The role of standing variation in geographic convergent adaptation

Peter L. Ralph*,1
email: pralph@usc.edu
and
Graham Coop*,2
email: gmcoop@ucdavis.edu

April 21, 2015

* These authors contributed equally to this work.
[1] Computational Biology and Bioinformatics
University of Southern California, Los Angeles, CA
[2] Center for Population Biology & Department of Evolution and Ecology
University of California – Davis, Davis, CA, 95616

Abstract

The extent to which populations experiencing shared selective pressures adapt through a shared genetic response is relevant to many questions in evolutionary biology. In a number of well studied traits and species, it appears that convergent evolution within species is common. In this paper, we explore how standing, genetic variation contributes to convergent genetic responses in a geographically spread population, extending our previous work on the topic. Geographically limited dispersal slows the spread of each selected allele, hence allowing other alleles – newly arisen mutants or present as standing variation – to spread before any one comes to dominate the population. When such alleles meet, their progress is substantially slowed – if the alleles are selectively equivalent, they mix slowly, dividing the species range into a random tessellation, which can be well understood by analogy to a Poisson process model of crystallization. In this framework, we derive the geographic scale over which a typical allele is expected to dominate, the time it takes the species to adapt as a whole, and the proportion of adaptive alleles that arise from standing variation. Finally, we explore how negative pleiotropic effects of alleles before an environment change can bias the subset of alleles that contribute to the species' adaptive response. We apply the results to the many geographically localized G6PD deficiency alleles thought to confer resistance to malaria, where the large mutational target size makes it a likely candidate for adaptation from standing variation, despite the selective cost of G6PD deficiency alleles in the absence of malaria. We find the numbers and geographic spread of these alleles matches our predictions reasonably well, consistent with the view that they arose from a combination of standing variation and new mutations since the advent of malaria. Our results suggest that much of adaptation may be geographically local even when selection pressures are homogeneous. Therefore, we argue that caution must be exercised when arguing that strongly geographically restricted alleles are necessarily the outcome of local adaptation. We close by discussing the implications of these results for ideas of species coherence and the nature of divergence between species.

1 Introduction

There are an increasing number of examples where different populations within a species have adapted to similar environments by means of independent genetic changes. In some cases this convergent evolution

is the result of quite distinct genetic changes, involving very different genes and pathways, despite shared selection pressures; in other cases independent adaptations are identical down the same nucleotide change (Jeong & Rienzo, 2014; Stern, 2013; Martin & Orgogozo, 2013; Conte et al., 2012). Such convergent evolution within populations has been seen for many carefully studied phenotypes across a range of species, including drug resistance in pathogens, resistance to pathogens or pesticides, and the molecular basis of pigmentation changes. The phrase "parallel evolution" is also used to refer to such convergent evolution; here we use these synonymously, as we are concerned with adaptation within a single species that can occur via a different or shared genetic routes (see Arendt & Reznick, 2008, for more discussion).

The issue of convergent adaptation within species touches on a number of important questions in evolutionary biology. These include the extent to which adaptation is shaped by pleiotropic constraints (Haldane, 1932; Orr, 2005), whether adaptation is mutation-limited (Bradshaw, 1991; Karasov et al., 2010), and to what degree species should be regarded as cohesive units. Convergent evolution also affects our ability to detect adaptation from population genomic data, since no single allele sweeps to fixation over the entire area affected by the selection pressure (Pennings & Hermisson, 2006b).

Convergent evolution can occur even within a well mixed population subject to a constant selection pressure, either through selection on multiple mutations present as standing variation within the population before selection pressures switch (Orr & Betancourt, 2001; Hermisson & Pennings, 2005), or due to multiple adaptive alleles that arise after selection pressures switch (Pennings & Hermisson, 2006a). Previous work has shown that a primary determinant of the probability that multiple alleles contribute to adaptation is the product of the population size and the mutation rate (see Messer & Petrov, 2013, for a review).

Spatial population structure, as caused for example by geographically limited dispersal, also increases the chance of convergent evolution. For example, geographically patchy selection pressures can lead to much higher probability of parallel adaptation than uniform pressures, since alleles are unable to spread through intervening populations (Ralph & Coop, 2014). In Ralph & Coop (2010) we formulated a simple model of convergent adaptation in a spatially spread population with local dispersal that is exposed to some novel, spatially homogeneous selection pressure. We assumed that a single mutational change was sufficient to adapt the population after the change in environment. Under this assumption, selected alleles arise and spread locally, as shown in Figure 1. If the geographic area is large enough multiple selected alleles can arise independently and spread before any one has spread across all of space. A somewhat analogous situation also arises in spatial models of clonal inference in asexuals (Gordo & Campos, 2006; Martens & Hallatschek, 2011; Otwinowski & Krug, 2014).

Under this model we previously derived the characteristic geographic scale over which multiple instances of the adaptive allele are expected to arise in parallel, a characteristic length expressed in terms the parameters of interest. In Ralph & Coop (2010) we assumed that there was no standing variation for the adaptive allele (e.g., because the allele was very strongly deleterious before the environmental change), so that parallel mutation must be due to multiple new mutations occurring after the environmental shift.

In this paper, we extend this spatial model to include standing variation present at mutation-selection balance before the selection pressures switch. Below, we show that convergent adaptation within a widespread species is likely to be common when ranges, population sizes, or mutational targets are large, as has already been seen for a number of traits. On this basis we argue that the genetics of adaptation may often be geographically local even when selection pressures are geographically broad, and that widespread selective sweeps should tend to occur only when adaptation is highly constrained (e.g. by small mutation rate or the need for a linked combination of alleles). We discuss the history, and implications, of this view for the evolutionary coherence of species and molecular evolution.

1.1 Model description

We assume that the species range is a large, homogeneous, one– or two–dimensional region. There are two selective classes – the *ancestral* type, and the *mutated* type (which will offer a fitness benefit after the environmental switch). We assume that separately arising mutations are distinguishable – either as selectively equivalent mutations, or by linked neutral variation. We assume that all alleles of the mutated type are selectively equivalent (both in terms how deleterious they were before, and how advantageous they

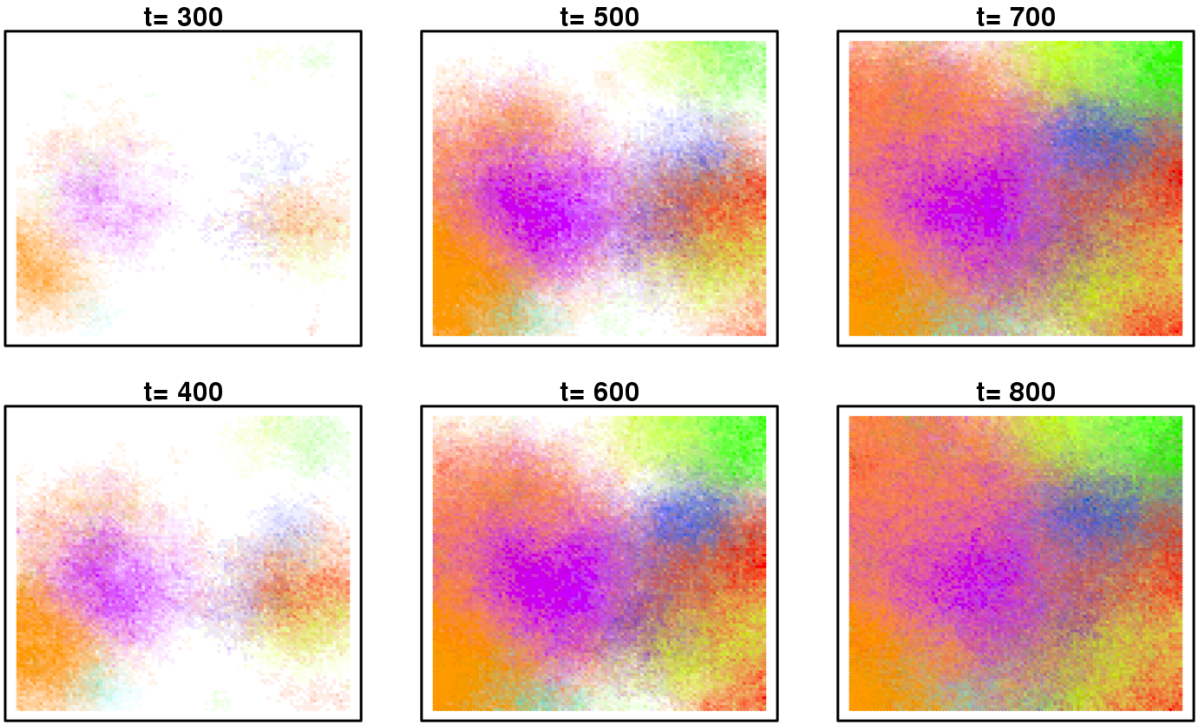

Figure 1: **Simulation of the geographic spread of new mutations**. Simulation on a grid of 101×101 populations with 100 haploid individuals in each, showing spread of a beneficial allele with advantage $s = 0.1$. Times are in numbers of generations, with probability of mutation per generation 0.3. Different colours show the different independent origins and spreads of selected alleles. Alleles quick establish and start to spread, but because they are selective equivalent once they spread into each other they only mix slowly (note the lack of change from generation 600 onward). More details of the simulation are given in Ralph & Coop (2010).

are after, the environmental shift). Note that these mutations do not have to arise at the same locus, just that they are selectively equivalent: these could be mutations arising at the same base pair, or knockout mutations of any one of a number of genes in a pathway, as long as carrying at least one of these alleles is sufficient to adapt an individual to the new environment.

We also suppose that the mutated type has been at a selective disadvantage for a sufficiently long enough time in the past to be at selection-mutation equilibrium, but at a certain time the selective regime changes, so that the mutated type has a selective advantage and quickly spreads to fixation. After fixation, alleles are either descended from families of mutants present as standing variation when the selective regime changed, or from new mutants arising since that time.

For concreteness, suppose that before time $t = 0$, the mutant type has fitness $1 - s_d$ relative to the neutral type (i.e. it produces on average $1 - s_d$ times the number of offspring per generation), and that after time $t = 0$, the mutant type has fitness $1 + s_b$, where $s_b > 0$ will usually be assumed to be small, and $0 < s_d < 1$. We assume that diploid fitness is additive, or at least that the important early dynamics are determined by the heterozygous fitness, with no reference to the fitness of the homozygote for the mutant alleles. Note that this means that we are ignoring the case of totally recessive alleles. We assume that the variance in offspring number is one, see Ralph & Coop (2010) for how this assumption can be relaxed. As for the other parameters, suppose that each offspring of a neutral parent is of the mutant type with probability μ, and

that the mean squared geographic distance between parent and child is σ^2. We assume that the dispersal kernel is not heavy tailed, i.e., falling into the Gaussian domain of attraction, see Ralph & Coop (2010) for discussion of heavy tailed kernels and Hallatschek & Fisher (2014) for recent progress on rapid spreading due to very long distance dispersal. The species occupies an area with mean density ρ of alleles per unit area (twice the number of diploid individuals per unit area).

Rates of origination of standing and new mutations We make use of the commonly used approximation that neglects competition between close relatives, treating the offspring of a new mutant that appears in an area not already occupied by the mutated type as a branching process in continuous time, measured in generations (average parental age at birth). After $t = 0$, the offspring of each individual mutant thus forms (approximately) a branching process with growth rate s_b, so each new mutant establishes locally with probability $p_s \approx 2s_b$. As in Ralph & Coop (2010), each new offspring has a very small probability of being a mutant and establishing locally, so the collections of times and locations at which mutants appear and establish locally is well–approximated by a Poisson process in space and time. By this we mean that there is a constant rate across space and time at which new mutations arise, and the occurrence of new mutations in any region of space-time is independent of that in non-overlapping intervals in which the mutated type has not already appeared. The rate of this Poisson process, i.e. the mean number of new mutants per unit area and per generation that appear and establish locally after $t = 0$ in areas not already occupied by the mutant type, is the product of the haploid population density, denoted ρ, the mutation rate (μ), and the probability of fixation p_s, or approximately

$$\lambda = 2s_b\mu\rho \tag{1}$$

In geographic areas already occupied by the mutant type, new mutations are effectively neutral, and so unlikely to establish over the short-time scales considered here, and are hence excluded.

Before $t = 0$, on the other hand, the allele is deleterious. The genetic descendants of each new mutation are (with high probability) doomed to extinction, but may persist for a short time (note that we have assumed that s_d is not too small). The times and locations of new mutations before $t = 0$ also will be well–approximated by a Poisson process with rate $\mu\rho$ since mutations are rare events. Therefore, the locations of all mutant families extant at $t = 0$ whose descendants are destined to fix locally is a thinning of the original process, and so, by the Poisson Mapping Theorem, is also a Poisson process. We define λ_0 to be the mean density of this process, i.e., the geographic density of standing variants that are present and escape loss when the environment shifts. If we assume that the descendants of at most only a few members of any extant mutant family at $t = 0$ will survive, and that these progenitors are near to each other in space, we can then treat each such family as equivalent to a single new mutation, but with somewhat larger probability of local establishment, following the environmental shift. (This approximation will be good if the logarithm of the size of each extant family is small relative to the establishment time, since then new families quickly "catch up" to the size of already-extant families, and the spatial distribution of each is small relative to the spread between the families.) To find λ_0, consider a mutation that arose T generations ago, and let Z_T be the number of its descendants at time 0. At time $t = 0$, when the environment shifts, there are Z_T individuals present with the mutation, and each has probability p_s of establishing, approximately independently. Therefore, the probability that at least one descendant of this mutation establish and fix locally is $1 - (1 - p_s)^{Z_T}$, the mean number of clusters of standing variants destined to fix locally, per unit area, is

$$\mu\rho \int_0^\infty \mathbb{E}\left[1 - (1 - p_s)^{Z_T}\right] dT.$$

For s_b small, using $p_s \approx 2s_b$ and that $\mathbb{E}[Z_T] = (1 - s_d)^T$, we know that $\mathbb{E}[(1 - 2s_b)^{Z_T}] \approx 1 - 2s_b\mathbb{E}[Z_T] =$

$1 - 2s_b(1 - s_d)^T$, resulting in the approximation

$$\lambda_0 = \mu\rho \int_0^\infty 2s_b(1 - s_d)^T dT \tag{2}$$

$$= \frac{2\mu\rho s_b}{-\log(1 - s_d)}. \tag{3}$$

$$= \frac{\lambda}{(-\log(1 - s_d))}. \tag{4}$$

Note that for small s_d this can also be found by taking the expected frequency under mutation-selection balance μ/s_d (Haldane, 1927, 1937) multiplying it by the population density to obtain the expected number of chromosomes per unit area carrying the deleterious allele, with each of these having a probability $2s_b$ of escaping loss (an analogous approach to that taken by Orr & Betancourt, 2001) .

Geographic spread of alleles Once an allele has become locally established it can begin to spread across space. We assume that the allele, once established, quickly settles down to spread spatially as a traveling wave of constant speed. The behavior of this wave of advance of a beneficial allele was first described by Fisher (1937) and Kolmogorov, Petrovskii & Piscunov (1937). Under reasonably general conditions, the speed of advance of this wave is $v = \sigma\sqrt{2s}$. See Ralph & Coop (2010) for a more thorough review of these travelling waves. Note that the speed of the wave will vary with details of the space that individuals migrate across (e.g. see Slatkin, 1976; Slatkin & Charlesworth, 1978, for comparisons to migration on discrete grids).

Putting it together. Now, we can put these ingredients together for a simple model of the geographic spread of alleles, a cartoon example of which is shown in Figure 2. Initially, when the selection pressures change at $t = 0$, a set of standing variants can start to spread having escaped loss through drift. The originating mutations of these variants are depicted by lightning bolts, and occur at a density λ_0 across space. They spread at velocity v, carving out cones in space-time. As these alleles proceed in their geographic spread, other new alleles can arise and become established in parallel, whose origins are indicated by stars. These new mutations arise and become established at rate λ.

As we outlined in Ralph & Coop (2010), this model of geographic convergent evolution, when $\lambda_0 = 0$, is analogous to a model of crystallization due to Kolmogorov (1937). In this model, nucleation sites form at random at a constant rate in time and space and initiate the radial growth of new crystals. After their initial spread, the different orientations of crystals form a random tessellation of space, whose properties have been studied by Møller (1992, 1995) and others (Bollobás & Riordan, 2008; Gilbert, 1962). The generalized version of this process, for non-constant wave speeds and inhomogeneous Poisson processes is known as the Kolmogorov–Johnson–Mehl–Avrami tessellation (Fanfoni & Tomellini, 1998). Our combined process with both standing variation and new mutation is a special case of the KJMA tessellation, where the spatial-temporally homogeneous Poisson origination process of new mutations, is supplemented by a single pulse of origination points at time zero with spatial density λ_0. (For the purposes of analogy, we could imagine that before time $t = 0$ the temperature is high enough that nucleation sites appear but do not persist long.)

If we ignore the effects of new mutation, then everything about the process is relatively simple: each point in space will be first reached by the wave whose origination point lies closest to it. This random tessellation of space is known as a Poisson-Voronoi tessellation (Møller, 1994) (i.e. the cells formed by assigning regions of space to the nearest point in a Poisson process). The properties of this tessellation by alleles is determined by the spatial locations of the initiation points, which are sampled from a spatially homogeneous Poisson process, independent of the rate of spatial spread. Introducing new mutations cause some qualitative changes beyond dependence on new parameters: the cells formed by a Voronoi tessellation have straight sides, but the introduction of new mutations cause these to curve (because the radii of the colliding circles differ; see Figure 1 of Ralph & Coop, 2010, for a graphical depiction of this point).

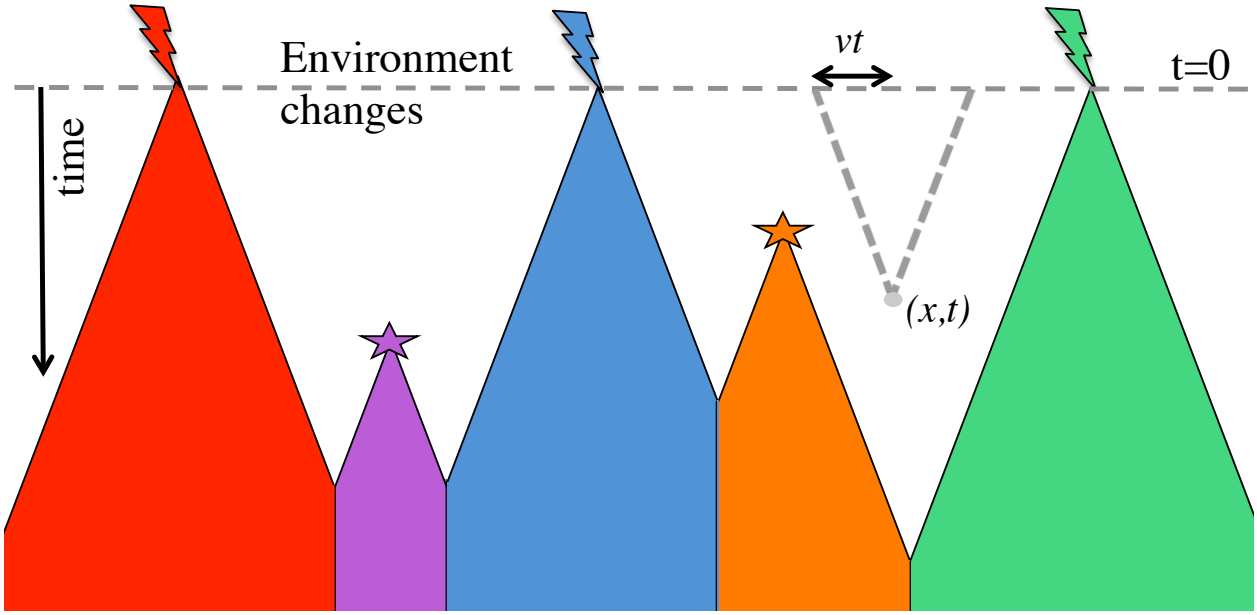

Figure 2: **Cartoon space-time diagram of the geographic spread of standing and new mutations**. Time runs down the page, and a single spatial dimension runs across the page. The environment switches at $t = 0$, following which adaptive alleles that successfully escape loss by drift spread locally from both standing variants (lightning bolts) and new mutations (stars). (Note that the standing variants arose at times prior to $t = 0$.) The mean spatial density of successful standing variants at $t = 0$ is λ_0, and the mean density in (space $\times$ time) of successful new mutations is λ. These successful alleles spread spatially outwards at speed v, so the space-time profile of their spread forms a cone. Any mutation that arises in an area already within a cone will not have a selective advantage. When these adaptive alleles meet, their rapid advance is halted because they no longer have an advantage, so their boundaries form straight lines on short time-scales (over longer time-scales, they will mix into each other). One of the ways we derive various results in the text is by taking a space-time point (x, t) and asking the probability that no successful alleles have spread there yet. The grey dashed line shows a cone radiating backward in space-time, originating at (x, t), with slope v; its radius at $t = 0$ is vt. This shows the location x is not yet adapted at time t because no successful alleles appear within this grey cone.

1.2 G6PD example

Below we describe a number of properties of our process, but we will first introduce a motivating example to provide some concrete numbers to use for illustrative purposes.

Over roughly the past ten thousand years, alleles conferring resistance to malaria have arisen in a number of genes and spread through human populations in areas where malaria has been endemic (Kwiatkowski, 2005). A number of these alleles appear to be examples of convergent adaptation, as different derived mutations in the same gene are seen in different individuals. For example, a number of changes that confer malaria resistance have been observed in the β-globin gene; and the sickle cell allele may plausibly have arisen by up to five independent occurrences of the same base pair mutation at different locations within Africa (Flint et al., 1998; Ralph & Coop, 2010). Another particularly impressive case of convergent evolution is presented by the numerous changes throughout the X-linked G6PD gene, with upward of 50 polymorphic variants (above 1% local frequency) having so far been described that lower the activity of the enzyme (Howes et al., 2013; Minucci et al., 2012). These alleles are now found at a combined frequency of around 8% frequency in malaria endemic areas, rarely exceeding 20% (Howes et al., 2012). Whether these all confer resistance to

malaria is unknown, but malaria is thought to be the primary driver of these polymorphisms (see Hedrick, 2011, for a general review). Three G6PD deficiency alleles are particularly common and relatively well studied: the $A-$ allele found in much of sub-Saharan Africa; the Med allele found in the Mediterranean and Middle East; and the Mahidol allele found Myanmar and Thailand. The protective effects of these G6PD alleles are complicated, and are likely heterogeneous across study populations and the form of malaria considered (see Manjurano et al., 2015; Malaria Genomic Epidemiology Network, 2014, for recent discussion). The $A-$ and Mahidol alleles are thought to offer some protective effects against *Plasmodium falciparum* and *P. vivax* in both heterozygote females and/or hemizygote males (Ruwende et al., 1995; Louicharoen et al., 2009; Manjurano et al., 2015). Haplotype-based analysis of genetic diversity surrounding $A-$, Med, and Mahidol suggest that they have spread over the past few thousand years (Tishkoff et al., 2001; Slatkin, 2008; Saunders et al., 2005; Louicharoen et al., 2009), consistent with the age of other known malaria resistance alleles. Population genetic analyses suggest that these three variants each have a hemizygote/heterozygote selection coefficient of $0.05 - 0.3$ (Tishkoff et al., 2001; Slatkin, 2008; Saunders et al., 2005; Louicharoen et al., 2009). This is in reasonable agreement with the selection coefficients calculated by Ruwende et al. (1995) on the basis of the present day levels of resistance to malaria due to the $A-$ allele.

Given such strong selection the alleles should have risen quickly to fixation, so their presence at intermediate frequency, over a broad geographic area, makes it a good candidate for a recently balanced polymorphism due to heterozygote advantage (note that the conditions for a balanced polymorphism are complicated by the hemizygosity of males, see Hedrick, 2011; Pamillo, 1979). Indeed, hemizygous males and homozygous females suffer from G6PD deficiency, and homozygote females may also not be protected against malaria (Manjurano et al., 2015; Malaria Genomic Epidemiology Network, 2014). The theory we use regarding the "wave of advance" (Fisher, 1937) applies as well in the case of heterozygote advantage (Aronson & Weinberger, 1975), with the selected allele spreading locally to the equilibrium frequency (rather than fixation). Therefore, our framework is applicable to the spread of G6PD, with speed determined by the advantage of heterozygotes when rare. We assume that before malaria became prevalent, G6PD deficiency alleles suffered a decrease in relative fitness of s_d in heterozygote and homozygotes females and hemizygote males. Assuming that the underlying causes and strength of this drop in fitness have not changed, we estimate that s_d has to have been upward of ~ 0.05 (if $s_b \geq 0.05$), in order to have resulted in the equilibrium frequency seen today in areas with endemic malaria (based on heterozygote advantage calculations for the X chromosome, results not shown, see also Ruwende et al., 1995).

The geographic area of Central and Eastern Asia with malaria is on the order of ten million square kilometers. In that area there are at least 15 common, clinically relevant variants (see Figure 3, from Howes et al., 2013). (These are type 2 variants that express at $< 50\%$ enzyme activity, predispose individuals to haemolytic anaemia, and are found in at least 10 localities; see Howes et al. (2013) for more details.) Therefore, the average width of an area occupied by an allele is $\sqrt{10^7/15} \approx 800$km. The coding region of G6PD is 515 codons long, and around 140 distinct deficiency alleles have been observed. Assuming a mutation rate of $\approx 10^{-8}$ per base pair per generation, we can take as an order-of-magnitude estimate $\mu \approx 10^{-6}$ per generation. The dispersal and demographic parameters of humans in the past few thousand years is unclear, particularly as we are concerned with the "effective" population density (i.e. population density divided by variance in offspring number). We therefore will use two reasonable values for the effective population density: $\rho = 2$ and 0.2 people per km^2, and three values for the dispersal distance: $\sigma = 10$, 50 and 100 kilometers per generation. Clearly, human migration has been shaped both by local dispersal and larger-scale expansions (see Pickrell & Reich, 2014, for a recent discussion), so these parameters only provide a rough view of the process.

1.3 The geographic resolution of adaptation from new and standing variation

In Ralph & Coop (2010), studying the model without standing variation, we defined a *characteristic length* which gave the spatial scale across which mutants with distinct origins would establish. This was proportional to the mean distance between neighboring established mutants, but had the advantage of being easier to calculate. Furthermore, the time scale over which adaptation occurred could be found by dividing the

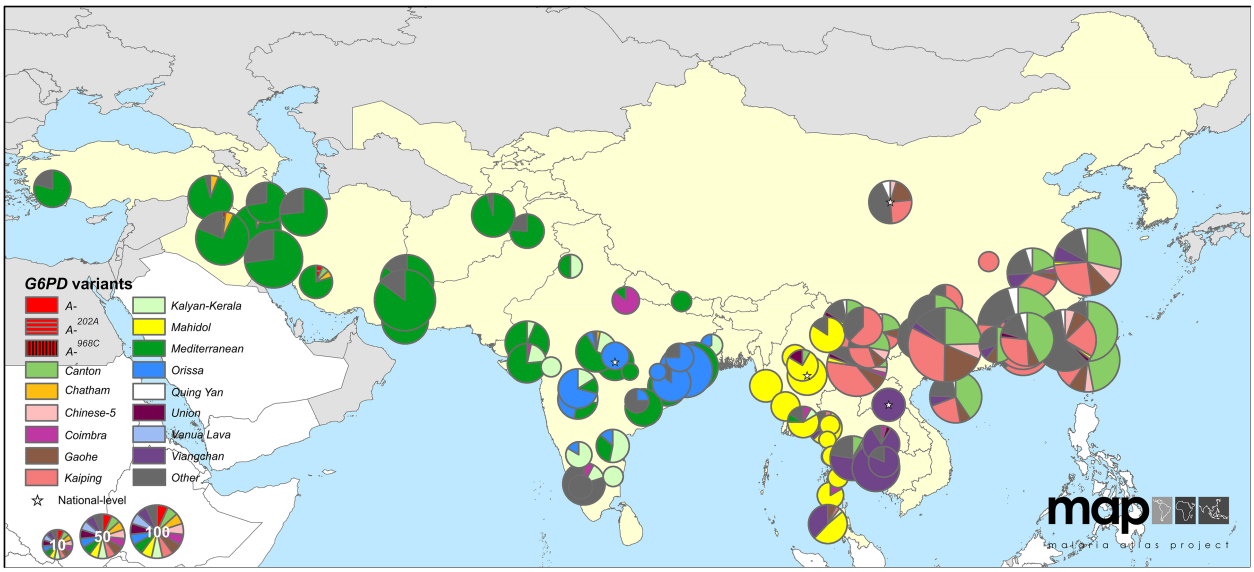

Figure 3: **Map of G6PD-deficiency allele frequencies across Asia.** The pie chart shows the frequency of G6PD-deficiency alleles. The size of the pie chart indicates the number of G6PD-deficient individuals sampled. Countries with endemic malaria are colored yellow. Figure taken from Howes et al. (2013) `http://www.malariajournal.com/content/12/1/418`.

characteristic length by the speed at which the mutants spread. We first define a similar characteristic length for this new model.

Suppose we fix our attention on a particular new mutation that happens to be the first to occur in some region. If it does not encounter other locally fixed, beneficial mutants, it will cover a distance L in time L/v. In doing so (in two-dimensions) it will have started from a point in space and spread out to cover a circular geographic region of area πL^2. The cylinder in space-time with this circle as a base and height L/v has volume of this cylinder is $\pi L^2 \times (L/v)$. Therefore, the number of other successfully established mutations that would have appeared in the circle it has covered up until this time is Poisson with mean $\lambda_0 \pi L^2 + \lambda \pi L^3/v$ in two dimensions (and $2\lambda_0 L + 2\lambda L^2/v$ in one dimension). Therefore, if we define χ for a two-dimensional model to be the unique positive solution to

$$\lambda_0 \pi \chi^2 + \lambda \pi \chi^3 / v = 1, \tag{5}$$

then χ gives the distance spread unobstructed by the descendants of a new mutant before it is expected that one other successful mutation would have arisen in the area covered so far. The explicit formula for χ in two dimensions can be found by rearranging eqn. (5) but is cumbersome; here we omit it. In one dimension things are a little prettier and the characteristic length is $\chi = (\sqrt{\lambda_0^2 + 2\lambda/v} - \lambda_0)/(2\lambda/v)$. Substituting expressions for λ_0, λ, and v from above, in two-dimensions we can rewrite eqn. (5) as

$$\frac{\sqrt{2s_b}}{-\log(1-s_d)} \chi^2 + \frac{1}{\sigma} \chi^3 = \frac{1}{\sqrt{2s_b} \rho \mu \pi} \tag{6}$$

From this we see that χ decreases with ρ and μ. Furthermore, for large σ, the characteristic length approaches the value we would obtain just from standing variation:

$$\chi = \sqrt{-\log(1-s_d)/2s_b \rho \mu \pi} + O(1/\sigma). \tag{7}$$

On the other hand, if the mutant allele is highly deleterious before $t = 0$, then standing variation is unim-

portant the characteristic length is approaches the value from Ralph & Coop (2010):

$$\chi = (\sigma/\rho\mu\pi\sqrt{2s_b})^{1/3} + O(1/\log(1-s_d)). \tag{8}$$

These two end points help build our intuition for the interaction of parameters in shaping the geographic scale of convergent evolution. By the above calculation, we know that the relevant mutations occur about distance χ apart, and occur within the first χ/v generations. Said another way, if we look in a circular region of space of radius χ over χ/v generations, we expect to find roughly one mutational origin.

In Figures 4 and 5, we show the range of characteristic lengths as a function of various parameters chosen to match the evolution of malaria resistance at G6PD. These curves match our intuition that higher population densities result in smaller characteristic lengths (as would higher mutation rates). Allowing standing variation increases the input of new alleles and so decreases the characteristic length below that predicted by only new mutations (i.e. equation (8), Ralph & Coop (2010)). In turn, increasing the prior deleterious effect of the allele (s_d) acts to increase the characteristic length until it reaches that predicted by new mutation alone. Higher dispersal distances lead to larger characteristic lengths, since more rapid geographic spread block other mutations from establishing. A larger selective advantage (s_b) acts in two conflicting ways: aiding the rapid geographic spread of established alleles, and also helping more independent copies to escape drift and become established. The effect of helpinglocally establish alleles wins out, since increasing the selective benefit s_b decreases the characteristic length. (This can be shown in general by differentiating (6) with respect to $z = \sqrt{s_b}$, showing that $\partial_z \chi = -(C_1 z\chi + C_2\chi^2)/(C_3 z^2 + C_4 z\chi) \leq 0$ for appropriate nonnegative constants C_{1-4}.) This effect is strongest when only standing variation contributes ($s_b^{-1/2}$, equation (7)), as in that case the speed of spread does not matter only the initial density of established alleles after the environmental shift. The dependence of the characteristic length on s_b when only new mutations contribute is much weaker (of order $s_b^{-1/6}$, equation (8)) Overall, the range of characteristic lengths observed are reasonably consistent with the average diameter of a G6PD variant in Eurasia of 800km, especially for the lower population density, as long as the fitness cost of G6PD-deficiency alleles before malaria (s_d) was high.

Finally, while the form of eqn. (6), specifying the characteristic length, is not particularly intuitive we can use it to ask when we should expect multiple adaptative alleles within a large geographic region where our selective pressure is present (thanks to Sam Yeaman for proposing this interpretation). Consider the case where this geographic area is a fairly regular shape of area G and diameter $\sqrt{G}$. Denote the total effective population size over this area by $N = \rho G$, and the standard deviation of dispersal as a fraction of the diameter of this area by $\sigma_G = \sigma/\sqrt{G}$. Measuring distance in units of the diameter of $\sqrt{G}$, we expect multiple mutations when $\chi < 1$ a condition which will be met when

$$N > \frac{\sigma_G}{\mu\pi\sqrt{2s_b}\left(1 - \sigma_G\sqrt{2s_b}/\log(1-s_d)\right)}, \tag{9}$$

this is found by rearranging eqn. (6) having set $\chi = 1$. This nicely shows that σ_G/μ is a primary determinant of the critical population density necessary for convergent evolution, when σ_G is small. As $\sigma_G s_b/s_d$ becomes large, e.g. because the allele is not too deleterious before the environmental shift, we get our standing variation only case where the dependence of σ_G drops out leaving the critical population size as $1/(2\mu\pi s_b/s_d)$

Note that the conflicting roles of ρ and σ mean that even in species where levels of neutral differentiation are low, geographic convergent adaptation may be common. This is because low levels of neutral genetic differentiation between geographic regions can be due to high population densities rather than high dispersal distances, and high population densities would allow convergent adaptation. As such geographic convergent evolution may be common even in species with little neutral population structure.

1.4 Time to adaptation

It is also straightforward to compute the mean time until adaptation. Imagine a geographic location, and let $\tau \geq 0$ be the time at which this location is first reached by some advantageous mutation. Then, as can be seen from the perspective of the grey dot in Figure 2, $\tau > t$ if and only if the cone with point at (x, t) and

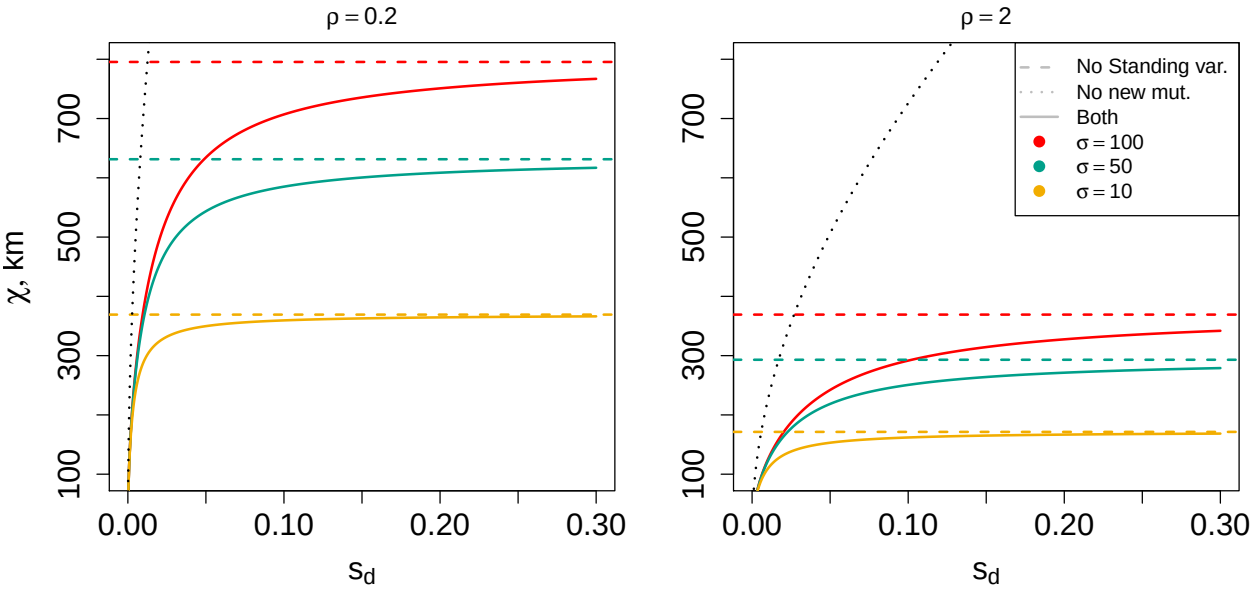

Figure 4: **Characteristic length, in kilometers, as a function of selective disadvantage**, compared to the corresponding quantity without standing variation (equation (8)) and to the quantity only considering standing variants (equation (7)). The other parameters are chosen to match those of G6PD, $\mu = 10^{-6}$ and $s_b = 0.05$. As the characteristic length with only standing variation is independent of dispersal distance it is plotted as a single black dotted line.

slope v extending back to $t = 0$ (light grey dashed lines) is empty of successful mutations. Since we assume that successful alleles arise as a Poisson process, in two dimensions

$$\mathbb{P}\{\tau > t\} = \exp\left(-\lambda_0 \pi v^2 t^2 - \lambda \pi v^2 t^3/3\right), \tag{10}$$

i.e., the combined probability that area of a circle of radius vt surrounding our point at $t = 0$ was free of successful standing variants, and no successful, new mutations arose in the cone (that has radius $r = vt$ at $t = 0$ and height $h = t$, and hence volume $\pi r^2 h/3$). Since $\mathbb{E}[\tau] = \int_0^\infty \mathbb{P}\{\tau > t\} dt$,

$$\mathbb{E}[\tau] = \int_0^\infty \exp\left(-\lambda_0 \pi v^2 t^2 - \lambda \pi v^2 t^3/3\right) dt. \tag{11}$$

For applications we evaluate this integral numerically.

In Figure 6 we show the mean time until adaptation for various values of the parameters chosen to match the case of adaptation at G6PD. Increasing σ and decreasing s_d lower the time to adaptation, as alleles spread geographically more quickly and are present as standing variation more often respectively. Increasing s_b strongly decreases the time to adaptation, as it both causes more alleles to escape drift and to rapidly spread. Given that the G6PD alleles likely spread over a few thousand years, i.e. less than a few hundred generations, this time scale seems quite plausible, except perhaps for the lowest dispersal distances.

1.5 The contribution of standing variation.

We can also ask in our framework to address what proportion of new adaptive variants arise from standing variation. We have defined λ_0 to be the mean density of standing variants that are present and escape loss when the environment shifts. We will define γ to be the mean density of newly arising alleles that spread having arisen in an area free of other adaptive alleles. Since the probability that a mutant arising at location

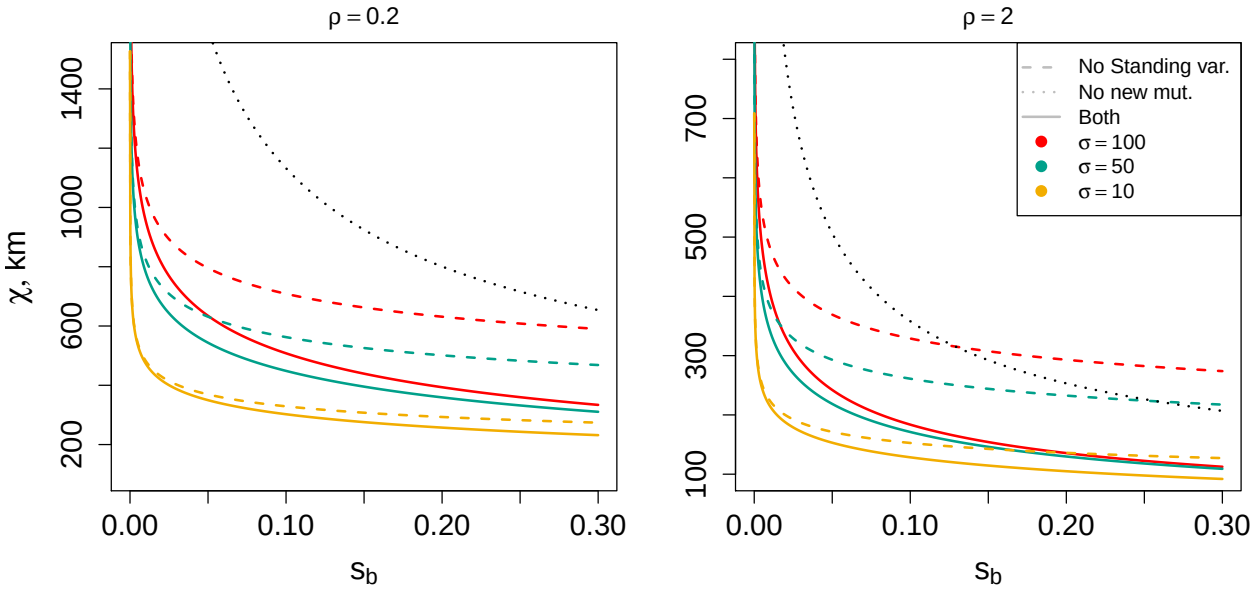

Figure 5: **Characteristic length**, in kilometers, as a function of selective advantage, at two population densities, holding the prior disadvantage of the allele fixed at $s_d = 0.05$ and the mutation rate fixed at $\mu = 10^{-6}$. We compare this to the corresponding quantity without standing variation (equation (8)) and to the quantity only considering standing variants (equation (7)).

x and time t is lucky enough to be born in a location not already occupied by mutants is $\mathbb{P}\{\tau > t\}$, we can see $\gamma = \int_0^\infty \lambda \mathbb{P}\{\tau > t\} dt$, and hence $\gamma = \lambda \mathbb{E}[\tau]$. Therefore, the mean proportion of adapted patches that come from standing variation is

$$\lambda_0/(\lambda_0 + \gamma) = \frac{1}{1 - \log(1/(1 - s_d))\mathbb{E}[\tau]}, \tag{12}$$

using the fact that $\lambda_0 = \lambda/\log(1/(1 - s_d))$. There are $\lambda_0 + \gamma$ patches per unit area, so the typical patch (informally, the patch around a randomly chosen successful mutation; formally, drawn from the Palm measure (Cox & Isham, 1980)) occupies area $1/(\lambda_0 + \gamma)$.

We can also find the mean proportion of space covered by standing variants (we will restrict ourselves to two dimensions). At time t a geographic location has not yet been reached by the mutation with probability given by equation (10). Given that it has not been reached by t, the probability that it will be reached by time $t + dt$ by a standing variant is approximately $2\lambda_0 \pi v^2 t dt$, which is λ_0 multiplied by the thin slice of extra area in our expanded circle at $t = 0$, which has gone from a radius vt to $v(t + dt)$. The corresponding probability that the point is reached by a new variant is $\lambda \pi v^2 t^2 dt$, which is λ multiplied by the sliver of extra volume in our space-time cone at time $t + dt$ compared to that at time t. Therefore, the mean proportion of space covered by standing variants, is

$$z_0 = \int_0^\infty 2\lambda_0 \pi v^2 t \, \exp\left(-\lambda_0 \pi v^2 t^2 - \lambda \pi v^2 t^3/3\right) dt, \tag{13}$$

this is the probability a given location is reached first by a standing variant (which follows from competing our two exponential waiting times). For applications we evaluate this integral numerically.

Furthermore, if we define a_0 to be the mean area occupied by a typical standing variant, then a_0 is given by the proportion of the range occupied by standing variants divided by the mean density of unique standing

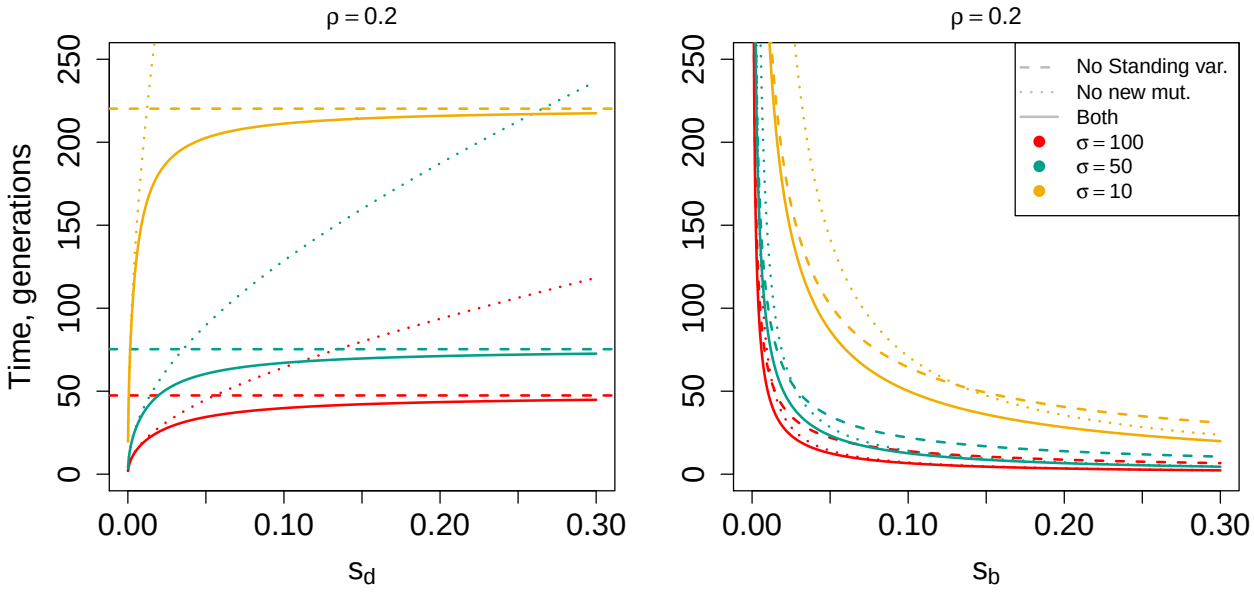

Figure 6: **Mean adaptation times** as a function of selective advantage and disadvantage. We compare these to the corresponding quantities without standing variation and only considering standing variants.

variants, i.e. $a_0 = z_0/\lambda_0$. We can solve for a_+, the corresponding mean area occupied by a given new variant, using the formula $a_0/a_+ = z_0/(1 - z_0)$.

In Figure 7 we show the proportion of alleles that spread from standing variation, and the proportion of geographic space covered by standing variants for parameters chosen to match our G6PD example. Even for relatively large deleterious costs prior to the environmental switch, standing variants still make up quite a large proportion of the adaptive alleles, and an even larger proportion of the range (they occupy a larger area than new mutations, since they get a head start).

1.6 Multiple variant types

Another problem that we can address with this work is the extent to which pleiotropy biases adaptation towards the repeated use of particular subset of loci (i.e. convergence at genetic level). While many alleles may confer the beneficial phenotype, not all will contribute equally to adaptation if they have negative pleiotropic consequences. There are at least two ways that negative pleiotropy can contribute to high rates of convergence when adapting if a single change is sufficient for adaptation. First, negative pleiotropic effects can reduce the overall beneficial selection coefficient of an allele in the new environment, making them unlikely to become established and slow to spread (and in the worst case making them deleterious). This first effect has been well studied by a number of authors (Orr, 2000; Otto, 2004; Welch & Waxman, 2003; Chevin et al., 2010) and its role in genetic convergence examined (Orr, 2005; Chevin et al., 2010; Unckless & Orr, 2009). A second contribution is that alleles that have less negative pleiotropy are more likely to be present as standing variation before the environmental shift, and so are more able to respond immediately.

Here we focus primarily on the second effect. Let's imagine for the moment that there are several classes of beneficial allele, all having the same beneficial selection coefficient (or at least that beneficial selection coefficients are similar enough that our selective exclusion approximation holds over the time scale on which we examine the process). Each class of mutations j has its own mutation rate μ_j and selective disadvantage $s_{d,j}$ prior to the environmental switch. As they have the same beneficial selection coefficient after the switch,

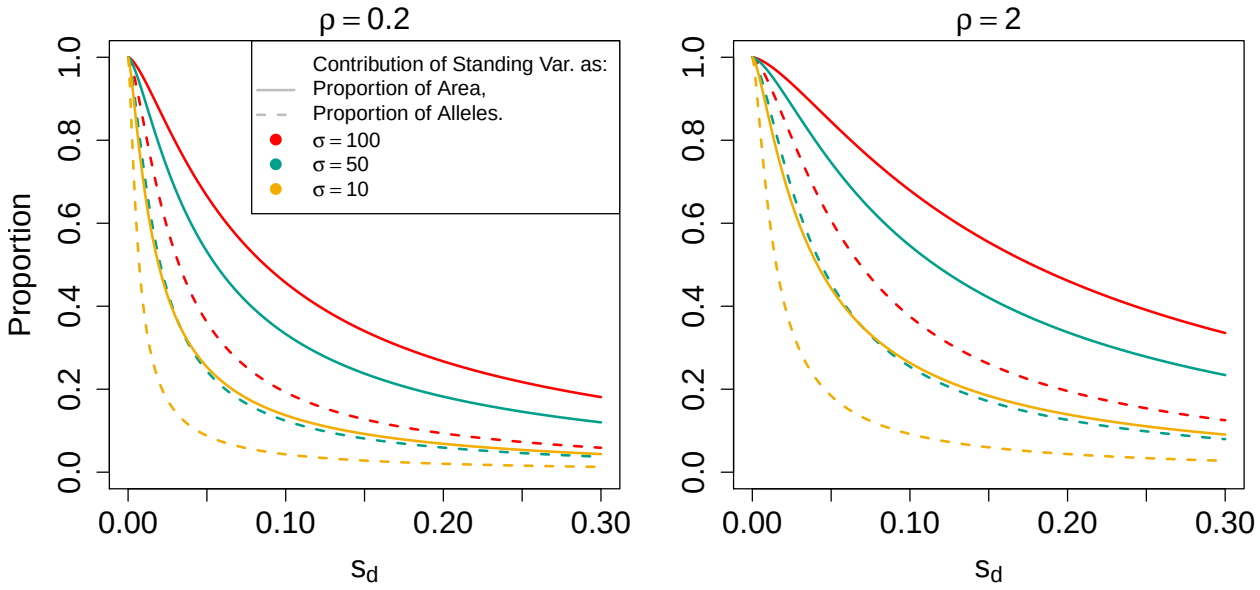

Figure 7: **Mean proportion of patches arising from standing variation,** as a function of selective disadvantage, at two population densities. The parameters given are for the G6PD example. Both panels show the expected proportion of patches that arise from standing variation (dotted lines, equation (12)), and the expected proportion of geographic space that is covered by adaptation from standing variation (solid lines, equation (13)). We set $\mu = 10^{-5}$ and $s_b = 0.05$ to match our G6PD example.

all of the waves travel outward at a rate v. Then, the density of type j standing variants per unit area and the input rate of *de novo* variants per unit area per generation are, respectively,

$$\lambda_{0,j} = \frac{2\mu_j \rho s_b}{-\log(1-s_{d,j})} \qquad \text{and} \qquad \lambda_j = 2\mu_j \rho s_b. \tag{14}$$

Using these rates, and an argument analogous to that used to derive equation (13), at the time when every location has been reached by an adaptive allele, the proportion of the species range covered by alleles of type j is

$$p_j = \int_0^\infty \left(2\lambda_{0,j}\pi v^2 t + \lambda_j \pi v^2 t^2\right) \exp\left(-\sum_k (\lambda_{0,k}\pi v^2 t^2 + \lambda_k \pi v^2 t^3/3)\right) dt. \tag{15}$$

If we only allow standing variation this collapses to

$$p_j = \frac{\mu_j/\log(1-s_{d,j})}{\sum_k \mu_k/\log(1-s_{d,k})}, \tag{16}$$

while if we only allow new variation, i.e. if all variants are highly deleterious before the environment switches, $p_j = \mu_j/(\sum_k \mu_k)$.

To illustrate some of the properties of this model, let's imagine the somewhat extreme scenario in which there is a single base pair at which a possible mutation is relatively free of negative pleiotropy (call this class 1); and a larger mutational target where changes have more serious pleiotropic consequences in the ancestral environment (class 2). We set $s_{d,1} \leq s_{d,2} = 0.05$ and $\mu_1 = 1 \times 10^{-8}$, assume that both classes share a beneficial selection coefficient of $s_b = 0.05$, and think of the second class of alleles as arising at one of ten, one hundred, or one thousand base pairs. We show the expected proportion of space covered by the rarer, class 1 mutations in Figure 8. As expected intuitively, the contribution of the rarer mutation

decreases as the mutational target of the second class becomes larger, and as the difference in the negative pleiotropic consequences of the two classes of alleles decreases. The case with standing variation only is the best case scenario for the rarer mutation, so its rate of introduction after $t = 0$ is necessarily lower. However, the standing-variation-only case does seem to provide a reasonable rule of thumb, especially for parameter combinations, such as high population densities and high dispersal distances, that increase the contribution of standing variation (and similarly for high s_b).

It is natural to also incorporate differences in the beneficial selection coefficients of the different classes of alleles, to allow for negative pleiotropic effects acting to suppress the advantage of an allele once the environment switches. One simple way to do this is to simply replace s_b with $s_{b,j}$ resulting in a class-specific new allele establishment rates (λ_j) and rates of spread (v_j). These then could be used in equation (15). For instance, we could extend the two class model above so that class one has the additional advantage of $s_{1,b} > s_{2,b}$. This would further increase the contribution of the rarer class, because this class would both overcome drift more often and spread more rapidly. However, a straightforward application of the logic of used to construct equation (15) fails once the different allelic types meet, since the assumption of selective exclusion no longer holds: alleles with higher s_b will spread, at a lower speed, into regions occupied by alleles with lower s_b. Given enough time, the most advantageous type (type 1, in this case) would spread everywhere, and so substituting multiple values for s_b in equation (15) would only provide a short-term approximation to a longer term dynamic. Even if the initial tessellation has formed with purely class 2 alleles, the first allele would have a selective advantage $\delta s = s_{b,1} - s_{b,2}$, and so would arise at rate $2\rho\mu_1\delta s$ and would spread at speed $\sigma\sqrt{\delta s}$. An extension of our Poisson process model could incorporate these effects, by thinning the Poisson process of establishing mutations by correctly, but is considerably less tractable. Whether allele 2 persists would depend on the linkage arrangements between loci. If the loci underlying allele 1 and 2 are unlinked, then allele 1 can spread without disrupting allele 2. However, if they are linked, the spread of allele 1 may push allele 2 out of the population. More complicated dynamics, including spatial Dobzhansky-Muller incompatibilities (Kondrashov, 2003; Ralph & Coop, 2010) could ensue if there are epistatic interactions between the alleles.

1.7 Local establishment and comparison to panmixia

In the above, we have assumed that once the mutation appears, conditional on eventual fixation, it begins to spread spatially at speed v instantly, effectively neglecting the time it must first spend escaping demographic stochasticity. In Ralph & Coop (2010) we addressed this by noting that there would be no change at all in our results if all mutations had to wait the same amount of time before fixing locally, and that this time was short relative to the time it took the wave to spread across the characteristic length; we then showed via simulation that this was reasonable in certain situations. In this section we examine this assumption in more detail, although mostly through heuristic arguments, and also compare the results above to the results without geographic structure of Pennings & Hermisson (2006a).

We are assuming that shortly after a new mutation appears, it can be approximated by a branching process growing at rate s_b until the point that it grows large enough to "feel" spatial structure, at which point it begins to spread as a more–or–less deterministic wave. Although we are not aware of good analysis of this transition, the relevant size of the branching process when spatial structure becomes important should be something close to $\sigma^2\rho$ (i.e. Wright's "local effective population size", Wright (1943)). Let Z_t be a continuous-time branching process with $Z_0 = 1$ and $\mathbb{E}[Z_t] = e^{s_b t}$. Then we know that there exists a random variable W such that $\lim_{t\to\infty} e^{-s_b t} Z_t = W$ almost surely, so that if τ is the time Z_t reaches size $\sigma^2\rho$, then $\sigma^2\rho = Z_\tau \approx e^{s_b \tau} W$ (Jagers, 1975). From this we know that $\tau \approx (1/s_b)(\log(\sigma^2\rho) - \log W)$; although more detailed information is available (e.g. a central limit theorem for τ, Nagaev (1971)), we will stick to the loose interpretation.

So, roughly speaking, we need to evaluate the importance of a delay of about $T = (1/s_b)\log(\sigma^2\rho)$. New mutations will appear and become established during this time if $2\rho\sigma^2\mu s_b \geq 1/T$, i.e. if $2\rho\sigma^2\mu \geq 1/\log(\sigma^2\rho)$. Our model will still be a good approximation, however, as long as T is short relative to the time a wave takes to spread between nearby mutational origins. This can be worked out, but it is simpler to note that if the

converse is true (i.e., that reaching local fixation is slow compared to the spread of the wave), the process is largely unaffected by spatial structure, and so the panmictic model is a good approximation for the true process.

Pennings & Hermisson (2006a) show that under a panmictic model with certain assumptions on the parameters, the number of independent origins due to both standing variation and new mutation seen in a sample of size n has approximately the Ewens distribution with parameters n and $\theta = 4N\mu$. As n increases, the total number of types seen grows as $\log n$. In our model, we can increase total population size by increasing either the density ρ or the total amount of area. In either case, the predicted number of distinct types grows linearly with n, much faster than under panmixia.

The results of Hermisson & Pennings (2005) and Pennings & Hermisson (2006a) suggest that in a panmictic population the number of independent alleles (and their frequencies) in a sample is nearly independent of s_b and s_d (although this breaks down with fluctuating population size, Wilson et al., 2014). In the panmictic model the lack of dependence on s_b comes about because while increasing s_b increases the rate at which independent mutations become established, it also accelerates the frequency gain of established alleles, hence decreasing the time period in which new alleles can arise and hope to be at significant frequency in the population. These two effects approximately cancel each other out leading to no strong effect of s_b on the number of independent alleles. Decreasing s_d increases the number of standing variants within a population, increasing the number of alleles that manage to establish and spread from standing variation (Hermisson & Pennings, 2005; Orr & Betancourt, 2001). However, having more established standing alleles acts to exclude the spread of new alleles that arise once the environment switches. These two opposing effects again cancel out, leading to little overall effect of s_d on the number of independent alleles. In contrast, our results show that the characteristic length (closely related to the density of independent alleles) depends on both s_b and s_d in a geographically spread population. Like the panmictic model, in our model, alleles also act to exclude each other; however, the geographic spread of an allele is slow compared to the initial exponential growth of an allele in a panmictic population. That means that the role of selection in helping alleles become established can dominate, leading to more independent origins, both by being weaker before and stronger after the environmental switch.

2 Discussion

When the geographical area where a species experiences a selection pressure is greater than our characteristic length we expect multiple independent alleles to arise and spread in that area. Our results suggest that convergent evolution among populations may be quite common, at least when population densities and mutational target sizes are not too small, and dispersal is limited across the range. While considering smaller mutational target sizes (e.g., less than the 100bp in our example) or lower population densities would lower the probability of convergence within a given geographical area, it would also act to considerably lengthen the time to adaptation. In such cases adaptation may simply fail to occur at all or populations may adapt via some other means (e.g., by using a broader mutational target).

Our inclusion of standing variation before environmental change greatly increase the probability of convergence within a species. While at face value this increase seems unsurprising, this relationship differs markedly from the case of multiple competing alleles in a panmictic population (see discussion above and in Hermisson & Pennings, 2005; Pennings & Hermisson, 2006a). Importantly, allowing standing variation may greatly lower the time until the species becomes adapted across the geographic range of the selection pressure. We have also shown that adaptation through standing variation biases the type of variation towards those alleles with fewer pleiotropic effects, since these are more common as standing variation before the environmental shift. This bias can in some cases easily overcome quite significant differences in mutational target sizes among loci allowing the same locus to be repeatedly the source of adaptation even if there are seemingly many different routes to adaptation. (See Figure 8.)

The confusing signal of geographic convergent evolution. As we have argued in Ralph & Coop (2010) the ease with which geographic convergent adaptation occurs means that we should incorporate it

more widely into our thinking about the genetic basis of adaptation. For example, the absence of European skin pigmentation alleles in ancient DNA from Europeans who lived several thousand years ago has led to the suggestion that these individuals had dark pigmentation (Olalde et al., 2014; Lazaridis et al., 2014; Wilde et al., 2014). However, given our results and the partially convergent basis of skin pigmentation between Europeans and East Asians (Norton et al., 2007; Edwards et al., 2010) it seems just as plausible that these ancient individuals adapted to high latitudes via a different complement of "light-skin" pigmentation alleles; to our knowledge, we have no strong evidence either way. Such convergence may considerably complicate the exploration of phenotypes and adaptation among populations using variants mapped in a limited set of populations (Berg & Coop, 2014).

More generally, if geographic convergence is common, we should often expect to see selected alleles that are strongly geographically restricted as they have simply not had time for neutral gene flow to spread them across the landscape. Such convergent alleles will be F_{ST}-outliers compared to older neutral variation, until there is sufficient time for migration to smooth out them out across the landscape. This pattern may be very hard to distinguish from local adaptation using population genomic approaches alone. This is especially problematic as boundaries between convergent alleles may often occur where gene flow rates are low, i.e. historical and ecological breaks, even if the alleles concerned have no bearing on the ecological differences across these breaks (see Bierne et al., 2011, for a wide-ranging discussion of how allelic differentiation may build along particular zones). We are rarely so fortunate as to know as much about the genetics, phenotypic distributions, and potential selection agents as we do for malaria resistance in humans. Therefore, we must be wary of mistaking the strange spatial distributions of particular alleles for adaptation to some very specific selection pressure (e.g. the distribution of the Mahidol allele in Figure 3), when they are simply elements of a larger geographic mosaic of alleles responding to a broadly shared selection pressure.

Each of these local sweeps will be associated with the haplotype on which the particular allele arose. Under the parameter regime we study, standing variants are still quite young, so we do not expect a strongly reduced hitchhiking effect. As such, following the initial period of adaptation, we should expect the population to be partitioned into a set of geographically restricted long haplotypes. Given sufficient time these haplotypes will mix together through migration and drift, potentially leading to a within population signal of a sweep from multiple independent mutations if our selected allele occurs at the same locus (Pennings & Hermisson, 2006b), or to multiple partial sweeps if the loci are scattered across the genome (Coop & Ralph, 2012).

Our results are predicated on the idea that adaptive variants are initially rare within populations, i.e. they are reasonably deleterious before the environment switches. In contrast, if they adapt via common variation even distant populations could have a shared basis of adaptation, e.g. previously neutral (or nearly neutral) variation, shared among populations. If many loci contribute to variation in a trait, then selection on any one allele may be weak, which might lead adaptation to use the same alleles in different populations. However, sufficiently differentiated populations may still adapt via different genetic routes, as the constellation of alleles that respond to selection quickest will be somewhat different due to drift among populations (Barton, 1989). Therefore, it may be the case that polygenic traits are even less susceptible to a shared genetic basis to adaptation across populations than simple traits. However, we currently lack good models and methods with which to test this.

Are species held together by widespread selective sweeps? Our results touch on an old debate on the evolutionary coherence of species. Mayr and many others have argued that species are coherent evolutionary units because they are united by shared gene flow (pages 521–522 in Mayr, 1963). However, this argument has been questioned by a number of authors based on relatively high levels of differentiation, and low rates of dispersal, in many species (Ehrlich & Raven, 1969; Levin, 1979). Even if gene flow is not high enough to prevent neutral differentiation or local adaptation, a number of authors have argued that species are cohesive if gene flow is high enough for globally selected alleles (and their hitchhiking haplotypes) to spread across entire species (see also Rieseberg & Burke, 2001; Morjan & Rieseberg, 2004; Ellstrand, 2014). At present, large scale genotyping and sequencing projects, along with more sophisticated methods, are highlighting ever more signals of gene flow between populations and species (Patterson et al., 2012; Sousa & Hey, 2013; Hellenthal et al., 2014). However, our work on geographic convergent adaptation (see

also Ralph & Coop, 2010, 2014) suggests that species should often adapt to widespread selection pressures through convergent evolution rather than waiting for a single allele to migrate across the range.

In support of this idea, putative recent selective sweeps seem to often be geographically restricted (Pickrell et al., 2009; Coop et al., 2009; Granka et al., 2012), rather than species-wide (but see Clark et al., 2007; Long et al., 2013, for a potential example). This is likely in part due to the relatively low incidence of widespread selection pressures, but as noted above even when we know of widespread selection pressures (e.g. malaria) the response is usually convergent, not shared, across large spatial scales. On the other hand, introgression of adaptive alleles across species and sub-species boundaries, suggests that selected alleles do sometimes spread despite low migration rates (see Hedrick, 2013, for a recent review). However, at least some of these cases may be caused by introgression of haplotype complexes consisting of many, tightly linked, beneficial alleles (that perhaps are inaccessible by mutation over reasonable time-scales for a population in a new environment). Currently we can only scan genomes for species-wide sweeps in those few organisms with population-scale sequence data, and so we do not know if these observations generalize to most species. This is rapidly changing, and will allow us to form a much improved picture of the relationship between the level of neutral population structure, and the age and geographic spread of selected alleles across many species.

Even if selective sweeps only bring alleles to fixation locally, they are still potentially a stronger homogenizing force than neutral mixing through migration. Under neutral mixing, the mean number of generations back to the most recent common ancestor is on order of the total effective population size. This quantity has not been worked out for a model with simultaneous local sweeps, but will be somewhat analogous to the "spatial Λ-Fleming–Viot" models of Barton et al. (2013b), in which local sweeps occur independently across the range. Lineages that are closely linked to the sweeping allele ($\sim v/\chi$ Morgans) will be moved towards the center of the sweep (a displacement $O(\chi)$), and pairs of lineages caught up in the same sweep could be forced to coalesce (see Barton et al., 2013a, for work on geographic hitchhiking). In this case lineages and alleles are literally hitchhiking across space. The overall rate of lineage movement and coalescence depends on the rate of sweeps, their geographic scale, and the rate of recombination, and could be calculated by combining the result presented here with Barton et al. (2013b) and Barton et al. (2013a). However, if geographic sweeps are common then this may substantially speed up the rate of mixing compared to neutral drift and migration.

How then do substitutions occur? If it is rare for gene flow to rapidly spread selected alleles across a species range, how then do selected alleles ever become fixed within species? Drift alone will act only slowly to sort variants within species into divergence among species. Slight selective differences in the pleiotropic effects (or linked background) among convergent alleles could allow one allele to press into areas occupied by other alleles. Furthermore, repeated bouts of adaptation in particular genomic regions may act to push a subset of previously selected alleles to fixation across the species range, through the spread of the genetic backgrounds on which they arise. However, it seems likely that this is a slow process compared to the initial rapid spread of selected alleles.

Speciation and extinction as phases of molecular evolution? One potential resolution is that many selected alleles achieve fixation, not through their own species-wide spread, but rather through subsequent large-scale changes in geographic range size induced by extirpation of the species over parts of its range (see Barton et al., 2013b, and references therein for how such a model could be constructed). Such drops in range size may fix, or radically change the species-wide frequency, of alleles previously restricted to small portion of a species range. Furthermore, many modes of speciation are proposed to occur through a geographically-limited subset of populations forming the basis of new species, e.g. the splitting off of part of the range through a vicariance event or dispersal of a subset of individuals. In this case, speciation will cause geographic assortment of polymorphic ancestral variation, again acting to fix variants within newly formed species that were previously polymorphic across ancestral species ranges.

Such ideas are not completely new and represent a perhaps logical consequence of an allopatric or parapatric view of the biogeography of speciation. However, it is worth revisiting this idea as geographically broad population genomic sampling allows us to return to themes in biogeography. Along similar lines, Futuyma

has argued that much of the adaptive differentiation within species, e.g. adaptation to local conditions, may be ephemeral and subject to loss due to local extinction and the mixing following the collapse of population structure (Futuyma, 2010, 1987). Futuyma offered this as an explanation of the pattern of punctuated equilibrium (Eldredge & Gould, 1972), and argued that the observation of stasis and rapid anagenesis associated with speciation were consistent with micro-evolution. Futuyma argued that despite rapid adaptation over short time-scales, we may observe morphological stasis in the fossil record as much of this adaptation is lost to local extinction and the collapse of population structure (see also Lieberman & Dudgeon, 1996; Eldredge et al., 2005; Futuyma, 2010). Furthermore, he suggests that speciation may act as ratchet to prevent the loss of differentiation, acting to maintain adaptive changes among populations, and prevent their loss by interbreeding. At face value the rate of species formation seems too low to contribute to this process. However, Rosenblum et al. (2012), and many others, have argued that the rate of speciation may well be quite high, but that the majority of incipient species do not persist long due to reabsorption or extinction. Changes in range size, due to local extinction, can also be very rapid on the time-scales over which alleles may spread on the landscape (Gaston, 2003; Hewitt, 1996). Repeated bouts of extinction and speciation will send waves of alleles to fixation along particular lineages.

Such a link between speciation and substitution would not imply that substitutions should necessarily be thought of as being clustered at splits in inferred phylogenies (see Pennell et al., 2014a; Venditti & Pagel, 2014; Pennell et al., 2014b, for a recent exchange on this). Neutral substitutions are unaffected by this process, because they accumulate in a clocklike manner along lineages, as dictated by the mutation rate, regardless of the geographic details of their polymorphic stage. Turning to the accumulation of adaptive substitutions, it is likely that splits in phylogenies are only a tiny proportion of all incipient speciation events, because extinction rates may be high (Rosenblum et al., 2012), and so every lineage has likely passed through many "speciation events" in addition to the observed ones. Under these assumptions, spatial polymorphisms could accumulate gradually in geographically restricted populations between the large-scale biogeographic events that cause their fixation or loss. This effectively decorrelates the time at which new alleles arise and when they fix in the species, an effect similar to that pointed out by Gillespie (1994).

This view would not imply that adaptive evolution or speciation is driven by the shifting balance or genetic revolutions (Wright, 1932; Mayr, 1954), whereby genetic drift allows populations to cross fitness valleys and substitute novel epistatic combinations. Although geographic lineage sorting via speciation and extinction can be thought of as very large-scale genetic drift events, in the models we study here the initial spread of alleles is due to selection, not drift (see also Futuyma, 1989, for discussion).

There is evidence that a reasonable fraction of genome-wide substitutions are fixed by positive selection in a number of species (most notably Drosophila, Sella et al., 2009). Under the geographic view of fixation, selection has played a strong role in the establishment of these alleles locally. As we get more broadly geographic population genomics sampling for a range of species we will have the opportunity to study whether the class of alleles that contribute to local differentiation are similar to those underlying species divergence, and the extent to which the answer to this depends on the age and type of population structure within species.

Finally, we close by noting that range expansion and speciation are obviously not separate from adaptive differentiation. The invasion of new geographic areas may lead to a burst of adaptive differentiation, at least in a subset of genes, and speciation may be associated with rapidly adaption to novel environments. Conversely, if the geographic spread of adaptive alleles within ranges is slow (e.g. if they only offer a local advantage or if they are selectively excluded) this may allow Dobzhansky-Muller incompatibilities to arise within species, effectively offering a mechanism for hybrid incompatibilities to evolve in parapatry, and fracturing the range (Bank et al., 2012; Kondrashov, 2003; Bierne et al., 2011). The alleles that act as components in many of the Dobzhansky-Muller incompatibilities studied to date are geographically restricted (see Cutter, 2012). Therefore, it seems possible that populations within species may often be tending towards speciation, and that as outlined here that this may drive some proportion of molecular divergence.

Acknowledgements

We thank Michael Turelli, Jon Seger, and the Coop lab for helpful conversations. We thank Florence Débarre, Douglas Futuyma, Michael Whitlock, Molly Przeworski, and Sam Yeaman for helpful comments on an earlier version of the paper. All figures and calculations were performed in R (R Core Team, 2014). The figure color paletes were generated using Karthik Ram's "Wes Anderson" R color palette `https://github.com/karthik/wesanderson`. This work was supported by grants from the National Science Foundation under Grant No. 1262645 to PLR and GC and by the National Institute of General Medical Sciences of the National Institutes of Health under award numbers NIH RO1GM83098 and RO1GM107374 to GC.

References

Arendt, J. & Reznick, D. 2008: Convergence and parallelism reconsidered: what have we learned about the genetics of adaptation? Trends Ecol Evol 23(1):26–32.

Aronson, D. & Weinberger, H. 1975: Nonlinear diffusion in population genetics, combustion, and nerve pulse propagation. Partial Differential Equations and Related Topics pages 5–49.

Bank, C.; Bürger, R. & Hermisson, J. 2012: The limits to parapatric speciation: Dobzhansky-Muller incompatibilities in a continent-island model. Genetics 191(3):845–863.

Barton, N. 1989: The divergence of a polygenic system subject to stabilizing selection, mutation and drift. Genetics Research 54:59–78.

Barton, N. H.; Etheridge, A. M.; Kelleher, J. & Véber, A. 2013a: Genetic hitchhiking in spatially extended populations. Theoretical Population Biology 87:75–89.

Barton, N. H.; Etheridge, A. M. & Véber, A. 2013b: Modelling evolution in a spatial continuum. Journal of Statistical Mechanics: Theory and Experiment 2013(01):P01002.

Berg, J. J. & Coop, G. 2014: A population genetic signal of polygenic adaptation. PLoS genetics 10(8):e1004412.

Bierne, N.; Welch, J.; Loire, E.; Bonhomme, F. & David, P. 2011: The coupling hypothesis: why genome scans may fail to map local adaptation genes. Molecular Ecology 20(10):2044–2072.

Bollobás, B. & Riordan, O. 2008: Percolation on random Johnson-Mehl tessellations and related models. Probab. Theory Related Fields 140(3-4):319–343.

Bradshaw, A. D. 1991: The Croonian lecture, 1991: Genostasis and the limits to evolution. Philosophical Transactions: Biological Sciences 333(1267):pp. 289–305.

Chevin, L. M.; Martin, G. & Lenormand, T. 2010: Fisher's model and the genomics of adaptation: restricted pleiotropy, heterogenous mutation, and parallel evolution. Evolution 64(11):3213–3231.

Clark, R. M.; Schweikert, G.; Toomajian, C.; Ossowski, S.; Zeller, G.; Shinn, P.; Warthmann, N.; Hu, T. T.; Fu, G.; Hinds, D. A.; Chen, H.; Frazer, K. A.; Huson, D. H.; Scholkopf, B.; Nordborg, M.; Ratsch, G.; Ecker, J. R. & Weigel, D. 2007: Common sequence polymorphisms shaping genetic diversity in *Arabidopsis thaliana*. Science 317(5836):338–342.

Conte, G. L.; Arnegard, M. E.; Peichel, C. L. & Schluter, D. 2012: The probability of genetic parallelism and convergence in natural populations. Proceedings of the Royal Society of London B: Biological Sciences .

Coop, G.; Pickrell, J. K.; Novembre, J.; Kudaravalli, S.; Li, J.; Absher, D.; Myers, R. M.; Cavalli-Sforza, L. L.; Feldman, M. W. & Pritchard, J. K. 2009: The role of geography in human adaptation. PLoS Genet 5(6).

Coop, G. & Ralph, P. 2012: Patterns of neutral diversity under general models of selective sweeps. Genetics 192(1):205–224.

Corbett-Detig, R. B.; Hartl, D. L. & Sackton, T. B. 2015: Natural selection constrains neutral diversity across A wide range of species. PLoS Biol 13(4):e1002112.

Cox, D. R. & Isham, V. 1980: Point processes. Chapman & Hall, London. Monographs on Applied Probability and Statistics.

Cutter, A. D. 2012: The polymorphic prelude to Bateson-Dobzhansky-Muller incompatibilities. Trends Ecol. Evol. (Amst.) 27(4):209–218.

Edwards, M.; Bigham, A.; Tan, J.; Li, S.; Gozdzik, A.; Ross, K.; Jin, L. & Parra, E. J. 2010: Association of the *oca2* polymorphism His615Arg with melanin content in east Asian populations: Further evidence of convergent evolution of skin pigmentation. PLoS Genet 6(3):e1000867.

Ehrlich, P. R. & Raven, P. H. 1969: Differentiation of populations. Science 165(3899):1228–1232.

Eldredge, N. & Gould, S. 1972: Punctuated equilibria: an alternative to phyletic gradualism. Models in paleobiology., (edited by T. J. M. Schopf). Freeman Cooper and Co.

Eldredge, N.; Thompson, J.; Brakefield, P.; Gavrilets, S.; Jablonski, D.; Jackson, J.; Lenski, R.; Lieberman, B.; McPeek, M. & Miller, W. 2005: The dynamics of evolutionary stasis. Paleobiology 31(2, S):133–145.

Ellstrand, N. C. 2014: Is gene flow the most important evolutionary force in plants? American journal of botany 101(5):737–753.

Fanfoni, M. & Tomellini, M. 1998: The Johnson-Mehl-Avrami-Kolmogorov model: A brief review. Il Nuovo cimento. D 20(7-8):1171–1182.

Fisher, R. 1937: The wave of advance of advantageous genes. Ann. Eugenics 7:353–369.

Flint, J.; Harding, R. M.; Boyce, A. J. & Clegg, J. B. 1998: The population genetics of the haemoglobinopathies. Baillieres Clin Haematol 11(1):1–51.

Futuyma, D. 1987: On the role of species in anagenesis. American Naturalist 130(3):465–473.

Futuyma, D. J. 1989: Speciational trends and the role of species in macroevolution. The American Naturalist 134(2):pp. 318–321.

Futuyma, D. J. 2010: Evolutionary constraint and ecological consequences. Evolution 64(7):1865–1884.

Gaston, K. J. 2003: The structure and dynamics of geographic ranges. Oxford University Press.

Gilbert, E. N. 1962: Random subdivisions of space into crystals. Ann. Math. Statist. 33:958–972.

Gillespie, J. H. 1994: Substitution processes in molecular evolution. II. Exchangeable models from population genetics. Evolution 48(4):pp. 1101–1113.

Gordo, I. & Campos, P. R. A. 2006: Adaptive evolution in a spatially structured asexual population. Genetica 127(1-3):217–229.

Granka, J. M.; Henn, B. M.; Gignoux, C. R.; Kidd, J. M.; Bustamante, C. D. & Feldman, M. W. 2012: Limited evidence for classic selective sweeps in African populations. Genetics 192(3):1049–1064.

Haldane, J. B. S. 1927: A mathematical theory of natural and artificial selection, part V: Selection and mutation. Mathematical Proceedings of the Cambridge Philosophical Society 23:838–844.

Haldane, J. B. S. 1932: The causes of evolution, pages 75–76. Longmans, Green, and Co, New York.

Haldane, J. B. S. 1937: The effect of variation of fitness. The American Naturalist 71(735):pp. 337–349.

Hallatschek, O. & Fisher, D. S. 2014: Acceleration of evolutionary spread by long-range dispersal. Proceedings of the National Academy of Sciences 111(46):E4911–E4919.

Hedrick, P. 2011: Population genetics of malaria resistance in humans. Heredity 107(4).

Hedrick, P. W. 2013: Adaptive introgression in animals: examples and comparison to new mutation and standing variation as sources of adaptive variation. Molecular ecology 22(18):4606–4618.

Hellenthal, G.; Busby, G. B.; Band, G.; Wilson, J. F.; Capelli, C.; Falush, D. & Myers, S. 2014: A genetic atlas of human admixture history. Science 343(6172):747–751.

Hermisson, J. & Pennings, P. S. 2005: Soft sweeps: molecular population genetics of adaptation from standing genetic variation. Genetics 169(4):2335–2352.

Hewitt, G. M. 1996: Some genetic consequences of ice ages, and their role in divergence and speciation. Biological Journal of the Linnean Society 58(3):247–276.

Howes, R. E.; Dewi, M.; Piel, F. B.; Monteiro, W. M.; Battle, K. E.; Messina, J. P.; Sakuntabhai, A.; Satyagraha, A. W.; Williams, T. N.; Baird, J. K. & Hay, S. I. 2013: Spatial distribution of G6PD deficiency variants across malaria-endemic regions. Malar. J. 12:418.

Howes, R. E.; Piel, F. B.; Patil, A. P.; Nyangiri, O. A.; Gething, P. W.; Dewi, M.; Hogg, M. M.; Battle, K. E.; Padilla, C. D.; Baird, J. K. & Hay, S. I. 2012: G6PD deficiency prevalence and estimates of affected populations in malaria endemic countries: a geostatistical model-based map. PLoS Med. 9(11):e1001339.

Jagers, P. 1975: Branching processes with biological applications. Wiley Series in Probability and Statistics: Applied Probability and Statistics Section Series. Wiley.

Jeong, C. & Rienzo, A. D. 2014: Adaptations to local environments in modern human populations. Current Opinion in Genetics & Development 29(0):1–8.

Karasov, T.; Messer, P. W. & Petrov, D. A. 2010: Evidence that adaptation in Drosophila is not limited by mutation at single sites. PLoS genetics 6(6):e1000924.

Kolmogorov, A.; Petrovskii, I. & Piscunov, N. 1937: A study of the equation of diffusion with increase in the quantity of matter, and its application to a biological problem. Byul. Moskovskogo Gos. Univ. 1(6):1–25.

Kolmogorov, A. N. 1937: On the statistical theory of metal crystallization. Izv. Akad. Nauk SSSR, Ser. Math 1:335–360.

Kondrashov, A. S. 2003: Accumulation of Dobzhansky-Muller incompatibilities within a spatially structured population. Evolution 57(1):151–153.

Kwiatkowski, D. P. 2005: How malaria has affected the human genome and what human genetics can teach us about malaria. Am. J. Hum. Genet. 77(2):171–192.

Lazaridis, I.; Patterson, N.; Mittnik, A.; Renaud, G.; Mallick, S.; Sudmant, P. H.; Schraiber, J. G.; Castellano, S.; Kirsanow, K.; Economou, C. et al. 2014: Ancient human genomes suggest three ancestral populations for present-day europeans. Nature .

Leffler, E. M.; Bullaughey, K.; Matute, D. R.; Meyer, W. K.; Ségurel, L.; Venkat, A.; Andolfatto, P. & Przeworski, M. 2012: Revisiting an old riddle: What determines genetic diversity levels within species? PLoS Biol 10(9):e1001388.

Levin, D. A. 1979: The nature of plant species. Science 204(4391):pp. 381–384.

Lieberman, B. S. & Dudgeon, S. 1996: An evaluation of stabilizing selection as a mechanism for stasis. Palaeogeography, Palaeoclimatology, Palaeoecology 127(1):229–238.

Long, Q.; Rabanal, F. A.; Meng, D.; Huber, C. D.; Farlow, A.; Platzer, A.; Zhang, Q.; Vilhjalmsson, B. J.; Korte, A.; Nizhynska, V.; Voronin, V.; Korte, P.; Sedman, L.; Mandakova, T.; Lysak, M. A.; Seren, U.; Hellmann, I. & Nordborg, M. 2013: Massive genomic variation and strong selection in *Arabidopsis thaliana* lines from Sweden. Nat. Genet. 45(8):884–890.

Louicharoen, C.; Patin, E.; Paul, R.; Nuchprayoon, I.; Witoonpanich, B.; Peerapittayamongkol, C.; Casademont, I.; Sura, T.; Laird, N. M.; Singhasivanon, P.; Quintana-Murci, L. & Sakuntabhai, A. 2009: Positively selected G6PD-Mahidol mutation reduces *Plasmodium vivax* density in Southeast Asians. Science 326(5959):1546–1549.

Malaria Genomic Epidemiology Network. 2014: Reappraisal of known malaria resistance loci in a large multicenter study. Nat Genet 46(11):1197–1204.

Manjurano, A.; Sepulveda, N.; Nadjm, B.; Mtove, G.; Wangai, H.; Maxwell, C.; Olomi, R.; Reyburn, H.; Riley, E. M.; Drakeley, C. J.; Clark, T. G. & MalariaGEN Consortium. 2015: African glucose-6-phosphate dehydrogenase alleles associated with protection from severe malaria in heterozygous females in Tanzania. PLoS Genet 11(2):e1004960.

Martens, E. A. & Hallatschek, O. 2011: Interfering waves of adaptation promote spatial mixing. Genetics 189(3):1045–1060.

Martin, A. & Orgogozo, V. 2013: The Loci of repeated evolution: a catalog of genetic hotspots of phenotypic variation. Evolution 67(5):1235–1250.

Mayr, E. 1954: Evolution as a Process, chapter "Change of genetic environment and evolution". Allen & Unwin, London.

Mayr, E. 1963: Animal Species and Evolution. Harvard University Press.

Messer, P. W. & Petrov, D. A. 2013: Population genomics of rapid adaptation by soft selective sweeps. Trends Ecol. Evol. (Amst.) 28(11):659–669.

Minucci, A.; Moradkhani, K.; Hwang, M. J.; Zuppi, C.; Giardina, B. & Capoluongo, E. 2012: Glucose-6-phosphate dehydrogenase (G6PD) mutations database: review of the "old" and update of the new mutations. Blood Cells Mol. Dis. 48(3):154–165.

Møller, J. 1992: Random Johnson-Mehl tessellations. Advances in Applied Probability 24(4):814–844.

Møller, J. 1994: Lectures on Random Voronoi Tessellations, chapter "Poisson-Voronoi tessellations". Lecture Notes in Statistics. Springer New York.

Møller, J. 1995: Generation of Johnson-Mehl crystals and comparative analysis of models for random nucleation. Advances in Applied Probability 27(2):367–383.

Morjan, C. L. & Rieseberg, L. H. 2004: How species evolve collectively: implications of gene flow and selection for the spread of advantageous alleles. Mol. Ecol. 13(6):1341–1356.

Nagaev, A. V. 1971: A limit theorem for a supercritical branching process. Mathematical notes of the Academy of Sciences of the USSR 9(5):338–342.

Norton, H. L.; Kittles, R. A.; Parra, E.; McKeigue, P.; Mao, X.; Cheng, K.; Canfield, V. A.; Bradley, D. G.; McEvoy, B. & Shriver, M. D. 2007: Genetic evidence for the convergent evolution of light skin in Europeans and east Asians. Mol Biol Evol 24(3):710–722.

Olalde, I.; Allentoft, M. E.; Sánchez-Quinto, F.; Santpere, G.; Chiang, C. W.; DeGiorgio, M.; Prado-Martinez, J.; Rodríguez, J. A.; Rasmussen, S.; Quilez, J. et al. 2014: Derived immune and ancestral pigmentation alleles in a 7,000-year-old mesolithic european. Nature 507(7491):225–228.

Orr, H. A. 2000: Adaptation and the cost of complexity. Evolution 54(1):13–20.

Orr, H. A. 2005: The probability of parallel evolution. Evolution 59(1):216–220.

Orr, H. A. & Betancourt, A. J. 2001: Haldane's sieve and adaptation from the standing genetic variation. Genetics 157(2):875–884.

Otto, S. P. 2004: Two steps forward, one step back: the pleiotropic effects of favoured alleles. Proc. Biol. Sci. 271(1540):705–714.

Otwinowski, J. & Krug, J. 2014: Clonal interference and Muller's ratchet in spatial habitats. Physical Biology 11(5):056003.

Pamillo, P. 1979: Genic variation at sex-linked loci: Quantification of regular selection models. Hereditas 91(1):129–133.

Patterson, N.; Moorjani, P.; Luo, Y.; Mallick, S.; Rohland, N.; Zhan, Y.; Genschoreck, T.; Webster, T. & Reich, D. 2012: Ancient admixture in human history. Genetics 192(3):1065–1093.

Pennell, M. W.; Harmon, L. J. & Uyeda, J. C. 2014a: Is there room for punctuated equilibrium in macroevolution? Trends in Ecology & Evolution 29(1):23–32.

Pennell, M. W.; Harmon, L. J. & Uyeda, J. C. 2014b: Speciation is unlikely to drive divergence rates. Trends in Ecology & Evolution 29(2):72–73.

Pennings, P. S. & Hermisson, J. 2006a: Soft sweeps II - molecular population genetics of adaptation from recurrent mutation or migration. Mol Biol Evol page msj117.

Pennings, P. S. & Hermisson, J. 2006b: Soft sweeps III: The signature of positive selection from recurrent mutation. PLoS Genet 2(12):e186.

Pickrell, J. K.; Coop, G.; Novembre, J.; Kudaravalli, S.; Li, J. Z.; Absher, D.; Srinivasan, B. S.; Barsh, G. S.; Myers, R. M.; Feldman, M. W. & Pritchard, J. K. 2009: Signals of recent positive selection in a worldwide sample of human populations. Genome Res 19(5):826–837.

Pickrell, J. K. & Reich, D. 2014: Toward a new history and geography of human genes informed by ancient DNA. Trends Genet. 30(9):377–389.

R Core Team. 2014: R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Ralph, P. & Coop, G. 2010: Parallel adaptation: One or many waves of advance of an advantageous allele? Genetics 186(2):647–668.

Ralph, P. L. & Coop, G. 2014: Convergent evolution during local adaptation to patchy landscapes. bioRxiv/006940 .

Rieseberg, L. H. & Burke, J. M. 2001: The biological reality of species: Gene flow, selection, and collective evolution. Taxon 50(1):pp. 47–67.

Rosenblum, E. B.; Sarver, B. A.; Brown, J. W.; Roches, S. D.; Hardwick, K. M.; Hether, T. D.; Eastman, J. M.; Pennell, M. W. & Harmon, L. J. 2012: Goldilocks Meets Santa Rosalia: An Ephemeral Speciation Model Explains Patterns of Diversification Across Time Scales. Evol Biol 39(2):255–261.

Ruwende, C.; Khoo, S. C.; Snow, R. W.; Yates, S. N.; Kwiatkowski, D.; Gupta, S.; Warn, P.; Allsopp, C. E.; Gilbert, S. C. & Peschu, N. 1995: Natural selection of hemi- and heterozygotes for G6PD deficiency in Africa by resistance to severe malaria. Nature 376(6537):246–249.

Saunders, M. A.; Slatkin, M.; Garner, C.; Hammer, M. F. & Nachman, M. W. 2005: The extent of linkage disequilibrium caused by selection on G6PD in humans. Genetics 171(3):1219–1229.

Sella, G.; Petrov, D. A.; Przeworski, M. & Andolfatto, P. 2009: Pervasive natural selection in the Drosophila genome? PLoS Genet. 5(6):e1000495.

Slatkin, M. 1976: The rate of spread of an advantageous allele in a subdivided population. Population Genetics and Ecology, (edited by S. Karlin & E. Nevo), pages 767–780. Academic Press, Inc; New York.

Slatkin, M. 2008: A Bayesian method for jointly estimating allele age and selection intensity. Genet Res (Camb) 90(1):129–137.

Slatkin, M. & Charlesworth, D. 1978: The spatial distribution of transient alleles in a subdivided population: a simulation study. Genetics 89(4):793–810.

Sousa, V. & Hey, J. 2013: Understanding the origin of species with genome-scale data: modelling gene flow. Nat. Rev. Genet. 14(6):404–414.

Stern, D. L. 2013: The genetic causes of convergent evolution. Nat Rev Genet 14(11):751–764.

Tishkoff, S. A.; Varkonyi, R.; Cahinhinan, N.; Abbes, S.; Argyropoulos, G.; Destro-Bisol, G.; Drousiotou, A.; Dangerfield, B.; Lefranc, G.; Loiselet, J.; Piro, A.; Stoneking, M.; Tagarelli, A.; Tagarelli, G.; Touma, E. H.; Williams, S. M. & Clark, A. G. 2001: Haplotype diversity and linkage disequilibrium at human G6PD: Recent origin of alleles that confer malarial resistance. Science 293(5529):455–462.

Unckless, R. L. & Orr, H. A. 2009: Dobzhansky-Muller incompatibilities and adaptation to a shared environment. Heredity (Edinb) 102(3):214–217.

Venditti, C. & Pagel, M. 2014: Plenty of room for punctuational change. Trends in Ecology & Evolution 29(2):71–72.

Welch, J. J. & Waxman, D. 2003: Modularity and the cost of complexity. Evolution 57(8):1723–1734.

Wilde, S.; Timpson, A.; Kirsanow, K.; Kaiser, E.; Kayser, M.; Unterländer, M.; Hollfelder, N.; Potekhina, I. D.; Schier, W.; Thomas, M. G. & Burger, J. 2014: Direct evidence for positive selection of skin, hair, and eye pigmentation in Europeans during the last 5,000 years. Proceedings of the National Academy of Sciences 111(13):4832–4837.

Wilson, B. A.; Petrov, D. A. & Messer, P. W. 2014: Soft selective sweeps in complex demographic scenarios. Genetics 198(2):669–684.

Wright, S. 1932: The roles of mutation, inbreeding, crossbreeding and selection in evolution. Proc. VI Int. Congr. Genet. .

Wright, S. 1943: Isolation by Distance. Genetics 28(2):114–138.

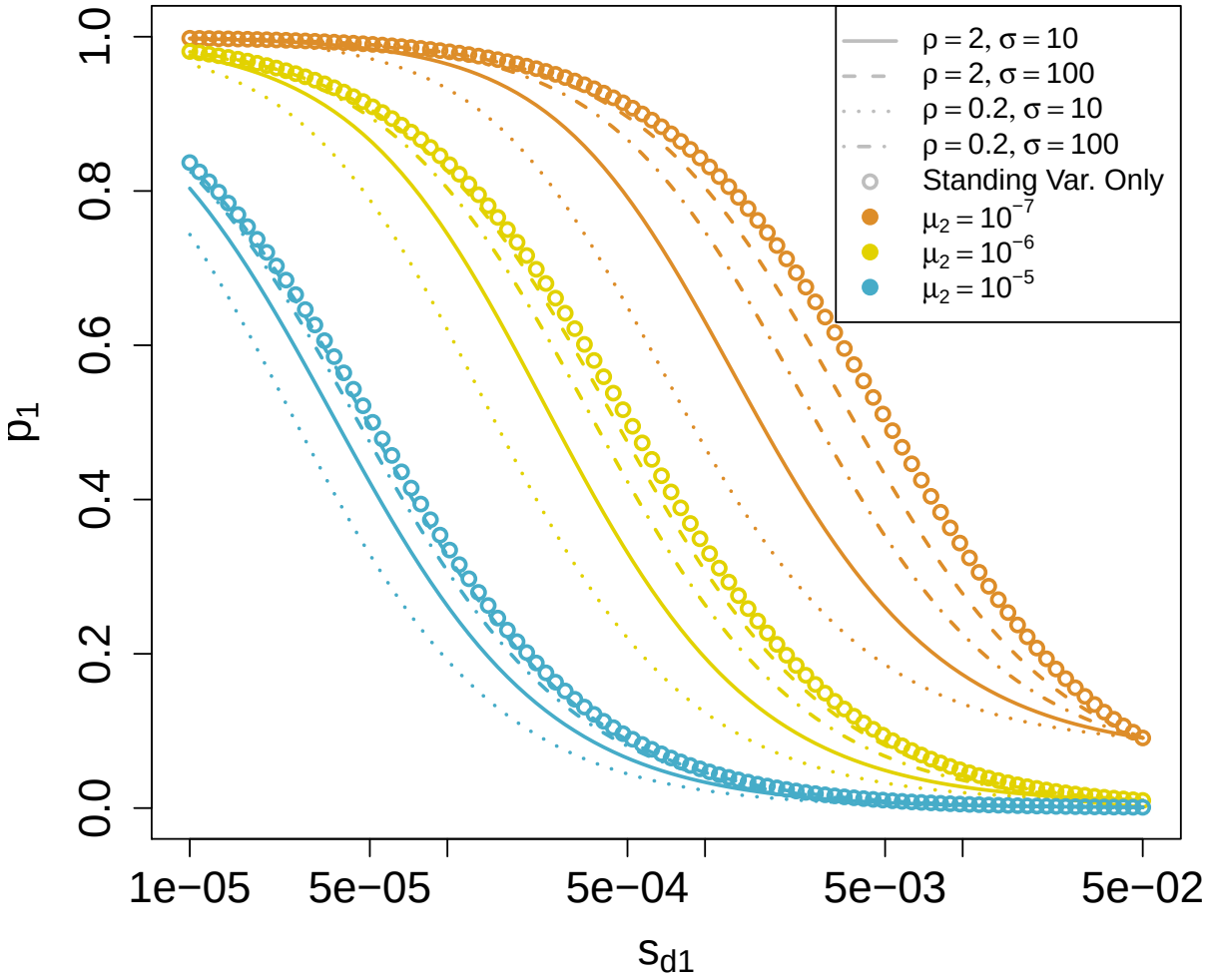

Figure 8: **Proportion of space covered by rarer but less negatively pleiotropic mutation**. Empty circles give the result for standing variation only, equation (16). Lines give the result allowing both standing and *de novo* mutations (equation (15)). Here we hold $s_b = s_{d,2} = 0.05$ and $\mu_1 = 1 \times 10^{-8}$.