Understanding rare and common diseases in the context of human evolution

Lluis Quintana-Murci

Abstract

The wealth of available genetic information is allowing the reconstruction of human demographic and adaptive history. Demography and purifying selection affect the purge of rare, deleterious mutations from the human population, whereas positive and balancing selection can increase the frequency of advantageous variants, improving survival and reproduction in specific environmental conditions. In this review, I discuss how theoretical and empirical population genetics studies, using both modern and ancient DNA data, are a powerful tool for obtaining new insight into the genetic basis of severe disorders and complex disease phenotypes, rare and common, focusing particularly on infectious disease risk.

Introduction

Intense research efforts have focused on identifying rare and common variants that increase disease risk in humans, for both rare and common diseases. Several, non-mutually exclusive models have been proposed to explain the functional properties of such variants and their contributions to pathological conditions, and this topic has been reviewed elsewhere [1–10]. These studies implicated multiple variants in disease susceptibility, but the relative importance of rare and common variants in phenotypic diversity, both benign and disease-related, has yet to be explored in detail [11]. We can use an evolutionary approach to tackle this question, as population genetics models can predict the allelic architecture of disease susceptibility [12, 13]. They are able to do so because rare and common disease-risk alleles are a subset of global human genetic diversity, and their occurrence, frequency, and population distribution is governed by evolutionary forces, such as mutation, genetic drift (e.g., migration, admixture, and changes in population size), and natural selection.

The plethora of genetic information generated in the last ten years, thanks largely to the publication of sequencing datasets for both modern human populations and ancient DNA samples [14–18], is making it possible to reconstruct the genetic history of our species, and to define the parameters characterizing human demographic history: expansion out of Africa, the loss of genetic diversity with increasing distance from Africa (i.e., the "serial founder effect"), demographic expansions over different time scales, and admixture with ancient hominins [16–21]. These studies are also revealing the extent to which selection has acted on the human genome, providing insight into the way in which selection removes deleterious variation and the potential of human populations to adapt to the broad range of climatic, nutritional, and pathogenic environments they have occupied [22–28]. It has thus become essential to dissect the role of selection, in its diverse forms and intensities, in shaping the patterns of population genetic diversity (Fig. 1a), not only to improve our understanding of human evolutionary history, but also to obtain insight into phenotypic diversity and differences in the risk of developing rare and common diseases [12, 13, 24, 29–32].

The removal of mutations deleterious to human health

Studies of the occurrence, frequency, and population distribution of deleterious mutations are of fundamental importance if we are to understand the genetic architecture of human disease. Theoretical and empirical population genetics studies have shown that most new mutations resulting in amino acid substitutions (non-synonymous) are rapidly culled from the population through purifying selection (Fig. 1a) [33, 34]. Indeed, the small number of non-synonymous variants observed relative to the rate of non-synonymous mutation indicates that most non-synonymous mutations are lethal or highly deleterious, strongly compromising the reproductive success of their carriers [34–36]. Purifying selection—the most common
Fig. 1 Modes in which selection or admixture can remove, maintain, or increase genetic diversity. **a** Schematic representation of the different types of natural selection. Purifying selection removes deleterious alleles (in black) from the population, and genes evolving under strong purifying selection are usually associated with rare, severe disorders. Conversely, mutations conferring a selective advantage (e.g., increased resistance to complex infectious disease) can increase in frequency in the population, or be maintained, through different forms of positive and balancing selection. Positive selection is represented here by the classic hard-sweep model where, following an environmental change, a newly arisen advantageous mutation or a mutation at very low frequency (in red) will be immediately targeted by positive selection and will ultimately reach fixation. Balancing selection is illustrated here by the case of heterozygote advantage (or overdominance), where the presence of heterozygotes (in blue) is favored in the population. **b** Long-term balancing selection. Advantageous genetic diversity can be maintained over long periods of time and survive speciation, resulting in "trans-species polymorphism" (represented by black and red arrows). In this example, a trans-species polymorphism that is present in the modern European population (where it has survived the known bottleneck out of Africa) is shared with other primates, such as chimpanzees and gorillas. **c** Modern humans can also acquire genetic diversity (whether advantageous or not) through admixture with other hominins, such as Neanderthals or Denisovans (Box 2). The green and blue arrows represent the direction and estimated magnitude of admixture between modern humans and Neanderthals and Denisovans, respectively (see [17]).
form of selection—refers to the selective removal of alleles that are deleterious, such as those associated with severe Mendelian disorders, or their maintenance at low population frequencies (i.e., mutation–selection balance) [32, 37]. The efficacy of purifying selection for eliminating deleterious mutations from a population depends not only on the selection coefficient(s), but also on population size (N), which determines the magnitude of genetic drift. Unlike highly deleterious mutations, variants subject to weaker selection (i.e., weakly deleterious mutations) behave like “nearly neutral mutations”; they may, therefore, reach relatively high population frequencies [38–40]. In large outbred populations, with low levels of drift, deleterious mutations will eventually be eliminated. By contrast, in small populations, deleterious mutations behave very much like neutral mutations and may be subject to strong drift, resulting in moderate-to-high frequencies, or even fixation [39].

**Rare variants are widespread in the human genome**

Recent deep-sequencing studies are showing a surprisingly high proportion of rare and low-frequency variants in different human populations [14, 15, 41–47]. The Exome Variant Server, for example, reports frequency information from 6515 exomes of individuals of African American and European American ancestry [46]. The most recent release of the 1000 Genomes Project, based on full-genome information for 2504 individuals from 26 populations from around the world, revealed that there was a large number of rare variants in the global dataset (~64 million autosomal variants have a frequency <0.5 %, and only ~8 million have a frequency >5%), with each individual genome harboring between 40,000 and 200,000 rare variants [15]. A more recent report of high-quality exome data from 60,706 individuals of diverse geographic ancestry, generated as part of the Exome Aggregation Consortium (ExAC), has provided unprecedented resolution for the analysis of low-frequency variants as well as an invaluable resource for the clinical interpretation of genetic variants observed in disease patients [47].

The contribution of rare variants to human disease is a matter of considerable debate, together with the distribution of these variants in the population, as they may underlie early-onset disease and increase susceptibility to common diseases [1, 44, 45, 48–50]. Most rare variants are private to a population, whereas common variants tend to be shared by different populations [51]. Rare variants, particularly those specific to a particular population, tend to have stronger deleterious effects than common variants [42, 52, 53]. Consequently, as shown by population genetics studies, most variants with large functional effects tend to be rare and private, and only a small proportion of variants with large effects are common to different populations. Genome-wide association studies (GWAS), which focus on common variants, have been only moderately successful in explaining the genetic basis of complex diseases [3]. Furthermore, theoretical studies have shown that a large proportion of the so-called “missing heritability” is explained by rare variants, particularly those that affect fitness as well as causing disease [54].

The increasing amount of sequence-based datasets available, both in basic and medically oriented research, is accelerating the investigation into the contribution of rare variants to disease susceptibility. In this context, diverse variant annotation tools and predictive algorithms have been developed to systematically evaluate the potential functional impacts of genetic variants (e.g., PolyPhen, SIFT, and GERP) [55–57], helping to prioritize the study of putative causal variants in further detail. These methods, which use different statistics and types of information, generally assess the “deleteriousness” of each genetic variant by considering different measures, such as evolutionary conservation scores, changes in amino acid sequence, or potential effect on protein function and structure [58]. Novel methods are increasingly being developed, providing improved power and resolution. For example, CADD, which integrates both evolutionary and functional importance, generates a single prediction from multiple annotation sources, including other variant effect predictors [59]. Likewise, MSC provides gene-level and gene-specific phenotypic impact cutoff values to improve the use of existing variant-level methods [60].

Quantification of the burden of deleterious, mostly rare, variants across human populations and an understanding of the ways in which this burden has been shaped by demographic history are now key issues in medical research, because they could help to optimize population sampling and, ultimately, to identify disease risk variants.

**Expansion out of Africa and the patterns of rare, deleterious variants**

The sizes of human populations have changed radically over the last 100,000 years, due to range expansions, bottlenecks, and rapid growth over different timescales [18–21]. Several studies have evaluated the impact of such demographic events on the distribution of deleterious variants and have shown that populations that have experienced bottlenecks, such as non-Africans, have higher proportions of deleterious variants of essential genes than African populations. This pattern has been interpreted as resulting from weaker purifying selection due to the Out-of-Africa bottleneck [45, 52, 61]. Nevertheless, an absolute increase in the number of rare functional variants has been observed in populations of African and European descent, relative to neutral expectations, due to the combined effects of an explosive expansion over recent millennia and weak purifying selection [41–46].
Furthermore, ~85% of known deleterious variants appear to have arisen during the last 5000 to 10,000 years, and these variants are enriched in mutations with a (relatively) large effect as there has not yet been sufficient time for selection to eliminate them from the population [46]. Furthermore, deleterious mutations in Europeans appear to have occurred after those in Africans (~3000 vs. 6200 years ago, respectively) [46], highlighting the effects of demographic history on the distribution of deleterious variants within the population.

However, some studies have suggested that demographic history may have a less straightforward impact on the mean burden of deleterious variants [62–64]. Simons and coworkers concluded that individual mutation load is insensitive to recent population history [64], and Do and coworkers suggested that selection is equally effective across human populations [62]. Several factors underlie these apparently conflicting conclusions, including differences in the choice of statistics and the features of genetic variation used to assess the burden of deleterious variation, and differences in the choice of predictive algorithms for defining deleteriousness, together with differences in the interpretations of the results; these factors have been reviewed in detail elsewhere [22, 65]. Nevertheless, all these studies converge to suggest that demographic history affects deleterious and neutral variants differently (Fig. 2), and that mutation and drift have stronger effects on the frequency of weakly deleterious mutations in bottlenecked populations than in large, expanding populations.

**Founder effects and bottlenecks increase the burden of deleterious variation**

Besides the impact of long-term population demographics (i.e., African vs. non-African populations) on the distribution of deleterious variants, a few studies have evaluated the effects of more recent, or stronger, changes in population demography. For example, it has been shown that French Canadians have both lower levels of diversity and a larger proportion of deleterious variants than the present-day French population. These findings highlight how a recent major change in population demographics (i.e., a small founder population of ~8500 French settlers subsequently growing by about 700-fold to attain its present size) can profoundly affect the population’s genetic landscape within as little as 400 years [66]. Likewise, the Finnish population, which experienced a recent population bottleneck estimated to have occurred ~4000 years ago, has larger proportions of rare deleterious alleles, including loss-of-function variants and complete gene knockouts, than other populations in Europe or of European descent [67].

Henn and coworkers investigated the consequences of a serial founder effect model for the distribution of deleterious mutations using a set of African populations and several groups located at different geographic distances

![Fig. 2](image-url)
from Africa [68]. Using explicit demographic models and considering different selection coefficients and dominance parameters, they found that non-African individuals carried larger proportions of deleterious alleles, mostly of modest effect, than African individuals, and that the number of homozygous deleterious genotypes carried by individuals increased with distance from Africa [68]. These results highlight the interaction between drift and purifying selection by showing that deleterious alleles previously maintained at low frequencies by purifying selection may have surfi to higher frequencies in populations at the edge of the wave expanding out of Africa, due to stronger drift [53, 68, 69]. Together, these studies suggest that demographic history has played a central role in shaping differences in the genetic architecture of disease between human populations through its effects on the frequency of deleterious alleles [64, 70].

**Favoring advantageous variants to increase adaptation**

Besides the interplay between drift and selection to remove deleterious mutations, other *de novo* or already existing variants can be advantageous and can increase in population frequency through various forms of positive and balancing selection [23–28, 71, 72]. Humans occupy diverse habitats and have gone through many different cultural and technological transitions; human populations have had to adapt to such shifts in habitat and mode of subsistence [25]. Dissecting the legacy of past genetic adaptation is thus key to identifying the regions of the genome underlying the broad morphological and physiological diversity observed across populations, and to increasing our understanding of the genetic architecture of adaptive phenotypes in health and disease.

**Positive selection targets mendelian and complex traits**

Positive selection can manifest in different guises: from the classic, hard-sweep model, in which a new mutation can confer an immediate fitness benefit (Fig. 1a), to alternative models of genetic adaptation, such as selection on standing variation or polygenic adaptation [73, 74], with each type of selection leaving a specific molecular signature in the targeted region (reviewed in [23, 26]). Most studies have focused on signals of positive selection according to the hard-sweep model, providing insight into the nature of adaptive phenotypes (see [23, 24, 26, 29, 31, 72, 75–77] and references therein). These phenotypes range from Mendelian traits (or almost so)—including the largely supported lactase persistence trait in various populations [78–82] and traits relating to infectious disease resistance (e.g., G6PD, DARC, FLIIT2 in particular (reviewed in [76])—to complex traits, such as skin pigmentation [83–86], adaptation to climate variables or high altitude [87–93], and the immune response and host–pathogen interactions [24, 29, 31, 77, 94–107]. These examples reveal the potent selective pressures that have been exerted by nutritional resources, climatic conditions, and infectious agents since humans first began to spread over the globe [29, 31, 72, 77, 96, 108].

Many selection signals were detected by candidate-gene approaches, based on *a priori* choices of the genes and functions to be investigated. However, a large number of genome-wide scans for positive selection have identified several hundred genomic regions displaying selection signals, consistent with the likely presence in these regions of beneficial, functional variants [28, 37, 109–124]. For example, Grossman and coworkers identified about 400 candidate regions subject to selection, using whole-genome sequencing data from the 1000 Genomes Project [28]. These regions either contain genes involved in skin pigmentation, metabolism and infectious disease resistance, or overlap with elements involved in regulatory functions, such as long intergenic noncoding RNAs and expression quantitative trait loci (eQTL). The presence of non-synonymous variants in less than 10 % of the candidate-selected regions suggests that regulatory variation has played a predominant role in recent human adaptation and phenotypic variation [28], as previously suggested [125–128].

The large number of studies searching for selection signals contrasts with the much smaller number of studies trying to determine when selection effects occurred [83, 129, 130]. Nevertheless, such studies could identify specific time periods corresponding to abrupt changes in environmental pressures. Studies aiming to date the lactase persistence allele in Europe have suggested that this allele was selected in farmers some 6000 to 11,000 years ago [79, 81, 95, 129, 130], although estimates based on ancient DNA point to a more recent time [131, 132] (see below). A recent study, using an approximate Bayesian computation framework, found that skin pigmentation alleles were generally much older than alleles involved in autoimmune disease risk, whose ages are consistent with selection during the spread of agriculture [129]. A report suggesting that many selective events targeting innate immunity genes have occurred in the last 6000 to 13,000 years [95] provides additional support for the notion that the adoption of agriculture and animal domestication modified human exposure to pathogens, leading to genetic adaptations of immune response functions.

Selection studies have thus increased our knowledge of the nature of several adaptive phenotypes at different timescales (Box 1), but the relative importance of selection according to the classic sweep model remains unclear. Several studies have reported the prevalence of classic sweeps for human adaptation to be non-negligible [28, 109–113, 115–118, 122], whereas others have suggested that such sweeps are rare and that the corresponding
Unlike other forms of selection, balancing selection can preserve functional diversity fluctuating over time and space in specific populations or species, and pleiotropy [27, 136, 137]. Unlike other forms of selection, balancing selection can maintain functional diversity over periods of millions of years because selection conditions remain constant over time and are strong enough to avoid the loss of selected polymorphisms due to drift. In some cases, polymorphisms subject to balancing selection can persist during speciation events, resulting in trans-species polymorphism (long-term balancing selection; Fig. 1b). In other cases, balancing selection may occur only in particular species or populations, owing to specific environmental pressures (see [27, 136] and references therein). Until a few years ago, evidence for the action of balancing selection was restricted to a few loci, including the sickle cell hemoglobin polymorphism (HbS), which protects against malaria in the heterozygous state [138], and several genes of the major histocompatibility complex (MHC, or HLA in humans), which presents intracellular peptides to cells involved in immune surveillance and triggers immune responses against diverse pathogens [139–141].

Recent studies, bolstered by the whole-genome sequence data published for humans and other species, have suggested that balancing selection is more prevalent than previously thought (see [27] for a review). Several studies searching for the occurrence of trans-species polymorphism have shown that advantageous variants in the human population may have been inherited from distant ancestral species [142–145]. For example, functional diversity in ABO blood group has been maintained across primates for millions of years, probably due to host–pathogen co-evolution [142]. Likewise, a scan of long-term balancing selection in the genomes of humans and chimpanzees has detected 125 regions containing trans-species polymorphisms, principally in genes involved in immune function, such as IGFBP7 and membrane glycoprotein genes; these findings suggest that there has long been functional variation in response to pressures exerted by pathogens in these species [144]. Other studies have searched for balancing selection within humans through the use of genome-wide approaches or by focusing on particular gene families. Selection signatures have been detected in multiple regions, including the KIR gene regions (KIR genes are known to co-evolve with their HLA ligands [146]), and regions encoding various molecules involved in cell migration, host defense, or innate immunity [146–155]. These studies indicate that, despite its low occurrence, balancing selection has maintained functional diversity at genes involved in functions relating to the immune response, as observed for other types of selection [24, 29, 31, 77, 103].

**Tracking selection signatures from ancient DNA data**

Population genetics methods can be used to estimate the approximate age and selection coefficient of adaptive mutations from data from modern human populations, with various degrees of confidence. However, the use of ancient human samples from different time periods is
making it possible to determine how rapidly the frequency of adaptive mutations has increased in populations. Until a few years ago, ancient DNA data were available only for single individuals or specimens, limiting the analysis to questions of comparative genomics. We learned a great deal about the degree of admixture between modern humans and ancient hominins, such as Neanderthals and Denisovans, a topic that has been reviewed elsewhere [16, 17, 156–158]. These studies have also revealed the existence of advantageous “archaic” variants in the genomes of modern humans [16, 158]. These variants, which were acquired through admixture with archaic humans, have improved adaptation and survival in modern humans (Fig. 1c, Box 2).

However, much less is known about genetic diversity levels in populations of modern humans from different eras, such as the Paleolithic and Neolithic periods. Deep sequencing is making it possible to sequence multiple samples per species or population, opening up new possibilities for the analysis of ancient DNA data within a population genetics framework (see [156] for a review). For example, in one recent study, 230 human samples from West Eurasia dating from between 8500 and 2300 years ago were sequenced [132]. The authors searched for abrupt changes in allele frequencies over time across the genome. They identified 12 loci containing variants with frequencies that rapidly increased over time, consistent with positive selection. The lactase persistence variant yielded one of the strongest signals and appeared to have reached appreciable frequencies in Europe only recently (less than 4000 years ago), as previously suggested [131]. The other strong signals identified were either directly or indirectly related to diet, corresponding to genes encoding proteins involved in fatty acid metabolism, vitamin D levels, and celiac disease, or corresponded to genes involved in skin pigmentation [132]. Interestingly, the authors also detected strong selection signals in immunity-related genes, such as the TLR1–TLR6–TLR10 gene cluster, which is essential for the induction of inflammatory responses and is associated with susceptibility to infectious diseases [159, 160]. Thus, ancient DNA studies can help us to understand the mode of selection following changes in human lifestyle, and the extent to which such selective events increased the frequency of functional alleles associated with specific traits or disease conditions [131, 132, 161, 162].

**Insight into rare and common diseases from natural selection**

Genes associated with Mendelian or complex diseases would be expected to be subject to unequal selective pressures. We can therefore use selection signatures to predict the involvement of genes in human disease [11, 12, 32, 37, 115, 163]. Mendelian disorders are typically severe, compromising survival and reproduction, and are caused by highly penetrant, rare deleterious mutations. Mendelian disease genes should therefore fit the mutation–selection balance model, with an equilibrium between the rate of mutation and the rate of risk allele removal by purifying selection [12]. The use of population genetics models is less straightforward when it comes to

---

**Box 2. Archaic admixture and disease risk in modern humans**

One source of adaptive genetic variation is variants from donor populations that have undergone introgression into recipient populations (or species) through interbreeding or admixture, a phenomenon known as adaptive introgression. During the dispersals of modern humans through Eurasia, they encountered other human forms, such as Neanderthals and Denisovans [16, 17, 156–158]. Analyses of whole-genome sequences from these hominins [191, 192] have shown that admixture between modern and archaic humans occurred, with modern Eurasian genomes comprising 1–6% of Neanderthal or Denisovan ancestry (Fig. 1c) [191, 193–197]. However, these estimations are averaged across the genome, and it has been shown that the degree of archaic ancestry in the modern human genome varies considerably [194, 198, 199]. A systematic search for regions of Neanderthal ancestry in the genomes of modern humans has shown that, globally, purifying selection has acted against Neanderthal introgression, particularly in protein-coding regions [194]. However, some regions of the genome can present a high degree of Neanderthal ancestry, which can be as high as 62% in Asian and 64% in European populations [194]. Some of the regions presenting the highest degree of Neanderthal ancestry have been found to overlap alleles that have been associated with phenotypes of medical relevance, such as lupus, biliary cirrhosis, Crohn’s disease, optic-disc size, smoking behavior, IL-18 levels, type 2 diabetes, and hypercoagulation [194, 200]. In some cases, the introgression of archaic segments into the genomes of modern humans appears to have been adaptive [194, 198]. Cases of adaptive introgression from Neanderthals or Denisovans have been reported for several genes (see [16] for a complete list), such as EPAS1, involved in human adaptation to life at high altitude [201], BNC2, involved in skin pigmentation [194, 198], and an increasing number of genes involved in immunity to infection, including several HLA genes [202], STAT2 [203], the OAS gene cluster [204, 205], and the TLR6–TLR1–TLR10 gene cluster [95, 206]. Collectively, these studies suggest that archaic admixture has been an important source of adaptive variation but also that modern humans have inherited archaic variation that today influences complex disease risk [158, 200].
predicting the genes involved in complex disease risk. Models of adaptive evolution based on positive or balancing selection apply to a few Mendelian traits or disorders, most notably, but not exclusively, those related to malaria resistance (reviewed in [76, 98]). However, the complex patterns of inheritance observed for common diseases, including incomplete penetrance, late onset and gene-by-environment interactions, make it more difficult to decipher the connection between disease risk and fitness [12].

**Purifying selection, rare variants, and severe disorders**

According to population genetics theory, strongly deleterious mutations are rapidly removed from the population by purifying selection, whereas mildly deleterious mutations generally remain present, albeit at low frequencies, depending on population sizes and fitness effects. Genome-wide studies are providing increasing amounts of support for these predictions, as “essential” genes—identified as such on the basis of association with Mendelian diseases or experimental evidence from model organisms—are enriched in signs of purifying selection [32, 37, 115, 164]. Purifying selection has also been shown to be widespread in regulatory variation, acting against variants with large effects on transcription, conserved noncoding regions of the genome, and genes that are central in regulatory and protein–protein interaction networks [8, 10, 165–171].

Mutations associated with Mendelian diseases or with deleterious effects on the phenotype of the organism are generally rare and display familial segregation, but such mutations may also be restricted to specific populations [11]. This restriction, in some cases, may be due to a selective advantage provided by the disease risk allele (e.g., the sickle cell allele in populations exposed to malaria [98]), but it mostly reflects a departure from the mutation–selection balance. Small population sizes or specific demographic events may randomly increase the frequency of some disease risk alleles, because too little time has elapsed for purifying selection to remove them from the population, as observed in French Canadians, Ashkenazi Jews, or Finns [11, 66, 67].

According to these principles of population genetics, searches for genes or functional elements evolving under strong purifying selection can be used to identify the genes of major relevance for survival, mutations of which are likely to impair function and lead to severe clinical phenotypes. In this context, the immune response and host defense functions appear to be the prime targets of purifying selection [37, 95, 102]. For example, a recent study based on whole-genome sequences from the 1000 Genomes Project estimated the degree to which purifying selection acted on ~1500 innate immunity genes. The genes of this class, taken as a whole, were found to have evolved under globally stronger purifying selection than the rest of the protein-coding genome [95]. This study also assessed the strength of selective constraints in the different innate immunity modules, organizing these constraints into a hierarchy of biological relevance, and providing information about the degree to which the corresponding genes were essential or redundant [95].

Population genetics has also facilitated the identification of immune system genes and signaling pathways that fulfill essential, non-redundant functions in host defense, variants of which are associated with severe, life-threatening infectious diseases (for examples, see [94, 95, 101, 106], and for reviews [29, 103, 172, 173]). This is well illustrated by the cases of STAT1 and TRAF3; they belong to the 1 % of genes presenting the strongest signals of purifying selection at the genomewide level [95], and mutations in these genes have been associated with severe viral and bacterial diseases, Mendelian susceptibility to mycobacterial disease, and herpes simplex virus 1 encephalitis [174, 175]. Using the paradigm of immunity and infectious disease risk, these studies highlight the value of population genetics as a complement to clinical and epidemiological genetic studies, for determining the biological relevance of human genes in *natura* and in predicting their involvement in human disease [29, 103, 173, 176].

**Genetic adaptation, common variants, and complex disease**

The relationship between selection and complex disease risk is less clear than for Mendelian disorders, but patterns are beginning to emerge. Genes associated with complex disease display signs of less pervasive purifying selection than Mendelian disease genes [32, 173], and are generally enriched in signals of positive selection [23, 28, 32, 37, 110, 122, 169]. There is also increasing evidence to suggest that genetic adaptations can alter complex disease susceptibility, and the population distribution of common susceptibility alleles is unlikely to result from neutral processes alone [12, 91, 177–179]. For example, the difference in susceptibility to hypertension and metabolic disorders between populations is thought to result from past adaptation to different environmental pressures [91, 179, 180]. Another study characterized the structure of complex genetic risk for 102 diseases in the context of human migration [178]. Differences between populations in the genetic risk of diseases such as type 2 diabetes, biliary liver cirrhosis, inflammatory bowel disease, systemic lupus erythematosus, and vitiligo could not be explained by simple genetic drift, providing evidence of a role for past genetic adaptation [178]. Likewise, Grossman and coworkers found overlaps between their candidate positively selected regions and genes associated with traits or diseases in
GWAS [28], including height, and multiple regions associated with infectious and autoimmune disease risks, including tuberculosis and leprosy.

Like purifying selection, positive selection is prevalent among genes related to immunity and host defense [24, 37, 95, 109, 112, 115, 181]. Notable examples of immunity-related genes evolving in an adaptive manner, through different forms of positive or balancing selection, and reported to be associated with complex traits or diseases include: TLR1 and TLR5, which have selection signals that seem to be related to decreases in NF-kB signaling in Europe and Africa, respectively [28, 94, 95]; many genes involved in malaria resistance in Africa and Southeast Asia [98, 100]; type-III interferon genes in Europeans and Asians, related to higher levels of spontaneous viral clearance [101, 182]; LARGE and IL21, which have been implicated in Lassa fever infectivity and immunity in West Africans [181]; and components of the NF-kB signaling pathway and inflammasome activation related to cholera resistance in a population from the Ganges river delta [97]. These cases of selection related to infectious disease and many others (see [29–31, 96, 103] for reviews and references therein) indicate that the pressures imposed by infectious disease agents have been paramount among the different threats faced by humans [183]. They also highlight the value of population genetics approaches in elucidating the variants and mechanisms underlying complex disease risk.

Changes in selective pressures and advantageous/deleterious variants

Most of the rare and common variants associated with susceptibility to disease in modern populations have emerged through neutral selection processes [184]. However, there is increasing evidence to suggest that, following changes in environmental variables or human lifestyle, alleles that were previously adaptive can become “maladaptive” and associated with disease risk [12, 13, 29, 30, 105]. For example, according to the popular “thrifty genotype” hypothesis based on epidemiological data, the high prevalence of type 2 diabetes and obesity in modern societies results from the selection of alleles associated with efficient fat and carbohydrate storage during periods of famine in the past. Increases in food abundance and a sedentary lifestyle have rendered these alleles detrimental [185]. The strongest evidence that past selection can lead to present-day maladaptation and disease susceptibility is provided by infectious and inflammatory disorders [12, 29–31, 77, 105]. According to the hygiene hypothesis, decreases in the diversity of the microbes we are exposed to, following improvements in hygiene and the introduction of antibiotics and vaccines, have led to an imbalance in the immune response, with alleles that helped us to fight infection in the past now being associated with a higher risk of inflammation or autoimmunity [105].

Population genetics studies have provided strong support for the hygiene hypothesis, by showing that genetic variants associated with susceptibility to certain autoimmune, inflammatory, or allergic diseases, such as inflammatory bowel disease, celiac disease, type 1 diabetes, multiple sclerosis, and psoriasis, also display strong positive selection signals [29, 30, 106, 186–188]. For example, genes conferring susceptibility to inflammatory diseases have been shown to be enriched in positive selection signals, with the selected loci forming a highly interconnected protein–protein interaction network, suggesting that a shared molecular function was adaptive in the past but now affects susceptibility to various inflammatory diseases [187]. Greater protection against pathogens is thought to be the most likely driver of past selection, but it has been suggested that other traits, such as anti-inflammatory conditions in utero, skin color, and hypoxic responses, might account for the past selective advantage of variants, contributing to the higher frequencies of chronic disease risk alleles in current populations [30]. Additional molecular, clinical, and epidemiological studies are required to support this hypothesis, but these observations highlight, more generally, the evolutionary trade-offs between past selection and current disease risk in the context of changes in environmental pressures and human lifestyle.

Conclusions and future directions

Population genetics offers an alternative approach, complementary to clinical and epidemiological genetic studies, for the identification of disease risk alleles/gens, the characterization of their properties, and the understanding of the relative contributions of human genetic variation to rare, severe disorders and complex disease phenotypes. Recent studies have shown that both ancient and recent demographic changes have modified the burden of rare, deleterious variants segregating in the population, whereas the population frequencies of other variants have increased because they conferred advantages in terms of better survival and reproduction.

These studies have made a major contribution, but further theoretical and empirical work is needed. Rare-variant studies should consider different fitness and dominance effects, epistatic interactions, and detailed demographic modeling to evaluate the potential impact of local changes in population size and admixture on the efficiency of purifying selection. Furthermore, rare-variant association studies involving complex traits or diseases should seek to account for the evolutionary forces that affect genetic architecture, such as selection and population demography, and integrate elaborated models of
population genetics that consider the relationship between allele frequency and effect size and the distribution of phenotypes, as recently reported [189]. Independently of the complex interactions between demography and selection, additional sequence-based studies are required to catalog rare variants in different worldwide populations (including isolated populations), focusing not only on point mutations but also on indels, inversions, or copy-number variation, and evaluate their contribution to disease risk.

Studies of genetic adaptation, particularly those aiming to make connections with disease in populations historically exposed to different environmental variables, should generate whole-genome data for different worldwide populations with greatly contrasting demographic histories, lifestyles, and subsistence strategies. There is also a need to develop and improve statistical approaches to facilitate the detection of positive selection following alternative modes of genetic adaptation, such as selection on standing variation, polygenic adaptation, and adaptive introgression. These selection studies, if combined with data for molecular phenotypes (e.g., gene expression, protein and metabolite levels, epigenetic marks) and organismal phenotypes (in health and disease), should provide great insight into adaptive phenotypes of major relevance in human evolution and the genetic architecture of rare and common human diseases.

**Abbreviations**

BNC2: Basonuclin-2; DARC: Duffy antigen/chemokine receptor; eQTL: expression quantitative trait loci; ExAC: Exome Aggregation Consortium; FUT2: Fucosyltransferase 2; G6PD: Glucose-6-phosphate dehydrogenase; GWAS: Genome-wide association studies; HbS: sickle cell haemoglobin; HLA: Human leukocyte antigen; IGFBP7: Insulin-like growth factor-binding protein 7; IL21: Interleukin 21; KIR: Killer-cell immunoglobulin-like receptors; LARGE: LARGE xylosyl- and glucuronyltransferase 1; MHC: Major histocompatibility complex; NF-kB: nuclear factor NF-kB; OAS1: O-Acetylserine; SDSL: singleton density score; STAT1: Signal transducer and activator of transcription 1; TLR: Toll-like receptors; TRAF3: Tumor necrosis factor receptor-associated factor

**Acknowledgements**

I would like to thank A. Kousathanas, M. Lopez, M. Rotival, M. Silvert, and J. Teixeira for helpful comments and discussions.

**Funding**

This work was supported by the Institut Pasteur, the Centre National de la Recherche Scientifique (CNRS), the French Government’s Investissement d’Avenir program, Laboratoire d’Excellence “Integrative Biology of Emerging Infectious Diseases” (grant no. ANR-10-LABX-62-BEID), and the European Research Council under the European Union’s Seventh Framework Programme (FP7/2007–2013)/ERC Grant Agreement No. 281297.

**Competing interests**

The author declares that he has no competing interests.

**Author details**

1. Human Evolutionary Genetics Unit, Department of Genomes & Genetics, Institut Pasteur, Paris 75015, France.
2. Centre National de la Recherche Scientifique, URA3012, Paris 75015, France.
3. Center of Bioinformatics, Biostatistics and Integrative Biology, Institut Pasteur, Paris 75015, France.

**Published online:** 07 November 2016

**References**

1. Pritchard JK. Are rare variants responsible for susceptibility to complex diseases? Am J Hum Genet. 2001;69:124–37.
2. Pritchard JK, Cox NJ. The allelic architecture of human disease genes: common disease-common variant... or not? Hum Mol Genet. 2002;11:2417–23.
3. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. Nature. 2009;461:747–53.
4. McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little JI, Ioannidis JP, et al. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nat Rev Genet. 2008;9:356–69.
5. Reich DE, Lander ES. On the allelic spectrum of human disease. Trends Genet. 2001;17:502–9.
6. Zwick ME, Cutler DJ, Chakravarti A. Patterns of genetic variation in Mendelian and complex traits. Annu Rev Genomics Hum Genet. 2000;1:387–407.
7. Schork NJ, Murray SS, Frazer KA, Topol EJ. Common vs. rare allele hypotheses for complex diseases. Curr Opin Genet Dev. 2009;19:212–9.
8. Bodmer W, Bonilla C. Common and rare variants in multifactorial susceptibility to common diseases. Nat Genet. 2008;40:695–701.
9. Goldstein DB. Common genetic variation and human traits. N Engl J Med. 2009;360:1696–8.
10. Zhu Q, Ge D, Maia JM, Zhu M, Petrovski S, Dickson SP, et al. A genome-wide comparison of the functional properties of rare and common genetic variants in humans. Am J Hum Genet. 2011;88:458–68.
11. Lu YF, Goldstein DB, Angrist M, Cavaliere G. Personalized medicine and human genetic diversity. Cold Spring Harb Perspect Med. 2014;4:a008581.
12. Di Rienzo A. Population genetics models of common diseases. Curr Opin Genet Dev. 2006;16:630–6.
13. Crespi BJ. The emergence of human-evolutionary medical genomics. Evol Appl. 2011;4:292–314.
14. Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, et al. An integrated map of genetic variation from 1,092 human genomes. Nature. 2012;491:56–65.
15. 1000 Genomes Project Consortium. A global reference for human genetic variation. Nature. 2013;514:68–74.
16. Racimo F, Sankararaman S, Nielsen R, Huerta-Sanchez E. Evidence for archaic adaptive introgression in humans. Nat Rev Genet. 2015;16:559–71.
17. Kelso J, Prufer K. Ancient humans and the origin of modern humans. Curr Opin Genet Dev. 2014;29:133–8.
18. Veeramah KR, Hannier MF. The impact of whole-genome sequencing on the reconstruction of human population structure. Nat Rev Genet. 2014;15:149–62.
19. Novembre J, Ramachandran S. Perspectives on human population structure at the cusp of the sequencing era. Annu Rev Genomics Hum Genet. 2011;12:245–74.
20. Henn BM, Cavalli-Sforza LL, Feldman MW. The great human expansion. Proc Natl Acad Sci U S A. 2012;109:17758–64.
21. Sousa V, Peischl S, Excoffier L. Impact of range expansions on current human genomic diversity. Curr Opin Genet Dev. 2014;29:22–30.
22. Lohmueller KE. The distribution of deleterious genetic variation in human populations. Curr Opin Genet Dev. 2014;29:139–46.
23. Nielsen R, Hellman I, Hubisz M, Bustamante C, Clark AG. Recent and ongoing selection in the human genome. Nat Rev Genet. 2007;8:587–68.
24. Sabeti PC, Schraiber JG, Fry B, Lohmueller KE, Marzドラマro V, Shih WS, et al. Positive natural selection in the human lineage. Science. 2006;312:1614–20.
25. Jeong C, Di Rienzo A. Adaptations to local environments in modern human populations. Curr Opin Genet Dev. 2014;29:1–5.
26. 1000 Genomes Project Consortium. Finding the missing heritability of complex diseases? Am J Hum Genet. 2002;69:124–7.
27. Pritchard JK, Cox NJ. The allelic architecture of human disease genes: common disease-common variant... or not? Hum Mol Genet. 2002;11:2417–23.
28. Brinkworth JF, Quintana-Murci L. From evolutionary genetics to human immunology: how selection shapes host defence genes. Curr Opin Immunol. 2011;23:17–30.
29. Brinkworth JF, Barreiro LB. The contribution of natural selection to present-day susceptibility to chronic inflammatory and autoimmune disease. Curr Opin Immunol. 2014;31:66–78.
30. Karlsson AK, Kvitikowski DP, Sabeti PC. Natural selection and infectious disease in human populations. Nat Rev Genet. 2014;15:379–93.
32. Bliksrøde, M., O., Hermann, L., Boyko, A.R., Indap, A., Kosiol, C., et al. Natural selection on genes that underlie human disease susceptibility. Curr Biol. 2008;18:883–9.

33. Eyre-Walker A, Keightley PD. High genomic deleterious mutation rates in hominids. Nature. 1999;397:344–7.

34. Kryukov GV, Pennacchio LA, Suryaev SR. Most rare missense alleles are deleterious in humans: Implications for complex disease and association studies. Am J Hum Genet. 2007;80:727–39.

35. Boyko AR, Williamson SH, Indap AR, Degenhardt JD, Hernandez RD, Lohmueller KE, et al. Assessing the evolutionary impact of amino acid mutations in the human genome. PLoS Genet. 2008;4:e1000883.

36. Eyre-Walker A, Keightley PD. The distribution of fitness effects of new mutations. Nat Rev Genet. 2007;8:610–8.

37. Bustamante CD, Felder-Alon A, Williamson S, Nielsen R, Hubisz MT, Glanowski S, et al. Natural selection on protein-coding genes in the human genome. Nature. 2005;437:1153–7.

38. Kimura M, Maruyama T, Crow JF. The mutation load in small populations. Genetics. 1963;48:1303–12.

39. Ohta T. Slightly deleterious mutant substitutions in evolution. Nature. 1973;246:96–8.

40. Akashi H, Osada N, Ohta T. Weak selection and protein evolution. Genetics. 2005;171:1551–69.

41. Tennessen JA, Bigham AW, O. et al. Proportionally more deleterious genetic variation in European than in East Asian and African populations. Nature. 2012;485:47–54.

42. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature. 2016;536:285–91.

43. Maher MC, Uricchio LH, Torgerson DG, Hernandez RD. Population genetics of rare variants and complex diseases. Hum Hered. 2012;74:118–30.

44. Simons YB, Turchin MC, Pritchard JK, Sella G. The deleterious mutation load is insensitive to recent population history. Nat Genet. 2014;46:220–4.

45. Lohmueller KE. The impact of population demography and selection on the genetic architecture of complex traits. PLoS Genet. 2014;10:e1004379.

46. Fu W, O. et al. The distribution and medical impact of loss-of-function variants in the Finnish founder population. PLoS Genet. 2014;10:e1004494.

47. Henn BM, Botigue LR, Bustamante CD, Clark AG, Gravel S. Estimating the mutation load in human genomes. Nat Rev Genet. 2015;16:333–43.

48. Casals F, Hodgkinson A, Hussin J, Idaghdour Y, Bruat V, de Maillard T, et al. Whole-exome sequencing reveals a rapid change in the frequency of rare functional variants in a founding population of humans. PLoS Genet. 2017;13:e1006815.

49. Lopatka S, Currier SJ, Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet. 2014;46:310–5.
189. Uricchio LH, Zaitlen NA, Ye CJ, Witte JS, Hernandez RD. Selection and explosive growth alter genetic architecture and hamper the detection of causal rare variants. Genome Res. 2016;26:863–73.

190. Field Y, Boyle EA, Telis N, Gao Zu, Gaulton KJ, Golan D, et al. Detection of human adaptation during the past 2,000 years. Science. Oct 13, 2016. Available from: https://www.ncbi.nlm.nih.gov/pubmed/27738015 [Epub ahead of print]

191. Prufer K, Racimo F, Patterson N, Jay F, Sankararaman S, Sawyer S, et al. The complete genome sequence of a Neanderthal from the Altai Mountains. Nature. 2014;505:43–9.

192. Meyer M, Kircher M, Gansauge MT, Li H, Racimo F, Mallick S, et al. A high-coverage genome sequence from an archaic Denisovan individual. Science. 2012;338:222–6.

193. Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, et al. A draft sequence of the Neandertal genome. Science. 2010;328:710–22.

194. Sankararaman S, Mallick S, Dannemann M, Prufer K, Kelso J, Paabo S, et al. The genomic landscape of Neandertal ancestry in present-day humans. Nature. 2014;507:354–7.

195. Reich D, Green RE, Kircher M, Krause J, Patterson N, Durand EY, et al. Genetic history of an archaic hominin group from Denisova Cave in Siberia. Nature. 2010;468:1053–60.

196. Reich D, Patterson N, Kircher M, Delfin F, Nandineni MR, Pugach I, et al. Denisovan admixture and the first modern human dispersals into Southeast Asia and Oceania. Am J Hum Genet. 2011;89:516–28.

197. Vernot B, Akey JM. Complex history of admixture between modern humans and Neandertals. Am J Hum Genet. 2015;96:448–53.

198. Vernot B, Akey JM. Resurrecting surviving Neandertal lineages from modern human genomes. Science. 2014;343:1017–21.

199. Sankararaman S, Mallick S, Patterson N, Reich D. The combined landscape of Denisovan and Neandertal ancestry in present-day humans. Curr Biol. 2016;26:1241–7.

200. Simonti CN, Vernot B, Bastarache L, Bottinger E, Carrell DS, Chisholm RL, et al. The phenotypic legacy of admixture between modern humans and Neandertals. Science. 2016;351:737–41.

201. Huerta-Sanchez E, Jin X, Anan, Bianba Z, Peter BM, Vinckenbosch N, et al. Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. Nature. 2014;512:194–7.

202. Abi-Rached L, Jobin MJ, Kulkarni S, McWhinnie A, Dalva K, Gragert L, et al. The shaping of modern human immune systems by multiregional admixture with archaic humans. Science. 2011;334:89–94.

203. Mendez FL, Watkins JC, Hammer MF. A haplotype at STAT2 introgressed from neandertals and serves as a candidate of positive selection in Papua New Guinea. Am J Hum Genet. 2012;91:265–74.

204. Mendez FL, Watkins JC, Hammer MF. Global genetic variation at OAS1 provides evidence of archaic admixture in Melanesian populations. Mol Biol Evol. 2012;29:1513–20.

205. Mendez FL, Watkins JC, Hammer MF. Neandertal origin of genetic variation at the cluster of OAS immunity genes. Mol Biol Evol. 2013;30:798–801.

206. Dannemann M, Andris AM, Keljo J. Introgresion of Neandertal- and Denisovan-like haplotypes contributes to adaptive variation in human Toll-like receptors. Am J Hum Genet. 2016;98:22–33.