

REVIEW

Received 24 Mar 2014 | Accepted 17 Sep 2014 | Published 27 Oct 2014

DOI: 10.1038/ncomms6281

On the unfounded enthusiasm for soft selective sweeps

Jeffrey D. Jensen^{1,2}

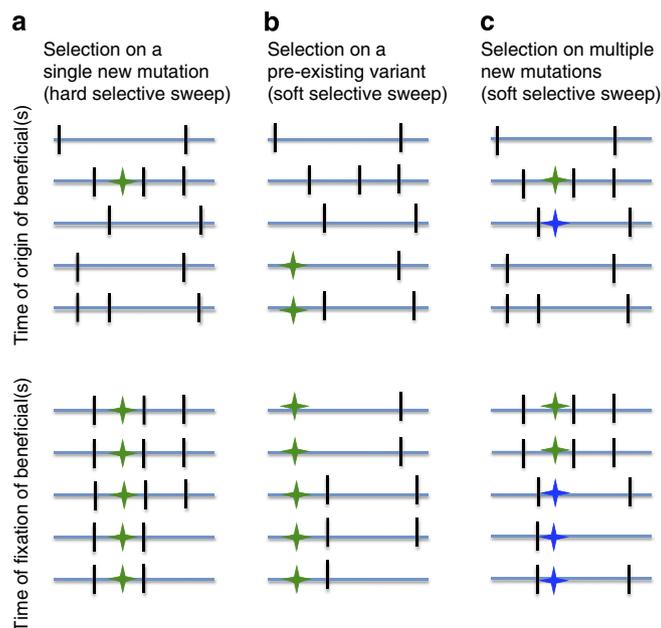
Underlying any understanding of the mode, tempo and relative importance of the adaptive process in the evolution of natural populations is the notion of whether adaptation is mutation limited. Two very different population genetic models have recently been proposed in which the rate of adaptation is not strongly limited by the rate at which newly arising beneficial mutations enter the population. However, empirical and experimental evidence to date challenges the recent enthusiasm for invoking these models to explain observed patterns of variation in humans and *Drosophila*.

Identifying the action of positive selection from genomic patterns of variation has remained a central focus in population genetics. This owes both to the importance of specific applications in fields ranging from ecology to medicine, but also to the desire to address more general evolutionary questions concerning the mode and tempo of adaptation. In this vein, the notion of a soft selective sweep has grown in popularity in the recent literature, and with this increasing usage the definition of the term itself has grown increasingly vague. A soft sweep does not reference a particular population genetic model *per se*, but rather a set of very different models that may result in similar genomic patterns of variation. Further, it is a term commonly used in juxtaposition with the notion of a hard selective sweep, the classic model in which a single novel beneficial mutation arises in a population and rises in frequency quickly to fixation. Patterns expected under the hard sweep model have been well described in the literature (see reviews of refs 1,2; Box 1), and consist of a reduction in variation surrounding the beneficial mutation owing to the fixation of the single haplotype carrying the beneficial, with resulting well-described skews in the frequency spectrum^{3–5} and in patterns of linkage disequilibrium^{6–8}. Indeed, a part of the recent popularity of soft sweeps comes from the seeming rarity of these expected hard sweep patterns in many natural populations (for example, see refs 9–11).

In terms of patterns of variation, the primary difference between soft and hard selective sweeps lies in the expected number of different haplotypes carrying the beneficial mutation or mutations, and thus in the expected number of haplotypes that hitchhike to appreciable frequency during the selective sweep, and which remain in the population at the time of fixation. This key difference results in different expectations in both the site frequency spectrum and in linkage disequilibrium, and thus in the many test statistics based on these patterns (see Box 1). Owing to this ambiguous definition, a number of models have been associated with producing a soft sweep pattern—including selection acting on previously segregating mutations, and multiple beneficials arising via mutation in quick succession (see review ref. 11 and Box 1).

¹School of Life Sciences, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne 1007, Switzerland. ²Swiss Institute of Bioinformatics (SIB), Lausanne 1007, Switzerland. Correspondence and requests for materials should be addressed to J.D.J. (email: jeffrey.jensen@epfl.ch).

Box 1 | Overview of two soft selective sweep models.



In contrast to a classic hard selective sweep (that is, selection on a single newly arising or rare beneficial mutation, **(a)**) I here discuss two models associated with soft selective sweeps. The first of these models, popularized by Hermisson and Pennings¹³, is one in which a given beneficial mutation previously segregated in the population neutrally (or at an appreciable frequency under mutation–selection balance), and thus existed on multiple haplotypes at the time of the selective shift in which the mutation became beneficial **(b)**. In this way, a single beneficial mutation may carry multiple haplotypes to intermediate frequency, while itself becoming fixed. Though Hermisson and Pennings¹³ associated this model with the term ‘soft sweep’, the model of selection on standing variation has been long considered in the literature. Orr and Betancourt¹² considered the model in some detail as will be discussed below, as did Innan and Kim⁸⁵. Indeed, Haldane⁸⁶ also discussed the possibility of selection acting on previously deleterious mutants segregating in the population.

A second commonly invoked model associated with soft selective sweeps, also popularized by Pennings and Hermisson⁸⁷, is one in which multiple beneficial mutations independently arise in short succession of one another—such that a second copy arises via mutation before the selective fixation of the first copy **(c)**. While this model was first envisioned as consisting of multiple identical beneficial mutations (that is, the identical change occurring at the same site), it has since been considerably expanded to include any mutation which produces an identical selective effect (for example, if all loss-of-function mutations produce an equivalent selective advantage, a large number of possible mutations may be considered as being identically beneficial⁸⁸). The similarity to the standing variation model described above, and thus their shared association with the notion of a soft sweep, is simply that these multiple beneficial mutations may arise on different haplotypes, and thus also sweep different genetic backgrounds to intermediate frequency.

In the cartoon above, the first row represents the time of the origin of the beneficial(s), in which five sampled chromosomes are shown with blue lines, each of which carries neutral mutations (black dashes) and some of which carry the beneficial mutation (given by the green star, with the blue star representing a second independently arising beneficial in the ‘multiple new mutation’ model). The second row represents the time of fixation of the beneficial mutation. In the hard sweep model **(a)**, the beneficial mutation as well as closely linked neutral variation has been brought to fixation, while more distant sites may only be brought to intermediate frequency owing to recombination. In the pre-existing variant model **(b)**, the beneficial mutation may carry the haplotypes on which it was segregating before the shift in selective pressure each to some intermediate frequency. In the multiple new mutation model **(c)**, each independent beneficial mutation may carry to intermediate frequency the haplotype on which it arose. Thus, the models **b** and **c** produce a qualitatively similar end result.

However, apart from shared expected patterns of variation, these two population genetic models are very different. Selection on standing variation requires that the beneficial mutation segregate at appreciable frequency in the preselection environment, whereas the multiple beneficial model requires a high mutation rate to the beneficial genotype. One important point that will be returned to throughout is the distinction between the relevance of these models themselves and the likelihood of these models resulting in a hard (that is, single haplotype) versus a soft (that is, multiple haplotype) selective sweep at the time of fixation. Below, I will discuss what is known from theory regarding these models, and what is known from experimental evolution and empirical population genetic studies regarding the values of the key parameters dictating their relevance. I conclude by arguing that the recent enthusiasm for invoking soft sweeps to

explain observed patterns of variation is likely to be largely unfounded in many cases.

Selection on standing variation

As described in Box 2, understanding the likelihood of a model of selection on standing variation requires knowing the frequency and fitness of beneficial mutations segregating in the population before becoming beneficial. Below, I will briefly review what is known from both experimental and empirical studies regarding these parameters in the handful of instances in which we have good inference.

Rare standing variants appear to contribute to adaptation. Orr and Betancourt¹² previously considered a model of selection on

standing variation and reached a similar result as Hermisson and Pennings¹³—namely, a soft sweep from standing variation only becomes feasible when the mutation has a non-zero probability of segregating at an appreciable frequency at the time of the selective shift (that is, the beneficial mutation was previously neutral or slightly deleterious and segregating under drift at relatively high frequency, it was maintained at appreciable frequency by balancing selection before the selective shift and so on). Indeed, they provide a direct calculation for the probability that multiple copies of the beneficial allele (X) exist, conditional on fixation of the beneficial mutation:

$$\Pr(X = i | X > 0) = \frac{e^{-4Nus_b/s_d} (4Nus_b/s_d)^i}{1 - e^{-4Nus_b/s_d}} \quad (1)$$

where s_d is the selection coefficient before the shift, s_b is the selection coefficient after the shift, and Nu is the population mutation rate.

Box 2 | Expectations and assumptions of a model of selection on standing variation.

A good deal is known from the theory literature regarding the likelihood of selection on standing variation, and the parameter space of relevance for this model. Kimura's⁸⁹ diffusion approximation gave the fixation probability of an allele (A) segregating in the population at frequency (x) with selective advantage (s_b):

$$\Pi_x(\alpha_b, h) \approx \frac{\int_0^x \exp[-\alpha_b(2hy + (1-2h)y^2)] dy}{\int_0^1 \exp[-\alpha_b(2hy + (1-2h)y^2)] dy}$$

where h is the dominance coefficient and $\alpha_b = 2N_e s_b$ (where N_e is the effective population size). Following Hermisson and Pennings¹³, if selection on the heterozygote is sufficiently strong (that is, $2h\alpha_b \gg (1-2h)/2h$), this may be approximated as:

$$\Pi_x(h\alpha_b) \approx \frac{1 - \exp[2h\alpha_b x]}{1 - \exp[2h\alpha_b]}$$

where for a new mutation (that is, $x = 1/2N$), we find Haldane's⁹⁰ result that the fixation probability is approximately twice the heterozygote advantage ($\Pi_{1/2N} \approx 2hs_b(N_e)/(N)$).

Comparing this result with the situation in which the beneficial allele was previously segregating neutrally (that is, $x > 1/2N$), Hermisson and Pennings¹³ obtained the following approximation:

$$\Pi_{\text{seg}}(h\alpha_b, N_e) \approx \frac{\ln(2h\alpha_b)}{\ln(2N_e)}$$

demonstrating that the fixation probability is much greater for beneficials that already begin at an appreciable frequency (because of their lower probability of being lost by drift)—though this condition may be misleading, as the segregation of a neutral allele at intermediate frequency is indeed already an unlikely event (which could be considered by integrating over the distribution of neutral variant frequencies). They further approximate the probability that the population adapts from standing variation (sgv) as:

$$P_{\text{sgv}}(h\alpha_b, h'\alpha_d, \Theta_\mu) = 1 - \exp(-\Theta_\mu \ln[1 + R_x])$$

where $R_x = 2h\alpha_b/(2h'\alpha_d + 1)$ measures the selective advantage of the allele in the new relative to the old environment (where α_b is the selective effect in the new environment and α_d in the previous environment, and h and h' are the dominance coefficients in the new and previous environments, respectively), and Θ_μ is the population mutation rate. Thus, for an allele at mutation-selection balance (that is, $x = \Theta_\mu/2h'\alpha_d$), this probability is simply $\approx 1 - \exp(-\Theta_\mu h\alpha_b/h'\alpha_d)$.

Thus, the key parameters for understanding the likelihood of a model of selection on standing variation involves knowing the preselection fitness of beneficial mutations (that is, before becoming beneficial), and relatedly the preselection frequency of beneficial mutations (again, before becoming beneficial).

With this, Orr and Betancourt made a notable observation that even if selection is acting on standing variation instead of a new mutation, a single copy is nonetheless surprisingly likely to sweep to fixation (that is, producing a hard, rather than a soft, sweep from standing variation). Indeed, they demonstrate that multiple-copy fixations become more likely than single-copy fixations from standing variation only when $4Nus_b/s_d > 1$. For reasonable parameter estimates, they calculate that the allele must be present in many copies in the population before obtaining an appreciable probability of sweeping multiple copies of the beneficial mutation to fixation. For example, from Orr and Betancourt, for $N = 1 \times 10^4$, $s_d = 0.05$, $s_b = 0.01$, $u = 10^{-5}$ and $h = 0.2$, 96% of the time a single copy will fix in the population, despite 20 copies segregating at mutation-selection balance before the shift in selection pressure. For these parameters, the population size must be in excess of $N = 1.5 \times 10^5$ (thus more than 300 copies segregating at mutation-selection balance) before multiple copies are more likely on fixation than a single copy.

Revisiting this model, Przeworski *et al.*¹⁴ more explicitly examined the frequency at which a mutation must be segregating before the shift in selection pressure, before multiple haplotypes would likely be involved in the selective sweep. They found that a hard sweep is likely when $x < 1/2N_e s_b$ (consistent with the simulated example from Orr and Betancourt above). Thus, taking the mutation-selection balance frequency given above, we may conclude that a hard sweep (that is, involving a single haplotype) is likely from standing variation when $\Theta_\mu/2h'\alpha_d < 1/2N_e s_b$ (see Fig. 1). An important distinction is again necessary here. While the parameter requirement mentioned above concerns the likelihood of a soft sweep from standing variation,

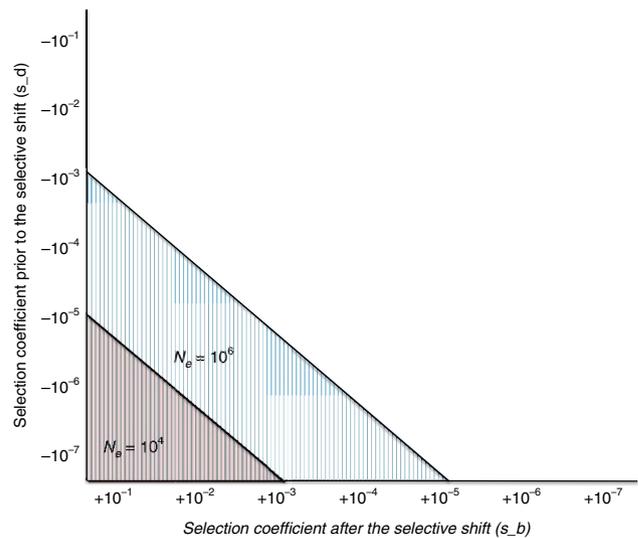


Figure 1 | The conditions under which a soft sweep from standing variation becomes possible. The y axis represents the selection coefficient before the selective shift (that is, given by negative selection coefficients) and the x axis is the selection coefficient after the shift in selective pressure (that is, given by positive selection coefficients). The area under each line represents the parameter space for which such a soft sweep is feasible for two different effective population sizes—one human like (10^4 , given by pink shading) and one *Drosophila* like (10^6 , given by vertical blue lines). As shown, the effect before the selective shift must be nearly neutral or weakly deleterious in order for the allele to segregate at an appreciable frequency, and the effect post selective shift must be strongly beneficial. As described in the text, this inference rests on the argument that $\Theta_\mu/2h'\alpha_d$ must be greater than $1/2N_e s_b$ for a soft selective sweep from standing variation to become likely, where here $\Theta_\mu = 10^{-8}$ and $h' = 1$.

it further suggests that we are unlikely to have statistical resolution when attempting to distinguish between a hard sweep on a new mutation versus a hard sweep on a rare previously standing variant.

As an empirical example of the above point, one of the most widely cited and discussed examples of selection on standing variation surrounds the *Eda* locus in Sticklebacks¹⁵. With evidence for selection reducing armour plating in freshwater populations compared with the ancestral heavily plated marine populations, the authors sequenced marine individuals to estimate the allele frequency of the freshwater adaptive low plate morphs, with estimates ranging from 0.2 to 3.8%. While the low plate morph is likely deleterious in marine populations (potentially suggesting that it is at mutation–selection balance), migration from the marine environment may indeed serve as an important source of variation for local freshwater hard sweeps. However, as noted by the authors, it is difficult to separate this hypothesis from that of local freshwater adaptation on new mutations, followed by back migration of locally adapted alleles into the marine population. Similarly, related arguments have been made for rare standing variation being responsible for the quick and persistent response of phenotypic traits to selection in the quantitative genetics literature (for a helpful review, see ref. 16). However, as with the above example, distinguishing rare standing ancestral variation from newly accumulating mutations has also been a topic of note¹⁷. Regardless of these caveats, hard sweeps from rare standing variants segregating at mutation–selection balance in ancestral populations, rather than on *de novo* mutations alone, appear to be an important and viable model of adaptation.

Quantifying the cost of beneficial mutations. On the basis of the simple and enlightening result of Orr and Betancourt, it is reasonable to ask, for cases in which we have reasonably strong functional evidence of adaptation, what we know about the value of $4Nus_b/s_d$ as this will dictate the likelihood of a hard versus a soft sweep from standing variation. There are two fields from which we may obtain insights—experimental–evolution studies in which the selective effects of mutations may be precisely measured under controlled environmental conditions and empirical population genetic studies in which inference can be drawn about the selective effect of functionally validated mutations in the presence and absence of a given selective pressure.

First, there is a rich literature in experimental evolution from which we can draw. In a recent evaluation of the distribution of fitness effects (DFE) in both the presence and absence of antibiotic in the bacterium *Pseudomonas fluorescens*, Kassen and Bataillon¹⁸ found that of the 665 resistance mutations isolated, greater than 95% were deleterious in the absence of the antibiotic treatment. In populations of yeast raised in both standard and challenging environments (in this case, high temperature and high salinity), Hietpas *et al.*¹⁹ identified a handful of beneficial mutations in each of the challenge environments, all of which were deleterious under standard conditions, with some even being lethal in the absence of the selective pressure. Foll *et al.*²⁰ in investigating the evolution of oseltamivir resistance mutations in the influenza A virus, identified 11 candidate resistance mutations, with the one functionally validated mutation (H274Y) having been demonstrated to be deleterious in the absence of drug pressure (see also refs 21,22).

Second, there are a small but increasing number of examples from natural populations where we have both a functionally validated beneficial mutation for which we understand the genotype–phenotype connection, as well as inference on the

selection coefficient both in the presence and absence of a given selective pressure. One such example is the evolution of cryptic colouration in wild populations of deer mice²³. In the Nebraska Sand Hills population, population genetic and functional evidence has been found for positive selection acting on a small number of mutations modifying different aspects of the cryptic phenotype, all contained within the *Agouti* gene region²⁴. Three lines of population genetic evidence suggest that selection began acting on these mutations when they arose (that is, selection on a *de novo* mutation): (1) the beneficial mutations appear to be carried on single haplotypes (though, as discussed above, selection on standing variation may indeed often only result in a single haplotype fixation), (2) the beneficial mutations have not been sampled off of the Sand Hills region (that is, the mutation is unlikely to have been segregating at appreciable frequency in the ancestral population before the formation of the Sand Hills) and (3) using an approximate Bayesian approach, the age of the selected mutation has been inferred to be younger than the geological age of the Sand Hills²³. In addition, ecological information pertaining to this phenotype exists as well. Performing a predation experiment with clay models, Linnen *et al.*²⁴ demonstrated a strong selective advantage of crypsis—with conspicuous models being subject to avian predation significantly more than cryptic models. This result suggests that if the beneficial phenotype currently present in the Sand Hills was indeed present in the ancestral population, it was likely to be strongly deleterious.

Other notable examples exist in the empirical literature as well. For example, Agren and Schemske²⁵ mapped quantitative trait loci for 398 recombinant inbred lines of *Arabidopsis* derived from crossing locally adapted lines from Sweden and Italy. Their results suggest a small number of locally adaptive genomic regions, and that in many cases the locally adaptive change was deleterious in the alternate environment. Performing a meta-analysis on a wide range of antibiotic-resistance mutations in pathogenic microbial populations, Melnyk *et al.*²⁶ found that across eight species and 15 drug treatments, resistance mutations were widely found to be deleterious in the absence of treatment (that is, in 19/21 examined studies). At the *Ace* locus of *D. melanogaster*, four described variants conferring varying degrees of pesticide resistance have been described, all of which are strongly deleterious in the absence of this pressure (with deleterious selection coefficients ranging from -5 to -20% ; see ref. 27).

Thus, given the required preselection frequency necessary to result in a soft rather than a hard sweep, it is fair to say that this combination of results provides poor support for the relevance of soft sweeps from standing variation in the populations examined. However, it is worth noting that such studies likely represent an ascertainment bias towards traits that strongly affect the phenotype, thus making them amenable for ecological and laboratory study. Assuming a relationship between the observed phenotypic and underlying selective effects, it may well be that beneficial mutations of small effects (which are more difficult to study and thus under-represented in the literature) may be those more likely to have only weakly deleterious effects in the absence of a given selective pressure.

Multiple competing beneficials

Box 3 describes the key parameters for understanding the likelihood of a model of competing beneficials—namely, the mutation rate to the beneficial genotype and the size of the mutational target available for creating an identical beneficial mutation. Below, I will briefly review what is known from both experimental and empirical studies regarding these parameters in the handful of instances in which we have good inference.

Box 3 | Expectations and assumptions of a model of multiple competing beneficial mutations.

As opposed to the standing variation model described in Box 2 in which there was a single mutational origin of the beneficial mutation, this model posits multiple beneficial mutations of independent origin arising in quick succession of one another. Importantly, these independent beneficial mutations must necessarily arise on different haplotypes. If the second independently arising identical beneficial mutation arises on the same background as the first, the resulting pattern would simply be that of a hard sweep as they would be impossible to distinguish (that is, a single haplotype would be swept).

Thus, the expected number of haplotypes and their frequency distributions is a necessary consideration. This expectation is given by Ewens' sampling formula (see ref. 91) for the case of no recombination. Given a mutation rate of Θ_b to the B allele, the probability to find k haplotypes occurring n_1, n_2, \dots, n_k times in a sample of size n is given as:

$$\Pr(n_1 \dots n_k | n, \Theta_b) = \frac{n!}{k! n_1 \dots n_k \Theta_b (\Theta_b + 1) \dots (\Theta_b + n - 1)} \Theta_b^k$$

For $k=1$ and $n_1=1$, Pennings and Hermisson⁸⁷ thus wrote the upper bound for the probability of such a soft sweep as:

$$P_{\text{soft},n} \leq 1 - \Pr(n | n, \Theta_b) = 1 - \prod_{i=1}^{n-1} \frac{i}{i + \Theta_b}$$

Hence, the probability of a soft sweep from multiple competing beneficials is naturally dependent on the beneficial population mutation rate. As they describe, the beneficial mutation rate must be extremely large before this model becomes likely (with $\Theta_b > 0.04$ before even two haplotypes would be swept with an appreciable probability, for $\alpha = 10,000$). In other words, within the relatively fast sojourn time of a beneficial mutation ($T_{\text{fix}} \approx 4N_e \log(\alpha)/\alpha$), multiple identical beneficial mutations must arise on separate haplotypes, escape drift (where, once again, the probability of fixation of each new independent mutation is given as $2hs_b(N_e)/(N)$) and ultimately fix in multiple copies.

Thus, the key parameter for understanding the likelihood of a model of competing beneficials involves knowing the mutation rate to the beneficial genotype. This latter clarification gives rise to an additional parameter—namely, the size of the mutational target available for creating an identical beneficial mutation. If only a single mutational change is possible, the target size is 1 bp. If, as given as an example in the Introduction, any loss-of-function mutation within a coding region produces an identical selective effect, the target size may be dozens of base pairs or more.

On the proportion of beneficial mutations. To understand the beneficial mutation rate is fundamentally to understand the DFE—that is, the proportions of newly arising mutations that are beneficial, neutral and deleterious. Characterizing this distribution has spawned a long and rich literature among both theoreticians and experimentalists. Fisher²⁸ had already considered the probability that a random mutation of a given phenotypic size would be beneficial, concluding that adaptations must consist primarily of small-effect mutations. Kimura²⁹ recognized one difficulty with this conclusion, noting that while small-effect mutations may be more likely to be adaptive, large-effect mutations have a higher probability of fixation. Thus, Kimura argued that, in fact, the intermediate-effect mutations may be most common in the adaptive process. Orr³⁰ gained an important additional insight—given any distribution of mutational effects, the distribution of factors fixed during an adaptive walk (that is, the sequential accumulation of beneficial mutations) is roughly exponential. An important by-product of this result is the notion that the first step of an adaptive walk may indeed be quite large (in agreement with Fisher's Geometric Model).

Efforts to quantify the shape of the DFE and characterize the beneficial mutation rate have come largely from the experimental

evolution literature. One common feature amongst this work is the use of extreme value theory (see review ref. 31). Because the DFE of new mutations is generally considered to be bimodal³²—consisting of a strongly deleterious mode and a nearly neutral mode—beneficial mutations represent the extreme tail of the mode centred around neutrality. One particular type of extreme value distribution—the Gumbel type (which contains a number of common distributions including normal, lognormal, gamma, and exponential)—has been of particular focus beginning with Gillespie³³.

Recently, experimental efforts have begun to better characterize the shape of the true underlying distribution in lab populations experiencing adaptive challenges (see review ref. 34). Though the fraction of beneficial mutations relative to the total mutation rate is indeed small, providing good support for the assumptions of extreme value theory, the exact shape of the beneficial distribution varies by study. Sanjuan *et al.*³⁵ found support for a gamma distribution using site-directed mutagenesis in vesicular stomatitis virus. Kassen and Bataillon¹⁸ found support for an exponential distribution assessing antibiotic-resistance mutations in *Pseudomonas*. Rokyta *et al.*³⁶ found support not for the Gumbel domain but rather for a distribution with a right-truncated tail (that is, suggesting that there is an upper bound on potential fitness effects), using two viral populations. MacLean and Buckling³⁷, again using *Pseudomonas*, argued that an exponential distribution well explained the data when the population was near optimum, but not when the population was far from optimum, owing to a long tail of strongly beneficial mutations. Schoustra *et al.*³⁸, working on the fungus *Aspergillus*, demonstrated that adaptive walks tend to be short, and characterized by an ever-decreasing number of available beneficial mutations with each mutational step taken. One important caveat in such experiments, however, is that they commonly begin from homogenous populations. Thus, while providing a good deal of insight into the underlying DFE, they are far from direct assessments of the relative role of single *de novo* beneficial mutations in adaptation.

Whole-genome time-sampled sequencing is also shedding light on the fraction of adaptive mutations. Examining resistance mutations in the influenza virus both in the presence and absence of oseltamivir (a common drug treatment), Foll *et al.*²⁰ identified the single and previously described resistance mutation (that is, H274Y (ref. 39)) as well as 10 additional putatively beneficial mutations based on duplicated experiments and population sequencing, suggesting a fraction of 11/13,588 potentially beneficial genomic sites in the presence of drug treatment, or 0.08% of the genome. But perhaps the most specific information currently available regarding beneficial mutation rates comes from experiments in which all mutations may be generated individually (as opposed to mutation-accumulation studies) and directly evaluated across different environmental conditions (see refs 19,40). Within this framework, Bank *et al.*⁴¹ recently evaluated all possible 560 individual mutations in a subregion of a yeast heat shock protein across six different environmental conditions (standard, as well as temperature and salinity variants), identifying few beneficials in the standard environment, and multiple beneficials associated with high salinity. To quantify this shift, the authors fit a Generalized Pareto Distribution, using the shape parameter (K) to summarize the changing DFE—with the Weibull domain fitting the less-challenging environments (that is, demonstrating that the DFE is right-bounded, suggesting that the populations are near optimum), and the Frechet domain fitting the challenging environment (that is, a heavy-tailed distribution owing to the presence of strongly beneficial mutations, potentially suggesting that the population is more distant from optimum).

Thus, despite some small but important differences between these conclusions, there is general support for a model in which newly arising mutations take a bimodal distribution, with the extreme right tail of this distribution representing putatively beneficial mutations. In other words, the beneficial mutation rate is likely a very small fraction of the total mutation rate. Given the requirements of the multiple competing beneficials model (that is, $\Theta_b > 0.04$), this would seemingly make the model only of relevance to populations of extremely large $N_e\mu$, as in perhaps certain viral populations (see ref. 42). Indeed, in an attempt to argue for the relevance of these models in *Drosophila*, Karasov *et al.*²⁷ claim an effective population size in *Drosophila* that is orders of magnitude larger than commonly believed (that is, $N_e > 10^8$), despite the great majority of empirical evidence to the contrary (see review ref. 43).

Small adaptive target size in natural populations. However, despite the above conclusion, if the mutational target size is not a single site, but rather a large collection of sites, this value may become more attainable for a wider array of species. As with the above section, the most abundant and reliably validated information comes from experimental evolution. However, this data is of course imperfect, as these studies do not necessarily reflect all potential available beneficial solutions (that is, mutation-accumulation studies can only draw inference on the mutations which happen to occur during the course of the experiment, and studies using direct-mutagenesis have thus far only evaluated sub-genomic regions). Returning to the examples given in the section above, we may ask what are the functional requirements of the identified beneficial mutations. In studying adaptation to the antibiotic rifampicin in the pathogen *Pseudomonas*, MacLean and Buckling³⁷ demonstrated that the beneficial mutations identified are consistent with known molecular interactions between rifampicin and RNA polymerase—as the antibiotic binds to a small pocket of the β -subunit of RNA polymerase, in which only 12 amino acid residues are involved in direct interaction. Wong *et al.*⁴⁴ investigated the genetics of adaptation to cystic fibrosis-like conditions in *Pseudomonas* both in the presence and absence of fluoroquinolone antibiotics, describing a small number of stereotypical resistance mutations in DNA gyrase. Examining the evolution of oseltamivir resistance in the influenza A virus, Foll *et al.*²⁰ described a similar story, in which a small handful of putatively beneficial resistance mutations are concentrated in haemagglutinin and neuraminidase, with the single-characterized resistance mutation being shown to alter the hydrophobic pocket of the neuraminidase active site, thus reducing affinity for drug.

Again considering the natural populations for which we have solid genotype–phenotype information and about which we understand something about the nature of adaptation acting on these mutations, let us consider a few examples. Describing widespread parallel evolution on armour plating in wild threespine sticklebacks, Colosimo *et al.*¹⁵ demonstrated the Ectodysplasin signalling pathways to be repeatedly targeted for modifications to this phenotype with a high degree of site-specific parallel evolution. Looking across 14 insect species that feed on cardenolide-producing plants, Zhen *et al.*⁴⁵ also noted repeated bouts of parallel evolution for dealing with this toxicity not only confined to the same alpha subunit of the sodium pump (ATP α), but in the great majority of cases to the same two amino acid positions. Examining adaptation to pesticide resistance in *Drosophila*, four specific point mutations in the *Ace* gene have been identified, which result in resistance to organophosphates and carbonates (see ref. 27). Cryptic colouration has also been a fruitful area, with specific mutations in the *Mcl1r* and *Agouti* gene regions having been described as the underlying cause of

adaptation for crypsis in mice of the Arizona/New Mexico lava flows⁴⁶, Nebraska Sand Hills²³ and the Atlantic coast⁴⁷, as well as in organisms ranging from the Siberian mammoth⁴⁸ to multiple species of lizards on the White Sands of New Mexico^{49,50} (and see review ref. 51 for further examples).

Thus, for the handful of convincing genotype-to-phenotype examples in the literature, the adaptive mutational target size appears small, a result which would appear to be biologically quite reasonable.

The effects of selection on linked sites. However, even if the mutational target size is sufficiently large such that a model of competing beneficials becomes feasible, it becomes necessary to consider interference between these segregating selected sites. It is helpful to consider three relevant areas of the parameter space: (a) beneficial mutations of identical selective effects arising in a low recombination rate region, potentially allowing for a soft sweep; (b) beneficial mutations of differing selective effects arising in a low recombination rate region, where the most strongly beneficial likely outcompetes the others producing a hard sweep; or (c) multiple beneficial mutations occurring in a high recombination rate region, allowing for a hard sweep of the most beneficial haplotype (that is, the recombinant carrying the most beneficial mutations).

Hill and Robertson⁵² explicitly considered the probability of fixation for two segregating beneficial mutations. Confirming the arguments of Fisher²⁸, they demonstrated that selection at one locus indeed interferes with selection at the alternate locus, reducing the probability of fixation at both sites—with the conclusion being that simultaneous selection at more than one site reduces the overall efficacy of selection (see also refs 53,54).

This effect is clearly a function of the amount of recombination between the selected sites. If the sites are independent there is no such effect, and if they are tightly linked the effect will be very strong (Fig. 2). While it is difficult to generalize this information, for the current empirical data available discussed above, it appears both likely and biologically reasonable to consider that mutations conferring identical selective effects may indeed be occurring within a narrow genomic region (for example, mutations within the drug-binding pocket haemagglutinin in influenza virus in response to drug, within RNA polymerase in *Pseudomonas* in response to antibiotic, at the *Agouti/Mcl1r* locus in vertebrates for colour modifications, at the *Ace* locus of *Drosophila* for pesticide resistance, at the *Eda* locus in Sticklebacks for armour modifications and so on).

Examining the extent of this effect by simulation, Comeron *et al.*⁵⁵ found that the effect becomes stronger as (1) the sites become more weakly beneficial, (2) the recombination rate is decreased and (3) the number of selected sites increases (consistent with the results of refs 56,57). However, as long as there is linkage between the sites, the probability of fixation decreases rapidly as the number of selected sites grows, even for very strong selection. Examining the effect of two competing beneficial mutations in the presence of recombination analytically, Yu and Etheridge⁵⁸ further demonstrate the relative likelihood of an ultimate single haplotype fixation.

Thus, while a large mutational target size may, in principle, increase the relevance of this model, it results in a scenario still requiring a large beneficial mutation rate, necessitates that these beneficial mutations escape initial stochastic loss and finally, owing to interference, results in a decreased probability of fixation for each competing beneficial relative to independence. Again invoking results from experimental evolution, in a highly informative recent study by Lee and Marx⁵⁹, the authors demonstrate the strong effects of clonal interference in

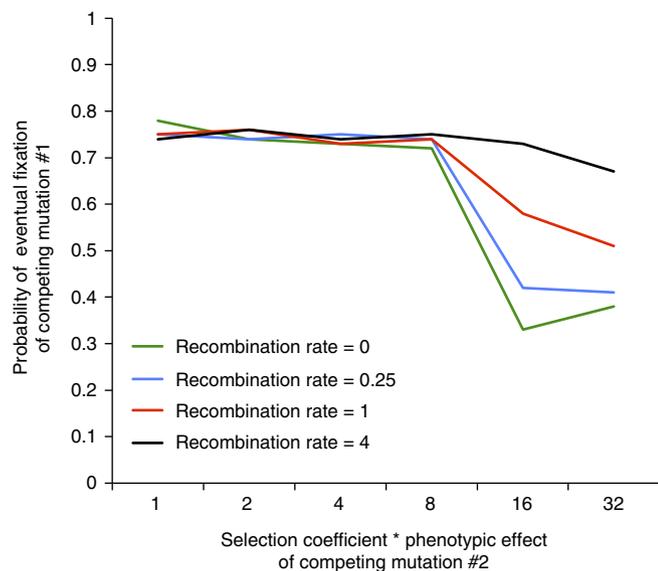


Figure 2 | Probability of fixation of a first beneficial mutation under a scenario of two competing beneficial mutations. Representation of the results of Hill and Robertson⁵² (modified from their Fig. 2). On the y axis is the probability of fixation of the first beneficial mutation, and on the x axis is the selection coefficient multiplied by the phenotypic effect of the second (and thus competing) beneficial mutation. Here the selective effect of mutation #1 is 8, on the scale given on the x axis, and each beneficial mutation is assumed to be at a 10% frequency in the population. The four lines represent four different recombination rates between the two sites. As shown, if the effect of the second site is weaker than the first (that is, < 8), the first beneficial retains a high probability of fixation and the second site will be lost unless it can recombine on to a common haplotype. However, as the effect of the second beneficial mutation becomes stronger, the probability of fixation of the first beneficial mutation decreases rapidly—a phenomenon that is increasingly strong with tighter genetic linkage between the sites. Thus, for rare to moderate recombination, and a strong beneficial effect of the second site, the probability of fixation of the first beneficial mutation is halved even for this simple case of only two competing beneficials.

replicated populations of *Methylobacterium extorquens*—identifying as many as 17 simultaneous beneficial mutations existing in a population which may rise in frequency initially, only to be lost owing to competition with an alternate and ultimately successful single beneficial mutation, in what they termed repeated ‘failed soft sweeps.’

As a natural population example, Hedrick⁶⁰ discusses the multiple identified malaria resistance variants identified in humans, and makes the case that in the continued presence of malaria, single variants are highly likely to ultimately fix at the cost of losing other competing and currently segregating beneficial resistance mutations, owing to measured selection differentials. Similarly, at the previously discussed *Ace* locus of *D. melanogaster*, a single of the four identified resistance mutations was found to confer 75% resistance to pesticide, two mutations confer 80% resistance and three mutations confer full resistance—again suggesting that a single haplotype carrying multiple beneficial mutations will likely ultimately result in a hard selective sweep. Both of these observations, along with the results of Lee and Marx⁵⁹, suggest that multiple competing beneficial mutations may indeed be a likely model, but a soft sweep from multiple beneficials is unlikely owing to non-equivalent selective effects between the mutations (or the haplotypes carrying these mutations). Thus, as with the model

of selection on standing variation above—the model of competing beneficial mutations itself has good empirical and experimental support for being relevant, but a hard sweep rather than a soft sweep appears as the more likely outcome given our current understanding of the parameters of relevance.

Perspective and future directions

Apart from the considerations discussed in the sections above, and conditional on the unsubstantiated assumption that adaptive fixations are common, the absence of hard sweep patterns in many natural populations has led some to conclude that soft sweeps must be the primary mechanism of adaptation, with a recent popularity for invoking these models in the human and *Drosophila* literature. However, as argued above, this assumption is poorly supported, and theoretical and experimental insights to date suggest that soft sweeps from standing variation or from multiple beneficial mutations for populations of this size are unlikely. This argument itself is of course somewhat circular, as quantifying the fraction of adaptively fixed mutations, and the proportion of newly arising beneficial mutations, is indeed one of the central focal points of population genetics, and is far from resolved as discussed. Thus, assuming a very large fraction of adaptive fixations to quantify the fraction of adaptive fixations is rather self-defeating.

A quite separate point has also been neglected in this literature. Namely, the power of existing tests of hard selective sweeps to identify these patterns within demographically complex populations (a category that certainly includes humans and *Drosophila*). Biswas and Akey⁶¹ examined the consistency between methods used for conducting genomic scans for beneficial mutations in humans. Results differed dramatically, ranging from 1,799 genes identified by Wang *et al.*⁶² to 27 genes identified by Altschuler *et al.*⁶³ Perhaps even more striking, of the six studies examined, there was virtually no overlap in the genes identified. For example, of the 1,799 genes identified by Wang *et al.*⁶², 125 overlap with the scan of Voight *et al.*⁶⁴, 47 from Carlson *et al.*⁶⁵, 5 from Altschuler *et al.*⁶², 4 from Nielsen *et al.*⁶⁶ and 40 from Bustamante *et al.*⁶⁷ In addition, the recent review by Crisci *et al.*², summarizing estimates of the rate of adaptive fixation in *Drosophila*, noted that the inferred genomic rate differs by two orders of magnitude between studies (from $\lambda = 1.0E - 12$ (ref. 68) to $\lambda = 1.0E - 11$ (refs 69,70) to $\lambda = 1.0E - 10$ (ref. 71)—where $2N_e\lambda$ is the rate of beneficial fixation per base pair per $2N_e$ generations).

Evaluating the performance of these statistics has thus remained an important question, and over the past decade numerous researchers have demonstrated low power under a wide range of neutral non-equilibrium models^{72–76}. More recently, Crisci *et al.*⁷⁷ specifically evaluated the ability of the most widely used and sophisticated tests of selection via simulation (Sweepfinder⁶⁶, SweeD⁷⁶ and OmegaPlus⁷⁸), to identify both complete and incomplete hard selective sweeps under a variety of demographic models of relevance for human and *Drosophila* populations. The results are troubling, with the true positive rate rarely exceeding 50% even under equilibrium models, and being considerably worse for models of moderate and severe population size reductions (Fig. 3). Furthermore, the false positive rate was often in excess of power, particularly for models of population bottlenecks. Though not conclusive, this indeed suggests a troubling potential interpretation for the lack of overlap between the above mentioned genomic scan studies.

If nothing else, these results demonstrate that the absence of evidence is not evidence of absence for the hard sweep model—implying that we only have minimal power to detect even very recent and very strong hard selective sweeps in these populations, and essentially no power for the great majority of the parameter

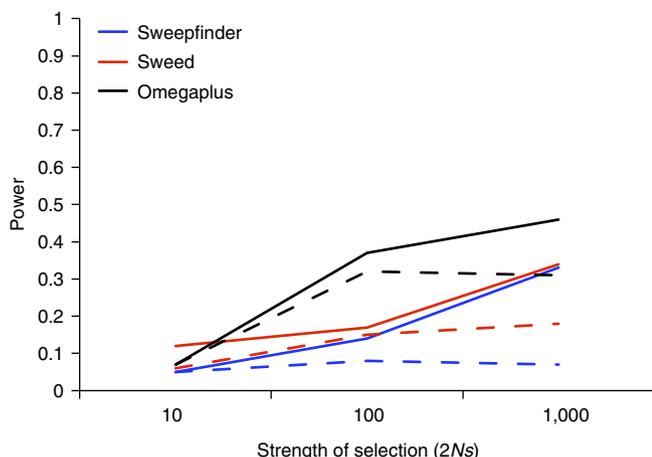


Figure 3 | Statistical power to detect hard selective sweeps.

Representation of the results of Crisci *et al.*⁷⁷ examining the power of three of the most commonly used approaches for detecting hard selective sweeps—Sweepfinder (blue), Sweed (red) and Omegaplus (black). On the y axis is the power of the test statistic and on the x axis is the strength of selection. Two models are plotted: (1) a selective sweep in an equilibrium population (given by the solid lines) and (2) a selective sweep in a moderately bottlenecked population, in which the population is reduced to 10% of its former size 0.2 $4N$ generations in the past (given by the dashed lines). In all cases, $N_e = 10^4$ and the selective event occurred 0.01 $4N$ generations in the past. As shown, even in equilibrium populations, the power scarcely reaches 50% even for strong selection, while in mildly bottlenecked populations the power approaches nominal levels for frequency spectrum based statistics (that is, Sweed and Sweepfinder) and drops to 30% for the better performing linkage disequilibrium-based approaches (that is, Omegaplus).

space. However, concerning these results may be, it is important that the field has made the effort to quantify the performance of the test statistics designed for detecting hard sweeps—defining Type-I and Type-II error and examining performance in demographic models both with and without selection. This scrutiny has yet to be brought to soft sweep expectations and statistics. Before these models can be reasonably invoked as explanations for observations in natural populations, we need to similarly understand the ability of neutral demographic models to replicate soft sweep patterns, quantify our ability to identify soft sweeps from standing variation and from multiple beneficial mutations in non-equilibrium populations and understand the effects of relaxing current assumptions involving linkage and epistasis (that is, for selection on standing variation, the assumption is made that a single beneficial mutation will have the same selective effects on all genetic backgrounds, and the multiple beneficial model assumes that there are no epistatic interactions between co-segregating mutations). Early efforts have been made in some of these areas, with recent work examining basic expectations of these models under fluctuating effective population sizes, resulting in a further description of how population size changes may result in the ultimate fixation of a single beneficial mutation⁷⁹.

In conclusion, the wide array of genomic patterns of variation that may be accounted for by models associated with soft selective sweeps has allowed adaptive explanations to proliferate in the literature, and be invoked for a larger subset of genomic data. However appealing this may be, these models in fact carry with them very specific and well-understood parameter requirements. Further, the ability of alternate models to produce these patterns needs to be more carefully weighed in future studies,

particularly given preliminary findings concerning similar patterns produced under both neutral demographic models⁷⁷ and models of background selection⁸⁰. Indeed, alternative models of positive selection have also been suggested to produce qualitatively similar patterns—including hard selective sweeps in subdivided populations exchanging migrants^{81–83} and polygenic adaptation⁸⁴.

Finally, while examples in the literature are accumulating in support of the models themselves (for example, selection on standing variation at the *Eda* locus of Sticklebacks or selection on multiple beneficials at the *Ace* locus of *Drosophila*), there is very little evidence of soft sweep fixations, with the best empirical and experimental examples to date almost universally pointing to hard sweep fixations under these models. This appears to primarily be owing to the low preselection allele frequency of the standing variants (which are seemingly often deleterious before the shift in selective pressure), and to the selective differential between competing beneficial mutations (or between the haplotypes carrying the beneficial mutations) resulting in the ultimate fixation of only a single haplotype. Thus, while the models themselves certainly deserve further attention, theoretical, empirical and experimental results to date suggest that the field ought to take greater caution when invoking soft sweep fixations, as hard sweep fixations (be it from models of selection on new mutations, standing variation or competing beneficial mutations) seem to remain as the most likely outcome across a wide parameter space relevant for many current populations of interest.

References

- Nielsen, R. Molecular signatures of natural selection. *Annu. Rev. Genet.* **39**, 197–218 (2005).
- Crisci, J., Poh, Y.-P., Bean, A., Simkin, A. & Jensen, J. D. Recent progress in polymorphism-based population genetic inference. *J. Hered.* **103**, 287–296 (2012).
- Braverman, J. M., Hudson, R. R., Kaplan, N. L., Langley, C. H. & Stephan, W. The hitchhiking effect on the site frequency spectrum of DNA polymorphisms. *Genetics* **140**, 783–796 (1995).
- Simonsen, K. L., Churchill, G. A. & Aquadro, C. F. Properties of statistical tests of neutrality for DNA polymorphism data. *Genetics* **141**, 413–429 (1995).
- Fay, J. C. & Wu, C.-I. Hitchhiking under positive Darwinian selection. *Genetics* **155**, 1405–1413 (2000).
- Stephan, W., Song, Y. S. & Langley, C. H. The hitchhiking effect on linkage disequilibrium between linked neutral loci. *Genetics* **172**, 2647–2663 (2006).
- McVean, G. The structure of linkage disequilibrium around a selective sweep. *Genetics* **175**, 1385–1406 (2007).
- Jensen, J. D., Thornton, K. R., Bustamante, C. D. & Aquadro, C. F. On the utility of linkage disequilibrium as a statistic for identifying targets of positive selection in non-equilibrium populations. *Genetics* **176**, 2371–2379 (2007).
- Pritchard, J. K., Pickrell, J. K. & Coop, G. The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. *Curr. Biol.* **20**, R208–R215 (2010).
- Hernandez, R. D. *et al.* Classic selective sweeps were rare in recent human evolution. *Science* **331**, 920–924 (2011).
- Messer, P. W. & Petrov, D. A. Population genomics of rapid adaptation by soft selective sweeps. *Trends Ecol. Evol.* **28**, 659–669 (2013).
- Orr, H. A. & Betancourt, A. J. Haldane's sieve and adaptation from standing genetic variation. *Genetics* **157**, 875–884 (2001).
- A highly significant early contribution to the selection on standing variation literature, the authors present a number of important results not yet fully appreciated in the empirical soft sweep literature – including the necessary pre-selection allele frequency necessary to result in a soft sweep fixation.*
- Hermisson, J. & Pennings, P. S. Soft sweeps: molecular population genetics of adaptation from standing genetic variation. *Genetics* **169**, 2335–2352 (2005).
- The first in a series of papers exploring soft sweeps, the authors develop both theory and expectations for a model of selection on standing variation.*
- Przeworski, M., Coop, G. & Wall, J. D. The signature of positive selection on standing genetic variation. *Evolution* **59**, 2312–2323 (2005).
- Colosimo, P. F. *et al.* Widespread parallel evolution in sticklebacks by repeated fixation of Ectodysplasin alleles. *Science* **307**, 1928–1933 (2005).
- Hill, W. G. Understanding and using quantitative genetic variation. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **365**, 73–85 (2010).

17. Keightley, P. D. & Hill, W. G. Quantitative genetic variation in body size of mice from new mutation. *Genetics* **131**, 693–700 (1992).
18. Kassen, R. & Bataillon, T. The distribution of fitness effects among beneficial mutations prior to selection in experimental populations of bacteria. *Nat. Genet.* **38**, 484–488 (2006).
19. Hietpas, R. T., Bank, C., Jensen, J. D. & Bolon, D. N. Shifting fitness landscapes in response to altered environments. *Evolution* **67**, 3512–3522 (2013).
20. Foll, M. *et al.* Influenza virus drug resistance: a time-sampled population genetics perspective. *PLoS Genet.* **10**, e1004185 (2014).
21. Ginting, T. E. *et al.* Amino acid changes in hemagglutinin contribute to the replication of oseltamivir-resistant H1N1 influenza viruses. *J. Virol.* **86**, 121–127 (2012).
22. Renzette, N. *et al.* Evolution of the influenza A virus genome during development of oseltamivir resistance *in vitro*. *J. Virol.* **88**, 272–281 (2014).
23. Linnen, C. R., Kingsley, E. P., Jensen, J. D. & Hoekstra, H. E. On the origin and spread of an adaptive allele in *Peromyscus* mice. *Science* **325**, 1095–1098 (2009).
24. Linnen, C. R. *et al.* Adaptive evolution of multiple traits through multiple mutations at a single gene. *Science* **339**, 1312–1316 (2013).
25. Agren, J. & Schemske, D. W. Reciprocal transplants demonstrate strong adaptive differentiation of the model organism *Arabidopsis thaliana* in its native range. *New Phytol.* **194**, 1112–1122 (2013).
26. Melnyk, A., Wong, A. & Kassen, R. The fitness costs of antibiotic resistance mutations. *Evol. Appl.* doi:10.1111/eva.12196 (2014).
- A helpful and informative overview of the antibiotic resistance literature pertaining to inferring fitness costs of resistance mutations in the absence of treatment.**
27. Karasov, T., Messer, P. W. & Petrov, D. A. Evidence that adaptation in *Drosophila* is not limited by mutation at single sites. *PLoS Genet.* **6**, e1000924 (2010).
28. Fisher, R. A. *The Genetical Theory of Natural Selection* (Clarendon Press, 1930).
29. Kimura, M. *The Neutral Theory of Molecular Evolution 1986* (Cambridge Univ. Press, 1983).
30. Orr, H. A. The population genetics of adaptation: the distribution of factors fixed during adaptive evolution. *Evolution* **52**, 935–949 (1998).
31. Orr, H. A. The genetic theory of adaptation: a brief history. *Nat. Rev. Genet.* **6**, 119–127 (2005).
32. Ohta, T. Slightly deleterious mutant substitutions in evolution. *Nature* **246**, 96–98 (1973).
33. Gillespie, J. H. The molecular clock may be an episodic clock. *Proc. Natl Acad. Sci. USA* **81**, 8009–8013 (1984).
34. Eyre-Walker, A. & Keightley, P. D. The distribution of fitness effects of new mutations. *Nat. Rev. Genet.* **8**, 610–618 (2007).
35. Sanjuan, R., Moya, A. & Elena, S. F. The distribution of fitness effects caused by single-nucleotide substitutions in an RNA virus. *Proc. Natl Acad. Sci. USA* **101**, 8396–8401 (2004).
36. Rokyta, D. R. *et al.* Beneficial fitness effects are not exponential for two viruses. *J. Mol. Evol.* **67**, 368–376 (2008).
37. Maclean, R. C. & Buckling, A. The distribution of fitness effects of beneficial mutations in *Pseudomonas aeruginosa*. *PLoS Genet.* **5**, e1000406 (2009).
38. Schoustra, S. E., Bataillon, T., Gifford, D. R. & Kassen, R. The properties of adaptive walks in evolving populations of fungus. *PLoS Biol.* **7**, e1000250 (2009).
39. Ives, J. A. *et al.* The H274Y mutation in the influenza A/H1N1 neuraminidase active site following oseltamivir phosphate treatment leaves virus severely compromised both *in vitro* and *in vivo*. *Antiviral Res.* **55**, 307–317 (2002).
40. Hietpas, R. T., Jensen, J. D. & Bolon, D. N. Experimental dissection of a fitness landscape. *Proc. Natl Acad. Sci. USA* **108**, 7896–7901 (2011).
41. Bank, C., Hietpas, R. T., Wong, A., Bolon, D. N. & Jensen, J. D. A Bayesian MCMC approach to assess the complete distribution of fitness effects of new mutations: uncovering the potential for adaptive walks in challenging environments. *Genetics* **196**, 841–852 (2014).
- Here the authors develop an approach for statistically characterizing the DFE and apply it to an experimental dataset consisting of all possible point mutations in a sub-genomic region—describing how the DFE and number/size of beneficial mutations change under differing environmental conditions.**
42. Pennings, P. S., Kryazhimskiy, S. & Wakeley, J. Loss and recovery of genetic diversity in adapting populations of HIV. *PLoS Genet.* **10**, e1004000 (2014).
43. Charlesworth, B. Effective population size and patterns of molecular evolution and variation. *Nat. Rev. Genet.* **10**, 195–205 (2009).
44. Wong, A., Rodrigue, N. & Kassen, R. Genomics of adaptation during experimental evolution of the opportunistic pathogen *Pseudomonas aeruginosa*. *PLoS Genet.* **8**, e1002928 (2012).
45. Zhen, Y., Aardema, M. L., Medina, E. M., Schumer, M. & Andolfatto, P. Parallel molecular evolution in an herbivore community. *Science* **6102**, 1634–1637 (2012).
46. Nachman, M. W., Hoekstra, H. E. & D'Agostino, S. L. The genetic basis of adaptive melanism in pocket mice. *Proc. Natl Acad. Sci. USA* **100**, 5268–5273 (2003).
47. Steiner, C. C., Weber, J. N. & Hoekstra, H. E. Adaptive variation in beach mice produced by two interacting pigmentation genes. *PLoS Biol.* **5**, e219 (2007).
48. Rompler, H. *et al.* Nuclear gene indicates coat-color polymorphism in mammoths. *Science* **313**, 62 (2006).
49. Rosenblum, E. B., Hoekstra, H. E. & Nachman, M. W. Adaptive reptile color variation and the evolution of the *Mcl1r* gene. *Evolution* **58**, 1794–1808 (2004).
50. Rosenblum, E. B., Rompler, H., Schoneberg, T. & Hoekstra, H. E. Molecular and functional basis of phenotypic convergence in white lizards at White Sands. *Proc. Natl Acad. Sci. USA* **107**, 2113–2117 (2010).
51. Manceau, M. V. S., Domingues, L., Linnen, C. R., Rosenblum, E. B. & Hoekstra, H. E. Convergence in pigmentation at multiple levels: mutations, genes and function. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **365**, 2439–2450 (2010).
52. Hill, W. G. & Robertson, A. The effect of linkage on limits to artificial selection. *Genet. Res.* **8**, 269–294 (1966).
- A classic paper examining the effects of linkage on the efficiency of selection, providing important results that must now be better grappled with in considering models of multiple co-segregating beneficial mutations.**
53. Felsenstein, J. The evolutionary advantage of recombination. *Genetics* **78**, 737–756 (1974).
54. Birky, Jr C. W. & Walsh, J. B. Effects of linkage on rates of molecular evolution. *Proc. Natl Acad. Sci. USA* **85**, 6414–6418 (1988).
55. Comeron, J. M., Willford, A. & Kliman, R. M. The Hill-Robertson effect: evolutionary consequences of weak selection and linkage in finite populations. *Heredity* **100**, 19–31 (2008).
- A helpful review of both data and theory pertaining to the effects of selection on linked sites.**
56. McVean, G. & Charlesworth, B. The effects of Hill-Robertson interference between weakly selected mutations on patterns of molecular evolution and variation. *Genetics* **155**, 929–944 (2000).
57. Comeron, J. M. & Kreitman, M. Population, evolutionary and genomic consequences of interference selection. *Genetics* **161**, 389–410 (2002).
58. Yu, F. & Etheridge, A. M. The fixation probability of two competing beneficial mutations. *Theor. Pop. Biol.* **78**, 36–45 (2010).
- In many ways building on the important work of Hill and Robertson, the authors take an analytical approach to examine a model of two competing beneficial mutations in the presence of recombination, and describe the probabilities of single vs. multiple mutational copies at the time of fixation.**
59. Lee, M.-C. & Marx, C. J. Synchronous waves of failed soft sweeps in the laboratory: remarkably rampant clonal interference of alleles at a single locus. *Genetics* **193**, 943–952 (2013).
- An informative experimental exploration of multiple competing beneficial mutations in methylobacterium – in which the authors identify multiple co-segregating beneficial mutations which ultimately result in a single-copy fixation (i.e., hard selective sweep) owing to non-equivalent selective effects.**
60. Hedrick, P. Population genetics of malaria resistance in humans. *Heredity* **107**, 283–304 (2011).
61. Biswas, S. & Akey, J. M. Genomic insights into positive selection. *Trends Genet.* **22**, 437–446 (2006).
62. Wang, E. T. *et al.* Global landscape of recent inferred Darwinian selection for *Homo sapiens*. *Proc. Natl Acad. Sci. USA* **103**, 135–140 (2006).
63. Altshuler, D. *et al.* A haplotype map of the human genome. *Nature* **437**, 1299–1320 (2005).
64. Voight, B. F., Kudravalli, S., Wen, X. & Pritchard, J. K. A map of recent positive selection in the human genome. *PLoS Biol.* **4**, e72 (2006).
65. Carlson, C. S. *et al.* Genomic regions exhibiting positive selection identified from dense genotype data. *Genome Res.* **15**, 1553–1565 (2005).
66. Nielsen, R. *et al.* Genomic scans for selective sweeps using SNP data. *Genome Res.* **15**, 1566–1575 (2005).
67. Bustamante, C. D. *et al.* Natural selection on protein coding genes in the human genome. *Nature* **437**, 1153–1157 (2005).
68. Macpherson, J. M., Sella, G., Davis, J. C. & Petrov, D. A. Genome-wide spatial correspondence between nonsynonymous divergence and neutral polymorphism reveals extensive adaptation in *Drosophila*. *Genetics* **177**, 2083–2099 (2007).
69. Li, H. & Stephan, W. Inferring the demographic history and rate of adaptive substitution in *Drosophila*. *PLoS Genet.* **2**, e166 (2006).
70. Jensen, J. D., Thornton, K. R. & Andolfatto, P. An approximate Bayesian estimator suggests strong recurrent selective sweeps in *Drosophila*. *PLoS Genet.* **4**, e1000198 (2008).
71. Andolfatto, P. Hitchhiking effects of recurrent beneficial amino acid substitutions in the *Drosophila melanogaster* genome. *Genome Res.* **17**, 1755–1762 (2007).
72. Przeworski, M. The signature of positive selection at randomly chosen loci. *Genetics* **160**, 1179 (2002).
73. Jensen, J. D., Kim, Y., Bauer DuMont, V., Aquadro, C. F. & Bustamante, C. D. Distinguishing between selective sweeps and demography using DNA polymorphism data. *Genetics* **170**, 1401–1410 (2005).
74. Teshima, K. M., Coop, G. & Przeworski, M. How reliable are empirical genomic scans for selective sweeps? *Genome Res.* **16**, 702–712 (2006).

75. Thornton, K. R. & Jensen, J. D. Controlling the false positive rate in multi-locus genome scans for selection. *Genetics* **175**, 737–750 (2007).
76. Pavlidis, P., Zivkovic, D., Stamatakis, A. & Alachiotis, N. SweeD: likelihood-based detection of selective sweeps in thousands of genomes. *Mol. Biol. Evol.* **30**, 2224–2234 (2013).
77. Crisci, J., Poh, Y.-P., Mahajan, S. & Jensen, J. D. On the impact of equilibrium assumptions on tests of selection. *Front. Genet.* **4**, 235 (2013).
78. Alachiotis, N., Stamatakis, A. & Pavlidis, P. OmegaPlus: a scalable tool for rapid detection of selective sweeps in whole-genome datasets. *Bioinformatics* **28**, 2274–2275 (2012).
79. Wilson, B. A., Petrov, D. A. & Meyser, P. W. Soft selective sweeps in complex demographic scenarios. *Genetics* 25060100 (2014).
80. Alves, I., Sramkova, H., Foll, M. & Excoffier, L. Genomic data reveal a complex making of humans. *PLoS Genet.* **8**, e1002837 (2012).
81. Slatkin, M. Gene flow and selection in a two-locus system. *Genetics* **81**, 787–802 (1975).
82. Kim, Y. & Maruki, T. Hitchhiking effects of a beneficial mutation spreading in a subdivided population. *Genetics* **189**, 213–226 (2011).
83. Coop, G. & Ralph, P. Parallel adaptation: one or many waves of advance of an advantageous allele? *Genetics* **186**, 647–668 (2010).
84. Pritchard, J. K. & DiRienzo, A. Adaptation – not by sweeps alone. *Nat. Rev. Genet.* **11**, 665–667 (2010).
85. Innan, H. & Kim, Y. Pattern of polymorphism after strong artificial selection in a domestication event. *Proc. Natl Acad. Sci. USA* **101**, 10667–10672 (2004).
86. Haldane, J. B. S. The cost of natural selection. *Genetics* **55**, 511–524 (1957).
87. Pennings, P. S. & Hermisson, J. Soft sweeps II – molecular population genetics of adaptation from recurrent mutation or migration. *Mol. Biol. Evol.* **23**, 1076–1084 (2006).
88. Pennings, P. S. & Hermisson, J. Soft sweeps III: the signature of positive selection from recurrent mutation. *PLoS Genet.* **2**, e186 (2006).
89. Kimura, M. Some problems of stochastic processes in genetics. *Ann. Math. Stat.* **28**, 882–901 (1957).
90. Haldane, J. B. S. The mathematical theory of natural and artificial selection. *Proc. Camb. Philos. Soc.* **23**, 838–844 (1927).
91. Ewens, W. J. *Mathematical Population Genetics* 2nd edn (Springer-Verlag, 2004).

Acknowledgements

I would like to thank Chip Aquadro, Roman Arguello, Dan Bolon, Margarida Cardoso Moreira, Brian Charlesworth, Laurent Excoffier, Adam Eyre-Walker, Joanna Kelley, Tim Kowalik, Anna-Sapfo Malaspina, Bret Payseur, Molly Przeworski, Nadia Singh, Wolfgang Stephan and Alex Wong for helpful comments and suggestions on an earlier version. I would also like to thank the authors of Melnyk, Wong and Kassen for sharing their manuscript while in review. I would finally like to thank members of the Jensen Lab for insightful comment and discussion throughout the writing process, in particular Claudia Bank, Anna Ferrer Admetlla, Matthieu Foll, Stefan Laurent, Louise Ormond, Cornelia Pokalyuk and Nick Renzette. J.D.J. is funded by grants from the Swiss National Science Foundation, and a European Research Council (ERC) Starting Grant.

Additional information

Competing financial interests: The author declares no competing financial interests.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

How to cite this article: Jensen, J. D. On the unfounded enthusiasm for soft selective sweeps. *Nat. Commun.* 5:5281 doi: 10.1038/ncomms6281 (2014).

The second contribution in the series of soft sweeps papers by the authors, this work explores theory and expectations under models with a rapid beneficial input in to the population via mutation or migration.