Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Genetics & Development

Impact of range expansions on current human genomic diversity

Vitor Sousa[1,2,3], Stephan Peischl[1,2,3] and Laurent Excoffier[1,2]

The patterns of population genetic diversity depend to a large extent on past demographic history. Most human populations are known to have gone recently through a series of range expansions within and out of Africa, but these spatial expansions are rarely taken into account when interpreting observed genomic diversity, possibly because they are difficult to model. Here we review available evidence in favour of range expansions out of Africa, and we discuss several of their consequences on neutral and selected diversity, including some recent observations on an excess of rare neutral and selected variants in large samples. We further show that in spatially subdivided populations, the sampling strategy can severely impact the resulting genetic diversity and be confounded by past demography. We conclude that ignoring the spatial structure of human population can lead to some misinterpretations of extant genetic diversity.

Addresses
[1] CMPG, Institute of Ecology and Evolution, 3012 Bern, Switzerland
[2] Swiss Institute of Bioinformatics, 105 Lausanne, Switzerland
[3] Contributed equally to the present work.

Corresponding author: Excoffier, Laurent (Laurent.excoffier@iee.unibe.ch)

Current Opinion in Genetics & Development 2014, **29**:22–30

This review comes from a themed issue on **Genetics of human origin**

Edited by **Aida M Andrés** and **Katja Nowick**

http://dx.doi.org/10.1016/j.gde.2014.07.007

0959-437X/Published by Elsevier Ltd.

Introduction

In the last five years our appreciation of human genetic diversity has changed considerably. Whereas this diversity was then assessed from the analysis of hundreds of thousands of SNPs obtained from microarray data [1–3], whole genome data are now available at low- [4,5] or high- [6] coverage for humans and some archaic hominins [7,8]. These data, together with new inference tools (see Box 1), offer the promise of getting a finer picture of human demographic history and to more reliably identify regions under selection.

Although range expansions have occurred in ancient and recent human history [9,10•,11•], they are rarely explicitly taken into account when interpreting observed patterns of genomic diversity. We therefore propose to discuss in this review the evidence for such expansions and their predicted effects on neutral and functional human diversity. We will also argue that recent observed patterns of human diversity, namely the excess of rare and potentially deleterious mutations in large samples [12,13,14•], or the relative excess of deleterious variants in non-Africans [15••,16•,17,18••] can be easily accounted for by models including range expansions and a spatial structure of the populations.

Evidence for range expansions out of Africa

There are several genetic evidences that global patterns of diversity have been shaped by a range expansion out of Africa, such as (1) an overall higher genetic diversity in Africa than in non-African populations [5,19,20]; (2) the observation that non-African diversity is to a large extent a subset of African diversity [21,22]; (3) gradients of decreasing genetic diversity within and among continents along the expansion axis [1,22,23••,24–26] compatible with a series of founders effects during human migrations [27]; (4) higher levels of linkage disequilibrium [4,28] and longer runs of homozygosity in non-African populations [29]; (5) lower diversity in recombination hot-spots in non-Africans [19,20]; and (6) a larger proportion of non-synonymous and potentially deleterious mutations in non-Africans [16•,17,18••]. The out-of-Africa model is also supported by patterns of diversity of human pathogens (e.g. *Helicobacter pylori* [30] or tuberculosis [31]), skull morphology [32], and linguistic diversity [33].

A new 'out of Africa with hybridization' model

The finding that 1–6% of the genome of modern humans is of archaic origin has deeply shaken the field of human evolution (see reviews of [10•,11•] and [34] for an alternative interpretation). However, since more than 95% of Eurasian genetic diversity still comes from Africa, the Out of Africa model is not fully invalidated, but it clearly needs to be changed into an 'Out of Africa with hybridization' model (see [35•]). Archaic introgression now needs to be taken into account when comparing populations, since introgressed populations will show higher levels of diversity than non-introgressed ones and will also appear as genetically more divergent than other populations [10•]. It has also been proposed that differential introgression could lead to clinal patterns of genetic variation away from Africa [36], but even though it is a theoretical possibility, the small differences in introgression levels existing between Eurasian populations [37] are

Box 1 Recent progress in genomic data production and analyses

Next Generation Sequencing (NGS) technologies [85,86] have led to a shift from gene-centred studies to whole genomes studies, and allowed the sequencing of ancient DNA in archaic hominins [9,42]. These are thus exciting times as these data provide information about levels of polymorphism (e.g. SNPs and indels), linkage disequilibrium (LD) and recombination rates along the genome, and this is available for many present-day human populations distributed across the world [4] or in archaic hominins like Neanderthals and Denisovans [7,8]. Detecting regions of the genome that are identical by descent (IBD) is also easier, and the distribution of the lengths of shared haplotypes between individuals is informative about the recent history of the populations [87,88,89]. Together, these data can lead to a better understanding of human demographic history, which is also important to learn about the effects of selection and about which evolutionary forces have been shaping our genomic diversity [90,91]. Advances in genomics have been accompanied by the development of better analytical tools to handle large genome-wide datasets (reviewed in [75]). Several inference methods are based on the site frequency spectrum (SFS) or on summaries of the SFS, which can be easily obtained from SNP data [86]. The SFS describes the proportion of sites in the genome that have a given frequency in a sample from one (1D SFS Figure 1) or multiple populations (joint SFS), and it contains information about relevant demographic parameters, such as effective sizes, migration rates, and admixture events [46,47]. Even though the SFS is obtained under the assumption that all SNPs are independent and hence ignores LD, statistical associations between SNPs can provide valuable information about demography [92]. Thus, there has been a growing interest on model-based approaches accounting for linkage, such as methods based on the distribution of lengths of 'runs-of-homozygosity' (ROH) in single individuals [93,94] and on the lengths of IBD haplotypes across individuals [87,89]. Another promising class of methods explicitly models recombination along the genome using hidden-Markov-models (HMM) [95–98], a technique that has been successfully used to infer past effective sizes (e.g. [8,96]) or to map the regions of the genome that have been introgressed from Neanderthal into modern humans [38••].

unlikely to have created the observed gradients. More importantly, studies of the distribution of introgressed Neanderthal DNA segments in Eurasians [38••,39] have suggested that the current high frequency of some fragments could have been driven by adaptive introgression, and that the occurrence of several deserts of introgression, particularly on the X chromosome, could be the mark a reduced fitness of some hybrids [38••,39]. Interestingly, such reduced fitness of hybrids had been postulated by an explicit modelling of hybridization during range expansions, where it was shown that massive amounts of introgression would be expected if hybridization had not been prevented by some form of selection at the pre-zygotic or post-zygotic stage [40]. This study also postulated that the similar low level of Neanderthal introgression was not due to a single pulse of gene flow in the Middle-east, as originally proposed [41,42], but rather by similar periods of interactions between modern humans in Europe and in Asia, implying an eastern extension of Neanderthal into Asia [40]. The finding of very different introgressed segments in Europeans and Asians [38••,39] also confirmed that there had been multiple admixture events in the Middle-East, Europe and Asia.

Consequences of range expansions

Despite these numerous evidence that genetic diversity has been mainly (but not only) shaped by range expansions out of Africa, spatial expansions are still largely ignored in many human genetic studies, where non-Africans are simply seen or modelled as a mere splitting population [13,43•,44–47]. A possible reason for ignoring spatial aspects of human genetic diversity is certainly due to the high complexity of spatially explicit models involving continuous migration events. However, despite their convenient mathematical properties, the use of simpler non spatially-explicit models might sometimes lead to different interpretation of observed patterns (as in the case of archaic admixture discussed above), but they have also failed to predict two potentially important phenomena occurring during range expansions: gene surfing and expansion load.

Gene surfing: When a population expands its density at the wave front is usually low and genetic drift is the predominant evolutionary force. New territories are usually colonized by a few individuals from the wave front, such that an expansion can be considered a series of repeated founder events [48,49]. As shown in Figure 1a, the genetic diversity of expanding populations is thus expected to decrease with distance from the origin [50], as observed in human populations [23••,51,52]. Genetic drift or founder events do not change the mean allele frequencies, and the loss of many low frequency variants is thus compensated by the increase in frequency of a few others. As first observed 10 years ago by Edmonds *et al.* [53••], new mutations occurring on the wave front can sometimes increase in frequency and then spread with the wave of advance and occupy a large portion of newly colonized territories, thus mimicking the effect of selective sweeps [54]. This phenomenon of gene surfing [55•] has been evidenced in a recent human range expansion [56] and has also been invoked to explain the lowering of human genetic diversity with distance from Africa [27,52,57]. Gene surfing also affects the SFS of expanding populations. Strong drift at the wave front leads to the rapid fixation or loss of neutral mutations during the expansion process, reducing the number of sites with intermediate allele frequencies. In recently colonized areas, we thus predict an excess of low and high frequency variants relative to source populations (Figure 1c,d), a phenomenon that could also potentially contribute to the observed excess of rare variants in human populations [16•].

Expansion load: Importantly, gene surfing can occur on standing variation, and it is not restricted to neutral mutations. Positively or negatively selected variants can indeed also surf to high frequency at the expansion front [18••,58–61]. Roughly speaking, unless very strongly selected (e.g. lethal), mutations evolve almost neutrally at the front of expanding populations because drift is so