Published in final edited form as:
Nat Rev Genet. 2013 April ; 14(4): 262–274. doi:10.1038/nrg3425.

Genomic signatures of selection at linked sites: unifying the disparity among species

Asher D. Cutter[1,2] and **Bret A. Payseur**[3]

[1]Department of Ecology and Evolutionary Biology, University of Toronto, Toronto, Ontario M5S 3B2, Canada

[2]Centre for the Analysis of Genome Evolution and Function, University of Toronto, Toronto, Ontario M5S 3B2, Canada

[3]Laboratory of Genetics, University of Wisconsin, Madison, Wisconsin 53706, USA

Abstract

Population genetics theory supplies powerful predictions about how natural selection interacts with genetic linkage to sculpt the genomic landscape of nucleotide polymorphism. Both the spread of beneficial mutations and removal of deleterious mutations act to depress polymorphism levels, especially in low-recombination regions. However, empiricists have documented extreme disparities among species. Here we characterize the dominant features that could drive variation in linked selection among species, including roles for selective sweeps being ‘hard’ or ‘soft’, and concealing by demography and genomic confounds. We advocate targeted studies of close relatives to unify our understanding of how selection and linkage interact to shape genome evolution.

Introduction

That natural selection shapes organismal phenotypes through inherited changes in DNA is an established cornerstone of biology. And yet, we still do not fully understand how natural selection leaves its imprint across the genome. How often do patterns of genetic variability implicate evolutionary constraint alongside repeated bouts of directional selection favoring new beneficial mutations? Studying the interaction of selection with recombination within genomes suggests a program for answering this question. The linear nature of strands of DNA imposes stronger linkage between physically proximate loci, the genetic consequences of which depend on patterns of recombination along the length of chromosomes. As a result, selection exerts its influence over greater physical distances in regions of chromosomes that experience low rates of recombination[1–3]. Population genetics theory also indicates that recombination improves the efficacy of natural selection when multiple linked loci simultaneously experience selective pressures[4, 5], so selective interference between them will be most prevalent in regions with low recombination rates. These powerful ideas have commanded extensive empirical exploration.

Correspondence to A.D.C. asher.cutter@utoronto.ca.

In 1992, Begun and Aquadro revolutionized population genetics by demonstrating a strong positive correlation between nucleotide diversity and local rates of crossing-over in *Drosophila melanogaster*[6] (see also [7, 8]) (Figure 1). This discovery was interpreted as evidence that recurring, strong positive selection dramatically shapes genomic diversity. In the two decades since, similar patterns have been reported in a diverse range of species (Table 1)[9] and for diverse marker types[10], raising the possibility of a common mechanism. But competing explanations for this intriguing pattern, such as the effects of selection removing linked deleterious variants and mutagenesis by recombination, have received empirical and theoretical support (Box 1). Moreover, some species reveal no such effects in their genomes, and changing notions about how adaptation typically proceeds at the molecular level raise questions about the ubiquity of the classic adaptive explanation for heterogeneity across the genome in patterns of genetic variation. Recent advances in genome-scale data collection and sophisticated analytical approaches fuel this debate and motivate a review of the evidence. It is important to understand why genomes differ in how selection and linkage shape the evolution of their genomes. The answer will help to reveal how phenotypic adaptation translates into genomic change and may help to explain why species are more similar in overall levels of diversity than is expected from neutral predictions[3, 11].

Box 1

Recombination-associated mutation and genomic heterogeneity in polymorphism

Selection is not the only possible cause of a positive association between levels of genetic diversity and rates of recombination across the genome[101]. The mutation rate might vary across the genome in a manner that yields more genetic variation in genomic regions with higher rates of recombination; after all, greater mutational input causes higher equilibrium levels of neutral diversity[128]. This selectively neutral effect could be coincidental, or causal, as some experiments with mitotically dividing yeast suggested that the molecular process of recombination could induce mutations[129, 130]. If present, it ought to manifest as elevated divergence between species at unconstrained sites in genomic regions that are subject to high recombination rates, as the rate of substitution is equal to the rate of mutation at equilibrium under selective neutrality[131]. This purely neutral, mutational explanation suffers its own complication, however, in that unaccounted ancestral polymorphism could produce a spurious signal of mutagenic recombination in the face of ongoing and historical selective effects[24, 47, 59, 132, 133]. The mutation-associated recombination mechanism was ruled out early on for *Drosophila melanogaster*[6], but was raised as a serious concern from correlations between divergence and recombination rate in human data[133, 134]. Moreover, there is emerging consensus that a general pattern of mutation-associated recombination is not evident in sequence-divergence data for most species[59, 135, 136], including humans[23, 24], provided that recombination rates themselves have not diverged too greatly between species. So, this selectively neutral alternative explanation generally appears insufficient to account for correlations between polymorphism and recombination at genome scales.

In this Review, we outline theoretical predictions about the relationship between genetic diversity and recombination rate. We then synthesize evidence across a range of organisms, focusing especially on recent population genomic studies, to examine the factors that can contribute to differences among species in the evidence for selection at linked sites. The disparities among species suggest that explicit comparative analysis of the magnitude and sign of selection at linked sites provides a promising avenue for unifying our understanding of how selection and linkage interact to shape genome evolution.

Selection at linked sites: a conceptual overview

When a new beneficial mutation arises in a population, it can rapidly rise in frequency to become fixed, so that, having 'swept through' the population, all individuals will henceforth carry the advantageous allele (Figure 2). Such a 'hard' selective sweep (Figure 2Aa) will also cause fixation of alleles that happen to be linked to the beneficial mutation, even if they have no fitness effects – this process is known as 'genetic hitchhiking'[3, 12]. This gives the classic signature of a hard selective sweep: a reduction in diversity in the vicinity of the selected locus. Following such a sweep, because the few polymorphisms present will typically be rare new mutations on the recently fixed genetic background, the signature presents as an excess of low-frequency variants or, equivalently, a negative skew in the site frequency spectrum relative to the distribution expected under neutrality[13]. Importantly, the prevalence of hitchhiking alleles will be greater when linkage is stronger (i.e. lower recombination rates). Many studies have examined such effects for chromosomes that are devoid of recombination (e.g. W chromosomes in birds, and Y and 'dot' chromosomes in *Drosophila* species[14]), but here we focus primarily on those portions of genomes that vary continuously in recombination rate. Should this process of hitchhiking be repeated again and again, then it is expected that neutral genetic variation will be systematically depleted in genomic regions with little recombination[1, 15], exactly the pattern of 'selection at linked sites' observed by Begun and Aquadro[6]. This reduction in genetic diversity is often described as a genomically localized reduction in effective population size (N_e), but we will generally avoid this inexact analogy because it does not fully capture the effects of linked selection[16–19].

An alternative selective mechanism that eliminates genetic variation in low-recombination regions is negative selection against recurrent deleterious mutations, known as 'background selection'[20, 21] (Figure 2Ad). Under background selection, neutral alleles that are linked to detrimental alleles are driven to extinction, with more drastic effects when recombination rates are low[2]. Empirically, this process is supported in a variety of organisms (Table 1)[21]. A virtue of the background selection explanation is that we know that most new mutations that affect fitness will exert detrimental effects, so negative selection against them is a perpetual force[22]. The inevitability and prevalence of deleterious mutations, and the effects on linked loci when they are removed by selection, argues that a background selection process should form part of the null evolutionary model for the genome when testing for any additional effects owing to recurrent selective sweeps[23, 24].

Additional predictions for linked selection

Recurrent bouts of positive directional selection alter other population genetic properties in addition to the overall amount of linked neutral polymorphism (Box 2). Because the effects of linked selection will be more pronounced in genomic regions where recombination is less frequent, it follows that measures of the skew in the site frequency spectrum, such as Tajima's D values[13], will correlate positively with the rate of recombination[25]. The *D. melanogaster* genome shows this effect[26]. Background selection against deleterious mutations can also generate such correlations under some realistic circumstances, such as when population sizes are small to moderate and selection strength is intermediate[20, 23, 27, 28].

Box 2

Detecting linked selection effects for weak versus strong targets

The greater efficacy of selection when linkage is low should extend to all modes of selection, including purifying selection against deleterious mutations that acts on replacement sites in genes and on synonymous sites for genes that are subject to translational selection on codon usage. Purifying selection dominates most parts of coding sequences, so by facilitating the elimination of deleterious mutations, high recombination regions might yield more slowly evolving genes (low d_N)[137]. However, strongly deleterious replacement-site mutations could have sufficiently large fitness effects to be removed effectively, regardless of local recombination regime, making d_N largely insensitive to recombination rate. By contrast, targets of weak selection should be particularly vulnerable[138]. Synonymous sites in highly expressed genes experience weak selection for translational efficiency, accuracy and stability that causes codon usage bias[139, 140]. In other words, selectively driven codon bias should be more pronounced in high-recombination regions of the genome[141]. In some cases, however, the selectively neutral process of biased gene conversion might yield this same empirical result[142].

Interference between sites that are subjected to very weak selection might also be detected in genomes at a very fine scale, such as for within-gene heterogeneity in codon usage bias[17]. Specifically, such weak Hill–Robertson interference (wHRi) can induce stronger codon usage bias near the edges of genes and exons by virtue of being linked to fewer selected sites than those codons in the middle of exons[17, 143]. However, similar patterns of codon usage bias could instead result from background selection on more strongly detrimental mutations[144] and selection on nearby non-coding regulatory regions has not been considered explicitly in this context. Notably, wHRi, background selection, genetic hitchhiking, and Muller's ratchet all are specific manifestations of the general Hill–Robertson effect resulting from how selection combines with linkage[17]. The reduced effective population size (N_e) imposed by both positive and negative selection is exacerbated in low-recombination regions, leading further selection to be impaired in such regions, although not all properties of Hill–Robertson effects can be captured by a simple re-scaling of N_e[16–19].

Presuming that changes to protein sequences often provide the driving basis for recurrent positive selection, we should expect that genes that rapidly diverge at replacement sites (also known as non-synonymous sites) (high d_N) will exhibit depressed polymorphism at linked neutral sites (e.g. low polymorphism at synonymous sites, π_S)[26, 29, 30]. While this prediction is unique to the influence of positive selection, whether it is likely to be observed in practice is sensitive to the joint effects of positive and negative selection[23]. Because genes represent high-density targets for fitness-affecting mutations, genomic regions that are crowded with coding sequences should generally exhibit stronger diversity-reducing effects of both positive and negative selection[31]. Thus, recurrent selection at linked sites should create a negative correlation between gene density and neutral polymorphism, after controlling for the effect of recombination rate. By the same token, lower levels of polymorphism and longer blocks of linkage disequilibrium ought to surround substitutions at selected sites (e.g. replacement sites), but not substitutions at unconstrained sites (e.g. synonymous sites) [32–35]. The rate at which polymorphism returns to normal with distance from such selected sites provides a means of quantifying the strength of selection, as well as its prevalence[34, 35]. However, when gene density and recombination rates covary, as is known for several species[36], these signatures of linked selection could become obscured.

In addition to the effects of selected sites on polymorphisms in linked loci that are selectively neutral, recombination alters the efficacy of selection at a given site that is linked to another target of selection[37]. This selective interference (known as 'Hill–Robertson interference')[4] should manifest in genomes in well-defined ways. First, recurrent positive selection favoring changes to amino acid sequences in coding genes will be facilitated when the selected targets are only loosely linked to other targets of selection. This could generate a positive correlation between rates of replacement site divergence (d_N) and the local rate of recombination in the genome, if linkage limits evolutionary responses to selection[38]. Similarly, the fraction of replacement site substitutions between species that were driven by positive selection (α) should be higher in genomic regions with more abundant recombination, as reported for *D. melanogaster*[39, 40]. Estimates of α can derive from application of the McDonald–Kreitman test that contrasts fixed differences and polymorphisms for neutral and selection-candidate site classes[41, 42]. In *Drosophila* species and several other organisms, α is estimated to be ~50% or even higher[26, 42, 43], although factors such as linkage to slightly deleterious mutations and weak selection on synonymous sites could inflate such estimates[44], and the relative rate of positively selected substitutions (ω_a) is a more directly relevant quantity[45]. It is also worth putting this seemingly high rate of selective divergence in the context of the perpetual elimination of deleterious mutations: in comparisons involving *Drosophila* species, typically <5% of replacement sites in genes have diverged, with >95% of replacement sites being conserved[42, 46].

The inconsistent evidence among species

Drosophila research has spearheaded much of empirical population genetics, and this also has been true for studying the influence of linked selection on genomes[6, 26, 47]. There is compelling evidence that recurrent positive selection has shaped genomic patterns of molecular evolution in multiple *Drosophila* species, consistent with an important influence of recurrent genetic hitchhiking[26, 48, 49]. It nevertheless remains unclear how positive and

negative selection each contribute in relative terms to genomic heterogeneity in *Drosophila*'s patterns of polymorphism, suggesting that sophisticated modeling that incorporates both positive and negative selection can play a role in further elucidating the processes of linked selection in this important system[23]. For example, it is primarily those genes in high-recombination regions that yield a signature of positive selection, as from McDonald–Kreitman tests, suggesting that hard sweeps might not fully explain reduced diversity in low-recombination regions[50]. Moreover, some analyses that apply a background selection process on the X-chromosome can reasonably explain patterns of polymorphism without any added contribution of widespread hitchhiking[51]. Regardless of such details, it is now clear that current inferences regarding recurrent hitchhiking effects for *Drosophila* species do not apply universally across organisms[52] (Table 1, Figure 3).

Many plants and yeast exhibit weak-at-best signatures of linked selection[53–58], and humans and selfing nematodes reveal particularly potent effects of background selection[23, 24, 34, 59–61] (Table 1). Although the interplay between selection and linkage was historically studied with datasets of polymorphism from tens of loci, it is now investigated using population genomic datasets for a range of species (Table 1). Therefore the persistence of differences among species is not simply an issue of statistical power. Indeed, population genomic analyses have revealed a clear-cut signal of recurrent genetic hitchhiking on genomic patterns of polymorphism only for *Drosophila* fruitflies[35, 47] and, in conjunction with background selection, in *Caenorhabditis elegans* nematodes[60, 62, 63] (Table 1). Such disparities across the tree of life argue that there must be fundamental differences among organisms in one or more factors that generate or mask the genome-wide effects of linked selection.

Factors underlying the disparities among species

Genomic signatures of selection at linked sites require the interplay of several factors, and differences among species in any of these variables could explain discrepant patterns (Figure 2B). Broadly, organisms might differ in: how selection operates in their genomes; their profile of recombination rate; and history of population size. Next we treat these factors in detail. We also address the potential for these and other characteristics to conceal, rather than result in, signatures of linked selection.

Differences in selection

Rates at which selected mutations arise

Species with a higher genomic density of selected mutations per unit of genetic distance should be more sensitive to selection at linked sites. The genomic deleterious mutation rate (U) can be estimated from the total mutation rate and the fraction of sites expected to be functional (for example, replacement sites). Direct estimates of mutation rates from sequencing of mutation-accumulation lines[64–68] or other sets of relatives[69] implicates inter-species differences in U that could produce heterogeneity in the strength of background selection. Estimating the rate of beneficial mutation is more difficult[70], but there are also hints of species differences in this quantity. Humans and some plants seem to have a smaller fraction of mutations driven to fixation by positive selection than do several other

species[42, 45, 71]. These differences reflect variability in the rate at which beneficial mutations arise or contrasting efficiencies of positive selection (see below). Such variability would generate discordant hitchhiking patterns.

The relative incidence of new beneficial and deleterious mutations in a population will depend on its proximity to the current fitness optimum: a higher fraction of mutations will be beneficial when the population lies further from the optimum. Recurrent bouts of positive versus negative selection in genomes might differ correspondingly among organisms. One challenge is to predict the degree to which either stabilizing selection or directional selection operate on phenotypes as a consequence of static versus fluctuating abiotic and biotic factors: are there organismal characteristics or ecological niches that predispose species to differ consistently in how regularly they undergo rounds of directional selection? Large-scale datasets have emboldened researchers to consider simultaneously the influence of positive and negative selection to help discern whether and how much positive selection is required to explain observed patterns of polymorphism above and beyond the persistent process of background selection. So far, however, only the human and *C. elegans* population genomic data have been subjected to thorough joint analysis, revealing background selection as a sufficient mechanism to explain widespread genomic patterns of neutral variation in humans[23] and a combined influence of positive and negative selection in *C. elegans* best explaining linked selection effects[63].

Mutation frequencies when selection begins

Selective sweeps are stronger when positive selection targets newer mutations because the beneficial mutation occurs on only one, or a small number, of genetic backgrounds. At present, few data are available to address whether species differ in the initial frequencies of selected mutations. A limiting input of beneficial mutations underlies the genetic hitchhiking model of positive directional selection, whereby selection acts iteratively on single, new beneficial mutations. This view also is encapsulated in modeling of genetic draft, in which linkage to positively selected loci overwhelms genetic drift in its effects on neutral polymorphism when the effective population size (N_e) is sufficiently large[72]. However, this notion of mutation-limited adaptation at the molecular level might not be the norm in nature. Hard sweeps are proposed to have been rare in recent human evolution[23, 34, 73, 74]. Positive selection might more commonly operate on standing genetic variation in the population[75]. Even for *Drosophila* species, the preeminent system for inferring hard sweeps, adaptation might not conform to the presumption of mutation-limited evolution[76], and experimental evolution in asexual systems indicates that adaptation is not solely mutation-limited[77, 78].

What happens when selection is not mutation-limited? When positive selection causes fixation of a beneficial allele that arose via multiple convergent mutations or that previously occurred in the population at mutation–drift balance, giving a so-called ‘soft’ selective sweep[79], then the traditional population genetic signatures of hitchhiking often are undetectable[75, 80]. In particular, soft selective sweeps yield only mild reductions in linked polymorphism (Figure 2Ab), alterations of the site frequency spectrum, and linkage disequilibrium[79, 81–83]. This theoretical result was presaged by the work of Ohta and Kimura[84]. In *Drosophila simulans*, one approach that can quantify only hard sweeps

suggests that ~13% of replacement-site substitutions were fixed in such a manner versus ~90% of such sites inferred to have fixed via either hard or soft sweeps[35, 43]. These values imply that soft sweeps could represent upwards of 85% of positively selected substitutions, but formal analyses are desperately needed to disentangle the relative contributions of hard and soft sweeps to divergence. Indeed, the relative incidence of hard and soft sweeps is a major unsolved problem. If soft sweeps are more common, then we might not expect positive selection to strongly depress levels of polymorphism in genomic regions with low rates of recombination. This possibility should motivate theoretical exploration, which has already begun for the case of partial sweeps[85]. Relaxing the classic hard sweep assumption that selected alleles fix rapidly generates a new suite of predictions[85]. For example, recurrent partial sweeps can leave the frequency spectrum unaffected even as levels of diversity are perturbed. A challenge is to determine why the incidence of hard versus soft sweeps in adaptation might differ among organisms, and whether it varies in predictable ways. One potential factor, discussed below, is population structure: species-wide soft sweeps could dominate when populations are subdivided[12, 86–88] (Box 3).

Box 3

Comparative hypothesis testing for selection at linked sites

Consider a hypothetical example to explore whether population structure obscures a signature of linked selection. If such masking occurs, then one would predict a negative relationship in a plot showing the strength of selection at linked sites (e.g. the correlation coefficient between recombination rate and neutral polymorphism, Corr[R, π]) as a function of within-species population differentiation (e.g. F_{st} or alternative metrics), as in the schematic shown in panel **a**, where each point on the plot represents a species (or, more formally, a phylogenetically independent contrast[125]). For example, contrasts of the rates of positively selected substitutions (ω_a) for taxa of differing effective populations size (N_e) have now been attempted in several studies[45, 71, 145]. More sophisticated analysis could test for correlated evolution directly along the phylogeny, relating population genetic parameters, genomic features, and other candidate variables[146]. Panel **b** illustrates the logic for a test for correlated character evolution on a phylogeny, with taxa having strong versus weak selection at linked sites (purple versus orange) contrasted with the state of a candidate explanatory factor such as effective population size (large red versus small blue). In the hypothetical collection of species depicted, lineages with strong effects of linked selection (purple) tend to have large population size (red).

Strength of selection, genetic architecture and dominance

Selective sweeps are expected to be most severe when selection is strong and adaptive mutations are dominant[15, 89]. By contrast, background selection will be most potent when selection is relatively weak and mutations are recessive, as long as mutations are deleterious enough to be removed efficiently from the population (i.e. when the selection coefficient (s) > $1/N_e$)[20, 21]. Multiple factors are likely to generate differences among species in the distribution of selection coefficients. These characteristics include the quality and variability of the environments species inhabit, and the distances of populations from their fitness

optima. The strength of selection in a genomic region also depends on the genetic architecture of the phenotype targeted by selection. For polygenic traits, selection intensity is spread across many variants and individual mutations experience weaker selection[90]. In combination with the attenuation of selection's magnitude across loci, evolutionary responses to such polygenic selection need not even result in fixation of alleles at any given locus[90, 91] (Figure 2Ac). The few theoretical studies to date on polygenic selective sweep effects suggest weaker signatures of selection at linked sites for these and other reasons[92, 93].

When a beneficial mutation is linked to deleterious mutations, those mildly deleterious mutations can hitchhike to fixation[94–96]. This raises the possibility that the fixed region will be susceptible to repeated bouts of positive selection as compensatory alleles arise and replace the detrimental alleles that had hitchhiked with the original beneficial mutation[18]. Moreover, the linkage to deleterious mutations slows the fixation process and increases the incidence of recombinant genetic backgrounds, which obscures the signatures of a selective sweep[95]. Given that slightly deleterious alleles are abundant in genomes of natural populations[97], this effect might contribute to some organisms having only weak signatures of selection at linked sites, even if hard sweeps represent the prevailing mode of positive selection. Even without positive selection, purifying selection against multiple linked targets similarly reduces the efficacy of selection, which can offset the diversity-reducing effects of classic models of background selection[16, 28]. At present, however, this mitigation of classic background selection effects, owing to Hill–Robertson interference, has been explored largely through simulation, and a complete mathematical association with recombination rate is not yet established[98], although fitness-class coalescent modeling appears promising[19].

Differences in dominance among species are perhaps harder to envision, but there is a general paucity of data available to address this issue. One aspect of organismal life history that could contribute to differences in the realized dominance among species is the extent to which selection acts during the haploid phase of the life cycle, which can be extensive in plants, algae and fungi. Similarly, the role of dominance will be less important in highly self-fertilizing species in which there are pronounced deficits of heterozygotes. In both haploids and highly selfing species, selection is akin to additive selection operating in outcrossing diploids, but twice as strong, effectively shifting the realized distribution of dominance coefficients. Conversely, species that retain substantial portions of their genome from historical episodes of polyploidization may experience even more masking of mutations, again contributing to differences among species in realized distributions of dominance[99].

Given the profound perturbation to neutral evolution that positive and negative selection both can exert on genomes on an ongoing basis, a sticky question presents itself: what is the appropriate null model for the genome[48]? A null model of neutrality is useful for simply detecting that some form of selection yields a genome-wide effect. However, positive selection often is the selective agent of most biological interest. Much effort has been devoted to distinguishing background selection from hitchhiking effects of recurrent selective sweeps [100], but this is difficult and perhaps a false dichotomy [101]; we advocate future efforts to integrate them. To this end, background selection should be taken into

account as part of a null model of molecular evolution, because deleterious mutations arise inexorably and in turn are subjected to purifying selection[23, 24]. One could envision even more elaborate null models, for example, to test specifically for the influence of hard sweeps by incorporating a genome-wide model of soft sweeps in addition to background selection. A challenge in defining the parameters for any selection model is to appropriately accommodate the distributions for coefficients of selection and dominance.

Differences in recombination profiles

Recombination rate

Just as recombination rate variation across genomes affects the signature of selection at linked sites, we expect observed differences in recombination rate among species[9] to create disparate patterns. Species with higher average levels of recombination should exhibit less evidence of linked selection at random sites in the genome, all else being equal. Focusing on species with population genomic data (Table 1), we observe mixed evidence for this trend. Some species with strong evidence for selection at linked sites, including *Drosophila melanogaster*, *Drosophila simulans*, *Caenorhabditis elegans* and *Caenorhabditis briggsae* have lower average meiotic crossover rates[102–104] than do some species with weak signatures, including *Arabidopsis thaliana*[32] and *Saccharomyces cerevisiae*[105]. On the other hand, *D. pseudoobscura* has a relatively high recombination rate[106] but still shows a strong pattern of linked selection, whereas humans exhibit a weaker signature of linked selection despite less crossing over[107]. Our ability to detect linked selection also depends critically on the variance of recombination rate across a genome. Chromosome number and architecture (such as the relative proportions of heterochromatic and euchromatic sequence) contribute to this variance and differ among species. Furthermore, local rates of recombination can evolve rapidly between closely related species[9, 106, 108, 109].

Because the likelihood that an allele exerts an influence on neighboring loci depends on the density of functionally important sites as well as genetic distance, genome-scale patterns will be sensitive to both of these factors. In particular, the lower gene density in genomic regions with low rates of recombination for many plant species might counteract genome-wide signatures of linked selection[36]. By contrast, *Caenorhabditis* nematodes have higher gene density in low-recombination regions, which likely contributes to especially strong signatures of linked selection in selfing species in this group[59, 60, 62, 63]. Consequently, we expect species with positive correlations between gene density and recombination rate to exhibit weaker evidence for selection at linked sites across their genomes.

Analyses relating selection and linkage implicitly presume that the mapping of recombination rate on the genome is itself fixed. Like any trait, however, recombination rates vary and evolve[9]. Consequently, linked selection could be obscured if polymorphism from one species is analyzed with recombination rates calculated from a different species. Moreover, individuals within a population can vary in their recombination propensities across the genome[9, 110–112]. If some organisms have high among-individual variability in recombination profiles, then recombination rates measured in a given map might be a poor predictor of population genetic parameters. With emerging maps of intra-species variation in recombination rate[113], this can be accounted for explicitly[106].

Mating system effects on effective recombination rates

Any factors that intensify genetic linkage will exacerbate the consequences of selective sweeps and background selection. One such factor is self-fertilization[20, 53], or other reproductive systems that reduce the incidence of outcrossing[114]. Selfing, which is common in both plants and animals, reduces the genetically effective rate of recombination across the genome as a byproduct of eliminating heterozygosity[115]. Because even modest levels of recombination can mitigate the signature of linked selection, selfing broadens the extent of the genome that can reveal its influence[20, 53, 116]. In addition, when linked selection occurs in selfers, it is more likely to enable the fixation of detrimental alleles that are linked to advantageous alleles[95], potentially contributing to genome degradation.

Differences in population size and history

Effective population size

Selection on beneficial and detrimental mutations is more efficient in large populations. As a result, the severity of hitchhiking and background selection ought to increase with N_e across species[12, 27]. Some evidence is consistent with this prediction; for example, in comparison with humans, *D. melanogaster* has a much larger N_e and shows stronger evidence of hitchhiking.

However, the effect of N_e on selection at linked sites is more complicated than it first appears. The transit time of a beneficial mutation grows with the population size (although only logarithmically), allowing more recombination to decouple the mutation from nearby neutral variants[12]. Recent theory also predicts positive selection to more often operate on standing variation when population mutation rates are high (i.e. in large populations)[79, 88], and the resulting soft sweeps induce a weaker hitchhiking effect than does selection on beneficial new variants[79, 81–83]. If the relative contribution of soft sweeps to adaptation does increase with N_e, then perhaps a genomic signature of linked selection will not correlate positively with population size after all. Even within a hard sweep framework, theory of genetic draft anticipates a threshold N_e above which polymorphism within genomes should be insensitive to further increases in population size[72]. Species vary widely in effective population size because of differences in demographic history, mating system and population structure[11, 117], indicating that it will be feasible to determine empirically how differences in N_e among species might contribute to differences in signatures of linked selection. These efforts should help determine whether the effects of linked selection are responsible for the relatively small range of inter-species variation in average nucleotide diversity despite the enormous range of census population sizes[3, 11].

Population demography and structure

The intricacies of a species' demographic history can mask a genome-wide signature of selection at linked sites[118] (Box 3). Population substructure is likely to be particularly subversive, especially when adaptation occurs locally within distinct subpopulations. Even if selection conforms to the hard sweep process, when it fixes alleles only within subpopulations, it will not yield a species-wide hitchhiking signature[12, 87, 118]. Analyzing population genetic data according to different sampling schemes (local population samples,

pooled samples, and scattered samples with a single individual per subpopulation) provides one promising approach for species that are subdivided into many subpopulations[119–121]. However, even when a new mutation is globally advantageous, it will arise in just a single subpopulation in a subdivided population, which can obscure the hard sweep signature at linked sites or make it heterogeneous across subpopulations[12, 86, 87]. Population structure can also facilitate parallel evolution and soft selective sweeps[88], with potentially similar obscuring effects. Some approaches exploit the population-specific nature of local adaptation in subdivided populations to rule out the influence of background selection[122, 123]. Expanding populations contain a greater relative abundance of new rare variants, so hard sweeps might be more prevalent in them compared to populations at equilibrium. Finally, some models of population demography that include high variance in reproductive success lead to a less-than-linear dependence of neutral polymorphism on population size[124], which might similarly affect the way in which reduced recombination rate mimics reduced effective population size. In addition, repeated population bottlenecks will increase stochastic variance across the genome, perhaps erasing historically present signatures of selection at linked sites.

Toward phylogenetic coherence

The inconsistent evidence for genome-wide signatures of selection at linked sites across species (Table 1, Figure 3) reveals a conspicuous gap in current understanding of genome evolution. Why do species have such disparate genomic patterns of polymorphism? Phylogenetic comparative methods provide an established framework for answering such questions. In addition to traditional traits[125], they have been advocated for understanding other issues in population genetics[11], genomics[126] and quantitative genetics[127] as these fields have matured to include large data sets from many species. Comparative approaches could prove to be powerful in evaluating the contributions of different factors to the detection, the sign and the magnitude of selection at linked sites in genomes (Figure 2B; Box 3).

Some first attempts in this vein have compared species of rice[36] and tomatoes[53], which suggested influences of domestication and mating system. Unfortunately, most current taxa with pertinent data are scattered among divergent lineages of eukaryotes (Table 1). A promising approach is to study closely related species that are known to differ in one or a few of the biological factors outlined here, so as to isolate the effects of particular traits. Some features will be simpler to explore than others, and all are contingent on recombination rate having been estimated for multiple related species from genetic and physical genome maps. In particular, it might prove more challenging to test directly for covarying genomic variables, such as gene density and recombination rate, because they are unlikely to have diverged greatly among close relatives. However, teasing apart the influence of factors such as mating system, the extent of population subdivision, and effective population size are eminently amenable to thoughtful construction of species sets for study (Box 3).

Conclusions

There is now compelling evidence that interactions between linkage and natural selection have shaped genome evolution in a wide variety of organisms (Table 1)[9]. However, species differ dramatically in the magnitude of effects exerted by linked selection, in how the patterns manifest in their genomes (e.g. correlations between recombination and polymorphism, gene density and polymorphism, site frequency spectrum and recombination), and in the prevailing form of selection (positive or negative). Importantly, few studies have attempted to incorporate simultaneously the influence of background selection and recurrent hitchhiking, despite it being well-appreciated that both processes likely operate simultaneously[101].

Moreover, analyses of recurrent hitchhiking have been predicated on models of hard selective sweeps, and yet emerging evidence suggests an important role of soft sweeps in adaptation[75]. It is not yet clear, however, whether recurrent soft selective sweeps could explain the observed genome-wide patterns of selection at linked sites. One solution will involve more detailed simulation studies of linked selection to better describe the quantitative effects of biologically relevant parameter differences between species. Examples of particularly incisive studies include measuring how signatures of selection at linked sites compare for a species at mutation-drift equilibrium versus a species undergoing a recent population expansion, quantifying how the rate of inbreeding interacts with both the strength of purifying selection and the frequency of selective sweeps to affect genome-wide patterns of linked selection, and addressing how genomes evolve through a mixture of recurrent soft sweeps and background selection.

It will also be crucial for future analyses to include background selection as part of the null model for selection at linked sites, layering models of recurrent selective sweeps on top to determine whether and how much the incorporation of positive selection improves the fit of these theoretical models to the actual genome-wide patterns of evolution[23, 63]. Adaptation and positive selection clearly occur, but how often across the tree of life do they yield sweeping population genomic signatures rather than perturbations that localize at particular loci? A fuller understanding of linked selection in evolution requires that we clarify and integrate the causes of disparities among organisms. The comparative method can play an important role in testing for the contributions of alternative factors that modulate the intensity of selection at linked sites in genomes, and its detection. Such a unified understanding across the tree of life will require substantial research investments, particularly in measuring species differences in biological characteristics that are expected to shape selection at linked sites.

Acknowledgments

A.D.C. is supported by grants from the Natural Sciences and Engineering Research Council of Canada and the National Institutes of Health USA, and from a Canada Research Chair. B. A. P. is supported by NIH grant HG004498. We thank Laurence Loewe and three anonymous reviewers for helpful comments.

Proposed Glossary Terms

Site frequency spectrum	The distribution of allele frequencies in a population; visualized as the histogram of counts of the number of alleles that have a given population frequency
Background selection	The elimination of neutral polymorphisms as a result of their linkage to deleterious mutations that are subject to purifying selection
Biased gene conversion (BGC)	Gene conversion is a non-reciprocal recombination process that causes one sequence to be overwritten with information from the other. BGC is when the two possible sequences act as donor templates with unequal probabilities
Effective population size (N_e)	Formulated by Wright in 1931, N_e reflects the size of an idealized population that would experience drift in the same way as the actual (census) population. N_e can be lower than census population size due to various factors, including variance in reproductive success, a history of population bottlenecks and inbreeding
Genetic hitchhiking	The process by which a neutral, or in some cases deleterious, mutation may change in population frequency owing to linkage with a selected mutation
Selective interference	When recombination fails to break down linkage disequilibrium between alleles at selected loci, the ability of selection to act on these alleles tends to be reduced
Linkage disequilibrium	A measure of whether alleles at two loci coexist in a population in a non-random fashion. Alleles that are in linkage disequilibrium are found together on the same haplotype more often than would be expected under free recombination
McDonald–Kreitman test	A statistical test used to compare between-species divergence and within-species polymorphism at replacement and synonymous sites to infer selection acting on proteins
Selection at linked sites	The interaction between natural selection and genetic linkage that can yield deviations from the levels of polymorphism, allele frequencies, and linkage disequilibria expected from neutral evolution alone
Selective sweep	The increase in frequency of a beneficial allele (and closely linked chromosomal segments via genetic hitchhiking) to fixation that is caused by positive selection
Directional selection	Selection that favours one allele over all other alleles of a gene. The frequency of this beneficial allele can rise or can be held in check by recurrent mutation

Neutral polymorphism	Alternative allelic variants with no selective difference between them, the dynamics of which are controlled mainly by genetic drift and migration. They can, however, be influenced by selection on nearby (linked) loci
Replacement site (also known as non-synonymous site)	Any nucleotide within a gene at which a point mutation can alter the encoded amino-acid sequence. Models of molecular evolution account for different possible degeneracies of such sites in codons
Synonymous site	Any site within a gene at which some or all possible point mutations, depending the corresponding codon's degeneracy, do not change the encoded amino acid. Changes at synonymous sites are often presumed to be selectively neutral
d_N	The rate of protein-coding sequence divergence, quantified as the number of non-synonymous substitutions per non-synonymous site
Mutation-accumulation lines	Unique genetic backgrounds created by multiple generations of controlled breeding in such a way as to minimize the action of natural selection and to maximize the retention of new mutations. They are used to identify spontaneous mutations and to study their phenotypic properties
Stabilizing selection	A type of natural selection that favours intermediate phenotypes, such as when the population is close to its fitness optimum with respect to the trait
Purifying selection	Natural selection against deleterious alleles that arise in a population, preventing their increase in frequency
Genetic draft	Stochastic fluctuations in allele frequencies in a population caused by repeated hard selective sweeps. Hypothesized to be the primary source of stochastic variation in allele frequencies in large populations, in which the sampling effects of genetic drift are relatively weak
Genetic drift	Random fluctuations through time in the allele frequencies of a population, caused by a sampling effect that is strongest in small populations. Drift can overwhelm the deterministic effects of natural selection if the selective differences between alleles are small
Mutation-limited adaptation	When mutational input into a population is sufficiently low, the rate of adaptation will be limited by the input of new beneficial mutations. This common theoretical assumption will be violated in real species with populations that are large, subdivided, or subject to frequent changes in selective regime

Standing genetic variation	Allelic variation that is currently segregating within a population from old mutation events, as opposed to alleles that just arose by new mutation events
Mutation–drift balance	The equilibrium between input of alleles into a subpopulation by migration and their loss by genetic drift. When there are many subpopulations, gene flow of alleles by migration is considered to introduce new alleles into any given subpopulation at a higher rate than does mutation
Population structure	The distribution of individuals into partially isolated, local subpopulations or demes that are interconnected by migration (gene flow)
Selection coefficient (*s*)	A parameter describing the difference in average fitness between two genotypes when fitness is measured relative to the average fitness of one of the genotypes (known as the reference genotype)
Genetic architecture	The number, identity, phenotypic effects and population frequencies of the mutations that contribute to phenotypic variation
Polygenic selection	Selection on a trait that has a genetic basis comprised of many gene loci (tens, hundreds or more). A given strength of selection on the phenotype will exert a weaker effect on any one locus when the trait is polygenic than when the trait is monogenic
Fitness-class coalescent	A version of structured coalescent models of evolution that traces how individuals descend by mutations through different fitness classes, rather than through time
Transit time	The duration of time that elapses from when an allele first experiences selection to when it becomes fixed in a population
Population bottleneck	A marked reduction in population size followed by the survival and expansion of a sample of the original population. It often results in the loss of genetic variation and a skewed site frequency spectrum
Muller's ratchet	The irreversible accumulation of deleterious mutations in asexual populations of finite size. The average load of mutations increases over generations because the class of individuals that carry the smallest number of mutant alleles is occasionally lost by genetic drift. In the absence of recombination or compensatory mutation, this class can never be re-created. The process is named after H. J. Muller, who described it in 1964
$\boldsymbol{F_{st}}$	A measure of population subdivision that indicates the proportion of genetic diversity found between populations relative to the amount within populations.

References

1. Wiehe THE, Stephan W. Analysis of a genetic hitchhiking model, and its application to DNA polymorphism data from *Drosophila melanogaster*. Mol Biol Evol. 1993; 10:842–854. [PubMed: 8355603]
2. Hudson RR, Kaplan NL. Deleterious background selection with recombination. Genetics. 1995; 141:1605–1617. [PubMed: 8601498]
3. Maynard Smith J, Haigh J. Hitch-hiking effect of a favorable gene. Genet Res. 1974; 23:23–35. The classic theoretical study describing the process of genetic hitchhiking. [PubMed: 4407212]
4. Hill WG, Robertson A. Effect of linkage on limits to artificial selection. Genet Res. 1966; 8:269–294. The classic model of how linked selected loci interfere with each other's ability to fix the favored alleles in the population. [PubMed: 5980116]
5. Felsenstein J. The evolutionary advantage to recombination. Genetics. 1974; 78:737–756. [PubMed: 4448362]
6. Begun DJ, Aquadro CF. Levels of naturally-occurring DNA polymorphism correlate with recombination rates in *Drosophila melanogaster*. Nature. 1992; 356:519–520. The seminal study demonstrating selection at linked sites yielded a general pattern on genetic diversity. [PubMed: 1560824]
7. Aguade M, Miyashita N, Langley CH. Reduced variation in the yellow-achaete-scute region in natural-populations of *Drosophila melanogaster*. Genetics. 1989; 122:607–615. [PubMed: 17246506]
8. Stephan W, Langley CH. Molecular genetic-variation in the centromeric region of the X-chromosome in 3 *Drosophila ananassae* populations. 1 Contrasts between the *vermilion* and *forked* loci. Genetics. 1989; 121:89–99. [PubMed: 2563714]
9. Smukowski CS, Noor MA. Recombination rate variation in closely related species. Heredity. 2011; 107:496–508. [PubMed: 21673743]
10. Massouras A, et al. Genomic variation and its impact on gene expression in *Drosophila melanogaster*. PLoS Genetics. 2012; 8:e1003055. [PubMed: 23189034]
11. Leffler EM, et al. Revisiting an old riddle: what determines genetic diversity levels within species? PLoS Biol. 2012; 10:e1001388. [PubMed: 22984349]
12. Barton NH. Genetic hitchhiking. Philos Trans R Soc B. 2000; 355:1553–1562.
13. Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics. 1989; 123:585–595. [PubMed: 2513255]
14. Frankham R. How closely does genetic diversity in finite populations conform to predictions of neutral theory? Large deficits in regions of low recombination. Heredity. 2012; 108:167–178. [PubMed: 21878983]
15. Kaplan NL, Hudson RR, Langley CH. The "hitchhiking effect" revisited. Genetics. 1989; 123:887–899. The important theoretical extension of genetic hitchhiking for a single episode selection to recurrent selective sweeps in a genome that varies along its length in the rate of recombination. [PubMed: 2612899]
16. Charlesworth B, Betancourt AJ, Kaiser VB, Gordo I. Genetic recombination and molecular evolution. Cold Spring Harb Symp Quant Biol. 2009; 74:177–186. [PubMed: 19734202]
17. Comeron JM, Williford A, Kliman RM. The Hill-Robertson effect: evolutionary consequences of weak selection and linkage in finite populations. Heredity. 2008; 100:19–31. [PubMed: 17878920]
18. Barton NH. Genetic linkage and natural selection. Philos Trans R Soc Lond B Biol Sci. 2010; 365:2559–2569. [PubMed: 20643746]
19. Walczak AM, Nicolaisen LE, Plotkin JB, Desai MM. The structure of genealogies in the presence of purifying selection: a fitness-class coalescent. Genetics. 2012; 190:753–779. [PubMed: 22135349]
20. Charlesworth B, Morgan MT, Charlesworth D. The effect of deleterious mutations on neutral molecular variation. Genetics. 1993; 134:1289–1303. This classic paper introduces the notion that background selection reduces neutral genetic variation disproportionately in low recombination regions, similar to the effects of recurrent selective sweeps. [PubMed: 8375663]

21. Charlesworth B. The effects of deleterious mutations on evolution at linked sites. Genetics. 2012; 190:5–22. [PubMed: 22219506]
22. Eyre-Walker A, Keightley PD. The distribution of fitness effects of new mutations. Nat Rev Genet. 2007; 8:610–618. [PubMed: 17637733]
23. Lohmueller KE, et al. Natural selection affects multiple aspects of genetic variation at putatively neutral sites across the human genome. PLoS Genetics. 2011; 7 An exemplary recent empirical study that incorporates background selection as part of the null model of molecular evolution shaping genome-wide patterns of genetic variatoin in humans.
24. McVicker G, Gordon D, Davis C, Green P. Widespread genomic signatures of natural selection in Hominid evolution. PLoS Genetics. 2009; 5:e1000471. [PubMed: 19424416]
25. Braverman JM, Hudson RR, Kaplan NL, Langley CH, Stephan W. The hitchhiking effect on the site frequency-spectrum of DNA polymorphisms. Genetics. 1995; 140:783–796. [PubMed: 7498754]
26. Sella G, Petrov DA, Przeworski M, Andolfatto P. Pervasive natural delection in the Drosophila genome? PLoS Genetics. 2009; 5:e1000495. [PubMed: 19503600]
27. Charlesworth D, Charlesworth B, Morgan MT. The pattern of neutral molecular variation under the background selection model. Genetics. 1995; 141:1619–1632. [PubMed: 8601499]
28. Kaiser VB, Charlesworth B. The effects of deleterious mutations on evolution in non-recombining genomes. Trends Genet. 2009; 25:9–12. [PubMed: 19027982]
29. Andolfatto P. Hitchhiking effects of recurrent beneficial amino acid substitutions in the *Drosophila melanogaster* genome. Genome Res. 2007; 17:1755–1762. [PubMed: 17989248]
30. Macpherson JM, Sella G, Davis JC, Petrov DA. Genomewide spatial correspondence between nonsynonymous divergence and neutral polymorphism reveals extensive adaptation in Drosophila. Genetics. 2007; 177:2083–2099. [PubMed: 18073425]
31. Payseur BA, Nachman MW. Gene density and human nucleotide polymorphism. Mol Biol Evol. 2002; 19:336–340. [PubMed: 11861892]
32. Kim S, et al. Recombination and linkage disequilibrium in *Arabidopsis thaliana*. Nat Genet. 2007; 39:1151–1155. [PubMed: 17676040]
33. Hinds DA, et al. Whole-genome patterns of common DNA variation in three human populations. Science. 2005; 307:1072–1079. [PubMed: 15718463]
34. Hernandez RD, et al. Classic selective sweeps were rare in recent human evolution. Science. 2011; 331:920–924. This empirical study emphasizes the possibility that hard sweeps might not best explain much adaptation in recent human evolution, suggesting that soft sweeps and polygenic selection are important alternatives. [PubMed: 21330547]
35. Sattath S, Elyashiv E, Kolodny O, Rinott Y, Sella G. Pervasive adaptive protein evolution apparent in diversity patterns around amino acid substitutions in *Drosophila simulans*. PLoS Genetics. 2011; 7:e1001302. [PubMed: 21347283]
36. Flowers JM, et al. Natural selection in gene-dense regions shapes the genomic pattern of polymorphism in wild and domesticated rice. Mol Biol Evol. 2012; 29:675–687. An important empirical illustration of the confounding influence of genomic variables in detecting selection at linked sites. [PubMed: 21917724]
37. Barton NH. Linkage and the limits to natural selection. Genetics. 1995; 140:821–841. [PubMed: 7498757]
38. Betancourt AJ, Presgraves DC. Linkage limits the power of natural selection in Drosophila. Proc Natl Acad Sci USA. 2002; 99:13616–13620. [PubMed: 12370444]
39. Presgraves DC. Recombination enhances protein adaptation in *Drosophila melanogaster*. Curr Biol. 2005; 15:1651–1656. [PubMed: 16169487]
40. Mackay TF, et al. The *Drosophila melanogaster* genetic reference panel. Nature. 2012; 482:173–178. [PubMed: 22318601]
41. McDonald JH, Kreitman M. Adaptive protein evolution at the *Adh* locus in Drosophila. Nature. 1991; 351:652–654. [PubMed: 1904993]
42. Eyre-Walker A, Keightley PD. Estimating the rate of adaptive molecular evolution in the presence of slightly deleterious mutations and population size change. Mol Biol Evol. 2009; 26:2097–2108. [PubMed: 19535738]

43. Andolfatto P, Wong KM, Bachtrog D. Effective population size and the efficacy of selection on the X chromosomes of two closely related Drosophila species. Genome Biol Evol. 2011; 3:114–128. [PubMed: 21173424]
44. Fay JC. Weighing the evidence for adaptation at the molecular level. Trends Genet. 2011; 27:343–349. [PubMed: 21775012]
45. Gossmann TI, Keightley PD, Eyre-Walker A. The effect of variation in the effective population size on the rate of adaptive molecular evolution in eukaryotes. Genome Biol Evol. 2012; 4:658–667. [PubMed: 22436998]
46. Ellegren H. Comparative genomics and the study of evolution by natural selection. Mol Ecol. 2008; 17:4586–4596. [PubMed: 19140982]
47. Begun DJ, et al. Population genomics: whole-genome analysis of polymorphism and divergence in *Drosophila simulans*. PLoS Biol. 2007; 5:e310. [PubMed: 17988176]
48. Hahn MW. Toward a selection theory of molecular evolution. Evolution. 2008; 62:255–265. [PubMed: 18302709]
49. Langley CH, et al. Genomic variation in natural populations of *Drosophila melanogaster*. Genetics. 2012; 192:533–598. A thorough population genomic analysis relating polymorphism characteristics to recombination profiles. [PubMed: 22673804]
50. Shapiro JA, et al. Adaptive genic evolution in the Drosophila genomes. Proc Natl Acad Sci USA. 2007; 104:2271–2276. [PubMed: 17284599]
51. Charlesworth B. The role of background selection in shaping patterns of molecular evolution and variation: evidence from variability on the Drosophila X chromosome. Genetics. 2012; 191:233–246. [PubMed: 22377629]
52. Wright SI, Andolfatto P. The impact of natural selection on the genome: emerging patterns in Drosophila and Arabidopsis. Annu Rev Ecol Evol Syst. 2008; 39:193–213.
53. Baudry E, Kerdelhue C, Innan H, Stephan W. Species and recombination effects on DNA variability in the tomato genus. Genetics. 2001; 158:1725–1735. [PubMed: 11514458]
54. Wright SI, et al. Testing for effects of recombination rate on nucleotide diversity in natural populations of *Arabidopsis lyrata*. Genetics. 2006; 174:1421–1430. [PubMed: 16951057]
55. Nordborg M, et al. The pattern of polymorphism in *Arabidopsis thaliana*. PLoS Biol. 2005; 3:e196. [PubMed: 15907155]
56. Cutter AD, Moses AM. Polymorphism, divergence, and the role of recombination in *Saccharomyces cerevisiae* genome evolution. Mol Biol Evol. 2011; 28:1745–1754. [PubMed: 21199893]
57. Tenaillon MI, et al. Patterns of diversity and recombination along chromosome 1 of maize (*Zea mays ssp mays* L.). Genetics. 2002; 162:1401–1413. [PubMed: 12454083]
58. Cao J, et al. Whole-genome sequencing of multiple *Arabidopsis thaliana* populations. Nat Genet. 2011; 43:956–963. [PubMed: 21874002]
59. Cutter AD, Choi JY. Natural selection shapes nucleotide polymorphism across the genome of the nematode *Caenorhabditis briggsae*. Genome Res. 2010; 20:1103–1111. [PubMed: 20508143]
60. Rockman MV, Skrovanek SS, Kruglyak L. Selection at linked sites shapes heritable phenotypic variation in *C. elegans*. Science. 2010; 330:372–376. An intriguing illustration of how selection at linked sites can affect phenotypic (gene expression) variation in addition to neutral nucleotide variation. [PubMed: 20947766]
61. Hobolth A, Dutheil JY, Hawks J, Schierup MH, Mailund T. Incomplete lineage sorting patterns among human, chimpanzee, and orangutan suggest recent orangutan speciation and widespread selection. Genome Res. 2011; 21:349–356. [PubMed: 21270173]
62. Cutter AD, Payseur BA. Selection at linked sites in the partial selfer *Caenorhabditis elegans*. Mol Biol Evol. 2003; 20:665–673. [PubMed: 12679551]
63. Andersen EC, et al. Chromosome-scale selective sweeps shape *Caenorhabditis elegans* genomic diversity. Nat Genet. 2012; 44:285–290. This empirical study demonstrates how selective sweeps can eliminate large swathes of genetic variability in self-fertilizing species, in addition to the effects of background selection. [PubMed: 22286215]
64. Baer CF, Miyamoto MM, Denver DR. Mutation rate variation in multicellular eukaryotes: causes and consequences. Nat Rev Genet. 2007; 8:619–631. [PubMed: 17637734]

65. Denver DR, et al. Variation in base-substitution mutation in experimental and natural lineages of Caenorhabditis nematodes. Genome Biol Evol. 2012; 4:513–522. [PubMed: 22436997]
66. Haag-Liautard C, et al. Direct estimation of per nucleotide and genomic deleterious mutation rates in Drosophila. Nature. 2007; 445:82–85. [PubMed: 17203060]
67. Lynch M, et al. A genome-wide view of the spectrum of spontaneous mutations in yeast. Proc Natl Acad Sci USA. 2008; 105:9272–9277. [PubMed: 18583475]
68. Ossowski S, et al. The rate and molecular spectrum of spontaneous mutations in *Arabidopsis thaliana*. Science. 2010; 327:92–94. [PubMed: 20044577]
69. Kong A, et al. Rate of de novo mutations and the importance of father's age to disease risk. Nature. 2012; 488:471–475. [PubMed: 22914163]
70. Schneider A, Charlesworth B, Eyre-Walker A, Keightley PD. A method for inferring the rate of occurrence and fitness effects of advantageous mutations. Genetics. 2011; 189:1427–1437. [PubMed: 21954160]
71. Gossmann TI, et al. Genome wide analyses reveal little evidence for adaptive evolution in many plant species. Mol Biol Evol. 2010; 27:1822–1832. [PubMed: 20299543]
72. Gillespie JH. Is the population size of a species relevant to its evolution? Evolution. 2001; 55:2161–2169. [PubMed: 11794777]
73. Coop G, et al. The role of geography in human adaptation. PLoS Genetics. 2009; 5:e1000500. [PubMed: 19503611]
74. Granka JM, et al. Limited evidence for classic selective sweeps in African populations. Genetics. 2012; 192:1049–1064. [PubMed: 22960214]
75. Pritchard JK, Pickrell JK, Coop G. The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. Curr Biol. 2010; 20:R208–R215. [PubMed: 20178769]
76. Karasov T, Messer PW, Petrov DA. Evidence that adaptation in Drosophila is not limited by mutation at single sites. PLoS Genetics. 2010; 6:e1000924. [PubMed: 20585551]
77. Miller CR, Joyce P, Wichman HA. Mutational effects and population dynamics during viral adaptation challenge current models. Genetics. 2011; 187:185–202. [PubMed: 21041559]
78. Lang GI, Botstein D, Desai MM. Genetic variation and the fate of beneficial mutations in asexual populations. Genetics. 2011; 188:647–661. [PubMed: 21546542]
79. Hermisson J, Pennings PS. Soft sweeps: molecular population genetics of adaptation from standing genetic variation. Genetics. 2005; 169:2335–2352. This important theoretical study models selection from standing genetic variation, to produce soft sweeps. [PubMed: 15716498]
80. Storz JF, Wheat CW. Integrating evolutionary and functional approaches to infer adaptation at specific loci. Evolution. 2010; 64:2489–2509. [PubMed: 20500215]
81. Rouzine IM, Coffin JM. Multi-site adaptation in the presence of infrequent recombination. Theor Popul Biol. 2010; 77:189–204. [PubMed: 20149814]
82. Przeworski M, Coop G, Wall JD. The signature of positive selection on standing genetic variation. Evolution. 2005; 59:2312–2323. [PubMed: 16396172]
83. Innan H, Kim Y. Pattern of polymorphism after strong artificial selection in a domestication event. Proc Natl Acad Sci USA. 2004; 101:10667–10672. [PubMed: 15249682]
84. Ohta T, Kimura M. Effect of selected linked locus on heterozygosity of neutral alleles (hitchhiking effect). Genet Res. 1975; 25:313–326. [PubMed: 1183812]
85. Coop G, Ralph P. Patterns of neutral diversity under general models of selective sweeps. Genetics. 2012; 192:205–224. This recent theoretical extension of recurrent positive selection to partialsweeps helps generalize the expectations for patterns of genetic variation in genomes. [PubMed: 22714413]
86. Kim Y, Maruki T. Hitchhiking effect of a beneficial mutation spreading in a subdivided population. Genetics. 2011; 189:213–226. [PubMed: 21705748]
87. Bierne N. The distinctive footprints of local hitchhiking in a varied environment and global hitchhiking in a subdivided population. Evolution. 2010; 64:3254–3272. [PubMed: 20550573]
88. Ralph P, Coop G. Parallel adaptation: one or many waves of advance of an advantageous allele? Genetics. 2010; 186:647–668. [PubMed: 20660645]

89. Teshima KM, Przeworski M. Directional positive selection on an allele of arbitrary dominance. Genetics. 2006; 172:713–718. [PubMed: 16219788]

90. Turchin MC, et al. Evidence of widespread selection on standing variation in Europe at height-associated SNPs. Nat Genet. 2012; 44:1015–1019. [PubMed: 22902787]

91. Pritchard JK, Di Rienzo A. Adaptation - not by sweeps alone. Nat Rev Genet. 2010; 11:665–667. [PubMed: 20838407]

92. Pavlidis P, Metzler D, Stephan W. Selective sweeps in multilocus models of quantitative traits. Genetics. 2012; 192:225–239. [PubMed: 22714406]

93. Chevin LM, Hospital F. Selective sweep at a quantitative trait locus in the presence of background genetic variation. Genetics. 2008; 180:1645–1660. [PubMed: 18832353]

94. Chun S, Fay JC. Evidence for hitchhiking of deleterious mutations within the human genome. PLoS Genetics. 2011; 7:e1002240. [PubMed: 21901107]

95. Hartfield M, Otto SP. Recombination and hitchhiking of deleterious alleles. Evolution. 2011; 65:2421–2434. [PubMed: 21884046]

96. Hadany L, Feldman MW. Evolutionary traction: the cost of adaptation and the evolution of sex. J Evol Biol. 2005; 18:309–314. [PubMed: 15715837]

97. Boyko AR, et al. Assessing the evolutionary impact of amino acid mutations in the human genome. PLoS Genetics. 2008; 4:e1000083. [PubMed: 18516229]

98. Zeng K, Charlesworth B. The joint effects of background selection and genetic recombination on local gene genealogies. Genetics. 2011; 189:251–266. [PubMed: 21705759]

99. Gerstein AC, Otto SP. Ploidy and the causes of genomic evolution. J Hered. 2009; 100:571–581. [PubMed: 19625454]

100. Stephan W. Genetic hitchhiking versus background selection: the controversy and its implications. Philos Trans R Soc Lond B Biol Sci. 2010; 365:1245–1253. [PubMed: 20308100]

101. Hudson RR. How can the low levels of DNA sequence variation in regions of the Drosophila genome with low recombination rates be explained? Proc Natl Acad Sci USA. 1994; 91:6815–6818. [PubMed: 8041702]

102. Fiston-Lavier AS, Singh ND, Lipatov M, Petrov DA. *Drosophila melanogaster* recombination rate calculator. Gene. 2010; 463:18–20. [PubMed: 20452408]

103. True JR, Mercer JM, Laurie CC. Differences in crossover frequency and distribution among three sibling species of Drosophila. Genetics. 1996; 142:507–523. [PubMed: 8852849]

104. Hillier LW, et al. Comparison of *C. elegans* and *C. briggsae* genome sequences reveals extensive conservation of chromosome organization and synteny. PLoS Biol. 2007; 5:e167. [PubMed: 17608563]

105. Mancera E, Bourgon R, Brozzi A, Huber W, Steinmetz LM. High-resolution mapping of meiotic crossovers and non-crossovers in yeast. Nature. 2008; 454:479–485. [PubMed: 18615017]

106. McGaugh SE, et al. Recombination modulates how selection affects linked sites in Drosophila. PLoS Biol. 2012; 10:e1001422. [PubMed: 23152720]

107. Kong A, et al. Fine-scale recombination rate differences between sexes, populations and individuals. Nature. 2010; 467:1099–1103. [PubMed: 20981099]

108. Dumont BL, White MA, Steffy B, Wiltshire T, Payseur BA. Extensive recombination rate variation in the house mouse species complex inferred from genetic linkage maps. Genome Res. 2011; 21:114–125. [PubMed: 20978138]

109. Auton A, et al. A fine-scale chimpanzee genetic map from population sequencing. Science. 2012; 336:193–198. [PubMed: 22422862]

110. Coop G, Wen X, Ober C, Pritchard JK, Przeworski M. High-resolution mapping of crossovers reveals extensive variation in fine-scale recombination patterns among humans. Science. 2008; 319:1395–1398. [PubMed: 18239090]

111. Sandor C, et al. Genetic variants in *REC8*, *RNF212*, and *PRDM9* influence male recombination in cattle. PLoS Genetics. 2012; 8:e1002854. [PubMed: 22844258]

112. Ross JA, et al. *Caenorhabditis briggsae* recombinant inbred line genotypes reveal inter-strain incompatibility and the evolution of recombination. PLoS Genetics. 2011; 7

113. Comeron JM, Ratnappan R, Bailin S. The many landscapes of recombination in *Drosophila melanogaster*. PLoS Genetics. 2012; 8:e1002905. An important experimental demonstration of how different genetic backgrounds yield different profiles of recombination along their genomes. [PubMed: 23071443]
114. Neher RA, Shraiman BI. Genetic draft and quasi-neutrality in large facultatively sexual populations. Genetics. 2011; 188:975–996. [PubMed: 21625002]
115. Charlesworth D, Wright SI. Breeding systems and genome evolution. Curr Opin Genet Dev. 2001; 11:685–690. [PubMed: 11682314]
116. Innan H, Stephan W. Distinguishing the hitchhiking and background selection models. Genetics. 2003; 165:2307–2312. [PubMed: 14704207]
117. Charlesworth B. Fundamental concepts in genetics: Effective population size and patterns of molecular evolution and variation. Nat Rev Genet. 2009; 10:195–205. [PubMed: 19204717]
118. Siol M, Wright SI, Barrett SCH. The population genomics of plant adaptation. New Phytol. 2010; 188:313–332. [PubMed: 20696011]
119. Wakeley J. Nonequilibrium migration in human history. Genetics. 1999; 153:1863–1871. [PubMed: 10581291]
120. Stadler T, Haubold B, Merino C, Stephan W, Pfaffelhuber P. The impact of sampling schemes on the site frequency spectrum in nonequilibrium subdivided populations. Genetics. 2009; 182:205–216. [PubMed: 19237689]
121. Cutter AD, Wang GX, Ai H, Peng Y. Influence of finite-sites mutation, population subdivision and sampling schemes on patterns of nucleotide polymorphism for species with molecular hyperdiversity. Mol Ecol. 2012; 21:1345–1359. [PubMed: 22320847]
122. Chen Y, Marsh BJ, Stephan W. Joint effects of natural selection and recombination on gene flow between *Drosophila ananassae* populations. Genetics. 2000; 155:1185–1194. [PubMed: 10880480]
123. Charlesworth B, Nordborg M, Charlesworth D. The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. Genet Res. 1997; 70:155–174. [PubMed: 9449192]
124. Eldon B, Wakeley J. Coalescent processes when the distribution of offspring number among individuals is highly skewed. Genetics. 2006; 172:2621–2633. [PubMed: 16452141]
125. Harvey, PH.; Pagel, MD. The Comparative Method in Evolutionary Biology. Oxford University Press; New York: 1991.
126. O'Brien SJ, et al. The promise of comparative genomics in mammals. Science. 1999; 286:458–462. 479–481. [PubMed: 10521336]
127. Steppan SJ, Phillips PC, Houle D. Comparative quantitative genetics: evolution of the G matrix. Trends Ecol Evol. 2002; 17:320–327.
128. Kimura M, Crow JF. The number of alleles that can be maintained in a finite population. Genetics. 1964; 49:725–738. [PubMed: 14156929]
129. Strathern JN, Shafer BK, McGill CB. DNA-synthesis errors associated with double-strand-break repair. Genetics. 1995; 140:965–972. [PubMed: 7672595]
130. Holbeck SL, Strathern JN. A role for REV3 in mutagenesis during double-strand break repair in *Saccharomyces cerevisiae*. Genetics. 1997; 147:1017–1024. [PubMed: 9383049]
131. Kimura M. Evolutionary rate at molecular level. Nature. 1968; 217:624–626. [PubMed: 5637732]
132. Stevison L, Noor M. Genetic and evolutionary correlates of fine-scale recombination rate variation in *Drosophila persimilis*. J Mol Evol. 2010; 71:332–345. [PubMed: 20890595]
133. Hellmann I, Ebersberger I, Ptak SE, Paabo S, Przeworski M. A neutral explanation for the correlation of diversity with recombination rates in humans. Am J Hum Genet. 2003; 72:1527–1535. [PubMed: 12740762]
134. Lercher MJ, Hurst LD. Human SNP variability and mutation rate are higher in regions of high recombination. Trends Genet. 2002; 18:337–340. [PubMed: 12127766]
135. Kulathinal RJ, Stevison LS, Noor MAF. The genomics of speciation in Drosophila: diversity, divergence, and introgression estimated using low-coverage genome sequencing. PLoS Genetics. 2009; 5

136. Huang SW, Friedman R, Yu N, Yu A, Li WH. How strong is the mutagenicity of recombination in mammals? Mol Biol Evol. 2005; 22:426–431. [PubMed: 15496551]
137. Pal C, Papp B, Hurst LD. Does the recombination rate affect the efficiency of purifying selection? The yeast genome provides a partial answer. Mol Biol Evol. 2001; 18:2323–2326. [PubMed: 11719582]
138. McVean GAT, Charlesworth B. A population genetic model for the evolution of synonymous codon usage: patterns and predictions. Genet Res. 1999; 74:145–158.
139. Duret L. Evolution of synonymous codon usage in metazoans. Curr Opin Genet Dev. 2002; 12:640–649. [PubMed: 12433576]
140. Hershberg R, Petrov DA. Selection on codon bias. Annu Rev Genet. 2008; 42:287–299. [PubMed: 18983258]
141. Kliman RM, Hey J. Reduced natural selection associated with low recombination in *Drosophila melanogaster*. Mol Biol Evol. 1993; 10:1239–1258. [PubMed: 8277853]
142. Marais G, Mouchiroud D, Duret L. Does recombination improve selection on codon usage? Lessons from nematode and fly complete genomes. Proc Natl Acad Sci USA. 2001; 98:5688–5692. [PubMed: 11320215]
143. Comeron JM, Kreitman M. Population, evolutionary and genomic consequences of interference selection. Genetics. 2002; 161:389–410. [PubMed: 12019253]
144. Loewe L, Charlesworth B. Background selection in single genes may explain patterns of codon bias. Genetics. 2007; 175:1381–1393. [PubMed: 17194784]
145. Strasburg JL, et al. Effective population size is positively correlated with levels of adaptive divergence among annual sunflowers. Mol Biol Evol. 2011; 28:1569–1580. [PubMed: 20952500]
146. Lartillot N, Poujol R. A phylogenetic model for investigating correlated evolution of substitution rates and continuous phenotypic characters. Mol Biol Evol. 2011; 28:729–744. [PubMed: 20926596]
147. Nachman MW, Bauer VL, Crowell SL, Aquadro CF. DNA variability and recombination rates at X-linked loci in humans. Genetics. 1998; 150:1133–1141. [PubMed: 9799265]
148. Bullaughey K, Przeworski M, Coop G. No effect of recombination on the efficacy of natural selection in primates. Genome Res. 2008; 18:544–554. [PubMed: 18199888]
149. Nachman MW. Patterns of DNA variability at X-linked loci in *Mus domesticus*. Genetics. 1997; 147:1303–1316. [PubMed: 9383072]
150. Takahashi A, Liu YH, Saitou N. Genetic variation versus recombination rate in a structured population of mice. Mol Biol Evol. 2004; 21:404–409. [PubMed: 14660688]
151. Rao Y, Sun L, Nie Q, Zhang X. The influence of recombination on SNP diversity in chickens. Hereditas. 2011; 148:63–69. [PubMed: 21561450]
152. Fang L, et al. Positive correlation between recombination rate and nucleotide diversity is shown under domestication selection in the chicken genome. Chinese Science Bulletin. 2008; 53:746–750.
153. Stump AD, et al. Centromere-proximal differentiation and speciation in *Anopheles gambiae*. Proc Natl Acad Sci USA. 2005; 102:15930–15935. [PubMed: 16247019]
154. Andolfatto P, Przeworski M. Regions of lower crossing over harbor more rare variants in African populations of *Drosophila melanogaster*. Genetics. 2001; 158:657–665. This empirical study demonstrates how population genetic features other than simply the amount of diversity are affected by selection at linked sites. [PubMed: 11404330]
155. Hey J, Kliman RM. Interactions between natural selection, recombination and gene density in the genes of Drosophila. Genetics. 2002; 160:595–608. [PubMed: 11861564]
156. Charlesworth B. Background selection and patterns of genetic diversity in *Drosophila melanogaster*. Genet Res. 1996; 68:131–149. [PubMed: 8940902]
157. Kulathinal RJ, Bennett SM, Fitzpatrick CL, Noor MA. Fine-scale mapping of recombination rate in Drosophila refines its correlation to diversity and divergence. Proc Natl Acad Sci USA. 2008; 105:10051–10056. [PubMed: 18621713]

158. Marais G, Piganeau G. Hill-Robertson interference is a minor determinant of variations in codon bias across *Drosophila melanogaster* and *Caenorhabditis elegans* genomes. Mol Biol Evol. 2002; 19:1399–1406. [PubMed: 12200468]

159. Denver DR, et al. A genome-wide view of *Caenorhabditis elegans* base-substitution mutation processes. Proc Natl Acad Sci USA. 2009; 106:16310–16324. [PubMed: 19805298]

160. Cutter AD. Multilocus patterns of polymorphism and selection across the X-chromosome of *Caenorhabditis remanei*. Genetics. 2008; 178:1661–1672. [PubMed: 18245859]

161. Noor MAF. Mutagenesis from meiotic recombination is not a primary driver of sequence divergence between Saccharomyces species. Mol Biol Evol. 2008; 25:2439–2444. [PubMed: 18723832]

162. Connallon T, Knowles LL. Recombination rate and protein evolution in yeast. BMC Evol Biol. 2007; 7:235. [PubMed: 18042299]

163. Weber CC, Hurst LD. Protein rates of evolution are predicted by double-strand break events, independent of crossing-over rates. Genome Biol Evol. 2009; 1:340–349. [PubMed: 20333203]

164. Kliman RM, Irving N, Santiago M. Selection conflicts, gene expression, and codon usage trends in yeast. J Mol Evol. 2003; 57:98–109. [PubMed: 12962310]

165. Dvorak J, Luo MC, Yang ZL. Restriction fragment length polymorphism and divergence in the genomic regions of high and low recombination in self-fertilizing and cross-fertilizing Aegilops species. Genetics. 1998; 148:423–434. [PubMed: 9475752]

166. Schmid KJ, Ramos-Onsins S, Ringys-Beckstein H, Weisshaar B, Mitchell-Olds T. A multilocus sequence survey in *Arabidopsis thaliana* reveals a genome-wide departure from a neutral model of DNA sequence polymorphism. Genetics. 2005; 169:1601–1615. [PubMed: 15654111]

167. Slotte T, et al. Genomic determinants of protein evolution and polymorphism in Arabidopsis. Genome Biol Evol. 2011; 3:1210–1219. [PubMed: 21926095]

168. Marais G, Charlesworth B, Wright SI. Recombination and base composition: the case of the highly self-fertilizing plant *Arabidopsis thaliana*. Genome Biol. 2004; 5:R45. [PubMed: 15239830]

169. Kraft T, Sall T, Magnusson-Rading I, Nilsson NO, Hallden C. Positive correlation between recombination rates and levels of genetic variation in natural populations of sea beet (*Beta vulgaris subsp maritima*). Genetics. 1998; 150:1239–1244. [PubMed: 9799275]

170. Roselius K, Stephan W, Stadler T. The relationship of nucleotide polymorphism, recombination rate and selection in wild tomato species. Genetics. 2005; 171:753–763. [PubMed: 16085701]

171. Stephan W, Langley CH. DNA polymorphism in Lycopersicon and crossing-over per physical length. Genetics. 1998; 150:1585–1593. [PubMed: 9832534]

172. Branca A, et al. Whole-genome nucleotide diversity, recombination, and linkage disequilibrium in the model legume *Medicago truncatula*. Proc Natl Acad Sci USA. 2011; 108:E864–E870. [PubMed: 21949378]

173. Molina J, et al. Molecular evidence for a single evolutionary origin of domesticated rice. Proc Natl Acad Sci USA. 2011; 108:8351–8356. [PubMed: 21536870]

174. Tenaillon MI, U'Ren J, Tenaillon O, Gaut BS. Selection versus demography: a multilocus investigation of the domestication process in maize. Mol Biol Evol. 2004; 21:1214–1225. [PubMed: 15014173]

175. Thuillet AC, et al. A weak effect of background selection on trinucleotide microsatellites in maize. J Hered. 2008; 99:45–55. [PubMed: 17962226]

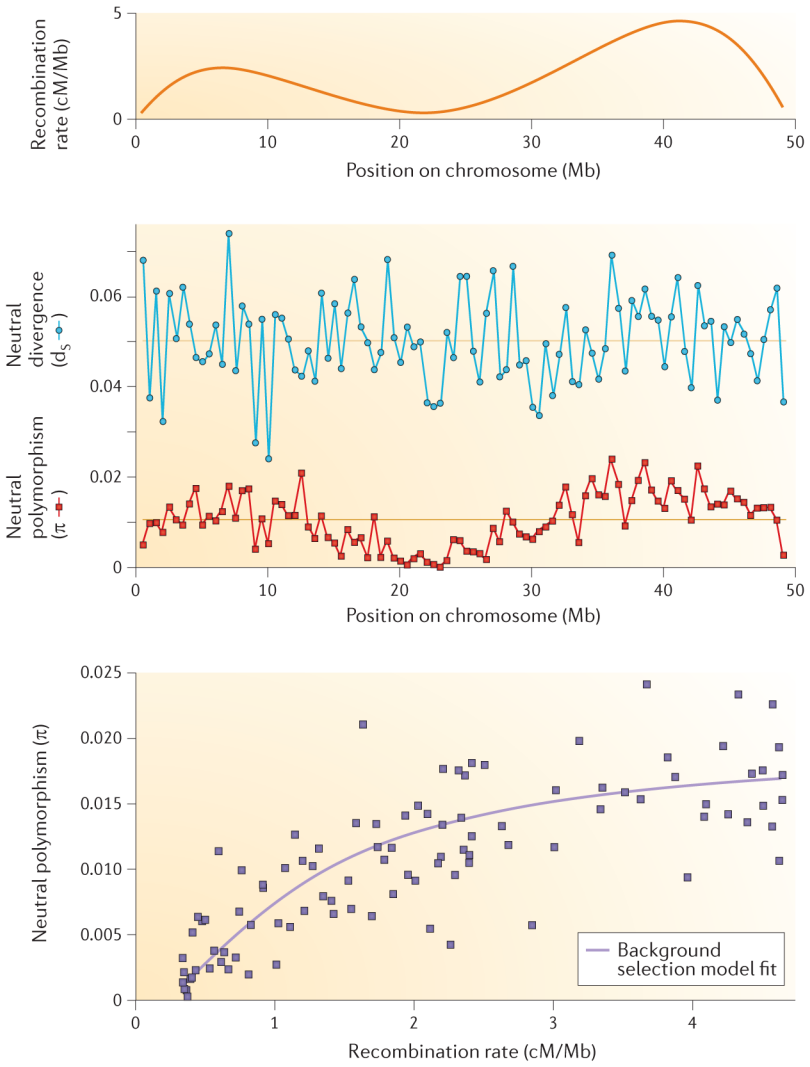

Figure 1. A hypothetical chromosome exhibiting a 'selection at linked sites' pattern

Given the hypothetical recombination rate profile across the chromosome (top panel) and random variation in mutation rate, reflected in divergence at unconstrained sites (blue points, middle panel), measures of neutral polymorphism at unconstrained sites are predicted by recurrent genetic hitchhiking and background selection to be lower in chromosomal regions with lower average recombination rates (red points, middle panel). Recurrent positive or negative selection would yield a positive association between polymorphism and recombination rate (bottom panel). These hypothetical data were generated assuming background selection, but recurrent hard sweeps would yield a qualitatively similar pattern.

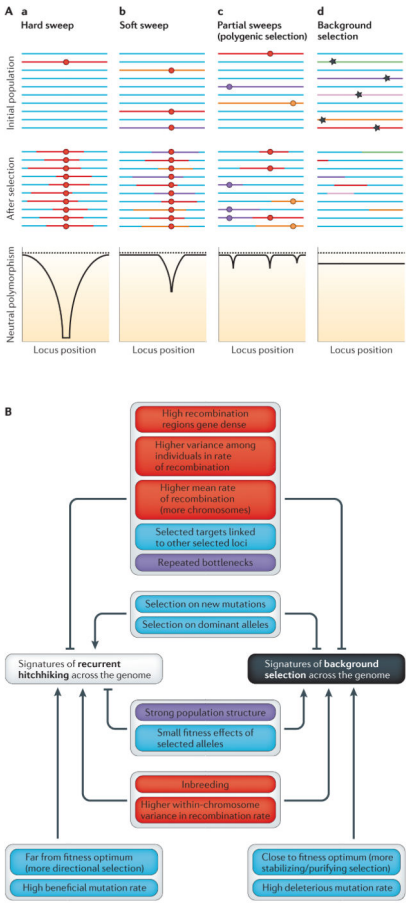

Figure 2. Modes of selection on linked genetic variation and factors affecting them

A | In the top two panel rows, horizontal lines represent haplotypes in a population at a particular genomic locus, with distinct colours representing different genetic backgrounds (compared to well-mixed reference genetic backgrounds indicated in blue). A hard sweep (Aa) involves the fixation of a single new beneficial mutation, whereas a soft sweep (Ab) involves fixation of selectively equivalent alleles that occur at intermediate frequency (i.e. on differing genetic backgrounds) in the population at the onset of selection. Adaptation by polygenic selection (Ac) causes an increase in the frequency of alleles at many loci, but selection does not necessarily drive fixation at any given locus (i.e. multiple partial sweeps). Background selection (Ad) eliminates deleterious mutations and the genetic backgrounds linked to them. Beneficial alleles are represented by circles and stars represent deleterious alleles. The chromosome segment represents a region with a uniform rate of recombination. Dotted lines in the bottom row of panels represent the equilibrium expectation for neutral polymorphism; solid curves show qualitative patterns of neutral polymorphism across the chromosome region. B Differences among species in a broad range of factors could vary the strength of observed signatures of selection at linked sites as caused by recurrent genetic hitchhiking or background selection. Arrows indicate factors associated with recombination (red), selection (blue) and population demography (purple) that tend to exacerbate the effects of linked selection; bars indicate mitigating factors. This diagram is meant as a

general summary, as the details of parameter values can change the influence of some exacerbating and mitigating factors.

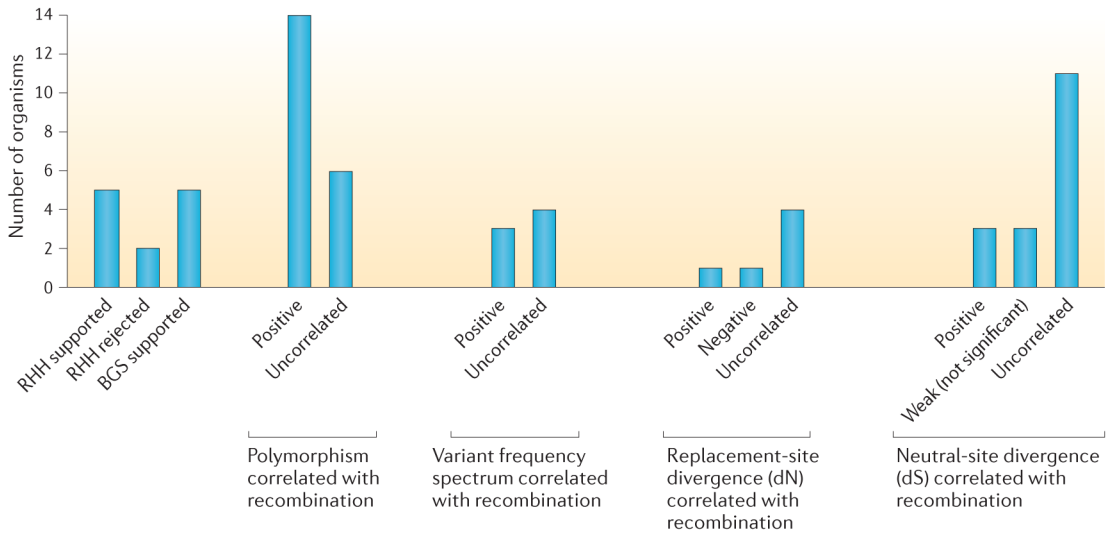

Figure 3. Taxonomic support for different signatures of selection at linked sites

Species differ in their support for how selection and recombination interact to modulate patterns of molecular evolution. Summary of species signatures of linked selection in Table 1; a given species may be included in multiple categories in the figure. Theory predicts recurrent genetic hitchhiking (RHH) to generate positive correlations of recombination rate with neutral polymorphism, the site frequency spectrum, and dN. Background selection (BGS) also is predicted to yield positive correlations of recombination rate with neutral polymorphism.

Table 1

Summary of organismal examples of selection at linked sites.*

		SLS inference[a]			neutral polymorphism correlated with[b]:			recombination rate correlated with[b]:				
organism	**genome-scale analysis[a]**	**strength**	**RHH**	**BGS**	**rec. rate**	**repl. site divergence**	**gene density**	**codon bias**	**site freq. spectrum**	**repl. site divergence**	**neutral site divergence**	**key references**
Mammals												
Homo sapiens	+	moderate	(+)	+	+		−		+	o	+	23, 24, 31, 134, 147, 148
Mus musculus					+						o	149, 150
Birds												
Gallus gallus					+				o		o	151, 152
Insects												
Anopheles gambiae			+		+				(+)			153
Drosophila melanogaster	+	strong	+	+	+	−		+	+	+	o	6, 10, 29, 39, 40, 49, 154–156
Drosophila persimilis					+			+			o	132
Drosophila pseudoobscura	+	moderate	(+)		+	−				o	o(+)	106, 135, 157
Drosophila simulans	+	strong	+		+	−	−				(+)	30, 47
Nematodes												
Caenorhabditis briggsae		strong		+	+				(+)	−	(+)	59
Caenorhabditis elegans	+	strong	+	+	+		−	o	+	(+)	o(+)	62, 63, 158, 159
Caenorhabditis remanei					o						(o)	160
Fungi												
Saccharomyces cerevisiae	+	weak			+	−	(+)	(+)	o	(−)	(−)	56, 137, 161–164
Plants												
Aegilops spp.					+							165
Arabidopsis thaliana	+	weak		+	o(+)	o	−	o	o	o		32, 55, 166–168
Arabidopsis lyrata		weak	−		o		o	o		o	o	54, 167

		SLS inference[a]			neutral polymorphism correlated with[b]:			recombination rate correlated with[b]:				
organism	genome-scale analysis[a]	strength	RHH	BGS	rec. rate	repl. site divergence	gene density	codon bias	site freq. spectrum	repl. site divergence	neutral site divergence	key references
Beta vulgaris					+						o	169
Lycopersicon spp.		weak		+	o(+)						o	53, 116, 170, 171
Medicago truncatula	+			+	+		−					172
Oryza sativa		weak	+	+	−		−				+	36, 173
Oryza rufipogon		moderate		+	o/−		−				+	36
Zea mays		weak	−	+	o(+)				o		o	57, 174, 175

*studies of non-recombining chromosomes not included;

a a '+' indicates affirmative evidence, '−' indicates refuting evidence;

b a '+' indicates positive correlation, '−' indicates negative correlation, 'o' indicates no correlation; values in parenthesis indicate weak or marginally-significant effect; RHH = recurrent genetic hitchhiking; BGS = background selection