Challenges in human genetic diversity: demographic history and adaptation

Patricia L. Balaresque, Stéphane J. Ballereau and Mark A. Jobling*

Department of Genetics, University of Leicester, University Road, Leicester LE1 7RH, UK

Received August 14, 2007; Revised August 14, 2007; Accepted August 22, 2007

Modern human genetic diversity is the result of demographic history, and selective effects that have acted to adapt different populations to their environments. Broad patterns of global diversity are well explained by geography, based on an out-of-Africa model of early human evolution. Genome-wide searches for signals of selection, plus studies of specific candidate loci and candidate phenotypes, have identified genes that show population differences due to adaptation to pathogens, climate, diet and possibly cognitive challenges. Some past adaptations are now maladaptive, and can lead to disease. However, the history of adaptation is complex, and adaptive explanations are often unsupported by hard evidence.

INTRODUCTION

The two-fold challenge of human population genetics is to explain current patterns of genetic diversity in terms of population history, and at the same time to understand the genetic basis of the phenotypic adaptations that allowed humans to colonize the globe so successfully. The first of these goals involves the study of loci that we hope are evolving neutrally, while the second focuses on the influence of natural selection. Traditionally, there has been a tension here, since it is difficult to be sure that any locus is neutral, and it can also be difficult to prove that patterns of diversity at a non-neutral locus are due to selection rather than demography.

In this review we discuss recent developments in the study of human genetic diversity in general, driven by the availability of large data sets on global samples, and then move on to address evidence for pathogen, climatic, dietary and cognitive genetic adaptation, many of which have consequences for disease susceptibility today.

THE BIG PICTURE: GLOBAL PATTERNS OF DIVERSITY

The complex patterns of genetic diversity in modern populations are the product of many layers of demographic and evolutionary events acting on different timescales, including colonizations, migrations, population expansions, mutation, genetic drift and selection (reviewed in 1). Excavating these genetic strata is not easy, but it is reasonable to suppose that early events when populations were small had disproportionately large effects on modern diversity. Fossil and archaeological evidence have been interpreted to support opposing models for the origins of modern humans (2), but most genetic data are best explained by a recent (<200 KYA (thousand years ago)) African origin, followed by migration of a small population out of Africa (Fig. 1A), and a replacement of archaic species without interbreeding (though see 3). Genetic diversity is generally greatest in Africa, where the roots of many locus-specific phylogenies lie (1).

Over the last two decades, genetic studies have focused on specific loci including the Y chromosome (4), mitochondrial DNA (5), and loci on the X chromosome and autosomes (6). However, each locus captures only one realization of the evolutionary process, and is potentially subject to unique selective forces; furthermore, the various studies analyse different sample sets, making a synthesis difficult.

More recent studies have exploited the availability of the CEPH Human Genome Diversity Panel (7,8), comprising 1064 DNAs from 54 widely distributed populations, typed with large numbers of markers including a set of over 1000 microsatellites and several thousand single nucleotide polymorphisms (SNPs). While selection may act on individual loci, the sheer number of markers should provide an unbiased picture of past events. Studies using these samples have shown the power of a geographically explicit approach (9–12) to explain global patterns. Geographical distance from East Africa, measured via plausible land migration routes, correlates remarkably well with declining gene diversity, explaining ~85% of its variance (12). A model

*To whom correspondence should be addressed. Tel: +44 01162523427; Fax: +44 01162523378; Email: maj4@le.ac.uk

© 2007 The Author(s)

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/2.0/uk/) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
Figure 1. Schematic representation of some important events in human evolution. (A) Modern human origins and early migrations. Occupation of new environments had adaptive consequences for skin pigmentation, and possibly for cognition. (B) Agricultural revolutions leading to population expansions and new disease challenges. Symbols indicate a selection of locally domesticated plants and animals. (C) Selected migrations and subsequent admixture in the last three thousand years, including consequences for infectious and other diseases in the New World (arrows indicate migrations; KYA: thousand years ago; C: century; figure based on material included in 1).
of human settlement history incorporating distance is compatible with the expansion of a founding population of ~1000 individuals, starting around 56 KYA (11). While the power of this model is greatest when applied to a large data set of apparently neutral markers, it also explains significant proportions of variance in selected genes (HLA Class I; 13) and a selected trait, craniofacial variation (14).

GENOME-WIDE PATTERNS OF SELECTION

The aim of the HapMap project (15,16) was not to help population geneticists, but to provide a genome-wide haplotype map based on common SNPs, to aid searches for disease susceptibility genes. HapMap phase I included only four populations (from Africa, Europe and East Asia), typed with ~1 000 000 SNPs, subject to considerable ascertainment bias. Nonetheless, interesting findings emerged, and in particular the identification of signals of natural selection (16), indicated by SNPs showing skewed frequencies between populations, segments of the genome in which diversity was particularly low (indicating ‘selective sweeps’), and extended haplotypes at a locus. Some of the identified loci were already known, including the lactase gene (LCT), which represents a target for past dietary adaptation, discussed below. However, several strongly selected loci are apparently devoid of genes, which emphasizes the difficulty of understanding the functions of genomic elements (17). The classes of genes lying in selected regions are broadly concordant between different studies (16,18,19), with significant over-representation of genes involved in chemosensory perception and olfaction, immunity, reproduction and fertility, and carbohydrate metabolism.

Differences between studies reflect differences in methodology, SNPs and, more interestingly: populations; as well as HapMap, one study includes the Perlegen data set (19), based on fewer markers and three populations (European-, Asian-, and African-Americans), all of which, to some degree, are admixed following recent migration. Migration involved new selective challenges, to which different populations responded in different ways. Loci from the various parental populations may be differentially selected (20), potentially providing useful clues to the genetics of, for example, disease resistance.

SIGNS OF ADAPTATION

Genetic adaptation to regional variation in pathogens, climate and diet has undoubtedly been important to our success. The genetic bases of adaptation are complex and characterized by successive and contradictory selective forces acting on the same genes through changing conditions. This complexity can make it difficult to prove adaptive explanations; the general approach has been to identify a geographically differentiated heritable trait in indigenous populations, then to propose a similarly differentiated environmental factor to explain it, preferably linked with a plausible biological mechanism. A key difficulty is that the mechanism may be plausible, but it often cannot be proved, which leads to debate and controversy.

Adaptation to pathogens

Infectious diseases have been a continuous cause of mortality, and, as humans have colonized new areas and undergone demographic changes, the spectrum of disease has varied through time; the development of agriculture (Fig. 1B) was particularly important in facilitating the transfer of pathogens from animals to humans (zoonosis, 21). Genetic variation has a major impact on both past and present susceptibility. This subject has been extensively studied (22), and cannot be summarized in this brief overview. The field is notable for clear examples of associations between gene variants and disease susceptibility or resistance, for which excellent epidemiological evidence, often backed up by laboratory studies, exists. Many of these relate to malaria (23), a disease that, in affected populations, has driven resistance variants to fixation (the Duffy O allele at the DARC gene), or to high frequency through balancing selection (e.g. Hbs, G6PD). The field provides good examples of the shifting nature of selection and adaptation. Homozygotes for a 32-bp deletion in the CCR5 chemokine receptor gene are resistant to HIV/AIDS, and the allele shows clear evidence of selection; however, HIV is a novel infectious agent, and cannot explain the highly geographically differentiated patterns of CCR5 Δ32. A more likely candidate is past smallpox infections (24,25).

Climatic adaptation

Early migration out of Africa exposed ancestral populations to colder environments, with less incident sunlight. The most obvious response is in pigmentation, due to the quantity, type, and distribution of melanin. Skin colour is strongly geographically differentiated, with darker-skinned populations concentrated in the tropics, and lighter-skinned populations in more northerly latitudes. In tropical climates the melanin in dark skins protects against sunburn by scattering and absorbing UV radiation; it may also limit photodegradation of nutrients such as folate. Where sunlight is low, depigmentation may be favoured because UV penetration is necessary for vitamin D synthesis (26).

A balance between these factors can largely explain the global pattern of pigmentation (27). However, there are discrepancies: while these could be due to recent migration or admixture, an alternative view (downplaying the importance of light skin in vitamin D synthesis) suggests that they may reflect sexual selection, in particular male preference for light-skinned females (28). The last idea predicts that sexual dimorphism should increase with increasing latitude, but in practice this is not borne out (29).

The genetic basis of pigmentation variation is complex (30). Candidate gene approaches based on information about melanosome development, abnormal pigmentation phenotypes and orthologs in other species, together with analyses of selection signals in HapMap, have identified a set of about twenty genes that are likely to play important roles (18,30–32). The finding of selection signatures in Europeans and East Asians suggests that there really is some selective advantage for light skin, and the fact that they are in different genes shows that evolutionary routes to similar phenotypes were distinct – convergent evolution (30,32,33).
Residence of ancestral populations in tropical Africa also necessitated heat adaptation, including cooling through efficient sweating. The considerable salt loss, combined with low dietary salt availability, led to selection for salt retention; at the same time, there was likely selection for increased arterial tone and cardiac contraction force when blood volume was depleted by water loss. After migration into temperate climates, these adaptations became maladaptive, and may be responsible for increased blood pressure.

In the CEPH-HGDP samples, the global distribution of the functional (heat-adapted) alleles of seven genes involved in blood pressure regulation shows a latitudinal cline (34) unmatched by neutral markers. Furthermore, a combination of latitude and the frequency of one of these alleles in the \( G \) protein \( \beta_3 \) subunit gene explains a remarkable 64% of global variation in blood pressure.

Initially, heat-adapted African populations expanded northward, undergoing selection for cold adaptation. Subsequently, cold-adapted north Asians expanded southward into the Americas less than 20 KYA, undergoing selection for heat adaptation, so that Native Americans show similar salt retention and cardiovascular phenotypes to Africans living at the same latitudes. In more recent migrations, particularly that of Africans to the temperate climate and high-salt environment of North America (Fig. 1C), previously adapted genes may be particularly maladapted, leading to high frequencies of hypertension.

**Dietary adaptation**

The diversity of environments occupied by hunter–gatherer humans after the early migrations are mirrored by a diversity of diets. Early heterogeneity of resources may still have an impact today: for example, as judged by modern Y chromosome diversity (36), population expansions occurred earlier in the northern part of East Asia, probably because of the abundant megafauna of the ‘Mammoth Steppe’, but later in the south, due to poorer resources.

In terms of influence on diet, the most important development in human prehistory was agriculture, beginning \( \sim 10 \) KYA in the Near East, with distinct varieties emerging later in China, the Americas and West Africa (Fig. 1B). Increased population densities, reduced dietary diversity, sedentary lifestyle and exposure to animal pathogens together represented a major set of challenges. A common view is that post-Neolithic humans are adapted, through a ‘thrifty genotype’ (37,38) to a hunter–gatherer lifestyle of feast and famine (39), and that the arrival of agriculture signalled the start of an era of dietary maladaptation, leading to high incidences of type 2 diabetes. Later colonization events, like those of the Pacific islands (Fig. 1C), may have involved particularly strong selection for thrifty genotypes (40) — possibly causing subsequent extreme levels of diabetes. These views have not gone unopposed, however (41,42), and recent studies of diabetes susceptibility loci suggest that reality is not so simple.

A variant in the transcription factor 7-like 2 gene (\( TCF7L2 \)) is responsible for 17–28% of the risk of type 2 diabetes in Europeans (43), but, contrary to expectations of the thrifty genotype hypothesis, is associated with reduced body mass index (BMI) in diabetics (44). In the HapMap samples, the frequency of another variant of the same gene, associated with increased BMI, has been driven by selection to near fixation (95%) in East Asians, with lower frequencies elsewhere. The ages of the variant in the different populations correspond approximately to the times of origin of agriculture, suggesting that it conferred some advantage in the post-agricultural environment. However, the nature of this advantage is unclear.

A clear example of genetic adaptation to cultural innovation is the selection of alleles of \( LCT \) permitting persistence of lactase expression into adulthood. This allows the drinking of milk without adverse effects, and the distribution of the phenotype correlates well with that of populations with a history of cattle domestication and milk drinking (45). In the HapMap samples (16), \( LCT \) in Europeans shows the strongest signal of positive selection, reflecting a powerful advantage that may have been more related to milk as a source of uninfected water than as a source of nutrition.

Studies in European populations identified a causative regulatory variant \( \sim 14 \) kb upstream of the \( LCT \) gene (46), with an estimated age of 2000–20 000 years (47). However, lactase-persistent populations elsewhere, including Africa, do not carry this variant. Studies of Tanzanians, Kenyans and Sudanese (48,49) reveal three further nearby variants causing lactase persistence. Examination of surrounding haplotypes show that the three African variants arose independently of each other and of the European variant (a further example of convergent evolution), within the last \( \sim 7000 \) years. The known variants still do not account for all of lactase persistence; so further examples are likely to exist.

Further dietary adaptations remain to be discovered, and signals of selection around genes involved in the metabolism of other carbohydrates, fat and alcohol (18) are interesting.

**Cognitive adaptation?**

While many factors have been crucial to the success of \( Homo sapiens \), the defining innovation has been culture — the capacity to communicate and transfer knowledge, and to deal with novel environments by creating new technologies, including the development and exploitation of new food resources. Some have made the argument that the out of Africa migration entailed novel challenges that favoured the selection of enhanced cognitive ability, and have supported this using comparisons of IQ and brain size (50). However, it is unclear why the cognitive challenges in this gradual migration should be greater than those facing non-migrants subsisting in the diverse and changing environments of Africa.

Genetic studies in this area differ from those in, for example, pigmentation, because the interest in a particular gene is stimulated not by any known phenotypic effect, but simply by its expression in the brain, and some unexpected pattern of population differentiation. The identification of an underlying selected phenotype may not be straightforward, or without controversy.

Undoubtedly, brain size (and presumably associated cognitive capabilities) increased rapidly in the human lineage over the past 3–4 million years. A long list of functional candidate brain genes has been produced (51), though few have been studied.
Most interest has focused on two genes that, when mutated, result in microcephaly – a small brain, but with normal neural architecture. Both abnormal spindle-like microcephaly associated (ASPM) and Microcephalin (MCPH1) show signatures of adaptive selection both during the emergence of the human lineage (52,53), and subsequently (54,55). Globally, both genes show young, high frequency haplotypes that are rare in Africa, and could reflect recent regional selection. There has been argument over whether these patterns could be explained by demographic processes rather than selection (56,57).

Selection on these genes is expected to be through some aspect of intelligence, rather than brain size (58). However, the common derived alleles for both genes are unlinked to standard measures of IQ (59), suggesting either that the gene variants were not being selected at all, or that the selected phenotype is something other than intelligence, as measured by the simple single metric of IQ. Information about the transcript and protein expression patterns of the different allelic variants would be helpful.

FUTURE PROMISE, FUTURE CHALLENGES

The global patterns of neutral genetic diversity have been sketched out, and the power of a geographical framework demonstrated. There are still gaps in the landscape, however, and still opportunities for regional studies in which barriers to gene flow may be subtle, including cultural aspects such as language, religion, mating practices, and social organization. Studies illuminating the tempo of genetic change in response to social change (60) will help us appreciate the validity (or otherwise) of making inferences about past populations from the diversities of modern ones.

We humans live in variable environments, and demonstrate non-uniform distributions of interesting phenotypes. Our genomes show tantalizing signals of selection around unsuspected loci (16,18,19), and differences between populations in genome-wide analyses of gene expression (61,62). A host of targets for investigation are suggested by these observations, and will include high altitude adaptation (63), variation in taste phenotypes (64) and its relation to dietary history, and the morphological adaptations well known to physical anthropologists (65): unravelling the web of evolutionary history, function and adaptation will be fascinating, but difficult.

Most of the studies described above focus on a specific phenotype, and analyse it in some set of populations. However, some phenotypes are difficult to measure, and combined with the difficulty of obtaining good global population samples, this can lead to a misunderstanding of past selective effects. On the other hand, global sample sets, like the CEPH-HGDP panel, are universally available and typed with many markers, but completely lack phenotypic information. Inventive researchers sometimes assume a phenotype (such as basic pigmentation 30), purely based on geography.

Ideally, we would possess an HGDP-like collection in which comprehensive phenotypic and life history information was available – a kind of global Biobank. However, the daunting practical and ethical problems faced by the HGDP itself (66,67), as well as by the UK Biobank project (68), suggest that such an undertaking would not be easy. Any volunteers?

ACKNOWLEDGEMENTS

Funding to pay the Open Access publication charges for this article was provided by the Wellcome Trust.

Conflict of Interest statement. None declared.

FUNDING

The authors are supported by the Wellcome Trust. M.A.J. is a Wellcome Trust Senior Fellow in Basic Biomedical Science (grant no. 057559).

REFERENCES
