From evolutionary genetics to human immunology: how selection shapes host defence genes

Luis B. Barreiro[*‡] *and Lluís Quintana-Murci*[*]

Abstract | Pathogens have always been a major cause of human mortality, so they impose strong selective pressure on the human genome. Data from population genetic studies, including genome-wide scans for selection, are providing important insights into how natural selection has shaped immunity and host defence genes in specific human populations and in the human species as a whole. These findings are helping to delineate genes that are important for host defence and to increase our understanding of how past selection has had an impact on disease susceptibility in modern populations. A tighter integration between population genetic studies and immunological phenotype studies is now necessary to reveal the mechanisms that have been crucial for our past and present survival against infection.

Innate immunity
Non-specific and evolutionarily ancient mechanisms that form the first line of host defence against infection. Innate immunity, which is inborn and does not involve memory, provides immediate defence and is found in all classes of plants and animals.

[*]*Human Evolutionary Genetics, Institut Pasteur, Centre National de la Recherche Scientifique URA3012, 25 Rue du Dr Roux, Paris 75015, France.*
[‡]*Department of Human Genetics, University of Chicago, 920 East 58th Street, Chicago, Illinois 60637, USA.*
Correspondence to L.Q.-M.
e-mail: quintana@pasteur.fr
doi:10.1038/nrg2698

"Nature is the best doctor: she cures three out of four illnesses, and she never speaks ill on her colleagues."
Louis Pasteur

We, contemporary humans, are indeed the descendants of those peoples that were 'cured', and selected, by nature in the past. The burden of infectious diseases has been massive throughout history, with life expectancy not exceeding 20–25 years of age until the advent of Pasteur's microbial theory of disease and the resulting control of infections by hygiene, vaccines and antibiotics[1]. Curiously, Charles Darwin never made substantial mention of infectious diseases as a driving force in natural selection, despite the fact that his contemporaries Louis Pasteur and Robert Koch discovered that microbes caused the most serious human diseases. It was only much later that John B. S. Haldane linked the two concepts — natural selection and infectious disease — with the proposal that red-blood-cell disorders (thalassaemia) could protect from malaria[2].

The evolutionary dynamics of host–pathogen interactions lead to constant selection for adaptation and counter-adaptation in the two competing species. Throughout evolution, animals and plants have developed complex immune defence mechanisms to combat microbial infections. Understanding the evolution of immune systems in general has been the focus of intense study by different disciplines, such as comparative immunology, evolutionary immunobiology and, more recently, ecological immunology[3–8]. In particular, comparative studies examining the molecular and functional features of immune systems in multiple organisms have greatly improved our knowledge of the origins and evolution of innate immunity and adaptive immunity, the selective pressures exerted on these systems, and the functional properties of the modules and molecules that are involved in innate and adaptive immune defence mechanisms (BOX 1). Here, we do not attempt a comprehensive overview of these topics, for which outstanding reviews have been published elsewhere[3–6,9,10]. Instead, we focus on what population and evolutionary genetic studies can tell us about the evolution of the human immune system.

Although interest in natural selection is not new[2,11], the past few years have seen an increase in studies that aim to characterize how selection, in its different forms and intensities (BOX 2), has targeted regions of the human genome[12–14]. The re-emergence of the field has been particularly bolstered by the advent of genome-wide surveys of genetic variation based on genotyping and resequencing data in human populations, the expanding repertoire of complete genome sequences from several species, and the development of theoretical models in population genetics. In this Review, we discuss some of the major findings regarding how natural selection has shaped the evolution of genes involved in immune defence mechanisms in humans. We must first point out that the definition of

Box 1 | **Main receptors of the human immune system**

Innate immunity[6,9,140] relies on pattern-recognition receptors (PRRs), which recognize conserved and largely invariant microbial molecules that are essential for microbial physiology (for example, lypopolysaccharides, flagellin and nucleic acids). These microbial molecules are often referred to as pathogen-associated molecular patterns (PAMPs)[9,10,140]. PRRs can be located on or in cells, such as macrophages, dendritic cells or natural killer cells, or can be secreted into the bloodstream and tissue fluids. After sensing a PAMP, PRRs trigger diverse mechanisms that initiate inflammatory and immune responses, and help to mount an adaptive immune response[10]. These mechanisms include opsonization, activation of complement cascades, phagocytosis, activation of proinflammatory pathways, release of antimicrobial peptides and induction of apoptosis.

There are several functionally distinct classes of PRRs, the best characterized being Toll-like receptors (TLRs)[9,141–143]. TLRs, which can be expressed on the cell surface or in intracellular compartments, detect microorganisms by sensing viral nucleic acids and several bacterial products[9,144]. At the cell surface, another family of PRRs, the C-type lectin receptors (CLRs), also bind and take up microbial components through the sensing of sugar motifs[145]. Bacteria and viruses that invade the cytosol are detected by two other families: the NOD-like receptors (NLRs)[146] and the RIG I-like receptors (RLRs)[147], which induce cytokine production and cell activation. In addition, secreted PRRs, such as collectins, ficolins and pentraxins, function in the circulation and in tissue fluids (acute-phase proteins and the complement system) and participate in the lysis or opsonization of microbes[10].

Adaptive immunity[4,5] is mediated by T cell receptors (TCRs) and B cell receptors (immunoglobulin receptors). The genes encoding the antigen receptors of T and B lymphocytes are assembled from variable and constant fragments through recombination-activating gene (RAG) protein-mediated somatic recombination, which yields a diverse repertoire of receptors[148]. This diversity is further increased by mechanisms such as non-templated nucleotide addition, gene conversion and (in the case of B cells) somatic hypermutation, which allow great variability in adaptive immune recognition[148]. There are two types of lymphocyte that express antigen receptors: conventional lymphocytes and innate-like lymphocytes. The antigen receptors of conventional lymphocytes are assembled essentially at random and their specificities are not predetermined[148], whereas the assembly process of antigen receptors of innate-like lymphocytes is restricted and their specificities are skewed towards a predefined set of ligands[149]. Microbial antigens are phagocyted by antigen-presenting cells (for example, dendritic cells) and their protein constituents are processed into antigenic peptides, which are presented at the cell surface to conventional T cells by major histocompatibility complex (MHC) class I and/or class II molecules[150,151]. Innate-like T cells recognize non-classical MHC molecules (also known as MHC class Ib molecules), which somehow function as PRRs and present microbial ligands to specialized T cells (for example, CD1 presents bacterial lipids). Conventional B cells recognize almost any antigen by binding to a specific three-dimensional molecular determinant (or epitope), and innate-like B cells produce antibodies with specificities that are biased towards some common bacterial polysaccharides or some self-antigens.

Because the innate and adaptive immune systems cooperate in their actions[5,10], a distinction between the two systems (and the molecules involved) is, in some cases, somewhat artificial[5].

an 'immunity' gene is broad and somewhat ambiguous. Several criteria can be used for defining immunity genes (for example, genes that are expressed specifically in immune tissues, those that show direct interaction with pathogens or their products, and so on)[15], but many genes that are not core elements of the immune system can also have a role in host defence. A clear example is the *HbS* 'sickle-cell' mutation in the β-globin (*HBB*) gene, which provides higher resistance to malaria[11]. Although *HBB* is not an integral part of the immune system *stricto sensu*, it unambiguously has a major role in immunity to infection and host defence. To avoid confusion, in this Review we refer to genes involved in immune defence mechanisms *lato sensu* as 'immunity-related genes'.

Adaptive immunity
A highly adaptable and specific immune response that can adjust to new pathogens (structures) and that retains a memory of prior exposure to them. Thought to have arisen in jawed vertebrates, the adaptive immune system is stimulated by the innate immune system.

We start with an overview of how comparative genomic studies suggest that humans and other primates may have differently adapted to pathogens. We then discuss data from population genetic studies in humans that show that immunity-related genes have been privileged targets of recent selection, followed by a discussion of possible causes of selective pressure. We also address how evolutionary genetic studies can provide important clues for delineating genes that have major roles in host defence and for predicting regions of the genome that are potentially associated with host susceptibility to infection or disease outcome. Finally, we discuss future directions in this field, such as the need to combine genotype data with detailed immunological phenotypes to better elucidate the immunological mechanisms that have been preferentially favoured in humans.

Differences between humans and other primates

There are many interesting differences in infectious disease frequency and severity between humans and non-human primates. For example, several medical conditions, such as HIV progression to AIDS, *Plasmodium falciparum* malaria, late complications in hepatitis B or C, and influenza A symptomatology, affect humans more severely than other primates[16,17]. This suggests interspecies differences in immune responses. In this section, we discuss how genome-wide comparative studies between primate species suggest that some of these differences might result from adaptive evolution.

Differences in protein-coding sequences. By comparing protein-coding sequences between species, comparative studies help to identify proteins that are rapidly evolving in humans and in other primates. These studies, which compare the ratio of functional changes (that is, amino acid substitutions) to neutral changes (that is, silent substitutions) through, for example, d_N/d_S methods[18] (BOX 3), have shown that compared with other protein classes, immunity-related proteins have been preferential targets of positive selection in the primate lineage[19–23], in mammals[24] and in other organisms[25]. Other protein classes that have been privileged targets of positive selection during primate evolution include olfaction, spermatogenesis and sensory perception proteins[21,23,24,26]. However, identifying the genetic basis of phenotypic traits that, for example, distinguish humans from chimpanzees requires finding genes that show accelerated evolution specifically in humans and/or chimpanzees since the time of the last common ancestor of the two species. The number of genes reported as having rapidly evolved in the human species as a whole varies from around 10 to 100 genes[19,22,24,27] (the variation reflects methodological differences among studies — that is, differences in statistical methods and/or the sets of species used).

Probably opposing common wisdom, little evidence exists that genes that have been specifically selected in the human lineage (or the chimpanzee lineage) are enriched for immunity-related functions[19,22,24,26–28]. However, this observation is likely to reflect the low power of d_N/d_S methods, particularly when it comes

McDonald–Kreitman test
The McDonald–Kreitman test compares the ratio of polymorphism (within-species variation) to divergence (between-species variation) at non-synonymous and synonymous sites.

Gene Ontology
A widely used classification system of gene functions and other gene attributes that uses a controlled vocabulary. The ontology covers three domains; cellular components, molecular functions and biological processes.

to detecting lineage-specific selective events between closely related species[24]. Indeed, using different, more powerful approaches, Nielsen *et al.*[23] and Bustamante *et al.*[20] (using a pairwise d_N/d_S method and a modified McDonald–Kreitman test, respectively) provided evidence that defence and/or immunity proteins have been privileged targets of positive selection since humans and chimpanzees started to diverge. The method used by Nielsen *et al.*[23] considers information from the human and chimpanzee lineages merged together, and the test used by Bustamante *et al.*[20] considers both polymorphic and divergence data. These analytical approaches give the methods increased power. However, the limitation of these two methods is that they do not allow the assignment of selection to a specific lineage.

For this Review, we assembled a set of immunity-related genes that were identified as having rapidly evolved in the human and/or chimpanzee lineage. We examined the results of six comparative genome-wide scans for selection in which lineage-specific selection could be assessed[19,22,24,26–28] and then performed Gene Ontology analyses on all of the genes that were reported as rapidly evolving in at least one study. We noted 84 immunity-related genes that showed rapid protein evolution: 17 genes seem to be positively selected only in the human lineage, 59 only in the chimpanzee lineage and 8 in both (see Supplementary information S1 (table)).

The rapid evolution observed at these genes makes them excellent candidates to account for the different way in which humans and chimpanzees respond to infection. Interestingly, among the 84 rapidly evolving immunity-related genes, 30 are HIV-interacting proteins ($p < 0.05$). Because the role of these HIV-interacting proteins is not limited to defence against HIV, the observed enrichment is likely to reflect past selection against older pathogens (for example, ancestral retroviral infections) that triggered immune response mechanisms that were similar to the present-day HIV response mechanisms. Regardless of the causative agent of selection, functional changes in some of these HIV-interacting proteins could, at least partially, explain why chimpanzees, contrarily to humans, can avoid progression to AIDS-like syndromes following HIV or simian immunodeficiency virus (SIV) infection (REFS 29,30, but see REF. 31). Some of the rapidly evolving genes that might account for differences in susceptibility to HIV and SIV between humans and chimpanzees include: the gene encoding the transcription factor HIV type I enhancer binding protein 3 (*HIVEP3*), which activates HIV gene expression by binding to the nuclear factor-κB motif of the HIV-1 long terminal repeat[32]; and chemokine (C-C motif) ligand 4-like 2 (*CCL4L2*), which is a paralogue of the *CCL3L* genes. Copy-number variation in *CCL3L* genes contributes to HIV susceptibility in humans[33] and other primates[34].

Humans and chimpanzees also differ in their susceptibility to *P. falciparum* malaria. In contrast to humans, chimpanzees seem to be immune to infection with this parasite, and although they can become infected by *Plasmodium reichnowii*, they do not seem to develop severe disease[16]. In this context, it is interesting to note the rapid evolution of glycophorin A (*GYPA*) and glycophorin C (*GYPC*) in different primate species (including humans)[24,35–37]. Because both proteins are known to mediate *P. falciparum* erythrocyte invasion[38,39], the distinctive patterns of selection observed at *GYPA* and *GYPC* during primate evolution could explain the increased resistance to *P. falciparum* malaria observed in chimpanzees compared with humans. Another genetic mechanism that is proposed to explain such a difference in resistance to *P. falciparum* malaria is the human-specific loss of *N*-glycolylneuraminic acid, a common primate sialic acid[40].

However, and more generally, the rapidly evolving immunity-related genes in the human and chimpanzee lineages (Supplementary information S1 (table)) can only partly explain the immunological phenotypic differences between the two species. Indeed, the aforementioned comparative studies[19,22,24,26–28] only detect selection that occurs in protein-coding regions. Because phenotypic differentiation between humans and chimpanzees may

Box 2 | **Types of natural selection and their molecular signatures**

Natural selection can take different forms, each of which has a different evolutionary outcome and leaves a distinctive signature in the genomic region targeted[12,13,152–154].

Negative selection
Also known as purifying selection, this refers to the selective removal of deleterious alleles from a population. This is probably the most pervasive form of natural selection acting on genomes. In humans, 38–75% of all new amino acid-altering mutations are estimated to be affected by moderate or strong negative selection[155,156]. The main consequence of negative selection is a local reduction of diversity and an increase of rare alleles when selection is not strong enough to completely eliminate deleterious variants from the population (that is, weak negative selection)[152]. The frequent removal of deleterious variants can also result in the occasional removal of neutral linked variation (particularly in low-recombining regions), a phenomenon known as background selection.

Positive selection
Also known as directional or Darwinian selection, this refers to selection acting upon newly arisen (or previously rare) advantageous mutations. When an advantageous mutation increases in frequency in the population as a result of positive selection, linked neutral variation will be dragged along with it — a process known as genetic hitchhiking. As a consequence, variation that is not associated with the selected allele is eliminated, resulting in a selective sweep that leads to an overall reduction of genetic diversity around the selected site[152]. Additional features include a skew in the distribution of allele frequencies towards an excess of rare and high-frequency derived alleles, and a transitory increase in the strength of linkage disequilibrium associated with the selected allele(s). These patterns can be detected by an increasing number of statistical tests, which have been extensively reviewed elsewhere[12,13,152–154] (BOX 3). Of note, most of these molecular signatures will be absent or very weak when the main substrate of selection is neutral standing genetic variation[157].

Balancing selection
This is a general type of selective regime that favours the maintenance of diversity in a population[152,158]. There are two main mechanisms by which balancing selection preserves polymorphism: heterozygote advantage (or overdominance) and frequency-dependent selection. Heterozygote advantage refers to a situation in which heterozygous individuals at a particular locus have a greater fitness than homozygotes (for example, the *HbS* 'sickle-cell' variant[11]). Frequency-dependent selection occurs when the fitness of a phenotype is dependent on its frequency relative to other phenotypes in a given population. For example, in negative frequency-dependent selection, the fitness of a phenotype decreases as it becomes more common. Contrary to what would be expected under a selective sweep, balancing selection (if it is not too recent) will lead to an excess of intermediate-frequency variants, which will result in increased levels of diversity[152–154,158].

Box 3 | **Statistical methods for detecting selection**

Statistical tests of neutrality can be roughly subdivided into: those that use divergence data between species (interspecies neutrality tests); those that use polymorphic data within a single species (intraspecies neutrality tests); and those that use data from both divergence between and polymorphism within species. Interspecies tests aim to detect old selective events (for example, adaptive events that participated in human speciation), whereas intraspecies tests detect more recent selective events that occurred no longer than $4N_e$ generations ago (N_e is the effective population size). Neutrality tests (for example, the McDonald–Kreitman test[159]) that use information from both polymorphic and divergence data can detect old as well as recent selection. We restrict our description to interspecies and intraspecies tests, as they are the most relevant to the findings discussed in this Review.

Interspecies neutrality tests

d_N/d_S test. This test detects selection in protein-coding loci by comparing the ratio of non-synonymous (d_N) to synonymous (d_S) substitutions[18]. In the absence of selection (that is, neutrality), synonymous and non-synonymous substitutions should occur at the same rate, and we expect $d_N/d_S = 1$. If non-synonymous variants are negatively selected, $d_N/d_S < 1$, and if they are positively selected, $d_N/d_S > 1$ (REFS 18,153). However, because of the constant action of negative selection at protein-coding loci, the amount of positively selected non-synonymous variants that is needed to elevate d_N/d_S above one is very high, and therefore this test has little power to detect positive selection when only one or a few non-synonymous variants have been selected. To overcome this problem, maximum-likelihood methods have been devised that allow for variation in d_N/d_S ratios among sites. If a distribution that allows values of $d_N/d_S > 1$ fits the data significantly better than a model that does not allow for such values, this is interpreted as evidence for positive selection[160,161]. These methods, however, can suffer from a high false-discovery rate[162].

Intraspecies neutrality tests

Site frequency spectrum-based methods. Natural selection can distort the distribution of allele frequencies in populations, therefore several methods have been developed to evaluate whether the site frequency spectrum (SFS) of mutations conforms to the expectations of the standard neutral model. These tests include Tajima's *D* test, Fu and Li's *D* and *F* tests, and Fay and Wu's *H* test (reviewed in REFS 153,154). The expected value of Tajima's *D* test and Fu and Li's *D* and *F* tests for populations that conform to a standard neutral model is zero. Significantly negative values for these statistics indicate an excess of low-frequency variants, which can result from population expansion, weak negative selection or positive selection. Significantly positive values for these statistics reflect an excess of intermediate-frequency alleles, which can result from population bottlenecks, structure and/or balancing selection. Fay and Wu's *H* statistic tests for an excess of high-frequency derived mutations, which is a hallmark of positive selection. Note, however, that the detection of natural selection using these tests is dependent upon a profound knowledge of the demographic history of the populations under study. More recently, more powerful composite-likelihood approaches have been developed that use the spatial pattern of the SFS to identify and locate selective sweeps (reviewed in REF. 163).

Population differentiation. The F_{ST} statistic examines variation in SNP allele frequencies between populations[164]. Under neutrality, F_{ST} is determined by genetic drift, which affects all loci across the genome similarly. Conversely, natural selection can cause local deviations in F_{ST} values in specific loci. For example, geographically restricted positive selection may lead to an increase in F_{ST} at a selected locus, whereas balancing, negative or species-wide directional selection may lead to decreased F_{ST} values[14]. F_{ST} has an advantage over SFS-based methods in that it can be SNP-specific and can theoretically unmask the genetic variants under selection.

Linkage disequilibrium-based neutrality tests. Some of the most powerful tests for detecting recent positive selection are based on the levels of linkage disequilibrium (LD) associated with particular haplotypes and/or alleles. These tests are aimed at detecting positive selection events that took place <30,000 years ago[13]. Popular LD-based methods include the long-range haplotype (LRH) test[90], the integrated haplotype score (iHS) test[53] and the LD decay (LDD) test[54]. These tests share a similar rationale: an allele that has a high population frequency and that is associated with unusually long-range LD is likely to have been targeted by recent positive selection. This is explained by the rapid increase in allele frequency of the advantageous allele, which means that recombination will not have enough time to substantially break down the haplotype on which the selected mutation arose.

Simian immunodeficiency virus
A type of retrovirus found in African non-human primates. Unlike HIV infections in humans, simian immunodeficiency virus infections in their natural hosts are usually non-pathogenic.

Copy-number variation
A class of DNA sequence variant (including deletions and duplications) in which the result is a departure from the expected diploid representation of DNA sequence.

Erythrocyte
A cell that contains haemoglobin and that can carry oxygen to the body. They are also known as red blood cells.

largely result from adaptive changes in gene regulation[41], comparative studies of gene expression in different primate species would greatly improve our understanding of how changes in gene regulation have been driven by selection. For example, Blekhman *et al.*[42] have recently shown the value of integrating interspecies expression data into an evolutionary framework to identify genes that have expression profiles that are consistent with the action of positive selection on gene regulation in liver, heart and kidney tissues. In the context of human immunity this type of study should now be extended to the most relevant immune cell populations (for example, monocytes, dendritic cells, T cells and B cells).

Differences in gene gains and losses. The amount of fixed nucleotide differences between humans and chimpanzees is small, and few genes show molecular evidence of rapid evolution at the level of protein-coding sequences in these species. However, primates — and humans in particular — exhibit a high rate of gene turnover (that is, gene gains and losses) relative to other mammals[43–45]. Interestingly, among the gene families that show rapid gene turnover, immune-response genes are over-represented[22,43]. In humans, the greatest expansions of immunity-related genes are those observed in the golgin subfamily A genes, which are involved in multiple autoimmune disorders, and in the immunoglobulin (Ig)

heavy chain variable region (*IGHV*) gene family. Humans have 49 golgin subfamily A gene copies, and chimpanzee and rhesus macaques have fewer than 30. Similarly, in terms of *IGHV* genes, humans have 10 more than chimpanzees and 11 more than rhesus macaques[43]. Other considerable gene expansions are observed in macaques, the most substantial being the gain of ~22 human leukocyte antigen (HLA)-related genes[22,43,45]. Whether such an accelerated rate of gene gains and losses is governed by neutral processes or positive selection remains an open question. However, the observation that gene families that show large gene expansions also show evidence for positive selection on their nucleotide sequence[22,43] suggests that adaptive evolution has had an important role in primate gene turnover, and could explain some of the differences in immune responses observed between humans and other primates.

Action of recent selection on immunity genes

Population genome-wide scans for selection. In contrast to the interspecies studies discussed above, which aim to detect old selective events (for example, adaptive events that participated in human speciation), population genetic approaches (that is, intraspecific studies) detect more recent selective events (BOX 3). Several genome-wide scans for positive and ongoing selection in humans have been performed to date; altogether, these have detected more than 5,000 genomic regions (encompassing ~14% of the genome; reviewed in REF. 14) that present at least one genomic signature of positive selection[14,46–54]. By inspecting these regions, we noted the presence of 360 genes with immune-related functions, 186 of which have been detected by two or more studies. These immunity-related genes are therefore the most likely to have been targeted by positive selection in humans (FIG. 1; Supplementary information S2 (table); Supplementary information S3 (box)). However, caution should be taken when claiming that any of these genes is a target of selection. Several of them are in strong linkage disequilibrium (LD) with each other or with other non-immunity genes, so the observed signatures of selection might reflect genetic hitchhiking. In addition, most of these genes have been identified using empirical outlier approaches that, unlike model-based approaches, are unable to estimate false positive rates (that is, the fraction of outlier genes that are not selected). Despite these limitations, this list of 186 immunity-related genes (Supplementary information S2 (table)) provides a set of genes that should be prioritized in future functional studies that aim to distinguish the targets of selection from 'evolutionary noise'.

The genomic regions that contain the set of genes identified by these genome-wide scans for selection[14,46–54] are not significantly enriched for immune-related functions when taken as a whole[14], but immunity-related genes are overrepresented when focusing on selective events that occurred more recently[53,54] (<30,000 years ago, as detected by the integrated haplotype score (iHS) and LD decay (LDD) tests, BOX 3). This observation suggests that our immune system has been particularly challenged during the more recent phases of human evolution, which could reflect the proposed burden of infectious diseases that are associated with the advent of agriculture at the beginning of the Neolithic period 10,000 years ago[55]. Farming resulted in a massive increase of human population sizes, a situation that may have facilitated the spread of certain infectious agents (for example, measles, rubella and pertussis) that require large human population densities to persist[56]. Inadequate sanitary practices and exposure to new zoonoses following the domestication of animals might also have contributed to the burden of infectious diseases[55]. Population genetic analyses of immunity-related genes in populations that differ in their mode of subsistence (that is, hunter-gatherers versus farmers) together with detailed epidemiological data could improve our understanding of how this major change in human lifestyle affected our genetic adaptation.

Pathogen-driven selection in the human genome. Intuitively, the most obvious cause of selection on immunity-related genes is predicted to be pathogen presence — that is, pathogen-driven selection. This hypothesis has been recently tested by three studies that searched for correlations between genetic variability in human populations worldwide and pathogen richness in the corresponding geographic regions[57–59]. The first focused on the HLA complex, the extraordinary genetic diversity of which is believed to be driven by balancing selection[60–62]. Prugnolle *et al.*[59] showed that populations from areas with high pathogen diversity show increased variability at HLA class I genes (particularly at the *HLA-B* gene) with respect to neutral diversity. Likewise, Fumagalli *et al.*[57,58] observed correlations between pathogen richness and diversity at some blood group antigen and interleukin (*IL*) genes. Consistently, population genetic analyses supported the action of balancing selection at some of these blood group antigen genes, such as *CD55*, *CD151*, solute carrier family 14 (urea transporter) member 1 (*SLC14A1*) and fucosyltransferase 2 (*FUT2*), and at some *IL* genes, such as IL1 family, member 5 (*IL1F5*), *IL1F7*, *IL1F10*, IL7 receptor (*IL7R*) and IL18 receptor accessory protein (*IL18RAP*)[57,58]. However, these studies assume that pathogen richness has been maintained unchanged over time and do not consider the virulence or pathogenicity of the different pathogens. Despite these limitations, they support the notion that the underlying selective pressure driving diversity at host immunity-related genes is pathogen presence.

However, inferring the specific pathogens that exert pressure on host genes remains a daunting task that requires precise knowledge of the nature of the host–pathogen interactions. It is further complicated by the facts that several pathogens are likely to have exerted pressure on the same genes and that selection can vary in form, intensity, time and space. For example, signatures of balancing and positive selection have been observed for chemokine (C-C motif) receptor 5 (*CCR5*) (REFS 63–65, but see REF. 66), variation in which confers higher resistance to HIV-1 and delays progression to AIDS[67,68]. The recent emergence of HIV-1 in the human population suggests that the resistance afforded by *CCR5* variation results from adaptive changes to older

Antigen
Any substance that the immune system can respond to by producing antibodies.

Linkage disequilibrium
The non-random association of alleles at two or more loci. The pattern of linkage disequilibrium in a genomic region is affected by mutation, recombination, genetic drift, natural selection and demographic history.

Genetic hitchhiking
The process by which a neutral, or in some cases deleterious, mutation may increase in population frequency owing to linkage with a positively selected mutation.

Outlier approaches
Popular approaches for detecting selection that identify loci that present extreme values for a given statistic from empirical genome-wide genetic data. It is assumed that these 'outlier loci' are enriched for loci targeted by selection.

Interleukins
A group of secreted proteins that are produced by immune cells and that can modulate immune and inflammatory responses.

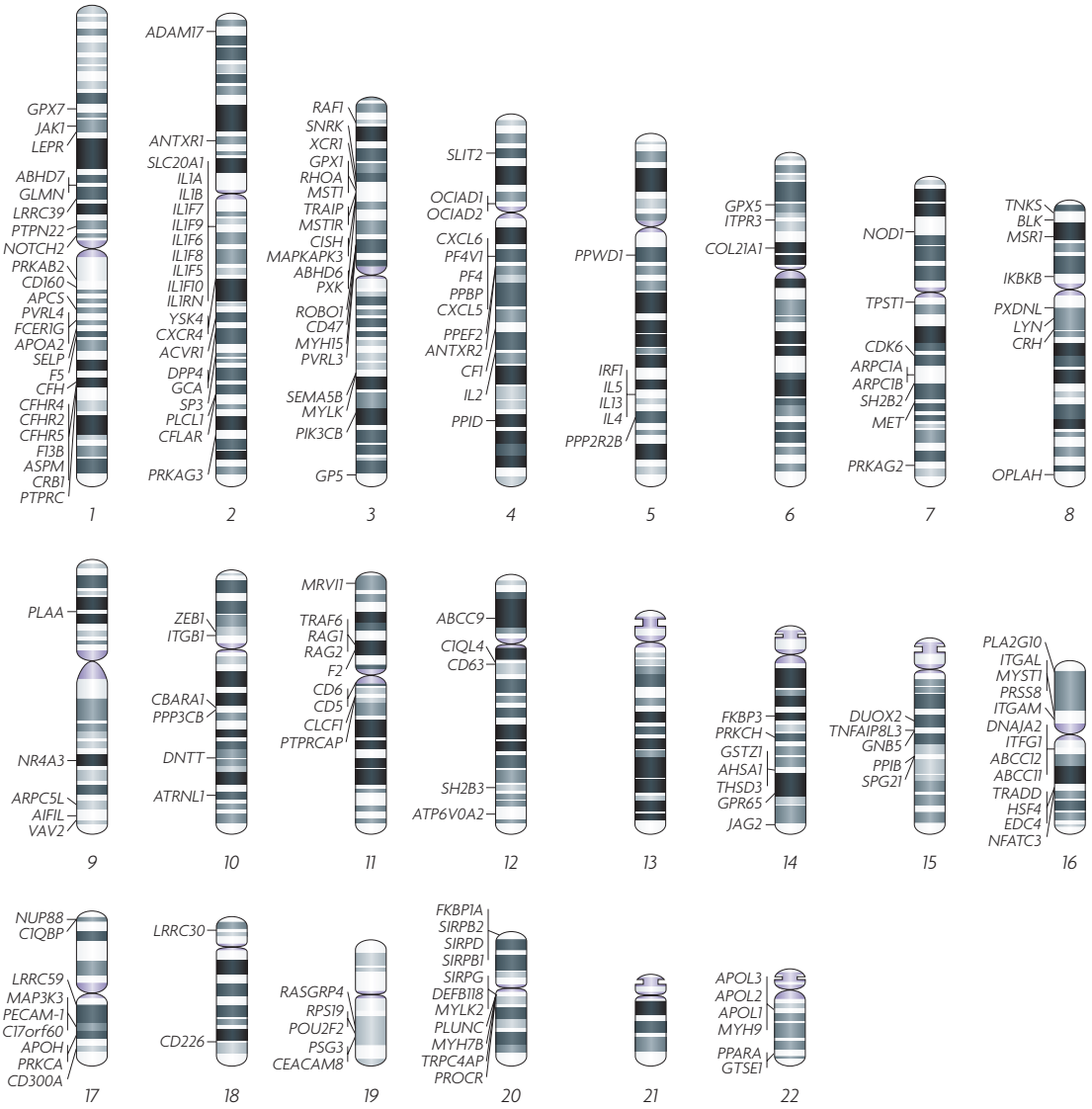
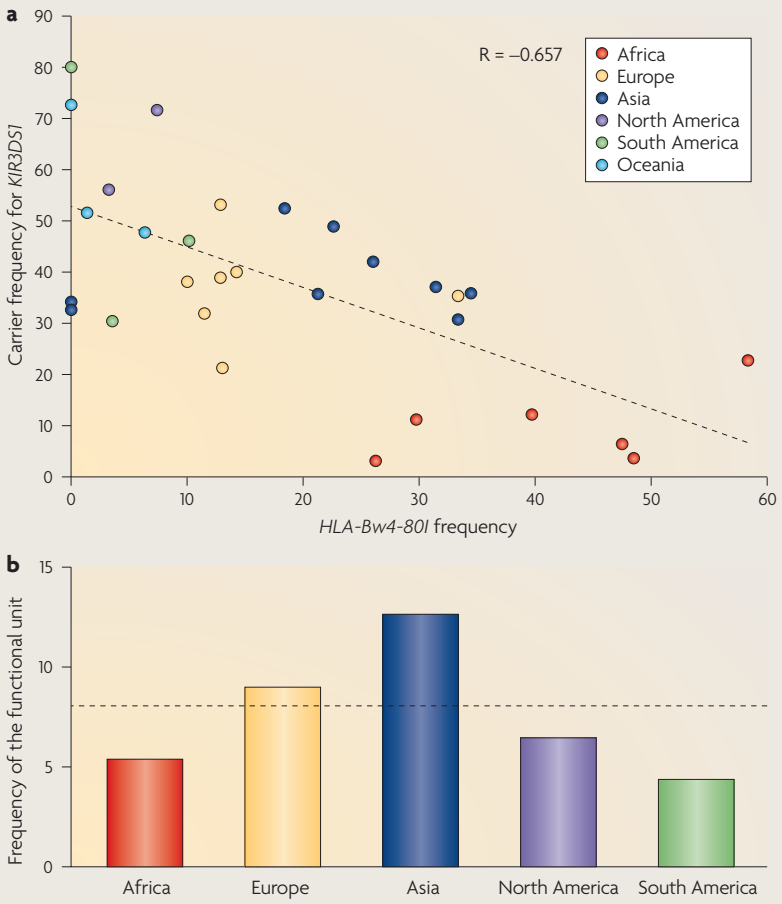


Figure 1 | **Genomic map of immunity genes that are candidates for positive selection.** Shown are immunity-related genes that have been reported by at least two different genome-wide scans for selection as presenting a signature of positive selection in at least one human population. Further information on the genes and a full list of all of the immunity-related genes reported by ten different genome-wide scans for selection are provided in Supplementary information S2 (table). X- and Y-linked genes are not reported because most genome-wide scans for selection performed to date have not considered the sex chromosomes in their analyses. Only genes that are consensually considered to be involved in 'immunity' or 'host defence' as defined by Gene Ontology analyses (see Supplementary information S3 (box)) are reported here, and therefore some 'non-classical' immunity-related genes that are also involved in host defence might have escaped our inclusion criteria. In addition, note that several well-documented cases of selection, such as glucose-6-phosphate dehydrogenase (*G6PD*), Duffy blood group chemokine receptor (*DARC*) and β-globin (*HBB*), are not reported here because they do not fall into the stringent thresholds of 'significance' defined by the genome-wide scans for selection. For example, the criterion used by Barreiro *et al.*[49] to identify genes under positive selection was that the gene should present at least one SNP with an overall $F_{ST} > 0.65$. Because the highest F_{ST} value at SNPs in *DARC* is 0.63, the Duffy antigen gene was not reported as a candidate for positive selection[49].

pathogens (for example, smallpox)[63,69]. Other types of selection might also influence the evolution of immunity-related genes — for example, epistatic selection, which has been recently described between killer cell Ig-like receptor (*KIR*) genes and their HLA ligands[70] (BOX 4), or sexual selection, which seems to have exacerbated the levels of diversity observed at *HLA* genes[71–73].

Selection imposed by malaria. Despite the difficulties in identifying the precise pathogens that exert pressure on individual genes, the *Plasmodium* species parasites — the causative agents of malaria in humans — remain an emblematic case for which there is strong clinical, epidemiological and evolutionary evidence to support their effects on human genome evolution. Malaria has

Box 4 | **Epistatic selection: co-evolution of immune players**



The immune system is composed of many elements that interact with one another in remarkably complex ways. The way in which these interactions occur will ultimately dictate an individual's immunological phenotype(s), and these interactions are therefore expected to be targets of natural selection. A simple example is that allele *A* of protein X will be selected if, and only if, it is found in combination with allele *B* of protein Y. This is called epistatic selection, which involves simultaneous positive selection of combinations of alleles at two or more sites[165]. Although epistatic effects have been convincingly demonstrated in model organisms[166], only recently have Single *et al.*[70] experimentally demonstrated the occurrence of epistatic selection in humans. This occurs between the killer cell immunoglobulin-like receptor (KIR) receptors and their human leukocyte antigen (HLA) ligands[70].

*KIR*s are a family of 14 closely linked genes that encode inhibitory and activating receptors that are mainly expressed by natural killer (NK) cells[167]. In humans, HLA class I molecules are the functional ligands for KIR receptors. After recognition of their specific HLA ligands, inhibitory receptors will downregulate NK cell activity, which impedes their ability to 'attack' healthy cells, whereas activating KIRs will lead to an enhancement of NK cell activation signals, which might be beneficial for fighting viral infection[167,168]. Like *HLA*s[60–62], *KIR*s have high levels of diversity that seem to be maintained by balancing selection[167,169]. Given the receptor–ligand relationship between certain KIR and HLA class I molecules, Single *et al.*[70] suggested that epistatic relationships between these polymorphic loci could exist. They genotyped a panel of 1,642 individuals from 30 geographically distinct populations to test for the presence or absence of several inactivating and activating KIR genes and their corresponding HLA ligands. Interestingly, they observed consistent negative correlations between the presence of activating KIR genes and their corresponding HLA ligand groups across populations. This was particularly prominent for the activating KIR gene *KIR3DS1* and the gene for its putative HLA ligand, *HLA-Bw4-80I* (see the figure, part **a**). Given that KIR and HLA genes map to different chromosomes, the correlations observed strongly support the co-evolution of KIR receptors and their HLA partners.

In evolutionary terms, the negative correlations indicate that there is negative selection against functional units that lead to NK cell activation (that is, the combination of activating KIRs and their specific ligands). This selection probably prevents exacerbated NK-induced immune responses against healthy cells or tissues (that is, the 'self'). However, if there is selection against high frequencies of these functional units, we may ask why, for example, have KIR3DS1 and/or its ligand not been lost? Indeed, evolutionary evidence suggests that *KIR3DS1* has been segregating in the population for more than 3 million years, which provides an astonishing example of balancing selection[169]. The answer seems to come from disease association data. Genotype combinations that lead to higher activation of NK cells tend to be beneficial for fighting viral infections such as HIV and hepatitis C[168]. These same genotypes, however, tend to increase susceptibility to autoimmunity and possibly to cancers that have an inflammatory component to their pathogenesis[168]. Natural selection seems to have ensured that HLA–KIR genotypes are found at a frequency equilibrium that guarantees an optimal balance between a stronger or quicker inflammatory response against pathogens (mainly viruses) and the avoidance of an exacerbated immune response against the 'self'. Indeed, although the frequency of *KIR3DS1* and *HLA-Bw4-80I* varies dramatically across populations (2–80% for *KIR3DS1*; see the figure, part **a**), the frequency of the functional KIR–HLA combination (*KIR3DS1–HLA-Bw4-80I*) is remarkably similar across major population groups[70] (~8% frequency; see the figure, part **b** — note that the frequency of the functional unit in Oceania is zero). The figure is modified, with permission, from REF. 70 © (2007) Macmillan Publishers Ltd. All rights reserved.

been, and remains, one of the leading causes of child mortality worldwide[74]. A growing number of human genes have been related to malaria resistance or severity, some of which exhibit strong signatures of natural selection[75]. These loci can be broadly divided into: genes that are related to erythrocyte metabolism, such as Duffy blood group chemokine receptor (*DARC*), glucose-6-phosphate dehydrogenase (*G6PD*) and *HBB*; genes that mediate cytoadherence by *P. falciparum*-infected erythrocytes, such as complement component (3b/4b) receptor 1 (*CR1*) and intercellular adhesion molecule 1 (*ICAM1*); and genes that are directly involved in immune responses, such as the *HLA* genes, interferon-γ (*IFNG*), tumour necrosis factor (*TNF*) and CD40 ligand (*CD40LG*).

The most well-known example is *HBB*, for which three amino acid changes — the *HbS*, *HbC* and *HbE* alleles — confer different levels of malaria protection[76]. *HbS* homozygotes suffer from sickle-cell disease, but heterozygotes have a tenfold reduced risk of severe malaria, which explains the high frequencies (up to 30%) of this allele in Africa; this is the most convincing case of heterozygote advantage reported so far[11]. Likewise, the relatively high frequencies of the protective *HbC* and *HbE* alleles in Western Africa and Southeast Asia, respectively, attest to the strong selective pressure exerted by malaria in different geographical areas[77–80]. Another example is *DARC*, which encodes a membrane protein used by *Plasmodium vivax* to enter erythrocytes[81]. The null mutation *FY*0*, which disrupts protein expression and imparts complete resistance to *P. vivax* infection in homozygotes[82], almost reaches fixation in many sub-Saharan African populations but is virtually absent elsewhere[83,84]. However, it is still unclear whether *P. vivax* has been the only selective agent driving the increase in frequency of the *FY*0* allele, as other pathogens or cell type-specific processes have also been proposed to have exerted selective pressure on *DARC*[84,85]. Finally, the genes *G6PD*, *CD40LG* and *CD36*, in which variation has been associated with differential susceptibility to *P. falciparum* malaria in Africa[86–88], have been targeted by recent positive selection, as shown by the long-range haplotypes that surround the protective mutations[89,90].

Interestingly, several malaria-resistance alleles at a number of genes (for example, *G6PD*, *HBB* and *CD40LG*) are estimated to have arisen recently, within the past 10,000 years[80,90–93]. Despite the methodological uncertainties in time estimates, all of the estimates coincide with the time at which the African *P. falciparum* population expanded dramatically[94] and the time of speciation of the African mosquito vector *Anopheles gambiae*[95] (<10,000 years ago). This suggests that the strong selective pressure imposed by *P. falciparum* on humans started (or at least increased) in the Neolithic period.

Do different microbes exert different pressures? The mode by which the human host adapts to pathogens depends on a large number of factors, including the type of microorganism, the different temporal and spatial presence of pathogens during evolution, their varying pathogenicity, the nature of the host–pathogen interaction, and the rate at which pathogens evolve. Theoretical considerations apart, is there any evidence that certain classes of pathogens exert stronger selective pressure on host genes? Two of the studies mentioned above argue that there is evidence of this[58,59]. Prugnolle *et al.*[59] found that the positive correlations between higher HLA-A and HLA-B diversity and pathogen richness are mainly explained by viruses rather than by the other pathogens considered (protozoa and obligate and facultative intracellular bacteria). This suggests that viruses have exerted the strongest selective pressure for diversity at HLA class I genes. Interestingly, a recent evolutionary study of the Toll-like receptor (*TLR*) gene family has concluded that viruses have exerted a stronger selective pressure than other pathogens on this family of innate immune receptors by constraining amino acid diversity at viral-recognition *TLR*s[96]. In addition, when examining how IL diversity correlates with the presence of micropathogens (viruses, bacteria, fungi and protozoa) and macropathogens (insects, arthropods and helminths (parasitic worms)), Fumagalli *et al.*[58] found that helminths have imposed the strongest pressure on a subset of the ILs studied (for example, *IL4*, *IL4R*, *IL10*, *IL19* and *IL20*). Because helminths can maintain themselves in small human groups, such as hunter-gatherers[97], the authors propose that parasitic worms have been a stable, long-standing threat to human survival. Systematic evolutionary studies of other families of immune-related genes, accompanied by increased knowledge of both the nature of host–pathogen interactions and the clinical and epidemiological significance of the different infectious diseases, are now needed to delineate the types of pathogens that have been most important in shaping the diversity of our immune responses. Likewise, it will be interesting to evaluate whether and how our long-standing contact with symbiotic microbes, such as those composing the mouth and the gut flora, has imposed a significant selective pressure in the shaping of our immune system.

Loss of function: neutral or advantageous?

Redundancy in host defence mechanisms. The evolutionary dissection of the extent to which selection has targeted the loss of function of some immunity-related genes can provide insights into the degree of redundancy in our immune system[98]. An illustrative example is provided by mannose-binding lectin (MBL, encoded by *MBL2*), which is a member of the collectin protein family that activates the lectin complement pathway[99]. MBL deficiency has been associated, although not conclusively[100], with increased susceptibility to several infectious diseases[101]. However, alleles that confer MBL deficiency are common worldwide (with a frequency of up to 30%)[100,102,103], which suggests either a protective effect of MBL deficiency or a redundant role of this lectin in host defence. Although the high frequency of *MBL2*-deficiency alleles was initially proposed to result from balancing selection in European-descent populations[103], a more recent study in multiple ethnologically well-defined populations has shown that

Fixation
The increase in the frequency of a genetic variant in a population to 100%.

Haplotype
The allelic composition over a contiguous chromosome stretch.

Complement pathway
The complement system is a complex protein cascade that is involved in both innate and adaptive immunity. Three pathways activate the complement system: the classical pathway, the alternative pathway and the lectin pathway.

the patterns of *MBL2* variation are compatible with neutrality[102]. Consistent with this, *MBL2*-deficiency alleles are not associated with unusually long-range haplotypes[102,104,105]. This lack of evidence for positive selection explaining the high frequencies of *MBL2*-deficiency alleles supports a largely redundant role for MBL in immunity to infection[102]. Alternative mechanisms of target recognition — for example, through the ficolins or the C1q-dependent classical pathway — may compensate for MBL deficiency[106,107].

A similar situation is observed for *TLR5*, which detects flagellated bacteria[108]. The *TLR5*^{R392STOP} mutation, which acts in a negative-dominant fashion[109], reaches high population frequencies of up to 23%[96,109–111]. Two independent studies[96,110] have detected no signals of positive selection in *TLR5*, which suggests that *TLR5*^{R392STOP} drifts neutrally in human populations in a similar way to *MBL2*. The evolutionary redundancy exhibited by TLR5 suggests that other mechanisms of flagellin recognition (for example, mediated by NOD-like receptor (NLR) family CARD domain-containing 4 (NLRC4))[112] may provide sufficient protection against infection. Future systematic studies of innate immunity receptors will allow us to better appreciate the degree of duality in microbe-sensing mechanisms[113] and will enable us to evaluate whether redundancy is a common, previously unappreciated feature of the immune system.

Selective advantage for gene inactivation. As posited by the rather counter-intuitive 'less is more' hypothesis[114], the death of a gene (pseudogenization) might, under certain circumstances, represent a major selective advantage. In the context of immunity, when a pathogen uses host immune receptors as a mechanism of cell entry and survival, mutations that inactivate those receptors are likely to represent a selective advantage for the host (assuming there are no major pleiotropic side effects). The case of positive selection for *CCR5* and *DARC* well illustrates this situation. As discussed above, mutations that impair the function of CCR5 (Δ*32-CCR5* deletion) and DARC (*FY*0*) obstruct the entrance of pathogens, resulting in increased resistance to HIV-1 (REF. 67) and *P. vivax*[82], respectively. Other examples of positive selection acting on loss-of-function mutations have been reported for caspase 12 (*CASP12*), serpin peptidase inhibitor clade A member 2 (*SERPINA2*), sialic acid-binding Ig-like lectin 12 (*SIGLEC12*) and some leukocyte Ig-like receptor genes[115–118]. Because all of these genes are involved in host defence mechanisms (although not exclusively), their loss may have been mediated through host–pathogen interactions.

CASP12 (which encodes a cysteine protease) is the best-documented example of a gene that is apparently being lost from the human species owing to host–pathogen interactions. The inactive form of *CASP12* is virtually fixed outside Africa and has a frequency of ~80% in African populations[117,119]. Population genetic analyses have revealed that the high frequency of the inactive form of CASP12 results from positive selection[44,117], with increased resistance to sepsis probably being the selective advantage that drives this process[117]. Indeed, individuals who express the inactive form of CASP12 have an increased inflammatory response to endotoxins and, when suffering from severe sepsis, their mortality rate is three times lower than that of patients who are homozygous for the active form (who present reduced inflammatory responses)[119]. The abrogated inflammatory response associated with the active form of CASP12 might largely be due to the inhibitory effect CASP12 has on CASP1 (which mediates the proteolytic activation of inflammatory cytokines, mainly IL1β)[120]. Under this model, it is interesting to note the possibly advantageous loss of ICEBERG in chimpanzees[21,121]. Like CASP12, ICEBERG prevents the activation of CASP1 and the subsequent generation of IL1β, which raises the interesting hypothesis that, by means of different mechanisms, humans and chimpanzees have followed two independent evolutionary routes to reduce sepsis risk.

Natural selection and immune-related disorders

By definition, natural selection targets alleles that change an individual's fitness. Among the diverse factors that affect fitness, disease susceptibility is arguably one of the most important. This notion is supported by recent studies showing that genes involved in Mendelian and complex diseases are privileged targets of purifying and positive selection, respectively[122,123]. An intuitive prediction is that genes that evolve under strong purifying selection are of major relevance for survival ('essential genes'), and mutations that impair their function are expected to lead to severe clinical phenotypes, or even death. This prediction seems to hold true for some immunity genes. For example, *TLR3*, myeloid differentiation 88 (*MYD88*) and IL1 receptor-associated kinase 4 (*IRAK4*), which are key members of the TLR pathway, are all evolving under strong purifying selection (REF. 96; L.B.B. and L.Q.M, unpublished data), and mutations that impair their function have been shown to lead to life-threatening infections[124–126].

For common infectious diseases that present complex patterns of inheritance (for example, tuberculosis and influenza) the situation is less straightforward. Because complex susceptibility to disease is probably controlled by genetic variations in multiple genes, each of them contributing modestly to the phenotype[127], each individual risk variant is expected to segregate in the population either neutrally or under the action of weaker purifying selection than that exerted on mutations that result in severe Mendelian disorders. Genes that are targeted by weak purifying selection are therefore ideal candidates for explaining complex susceptibility to infectious diseases at the population level[128]. An illustrative example is provided by *TLR4*: the amino acid diversity observed at *TLR4* is controlled by weak purifying selection[96,129] and, as predicted, these polymorphisms seem to be associated with increased susceptibility to several infectious diseases[130].

Genes that are targeted by positive selection are also thought to contribute to differences in disease susceptibility or outcome. In the context of immunity-related genes, an obvious, arguably naive, expectation is that positively selected alleles increase protection against infectious disease. This is unquestionably true for various genes that are involved in malaria resistance or severity, as shown

Ficolins
A group of humoral proteins that contain a collagen-like domain and a fibrinogen-like domain. They can bind carbohydrate molecules on pathogens, apoptotic and necrotic cells to activate the lectin complement pathway.

Caspases
Enzymes that play an important part in the immune system by activating numerous cellular proteins that are involved in apoptosis, necrosis, immunity and inflammation. They are synthesized as inactive procaspases that are later activated by proteolytic cleavage into active caspases.

Fitness
A measure of the capacity of an organism to survive and reproduce.

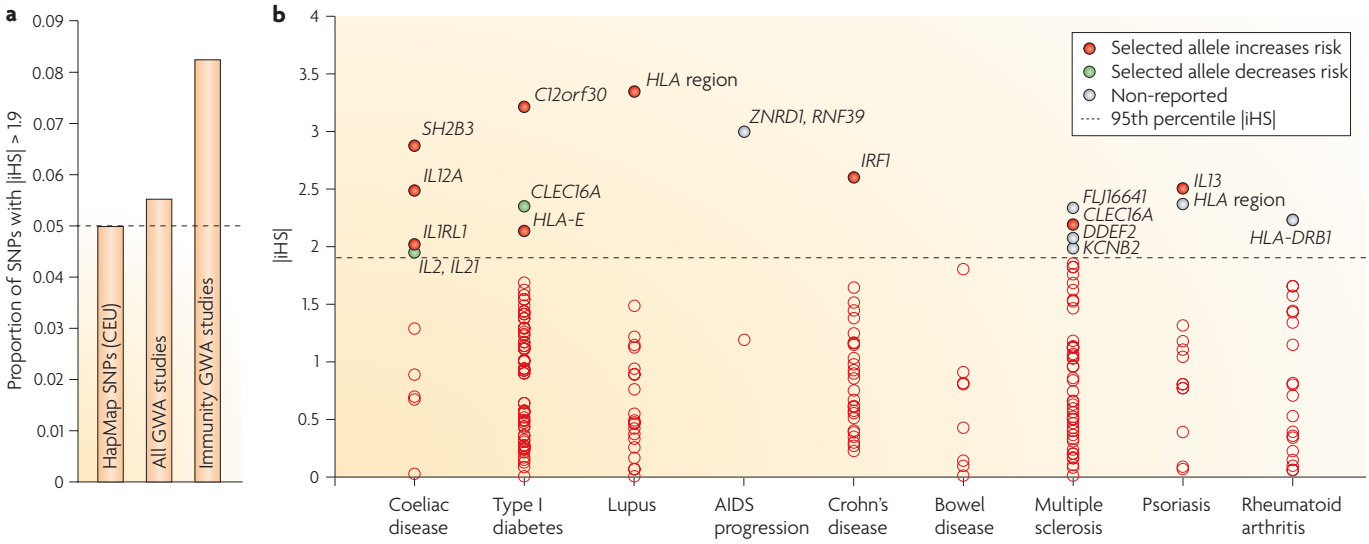


Figure 2 | **Positive selection targeting SNPs associated with disease. a** | The proportion of SNPs that present evidence for recent positive selection, as attested by the integrated haplotype score (iHS) statistic[53] (BOX 3). iHS scores were obtained from the Haplotter genome browser. The *y* axis represents the proportion of SNPs that present iHS values that are suggestive of recent positive selection (|iHS| = 1.9 corresponds to the 95th percentile of HapMap phase II genome-wide distribution for Europeans). The proportion of SNPs is shown for European HapMap phase II SNPs (individuals of European descent from Utah (CEU) population), for European HapMap phase II SNPs that have been associated with disease by genome-wide association (GWA) studies, and for European HapMap phase II SNPs that have been associated with immune-related disorders (that is, infectious disease and autoimmune or inflammatory disorders) by GWA studies. The analyses are restricted to European HapMap data because most GWA studies have been performed in populations of European descent. **b** | iHS values for SNPs associated with several immune-related diseases by different GWA studies. For most diseases, there is at least one associated SNP that seems to lie on a positively selected haplotype (Supplementary information S3 (box)). 'Non-reported' stands for SNPs for which no information is available about whether the selected allele is associated with increased resistance or susceptibility to disease. *CLEC16A*, C-type lectin domain family 16, member A; *DDEF2*, development and differentiation enhancing factor 2 (also known as *ASAP2*); *FLJ16641*, also known as leucine, glutamate and lysine rich 1 (*LEKR1*); *HLA*, human leukocyte antigen; *IL*, interleukin; *IL1RL1*, interleukin 1 receptor-like 1; *IRF1*, interferon regulatory factor 1; *RNF39*, ring finger protein 39; *ZNRD1*, zinc ribbon domain-containing 1.

by the protection afforded against *P. falciparum* and/or *P. vivax* malaria by positively selected mutations at *DARC*, *HBB* or *G6PD* (discussed above), among others[75]. Another example comes from the increased protection against tuberculosis[131] and leprosy reversal reaction[132] conferred by a low-responsiveness *TLR1* mutation, which has been positively selected among Europeans[96]. Interestingly, a recent genome-wide scan for selection has shown that the SNP that shows the greatest differentiation among Western Eurasian populations is located in the *TLR10–TLR6–TLR1* cluster, which indicates that the event of positive selection observed for the low-responsiveness *TLR1* mutation was probably restricted to specific Western Eurasian populations[133].

Together, these examples show a link between positively selected mutations and disease resistance, but can these observations be extended to the whole human genome? We addressed this question by searching for signatures of recent positive selection in ~1,500 SNPs that were found to be associated with different diseases by genome-wide association studies (GWA studies) (see the National Human Genome Research Institute catalogue of published GWA studies). We retrieved from Haplotter the iHS scores (BOX 3) of each disease-associated SNP and searched for an enrichment in signals of positive selection at these SNPs (that is, an increased proportion of elevated iHS values) in comparison with genome-wide expectations. Overall, these SNPs do not seem to be significantly enriched for signatures of positive selection (FIG. 2), an observation that might not be surprising given that most GWA studies have searched for associations with late-onset diseases that have theoretically little impact on fitness. However, when we restricted the analysis to SNPs that are associated with immunity-related phenotypes (~230 SNPs from 46 GWA studies; see Supplementary information S3 (box)) — mainly autoimmune diseases and some infectious diseases — we observed an enrichment for signals of recent selection ($p = 0.015$) (FIG. 2). The signatures of positive selection at SNPs associated with coeliac disease provide an interesting example; recent positive selection has targeted four out of the nine known loci associated with this disease (all loci are unlinked) (FIG. 2). Coeliac disease is an inflammatory disorder with an autoimmune component that is caused by an unbalanced inflammatory reaction to gliadin, a gluten protein found in wheat. The fact that wheat consumption started to spread extensively after the introduction of agriculture

Leprosy reversal reaction
A syndrome that is characterized by the rapid activation of a T helper 1 inflammatory response to *Mycobacterium leprae*. This reaction can cause substantial morbidity.

Genome-wide association studies
Studies in which associations between genetic variation and a phenotype or trait of interest are identified by genotyping cases (for example, diseased individuals) and controls (for example, healthy individuals) for a set of genetic variants that capture variation across the entire genome.

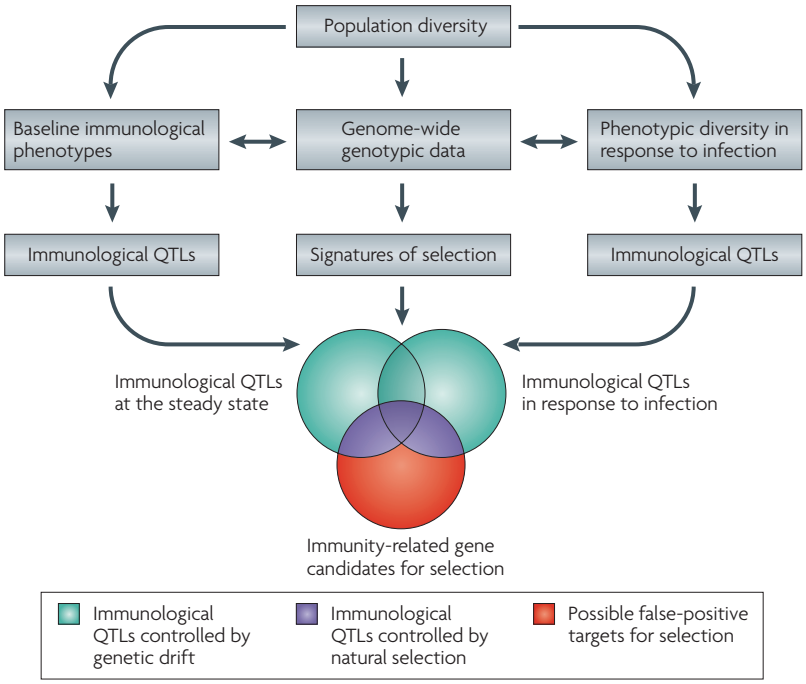


Figure 3 | **Towards the identification of adaptive immunological phenotypes.** To reveal immunological phenotypes that have been actual targets of selection, a promising strategy is to identify genomic regions that present both molecular signatures of selection and immunological quantitative trait loci (QTLs) — that is, SNPs that are associated with different immunological phenotypes. Immunological phenotypes, such as mRNA or protein levels, can be measured at the steady state (that is, with no stimulation) and under stress conditions (that is, *in vitro* infection with different pathogens). The quantitative measurement of these immunological traits can then be correlated with genome-wide sequencing or genotyping data to identify immunological QTLs. Genes that present molecular signatures of selection in addition to immunological QTLs would correspond to real targets of selection (purple area in the Venn diagram). Genes that present no signatures of selection but that are associated with immunological QTLs probably reflect variations in immune responses that have little or no impact on fitness (green area in the Venn diagram). Finally, genes that present molecular signatures of selection but that are non-QTLs (that is, variation in these genes has no impact on phenotypes) might represent false-positive targets of selection (red area in the Venn diagram). Note, however, that this latter group of genes could be under selection if the selected genetic variants are fixed in the population (no phenotypic variation) or if they are associated with phenotypes that are not measured (or identifiable) under the designed experimental setting (for example, a genetic variant that affects the immune response of an immune cell type that has not been tested).

again provides support for links between farming (for example, the domestication of wheat) and increased selection, in ways that need further investigation.

Another remarkable observation is that, in almost all cases, the positively selected alleles or haplotypes do not confer protection against disease but instead increase risk[128] (FIG. 2), particularly for autoimmune or inflammatory disorders (note, however, that most GWA studies have focused on autoimmune or inflammatory rather than infectious diseases). This intriguing observation raises a number of questions. For example, what is the underlying selective advantage conferred by these haplotypes? Have the haplotypes that are today associated with increased risk for autoimmune or inflammatory disorders conferred increased resistance to past or present infectious agents? How do these selected haplotypes affect immunological phenotypes? Do they increase pro-inflammatory responses, which could explain why they increase susceptibility to autoimmune and inflammatory disorders, or instead do they dampen the levels of pro-inflammatory responses? Additional GWA studies for infectious diseases, together with the functional characterization of these positively selected alleles and haplotypes, will be required to delineate the selective advantage conferred by these genetic variants. However, it is tempting to speculate that the incidence of autoimmune disorders in modern societies could reflect, at least partially, the consequences of past selection for increased immune responses to combat infection.

Conclusions and future perspectives

The field of evolutionary genetics of immunity investigates how natural selection has shaped the variability of host defence genes in present-day human populations, providing an indispensable complement to clinical and epidemiological genetics[12,13,98,127]. Despite the inherent limitations associated with genome-wide genotyping data sets and methods of detecting selection (see REF. 12 for an extensive review on these problems), multiple regions of the genome, many of which encompass immunity-related genes (FIG. 1; Supplementary information S2 (table)), exhibit compelling molecular signatures of selection. The upcoming availability of whole-genome sequences from thousands of ethnically diverse individuals (see, for example, the 1000 Genomes Project) will provide a more complete picture of genetic diversity, including low-frequency variants, segregating around the world. Together with improved statistical methods, these data will provide a more detailed, robust map of selection in the human genome. However, natural selection acts on phenotypes and therefore, once immunity-related genes that present robust signatures of selection at the molecular level have been identified, we will have to turn towards a 'phenomics' approach to identify the true targets of selection and ultimately understand selection at the organismal level (FIG. 3). This approach will provide new insights into the immunological mechanisms that underlie the patterns of selection observed at the molecular level, and will help to settle questions on the relationship between adaptive changes and present-day increased resistance or susceptibility to infectious diseases. More generally, the integration of genetic and phenotypic data will enable us to quantify the degree of naturally occurring variation in immune responses and tackle questions regarding the definition of a 'healthy' immune response, inter-individual and inter-population variation in immune responses, and the extent to which this phenotypic variation is shaped by natural selection.

Therefore, a lot more work is still needed. This work can take advantage of several technologies (for example, expression microarrays, RNA–seq and mass spectrometry) that allow assessment of the levels of inter-individual phenotypic variation at the genome-wide level. We are therefore in an exciting era in which it is now possible to correlate genome-wide genetic data with gene expression phenotypes and other quantitative traits (such as levels

Expression quantitative trait loci
Regions of the genome at which genetic allelic variation is associated with variation in gene expression.

Quantitative trait loci
Regions of the genome at which genetic variation is associated with a particular phenotypic characteristic or trait (for example, height, weight or skin colour).

of circulating proteins, metabolite abundance or patterns of methylation) by mapping expression quantitative trait loci (eQTLs)[134] and quantitative trait loci (QTLs), respectively. In this context, recent genome-wide SNP and expression data in human populations have augmented our knowledge of how genetic variation affects gene expression[135,136] and how changes in gene expression have been targeted by selection[137]. However, the observation that variants that affect gene expression tend to act in a cell type-specific manner[138] highlights the need to explore the contribution of genetic variation to inter-individual differences in immunological phenotypes in specific cell types. In this context, studies that examine the levels of thousands of transcripts and proteins (for example, cytokines or chemokines) both at steady state (in 'health') and under stressed conditions (in 'disease'; after *in vitro* infection with different pathogens) in different immunological cell types (such as dendritic cells or $CD4^+$ or $CD8^+$ T cells) (FIG. 3) would undoubtedly provide tremendous insights into how the genotype influences the immunological phenotype under different settings (for an example, see REF. 139). In addition, these studies will pave the way for a better understanding of how 'altered' immune responses may correlate with different disease states. Finally, because all of the different intermediate phenotypes to be tested represent potential substrates for natural selection, it is sensible to search for genomic regions that present molecular signatures of natural selection in addition to immunological QTLs, as they might contain functionally important genes and variants that are involved in host immunity to infection and disease outcome. This integrative approach will increase our understanding of how our immune system has adapted to pathogen pressures over time and will reveal immunological mechanisms that have been crucial for our past and present survival.

1. Casanova, J. L. & Abel, L. Inborn errors of immunity to infection: the rule rather than the exception. *J. Exp. Med.* **202**, 197–201 (2005).
2. Haldane, J. B. S. Disease and evolution. *Ric. Sci.* **19** (Suppl. A), 68–76 (1949).
3. Ausubel, F. M. Are innate immune signaling pathways in plants and animals conserved? *Nature Immunol.* **6**, 973–979 (2005).
4. Cooper, M. D. & Alder, M. N. The evolution of adaptive immune systems. *Cell* **124**, 815–822 (2006).
5. Flajnik, M. F. & Du Pasquier, L. Evolution of innate and adaptive immunity: can we draw a line? *Trends Immunol.* **25**, 640–644 (2004).
6. Kimbrell, D. A. & Beutler, B. The evolution and genetics of innate immunity. *Nature Rev. Genet.* **2**, 256–267 (2001).
7. Lazzaro, B. P. & Little, T. J. Immunity in a variable world. *Phil. Trans. R. Soc. Lond. B* **364**, 15–26 (2009).
8. Schulenburg, H., Kurtz, J., Moret, Y. & Siva-Jothy, M. T. Introduction. Ecological immunology. *Phil. Trans. R. Soc. Lond. B* **364**, 3–14 (2009).
9. Akira, S., Uematsu, S. & Takeuchi, O. Pathogen recognition and innate immunity. *Cell* **124**, 783–801 (2006).
10. Medzhitov, R. Recognition of microorganisms and activation of the immune response. *Nature* **449**, 819–826 (2007).
An integrated view of the innate and adaptive components of the immune system, how they work and interact, and their role in host defence.
11. Allison, A. C. Protection afforded by sickle-cell trait against subtertian malareal infection. *BMJ* **1**, 290–294 (1954).
The seminal paper that showed that 'sickle-cell' heterozygotes have increased protection against malaria.
12. Nielsen, R., Hellmann, I., Hubisz, M., Bustamante, C. & Clark, A. G. Recent and ongoing selection in the human genome. *Nature Rev. Genet.* **8**, 857–868 (2007).
13. Sabeti, P. C. *et al.* Positive natural selection in the human lineage. *Science* **312**, 1614–1620 (2006).
14. Akey, J. M. Constructing genomic maps of positive selection in humans: where do we go from here? *Genome Res.* **19**, 711–722 (2009).
References 12–14 are excellent reviews on how recent positive selection has targeted the human genome. They discuss the major findings of genome-wide scans for selection and their strengths and weaknesses.
15. Trowsdale, J. & Parham, P. Mini-review: defense strategies and immunity-related genes. *Eur. J. Immunol.* **34**, 7–17 (2004).
16. Varki, A. A chimpanzee genome project is a biomedical imperative. *Genome Res.* **10**, 1065–1070 (2000).
17. Varki, A. & Altheide, T. K. Comparing the human and chimpanzee genomes: searching for needles in a haystack. *Genome Res.* **15**, 1746–1758 (2005).
18. Yang, Z. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Mol. Biol. Evol.* **15**, 568–573 (1998).
19. Arbiza, L., Dopazo, J. & Dopazo, H. Positive selection, relaxation, and acceleration in the evolution of the human and chimp genome. *PLoS Comput. Biol.* **2**, e38 (2006).
20. Bustamante, C. D. *et al.* Natural selection on protein-coding genes in the human genome. *Nature* **437**, 1153–1157 (2005).
This paper assesses the genome-wide impact of weak negative and positive selection on protein-coding regions since humans and chimpanzees started to diverge.
21. The Chimpanzee Sequencing and Analysis Consortium. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* **437**, 69–87 (2005).
22. Gibbs, R. A. *et al.* Evolutionary and biomedical insights from the rhesus macaque genome. *Science* **316**, 222–234 (2007).
23. Nielsen, R. *et al.* A scan for positively selected genes in the genomes of humans and chimpanzees. *PLoS Biol.* **3**, e170 (2005).
24. Kosiol, C. *et al.* Patterns of positive selection in six mammalian genomes. *PLoS Genet.* **4**, e1000144 (2008).
25. International Chicken Genome Sequencing Consortium. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* **432**, 695–716 (2004).
26. Clark, A. G. *et al.* Inferring nonneutral evolution from human–chimp–mouse orthologous gene trios. *Science* **302**, 1960–1963 (2003).
The first genome-wide analysis of genes evolving under positive selection in the human evolutionary lineage.
27. Bakewell, M. A., Shi, P. & Zhang, J. More genes underwent positive selection in chimpanzee evolution than in human evolution. *Proc. Natl Acad. Sci. USA* **104**, 7489–7494 (2007).
28. Vamathevan, J. J. *et al.* The role of positive selection in determining the molecular cause of species differences in disease. *BMC Evol. Biol.* **8**, 273 (2008).
29. Sharp, P. M., Shaw, G. M. & Hahn, B. H. Simian immunodeficiency virus infection of chimpanzees. *J. Virol.* **79**, 3891–3902 (2005).
30. Silvestri, G. Immunity in natural SIV infections. *J. Intern. Med.* **265**, 97–109 (2009).
31. Keele, B. F. *et al.* Increased mortality and AIDS-like immunopathology in wild chimpanzees infected with SIVcpz. *Nature* **460**, 515–519 (2009).
32. Seeler, J. S., Muchardt, C., Suessle, A. & Gaynor, R. B. Transcription factor PRDII-BF1 activates human immunodeficiency virus type 1 gene expression. *J. Virol.* **68**, 1002–1009 (1994).
33. Gonzalez, E. *et al.* The influence of *CCL3L1* gene-containing segmental duplications on HIV-1/AIDS susceptibility. *Science* **307**, 1434–1440 (2005).
34. Degenhardt, J. D. *et al.* Copy number variation of *CCL3*-like genes affects rate of progression to simian-AIDS in rhesus macaques (*Macaca mulatta*). *PLoS Genet.* **5**, e1000346 (2009).
35. Baum, J., Ward, R. H. & Conway, D. J. Natural selection on the erythrocyte surface. *Mol. Biol. Evol.* **19**, 223–229 (2002).
36. Wang, H. Y., Tang, H., Shen, C. K. & Wu, C. I. Rapidly evolving genes in human. I. The glycophorins and their possible role in evading malaria parasites. *Mol. Biol. Evol.* **20**, 1795–1804 (2003).
37. Wilder, J. A., Hewett, E. K. & Gansner, M. E. Molecular evolution of *GYPC*: evidence for recent structural innovation and positive selection in humans. *Mol. Biol. Evol.* 13 Aug 2009 (doi:10.1093/molbev/msp183).
38. Maier, A. G. *et al. Plasmodium falciparum* erythrocyte invasion through glycophorin C and selection for Gerbich negativity in human populations. *Nature Med.* **9**, 87–92 (2003).
39. Sim, B. K., Chitnis, C. E., Wasniowska, K., Hadley, T. J. & Miller, L. H. Receptor and ligand domains for invasion of erythrocytes by *Plasmodium falciparum*. *Science* **264**, 1941–1944 (1994).
40. Martin, M. J., Rayner, J. C., Gagneux, P., Barnwell, J. W. & Varki, A. Evolution of human–chimpanzee differences in malaria susceptibility: relationship to human genetic loss of *N*-glycolylneuraminic acid. *Proc. Natl Acad. Sci. USA* **102**, 12819–12824 (2005).
41. King, M. C. & Wilson, A. C. Evolution at two levels in humans and chimpanzees. *Science* **188**, 107–116 (1975).
42. Blekhman, R., Oshlack, A., Chabot, A. E., Smyth, G. K. & Gilad, Y. Gene regulation in primates evolves under tissue-specific selection pressures. *PLoS Genet.* **4**, e1000271 (2008).
43. Hahn, M. W., Demuth, J. P. & Han, S. G. Accelerated rate of gene gain and loss in primates. *Genetics* **177**, 1941–1949 (2007).
44. Wang, X., Grus, W. E. & Zhang, J. Gene losses during human origins. *PLoS Biol.* **4**, e52 (2006).
45. Dumas, L. *et al.* Gene copy number variation spanning 60 million years of human and primate evolution. *Genome Res.* **17**, 1266–1277 (2007).
46. Carlson, C. S. *et al.* Genomic regions exhibiting positive selection identified from dense genotype data. *Genome Res.* **15**, 1553–1565 (2005).
47. Kelley, J. L., Madeoy, J., Calhoun, J. C., Swanson, W. & Akey, J. M. Genomic signatures of positive selection in humans and the limits of outlier approaches. *Genome Res.* **16**, 980–989 (2006).
48. Williamson, S. H. *et al.* Localizing recent adaptive evolution in the human genome. *PLoS Genet.* **3**, e90 (2007).
49. Barreiro, L. B., Laval, G., Quach, H., Patin, E. & Quintana-Murci, L. Natural selection has driven population differentiation in modern humans. *Nature Genet.* **40**, 340–345 (2008).

50. Kimura, R., Fujimoto, A., Tokunaga, K. & Ohashi, J. A practical genome scan for population-specific strong selective sweeps that have reached fixation. *PLoS ONE* **2**, e286 (2007).
51. Sabeti, P. C. *et al.* Genome-wide detection and characterization of positive selection in human populations. *Nature* **449**, 913–918 (2007).
52. Tang, K., Thornton, K. R. & Stoneking, M. A new approach for using genome scans to detect recent positive selection in the human genome. *PLoS Biol.* **5**, e171 (2007).
53. Voight, B. F., Kudaravalli, S., Wen, X. & Pritchard, J. K. A map of recent positive selection in the human genome. *PLoS Biol.* **4**, e72 (2006).
This paper reports a genome-wide scan, based on haplotype structure information, for recent and ongoing positive selection in human populations.
54. Wang, E. T., Kodama, G., Baldi, P. & Moyzis, R. K. Global landscape of recent inferred Darwinian selection for *Homo sapiens*. *Proc. Natl Acad. Sci. USA* **103**, 135–140 (2006).
55. Wolfe, N. D., Dunavan, C. P. & Diamond, J. Origins of major human infectious diseases. *Nature* **447**, 279–283 (2007).
56. Anderson, R. M. & May, R. M. *Infectious Diseases of Humans: Dynamics and Control* (Oxford Univ. Press, 1991).
57. Fumagalli, M. *et al.* Widespread balancing selection and pathogen-driven selection at blood group antigen genes. *Genome Res.* **19**, 199–212 (2009).
58. Fumagalli, M. *et al.* Parasites represent a major selective force for interleukin genes and shape the genetic predisposition to autoimmune conditions. *J. Exp. Med.* **206**, 1395–1408 (2009).
59. Prugnolle, F. *et al.* Pathogen-driven selection and worldwide HLA class I diversity. *Curr. Biol.* **15**, 1022–1027 (2005).
The first study to correlate the worldwide genetic diversity at HLA class I genes with pathogen presence.
60. Hughes, A. L. & Nei, M. Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature* **335**, 167–170 (1988).
61. Hedrick, P. W., Whittam, T. S. & Parham, P. Heterozygosity at individual amino acid sites: extremely high levels for HLA-A and -B genes. *Proc. Natl Acad. Sci. USA* **88**, 5897–5901 (1991).
62. Takahata, N., Satta, Y. & Klein, J. Polymorphism and balancing selection at major histocompatibility complex loci. *Genetics* **130**, 925–938 (1992).
References 60–62 describe how the extraordinary levels of genetic diversity observed at the major histocompatibility complex may result from the action of balancing selection.
63. Bamshad, M. J. *et al.* A strong signature of balancing selection in the 5′ *cis*-regulatory region of *CCR5*. *Proc. Natl Acad. Sci. USA* **99**, 10539–10544 (2002).
64. Novembre, J., Galvani, A. P. & Slatkin, M. The geographic spread of the *CCR5 Δ32* HIV-resistance allele. *PLoS Biol.* **3**, e339 (2005).
65. Stephens, J. C. *et al.* Dating the origin of the *CCR5-Δ32* AIDS-resistance allele by the coalescence of haplotypes. *Am. J. Hum. Genet.* **62**, 1507–1515 (1998).
66. Sabeti, P. C. *et al.* The case for selection at *CCR5-Δ32*. *PLoS Biol.* **3**, e378 (2005).
67. Arenzana-Seisdedos, F. & Parmentier, M. Genetics of resistance to HIV infection: role of co-receptors and co-receptor ligands. *Semin. Immunol.* **18**, 387–403 (2006).
68. Gonzalez, E. *et al.* Race-specific HIV-1 disease-modifying effects associated with *CCR5* haplotypes. *Proc. Natl Acad. Sci. USA* **96**, 12004–12009 (1999).
69. Lalani, A. S. *et al.* Use of chemokine receptors by poxviruses. *Science* **286**, 1968–1971 (1999).
70. Single, R. M. *et al.* Global diversity and evidence for coevolution of *KIR* and *HLA*. *Nature Genet.* **39**, 1114–1119 (2007).
71. Chaix, R., Cao, C. & Donnelly, P. Is mate choice in humans MHC-dependent? *PLoS Genet.* **4**, e1000184 (2008).
72. Ober, C. *et al.* HLA and mate choice in humans. *Am. J. Hum. Genet.* **61**, 497–504 (1997).
73. Parham, P. & Ohta, T. Population biology of antigen presentation by MHC class I molecules. *Science* **272**, 67–74 (1996).
74. Snow, R. W., Guerra, C. A., Noor, A. M., Myint, H. Y. & Hay, S. I. The global distribution of clinical episodes of *Plasmodium falciparum* malaria. *Nature* **434**, 214–217 (2005).
75. Kwiatkowski, D. P. How malaria has affected the human genome and what human genetics can teach us about malaria. *Am. J. Hum. Genet.* **77**, 171–192 (2005).
The most comprehensive review of the genetic basis of resistance to malaria in humans and the selective pressures that are exerted by malaria in the human genome.
76. Flint, J., Harding, R. M., Boyce, A. J. & Clegg, J. B. The population genetics of the haemoglobinopathies. *Baillieres Clin. Haematol.* **11**, 1–51 (1998).
77. Agarwal, A. *et al.* Hemoglobin C associated with protection from severe malaria in the Dogon of Mali, a West African population with a low prevalence of hemoglobin S. *Blood* **96**, 2358–2363 (2000).
78. Modiano, D. *et al.* Haemoglobin C protects against clinical *Plasmodium falciparum* malaria. *Nature* **414**, 305–308 (2001).
79. Chotivanich, K. *et al.* Hemoglobin E: a balanced polymorphism protective against high parasitemias and thus severe *P. falciparum* malaria. *Blood* **100**, 1172–1176 (2002).
80. Ohashi, J. *et al.* Extended linkage disequilibrium surrounding the hemoglobin E variant due to malarial selection. *Am. J. Hum. Genet.* **74**, 1198–1208 (2004).
81. Chitnis, C. E. & Miller, L. H. Identification of the erythrocyte binding domains of *Plasmodium vivax* and *Plasmodium knowlesi* proteins involved in erythrocyte invasion. *J. Exp. Med.* **180**, 497–506 (1994).
82. Tournamille, C., Colin, Y., Cartron, J. P. & Le Van Kim, C. Disruption of a GATA motif in the Duffy gene promoter abolishes erythroid gene expression in Duffy-negative individuals. *Nature Genet.* **10**, 224–228 (1995).
83. Cavalli-Sforza, L. L., Menozzi, P. & Piazza, A. *The History and Geography of Human Genes* (Princeton Univ. Press, 1994).
84. Hamblin, M. T. & Di Rienzo, A. Detection of the signature of natural selection in humans: evidence from the Duffy blood group locus. *Am. J. Hum. Genet.* **66**, 1669–1679 (2000).
85. Hamblin, M. T., Thompson, E. E. & Di Rienzo, A. Complex signatures of natural selection at the Duffy blood group locus. *Am. J. Hum. Genet.* **70**, 369–383 (2002).
86. Aitman, T. J. *et al.* Malaria susceptibility and *CD36* mutation. *Nature* **405**, 1015–1016 (2000).
87. Ruwende, C. & Hill, A. Glucose-6-phosphate dehydrogenase deficiency and malaria. *J. Mol. Med.* **76**, 581–588 (1998).
88. Sabeti, P. *et al.* CD40L association with protection from severe malaria. *Genes Immun.* **3**, 286–291 (2002).
89. The International HapMap Consortium. A haplotype map of the human genome. *Nature* **437**, 1299–1320 (2005).
90. Sabeti, P. C. *et al.* Detecting recent positive selection in the human genome from haplotype structure. *Nature* **419**, 832–837 (2002).
91. Currat, M. *et al.* Molecular analysis of the β-globin gene cluster in the Niokholo Mandenka population reveals a recent origin of the β^S Senegal mutation. *Am. J. Hum. Genet.* **70**, 207–223 (2002).
92. Saunders, M. A., Hammer, M. F. & Nachman, M. W. Nucleotide variability at *G6pd* and the signature of malarial selection in humans. *Genetics* **162**, 1849–1861 (2002).
93. Tishkoff, S. A. *et al.* Haplotype diversity and linkage disequilibrium at human *G6PD*: recent origin of alleles that confer malarial resistance. *Science* **293**, 455–462 (2001).
94. Joy, D. A. *et al.* Early origin and recent expansion of *Plasmodium falciparum*. *Science* **300**, 318–321 (2003).
95. Coluzzi, M., Sabatini, A., della Torre, A., Di Deco, M. A. & Petrarca, V. A polytene chromosome analysis of the *Anopheles gambiae* species complex. *Science* **298**, 1415–1418 (2002).
96. Barreiro, L. B. *et al.* Evolutionary dynamics of human Toll-like receptors and their different contributions to host defense. *PLoS Genet.* **5**, e1000562 (2009).
The first evolutionary dissection of how natural selection has targeted the *TLR* family members in humans.
97. Dobson, A. in *The Cambridge Encyclopedia of Human Evolution* (eds Jones, S., Martin, R. & Pilbeam, D.) 411–420 (Cambridge Univ. Press, 1992).
98. Quintana-Murci, L., Alcais, A., Abel, L. & Casanova, J. L. Immunology *in natura*: clinical, epidemiological and evolutionary genetics of infectious diseases. *Nature Immunol.* **8**, 1165–1171 (2007).
99. Garred, P., Larsen, F., Seyfarth, J., Fujita, R. & Madsen, H. O. Mannose-binding lectin and its genetic variants. *Genes Immun.* **7**, 85–94 (2006).
100. Casanova, J. L. & Abel, L. Human mannose-binding lectin in immunity: friend, foe, or both? *J. Exp. Med.* **199**, 1295–1299 (2004).
101. Eisen, D. P. & Minchinton, R. M. Impact of mannose-binding lectin on susceptibility to infectious diseases. *Clin. Infect. Dis.* **37**, 1496–1505 (2003).
102. Verdu, P. *et al.* Evolutionary insights into the high worldwide prevalence of *MBL2* deficiency alleles. *Hum. Mol. Genet.* **15**, 2650–2658 (2006).
103. Bernig, T. *et al.* Sequence analysis of the mannose-binding lectin (*MBL2*) gene reveals a high degree of heterozygosity with evidence of selection. *Genes Immun.* **5**, 461–476 (2004).
104. Mukherjee, S., Sarkar-Roy, N., Wagener, D. K. & Majumder, P. P. Signatures of natural selection are not uniform across genes of innate immune system, but purifying selection is the dominant signature. *Proc. Natl Acad. Sci. USA* **106**, 7073–7078 (2009).
105. Walsh, E. C. *et al.* Searching for signals of evolutionary selection in 168 genes related to immune function. *Hum. Genet.* **119**, 92–102 (2006).
106. Lynch, N. J. *et al.* L-ficolin specifically binds to lipoteichoic acid, a cell wall constituent of Gram-positive bacteria, and activates the lectin pathway of complement. *J. Immunol.* **172**, 1198–1202 (2004).
107. Roos, A. *et al.* Antibody-mediated activation of the classical pathway of complement may compensate for mannose-binding lectin deficiency. *Eur. J. Immunol.* **34**, 2589–2598 (2004).
108. Hayashi, F. *et al.* The innate immune response to bacterial flagellin is mediated by Toll-like receptor 5. *Nature* **410**, 1099–1103 (2001).
109. Hawn, T. R. *et al.* A common dominant TLR5 stop codon polymorphism abolishes flagellin signaling and is associated with susceptibility to legionnaires' disease. *J. Exp. Med.* **198**, 1563–1572 (2003).
110. Wlasiuk, G., Khan, S., Switzer, W. M. & Nachman, M. W. A history of recurrent positive selection at the Toll-like receptor 5 in primates. *Mol. Biol. Evol.* **26**, 937–949 (2009).
111. Hawn, T. R. *et al.* A stop codon polymorphism of Toll-like receptor 5 is associated with resistance to systemic lupus erythematosus. *Proc. Natl Acad. Sci. USA* **102**, 10593–10597 (2005).
112. Franchi, L. *et al.* Cytosolic flagellin requires Ipaf for activation of caspase-1 and interleukin 1β in salmonella-infected macrophages. *Nature Immunol.* **7**, 576–582 (2006).
113. Miao, E. A., Andersen-Nissen, E., Warren, S. E. & Aderem, A. TLR5 and Ipaf: dual sensors of bacterial flagellin in the innate immune system. *Semin. Immunopathol.* **29**, 275–288 (2007).
114. Olson, M. V. When less is more: gene loss as an engine of evolutionary change. *Am. J. Hum. Genet.* **64**, 18–23 (1999).
115. Hirayasu, K. *et al.* Evidence for natural selection on leukocyte immunoglobulin-like receptors for HLA class I in Northeast Asians. *Am. J. Hum. Genet.* **82**, 1075–1083 (2008).
116. Seixas, S. *et al.* Sequence diversity at the proximal 14q32.1 *SERPIN* subcluster: evidence for natural selection favoring the pseudogenization of *SERPINA2*. *Mol. Biol. Evol.* **24**, 587–598 (2007).
117. Xue, Y. *et al.* Spread of an inactive form of caspase-12 in humans is due to recent positive selection. *Am. J. Hum. Genet.* **78**, 659–670 (2006).
An intriguing example of how the loss of an immunity-related gene can represent a selective advantage in humans.
118. Yngvadottir, B. *et al.* A genome-wide survey of the prevalence and evolutionary forces acting on human nonsense SNPs. *Am. J. Hum. Genet.* **84**, 224–234 (2009).
119. Saleh, M. *et al.* Differential modulation of endotoxin responsiveness by human caspase-12 polymorphisms. *Nature* **429**, 75–79 (2004).
120. Scott, A. M. & Saleh, M. The inflammatory caspases: guardians against infections and sepsis. *Cell Death Differ.* **14**, 23–31 (2007).
121. Perry, G. H. *et al.* Copy number variation and evolution in humans and chimpanzees. *Genome Res.* **18**, 1698–1710 (2008).
122. Blekhman, R. *et al.* Natural selection on genes that underlie human disease susceptibility. *Curr. Biol.* **18**, 883–889 (2008).
123. Nielsen, R. *et al.* Darwinian and demographic forces affecting human protein coding genes. *Genome Res.* **19**, 838–849 (2009).

124. Picard, C. *et al.* Pyogenic bacterial infections in humans with IRAK-4 deficiency. *Science* **299**, 2076–2079 (2003).
125. von Bernuth, H. *et al.* Pyogenic bacterial infections in humans with MyD88 deficiency. *Science* **321**, 691–696 (2008).
126. Zhang, S. Y. *et al.* TLR3 deficiency in patients with herpes simplex encephalitis. *Science* **317**, 1522–1527 (2007).
127. Casanova, J. L. & Abel, L. Human genetics of infectious diseases: a unified theory. *EMBO J.* **26**, 915–922 (2007).
An important review of the genetic basis of infectious diseases in humans, bridging the dysfunctional gap between Mendelian genetics and population-based complex genetics.
128. Di Rienzo, A. Population genetics models of common diseases. *Curr. Opin. Genet. Dev.* **16**, 630–636 (2006).
129. Smirnova, I., Hamblin, M. T., McBride, C., Beutler, B. & Di Rienzo, A. Excess of rare amino acid polymorphisms in the Toll-like receptor 4 in humans. *Genetics* **158**, 1657–1664 (2001).
130. Schroder, N. W. & Schumann, R. R. Single nucleotide polymorphisms of Toll-like receptors and susceptibility to infectious disease. *Lancet Infect. Dis.* **5**, 156–164 (2005).
131. Ma, X. *et al.* Full-exon resequencing reveals Toll-like receptor variants contribute to human susceptibility to tuberculosis disease. *PLoS ONE* **2**, e1318 (2007).
132. Misch, E. A. *et al.* Human TLR1 deficiency is associated with impaired mycobacterial signaling and protection from leprosy reversal reaction. *PLoS Negl. Trop. Dis.* **2**, e231 (2008).
133. Pickrell, J. K. *et al.* Signals of recent positive selection in a worldwide sample of human populations. *Genome Res.* **19**, 826–837 (2009).
134. Gilad, Y., Rifkin, S. A. & Pritchard, J. K. Revealing the architecture of gene regulation: the promise of eQTL studies. *Trends Genet.* **24**, 408–415 (2008).
135. Dimas, A. S. *et al.* Modifier effects between regulatory and protein-coding variation. *PLoS Genet.* **4**, e1000244 (2008).
136. Stranger, B. E. *et al.* Population genomics of human gene expression. *Nature Genet.* **39**, 1217–1224 (2007).
137. Kudaravalli, S., Veyrieras, J. B., Stranger, B. E., Dermitzakis, E. T. & Pritchard, J. K. Gene expression levels are a target of recent natural selection in the human genome. *Mol. Biol. Evol.* **26**, 649–658 (2009).
138. Dimas, A. S. *et al.* Common regulatory variation impacts gene expression in a cell type-dependent manner. *Science* **325**, 1246–1250 (2009).
139. Ko, D. C. *et al.* A genome-wide *in vitro* bacterial-infection screen reveals human variation in the host response associated with inflammatory disease. *Am. J. Hum. Genet.* **85**, 214–227 (2009).
140. Janeway, C. A. Jr & Medzhitov, R. Innate immune recognition. *Annu. Rev. Immunol.* **20**, 197–216 (2002).
141. Beutler, B. *et al.* Genetic analysis of host resistance: Toll-like receptor signaling and immunity at large. *Annu. Rev. Immunol.* **24**, 353–389 (2006).
142. Medzhitov, R. Toll-like receptors and innate immunity. *Nature Rev. Immunol.* **1**, 135–145 (2001).
143. Leulier, F. & Lemaitre, B. Toll-like receptors — taking an evolutionary approach. *Nature Rev. Genet.* **9**, 165–178 (2008).
An outstanding overview of the origins, evolution and different functions of TLRs across the animal kingdom.
144. Kawai, T. & Akira, S. Innate immune recognition of viral infection. *Nature Immunol.* **7**, 131–137 (2006).
145. Geijtenbeek, T. B., van Vliet, S. J., Engering, A., 't Hart, B. A. & van Kooyk, Y. Self- and nonself-recognition by C-type lectins on dendritic cells. *Annu. Rev. Immunol.* **22**, 33–54 (2004).
146. Fritz, J. H., Ferrero, R. L., Philpott, D. J. & Girardin, S. E. Nod-like proteins in immunity, inflammation and disease. *Nature Immunol.* **7**, 1250–1257 (2006).
147. Takeuchi, O. & Akira, S. Recognition of viruses by innate immunity. *Immunol. Rev.* **220**, 214–224 (2007).
148. Schatz, D. G., Oettinger, M. A. & Schlissel, M. S. V(D)J recombination: molecular biology and regulation. *Annu. Rev. Immunol.* **10**, 359–383 (1992).
149. Bendelac, A., Bonneville, M. & Kearney, J. F. Autoreactivity by design: innate B and T lymphocytes. *Nature Rev. Immunol.* **1**, 177–186 (2001).
150. Klein, J. *Natural History of the Major Histocompatibility Complex* (Wiley, New York, 1986).
151. McDevitt, H. O. Discovering the role of the major histocompatibility complex in the immune response. *Annu. Rev. Immunol.* **18**, 1–17 (2000).
152. Hurst, L. D. Genetics and the understanding of selection. *Nature Rev. Genet.* **10**, 83–93 (2009).
153. Nielsen, R. Molecular signatures of natural selection. *Annu. Rev. Genet.* **39**, 197–218 (2005).
154. Kreitman, M. Methods to detect selection in populations with applications to the human. *Annu. Rev. Genomics Hum. Genet.* **1**, 539–559 (2000).
References 152–154 are thorough reviews on how natural selection acts, its signatures and implications, and the different methods used for detecting selection based on population genetic data.
155. Eyre-Walker, A. & Keightley, P. D. High genomic deleterious mutation rates in hominids. *Nature* **397**, 344–347 (1999).
156. Kryukov, G. V., Pennacchio, L. A. & Sunyaev, S. R. Most rare missense alleles are deleterious in humans: implications for complex disease and association studies. *Am. J. Hum. Genet.* **80**, 727–739 (2007).
157. Przeworski, M., Coop, G. & Wall, J. D. The signature of positive selection on standing genetic variation. *Evolution* **59**, 2312–2323 (2005).
158. Charlesworth, D. Balancing selection and its effects on sequences in nearby genome regions. *PLoS Genet.* **2**, e64 (2006).
159. McDonald, J. H. & Kreitman, M. Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature* **351**, 652–654 (1991).
160. Nielsen, R. & Yang, Z. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* **148**, 929–936 (1998).
161. Yang, Z. & Bielawski, J. P. Statistical methods for detecting molecular adaptation. *Trends Ecol. Evol.* **15**, 496–503 (2000).
162. Guindon, S., Black, M. & Rodrigo, A. Proceedings of the SMBE Tri-National Young Investigators' Workshop 2005. Control of the false discovery rate applied to the detection of positively selected amino acid sites. *Mol. Biol. Evol.* **23**, 919–926 (2006).
163. Thornton, K. R., Jensen, J. D., Becquet, C. & Andolfatto, P. Progress and prospects in mapping recent selection in the genome. *Heredity* **98**, 340–348 (2007).
164. Weir, C. L. & Cockerham, C. C. Estimating F-statistics for the analysis of population structure. *Evolution* **38**, 1358–1370 (1984).
165. Hartl, D. L. & Clark, A. G. *Principles of Population Genetics* (Sinauer Associates Inc., Sunderland, 2007).
166. Phillips, P. C. Epistasis — the essential role of gene interactions in the structure and evolution of genetic systems. *Nature Rev. Genet.* **9**, 855–867 (2008).
167. Vilches, C. & Parham, P. KIR: diverse, rapidly evolving receptors of innate and adaptive immunity. *Annu. Rev. Immunol.* **20**, 217–251 (2002).
168. Rajagopalan, S. & Long, E. O. Understanding how combinations of HLA and KIR genes influence disease. *J. Exp. Med.* **201**, 1025–1029 (2005).
169. Norman, P. J. *et al.* Unusual selection on the KIR3DL1/S1 natural killer cell receptor in Africans. *Nature Genet.* **39**, 1092–1099 (2007).

Acknowledgements
We would like to thank L. Abel, M. Ben-Ali, J. L. Casanova, Y. Gilad, B. Jabri, O. Neyrolles, E. Patin, G. Perry, M. Przeworski, H. Quach and P. Sansonetti for helpful comments and discussions, and J. Akey, M. Carrington and R. Single for sharing data that helped the writing of this Review. L.B.B. is supported by a Human Frontier Science Program postdoctoral fellowship. This work was supported by the Institut Pasteur, the Centre National de la Recherche Scientifique, the Fondation pour la Recherche Médicale, and an Agence Nationale de la Recherche research grant (ANR-08-MIEN-009-01) to L.Q.-M.

Competing interests statement
The authors declare no competing financial interests.

DATABASES
Entrez Gene: http://www.ncbi.nlm.nih.gov/gene
CASP12 | *CCL4L2* | *CCR5* | *CD40LG* | *DARC* | *G6PD* | *GYPA* | *GYPC* | *HBB* | *HIVEP3* | *HLA-B* | *MBL2* | *TLR1* | *TLR3* | *TLR4* | *TLR5*

FURTHER INFORMATION
Lluís Quintana-Murci's laboratory: http://www.pasteur.fr/research/heg
1000 Genomes Project: http://www.1000genomes.org
Haplotter genome browser: http://hg-wen.uchicago.edu/selection/haplotter.htm
Human Genome Diversity Project selection browser: http://hgdp.uchicago.edu/cgi-bin/gbrowse/HGDP
Innate Immunity Program in Genomics Applications: http://www.pharmgat.org/IIPGA2/index_html
International HapMap Project: http://www.hapmap.org
National Human Genome Research Institute catalogue of published GWA studies: http://www.genome.gov/26525384
National Institute of Environmental Health Sciences Environmental Genome Project: http://egp.gs.washington.edu
SeattleSNPs database: http://pga.gs.washington.edu

SUPPLEMENTARY INFORMATION
See online article: S1 (table) | S2 (table) | S3 (box)
ALL LINKS ARE ACTIVE IN THE ONLINE PDF