Spatial patterns of variation due to natural selection in humans

John Novembre and Anna Di Rienzo‡*

Abstract | Empowered by technology and sampling efforts designed to facilitate genome-wide association mapping, human geneticists are now studying the geography of genetic variation in unprecedented detail. With high genomic coverage and geographic resolution, these studies are identifying loci with spatial signatures of selection, such as extreme levels of differentiation and correlations with environmental variables. Collectively, patterns at these loci are beginning to provide new insights into the process of human adaptation. Here, we review the challenges of these studies and emerging results, including how human population structure has influenced the response to novel selective pressures.

Although many this year are celebrating the important anniversaries of Darwin's birth (1809) and his publication of *On the Origin of Species* (1869), it is worth recalling that a major intervening event and intellectual milestone for Darwin was his voyage on the HMS *Beagle*. Indeed, both Wallace and Darwin came to understand natural selection not by studying the flora and fauna of a single region but by comparing patterns of variation across geographic regions.

The potential of geographic studies of genetic variation in humans has been recognized for some time, and progress in tapping this potential has benefited from sequential advances in technologies for surveying genetic variation. More than 40 years ago, the innovation of protein electrophoresis allowed researchers to survey allele frequency variation in worldwide population samples at multiple loci[1,2]. Some of the best understood examples of natural selection in humans were discovered in this era, such as the correlation between the geographic distribution of malaria and the sickle cell allele[3] and the extreme geographic differentiation of the null allele of the Duffy blood group[1]. These early studies were limited to loci that could be easily surveyed using classical assays, but these limitations are increasingly falling aside. High-throughput genotyping and sequencing are allowing population geneticists to survey variation on a genomic scale in large worldwide population samples.

The scale of modern data makes it possible to apply new, powerful methods for detecting loci under selection by using data from multiple populations. Theoretical population genetic models have provided some insights into possible signatures of natural selection (BOX 1). For example, if an advantageous variant is due to a novel mutation (as opposed to standing variation), it will initially be present at a single geographic location and will spread outwards from that point in what R.A. Fisher referred to as a 'wave of advance' of the advantageous allele[4]. If selection intensities vary over space, correlations between environmental variables and allele frequencies can arise. The exact outcome in each case depends on various factors, such as how the selective advantage of an allele varies across space and the dispersal patterns in a population (FIGS 1,2). Several methods have emerged for detecting these spatial signatures of selection.

The investigation of spatial patterns at loci under selection can give insights into fundamental questions about geographically variable traits in humans and how humans have evolved in response to novel selective pressures. Although humans are overwhelmingly genetically similar, marked geographic patterns have been observed for many heritable traits. Such traits include disease risk, pathogen resistance and variable drug response, in addition to physiological characteristics, such as skin pigmentation and body mass. In some cases, correlations with environmental variables suggest that natural selection shaped the global distribution of a trait; examples include correlations between skin pigmentation and latitude[5], lactose tolerance and milk consumption[6–8], and body mass and weather temperature[9,10]. The study of variation for these traits prompts many questions. For example, what combination of selection and dispersal led to the current geographic distributions of alleles at these loci? Which loci have geographic patterns of variation that suggest the recent impact of natural selection? And how often do different populations use different

**Department of Ecology and Evolutionary Biology, Interdepartmental Program in Bioinformatics, University of California-Los Angeles, Los Angeles, California 90095-1606, USA.*
‡Department of Human Genetics, University of Chicago, Chicago, Illinois 60637, USA.
e-mails: jnovembre@ucla.edu; dirienzo@uchicago.edu
doi:10.1038/nrg2632
Published online
13 October 2009

Box 1 | **Classical models of selection in structured populations**

Models of selection in structured populations have a long history in theoretical population genetics. Some of the classical questions addressed with these models include: at what speed does an advantageous allele spread geographically through a population? Under what conditions does gene flow homogenize populations sufficiently to prevent local adaptation? And is it plausible that observed genetic variation can be explained by spatially heterogeneous selection? In pursuing these questions, theoreticians developed a collection of mathematical models to describe selection in structured populations.

Dynamic spread of an allele

One of the earliest models was Fisher's 'wave of advance' model[4]. Fisher studied a model in which one region of the habitat is fixed for an advantageous allele and the allele is spreading into territory in which it was formerly absent. Fisher described the moving gradient (that is, the 'wave') of allele frequency that can form (FIG. 1a) and derived the minimum speed of the travelling wave. This minimum speed is of interest for theoretical arguments about the plausibility of geographically distributed populations evolving cohesively at adaptively evolving loci. More recent work has focused on studying the dynamics of the wave in models that include genetic drift[45,47,87].

Stable polymorphisms

Whereas Fisher's model focuses on the non-equilibrium dynamics of a spreading advantageous allele, others considered the problem of whether, at equilibrium, genetic variation could be maintained by spatially varying selection. This interest in stable polymorphisms was driven partly by a desire to explain situations in nature in which genetic variation is stable in a population over time; such polymorphisms are also more easily observed in empirical surveys. The Levene model[88] was one of the first to consider variation in selection coefficients across environments. In its original form it did not include spatially restricted migration, so it presented a particularly challenging scenario for producing stable polymorphisms[88–90] (for a review, see REF. 91). Models that include spatial variation in selective pressures in addition to spatially restricted migration have a much broader range of parameter values and therefore favour stable polymorphisms. A common feature of these models is the formation of gradients (or 'clines') in allele frequency over space. The most well-studied models consider abrupt or continuous environmental transitions between two habitats with differing selection regimes[34,64,92–100] (FIG. 2a,b).

variants to respond to the same selective pressures? Answering these evolutionary questions is relevant for understanding the genetic basis and evolutionary origins of human phenotypic variation, a task that is particularly important for biomedical traits.

In this Review, we describe methods for studying selection based on spatial patterns of genetic variation, keeping in mind the difficulties of distinguishing the outcome of neutral demographic processes and selective events. We first review aspects of background spatial patterns of variation that might confound and/or strongly influence the outcome of selection in humans. We then turn to approaches for using spatial patterns to study signatures of selection and review some of the established and emerging results of these studies. In particular, we outline insights into the sources of variation used by human populations to adapt to novel selective pressures.

Background patterns due to demographic history

Recent large-scale studies of human variation have greatly increased our understanding of the geographic structure of human populations[11–19]. Before investigating the possible spatial signatures of selection, it is crucial to understand the background spatial patterns that are created by human demographic processes.

Clusters and clines. Genome-wide studies using the Human Genome Diversity Project (HGDP) panel[17–21] have revealed evidence for clusters and clines in human genetic data.

At the global scale, there is evidence for at least five major genetic clusters that correspond to broad continental regions[17–19]. The change in the proportion of ancestry across the boundary regions tends to be gradual, and individuals from these regions have ancestry from groups on both sides of the boundary[20]. Although there has been some debate as to whether these clusters are due to uneven sampling[20,22], there are also biological explanations for this pattern. One is isolation by distance (FIG. 3a), which is accentuated by geographic barriers, and the other is secondary contact (FIG. 3b) between differentiated ancestral populations.

In addition to continental-scale clusters, clines of allele frequencies are common in humans: a graded change in allele frequencies is evident in analyses of the relationship between genetic and geographic distance in regions of the world and at a global scale[1,21,23]. Geographic clines may have resulted from the spatial patterning that occurred as humans expanded out of Africa (see below), but they may also have arisen from a long-standing history of spatially restricted gene flow. Taken together, these results imply that although there are continental-scale clusters, allele frequencies change gradually on small geographic scales.

The expansion out of Africa and allele surfing. Studies of the HGDP panel have also highlighted a gradual decrease in genetic diversity levels as a function of the distance from sub-Saharan Africa, a result that is consistent with the serial-founder model of human expansion out of sub-Saharan Africa[21,24]. Importantly, this model can give rise to the 'allele surfing' phenomenon, which produces cline patterns (FIG. 3c) and geographic patches called 'sectors'[25–28]. This phenomenon is a result of the intense amount of genetic drift that occurs at the leading edge of a population expansion. As one set of population founders is further sub-sampled to produce a new set of founders, alleles that are at low frequency in the ancestral population, or new mutations that occur during the expansion, can rapidly rise to high frequency in the newly colonized populations (that is, alleles 'surf' the expansion wave).

All of the above genome-wide patterns are crucially important in assessing and interpreting the evidence for selection in structured populations. This is because some of these patterns closely resemble those expected under specific models of selection. For example, latitudinal clines of allele frequencies could reflect migration patterns or the action of a selective pressure correlated with latitude (for example, ultraviolet (UV) radiation). Similarly, alleles that have surfed may look similar to alleles that have recently undergone positive selection[29,30]. Therefore, a rigorous assessment of the evidence for adaptation requires that background spatial patterns of variation be incorporated in statistical tests of neutrality so that the effects of selection can be distinguished from those of population history alone.

Isolation by distance
A model in which the amount of gene flow between two locations decreases as a function of distance. At equilibrium, this model predicts that genetic differentiation increases as a function of geographic distance. Sometimes the term refers simply to this emergent pattern, rather than the model.

Secondary contact
When two populations that have ceased to exchange migrants begin to re-exchange migrants with one another. In cases in which the populations exchange migrants along a frontier, this boundary is known as a secondary contact zone.

Gene flow
The movement of genes among populations. Often expressed as the proportion of gene copies (or breeding individuals) that are immigrants from a different population.

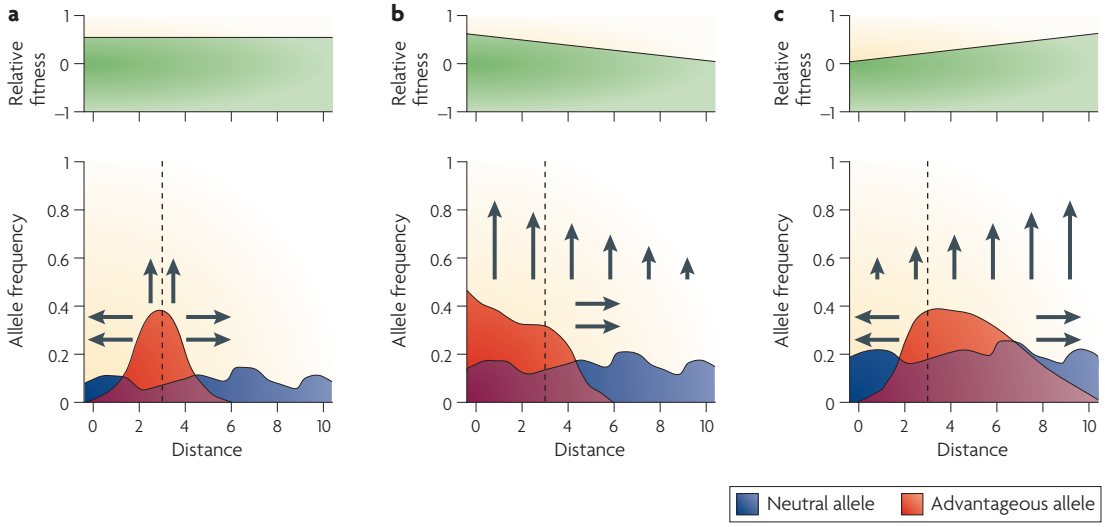

Figure 1 | **The 'wave of advance' spread of a globally advantageous mutation.** Arrows indicate how the allele frequencies of a selected allele (red) are expected to change over time, depending on the pattern of selective advantage of the allele (indicated in green above each graph). Vertical arrows represent the expected magnitude of increase due to selection. Horizontal arrows represent how dispersal homogenizes allele frequencies across space. For every selected allele, a representative neutral allele (blue) of similar average frequency is shown for comparison. In each case the allele arose at location 3 on the *x* axis (marked with a vertical dashed line); the spread will continue until the selected allele is at frequency 1 across the whole habitat. **a** | Uniform selective advantage across space. If the new variant is identically advantageous everywhere, then as the variant increases in frequency it will become exceptionally concentrated around its geographic origin relative to a neutral variant of the same age. One effect of this is the creation of transiently enhanced levels of divergence among populations and clines in allele frequencies that reflect the geographic origin of the allele. **b**,**c** | Non-uniform advantage across space. In scenario **b**, the novel allele is introduced to the regions in which it is most advantageous and rapidly increases in frequency in those regions. This can lead to transient correlations between allele frequency and the environmental factor that drives positive selection. By contrast, in scenario **c**, the novel allele arises in an area distant from where it is most advantageous. It will increase in frequency locally before spreading outwards, and its distribution will carry a strong signature of its geographic origin and be less reflective of spatial variation in selective advantage. These models assume selection acting on new mutations, which may not be the prevailing scenario in humans. Selection on pre-existing variation will complicate these simple scenarios.

Serial-founder model
A model of how novel habitats are colonized in which a source population is first sub-sampled to choose founders who will colonize a neighbouring unoccupied space. This sub-sampling process, which results in a population bottleneck, is repeated sequentially as the population further expands into unoccupied space.

Genetic drift
The fluctuations in allele frequency through time that occur owing to chance.

Variance
A measure of the dispersion of a random variable around its mean value.

Coalescent simulation
An efficient and flexible approach for simulating population genetic data. Ancestral lineages are traced backwards in time, and events in which ancestral lineages have common ancestors (coalescent events) are recorded.

Power
The frequency with which a statistical test rejects the null hypothesis given an alternative hypothesis.

Spatial approaches for detecting selection

F_{ST}-based approaches. Wright's fixation index, F_{ST}, provides a measure of the amount of genetic differentiation among populations[31,32]. This simple statistic has been used to study spatial selection based on the insights of Lewontin and Krakauer[33] (although traces of the idea can be found in several papers at least as early as 1948 (REFS 34,35)). The key insight is that the expected differentiation of allele frequencies between populations is the same at all neutral loci and is determined principally by demographic processes. However, loci that have undergone selection in one population but not another are expected to show higher levels of differentiation (that is, higher F_{ST} values). Importantly, such high F_{ST} values can arise as a result of the transient wave of advance dynamics of a globally advantageous allele (FIG. 1a) or of local selective pressures (FIGS 1b,2). At the other extreme, alleles that are maintained at an equilibrium frequency by balancing selection are expected to exhibit less differentiation (that is, lower F_{ST} values) than is expected at neutral loci (FIG. 2d).

Lewontin and Krakauer considered data from many independent loci and asked whether the variance across loci of the F_{ST} statistic was greater than expected under a model in which all loci evolve neutrally. A difficulty with this approach is that a model of population history has to be specified in order to develop expectations for the F_{ST} distribution under neutrality. An early analysis of human allele frequency data that incorporated a complex demographic model found evidence for an excess of high and low values of F_{ST} relative to expectations[36]. This type of approach has been extended to thousands of SNPs and coupled with coalescent simulations of a population structure model[37], and again it showed an excess of both high and low values of F_{ST}, which is consistent with both adaptive divergence and balancing selection.

Recently, it has become more common to take a complementary approach that avoids overly simplified models of human history. One group of approaches relies on defining sets of functionally related SNPs and then asking whether the distribution of F_{ST} values for these SNPs is substantially different from that for other, putatively neutral, SNPs (for example, by comparing the distributions for genic versus non-genic SNPs). This approach may have the power to detect weak or recent selective pressures that result in small changes in allele frequencies at a sufficiently large number of loci. However, defining groups of functionally related SNPs or genes is not straightforward and, because the number of possible sets is large, the power may be low (owing to

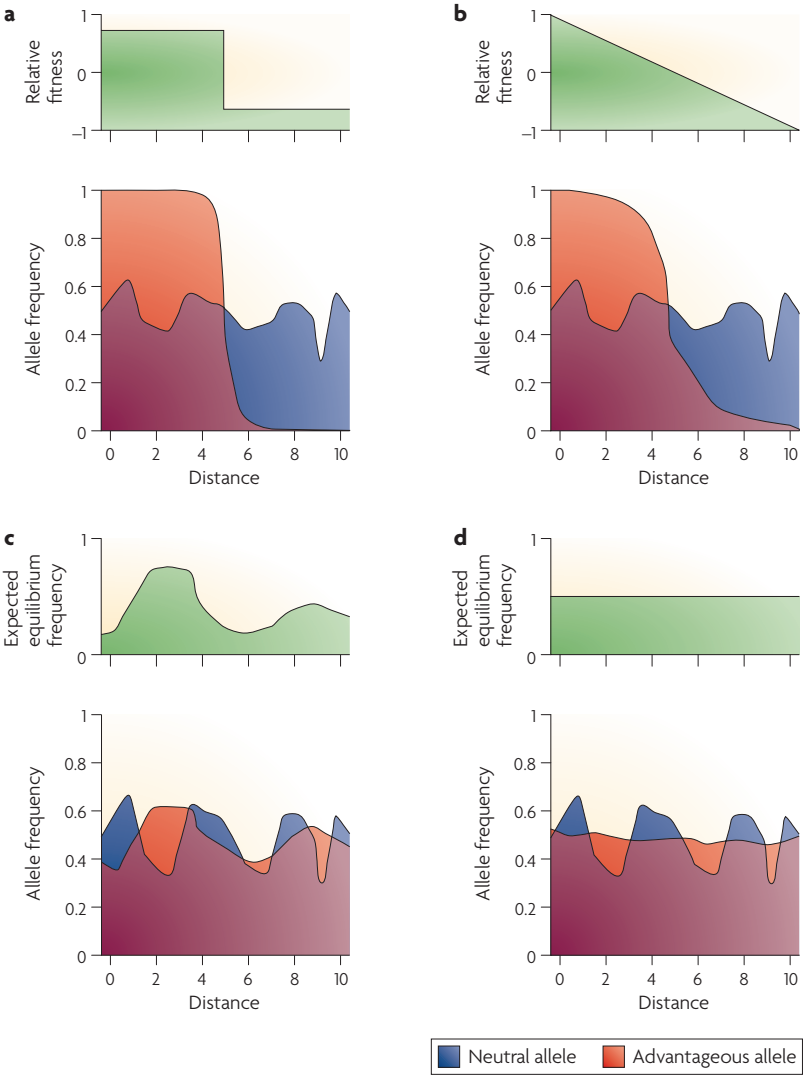

Figure 2 | **Spatially varying selection and stable polymorphisms.** The allele frequency dynamics of a novel variant (red) in a population when the variant is advantageous in some locations and disadvantageous in others (**a**,**b**) or when local balancing selection (for example, due to heterozygote advantage) is operating (**c**,**d**). These scenarios give rise to a stable polymorphism (in which the novel and ancestral variants persist in the population). In these models the novel variant will not replace the ancestral variant — the novel variant will simply become more common in the regions it can disperse to and in which it is advantageous. For every selected allele, a representative neutral allele of a similar average frequency (blue) is shown for reference. The pattern of selective advantage (**a**,**b**) or frequency (**c**,**d**) of the allele is shown in green. **a** | The allele is favoured in some patches and disfavoured in others. In this situation, the allele is fixed in the geographic regions in which it is advantageous and absent in regions in which it is disadvantageous, and there are clines of frequency along the contact points between the two regions. **b** | The allele is favoured in one geographic extreme and disfavoured in the opposite extreme. If the transition from being advantageous to being disadvantageous occurs across a geographic range, rather than being abrupt, broader clines are expected. **c** | Local balancing selection that varies in intensity across space. When selection intensities vary, the local equilibrium frequencies will also vary depending on the environmental factors driving selection, and the allele frequencies will correlate with non-transient environmental factors. A classic example is the sickle cell mutation, which is found at high frequency in regions in which malaria is endemic and decreases in frequency as the prevalence of malaria decreases. **d** | Local balancing selection that is constant across space. Balancing selection can lead to exceptionally constant allele frequencies over space. Scenarios **a**–**c** will generate correlations between allele frequency and environmental factors underlying variation in selection.

multiple-testing correction). Recent genome-wide surveys have shown that the extreme tail of the F_{ST} distribution contains a substantial excess of genic relative to non-genic SNPs and of non-synonymous and 5′ UTR SNPs relative to non-genic SNPs[38,39].

These results suggest that the approach can also be used to identify the individual targets of adaptations. If selection is strong, the loci under selection should exhibit extreme levels of differentiation relative to neutral loci and therefore can be identified through outlier approaches (FIG. 4). With such outlier approaches, the evidence for selection at specific loci is assessed by calculating the proportion of SNPs in a genome-wide distribution that have F_{ST} values that are extreme compared with the value of the SNP of interest (this proportion is sometimes referred to as an 'empirical *p*-value', although it is perhaps better understood as a transformed rank statistic than as a proper statistical *p*-value). Related model-based approaches test for abnormally differentiated loci using the island model[40,41], although these methods are limited in that they can mistake neutral loci as being under selection when population structure is hierarchical[42].

Several individual variants that underlie phenotypes that are known to be advantageous have extreme allele frequency differentiation. Examples include several SNPs in skin pigmentation genes (for example, solute carrier family 24, member 5 (*SLC24A5*)[43], KIT ligand (*KITLG*)[44] and melanocortin 1 receptor (*MC1R*)[39]) or in immune response genes (for example, *FY* and toll-like receptor 6 (*TLR6*)[39]).

Caveats to interpreting the results of F_{ST} ***analyses.*** Although the extremely differentiated variants described above provide plausible examples of advantageous alleles, it is conceivable that many selective pressures in humans are not strong and therefore would be missed by outlier approaches.

Several other caveats apply to the interpretation of the results of F_{ST} analyses. One possible complication is the impact of background selection, whereby strong purifying selection acts repeatedly on a locus. Because strong deleterious alleles are quickly eliminated, it is as if neighbouring loci exist in a population with a much smaller population size (that is, neighbouring loci have a lower effective population size). This enhances the rate of genetic drift and hence differentiation, which results in an excess of high F_{ST} values in comparison with strictly neutral loci[45,46]. Because background selection is likely to act more strongly in genic compared with non-genic regions, negative rather than positive selection could underlie part or all of the observed excess of high F_{ST} for genic relative to non-genic SNPs. To take the effect of background selection into account, different classes of SNPs in coding regions should be compared. At a qualitative level, two recent studies showed a higher enrichment of non-synonymous SNPs than synonymous SNPs in the upper tail of the F_{ST} distribution[38,39], which provides some evidence against background selection as an explanation for the enrichment in high F_{ST} values.

There is a second potential complication — theoretical work has shown that, under some parameter values, the

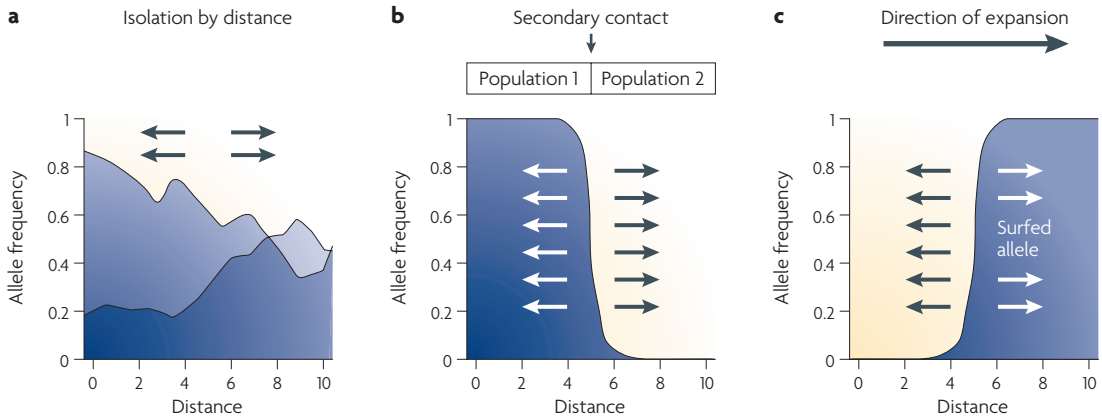

Figure 3 | **Neutral scenarios that produce clines in allele frequencies. a** | Isolation by distance. Under models of isolation by distance, many neutral alleles will show cline patterns, especially along geographic axes with the least gene flow. **b** | Secondary contact between two populations. With secondary contact, neutral alleles will transiently show a cline pattern at the contact zone between the two populations. Note that the allele frequency in the two source populations need not be 0 and 1, as shown here; clines along the secondary contact zone will form even if the allele frequency difference between the two populations is more modest. **c** | Expansion into new territory giving rise to serial founder effects and 'allele surfing'. In this extreme allele surfing scenario, the allele frequency has increased to fixation immediately after expansion. In each case, arrows indicate that none of these patterns is stable — dispersal and drift will erode the clines over time. In all three models, cline patterns that arise can potentially be confused with cline patterns that are expected as a result of selection.

expected levels of differentiation of a spreading adaptive mutation at the selected site are lower (rather than higher) than at neutral loci[47]. However, it is arguable whether the parameter range under which this effect occurs is relevant to humans.

Finally, several factors might bias levels of differentiation in ways that confound F_{ST}-based methods, especially when comparing F_{ST} distributions among large sets of SNPs (for example, genic versus non-genic comparisons). Caution may be necessary to avoid biases due to between-population differences in the frequency spectrum, systematic differences in SNP ascertainment, and genomic features that might influence levels of differentiation (for example, the percentage of GC content and the rates of recombination).

Geographic clines and correlations of allele frequencies with environmental variables. In natural populations of many species, several quantitative traits are distributed clinally. Other traits are correlated with specific environmental variables (for example, temperature), which mirror the selective pressures acting on the phenotypes themselves. Studies of protein and DNA polymorphisms, mainly in fruitflies and humans, have shown that in many cases allele frequencies have spatial patterns that parallel those in adaptive phenotypes, which raises the possibility that gradual changes in allele frequency across space signal adaptations to continuous local environments (FIGS 1b,2b,2c). A notable example is a selected polymorphism in the *Drosophila melanogaster* alcohol dehydrogenase (*Adh*) locus that correlates with latitude in both hemispheres[48]. Consistent with the idea that this spatial pattern reflects adaptations to varying climate, the clines for the *Adh* polymorphism have shifted over the past 20 years in response to climate change[49].

Until recently, analyses of geographic clines did not take into account the effect of population history in assessing the evidence for selection (this is also true for the recently proposed SAM[50]). By contrast, most recent studies (BOX 2; FIG. 5) have used background spatial patterns of allele frequency to guide a null expectation by comparing the correlations between test loci and geographic or environmental variables with the correlations between large collections of unlinked control loci and the same geographic or environmental variables. Spatial evidence for adaptations to continuous environments is inferred if there is an excess of high correlations at test loci compared with the control loci. Using this approach, a study showed that a set of candidate susceptibility alleles for hypertension are more strongly correlated with latitude than hundreds of microsatellites and many SNPs[51,52]. Another study showed that SNPs in candidate genes for the metabolic syndrome are more strongly correlated with climate variables than control SNPs[53]. In an attempt to adjust for the background genetic structure of human populations, this study estimated a null model for the covariance of allele frequencies among populations; it then assessed the evidence for selection by testing whether a linear relationship between a genic SNP and a continuous environmental variable provided a better explanation for the data than the null model.

As for F_{ST}-based approaches, outlier approaches can be applied to the analysis of latitudinal clines and of correlations with environmental variables. In this case, the test statistic measures the evidence for a correlation between allele frequency and latitude or environmental variables (BOX 2).

Importantly, geographic clines are also expected to arise under a neutral model of isolation by distance or if two separated populations have recently come

Multiple-testing correction
When many statistical tests are conducted simultaneously, some tests are expected to have low p-values under the null hypothesis, and therefore a correction is necessary to compensate for this effect.

Island model
A model of population structure in which several island populations exchange migrants symmetrically.

Purifying selection
When natural selection removes novel deleterious mutations from a population.

Effective population size
The population size needed to predict how a locus would evolve (in accordance with the idealized Wright–Fisher model of population genetics) with respect to a property (typically genetic drift). In many complex scenarios, the behaviour of a locus can be predicted with an appropriate effective population size.

Metabolic syndrome
A combination of traits related to type 2 diabetes, obesity, hypertension and altered lipid levels. It is a major risk factor for cardiovascular disease.

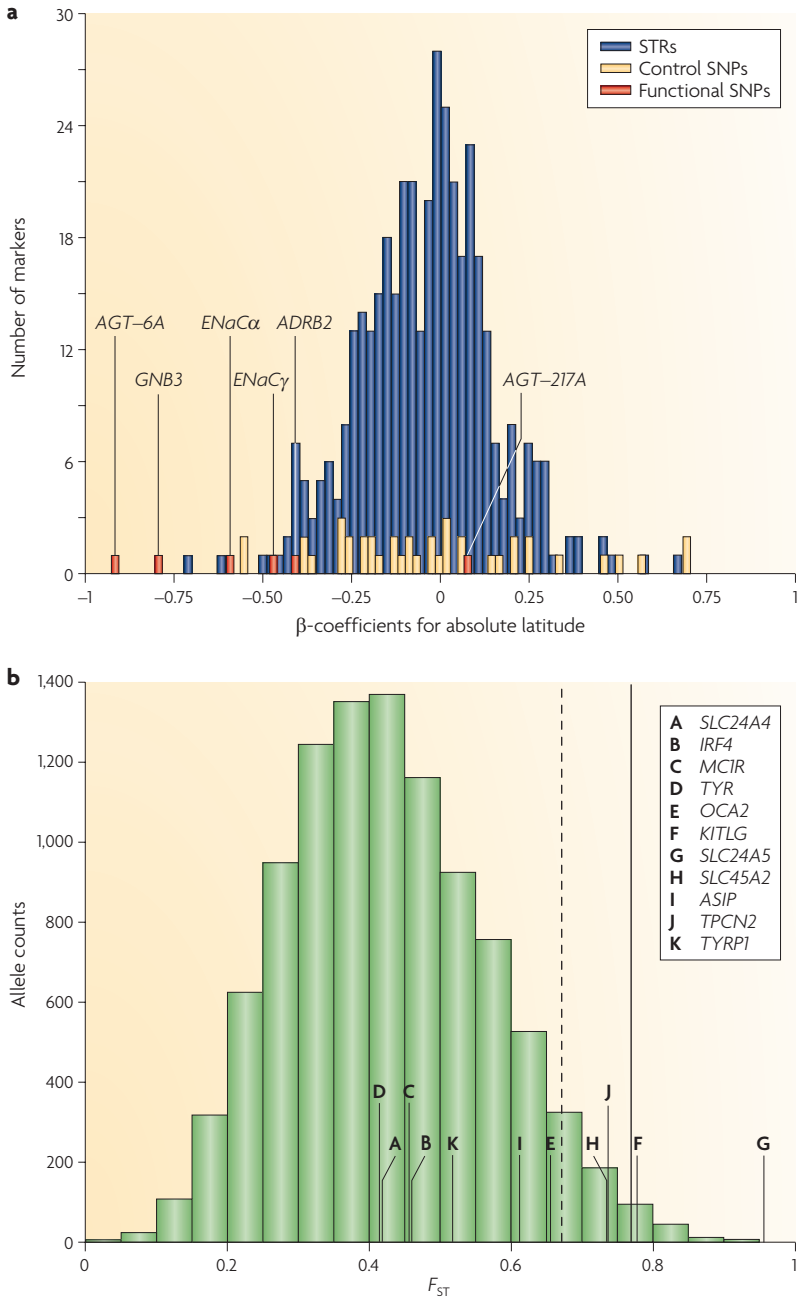

Figure 4 | **Outlier approaches for identifying candidate targets of selection.** Outlier approaches simply rank all SNPs from a large-scale survey based on the value of a test statistic (for example, F_{ST}) and then take all of the SNPs above a specified cut-off point as targets of selection. This is based on the assumption that selection is sufficiently strong to generate extreme spatial patterns compared with the rest of the genome. The power and accuracy of these approaches depends on a number of variables, including the proportion of loci affected by selection and the strength of selection[105]. In a similar vein, a specific candidate SNP can be evaluated against a large collection of SNPs to determine whether the spatial pattern of the candidate SNP is unusual relative to the rest of the genome. **a** | Correlation between allele frequency and latitude for candidate susceptibility SNPs for hypertension (red) compared with random short tandem repeats (STRs; blue) and SNPs (yellow)[52]. **b** | Differentiation of allele frequency in loci (A–K) involved in natural variation in skin pigmentation compared with a large collection of random SNPs (green bars)[75]. The dashed line shows the position beyond which 5% of the random SNPs fall, and the solid line the position beyond which 1% of the random SNPs fall. Part **a** is reproduced from REF. 52. Part **b** is modified, with permission, from REF. 75 © (2009) Cold Spring Harbor Laboratory Press.

into secondary contact (FIG. 3b,c). Moreover, correlations between allele frequencies and environmental variables may arise under neutrality if populations happen to be structured in the same way as the environmental variables (FIG. 3a). If this is not the case, correlations between allele frequencies and environmental variables (FIG. 2c) may provide more robust signals of spatially varying selection than correlations between allele frequencies and geographic clines.

Comparing within-population diversity among populations. If a recent selective sweep has taken place in one population but not in another (owing to local adaptation or a wave of advance), there will be a paucity of variation near the selected site in the population in which the sweep took place. This reduction in variation can be detected by comparing locus-specific diversity levels between the populations[54–57]. Two recent approaches, cross-population extended haplotype homozygosity (XP-EHH)[58] and ln(Rsb)[59], which are based on comparing the levels of local haplotype diversity between two populations, aim to exploit the signature of positive selection on patterns of linked variation. These methods are powerful in cases in which selection drives the allele to fixation or near-fixation in one population but not in the other (as opposed to the integrated haplotype score (iHS) method, which uses only one population and loses power once the beneficial allele becomes nearly fixed[60]). A useful property of these methods is that they can point to the haplotype carrying the advantageous allele and the population in which the adaptation occurred. Signals detected through such approaches include variation at the lactase (*LCT*) locus[61], a well-established target of selection, at the like-glycosyltransferase (*LARGE*) and dystrophin (*DMD*) genes, which have a role in Lassa virus infection[58], and at the ectodysplasin A receptor (*EDAR*) gene[58], which is involved in the development of hair, teeth and exocrine glands.

The history of specific adaptive variants

Most contemporary research focuses on identifying loci that have undergone recent positive selection, but further insights will come from detailed follow-up studies on the spatial distribution of alleles that are putatively adaptive.

If a geographically localized signature is due to a partial sweep, a wave of advance model can be used to estimate the relative strength of positive selection from the spatial distribution of the selected allele. Qualitatively, if a selected allele is at a high frequency near its origin but has not spread out broadly, selection is inferred to be strong relative to dispersal, whereas if the allele disperses broadly before reaching a high frequency near its origin, selection is likely to be weak relative to dispersal. Quantitatively fitting a wave of advance model to spatial data allows the estimation of the ratio of the strength of positive selection to dispersal. Further interpretation of this ratio requires additional information; knowledge of dispersal parameters can provide insights into selection, and information about selection intensity can be used to estimate dispersal. This analysis was

Box 2 | **Recent studies of geographic clines and environmental correlations**

The newly available genome-scale data sets in densely sampled human populations have reawakened interest in the study of geographic clines and correlations between allele frequencies and environmental variables. Recent studies have focused primarily on genes that are plausible candidate susceptibility loci for common human diseases or disease-related traits that have marked inter-ethnic differences in prevalence. These approaches are bolstered by adaptive hypotheses that explain the epidemiology of these phenotypes. For example, the higher prevalence of hypertension, and in particular salt-sensitive hypertension, in African Americans compared with European Americans was hypothesized to reflect adaptations to hot equatorial climates in ancestral African populations[101]. Accordingly, candidate susceptibility variants for hypertensions showed strong clines with latitude in the Human Genome Diversity Project panel[51,52].

These approaches are now being used to inform the identification of polymorphic variants with effects on gene function. An early study of the *TP53* gene, which is a master sensor of stress, detected a latitudinal cline for a common amino acid variant in worldwide populations[102]. Recently, the same polymorphism was found to be correlated with cold winter temperature in East Asians, and a SNP in the murine double minute 2 (*MDM2*) gene — the protein product of which interacts with p53 — was strongly associated with ultraviolet radiation intensity in the same populations[103]. Although there is a plausible biological link between variations in stress response and adaptations to local environments, it will be important to determine whether these observations stand out against background spatial patterns of variation or are consistent with neutral evolution. A role for spatially varying selection in shaping the stress response was recently proposed for the serum and glucocorticoid regulated kinase 1 (*SGK1*) gene. An upstream *SGK1* variant that affects glucocorticoid receptor–DNA binding and glucocorticoid-mediated induction of *SGK1* expression was identified based on unusual levels of allele frequency divergence between populations[104].

Interestingly, clines of genetic diversity have been observed in the human leukocyte antigen (*HLA*) region[81]. More specifically, a significant correlation between *HLA* diversity and pathogen richness was detected, which is consistent with pathogen-driven selection acting on the *HLA* genes. These results suggest a role for balancing selection in adaptations to local environments. If this is the case, some polymorphisms that show allele frequency clines may be maintained by balancing selection that varies continuously in intensity; this variation is in turn linked to a gradual variation in the allele frequencies at equilibrium. Indeed, some variants (for example, *RPTOR* variants) show extremely strong correlations with environmental variables without marked differences in allele frequencies between the extremes of the geographic range[53], as might be expected under strong balancing selection but not under strong directional selection (FIG. 5).

Selective sweep
When a mutation with a beneficial fitness effect arises in a population, natural selection will rapidly increase the frequency of the mutation to a high frequency (partial sweep) or to fixation (complete sweep), which results in a reduction of diversity at and around the selected locus.

RPTOR
(Regulatory-associated protein of mammalian target of rapamycin). The complex between the *RPTOR* gene product and the target of rapamycin is the central component of a nutrient- and hormone-sensitive signalling pathway that regulates cell growth.

attempted for the human chemokine (C-C motif) receptor 5 (*CCR5*) *Δ32* HIV resistance allele[62]; however, the interpretation of the results was complicated by inaccuracies in the estimated recombination rates, which in turn led to incorrect inferences about the age of the allele[63]. Although this approach might have broad applicability, the current wave of advance models are still relatively simple in that, for example, they assume spatially and temporally homogenous dispersal and completely deterministic allele frequency change.

In cases in which the allele in question shows what seems to be a simple geographic cline (FIG. 2), the width of the cline can be used to estimate the strength of selection[64]. One complication is that these methods are only feasible if the cline is known to be due to an abrupt change (FIG. 2a) or a linear gradient (FIG. 2b) in selective pressure. Observations of environmental factors related to the selective pressure may help to distinguish models for the geographic cline (for example, if environmental proxies for the selective pressure are clinal, one might infer that the selective pressure is also clinal).

Insights into the sources of adaptive variation

To adapt to a new selective pressure, a population must: wait for an adaptive variant to be introduced by mutation; wait for an adaptive variant to be introduced via dispersal from a neighbouring population; or use standing variation (for example, wait for a previously neutral segregating variant to become advantageous). Spatial patterns of selected alleles are beginning to provide insights into the importance of these three sources of adaptive genetic variation for human populations.

Multiple mutations underlying adaptive phenotypes. Strikingly, many of the best studied adaptive phenotypes in humans have revealed multiple mutations that confer an advantage to the same or similar selective pressures. Examples include lactase persistence[65–67], skin pigmentation[68] and malaria resistance polymorphisms[69–71]. For instance, a form of malaria resistance caused by glucose-6-phosphate dehydrogenase (G6PD) deficiency is due to two main protein mutations, $G6PD^{A-}$ and $G6PD^{Med}$, which account for the vast majority of G6PD deficiency in sub-Saharan Africa and Europe, respectively. Lactase persistence has a similar pattern at the continental scale, whereby different mutations are found in Europe, Africa and the Middle East[65–67]. Moreover, in Africa and the Middle East, multiple mutations have been shown to contribute to lactose tolerance (note that some of these are shared between populations as if they arose from shared ancestral standing variation, see below). A possible explanation is that the spread of adaptive alleles across continental regions is slow enough and mutation rates to adaptive alleles are high enough to allow novel adaptive mutations to arise in distinct geographic regions before any single variant spreads globally[39]. Indeed, according to theoretical work, population-specific mutations are likely to arise if the population structure is strong and if multiple mutations at multiple loci can give rise to the adaptive trait (that is, if there is genetic redundancy[72,73]). If genetic heterogeneity is common for adaptive phenotypes, it poses challenges for replicating results between populations in association studies and for detecting selection[39,73,74]. Further studies will help to establish whether population-specific adaptive mutations are the rule or whether they are seen only for the extreme examples of selection that have been studied thus far.

Dispersal, demography and the geographic distribution of selected alleles. Two recent analyses of spatial patterns in SNP data from the HGDP and HapMap populations have shown that the most strongly differentiated alleles (and hence the most likely targets of selection) in continental populations are distributed geographically in patterns similar to those expected for neutral genetic variation[39,75]. These results support the notion that population structure is important in shaping the dispersal of selected alleles and, in turn, the outcome of natural selection in humans. Presumably, adaptive mutations have been slow to spread in humans because of low levels of long-range dispersal and because the variant might still be lost by drift in the new population, even if it disperses (especially if it has a weak selective advantage).

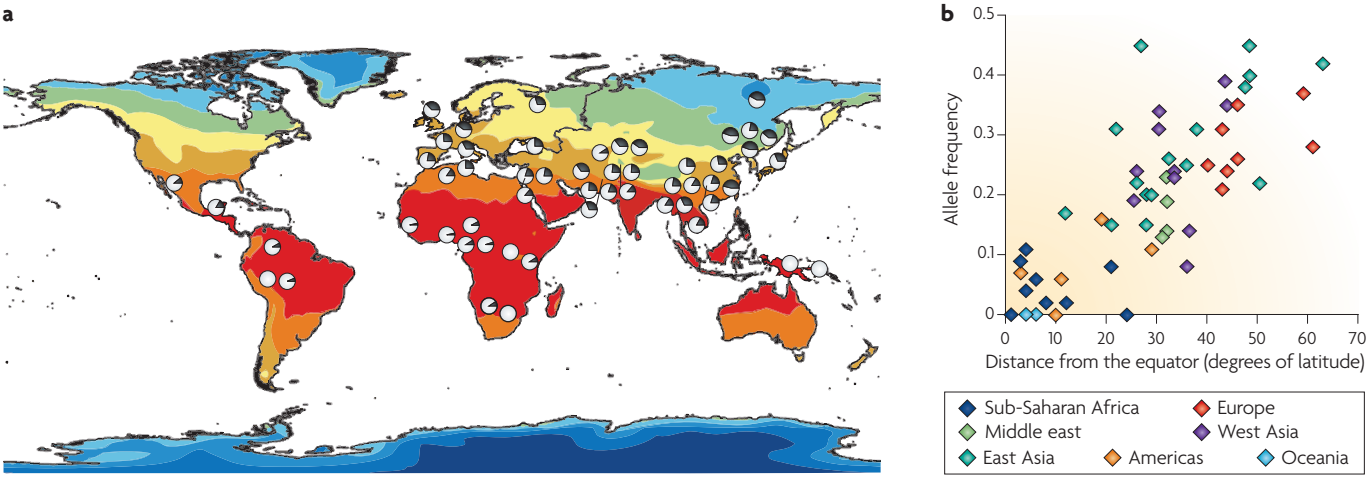

Figure 5 | **Correlation of a SNP in the *RPTOR* gene with environmental variables. a** | Pie charts show the frequency of the derived allele at SNP rs12946049 in the *RPTOR* gene in the Human Genome Diversity Project panel[53]. The colours, ranging from dark blue (cold) to red (hot), represent the maximum temperature in the winter. **b** | Allele frequency at rs12946049 as a function of distance from the equator. Qualitatively, the correlation is convincing for two reasons. First, the variation correlates strongly with an environmental feature in ways that depart from background spatial patterns (for example, distantly related populations in the tropical Americas, Oceania and sub-Saharan Africa share the same environment and have similar allele frequencies). Second, the correlation exists in multiple world regions. Figure is modified from REF. 53.

These studies also suggest that the timing of human population expansions may be important in determining the dispersal and distribution of adaptive alleles. For instance, many alleles at high frequency in East Asian HGDP populations are also found in the Americas, presumably because alleles that were selected in East Asian populations before the colonization of America would have been carried into the Americas at a high frequency. The severity of bottlenecks might also affect the distribution of selected alleles and might explain why more loci were observed as fixed differences between the Han Chinese and Yoruban populations than between the Centre d'Etude du Polymorphisme Humain (CEPH) European and Yoruban populations[39].

A major criticism of the conclusions of these studies is that many of the extremely differentiated SNPs may be neutral[30] and most truly adaptive variants may not be detected because they involve subtle allele frequency shifts or selection on standing variation (see below). Nonetheless, the patterns described above are also observed for a number of extremely differentiated alleles with known phenotypic and fitness consequences, which implies that dispersal has been limiting at least for this set of adaptive variants. For example, several skin pigmentation variants reached high frequency in Western Eurasia but not in East Asia (and vice versa), even though the two groups of populations experience similar degrees of UV radiation[39,76,77].

Selection on standing variation. Owing to the recent origin and dispersal of human populations, there is abundant shared variation among populations. It is plausible that some of that shared variation became adaptive with the onset of environmental challenges posed by new diet, habitat and pathogen pressures. As a result, different populations may have responded to similar selective pressures using the same or different standing alleles. This may create scenarios in which there is limited power to detect signals of selection[39,73,74]. For example, if the trait is complex, adaptation can occur through small shifts in allele frequencies at multiple loci. Even if an adaptive trait is simple and monogenic, the canonical signatures of enhanced linkage disequilibrium (LD) and reduced diversity due to a selective sweep are substantially weakened when positive selection acts on standing variation[78,79]. These systematic deficiencies in power make it challenging to assess the importance of standing variation in the response to selection.

This challenge can be addressed by the development of models and methods for studying selection on standing variation. Even when the LD- and diversity-based signatures are weak[78,79], the spatial patterns expected at the selected site itself can still be quite strong for standing variation. For example, the scenarios in FIG. 2 do not depend on whether selection began on standing or new variation. Whole-genome resequencing will allow all sites to be interrogated for spatial signatures of selection without the current reliance on LD signatures and will therefore allow more examples of standing variation to be discovered.

Future directions

We have discussed various approaches that are providing novel insights into spatial variation, but further progress in the field will depend on new developments.

Advancing models of selection for humans. The classical models of selection in structured populations (BOX 1) lack several features that may have important

Genetic heterogeneity
The production of a similar phenotype by different mutations at either the same locus or different loci in different individuals.

Linkage disequilibrium
The non-random association of alleles carried at different loci. If a particular combination of alleles on a chromosome is found more or less frequently than expected (assuming independence among loci), then linkage disequilibrium is said to exist. It can arise for various reasons (novel mutations, genetic drift, natural selection and admixture) but recombination is the main process that removes it.

effects on the dynamics of selection in human populations. Future models should account for: recent spatial expansions with serial founder effects; changes in selective pressures over various timescales (for example, changes associated with climate change, the colonization of new habitats during the expansion out of East Africa, the development of agriculture, modern health technologies and ongoing pathogen emergence); spatial complexities, such as irregularly shaped habitats, fine-scaled variation in selection pressures, and geographically and temporally varying dispersal parameters; and the potential for multiple mutations and/or standing variation to underlie the adaptive response to a novel selective pressure.

Building more insightful models will be computationally challenging. Carefully describing the basic properties of even the classical models reviewed in BOX 1 is still an ongoing research area[80], largely because the simple classical models alone present serious mathematical challenges. The inclusion of the features listed above will require innovative theoretical approaches and possibly large-scale numerical approximation techniques similar to those used in other fields that deal inherently with complex spatial processes (for example, meteorology and oceanography).

Full resequencing data. A major limitation of SNP genotyping data sets is the ascertainment bias that is introduced in the selection of SNPs to be surveyed. This is particularly problematic in spatial studies of selection because the most interesting variants may be geographically restricted and hence may not be variable in the population panel used for SNP discovery. The availability of full resequencing data will overcome this limitation. Further, such data will address whether polymorphism levels at specific loci (such as the human leukocyte antigen (*HLA*) region[81]) vary as a function of specific environmental variables. Small-scale resequencing studies have already identified interesting spatial patterns of non-synonymous variants for candidate selection targets and susceptibility loci for clinical phenotypes. These loci include: the *MC1R* locus, which is involved in skin pigmentation[82]; the angiopoietin-like 4 (*ANGPTL4*) gene, which is known to influence plasma lipid levels[83]; and the *N*-acetyltransferase 2 (*NAT2*) gene, which codes for a drug-metabolizing enzyme[84,85]. Next-generation sequencing technologies will help to reveal the fine texture of human sequence variation across geographic regions and address questions about human history and selection with unprecedented detail.

Subtle allele frequency shifts due to selection. Analyses of spatial patterns of variation have focused on extremely differentiated variants and so have probably missed loci that show more subtle allele frequency differences or weaker environmental correlations. Future studies will be most fruitful if they are coupled with a detailed characterization of human environmental diversity, such as information about climate, diet composition, pathogen diversity and load, and modes of subsistence. Indeed, previous studies of specific environmental features that are directly relevant to human physiology and health have successfully identified regionally advantageous variants and have established paradigmatic examples (for example, the link between malarial resistance and sickle cell anaemia). Perhaps because of the need for fine-scale geographic sampling and environmental data, analyses of environmental correlations have been applied only to individual candidate genes rather than at the genome scale. These approaches are most powerful if they compare populations that live in different environments in the same continental cluster with populations in different clusters; allele frequency differences that are consistently observed between environments in multiple clusters can be attributed to environmental adaptations as opposed to divergence between populations. This study design was recently applied to copy number variation in the amylase gene, which seems to have increased and decreased in frequency in response to selective pressures for high or low starch diets in disparate regions of the world[86].

Relevance to biomedical research

To the extent that genetic variation contributes to health disparities, it will be interesting to determine how often risk variants for diseases with marked differentiation across ethnic groups result from neutral or adaptive processes and, if adaptive, whether the alleles were globally or locally advantageous. The convergent evolution of the light pigmentation phenotype in Asia and Europe[39,68,76] reminds us that if a variant that influences a phenotype occurs at different frequencies, it does not necessarily imply that the phenotype has different prevalences across populations. Furthermore, these patterns argue that biological mechanisms underlying a given adaptive phenotype may differ across populations. Common diseases may provide examples of this, as the subtypes of a given disease (for example, triple-negative breast cancer and salt-sensitive hypertension) often vary in prevalence across populations, therefore pointing to different pathways underlying the same disease in different populations. Additionally, studies of spatial patterns of human genetic variation are beginning to give insights into the extent to which the socio-political (for example, census) categories used in biomedical research coincide with the spatial distribution of selective pressures and of selected alleles. The alignment of the spatial distribution of many selected alleles with continental clusters[39,75] supports claims that continental clusters might be useful for describing and understanding the distribution of medically relevant phenotypes. However, the observed correlation between geographic or environmental variables and the frequency of some susceptibility alleles suggests that some genetic variation may follow specific aspects of the environment, rather than being organized according to the geographically defined clusters[51–53]. Finally, a deeper understanding of the evolutionary processes that generate genetic differences between populations will hopefully prevent misinterpretations of spatial patterns of genetic variation.

Convergent evolution
The evolution of similar traits by independent processes in individuals with no common ancestry. It usually indicates evolutionary adaptation to similar environmental conditions.

Triple-negative breast cancer
A subtype of breast cancer in which cells lack the oestrogen receptor, progesterone receptor and human epidermal growth factor receptor 2. Although breast cancer overall is more common among women of European ancestry, triple-negative cases occur more frequently in post-menopausal women of African ancestry. These cancers are also more aggressive and resistant to current treatment than those that express these receptors.

Salt-sensitive hypertension
Inter-individual variation in blood pressure changes in response to high or low sodium intake. Hypertensive subjects whose blood pressure increases more than a specified proportion upon salt loading are defined as salt sensitive.

1. Cavalli-Sforza, L. L., Menozzi, P. & Piazza, A. *History and Geography of Human Genes* (Princeton Univ. Press, 1994).
2. Roychoudhury, A. K. & Nei, M. *Human Polymorphic Genes: World Distribution* (Oxford Univ. Press, 1988).
3. Haldane, J. B. S. The rate of mutation of human genes. *Hereditas* **35** (Suppl. 1), 267–272 (1949).
4. Fisher, R. The wave of advance of advantageous genes. *Ann. Eugen.* **7**, 355–369 (1937).
A classic paper that establishes a reaction–diffusion model for the spread of an advantageous allele and uses it to calculate the speed of the expanding wave of advance.
5. Roberts, D. F. Human pigmentation: its geographical and racial distribution and biological significance. *J. Soc. Cosmet. Chem.* **28**, 329–342 (1977).
6. Simoons, F. J. Primary adult lactose intolerance and the milking habit: a problem in biologic and cultural interrelations. II. A culture historical hypothesis. *Am. J. Dig. Dis.* **15**, 695–710 (1970).
7. Simoons, F. J. Primary adult lactose intolerance and the milking habit: a problem in biological and cultural interrelations. I. Review of the medical research. *Am. J. Dig. Dis.* **14**, 819–836 (1969).
8. Cavalli-Sforza, L. L. Analytic review: some current problems of human population genetics. *Am. J. Hum. Genet.* **25**, 82–104 (1973).
9. Katzmarzyk, P. T. & Leonard, W. R. Climatic influences on human body size and proportions: ecological adaptations and secular trends. *Am. J. Phys. Anthropol.* **106**, 483–503 (1998).
10. Roberts, D. F. *Climate and Human Variability* (Cummings, Menlo Park, 1978).
11. Friedlaender, J. S. *et al.* The genetic structure of Pacific Islanders. *PLoS Genet.* **4**, e19 (2008).
12. Wang, S. *et al.* Genetic variation and population structure in native Americans. *PLoS Genet.* **3**, e185 (2007).
13. Nelson, M. R. *et al.* The Population Reference Sample, POPRES: a resource for population, disease, and pharmacological genetics research. *Am. J. Hum. Genet.* **83**, 347–358 (2008).
14. Novembre, J. *et al.* Genes mirror geography within Europe. *Nature* **456**, 98–101 (2008).
15. Lao, O. *et al.* Correlation between genetic and geographic structure in Europe. *Curr. Biol.* **18**, 1241–1248 (2008).
16. Tishkoff, S. A. *et al.* The genetic structure and history of Africans and African Americans. *Science* **324**, 1035–1044 (2009).
17. Li, J. Z. *et al.* Worldwide human relationships inferred from genome-wide patterns of variation. *Science* **319**, 1100–1104 (2008).
18. Rosenberg, N. A. *et al.* Genetic structure of human populations. *Science* **298**, 2381–2385 (2002).
19. Jakobsson, M. *et al.* Genotype, haplotype and copy-number variation in worldwide human populations. *Nature* **451**, 998–1003 (2008).
20. Rosenberg, N. A. *et al.* Clines, clusters, and the effect of study design on the inference of human population structure. *PLoS Genet.* **1**, e70 (2005).
21. Ramachandran, S. *et al.* Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. *Proc. Natl Acad. Sci. USA* **102**, 15942–15947 (2005).
22. Serre, D. & Paabo, S. Evidence for gradients of human genetic diversity within and among continents. *Genome Res.* **14**, 1679–1685 (2004).
23. Handley, L. J., Manica, A., Goudet, J. & Balloux, F. Going the distance: human population genetics in a clinal world. *Trends Genet.* **23**, 432–439 (2007).
A thoughtful review about the effects of population history on spatial patterns of neutral variation in humans.
24. Prugnolle, F., Manica, A. & Balloux, F. Geography predicts neutral genetic diversity of human populations. *Curr. Biol.* **15**, R159–R160 (2005).
25. Edmonds, C. A., Lillie, A. S. & Cavalli-Sforza, L. L. Mutations arising in the wave front of an expanding population. *Proc. Natl Acad. Sci. USA* **101**, 975–979 (2004).
26. Vlad, M. O., Cavalli-Sforza, L. L. & Ross, J. Enhanced (hydrodynamic) transport induced by population growth in reaction–diffusion systems with application to population genetics. *Proc. Natl Acad. Sci. USA* **101**, 10249–10253 (2004).
27. Klopfstein, S., Currat, M. & Excoffier, L. The fate of mutations surfing on the wave of a range expansion. *Mol. Biol. Evol.* **23**, 482–490 (2006).
28. Excoffier, L. & Ray, N. Surfing during population expansions promotes genetic revolutions and structuration. *Trends Ecol. Evol.* **23**, 347–351 (2008).
29. Currat, M. *et al.* Comment on 'Ongoing adaptive evolution of ASPM, a brain size determinant in *Homo sapiens*' and 'Microcephalin, a gene regulating brain size, continues to evolve adaptively in humans'. *Science* **313**, 172 (2006); author reply in **313**,172 (2006).
30. Hofer, T., Ray, N., Wegmann, D. & Excoffier, L. Large allele frequency differences between human continental groups are more likely to have occurred by drift during range expansions than by selection. *Ann. Hum. Genet.* **73**, 95–108 (2009).
31. Wright, S. The genetical structure of populations. *Ann. Eugen.* **15**, 323–354 (1951).
32. Holsinger, K. E. & Weir, B. S. Genetics in geographically structured populations: defining, estimating and interpreting F_{ST}. *Nature Rev. Genet.* **10**, 639–650 (2009).
33. Lewontin, R. C. & Krakauer, J. Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics* **74**, 175–195 (1973).
34. Haldane, J. B. The theory of a cline. *J. Genet.* **48**, 277–284 (1948).
35. Cavalli-Sforza, L. L. Population structure and human evolution. *Proc. R. Soc. Lond. B* **164**, 362–379 (1966).
36. Bowcock, A. M. *et al.* Drift, admixture, and selection in human evolution: a study with DNA polymorphisms. *Proc. Natl Acad. Sci. USA* **88**, 839–843 (1991).
37. Akey, J. M., Zhang, G., Zhang, K., Jin, L. & Shriver, M. D. Interrogating a high-density SNP map for signatures of natural selection. *Genome Res.* **12**, 1805–1814 (2002).
38. 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).
An F_{ST}-based analysis of patterns of differentiation in the HapMap phase II data that reveals evidence for adaptive genetic divergence among human populations.
39. Coop, G. *et al.* The role of geography in human adaptation. *PLoS Genet.* **5**, e1000500 (2009).
A synthetic overview of patterns of variation in the HGDP and HapMap data that argues that human demography has had a strong effect on the geographic distribution of selected alleles.
40. Beaumont, M. A. & Balding, D. J. Identifying adaptive genetic divergence among populations from genome scans. *Mol. Ecol.* **13**, 969–980 (2004).
41. Beaumont, M. A. & Nichols, R. A. Evaluating loci for use in the genetic analysis of population structure. *Proc. R. Soc. Lond. B* **263**, 1619–1626 (1996).
42. Excoffier, L., Hofer, T. & Foll, M. Detecting loci under selection in a hierarchically structured population. *Heredity* **103**, 285–298 (2009).
43. Lamason, R. L. *et al.* SLC24A5, a putative cation exchanger, affects pigmentation in zebrafish and humans. *Science* **310**, 1782–1786 (2005).
44. Williamson, S. H. *et al.* Localizing recent adaptive evolution in the human genome. *PLoS Genet.* **3**, e90 (2007).
45. Charlesworth, B., Nordborg, M. & Charlesworth, D. The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. *Genet. Res.* **70**, 155–174 (1997).
46. Hu, X. S. & He, F. Background selection and population differentiation. *J. Theor. Biol.* **235**, 207–219 (2005).
47. Santiago, E. & Caballero, A. Variation after a selective sweep in a subdivided population. *Genetics* **169**, 475–483 (2005).
48. Berry, A. & Kreitman, M. Molecular analysis of an allozyme cline: alcohol dehydrogenase in *Drosophila melanogaster* on the east coast of North America. *Genetics* **134**, 869–893 (1993).
49. Umina, P. A., Weeks, A. R., Kearney, M. R., McKechnie, S. W. & Hoffmann, A. A. A rapid shift in a classic clinal pattern in *Drosophila* reflecting climate change. *Science* **308**, 691–693 (2005).
50. Joost, S. *et al.* A spatial analysis method (SAM) to detect candidate loci for selection: towards a landscape genomics approach to adaptation. *Mol. Ecol.* **16**, 3955–3969 (2007).
51. Thompson, E. E. *et al. CYP3A* variation and the evolution of salt-sensitivity variants. *Am. J. Hum. Genet.* **75**, 1059–1069 (2004).
52. Young, J. H. *et al.* Differential susceptibility to hypertension is due to selection during the out-of-Africa expansion. *PLoS Genet.* **1**, e82 (2005).
An analysis of candidate risk variants for hypertension, which are shown to have unusually strong correlations with latitude and climate variables relative to random genomic loci.
53. Hancock, A. M. *et al.* Adaptations to climate in candidate genes for common metabolic disorders. *PLoS Genet.* **4**, e32 (2008).
An analysis of variation in candidate genes for metabolic syndrome. The study uses a novel method that accounts for population structure when testing correlations with climate variables.
54. Schlotterer, C. A microsatellite-based multilocus screen for the identification of local selective sweeps. *Genetics* **160**, 753–763 (2002).
55. Kauer, M. O., Dieringer, D. & Schlotterer, C. A microsatellite variability screen for positive selection associated with the 'out of Africa' habitat expansion of *Drosophila melanogaster*. *Genetics* **165**, 1137–1148 (2003).
56. Storz, J. F., Payseur, B. A. & Nachman, M. W. Genome scans of DNA variability in humans reveal evidence for selective sweeps outside of Africa. *Mol. Biol. Evol.* **21**, 1800–1811 (2004).
57. Marshall, J. M. & Weiss, R. E. A Bayesian heterogeneous analysis of variance approach to inferring recent selective sweeps. *Genetics* **173**, 2357–2370 (2006).
58. Sabeti, P. C. *et al.* Genome-wide detection and characterization of positive selection in human populations. *Nature* **449**, 913–918 (2007).
59. 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).
60. 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).
61. Bersaglieri, T. *et al.* Genetic signatures of strong recent positive selection at the lactase gene. *Am. J. Hum. Genet.* **74**, 1111–1120 (2004).
62. Novembre, J., Galvani, A. P. & Slatkin, M. The geographic spread of the CCR5 Δ32 HIV-resistance allele. *PLoS Biol.* **3**, e339 (2005).
An application of the wave of advance model to the geographic distribution of a variant that confers resistance to HIV infection.
63. Sabeti, P. C. *et al.* The case for selection at *CCR5-Δ32*. *PLoS Biol.* **3**, e378 (2005).
64. Endler, J. A. *Geographic Variation, Speciation and Clines* (Princeton Univ. Press, 1977).
65. Enattah, N. S. *et al.* Identification of a variant associated with adult-type hypolactasia. *Nature Genet.* **30**, 233–237 (2002).
66. Tishkoff, S. A. *et al.* Convergent adaptation of human lactase persistence in Africa and Europe. *Nature Genet.* **39**, 31–40 (2007).
67. Enattah, N. S. *et al.* Independent introduction of two lactase-persistence alleles into human populations reflects different history of adaptation to milk culture. *Am. J. Hum. Genet.* **82**, 57–72 (2008).
68. Norton, H. L. *et al.* Genetic evidence for the convergent evolution of light skin in Europeans and East Asians. *Mol. Biol. Evol.* **24**, 710–722 (2007).
69. Cappellini, M. D. & Fiorelli, G. Glucose-6-phosphate dehydrogenase deficiency. *Lancet* **371**, 64–74 (2008).
70. Flint, J., Harding, R. M., Boyce, A. J. & Clegg, J. B. The population genetics of the haemoglobinopathies. *Baillieres Clin. Haematol.* **11**, 1–51 (1998).
71. Hill, A. V. Molecular epidemiology of the thalassaemias (including haemoglobin E). *Baillieres Clin. Haematol.* **5**, 209–238 (1992).
72. Goldstein, D. B. & Holsinger, K. E. Maintenance of polygenic variation in spatially structured populations — roles for local mating and genetic redundancy. *Evolution* **46**, 412–429 (1992).
73. Kelly, J. K. Geographical variation in selection, from phenotypes to molecules. *Am. Nat.* **167**, 481–495 (2006).
An insightful simulation study of how spatially varying selection affects neutral sequence variation that is linked to the quantitative trait loci that underlie the selected trait.
74. Latta, R. G. Differentiation of allelic frequencies at quantitative trait loci affecting locally adaptive traits. *Am. Nat.* **151**, 283–292 (1998).

75. Pickrell, J. K. *et al.* Signals of recent positive selection in a worldwide sample of human populations. *Genome Res.* **19**, 826–837 (2009).
76. Myles, S., Somel, M., Tang, K., Kelso, J. & Stoneking, M. Identifying genes underlying skin pigmentation differences among human populations. *Hum. Genet.* **120**, 613–621 (2007).
77. Norton, H. L. *et al.* Genetic evidence for the convergent evolution of light skin in Europeans and East Asians. *Mol. Biol. Evol.* **24**, 710–722 (2007).
78. Przeworski, M., Coop, G. & Wall, J. D. The signature of positive selection on standing genetic variation. *Evolution* **59**, 2312–2323 (2005).
79. Hermisson, J. & Pennings, P. S. Soft sweeps: molecular population genetics of adaptation from standing genetic variation. *Genetics* **169**, 2335–2352 (2005).
80. Nagylaki, T. & Lou, Y. Evolution under multiallelic migration-selection models. *Theor. Popul. Biol.* **72**, 21–40 (2007).
81. Prugnolle, F. *et al.* Pathogen-driven selection and worldwide HLA class I diversity. *Curr. Biol.* **15**, 1022–1027 (2005).
82. Harding, R. M. *et al.* Evidence for variable selective pressures at *MC1R*. *Am. J. Hum. Genet.* **66**, 1351–1361 (2000).
83. Romeo, S. *et al.* Population-based resequencing of *ANGPTL4* uncovers variations that reduce triglycerides and increase HDL. *Nature Genet.* **39**, 513–516 (2007).
84. Luca, F. *et al.* Multiple advantageous amino acid variants in the *NAT2* gene in human populations. *PLoS ONE* **3**, e3136 (2008).
85. Patin, E. *et al.* Deciphering the ancient and complex evolutionary history of human arylamine N-acetyltransferase genes. *Am. J. Hum. Genet.* **78**, 423–436 (2006).
86. Perry, G. H. *et al.* Diet and the evolution of human amylase gene copy number variation. *Nature Genet.* **39**, 1256–1260 (2007).
The application of a population genetics approach to copy number variation in the amylase gene to identify adaptive variation in response to dietary changes.
87. Slatkin, M. & Wiehe, T. Genetic hitch-hiking in a subdivided population. *Genet. Res.* **71**, 155–160 (1998).
88. Levene, H. Genetic equilibrium when more than one ecological niche is available. *Am. Nat.* **87**, 331–333 (1953).
89. Hoekstra, R. F., Bijlsma, R. & Dolman, A. J. Polymorphism from environmental heterogeneity — models are only robust if the heterozygote is close in fitness to the favored homozygote in each environment. *Genet. Res.* **45**, 299–314 (1985).
90. Smith, J. M. & Hoekstra, R. Polymorphism in a varied environment: how robust are the models? *Genet. Res.* **35**, 45–57 (1980).
91. Barton, N. H. & Clark, A. G. in *Population Biology: Ecological and Evolutionary Viewpoints* (eds Wöhrmann, K. & Jain, S. K.) 115–173 (Springer, Berlin, 1990).
92. Fisher, R. A. Gene frequencies in a cline determined by selection and diffusion. *Biometrics* **6**, 353–361 (1950).
93. Slatkin, M. Gene flow and selection in a cline. *Genetics* **75**, 733–756 (1973).
94. Slatkin, M. Gene flow and selection in a two-locus system. *Genetics* **81**, 787–802 (1975).
95. May, R. M., Endler, J. A. & Mcmurtrie, R. E. Gene frequency clines in presence of selection opposed by gene flow. *Am. Nat.* **109**, 659–676 (1975).
96. Nagylaki, T. Conditions for existence of clines. *Genetics* **80**, 595–615 (1975).
97. Nagylaki, T. Clines with variable migration. *Genetics* **83**, 867–886 (1976).
98. Nagylaki, T. Clines with asymmetric migration. *Genetics* **88**, 813–827 (1978).
99. Endler, J. A. Gene flow and population differentiation. *Science* **179**, 243–250 (1973).
100. Slatkin, M. & Maruyama, T. Genetic drift in a cline. *Genetics* **81**, 209–222 (1975).
101. Gleibermann, L. Blood pressure and dietary salt in human populations. *Ecol. Food Nutr.* **2**, 143–156 (1973).
102. Beckman, G. *et al.* Is p53 polymorphism maintained by natural selection? *Hum. Hered.* **44**, 266–270 (1994).
103. Shi, H. *et al.* Winter temperature and UV are tightly linked to genetic changes in the p53 tumor suppressor pathway in Eastern Asia. *Am. J. Hum. Genet.* **84**, 534–541 (2009).
104. Luca, F. *et al.* Adaptive variation regulates the expression of the human *SGK1* gene in response to stress. *PLoS Genet.* **5**, e1000489 (2009).
105. Teshima, K. M., Coop, G. & Przeworski, M. How reliable are empirical genomic scans for selective sweeps? *Genome Res.* **16**, 702–712 (2006).

Acknowledgements
We would like to thank M. Przeworski, G. Coop and members of our laboratories for discussions and critical comments on the manuscript. A.D. acknowledges research support from the US National Institutes of Health (GM79558 and DK56670) and J.N. acknowledges support from the Searle Scholars Program and the US National Science Foundation (0733033).

FURTHER INFORMATION
John Novembre's homepage: http://www.eeb.ucla.edu/Faculty/Novembre
Anna Di Rienzo's homepage: http://www.genes.uchicago.edu/dirienzo.html
1000 Genomes Project: http://www.1000genomes.org
Allele Frequency Database (ALFRED): http://alfred.med.yale.edu
Database of Genotypes and Phenotypes (dbGaP): http://www.ncbi.nlm.nih.gov/gap
Database of SNPs (dbSNP): http://www.ncbi.nlm.nih.gov/SNP
Human Genome Diversity Project: http://www.cephb.fr/en/hgdp/diversity.php
Human Genome Diversity Project Selection Browser: http://hgdp.uchicago.edu/cgi-bin/gbrowse/HGDP
International HapMap Project: http://www.hapmap.org
National Center for Biotechnology Information: http://www.ncbi.nlm.nih.gov
Protein Analysis Through Evolutionary Relationships (PANTHER) database: http://www.pantherdb.org

ALL LINKS ARE ACTIVE IN THE ONLINE PDF