious *MBL2* alleles is the result of genetic drift and that MBL is largely redundant in immunity to infection in the natural setting (Verdu et al. 2006). All these examples illustrate how the integration of evolutionary data into an immunological frame-

work can provide important insights into genes playing a major biological role for past and present survival and highlight host pathways playing an essential role in pathogen resistance.

Natural selection and climate adaptation

Early migrations out of Africa exposed ancestral populations to colder environments, with less incident sunlight. The most obvious response is in pigmentation, due to the quantity, type and distribution of melanin. One explanation would be that darker skin is likely to be favoured in regions of high UVR for its protection against sunburn and skin cancer (Robins 1991; Izagirre et al. 2006), whereas lighter skin permit sufficient vitamin D photosynthesis (Jablonski and Chaplin 2000). An alternative explanation is based on sexual selection (Darwin 1871; Aoki 2002) but has not been tested. From mouse studies, approximately 100 genes are known to be involved in the determination of pigmentation (Bennett and Lamoreux 2003). Among these genes, quantitative genetic studies have estimated that 10 genes may be involved in the variation of pigmentation in humans (Lynch and Walsh 1998). More recently, evolutionary genetic approaches have shown the impact of natural selection on differences in skin pigmentation. Different population studies have identified ~10 genes exhibiting signatures of natural selection [review in McEvoy et al. (2006)]. Interestingly, it has been shown that convergent evolution has occurred in relation to lighter pigmentation in Europeans and East Asians (Norton et al. 2007).

Other adaptations to climate conditions can have a more important impact in human health. Because the human species first arose in equatorial Africa, the ability to tolerate high temperatures was likely under strong selective pressure and may have become deleterious in the colder environments humans encountered after spreading out of Africa. An interesting example illustrating this type of selective pressure concerns sodium homeostasis. High sodium retention may be necessary in hot, humid environments, whereas it may be deleterious in cooler environments. The study of patterns of selection in genes involved in salt and water retention has shown that variants in the *CYP3A4* and *AGT* genes, which are implicated in hypertension, show remarkable interethnic differences and a strong correlation with latitude (Thompson et al. 2004). A more recent study has tested the hypothesis that climate has shaped variation in metabolism genes by genotyping more than 800 SNPs at 82 candidate genes for common metabolic disorders in 54 worldwide populations (Hancock et al. 2008). They tested for correlations between the frequency distribution of these SNPs with climate variables and compared the frequency distribution

of 'metabolic SNPs' with that expected under neutrality (i.e. allelic distribution at 'neutral' control SNPs). For nearly all climate variables, several 'metabolic SNPs' show stronger, and significant correlations than control SNPs. This subset of highly correlated SNPs differs according to the climate variables. This study not only has shown the occurrence of genetic adaptation to climate but also has identified new genes and variants that are good candidates to be used in association studies searching for genes involved in energy metabolism disorders. Among the strongest candidates were several SNPs in *LEPR*, *R109K*, *FABP2*, which had previously been associated with phenotypes directly related to cold tolerance. However, associations with climate may hide other associations between genetics and environmental factors. For example, different climate conditions may be associated with differences in food intake and pathogen presence (Guernier et al. 2004).

Genome-wide approaches to detect natural selection

Before the advent of large datasets of genomewide variation, the best examples of selection acting in humans were discovered in candidate gene studies where there was an *a priori* hypothesis of selection. Hence, very little was known about how widespread natural selection is at the genome-wide scale and about whether some kinds of genes or biological processes are particularly most involved in the adaptation of modern humans. It was also unclear the extent to which selective events have been geographically restricted, as opposed to taking place in all populations. Genome-wide studies of natural selection in humans have now become possible because of the availability of a near-complete sequence of the human genome, together with an increasing number of genome sequences for other species, and large catalogues of human genetic variation, such as those provided by the International Haplotype Map (HapMap) Project (Frazer et al. 2007), Perlegen Sciences (Hinds et al. 2005) and HGDP-CEPH (Human Genome Diversity Project-Centre d'Etude du Polymorphisme Humain) (Jakobsson et al. 2008; Li et al. 2008). The availability of such large-scale SNP data sets clearly makes it possible to provide detailed selection maps in humans and other organisms. The robust detection of natural selection in the human genome using these datasets is however complicated by the varying levels of biases in the choice of genetic markers studied (ascertainment biases) and by the mimicking effects that some demographic processes (e.g. expansions in Africa, bottlenecks among non-Africans) and natural selection have in the shaping of the patterns of genetic variation (Nielsen 2005; Nielsen et al. 2007). Despite these potential caveats, several genomewide studies have identified genes that may be good candidates for

selected loci because of their outlier behaviour with respect to the rest of the genome (Carlson et al. 2005; Voight et al. 2006; Frazer et al. 2007; Sabeti et al. 2007; Tang et al. 2007; Barreiro et al. 2008). Specifically, these genes exhibit increased levels of LD, reduced or enhanced levels of variability, increased levels of population differentiation, or skewed allele frequency spectra.

These studies have provided with blind classes of candidate genes under selection (Carlson et al. 2005; Voight et al. 2006; Frazer et al. 2007; Sabeti et al. 2007; Tang et al. 2007; Barreiro et al. 2008), with significant overrepresentation of genes involved in chemosensory perception and olfaction, immunity, reproduction and fertility, carbohydrate metabolism, and morphological traits. Some genes appear to be recurrently emerging from the different genome scans for selection. Among others, the patterns of variation of the *EDAR* gene are particularly interesting (Sabeti et al. 2007; Bryk et al. 2008). *EDAR* has a central role in generation of the primary hair follicle pattern, and mutations in *EDAR* cause hypohidrotic ectodermal dysplasia (HED) in humans and mice, characterized by defects in the development of hair, teeth and exocrine glands (Monreal et al. 1999; Mou et al. 2006). A mutation encoding a V370A substitution in *EDAR* is near fixation in Asia and absent in Europe and Africa; this polymorphism is notable because it is highly differentiated between the Asian and other continental populations (the 3rd most differentiated among 15 816 nonsynonymous SNPs), and also within Asian populations (Sabeti et al. 2007; Barreiro et al. 2008; Bryk et al. 2008). The V370A polymorphism, proposed to be the target of selection, lies within *EDAR*'s highly conserved death domain, the location of the majority of *EDAR* polymorphisms causing HED. In addition, it has been recently shown that this polymorphism was driven to high frequency in East Asia by positive selection prior to 10 000 years ago (Bryk et al. 2008). In humans, selective pressures on *EDAR* favouring changes in body temperature regulation and hair follicle density in response to colder climates may have influenced tooth shape, although this trait probably does not affect population fitness. This anecdotal example shows how 'phenotypic hitchhiking' in genes under positive selection may have substantially increased the observed number of physiological and morphological traits differentiating modern human populations (Barreiro et al. 2008).

However, leaving apart a number of genes that recurrently appear as being under selection from the different studies, there are large differences in the specific lists of candidate genes being reported by the different studies. These differences reflect (i) different methods and criteria used to detect selection (population differentiation, LD-based studies, allele frequency spectra, etc.) each of them

being more or less adapted to detect selection at a given different time scale, for a critic of some of them see (Hughes 2007), (ii) number of SNPs used in the analyses and (iii) differences in the populations used for the analyses (e.g. HapMap uses four general continental populations, Perlegen uses three populations – European, Asian- and African-Americans – known to be to some degree admixed and HGDP-CEPH uses more than 50 populations, most of them presenting idiosyncratic population histories). Detailed studies using large population datasets, applying different methods to detect selection using different aspects of the data and making use of massive sequence-based data, which will reduce the potential caveats related to SNP ascertainment bias, are now required.

Conclusions and perspectives

Traditionally, several different, yet complementary approaches to the identification of genetic variation important in the course of disease have been taken. By far the most common approach has been to look for association in candidate genes using case–control studies. In addition, family-based approaches have also become widely used, unmasking large numbers of loci associated with susceptibility to diseases. In this context, evolutionary approaches are complementary to the more classical clinical and epidemiological studies. Indeed, the use of an evolutionary perspective to better understand human disease dynamics is not a new field. For example, a clear illustration of this comes from Haldane's 1949 proposal that genetic variation in globin genes might be driven by providing malarial resistance (Haldane 1949), and that similar forces from other pathogens could maintain great biochemical diversity. The interest of the evolutionary approach can be double: (i) it can confirm, or refute, the biological relevance of a candidate gene in human survival, (ii) using whole genome approaches, it can unmask genes having been of major biological importance for our species, either at the level of our species as a whole or involved in local adaptation. In addition, assessing whether and how human genes have been targeted by natural selection can also be very useful to predict the effects of these genes in human diseases, namely the evolutionary fate of alleles underlying Mendelian and complex (Zwick et al. 2000; Pritchard and Cox 2002; Di Rienzo 2006). Mendelian disease alleles are not transmitted from deceased children, accounting for their relatively low frequency in the general population (i.e. <1%). The corresponding genes are under strong purifying selection, a selective regime that eliminates virtually all newly-occurring amino-acid substitutions. Indeed, at the genome-wide scale, it has been shown that most Mendelian-disease genes (non-infectious and infectious diseases) are under

purifying selection, especially when the disease mutations are dominant (Bustamante et al. 2005; Barreiro et al. 2006; Blekhman et al. 2008). By contrast, complex disease alleles have lower penetrance and consequently can reach higher frequencies in the population (i.e. >5–10%) (Pritchard and Cox 2002; Di Rienzo 2006; Kryukov et al. 2007). Moreover, alleles contributing to late-onset diseases will be under lower selection pressure. In this case selection can occur after the end of the reproductive period owing to the selective pressure of parental care and to the variance at the age of onset of the disease (Pavard and Metcalf 2007). The corresponding genes will therefore exhibit signatures of weaker purifying selection. They are globally under less selective constraints than Mendelian-disease genes, and the selection coefficient of mutations associated with complex disease risk is smaller than for Mendelian mutations (Blekhman et al. 2008). In this context, the definition of the effects of natural selection in the human genome is not only of major importance to distinguish genes having played an essential role in human survival from those exhibiting a rather redundant role, but also to understand the mode in which variation at these genes will contribute to human disease. Indeed, many of the genes, where signature of selection has been detected, do not have very large effects on diseases when allelic variation is present (Weiss 2008). They account for 2–10% of the variation in the disease. Even classical examples like sickle cell anaemia account for less variation in malaria than one would think (Mackinnon et al. 2005). Now it is time to reinforce the powerful predictions of the evolutionary approach. The massive genotype datasets should be accompanied by phenotypic data, such as measures of the circulating levels of insulin, cholesterol, immunoglobulins, interleukins, etc, so as to fill the gap between descriptive genetics datasets and phenotypic consequences. This integrative approach should shed light on the phenotypes that have participated in human adaptation as well as those involved in human maladaptation and disease.

Acknowledgements

The authors thank two anonymous reviewers for helpful comments on the manuscript.

Literature cited

- Ackerman, H., S. Usen, M. Jallow, F. Sisay-Joof, M. Pinder, and D. P. Kwiatkowski. 2005. A comparison of case-control and family-based association methods: the example of sickle-cell and malaria. *Annals of Human Genetics* **69**:559–565.
- Aoki, K. 2002. Sexual selection as a cause of human skin colour variation: Darwin's hypothesis revisited. *Annals of Human Biology* **29**:589–608.

- Austerlitz, F., L. Kalaydjieva, and E. Heyer. 2003. Detecting population growth, selection and inherited fertility from haplotypic data in humans. *Genetics* **165**:1579–1586.
- Bamshad, M., and S. P. Wooding. 2003. Signatures of natural selection in the human genome. *Nature Reviews. Genetics* **4**:99–111.
- Bamshad, M. J., S. Mummidi, E. Gonzalez, S. S. Ahuja, D. M. Dunn, W. S. Watkins, S. Wooding *et al.* 2002. A strong signature of balancing selection in the 5' cis-regulatory region of CCR5. *Proceedings of National Academy of Sciences of the United States of America* **99**:10539–10544.
- Barreiro, L. B., E. Patin, O. Neyrolles, H. M. Cann, B. Gicquel, and L. Quintana-Murci. 2005. The heritage of pathogen pressures and ancient demography in the human innate-immunity CD209/CD209L region. *American Journal of Human Genetics* **77**:869–886.
- Barreiro, L. B., O. Neyrolles, C. L. Babb, L. Tailleux, H. Quach, K. McElreavey, P. D. Helden *et al.* 2006. Promoter variation in the DC-SIGN-encoding gene CD209 is associated with tuberculosis. *PLoS Medicine* **3**:e20.
- Barreiro, L. B., G. Laval, H. Quach, E. Patin, and L. Quintana-Murci. 2008. Natural selection has driven population differentiation in modern humans. *Nature Genetics* **40**:340–345.
- Beja-Pereira, A., G. Luikart, P. R. England, D. G. Bradley, O. C. Jann, G. Bertorelle, A. T. Chamberlain *et al.* 2003. Gene-culture coevolution between cattle milk protein genes and human lactase genes. *Nature Genetics* **35**:311–313.
- Bellamy, R. 2003. Susceptibility to mycobacterial infections: the importance of host genetics. *Genes and Immunity* **4**:4–11.
- Bennett, D. C., and M. L. Lamoreux. 2003. The color loci of mice – a genetic century. *Pigment Cell Research* **16**:333–344.
- Benyshek, D. C., and J. T. Watson. 2006. Exploring the thrifty genotype's food-shortage assumptions: a cross-cultural comparison of ethnographic accounts of food security among foraging and agricultural societies. *American Journal of Physical Anthropology* **131**:120–126.
- Bersaglieri, T., P. C. Sabeti, N. Patterson, T. Vanderploeg, S. F. Schaffner, J. A. Drake, M. Rhodes *et al.* 2004. Genetic signatures of strong recent positive selection at the lactase gene. *American Journal of Human Genetics* **74**:1111–1120.
- Beutler, E. 1994. G6PD deficiency. *Blood* **84**:3613–3636.
- Blekhman, R., O. Man, L. Herrmann, A. R. Boyko, A. Indap, C. Kosiol, C. D. Bustamante *et al.* 2008. Natural selection on genes that underlie human disease susceptibility. *Current Biology* **18**:883–889.
- Bradbury, J. 2004. Ancient footsteps in our genes: evolution and human disease. *Gene variants selected during evolution may underlie many common diseases. Lancet* **363**:952–953.
- Bryk, J., E. Hardouin, I. Pugach, D. Hughes, R. Strotmann, M. Stoneking, and S. Myles. 2008. Positive selection in East Asians for an EDAR allele that enhances NF-kappaB activation. *PLoS ONE* **3**:e2209.
- Bustamante, C. D., A. Fledel-Alon, S. Williamson, R. Nielsen, M. T. Hubisz, S. Glanowski, D. M. Tanenbaum *et al.* 2005. Natural selection on protein-coding genes in the human genome. *Nature* **437**:1153–1157.
- Cambi, A., M. Koopman, and C. G. Figdor. 2005. How C-type lectins detect pathogens. *Cell Microbiology* **7**:481–488.
- Carlson, C. S., D. J. Thomas, M. A. Eberle, J. E. Swanson, R. J. Livingston, M. J. Rieder, and D. A. Nickerson. 2005. Genomic regions exhibiting positive selection identified from dense genotype data. *Genome Research* **15**:1553–1565.
- Carrington, M., T. Kissner, B. Gerrard, S. Ivanov, S. J. O'Brien, and M. Dean. 1997. Novel alleles of the chemokine-receptor gene CCR5. *American Journal of Human Genetics* **61**:1261–1267.
- Casanova, J. L., and L. Abel. 2002. Genetic dissection of immunity to mycobacteria: the human model. *Annual Review of Immunology* **20**:581–620.
- Cavalli-Sforza, L. L., and M. W. Feldman. 2003. The application of molecular genetic approaches to the study of human evolution. *Nature Genetics* **33**(Suppl.):266–275.
- Colagiuri, S., and J. Brand Miller. 2002. The 'carnivore connection' – evolutionary aspects of insulin resistance. *European Journal of Clinical Nutrition* **56**(Suppl. 1):S30–S35.
- Cooke, G. S., and A. V. Hill. 2001. Genetics of susceptibility to human infectious disease. *Nature Reviews. Genetics* **2**:967–977.
- Darwin, C. 1871. *The Descent of Man, and Selection in Relation to Sex*. D. Appleton and company, New York.
- Dean, M., M. Carrington, C. Winkler, G. A. Huttley, M. W. Smith, R. Allikmets, J. J. Goedert *et al.* 1996. Genetic restriction of HIV-1 infection and progression to AIDS by a deletion allele of the CKR5 structural gene. Hemophilia Growth and Development Study, Multicenter AIDS Cohort Study, Multicenter Hemophilia Cohort Study, San Francisco City Cohort, ALIVE Study. *Science* **273**:1856–1862.
- Di Rienzo, A. 2006. Population genetics models of common diseases. *Current Opinion in Genetics and Development* **16**:630–636.
- Di Rienzo, A., and R. R. Hudson. 2005. An evolutionary framework for common diseases: the ancestral-susceptibility model. *Trends in Genetics* **21**:596–601.
- Enattah, N. S., T. Sahi, E. Savilahti, J. D. Terwilliger, L. Peltonen, and I. Jarvela. 2002. Identification of a variant associated with adult-type hypolactasia. *Nature Genetics* **30**:233–237.
- Frazer, K. A., D. G. Ballinger, D. R. Cox, D. A. Hinds, L. L. Stuve, R. A. Gibbs, J. W. Belmont *et al.* 2007. A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449**:851–861.
- Freeman, H., and R. D. Cox. 2006. Type-2 diabetes: a cocktail of genetic discovery. *Human Molecular Genetics* **15**:R202–R209.
- Gonzalez, E., R. Dhanda, M. Bamshad, S. Mummidi, R. Geevarghese, G. Catano, S. A. Anderson *et al.* 2001. Global survey of genetic variation in CCR5, RANTES, and MIP-1alpha: impact on the epidemiology of the HIV-1 pandemic.

- Proceedings of National Academy of Sciences of the United States of America **98**:5199–5204.
- Guernier, V., M. E. Hochberg, and J. F. Guegan. 2004. Ecology drives the worldwide distribution of human diseases. *PLoS Biology* **2**:e141.
- Haldane, J. B. S. 1949. Disease and evolution. *Ric Sci Suppl. A* **19**:68–76.
- Hancock, A. M., D. B. Witonsky, A. S. Gordon, G. Eshel, J. K. Pritchard, G. Coop, and A. Di Rienzo. 2008. Adaptations to climate in candidate genes for common metabolic disorders. *PLoS Genetics* **4**:e32.
- Harpending, H., and A. Rogers. 2000. Genetic perspectives on human origins and differentiation. *Annual Review of Genomics and Human Genetics* **1**:361–385.
- Hill, A. V., C. E. Allsopp, D. Kwiatkowski, N. M. Anstey, P. Twumasi, P. A. Rowe, S. Bennett *et al.* 1991. Common west African HLA antigens are associated with protection from severe malaria. *Nature* **352**:595–600.
- Hinds, D. A., L. L. Stuve, G. B. Nilsen, E. Halperin, E. Eskin, D. G. Ballinger, K. A. Frazer *et al.* 2005. Whole-genome patterns of common DNA variation in three human populations. *Science* **307**:1072–1079.
- Hughes, A. L. 2007. Looking for Darwin in all the wrong places: the misguided quest for positive selection at the nucleotide sequence level. *Heredity* **99**:364–373.
- Hughes, A. L., and F. Verra. 2001. Very large long-term effective population size in the virulent human malaria parasite *Plasmodium falciparum*. *Proceedings. Biological Sciences* **268**:1855–1860.
- Hughes, A. L., M. K. Hughes, C. Y. Howell, and M. Nei. 1994. Natural selection at the class II major histocompatibility complex loci of mammals. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **346**:359–366; discussion 366–357.
- Ingram, C. J., M. F. Elamin, C. A. Mulcare, M. E. Weale, A. Tarekegn, T. O. Raga, E. Bekele *et al.* 2007. A novel polymorphism associated with lactose tolerance in Africa: multiple causes for lactase persistence? *Human Genetics* **120**:779–788.
- Itzagirre, N., I. Garcia, C. Junquera, C. de la Rúa, and S. Alonso. 2006. A scan for signatures of positive selection in candidate loci for skin pigmentation in humans. *Molecular Biology and Evolution* **23**:1697–1706.
- Jablonski, N. G., and G. Chaplin. 2000. The evolution of human skin coloration. *Journal of Human Evolution* **39**:57–106.
- Jakobsson, M., S. W. Scholz, P. Scheet, J. R. Gibbs, J. M. VanLiere, H. C. Fung, Z. A. Szpiech *et al.* 2008. Genotype, haplotype and copy-number variation in worldwide human populations. *Nature* **451**:998–1003.
- Kimura, M. 1968. Evolutionary rate at the molecular level. *Nature* **217**:624–626.
- Kryukov, G. V., L. A. Pennacchio, and S. R. Sunyaev. 2007. Most rare missense alleles are deleterious in humans: implications for complex disease and association studies. *American Journal of Human Genetics* **80**:727–739.
- Kwiatkowski, D. P. 2005. How malaria has affected the human genome and what human genetics can teach us about malaria. *American Journal of Human Genetics* **77**:171–192.
- Lalani, A. S., J. Masters, W. Zeng, J. Barrett, R. Pannu, H. Everett, C. W. Arendt *et al.* 1999. Use of chemokine receptors by poxviruses. *Science* **286**:1968–1971.
- Li, J. Z., D. M. Absher, H. Tang, A. M. Southwick, A. M. Casto, S. Ramachandran, H. M. Cann *et al.* 2008. Worldwide human relationships inferred from genome-wide patterns of variation. *Science* **319**:1100–1104.
- Lienhardt, C. 2001. From exposure to disease: the role of environmental factors in susceptibility to and development of tuberculosis. *Epidemiologic Reviews* **23**:288–301.
- Lynch, M., and B. Walsh. 1998. *Genetics and Analysis of Quantitative Traits*. Sinauer, Sunderland, MA.
- Mackinnon, M. J., T. W. Mwangi, R. W. Snow, K. Marsh, and T. N. Williams. 2005. Heritability of malaria in Africa. *PLoS Medicine* **2**:e340.
- Malecki, M. T. 2005. Genetics of type 2 diabetes mellitus. *Diabetes Research and Clinical Practice* **68**(Suppl. 1):S10–S21.
- McEvoy, B., S. Beleza, and M. D. Shriver. 2006. The genetic architecture of normal variation in human pigmentation: an evolutionary perspective and model. *Human Molecular Genetics* **15**:R176–R181.
- Monreal, A. W., B. M. Ferguson, D. J. Headon, S. L. Street, P. A. Overbeek, and J. Zonana. 1999. Mutations in the human homologue of mouse *dl* cause autosomal recessive and dominant hypohidrotic ectodermal dysplasia. *Nature Genetics* **22**:366–369.
- Mou, C., B. Jackson, P. Schneider, P. A. Overbeek, and D. J. Headon. 2006. Generation of the primary hair follicle pattern. *Proceedings of National Academy of Sciences of the United States of America* **103**:9075–9080.
- Neel, J. V. 1962. Diabetes mellitus: a “thrifty” genotype rendered detrimental by “progress”? *American Journal of Human Genetics* **14**:353–362.
- Neel, J. V. 1999. Diabetes mellitus: a “thrifty” genotype rendered detrimental by “progress”? 1962. *Bulletin of the World Health Organization* **77**:694–703; discussion 692–693.
- Neyrolles, O., B. Gicquel, and L. Quintana-Murci. 2006. Towards a crucial role for DC-SIGN in tuberculosis and beyond. *Trends in Microbiology* **14**:383–387.
- Nielsen, R. 2005. Molecular signatures of natural selection. *Annual Review of Genetics* **39**:197–218.
- Nielsen, R., I. Hellmann, M. Hubisz, C. Bustamante, and A. G. Clark. 2007. Recent and ongoing selection in the human genome. *Nature Reviews. Genetics* **8**:857–868.
- Norton, H. L., R. A. Kittles, E. Parra, P. McKeigue, X. Mao, K. Cheng, V. A. Canfield *et al.* 2007. Genetic evidence for the convergent evolution of light skin in Europeans and East Asians. *Molecular Biology and Evolution* **24**:710–722.
- Ohta, T. 1991. Role of diversifying selection and gene conversion in evolution of major histocompatibility complex loci. *Proceedings of National Academy of Sciences of the United States of America* **88**:6716–6720.

- Olson, S. 2002. Population genetics. Seeking the signs of selection. *Science* **298**:1324–1325.
- Pavard, S., and C. J. Metcalf. 2007. Negative selection on BRCA1 susceptibility alleles sheds light on the population genetics of late-onset diseases and aging theory. *PLoS ONE* **2**:e1206.
- Perry, G. H., N. J. Dominy, K. G. Claw, A. S. Lee, H. Fiegler, R. Redon, J. Werner *et al.* 2007. Diet and the evolution of human amylase gene copy number variation. *Nature Genetics* **39**:1256–1260.
- Pritchard, J. K., and N. J. Cox. 2002. The allelic architecture of human disease genes: common disease-common variant...or not? *Human Molecular Genetics* **11**:2417–2423.
- Quintana-Murci, L., A. Alcais, L. Abel, and J. L. Casanova. 2007. Immunology in natura: clinical, epidemiological and evolutionary genetics of infectious diseases. *Nature Immunology* **8**:1165–1171.
- Robins, A. H. 1991. *Biological Perspectives on Human Pigmentation: Cambridge Studies in Biological Anthropology*, Vol. 7. pp. 253. Cambridge University Press, Cambridge.
- Sabeti, P. C., D. E. Reich, J. M. Higgins, H. Z. Levine, D. J. Richter, S. F. Schaffner, S. B. Gabriel *et al.* 2002. Detecting recent positive selection in the human genome from haplotype structure. *Nature* **419**:832–837.
- Sabeti, P. C., E. Walsh, S. F. Schaffner, P. Varilly, B. Fry, H. B. Hutcheson, M. Cullen *et al.* 2005. The case for selection at CCR5-Delta32. *PLoS Biology* **3**:e378.
- Sabeti, P. C., S. F. Schaffner, B. Fry, J. Lohmueller, P. Varilly, O. Shamovsky, A. Palma *et al.* 2006. Positive natural selection in the human lineage. *Science* **312**:1614–1620.
- Sabeti, P. C., P. Varilly, B. Fry, J. Lohmueller, E. Hostetter, C. Cotsapas, X. Xie *et al.* 2007. Genome-wide detection and characterization of positive selection in human populations. *Nature* **449**:913–918.
- Sladek, R., G. Rocheleau, J. Rung, C. Dina, L. Shen, D. Serre, P. Boutin *et al.* 2007. A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature* **445**:881–885.
- Stephens, J. C., D. E. Reich, D. B. Goldstein, H. D. Shin, M. W. Smith, M. Carrington, C. Winkler *et al.* 1998. Dating the origin of the CCR5-Delta32 AIDS-resistance allele by the coalescence of haplotypes. *American Journal of Human Genetics* **62**:1507–1515.
- Tailleux, L., O. Schwartz, J. L. Herrmann, E. Pivert, M. Jackson, A. Amara, L. Legres *et al.* 2003. DC-SIGN is the major *Mycobacterium tuberculosis* receptor on human dendritic cells. *Journal of Experimental Medicine* **197**:121–127.
- Tang, K., K. R. Thornton, and M. Stoneking. 2007. A new approach for using genome scans to detect recent positive selection in the human genome. *PLoS Biology* **5**:e171.
- Thompson, E. E., H. Kuttab-Boulos, D. Witonsky, L. Yang, B. A. Roe, and A. Di Rienzo. 2004. CYP3A variation and the evolution of salt-sensitivity variants. *American Journal of Human Genetics* **75**:1059–1069.
- Tishkoff, S. A., R. Varkonyi, N. Cahinhinan, S. Abbes, G. Argyropoulos, G. Destro-Bisol, A. Drousiotou *et al.* 2001. Haplotype diversity and linkage disequilibrium at human G6PD: recent origin of alleles that confer malarial resistance. *Science* **293**:455–462.
- Tishkoff, S. A., F. A. Reed, A. Ranciaro, B. F. Voight, C. C. Babbitt, J. S. Silverman, K. Powell *et al.* 2007. Convergent adaptation of human lactase persistence in Africa and Europe. *Nature Genetics* **39**:31–40.
- Vallender, E. J., and B. T. Lahn. 2004. Positive selection on the human genome. *Human Molecular Genetics* **2**:R245–R254.
- Verdu, P., L. B. Barreiro, E. Patin, A. Gessain, O. Cassar, J. R. Kidd, K. K. Kidd *et al.* 2006. Evolutionary insights into the high worldwide prevalence of MBL2 deficiency alleles. *Human Molecular Genetics* **15**:2650–2658.
- Verrelli, B. C., J. H. McDonald, G. Argyropoulos, G. Destro-Bisol, A. Froment, A. Drousiotou, G. Lefranc *et al.* 2002. Evidence for balancing selection from nucleotide sequence analyses of human G6PD. *American Journal of Human Genetics* **71**:1112–1128.
- Voight, B. F., S. Kudravalli, X. Wen, and J. K. Pritchard. 2006. A map of recent positive selection in the human genome. *PLoS Biology* **4**:e72.
- Weiss, K. M. 2008. Tilting at quixotic trait loci (QTL): an evolutionary perspective on genetic causation. *Genetics* **179**:1741–1756.
- Zeggini, E., L. J. Scott, R. Saxena, B. F. Voight, J. L. Marchini, T. Hu, P. I. de Bakker *et al.* 2008. Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nature Genetics* **40**:638–645.
- Zwick, M. E., D. J. Cutler, and A. Chakravarti. 2000. Patterns of genetic variation in Mendelian and complex traits. *Annual Review of Genomics and Human Genetics* **1**:387–407.