

Genetic surfing in human populations: from genes to genomes

Stephan Peischl^{1,2}, Isabelle Dupanloup^{2,3}, Lars Bosshard^{2,3}, Laurent Excoffier^{2,3}

¹ CMPG, Institute of Ecology and Evolution, Baltzerstrasse 6, 3012 Berne, Switzerland

² Interfaculty Bioinformatic Units, Baltzerstrasse 6, 3012 Berne, Switzerland

³ Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland

Short title:

Genetic surfing in humans

Keywords:

Founder effect, mutation load, expansion load, human expansions, human evolution

Correspondence

Stephan Peischl & Laurent Excoffier

Baltzerstrasse 6, 3012 Berne, Switzerland

Email: stephan.peischl@bioinformatics.unibe.ch & laurent.excoffier@iee.unibe.ch

Abstract

Genetic surfing describes the spatial spread and increase in frequency of variants that are not lost by genetic drift and serial migrant sampling during a range expansion. Genetic surfing does not modify the total number of derived alleles in a population or in an individual genome, but it leads to a loss of heterozygosity along the expansion axis, implying that derived alleles are more often in homozygous state. Genetic surfing also affects selected variants on the wave front, making them behave almost like neutral variants during the expansion. In agreement with theoretical predictions, human genomic data reveals an increase in recessive mutation load with distance from Africa, an expansion load likely to have developed during the expansion of human populations out of Africa.

Main text

Since human populations have spread over the planet by a series of spatial range expansions, it is important to understand the consequences of these expansions on our genomic diversity. In a seminal paper, Cavalli-Sforza and colleagues [1] have shown that new mutations occurring on the edge of spatial expansion can increase in frequency and spread over a large portion of the newly colonized territories, far from their place of origin where they can even disappear. This increase in frequency of rare variants was later coined allele surfing [2], but we shall use here the more generic term of genetic surfing. It applies not only to new mutations, but to any standing variants happening to be on or close to the wave front of a range expansion [3-5], or to alleles introgressing from other species [6,7]. The genetic consequences of range expansions for single neutral loci are fairly well understood, and they have been described in several reviews [see e.g. 3,8-11]. In the present review, we will thus focus more on the causes and effects of surfing at the genomic level rather than at the single locus level, and contrast the surfing of neutral and selected variants, as range expansions can also affect the latter [12-15].

Mechanism and consequences of genetic surfing

Genetic surfing results from the stochastic evolution of allele frequencies on expanding wave fronts where population densities are low and genetic drift is strong. It has been described in models of continuous habitat where population densities are monotonously declining towards the front of the expansion [4,5,14,16,17] (**Fig. 1A**), and in models with discrete habitats where the expansion proceeds by series of serial founder effects [18-23] (**Fig. 1B**). Note that from a modelling perspective, genetic surfing is conceptually similar to the hitchhiking of neutral (or deleterious) alleles with beneficial variants that sweep through a spatially structured population [24,25]. Due to the sampling of individuals colonizing new territories and high rates of genetic drift on the front, some derived alleles are lost at some loci, and some alleles are increasing in frequencies at others. It is this increase in frequency over time and space at those loci where derived alleles are not lost that has been called genetic surfing. It is difficult to define genetic surfing more precisely than this propagation of alleles conditional on non-loss [19], as some alleles can spread over areas of variable size and reach variable frequencies [2]. In extreme cases, derived alleles can fix on the wave front and surf over very large distances, leading to sectors of no or very low diversity in two

dimensional expansions [4,5] (**Fig. 1C**). Range expansions thus globally lead to a decrease of genetic diversity along the expansion axis [26] due to the stochastic loss of variants. Genetic surfing is favored by low effective sizes on the front, low migration rate from the back of the wave, and high levels of growth on the wave front [2,19,27], but Allee effects and long distance dispersal can limit or prevent surfing and preserve diversity [28-34].

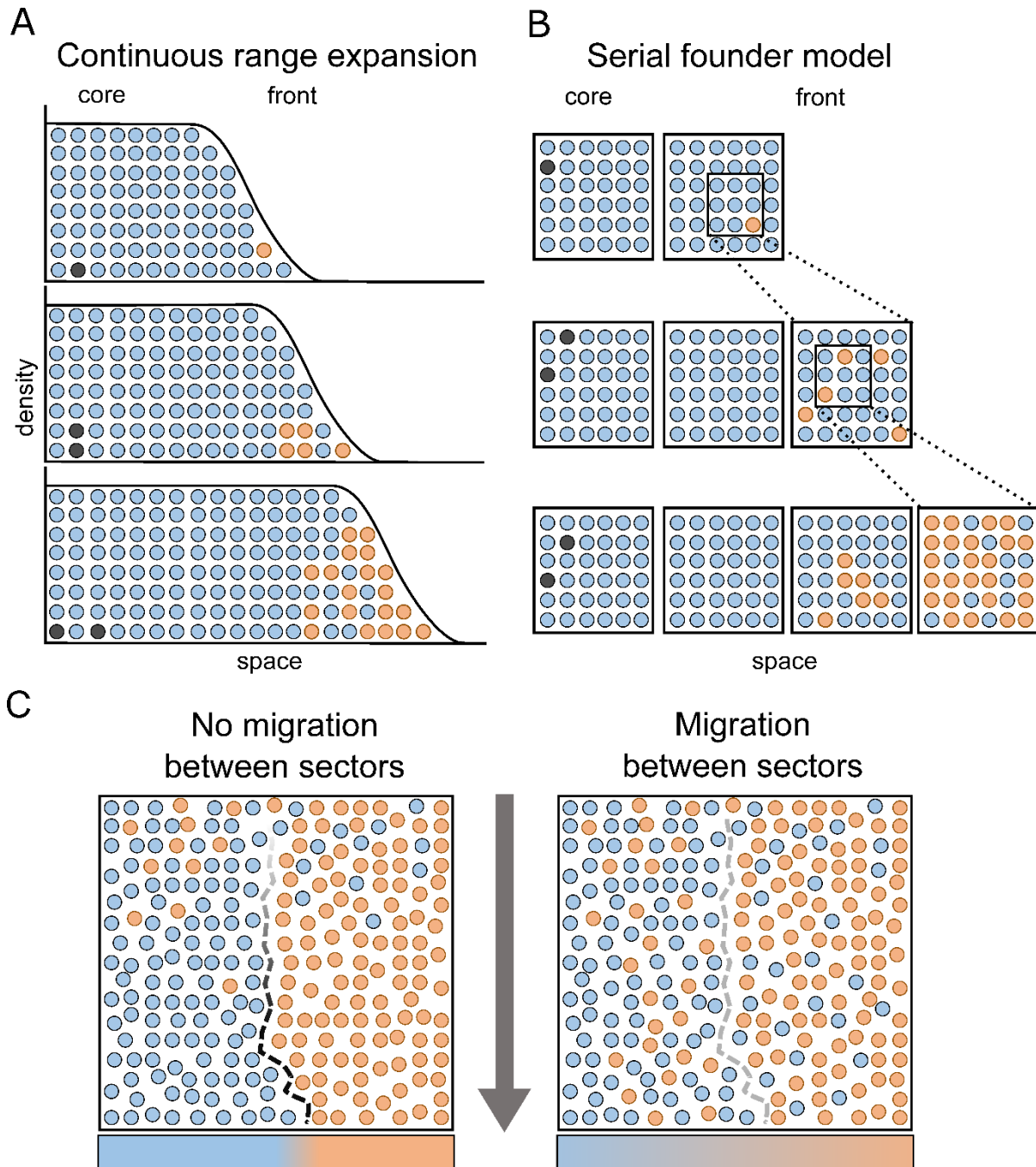


Figure 1. Allele frequency changes due to genetic surfing. **A:** During a range expansion in a continuous one dimensional habitat, a new mutation appearing on the wave front (orange) can increase in frequency and surf on the wave of advance, reaching high frequencies in newly occupied territories. **B:** Discrete serial founder model of range expansion. A given number of founders are sampled from the leading deme to colonize an empty deme on the right. These founders then reproduce, and a new set of founders is selected to colonize the next deme. Note that in both models A and B, a new mutation appearing in the core of the range (black) has a much lower probability to increase in frequency than a new mutation in the front (orange), because individuals in the core have less descendants than those on the front. **C:** Two-dimensional expansions from top to bottom, starting from a well-mixed population. Without migration between colonized areas (left pane), the expansion leads to the (local) fixation of a random neutral allele by surfing and thus to the creation of sectors of low diversity. The sectors are delineated by a dashed line whose shade of grey indicates the intensity of the differentiation between sectors. With lateral migrations between colonized areas, the sectors do not emerge as clearly, and a transitory gradient of diversity perpendicular to the expansion axis can establish (bottom row).

Surfing of standing vs. new mutations

At neutral loci, the overall frequency of derived alleles in the genome that are already present at the onset of a range expansion (standing variants) should not change under the sole effect of drift, founder effects, or bottlenecks because the loss of derived alleles at some loci is compensated by the increase in frequency of derived alleles at other loci. Individuals on the front of an expansion should therefore have a smaller number of polymorphic sites than individuals from the core, but they should be more often homozygote for derived alleles. Overall, they should thus have similar numbers of derived alleles in their genome than individuals in the core (**Fig. 2**). Front individuals should also have higher frequencies of derived alleles at sites shared with core individuals, a property that has been exploited to polarize the direction of range expansions and to locate their geographic origin [35]. Individuals have a higher chance to transmit their genes to later generations if they are present on a wave front, because they have a higher reproductive success than individuals from the core [36]. It may therefore appear as if genetic surfing is a non-neutral process because any given allele that finds itself on the wave front experiences an immediate increase in absolute frequency, due to the higher growth rate on the expansion front than in the core. However, this apparent increase in absolute fitness is only visible if we condition mutations to occur in founding individuals, and vanishes if we take into account

that many mutations at the wave front are not sampled during later founder events (**Figs. 1A** and **B**, see also $s = 0$ in **Fig. 3**).

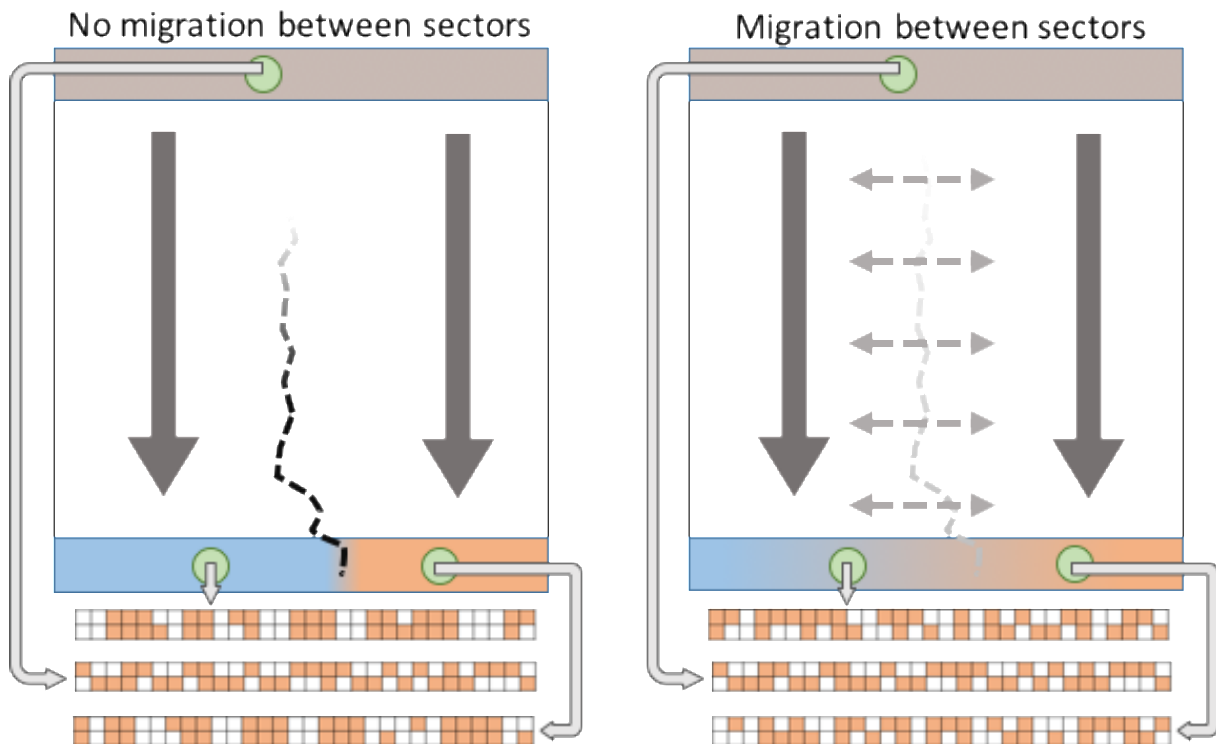


Figure 2. Illustration of the type of genomic diversity expected in core and front individuals after surfing under scenarios represented in Fig. 1C. Vertical arrows represent the direction of the expansion, whereas dashed horizontal arrows in the right pane represent lateral gene flow between neighboring demes. We show here the putative genomic diversity of diploid individuals genotyped at 30 loci either sampled from the core (top green circle) or from the front (bottom green circles). Derived alleles are denoted as orange squares. The expected total number of derived alleles is the same in all individuals, but due to surfing, front individuals have less sites with derived alleles, and thus more loci with homozygous derived alleles than typical core individuals. Derived allele frequencies at sites shared between front and core populations are therefore larger in front than in core populations [19]. In presence of migrations between sectors (right pane), the number of polymorphic sites on the front is higher than without lateral gene flow, and heterozygosity is restored more rapidly.

Gradients of diversity due to surfing

The fact that genetic diversity should decrease by allele loss almost linearly with distance from the origin of an expansion [18,20] has been used to identify the origin of human expansions out of Africa. Early studies of microsatellite diversity have revealed an almost perfect linear decrease of heterozygosity with distance from Ethiopia [21,37], which has been confirmed with genomic data [e.g.38]. Interestingly, human range expansions are also

associated with gradients of linkage disequilibrium [39,40], recombinational diversity [41], length of runs of homozygosity (ROHs) [42], allele frequencies [35,43], as well as a right shift or a flattening of the site frequency spectrum [38,44]. Given these multiple evidence for the creation of gradients of diversity along expansion axes, one would expect that the largest extent of differentiation should be between the source and the edge of an expansion. However, PCA analyses of simulated data reveal that the main axis of differentiation after an expansion runs perpendicular to the expansion axis [45], which is explained by the differentiation existing between surfing sectors along the expansion axis (see **Fig. 1C**). Consequently, the classical PC analyses of human European diversity showing gradients of differentiation from the Near-East to North Western Europe [46] are unlikely to be the signature of a Neolithic demic diffusion [45], but could rather reflect re-expansions (e.g. post-LGM) from different refuge areas in Southern Europe [47].

Surfing or adaptive evolution?

A striking feature of genome scans for positive selection in humans is that they reveal many candidate SNPs that have very similar geographic distributions, with high contrasts observed between continental groups [48,49], and where Africa and America show the largest number of loci with extreme frequency differences [50]. Simulations have shown that such large continental differences could simply arise from surfing of neutral variants during the out of Africa expansion [51]. Therefore, previous claims of signals of adaptation based mainly on the spatial distribution of allele frequencies [e.g. 52,53] should probably be taken with great caution or revised [54], and could partly explain some lack of consistency between the outcome of various genome scans [55]. It has also been recently recognized that there was an overall lack of evidence for strong signals of selective sweeps in the human genome [56-58], with only a handful of regions with fixed differences between continents [59]. It is difficult to understand why positive selection, if active, has not led to more fixations of beneficial mutations in human populations even under polygenic selection models [60]. A low number of fixed difference due to genetic surfing between the source of human expansions (Africa) and the edges of expansions is equally puzzling, as one would expect that surfing mutations should quickly fix on wave fronts (**Fig. 3**) [1,2,19]. However, mutations can surf in large populations without necessarily being fixed on the front [2], and even if different mutations fix in different sectors, migration between parallel sectors can lead to

intermediate final frequencies in newly colonized areas (**Fig. 1C**). In addition, some long distance dispersal events between demes in the wake of the expansion might have reshuffled diversity and limited the fixation of derived alleles in non-African populations [61].

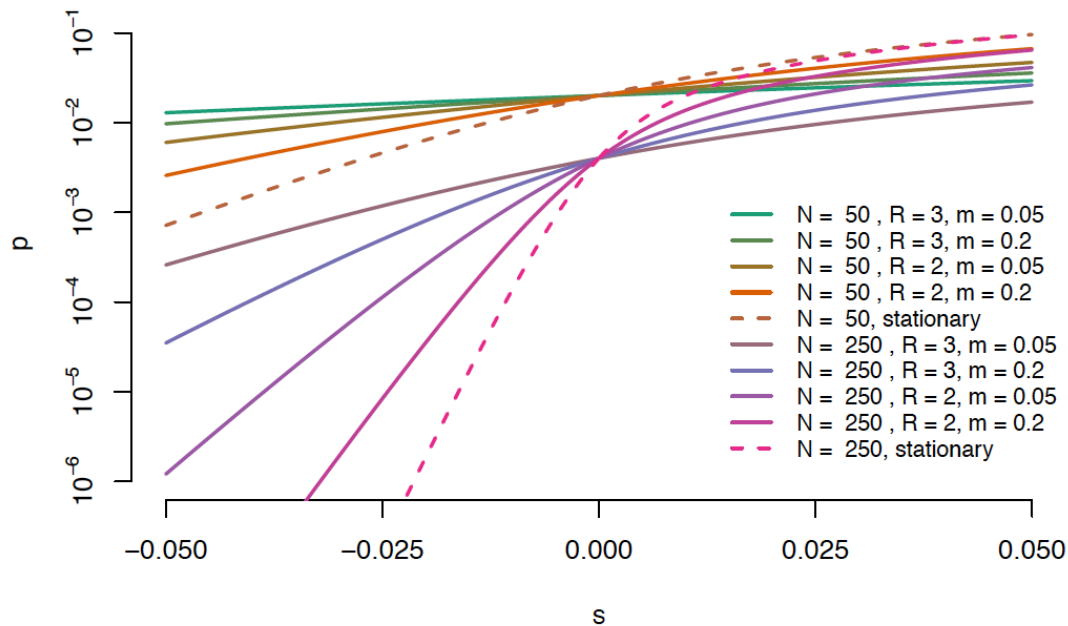


Figure 3. Probability of fixation on the wave front for a new mutation occurring at the wave front during a one dimensional range expansion (serial founder model) for deleterious, neutral and advantageous variants (figure based on equation (1) in ref [15]). Mutations are assumed to occur in a fully colonized deme just before the sampling of founders such that they have an initial frequency of $1/2N$ (**Fig 1B**). The dashed line represents the probability of fixation in a single stationary deme, which is smaller for deleterious mutations and larger for beneficial ones.

Surfing of deleterious mutations leads to an expansion load

Genetic surfing is not restricted to neutral variants. The probability that a mutation establishes itself on the expansion front and spreads with the expanding wave is largely driven by genetic drift, which makes the rate at which beneficial, neutral or deleterious mutations are established at the expansion front more similar to each other relative to what is expected in the core [15] (**Fig. 3**). Mildly positively or negatively selected variants therefore mostly behave like neutral variants on the front of the expansion. Deleterious variants can therefore temporarily surf on the expansion wave due to the high growth rate

prevailing in front populations [12]. In asexual (non-recombining) organisms with high mutation rates, genetic surfing can even lead to an accumulation of deleterious mutations at the expansion front [14]. Interestingly, and perhaps somewhat counter-intuitively, beneficial mutations that occur on the front of an expansion have a higher probability to become permanently established in the whole population (even if they fail to surf) as compared to mutations that occur in the core [13]. Indeed, beneficial mutations occurring on the wave front quickly increase in absolute frequency due to high growth rates in these marginal populations (**Fig. 1A, 1B**), which lowers the probability that they are initially lost. This effect is similar to what has been observed in demographically growing well-mixed populations [62]. The spread of beneficial mutations in the core has, however, little impact on the evolution at the expansion front because beneficial mutations sweep through the core at a speed [25] that is much lower than that of the expanding wave [14,63,64]. Since beneficial mutations that establish on the front spread much faster than beneficial mutations in the core [14,65], they provide opportunity for rapid adaptation during early stages of an expansion [66]. In the long run, however, genetic drift at the expansion front dominates and, since beneficial mutations occur at a lower rate than deleterious ones [67], the overall fitness of the populations should decrease along the expansion axis under a broad range of conditions [15]. The resulting elevated mutational burden of expanding species has been coined expansion load [15]. Expansion load generally builds up fastest under conditions that favor genetic surfing, such as small local carrying capacities, high growth rates, and low migration rates [2], which contributes to making the probabilities with which deleterious and beneficial alleles fix on the wave front relatively similar to that of neutral alleles (**Fig. 3**).

In the absence of epistasis or dominance, only new mutations contribute to expansion load [63], and the resulting decrease in fitness is relatively slow. If some deleterious mutations are (partially) recessive [63], the increase in homozygosity of standing deleterious variants along the expansion axis leads to a very rapid increase of the (recessive) mutation load [63]. Another interesting consequence of recessivity is the establishment of heterozygosity-fitness correlations during 2D expansions: migration between sectors in which different mutations have established (see **Fig. 1C and Fig 2**) creates a ridge of high fitness “hybrids” in which the effects of deleterious mutations are partially masked [63]. Whereas this effect has been observed in plants after a range expansion [68], it remains to be shown in humans.

Even though expansion load is a transient phenomenon, simulations show that the increased mutation load can persist for hundreds to thousands of generations after the expansion has ended [15]. The rate of purging of deleterious mutations depends on factors such as selection coefficients, the degree of dominance and the amount of gene flow between populations [15,63,69-71].

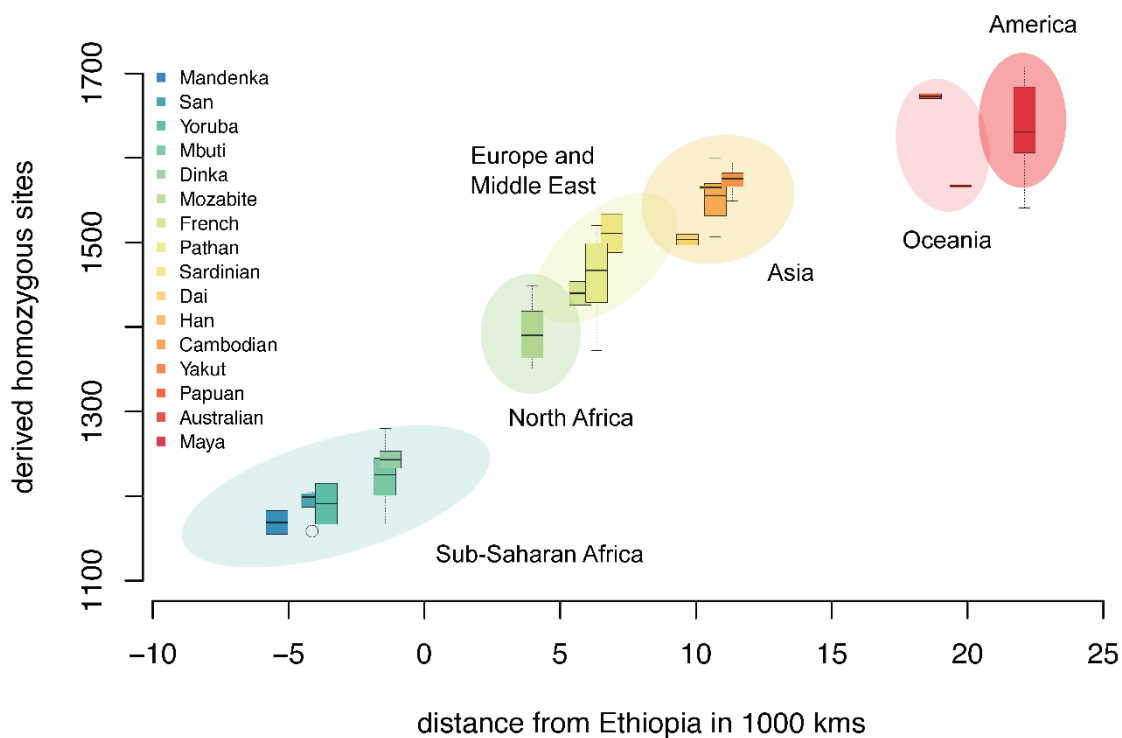


Figure 4. Recessive mutation load as a function of the least-cost distance from Addis-Ababa (Ethiopia). Negative distances denote sub-Saharan-African populations. Recessive mutation load is measured as the number of sites homozygote for derived mutation at sites that have a GERP rejected substitution score [72] larger than 4, indicative of relatively strong evolutionary constraints. Reported populations are from [44], complemented by samples from [73,74], all individuals examined for the same 44Mb of exomic sequences without missing data (see [44] for details).

Evidence for expansion load in human populations

Evidence for expansion load is generally difficult to get as fitness is difficult to measure directly. Ideally, one would need to evidence the spread over a wide area of attested deleterious mutations in populations with a history of range expansions. In Europe, a couple

of disease mutations are increasing in frequencies along a South-East to North-West gradient (e.g. *HFE*-C282Y associated to hemochromatosis [75] or Δ F508 associated to cystic fibrosis [76]), and a *CPT1A* mutation associated with hypoglycemia and high infant mortality has been shown recently to have greatly increased in frequency in arctic populations [77]. Interestingly, in the three cases mentioned above, the spread of these deleterious variants over long distances has been explained by past episodes of positive selection, but genetic surfing of only deleterious variants seems an equally reasonable and more parsimonious hypothesis.

Another way to approximate mutation load from sequence data is to estimate the degree of evolutionary conservation of mutations at particular positions in the genome (see e.g., ref [78]). Using such an approach, a clear gradient of homozygosity at predicted deleterious sites has been observed along the human expansion axis out of Africa [44] (**Fig. 4**), in keeping with predictions [63]. Furthermore, this cline becomes weaker with increasing deleteriousness of mutations [44], suggesting that the evolution of mildly and moderately deleterious mutations are dominated by strong genetic drift during the expansion, whereas selection still dominates the evolution of strongly deleterious mutations, as predicted by theory [15].

The impact of demographic history on mutation load is still a hotly debated topic [44,71,78-83]. When focusing on the number of predicted deleterious alleles per individuals, which is equivalent to considering that load is due to co-dominant mutations, no differences were found between Europeans, Americans of European ancestry or African-Americans [80,81,84]. However, the relative proportion of alleles predicted deleterious is significantly higher in Europeans as compared to Sub-saharan Africans [15,79]. These apparently contradictory results can be understood from **Fig. 2**, by realizing that while each individual carries roughly similar numbers of derived alleles before and after surfing, these alleles are distributed over fewer loci in populations having gone through a range expansion (e.g. Europeans) as compared to core populations (e.g. Sub-Saharan Africans), and therefore reach higher frequencies in front populations [19]. Deleterious alleles that are kept at very low frequency in core populations and that have increased in frequency after surfing have a higher probability of being detected in a sample of individuals on the front, contributing to

the observation of a larger proportion of non-synonymous and potentially deleterious alleles in non-African populations [15,38,79].

The total genetic load of an individual and its fitness will also highly depend on the distribution of dominance coefficients in the genome [44,71]. If all mutations were co-dominant, one would expect no or little differences in genetic load between populations, because mildly deleterious mutations have mainly evolved as neutral mutations during the expansion (or during a bottleneck [81]). However, the mutation load should be clearly larger in front populations if some of the surfing deleterious mutations are partially recessive. The extent of this load excess in front populations is difficult to assess since recessivity appears to be correlated with deleteriousness [85]. The most deleterious variants, which should be the most recessive ones, should not be influenced much by surfing since their frequencies are maintained consistently low in all populations by selection [44]. The total genomic frequency of very mildly deleterious variants and their effect on load should also not vary due to surfing as they should be almost co-dominant. It is thus likely that the difference in load between populations should be mainly due to partially recessive variants of intermediate effects (e.g. [71]). These predictions offer interesting areas of research in pheno-genomic diversity, showing the need of better methods to assess the phenotypic effect of mutations in humans, and to distinguish between the effects of somatic and germ-line mutations [86].

Acknowledgements

ID and LB were partially supported by a Swiss NSF grant No 31003A-143393 to LE.

References

1. Edmonds CA, Lillie AS, Cavalli-Sforza LL: **Mutations arising in the wave front of an expanding population.** *PNAS* 2004, **101**:975-979.
**** Seminal paper introducing the phenomenon of genetic surfing, described here as the spread and increase in frequency of new mutations over a wide geographic area in the newly colonized territories.**
2. Klopstein S, Currat M, Excoffier L: **The fate of mutations surfing on the wave of a range expansion.** *Mol Biol Evol* 2006, **23**:482-490.
**** Detailed simulation study of the fate of mutations spreading during two-dimensional population expansions. Introduction of the term "surfing" to describe the process discovered by Cavalli-Sforza. New mutations are shown to have a much larger chance to survive if they occur on the edge of the expansion than in the core. Genetic surfing is also shown to occur when large populations (e.g. Neolithic) expand, but that successful mutations rarely fix in new territories in that case, contrary to what happens during the expansion of small (e.g. Paleolithic) populations.**
3. Excoffier L, Ray N: **Surfing during population expansions promotes genetic revolutions and structuration.** *Trends Ecol Evol* 2008, **23**:347-351.
4. Hallatschek O, Hersen P, Ramanathan S, Nelson DR: **Genetic drift at expanding frontiers promotes gene segregation.** *Proc Natl Acad Sci U S A* 2007, **104**:19926-19930.
5. Hallatschek O, Nelson DR: **Gene surfing in expanding populations.** *Theor Popul Biol* 2008, **73**:158-170.
6. Currat M, Ruedi M, Petit RJ, Excoffier L: **The hidden side of invasions: Massive introgression by local genes.** *Evolution* 2008, **62**:1908-1920.
7. Excoffier L, Quilodran CS, Currat M: **Models of hybridization during range expansions and their application to recent human evolution.** In *Cultural Developments in the Eurasian Paleolithic and the Origin of Anatomically Modern Humans*. Edited by Derevianko AP, Shunkov MV: Dept. of the Institute of Archaeology and Ethnography of the Siberian Branch of the Russian Academy of Sciences; 2014:122-137.
8. Excoffier L, Foll M, Petit RJ: **Genetic Consequences of Range Expansions.** *Annual Review of Ecology Evolution and Systematics* 2009, **40**:481-501.
9. Waters JM, Fraser CI, Hewitt GM: **Founder takes all: density-dependent processes structure biodiversity.** *Trends Ecol Evol* 2013, **28**:78-85.
10. Currat M: **Consequences of population expansions on European genetic diversity.** In *Population Dynamics in Prehistory and Early History: New Approaches Using Stable Isotopes and Genetics*. Edited by Kaiser E, Burger J, Schier W: Walter de Gruyter & Co; 2012:3-16.
11. Sousa V, Peischl S, Excoffier L: **Impact of range expansions on current human genomic diversity.** *Curr Opin Genet Dev* 2014, **29**:22-30.
12. Travis JM, Munkemuller T, Burton OJ, Best A, Dytham C, Johst K: **Deleterious mutations can surf to high densities on the wave front of an expanding population.** *Mol Biol Evol* 2007, **24**:2334-2343.
13. Lehe R, Hallatschek O, Peliti L: **The rate of beneficial mutations surfing on the wave of a range expansion.** *PLoS Comput Biol* 2012, **8**:e1002447.
14. Hallatschek O, Nelson DR: **Life at the front of an expanding population.** *Evolution* 2010, **64**:193-206.
**** Detailed analysis of the fate of beneficial and deleterious mutations happening on the edge of continuous two-dimensional spatial expansions in non-recombining (micro)organisms. The angle of opening sectors containing beneficial variants is shown to**

be proportional to the square root of their selection coefficients. Wave front are predicted to be loaded with deleterious mutations that are transiently surfing on the expanding wave. First suggestion that non-African human populations could have a higher load due to their history of expansion.

15. Peischl S, Dupanloup I, Kirkpatrick M, Excoffier L: **On the accumulation of deleterious mutations during range expansions.** *Mol Ecol* 2013, **22**:5972-5982.
**** First paper describing the build-up of an expansion load due to the surfing of new deleterious mutations during a range expansion, which leads to a long-lasting fitness gradient between the core of the range and its edge. An analytical study under a serial founder model shows that i) too deleterious mutations will not surf implying that the expansion load is only due to deleterious mutations of small or intermediate effects, and ii) the probability of fixation of deleterious mutations on the wave front is much higher than in the core and that of beneficial mutations is much lower, and not much higher than that of neutral mutations. .**
16. Vlad MO, Cavalli-Sforza LL, Ross J: **Enhanced (hydrodynamic) transport induced by population growth in reaction-diffusion systems with application to population genetics.** *Proc Natl Acad Sci U S A* 2004, **101**:10249-10253.
17. Neher RA, Hallatschek O: **Genealogies of rapidly adapting populations.** *Proc Natl Acad Sci U S A* 2013, **110**:437-442.
18. DeGiorgio M, Degnan JH, Rosenberg NA: **Coalescence-time distributions in a serial founder model of human evolutionary history.** *Genetics* 2011, **189**:579-593.
19. Slatkin M, Excoffier L: **Serial Founder Effects During Range Expansion: A Spatial Analog of Genetic Drift.** *Genetics* 2012, **191**:171-181.
20. DeGiorgio M, Jakobsson M, Rosenberg NA: **Explaining worldwide patterns of human genetic variation using a coalescent-based serial founder model of migration outward from Africa.** *Proc Natl Acad Sci U S A* 2009, **106**:16057-16062.
21. Ramachandran S, Deshpande O, Roseman CC, Rosenberg NA, Feldman MW, Cavalli-Sforza LL: **Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa.** *PNAS* 2005, **102**:15942-15947.
22. Henn BM, Cavalli-Sforza LL, Feldman MW: **The great human expansion.** *PNAS* 2012, **109**:17758-17764.
23. Deshpande O, Batzoglou S, Feldman MW, Cavalli-Sforza LL: **A serial founder effect model for human settlement out of Africa.** *Proc Biol Sci* 2009, **276**:291-300.
24. Barton NH, Etheridge AM, Kelleher J, Veber A: **Genetic hitchhiking in spatially extended populations.** *Theoretical Population Biology* 2013, **87**:75-89.
25. Fisher RA: **The wave of advance of advantageous genes.** *Annals of eugenics* 1937, **7**:355-369.
26. Austerlitz F, Jung-Muller B, Godelle B, Gouyon P-H: **Evolution of coalescence times, genetic diversity and structure during colonization.** *Theor. Popul. Biol.* 1997, **51**:148-164.
27. Miller JR: **Survival of mutations arising during invasions.** *Evolutionary Applications* 2010, **3**:109-121.
28. Arenas M, Ray N, Currat M, Excoffier L: **Consequences of range contractions and range shifts on molecular diversity.** *Mol Biol Evol* 2012, **29**:207-218.
29. Roques L, Garnier J, Hamel F, Klein EK: **Allee effect promotes diversity in traveling waves of colonization.** *Proc Natl Acad Sci U S A* 2012, **109**:8828-8833.
30. Bonnefon O, Coville J, Garnier J, Hamel F, Roques L: **The spatio-temporal dynamics of neutral genetic diversity.** *Ecological Complexity* 2014, **20**:282-292.
31. Alves I, Arenas M, Currat M, Hanulova AS, Sousa VC, Ray N, Excoffier L: **Long distance dispersal shaped patterns of human genetic diversity in Eurasia.** *Mol Biol Evol* 2015.
32. Fayard J, Klein EK, Lefevre F: **Long distance dispersal and the fate of a gene from the colonization front.** *Journal of Evolutionary Biology* 2009, **22**:2171-2182.

33. Bialozyt R, Ziegenhagen B, Petit RJ: **Contrasting effects of long distance seed dispersal on genetic diversity during range expansion.** *J Evol Biol* 2006, **19**:12-20.
34. Ray N, Excoffier L: **A first step towards inferring levels of long-distance dispersal during past expansions.** *Mol Ecol Resour* 2010, **10**:902-914.
35. Peter BM, Slatkin M: **Detecting range expansions from genetic data.** *Evolution* 2013, **67**:3274-3289.
36. Moreau C, Bherer C, Vezina H, Jomphe M, Labuda D, Excoffier L: **Deep human genealogies reveal a selective advantage to be on an expanding wave front.** *Science* 2011, **334**:1148-1150.
37. Prugnolle F, Manica A, Balloux F: **Geography predicts neutral genetic diversity of human populations.** *Curr Biol* 2005, **15**:R159-160.
38. Li JZ, Absher DM, Tang H, Southwick AM, Casto AM, Ramachandran S, Cann HM, Barsh GS, Feldman M, Cavalli-Sforza LL, et al.: **Worldwide human relationships inferred from genome-wide patterns of variation.** *Science* 2008, **319**:1100-1104.
39. Jakobsson M, Scholz SW, Scheet P, Gibbs JR, VanLiere JM, Fung HC, Szpiech ZA, Degnan JH, Wang K, Guerreiro R, et al.: **Genotype, haplotype and copy-number variation in worldwide human populations.** *Nature* 2008, **451**:998-1003.
40. Bosch E, Laayouni H, Morcillo-Suarez C, Casals F, Moreno-Estrada A, Ferrer-Admetlla A, Gardner M, Rosa A, Navarro A, Comas D, et al.: **Decay of linkage disequilibrium within genes across HGDP-CEPH human samples: most population isolates do not show increased LD.** *BMC Genomics* 2009, **10**:338.
41. Mele M, Javed A, Pybus M, Zalloua P, Haber M, Comas D, Netea MG, Balanovsky O, Balanovska E, Jin L, et al.: **Recombination gives a new insight in the effective population size and the history of the old world human populations.** *Mol Biol Evol* 2012, **29**:25-30.
42. Pemberton TJ, Absher D, Feldman MW, Myers RM, Rosenberg NA, Li JZ: **Genomic patterns of homozygosity in worldwide human populations.** *Am J Hum Genet* 2012, **91**:275-292.
43. Currat M, Excoffier L: **The effect of the Neolithic expansion on European molecular diversity.** *Proceedings of the Royal Society London B* 2005, **272**:679-688.
44. Henn BM, Botigue LR, Peischl S, Dupanloup I, Lipatov M, Maples BK, Martin AR, Musharoff S, Cann H, Snyder MP, et al.: **Distance from sub-Saharan Africa predicts mutational load in diverse human genomes.** *Proc Natl Acad Sci U S A* 2016, **113**:E440-449.

**** Detailed genomic analysis of the mutation load of seven human populations arranged along the expansion axis out of Africa. The mutation load due non-extremely deleterious mutation increases linearly with distance from Africa, suggesting it has been created during the expansion process, and thus giving a strong support for the expansion load hypothesis in humans. Most of this load seems due to the increase in frequency of standing deleterious variants.**
45. Francois O, Currat M, Ray N, Han E, Excoffier L, Novembre J: **Principal component analysis under population genetic models of range expansion and admixture.** *Mol Biol Evol* 2010, **27**:1257-1268.
46. Menozzi P, Piazza A, Cavalli-Sforza LL: **Synthetic maps of human gene frequencies in Europeans.** *Science* 1978, **201**:786-792.
47. Arenas M, Francois O, Currat M, Ray N, Excoffier L: **Influence of admixture and paleolithic range contractions on current European diversity gradients.** *Mol Biol Evol* 2013, **30**:57-61.
48. Coop G, Pickrell JK, Novembre J, Kudaravalli S, Li J, Absher D, Myers RM, Cavalli-Sforza LL, Feldman MW, Pritchard JK: **The role of geography in human adaptation.** *PLoS Genet* 2009, **5**:e1000500.
49. Scheinfeldt LB, Biswas S, Madeoy J, Connelly CF, Akey JM: **Clusters of adaptive evolution in the human genome.** *Front Genet* 2011, **2**:50.
50. 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.** *Annals of Human Genetics* 2009, **73**:95-108.

51. Currat M, Excoffier L, Maddison W, Otto SP, Ray N, Whitlock MC, Yeaman S: **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* 2006, **313**:172; author reply 172.
52. Evans PD, Gilbert SL, Mekel-Bobrov N, Vallender EJ, Anderson JR, Vaez-Azizi LM, Tishkoff SA, Hudson RR, Lahn BT: **Microcephalin, a Gene Regulating Brain Size, Continues to Evolve Adaptively in Humans**. *Science* 2005, **309**:1717-1720.
53. Pickrell JK, Coop G, Novembre J, Kudaravalli S, Li JZ, Absher D, Srinivasan BS, Barsh GS, Myers RM, Feldman MW, et al.: **Signals of recent positive selection in a worldwide sample of human populations**. *Genome Res* 2009, **19**:826-837.
54. Hofer T, Foll M, Excoffier L: **Evolutionary forces shaping genomic islands of population differentiation in humans**. *BMC Genomics* 2012, **13**:107.
55. Akey JM: **Constructing genomic maps of positive selection in humans: where do we go from here?** *Genome Res* 2009, **19**:711-722.
56. Hernandez RD, Kelley JL, Elyashiv E, Melton SC, Auton A, McVean G, Genomes P, Sella G, Przeworski M: **Classic selective sweeps were rare in recent human evolution**. *Science* 2011, **331**:920-924.
57. Genomes Project C, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, et al.: **A global reference for human genetic variation**. *Nature* 2015, **526**:68-74.
58. Granka JM, Henn BM, Gignoux CR, Kidd JM, Bustamante CD, Feldman MW: **Limited evidence for classic selective sweeps in African populations**. *Genetics* 2012, **192**:1049-1064.
59. Genomes Project C, Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, Gibbs RA, Hurles ME, McVean GA: **A map of human genome variation from population-scale sequencing**. *Nature* 2010, **467**:1061-1073.
60. Pritchard JK, Pickrell JK, Coop G: **The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation**. *Current biology* 2010, **20**:R208-R215.
61. Alves I, Arenas M, Currat M, Sramkova Hanulova A, Sousa VC, Ray N, Excoffier L: **Long-Distance Dispersal Shaped Patterns of Human Genetic Diversity in Eurasia**. *Mol Biol Evol* 2016, **33**:946-958.
*** First evidence that long distance dispersal has been important during human expansions, potentially limiting the effect of genetic surfing and preserving diversity in newly colonized territories.**
62. Otto SP, Whitlock MC: **The probability of fixation in populations of changing size**. *Genetics* 1997, **146**:723-733.
63. Peischl S, Excoffier L: **Expansion load: recessive mutations and the role of standing genetic variation**. *Mol Ecol* 2015, **24**:2084-2094.
*** Description of the expansion load at recessive loci. The surfing of standing recessive mutations leads to the development of a recessive load in a few generations after the start of the expansion. The surfing of different recessive mutations in different sectors can create a heterozygosity-fitness correlation in demes receiving migrants from both sectors on the edge of the expansion.**
64. Skellam JG: **Random dispersal in theoretical populations**. *Biometrika* 1951, **38**:196-218.
65. Peischl S, Kirkpatrick M, Excoffier L: **Expansion load and the evolutionary dynamics of a species range**. *Am Nat* 2015, **185**:E81-93.
*** Study of expansion load in case of hard selection, leading to a slow down of the speed of expansion with distance from the source, which can impose limits to the range of a species in absence of recombination.**
66. Gralka M, Stiewe F, Farrell F, Moebius W, Waclaw B, Hallatschek O: **Allele Surfing Promotes Microbial Adaptation from Standing Variation**. *bioRxiv* 2016:049353.

67. Eyre-Walker A, Keightley PD: **The distribution of fitness effects of new mutations.** *Nature Reviews Genetics* 2007, **8**:610-618.
68. Keller S, Fields P, Berardi A, Taylor D: **Recent admixture generates heterozygosity–fitness correlations during the range expansion of an invading species.** *Journal of evolutionary biology* 2014, **27**:616-627.
69. Balick DJ, Do R, Cassa CA, Reich D, Sunyaev SR: **Dominance of Deleterious Alleles Controls the Response to a Population Bottleneck.** *PLoS Genet* 2015, **11**:e1005436.
70. Kirkpatrick M, Jarne P: **The Effects of a Bottleneck on Inbreeding Depression and the Genetic Load.** *Am Nat* 2000, **155**:154-167.
71. Gravel S: **When is Selection Effective?** *Genetics* 2016.
**** Important paper disentangling the concepts of mutation load and of selection efficacy. The paper discusses several commonly used measures of selection efficacy and provides guidelines of how to estimate the efficacy of selection based on allele frequency data. The theoretical results are applied to human data and show that there are indeed a present-day differences in the efficacy of selection.**
72. Davydov EV, Goode DL, Sirota M, Cooper GM, Sidow A, Batzoglou S: **Identifying a High Fraction of the Human Genome to be under Selective Constraint Using GERP plus.** *Plos Computational Biology* 2010, **6**.
73. Prufer K, Racimo F, Patterson N, Jay F, Sankararaman S, Sawyer S, Heinze A, Renaud G, Sudmant PH, de Filippo C, et al.: **The complete genome sequence of a Neanderthal from the Altai Mountains.** *Nature* 2014, **505**:43-+.
74. Reich D, Green RE, Kircher M, Krause J, Patterson N, Durand EY, Viola B, Briggs AW, Stenzel U, Johnson PLF, et al.: **Genetic history of an archaic hominin group from Denisova Cave in Siberia.** *Nature* 2010, **468**:1053-1060.
75. Distant S, Robson KJ, Graham-Campbell J, Arnaiz-Villena A, Brissot P, Worwood M: **The origin and spread of the HFE-C282Y haemochromatosis mutation.** *Hum Genet* 2004, **115**:269-279.
76. Lao O, Andres AM, Mateu E, Bertranpetit J, Calafell F: **Spatial patterns of cystic fibrosis mutation spectra in European populations.** *Eur J Hum Genet* 2003, **11**:385-394.
77. Clemente FJ, Cardona A, Inchley CE, Peter BM, Jacobs G, Pagani L, Lawson DJ, Antao T, Vicente M, Mitt M, et al.: **A Selective Sweep on a Deleterious Mutation in CPT1A in Arctic Populations.** *Am J Hum Genet* 2014, **95**:584-589.
78. Henn BM, Botigue LR, Bustamante CD, Clark AG, Gravel S: **Estimating the mutation load in human genomes.** *Nat Rev Genet* 2015, **16**:333-343.
*** Comprehensive review of potential causes, measures and extent of mutation load in human populations. Underlines the difficulty of estimating mutation load without knowing exactly the distribution of selective and dominance coefficients. .**
79. Lohmueller KE, Indap AR, Schmidt S, Boyko AR, Hernandez RD, Hubisz MJ, Sninsky JJ, White TJ, Sunyaev SR, Nielsen R: **Proportionally more deleterious genetic variation in European than in African populations.** *Nature* 2008, **451**:994-997.
80. Do R, Balick D, Li H, Adzhubei I, Sunyaev S, Reich D: **No evidence that selection has been less effective at removing deleterious mutations in Europeans than in Africans.** *Nature genetics* 2015, **47**:126-131.
81. Simons YB, Turchin MC, Pritchard JK, Sella G: **The deleterious mutation load is insensitive to recent population history.** *Nature genetics* 2014, **46**:220.
82. Lohmueller KE: **The distribution of deleterious genetic variation in human populations.** *Current opinion in genetics & development* 2014, **29**:139-146.
83. Sella G: **this issue.** *Curr Opin Genet Dev* 2016.
84. Fu W, Gittelman RM, Bamshad MJ, Akey JM: **Characteristics of neutral and deleterious protein-coding variation among individuals and populations.** *The American Journal of Human Genetics* 2014, **95**:421-436.

85. Agrawal AF, Whitlock MC: **Inferences about the distribution of dominance drawn from yeast gene knockout data.** *Genetics* 2011, **187**:553-566.
86. Shendure J, Akey JM: **The origins, determinants, and consequences of human mutations.** *Science* 2015, **349**:1478-1483.