Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Immunology

Human genome variability, natural selection and infectious diseases

Matteo Fumagalli[1] and Manuela Sironi[2]

The recent availability of large-scale sequencing DNA data allowed researchers to investigate how genomic variation is distributed among populations. While demographic factors explain genome-wide population genetic diversity levels, scans for signatures of natural selection pinpointed several regions under non-neutral evolution. Recent studies found an enrichment of immune-related genes subjected to natural selection, suggesting that pathogens and infectious diseases have imposed a strong selective pressure throughout human history. Pathogen-mediated selection often targeted regulatory sites of genes belonging to the same biological pathway. Results from these studies have the potential to identify mutations that modulate infection susceptibility by integrating a population genomic approach with molecular immunology data and large-scale functional annotations.

Addresses
[1] UCL Genetics Institute, Department of Genetics, Evolution and Environment, University College London, Gower Street, London WC1E 6BT, United Kingdom
[2] Bioinformatics – Scientific Institute IRCCS E.MEDEA, 23842 Bosisio Parini, Italy

Corresponding author: Fumagalli, Matteo (m.fumagalli@ucl.ac.uk)

Current Opinion in Immunology 2014, **30**:9–16

This review comes from a themed issue on **Immunogenetics and transplantation**

Edited by **Luis B Barreiro** and **Lluis Quintana-Murci**

For a complete overview see the Issue and the Editorial

Available online 29th May 2014

http://dx.doi.org/10.1016/j.coi.2014.05.001

Introduction

Over the last couple of years, our understanding of human genetic diversity has dramatically increased. The introduction of high-throughput sequencing machines, along with the development of sophisticated analytical tools, has allowed researchers to obtain and analyze unprecedentedly large amount of DNA data [1,2].

Here we discuss how these datasets can be exploited to elucidate the role evolution and natural selection played in shaping the distribution of genetic diversity across human populations. We review currently employed methods to detect signatures of natural selection in the human genome, and discuss notable examples of pathogen-driven selective pressure and their implications in terms of human immunology and infectious disease epidemiology.

Population genomic variability

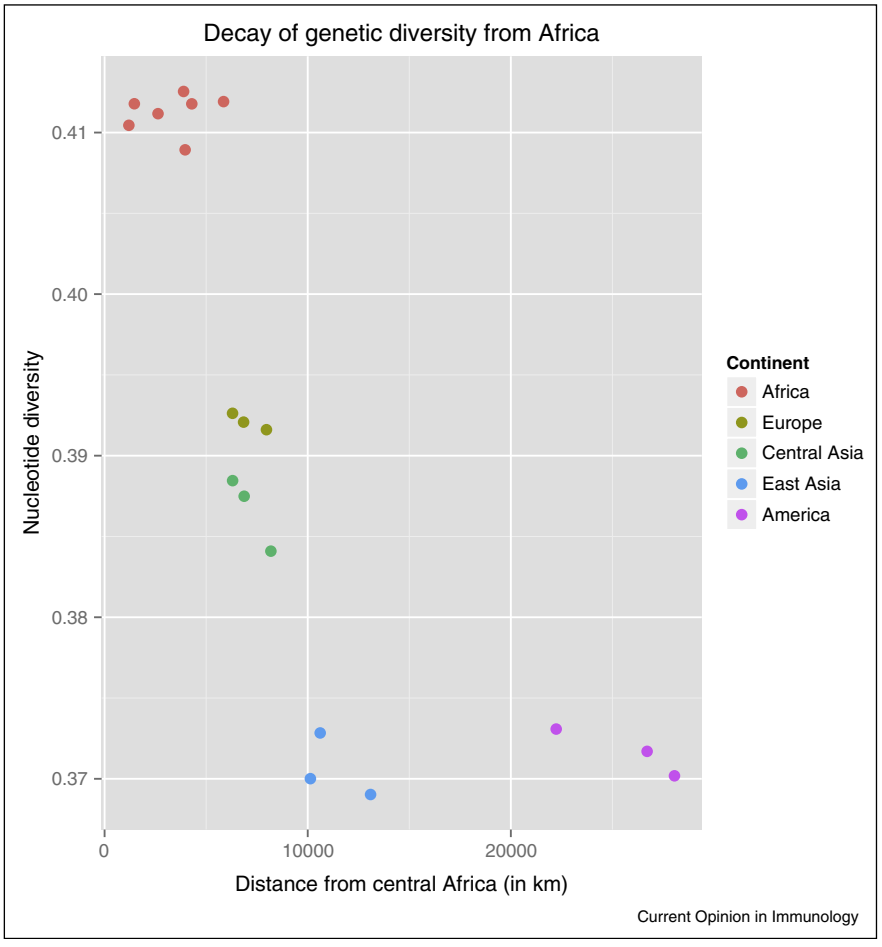
Whole-genome sequencing of multiple individuals led to the identification of common and rare genetic variants in the human genome. The 1000 Genomes Project (1000G) identified and characterized almost 40 millions Single Nucleotide Polymorphisms (SNPs) by DNA sequencing of more than 1000 individuals belonging to 15 different populations [3•]. This study builds upon previous projects aiming at characterizing human genetic variation, such as the HapMap Consortium [4], the HGDP-CEPH Panel resource [5], and the NHLBI-Exome Sequencing Project [6].

These datasets can be analyzed to infer which factors shaped human genome diversity and phenotypic variation. Patterns of worldwide genetic variation are compatible with an origin of modern humans in Africa, followed by a series of expansions, bottlenecks, and migrations between populations [7,8]. The observation of an overall decay of genetic diversity along axes of migration [9], together with a negative correlation between nucleotide variation and geographical distance from Africa [5] (Figure 1), suggests that demographic events played a major role shaping worldwide patterns of human genetic diversity. More recently, availability of sequenced DNA from ancient humans [10•,11] and archaic hominids [12,13] elucidated past admixture events and possible colonization routes.

Nonetheless, during this period of migratory events, humans have been exposed to new environments, namely to different climate conditions, pathogens, and food availability, to which they have been forced to adapt, through the action of natural selection.

Lactase persistence is one of the most notable examples of genetic adaptation, with a polymorphism promoting the expression of *LCT*, the gene encoding the enzyme lactase, being at high frequency in European populations and mostly absent elsewhere [14]. Another stunning case of natural selection in the human genome is represented by adaptation to high altitude in Tibetans. A SNP in *EPAS1*, a gene related to response to hypoxia and hemoglobin concentration, exhibits a dramatic difference in frequency between Tibetans and Han Chinese [15], despite the two populations being genetically close to

Figure 1



Decay of nucleotide diversity with geographical distance from central Africa. For each country, the average nucleotide diversity is plotted against the geographical distance from central Africa. We analyzed allele frequencies of more than 500 000 SNPs in 20 countries, using the HGDP-CEPH (http://www.cephb.fr/en/hgdp) and HapMap (http://hapmap.ncbi.nlm.nih.gov/) databases. We calculated the shortest distance between each country and central Africa on landmass routes.

each other. Further examples of genes with strong evidence of being subjected to natural selection are reviewed and discussed in [16,17].

Genomic signatures of natural selection

Detecting loci in the human genome that have been targeted by natural selection has a two-fold importance. Firstly, we can infer which selective events have shaped genome diversity most, and learn about past evolutionary events that characterized human history. Secondly, loci under natural selection are more likely to harbor functional variants, and therefore can be prioritized in screenings for association with susceptibility or resistance to diseases and infections. Indeed, genetic variants that are favorable to the carrier tend to increase in frequency in the population (a process known as *positive selection*), while deleterious mutations tend to be eliminated (*negative selection*).

Comparing orthologous genes among primate species is an effective approach to detect positive selection acting over long evolutionary timescales. On the other hand, comparing genetic variation within human populations may shed light onto more recent adaptive events. Such rapid selective pressures can shift allele frequencies of SNPs in proximity of the selected allele, causing a local reduction of nucleotide diversity (a phenomenon known as *selective sweep*). Various parameters, including the strength of selection, time of onset of the beneficial mutation, current allele frequency of targeted variant, and recombination rate, affect the extent of genetic variation reduction around the selected site (Figure 2).

Over the last years, many genome-wide scans for selection have been performed by looking at regions with a reduction of haplotype diversity and increased homozygosity [18,19,20]. Power to detect selection can be gained