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

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

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 the 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 differences in linked selection among species — including roles for selective sweeps being 'hard' or 'soft' — and the concealing effects of demography and confounding genomic variables. We advocate targeted studies of closely related species to unify our understanding of how selection and linkage interact to shape genome evolution.

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 mutations.

[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, Madison, Wisconsin 53706, USA.*
Correspondence to A.D.C.
e-mail:
asher.cutter@utoronto.ca
doi:10.1038/nrg3425
Published online
12 March 2013

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 favouring new beneficial mutations? Studying the interaction of selection with recombination within genomes suggests a programme 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.

In 1992, Begun and Aquadro[6] revolutionized population genetics by demonstrating a strong positive correlation between nucleotide diversity and local rates of crossing-over in *Drosophila melanogaster* (see also REFS 7,8) (FIG. 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[9] (TABLE 1) 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 would be expected from neutral predictions[3,11].

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

© 2013 Macmillan Publishers Limited. All rights reserved

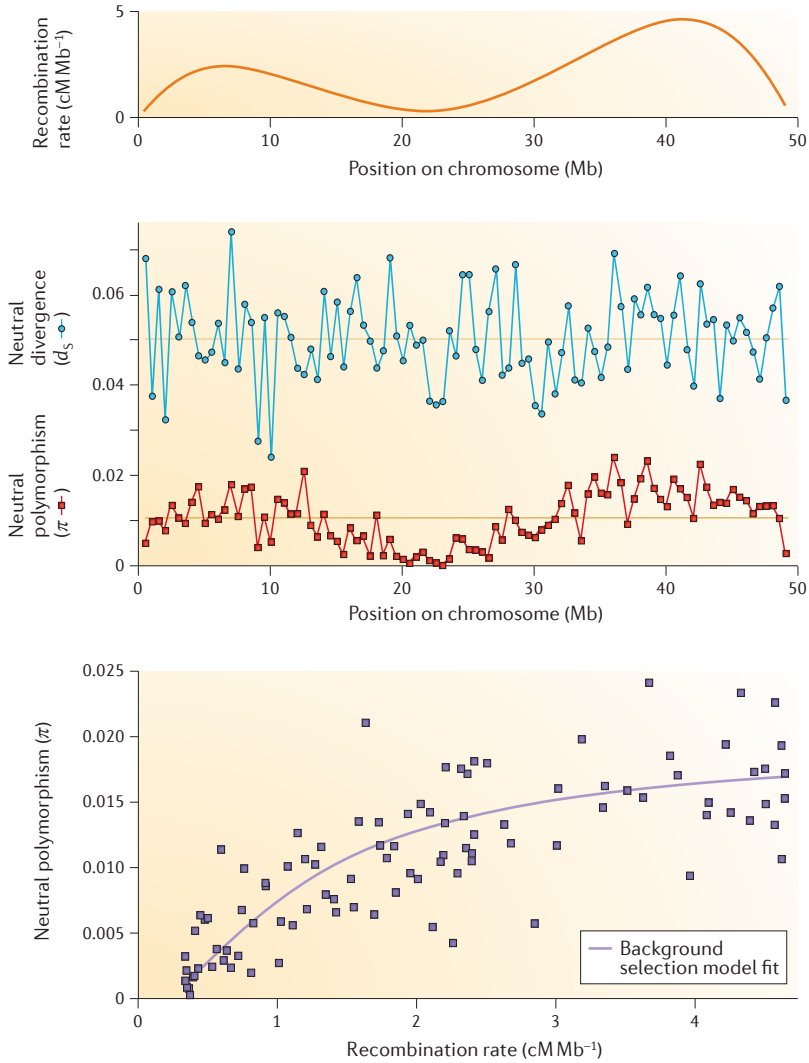


Figure 1 | **A hypothetical chromosome exhibiting a 'selection at linked sites' pattern.** Given the hypothetical recombination rate profile across the chromosome (top panel) and random variation in mutation rate, which is reflected in divergence at unconstrained sites (blue points, middle panel), measures of neutral polymorphism at unconstrained sites are predicted by recurrent genetic hitch-hiking 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. cM, centimorgans; Mb, megabases.

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 (FIG. 2). Such a 'hard' selective sweep (FIG. 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 hitch-hiking[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, as a negative skew in the site frequency spectrum relative to the distribution expected under neutrality[13]. Importantly, the prevalence of hitch-hiking alleles will be greater when linkage is stronger (that is, lower recombination rates). Many studies have examined such effects for chromosomes that are devoid of recombination (for example, W chromosomes in birds, and Y and 'dot' chromosomes in *Drosophila* spp.[14]), but here we primarily focus on those portions of genomes that vary in recombination rate along their length. Should this process of hitch-hiking 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; this is known as background selection[20,21] (FIG. 2Ad). Under background selection, neutral alleles that are linked to detrimental alleles are driven to extinction, and more drastic effects occur when recombination rates are low[2]. Empirically, this process is supported in various organisms[21] (TABLE 1). 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, argue that a background selection process should form a 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 in which recombination is less frequent, it follows that measures of the skew in the site frequency spectrum, such as Tajima's *D* values[13], will positively correlate with the rate of recombination[25]. The *D. melanogaster* genome shows this effect[26]. Background selection against deleterious mutations can also

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.

© 2013 Macmillan Publishers Limited. All rights reserved

Table 1 | **Summary of organismal examples of selection at linked sites***

Organism	Genome-scale analysis‡	SLS inference‡			Neutral polymorphism correlated with§:			Recombination rate correlated with§:				Key refs
		Strength	RHH	BGS	Recombination rate	Replacement site divergence	Gene density	Codon bias	Site frequency spectrum	Replacement site divergence	Neutral site divergence	
Mammals												
Homo sapiens	+	M	(+)	+	+		–		+	0	+	23,24,31, 134,147, 148
Mus musculus					+						0	149,150, 151
Birds												
Gallus gallus					+				0		0	152,153
Insects												
Anopheles gambiae			+		+				(+)			154
Drosophila melanogaster	+	S	+	+	+	–		+	+	+	0	6,10,29, 39,40,49, 155–157
Drosophila persimilis					+			+			0	132
Drosophila pseudoobscura	+	M	(+)		+	–				0	0(+)	106,135, 158
Drosophila simulans	+	S	+		+	–	–				(+)	30,47
Nematodes												
Caenorhabditis briggsae		S		+	+				(+)	–	(+)	59
Caenorhabditis elegans	+	S	+	+	+		–	0	+	(+)	0(+)	62,63, 159,160
Caenorhabditis remanei					0						(0)	161
Fungi												
Saccharomyces cerevisiae	+	W			+	–	(+)	(+)	0	(–)	(–)	56,137, 162–165
Plants												
Aegilops spp.					+							166
Arabidopsis thaliana	+	W		+	0(+)	0	–	0	0	0		32,55, 167–169
Arabidopsis lyrata		W		–	0		0	0		0	0	54,168
Beta vulgaris					+						0	170
Lycopersicon spp.		W		+	0(+)						0	53,116, 171,172
Medicago truncatula	+			+	+		–					173
Oryza sativa		W	+	+	–		–				+	36,174
Oryza rufipogon		M		+	0/–		–				+	36
Zea mays		W	–	+	0(+)				0		0	57,175,176

*Studies of non-recombining chromosomes not included. ‡'+' indicates affirmative evidence, '–' indicates refuting evidence. §'+' indicates positive correlation, '–' indicates negative correlation, '0' indicates no correlation; values in parentheses indicate a weak or marginally significant effect. BGS, background selection; M, moderate; RHH, recurrent genetic hitch-hiking; S, strong; SLS, selection at linked sites; W, weak.

© 2013 Macmillan Publishers Limited. All rights reserved

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 to be insufficient to account for correlations between polymorphism and recombination at genome scales.

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 by genetic hitch-hiking) to fixation that is caused by positive selection.

Genetic hitch-hiking
The process by which a neutral, or in some cases deleterious, mutations may change in population frequency owing to linkage with a selected mutation.

Site frequency spectrum
The distribution of allele frequencies in a population; it is visualized as the histogram of counts of the number of alleles that have a given population frequency.

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 owing to various factors, including variance in reproductive success, a history of population bottlenecks and inbreeding.

Background selection
The elimination of neutral polymorphisms as a result of their linkage to deleterious mutations that are subject to purifying selection.

Neutral polymorphism
Alternative allelic variants with no selective difference between them, the dynamics of which are mainly controlled by genetic drift and migration. They can, however, be influenced by selection on nearby (linked) loci.

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].

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 nonsynonymous sites; high d_N) will exhibit depressed polymorphism at linked neutral sites (for example, low polymorphism at synonymous sites (π_S))[26,29,30]. Although 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 (for example, replacement sites) but not substitutions at unconstrained sites (for example, 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 favouring 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, which contrasts fixed differences and polymorphisms for neutral and selection candidate site classes[41,42]. In *Drosophila* spp. 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* spp., typically <5% of replacement sites in genes have diverged, and >95% of replacement sites are conserved[42,46].

The inconsistent evidence among species

Drosophila research has spearheaded much of empirical population genetics, and this has also 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 species of *Drosophila*, which is consistent with an important influence of recurrent genetic hitch-hiking[26,48,49]. It nevertheless remains unclear how positive and negative selection each contribute in relative terms to genomic heterogeneity in patterns of polymorphism in *Drosophila* spp., suggesting that sophisticated modelling incorporating both positive and negative selection can have 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

© 2013 Macmillan Publishers Limited. All rights reserved

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 hitch-hiking[51]. Regardless of such details, it is now clear that current inferences regarding recurrent hitch-hiking effects for *Drosophila* spp. do not universally apply across organisms[52] (FIG. 3; TABLE 1).

Many plants and yeast exhibit weak-at-best signatures of linked selection[53–58], and humans and self-fertilizing 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 data sets of polymorphism from tens of loci, it is now investigated using population genomic data sets 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 hitch-hiking on genomic patterns of polymorphism only for *Drosophila* spp.[35,47] and, in conjunction with background selection, in *Caenorhabditis elegans*[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 (FIG. 2B). Broadly, organisms might differ in: how selection operates in their genomes; their profile of recombination rate; and the 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 to 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] implicate 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 hitch-hiking patterns.

Figure 2 | **Modes of selection on linked genetic variation and factors affecting them. A** | In the top two panels, horizontal lines represent haplotypes in a population at a particular genomic locus, and distinct colours represent different genetic backgrounds (compared with 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 (that is, 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 (that is, 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 hitch-hiking 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.

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 operates 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 data sets 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].

Replacement sites
Also known as nonsynonymous sites, these are any nucleotides 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.

d_N
The rate of protein-coding sequence divergence that is quantified as the number of nonsynonymous substitutions per nonsynonymous site.

Synonymous sites
Any sites within a gene at which some or all possible point mutations, depending on the degeneracy of the corresponding codon, do not change the encoded amino acid. Changes at synonymous sites are often presumed to be selectively neutral.

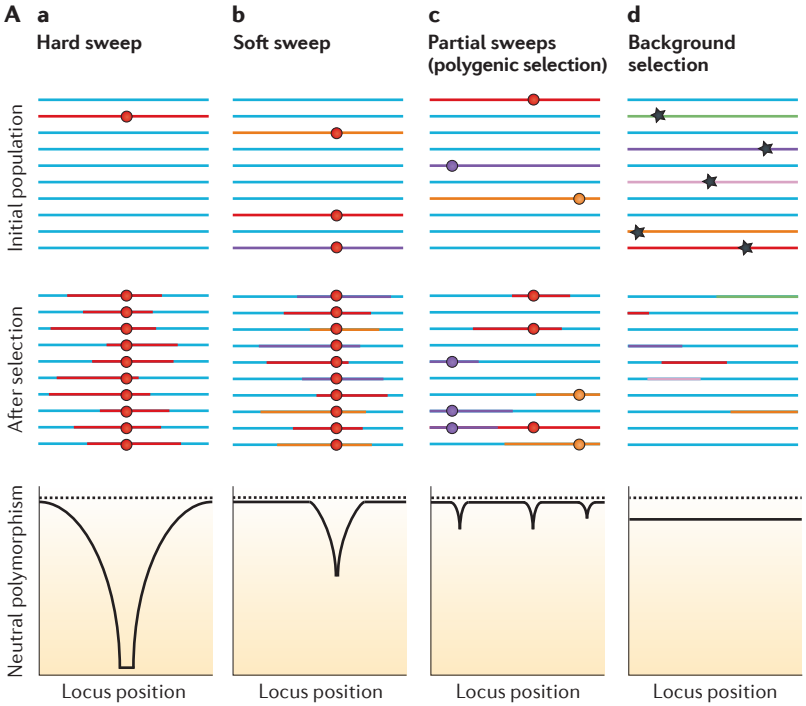
Linkage disequilibrium
A measure of whether alleles at two loci coexist in a population in a nonrandom fashion. Alleles that are in linkage disequilibrium are found together on the same haplotype more often than would be expected under free recombination.

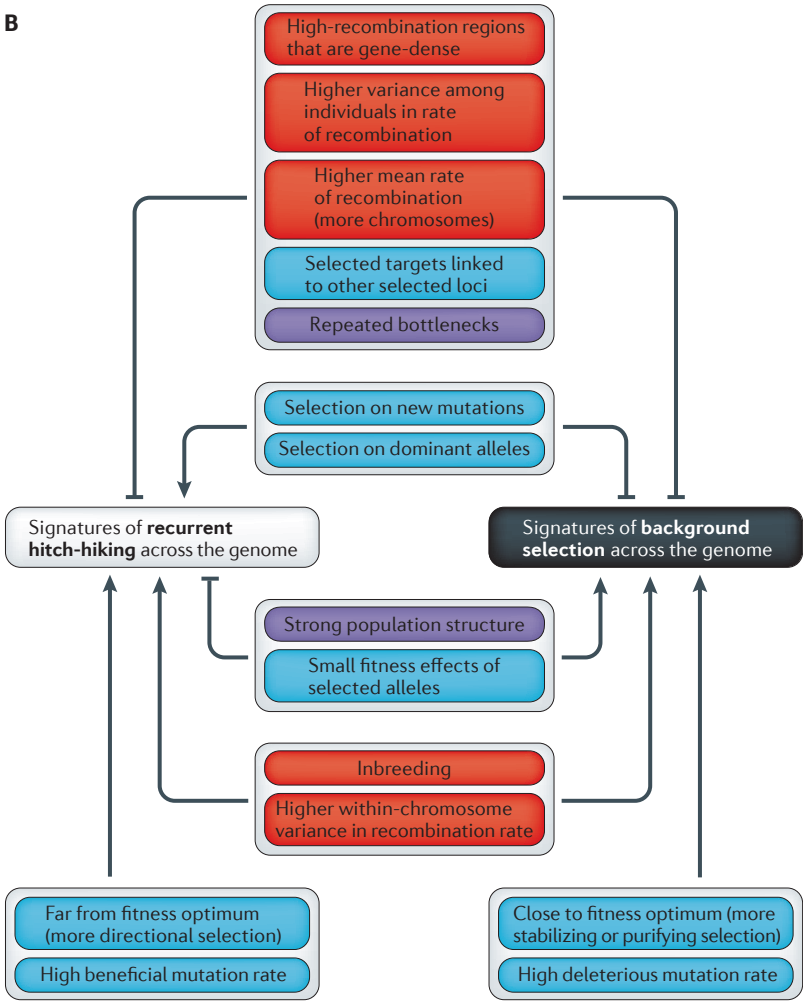
McDonald–Kreitman test
A statistical test used to compare between-species divergence and within-species polymorphism at replacement and synonymous sites to infer selection acting on proteins.

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.

© 2013 Macmillan Publishers Limited. All rights reserved





Mutation frequencies when selection begins. Selective sweeps are stronger when positive selection targets newer mutations because the beneficial mutation occurs on only one genetic background or on a small number of them. 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 hitch-hiking model of positive directional selection, whereby selection iteratively acts on single, new beneficial mutations. This view also is encapsulated in modelling 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* spp., which are the pre-eminent systems 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 through 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 hitch-hiking are often undetectable[75,80]. In particular, soft selective sweeps yield only mild reductions in linked polymorphism (FIG. 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 through 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 depress levels of polymorphism strongly 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 are rapidly driven to fixation 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).

© 2013 Macmillan Publishers Limited. All rights reserved

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 effectively removed, 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, resulting in 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 interference from selection on nearby non-coding regulatory regions has not been explicitly considered in this context. Notably, wHRi, background selection, genetic hitch-hiking and Muller's ratchet are all 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 rescaling of N_e[16–19].

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 compared with smaller populations.

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.

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 efficiently removed from the population (that is, 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 that 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 the magnitude of selection across loci, evolutionary responses to such polygenic selection need not even result in fixation of alleles at any given locus[90,91] (FIG. 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 hitch-hike 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 hitch-hiked with the original beneficial mutation[18]. Moreover, the linkage to deleterious mutations slows the fixation process and increases the incidence of recombinant genetic backgrounds, obscuring 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, and this 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 largely been explored through simulation, and a complete mathematical association with recombination rate has not yet been established[98], although fitness-class coalescent modelling appears to be 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 self-fertilizing species, selection is akin to additive selection operating in out-crossing diploids but is 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 both positive and negative selection 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 is often the selective agent that is of most biological interest. Much effort has been devoted to distinguishing background selection from hitch-hiking effects of recurrent selective sweeps[100], but this is difficult and is perhaps a false dichotomy[101]; we advocate future efforts to integrate them. To this end, background selection should be taken into account as a part of a null model of molecular evolution, because deleterious mutations arise inexorably and in turn are subjected to purifying selection[23,24]. Even more elaborate null models could be envisioned: 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 accommodate appropriately the distributions for coefficients of selection and dominance.

© 2013 Macmillan Publishers Limited. All rights reserved

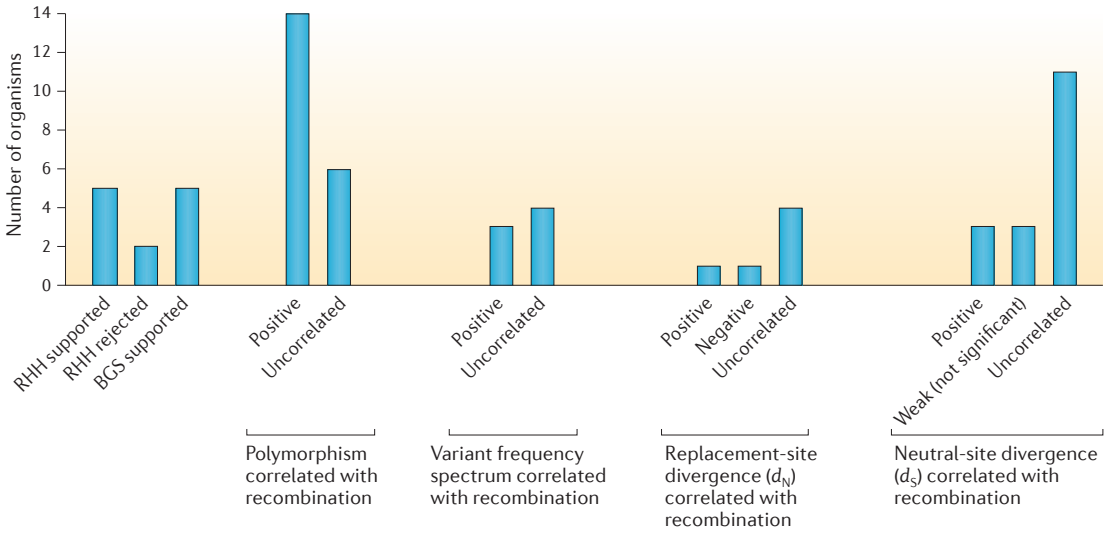


Figure 3 | **Taxonomic support for different signatures of selection at linked sites.** Species differ in their support for how selection and recombination interact to modulate patterns of molecular evolution. A summary is given of species signatures of linked selection for those species listed in TABLE 1; a given species may be included in multiple categories in the figure. Theory predicts recurrent genetic hitch-hiking (RHH) to generate positive correlations of recombination rate with neutral polymorphism, the site frequency spectrum and d_N. Background selection (BGS) is also predicted to yield positive correlations of recombination rate with neutral polymorphism.

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 population by mutation and their loss by genetic drift.

Population structure
The distribution of individuals into partially isolated, local subpopulations or demes that are interconnected by migration (gene flow).

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.

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 recreated. The process is named after H. J. Muller, who described it in 1964.

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.

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 *D. melanogaster*, *D. simulans*, *C. 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]. Conversely, *Drosophila 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 showing less crossing over[107]. Our ability to detect linked selection also crucially depends 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 rapidly evolve between closely related species[9,106,108,109].

Because the likelihood that an allele exerts an influence on neighbouring 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* spp. have a higher gene density in low-recombination regions, which probably contributes to especially strong signatures of linked selection in self-fertilizing 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 analysed 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 explicitly accounted for[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]. Self-fertilization, which is common in both plants and animals, reduces the genetically effective rate of recombination across the genome as a by-product of eliminating heterozygosity[115]. Because even modest levels of recombination can mitigate the signature of linked selection, self-fertilization broadens the extent

© 2013 Macmillan Publishers Limited. All rights reserved

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.

Purifying selection
Natural selection against deleterious alleles that arise in a population, preventing their increase in frequency.

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.

of the genome that can reveal its influence[20,53,116]. In addition, when linked selection occurs in self-fertilizers, it is more likely to facilitate 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 hitch-hiking 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 hitch-hiking.

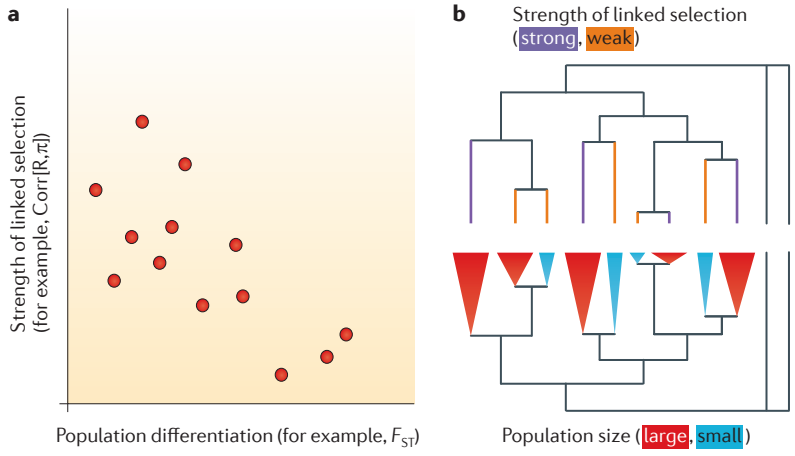
However, the effect of N_e on selection at linked sites is more complicated than it first appears. The transit time of a beneficial mutation grows with the population size (although only logarithmically), allowing more recombination to decouple the mutation from nearby neutral variants[12]. Recent theory also predicts that positive selection operates more often on standing variation when population mutation rates are high (that is, in large populations)[79,88], and the resulting soft sweeps induce a weaker hitch-hiking 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 positively correlate with population size after all. Even within a hard sweep framework, the 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 widely vary in N_e 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 to determine whether the effects of linked selection are responsible for the relatively small range of interspecies 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 hitch-hiking signature[12,87,118]. Analysing population genetic data according to different sampling schemes (such as 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], resulting in 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 with 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 N_e. In addition, repeated population bottlenecks will increase stochastic variance across the genome, perhaps erasing historically present signatures of selection at linked sites.

Towards phylogenetic coherence

The inconsistent evidence for genome-wide signatures of selection at linked sites across species (FIG. 3; TABLE 1) reveals a conspicuous gap in the current understanding of genome evolution. Why do species have such disparate genomic patterns of polymorphism? Phylogenetic comparative methods provide an established framework

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 a negative relationship would be predicted in a plot showing the strength of selection at linked sites (for example, the correlation coefficient between recombination rate and neutral polymorphism (Corr[R,π])) as a function of within-species population differentiation (for example, F_{ST} or alternative metrics), as in the schematic shown in panel **a** of the figure, in which 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 directly test for correlated evolution along the phylogeny, relating population genetic parameters, genomic features and other candidate variables[146]. Panel **b** of the figure illustrates the logic for a test for correlated character evolution on a phylogeny; taxa have 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).

© 2013 Macmillan Publishers Limited. All rights reserved

Transit time
The duration of time that elapses from when an allele first experiences selection to when it becomes fixed in a population.

F_{ST}
A measure of population subdivision that indicates the proportion of genetic diversity found between populations relative to the amount within populations.

Population bottlenecks
Marked reductions 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.

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 (BOX 3; FIG. 2B).

Some first attempts in this vein have compared species of rice[36] and tomatoes[53], and these 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 to be more challenging to test directly for co-varying genomic variables, such as gene density and recombination rate, because they are unlikely to have greatly diverged among close relatives. However, teasing apart the influence of factors such as mating system, the extent of population subdivision and N_e 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[9] (TABLE 1). However, species dramatically differ in the magnitude of effects exerted by linked selection, in how the patterns manifest in their genomes (for example, 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 hitch-hiking, despite it being well-appreciated that both processes probably operate simultaneously[101].

Moreover, analyses of recurrent hitch-hiking 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 describe better 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 a 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 have 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.

1. Wiehe, T. H. E. & Stephan, W. Analysis of a genetic hitchhiking model, and its application to DNA polymorphism data from *Drosophila melanogaster*. *Mol. Biol. Evol.* **10**, 842–854 (1993).
2. Hudson, R. R. & Kaplan, N. L. Deleterious background selection with recombination. *Genetics* **141**, 1605–1617 (1995).
3. Maynard Smith, J. & Haigh, J. Hitch-hiking effect of a favorable gene. *Genet. Res.* **23**, 23–35 (1974).
This is the classic theoretical study that describes the process of genetic hitch-hiking.
4. Hill, W. G. & Robertson, A. Effect of linkage on limits to artificial selection. *Genet. Res.* **8**, 269–294 (1966).
Here, the classic model is presented of how linked selected loci interfere with each other's ability to fix the favoured alleles in the population.
5. Felsenstein, J. The evolutionary advantage to recombination. *Genetics* **78**, 737–756 (1974).
6. Begun, D. J. & Aquadro, C. F. Levels of naturally-occurring DNA polymorphism correlate with recombination rates in *Drosophila melanogaster*. *Nature* **356**, 519–520 (1992).
This seminal study demonstrates that selection at linked sites yields a general pattern on genetic diversity.
7. Aguade, M., Miyashita, N. & Langley, C. H. Reduced variation in the yellow-achaete-scute region in natural populations of *Drosophila melanogaster*. *Genetics* **122**, 607–615 (1989).
8. Stephan, W. & Langley, C. H. 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* **121**, 89–99 (1989).
9. Smukowski, C. S. & Noor, M. A. Recombination rate variation in closely related species. *Heredity* **107**, 496–508 (2011).
10. Massouras, A. *et al.* Genomic variation and its impact on gene expression in *Drosophila melanogaster*. *PLoS Genet.* **8**, e1003055 (2012).
11. Leffler, E. M. *et al.* Revisiting an old riddle: what determines genetic diversity levels within species? *PLoS Biol.* **10**, e1001388 (2012).
12. Barton, N. H. Genetic hitchhiking. *Phil. Trans. R. Soc. B* **355**, 1553–1562 (2000).
13. Tajima, F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585–595 (1989).
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* **108**, 167–178 (2012).
15. Kaplan, N. L., Hudson, R. R. & Langley, C. H. The "hitchhiking effect" revisited. *Genetics* **123**, 887–899 (1989).
The important theoretical extension of genetic hitch-hiking is described here — from a single episode of selection to recurrent selective sweeps — in a genome that has variable recombination rates across different genomic regions.
16. Charlesworth, B., Betancourt, A. J., Kaiser, V. B. & Gordo, I. Genetic recombination and molecular evolution. *Cold Spring Harb. Symp. Quant. Biol.* **74**, 177–186 (2009).
17. Comeron, J. M., Williford, A. & Kliman, R. M. The Hill–Robertson effect: evolutionary consequences of weak selection and linkage in finite populations. *Heredity* **100**, 19–31 (2008).
18. Barton, N. H. Genetic linkage and natural selection. *Phil. Trans. R. Soc.* **365**, 2559–2569 (2010).

© 2013 Macmillan Publishers Limited. All rights reserved

19. Walczak, A. M., Nicolaisen, L. E., Plotkin, J. B. & Desai, M. M. The structure of genealogies in the presence of purifying selection: a fitness-class coalescent. *Genetics* **190**, 753–779 (2012).
20. Charlesworth, B., Morgan, M. T. & Charlesworth, D. The effect of deleterious mutations on neutral molecular variation. *Genetics* **134**, 1289–1303 (1993).
This classic paper introduces the notion that background selection reduces neutral genetic variation disproportionately in low-recombination regions, similarly to the effects of recurrent selective sweeps.
21. Charlesworth, B. The effects of deleterious mutations on evolution at linked sites. *Genetics* **190**, 5–22 (2012).
22. Eyre-Walker, A. & Keightley, P. D. The distribution of fitness effects of new mutations. *Nature Rev. Genet.* **8**, 610–618 (2007).
23. Lohmueller, K. E. *et al.* Natural selection affects multiple aspects of genetic variation at putatively neutral sites across the human genome. *PLoS Genet.* **7**, e1002326 (2011).
This exemplary recent empirical study incorporates background selection as a part of the null model of molecular evolution shaping genome-wide patterns of genetic variation in humans.
24. McVicker, G., Gordon, D., Davis, C. & Green, P. Widespread genomic signatures of natural selection in Hominid evolution. *PLoS Genet.* **5**, e1000471 (2009).
25. 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).
26. Sella, G., Petrov, D. A., Przeworski, M. & Andolfatto, P. Pervasive natural selection in the *Drosophila* genome? *PLoS Genet.* **5**, e1000495 (2009).
27. Charlesworth, D., Charlesworth, B. & Morgan, M. T. The pattern of neutral molecular variation under the background selection model. *Genetics* **141**, 1619–1632 (1995).
28. Kaiser, V. B. & Charlesworth, B. The effects of deleterious mutations on evolution in non-recombining genomes. *Trends Genet.* **25**, 9–12 (2009).
29. Andolfatto, P. Hitchhiking effects of recurrent beneficial amino acid substitutions in the *Drosophila melanogaster* genome. *Genome Res.* **17**, 1755–1762 (2007).
30. Macpherson, J. M., Sella, G., Davis, J. C. & Petrov, D. A. Genomewide spatial correspondence between nonsynonymous divergence and neutral polymorphism reveals extensive adaptation in *Drosophila*. *Genetics* **177**, 2083–2099 (2007).
31. Payseur, B. A. & Nachman, M. W. Gene density and human nucleotide polymorphism. *Mol. Biol. Evol.* **19**, 336–340 (2002).
32. Kim, S. *et al.* Recombination and linkage disequilibrium in *Arabidopsis thaliana*. *Nature Genet.* **39**, 1151–1155 (2007).
33. Hinds, D. A. *et al.* Whole-genome patterns of common DNA variation in three human populations. *Science* **307**, 1072–1079 (2005).
34. Hernandez, R. D. *et al.* Classic selective sweeps were rare in recent human evolution. *Science* **331**, 920–924 (2011).
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.
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 Genet.* **7**, e1001302 (2011).
36. Flowers, J. M. *et al.* Natural selection in gene-dense regions shapes the genomic pattern of polymorphism in wild and domesticated rice. *Mol. Biol. Evol.* **29**, 675–687 (2012).
Here, the authors describe an important empirical demonstration of the confounding influence of genomic variables in detecting selection at linked sites.
37. Barton, N. H. Linkage and the limits to natural selection. *Genetics* **140**, 821–841 (1995).
38. Betancourt, A. J. & Presgraves, D. C. Linkage limits the power of natural selection in *Drosophila*. *Proc. Natl Acad. Sci. USA* **99**, 13616–13620 (2002).
39. Presgraves, D. C. Recombination enhances protein adaptation in *Drosophila melanogaster*. *Curr. Biol.* **15**, 1651–1656 (2005).
40. Mackay, T. F. *et al.* The *Drosophila melanogaster* genetic reference panel. *Nature* **482**, 173–178 (2012).
41. McDonald, J. H. & Kreitman, M. Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature* **351**, 652–654 (1991).
42. Eyre-Walker, A. & Keightley, P. D. Estimating the rate of adaptive molecular evolution in the presence of slightly deleterious mutations and population size change. *Mol. Biol. Evol.* **26**, 2097–2108 (2009).
43. Andolfatto, P., Wong, K. M. & Bachtrog, D. Effective population size and the efficacy of selection on the X chromosomes of two closely related *Drosophila* species. *Genome Biol. Evol.* **3**, 114–128 (2011).
44. Fay, J. C. Weighing the evidence for adaptation at the molecular level. *Trends Genet.* **27**, 343–349 (2011).
45. Gossmann, T. I., Keightley, P. D. & Eyre-Walker, A. The effect of variation in the effective population size on the rate of adaptive molecular evolution in eukaryotes. *Genome Biol. Evol.* **4**, 658–667 (2012).
46. Ellegren, H. Comparative genomics and the study of evolution by natural selection. *Mol. Ecol.* **17**, 4586–4596 (2008).
47. Begun, D. J. *et al.* Population genomics: whole-genome analysis of polymorphism and divergence in *Drosophila simulans*. *PLoS Biol.* **5**, e310 (2007).
48. Hahn, M. W. Toward a selection theory of molecular evolution. *Evolution* **62**, 255–265 (2008).
49. Langley, C. H. *et al.* Genomic variation in natural populations of *Drosophila melanogaster*. *Genetics* **192**, 533–598 (2012).
This is a thorough population genomic analysis relating polymorphism characteristics to recombination profiles.
50. Shapiro, J. A. *et al.* Adaptive genic evolution in the *Drosophila* genomes. *Proc. Natl Acad. Sci. USA* **104**, 2271–2276 (2007).
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* **191**, 233–246 (2012).
52. Wright, S. I. & Andolfatto, P. The impact of natural selection on the genome: emerging patterns in *Drosophila* and *Arabidopsis*. *Annu. Rev. Ecol. Evol. Syst.* **39**, 193–213 (2008).
53. Baudry, E., Kerdelhue, C., Innan, H. & Stephan, W. Species and recombination effects on DNA variability in the tomato genus. *Genetics* **158**, 1725–1735 (2001).
54. Wright, S. I. *et al.* Testing for effects of recombination rate on nucleotide diversity in natural populations of *Arabidopsis lyrata*. *Genetics* **174**, 1421–1430 (2006).
55. Nordborg, M. *et al.* The pattern of polymorphism in *Arabidopsis thaliana*. *PLoS Biol.* **3**, e196 (2005).
56. Cutter, A. D. & Moses, A. M. Polymorphism, divergence, and the role of recombination in *Saccharomyces cerevisiae* genome evolution. *Mol. Biol. Evol.* **28**, 1745–1754 (2011).
57. Tenaillon, M. I. *et al.* Patterns of diversity and recombination along chromosome 1 of maize (*Zea mays* ssp. *mays* L.). *Genetics* **162**, 1401–1413 (2002).
58. Cao, J. *et al.* Whole-genome sequencing of multiple *Arabidopsis thaliana* populations. *Nature Genet.* **43**, 956–963 (2011).
59. Cutter, A. D. & Choi, J. Y. Natural selection shapes nucleotide polymorphism across the genome of the nematode *Caenorhabditis briggsae*. *Genome Res.* **20**, 1103–1111 (2010).
60. Rockman, M. V., Skrovanek, S. S. & Kruglyak, L. Selection at linked sites shapes heritable phenotypic variation in *C. elegans*. *Science* **330**, 372–376 (2010).
This is an intriguing demonstration of how selection at linked sites can affect phenotypic (gene expression) variation in addition to neutral nucleotide variation.
61. Hobolth, A., Dutheil, J. Y., Hawks, J., Schierup, M. H. & Mailund, T. Incomplete lineage sorting patterns among human, chimpanzee, and orangutan suggest recent orangutan speciation and widespread selection. *Genome Res.* **21**, 349–356 (2011).
62. Cutter, A. D. & Payseur, B. A. Selection at linked sites in the partial selfer *Caenorhabditis elegans*. *Mol. Biol. Evol.* **20**, 665–673 (2003).
63. Andersen, E. C. *et al.* Chromosome-scale selective sweeps shape *Caenorhabditis elegans* genomic diversity. *Nature Genet.* **44**, 285–290 (2012).
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.
64. Baer, C. F., Miyamoto, M. M. & Denver, D. R. Mutation rate variation in multicellular eukaryotes: causes and consequences. *Nature Rev. Genet.* **8**, 619–631 (2007).
65. Denver, D. R. *et al.* Variation in base-substitution mutation in experimental and natural lineages of *Caenorhabditis* nematodes. *Genome Biol. Evol.* **4**, 513–522 (2012).
66. Haag-Liautard, C. *et al.* Direct estimation of per nucleotide and genomic deleterious mutation rates in *Drosophila*. *Nature* **445**, 82–85 (2007).
67. Lynch, M. *et al.* A genome-wide view of the spectrum of spontaneous mutations in yeast. *Proc. Natl Acad. Sci. USA* **105**, 9272–9277 (2008).
68. Ossowski, S. *et al.* The rate and molecular spectrum of spontaneous mutations in *Arabidopsis thaliana*. *Science* **327**, 92–94 (2010).
69. Kong, A. *et al.* Rate of *de novo* mutations and the importance of father's age to disease risk. *Nature* **488**, 471–475 (2012).
70. Schneider, A., Charlesworth, B., Eyre-Walker, A. & Keightley, P. D. A method for inferring the rate of occurrence and fitness effects of advantageous mutations. *Genetics* **189**, 1427–1437 (2011).
71. Gossmann, T. I. *et al.* Genome wide analyses reveal little evidence for adaptive evolution in many plant species. *Mol. Biol. Evol.* **27**, 1822–1832 (2010).
72. Gillespie, J. H. Is the population size of a species relevant to its evolution? *Evolution* **55**, 2161–2169 (2001).
73. Coop, G. *et al.* The role of geography in human adaptation. *PLoS Genet.* **5**, e1000500 (2009).
74. Granka, J. M. *et al.* Limited evidence for classic selective sweeps in African populations. *Genetics* **192**, 1049–1064 (2012).
75. 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).
76. 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).
77. Miller, C. R., Joyce, P. & Wichman, H. A. Mutational effects and population dynamics during viral adaptation challenge current models. *Genetics* **187**, 185–202 (2011).
78. Lang, G. I., Botstein, D. & Desai, M. M. Genetic variation and the fate of beneficial mutations in asexual populations. *Genetics* **188**, 647–661 (2011).
79. Hermisson, J. & Pennings, P. S. Soft sweeps: molecular population genetics of adaptation from standing genetic variation. *Genetics* **169**, 2335–2352 (2005).
This important theoretical study models selection from standing genetic variation, to produce soft sweeps.
80. Storz, J. F. & Wheat, C. W. Integrating evolutionary and functional approaches to infer adaptation at specific loci. *Evolution* **64**, 2489–2509 (2010).
81. Rouzine, I. M. & Coffin, J. M. Multi-site adaptation in the presence of infrequent recombination. *Theor. Popul. Biol.* **77**, 189–204 (2010).
82. Przeworski, M., Coop, G. & Wall, J. D. The signature of positive selection on standing genetic variation. *Evolution* **59**, 2312–2323 (2005).
83. Innan, H. & Kim, Y. Pattern of polymorphism after strong artificial selection in a domestication event. *Proc. Natl Acad. Sci. USA* **101**, 10667–10672 (2004).
84. Ohta, T. & Kimura, M. Effect of selected linked locus on heterozygosity of neutral alleles (hitch-hiking effect). *Genet. Res.* **25**, 313–326 (1975).
85. Coop, G. & Ralph, P. Patterns of neutral diversity under general models of selective sweeps. *Genetics* **192**, 205–224 (2012).
This recent theoretical extension of recurrent positive selection to partial sweeps helps to generalize the expectations for patterns of genetic variation in genomes.
86. Kim, Y. & Maruki, T. Hitchhiking effect of a beneficial mutation spreading in a subdivided population. *Genetics* **189**, 213–226 (2011).
87. Bierne, N. The distinctive footprints of local hitchhiking in a varied environment and global hitchhiking in a subdivided population. *Evolution* **64**, 3254–3272 (2010).

© 2013 Macmillan Publishers Limited. All rights reserved

88. Ralph, P. & Coop, G. Parallel adaptation: one or many waves of advance of an advantageous allele? *Genetics* **186**, 647–668 (2010).
89. Teshima, K. M. & Przeworski, M. Directional positive selection on an allele of arbitrary dominance. *Genetics* **172**, 713–718 (2006).
90. Turchin, M. C. *et al.* Evidence of widespread selection on standing variation in Europe at height-associated SNPs. *Nature Genet.* **44**, 1015–1019 (2012).
91. Pritchard, J. K. & Di Rienzo, A. Adaptation — not by sweeps alone. *Nature Rev. Genet.* **11**, 665–667 (2010).
92. Pavlidis, P., Metzler, D. & Stephan, W. Selective sweeps in multilocus models of quantitative traits. *Genetics* **192**, 225–239 (2012).
93. Chevin, L. M. & Hospital, F. Selective sweep at a quantitative trait locus in the presence of background genetic variation. *Genetics* **180**, 1645–1660 (2008).
94. Chun, S. & Fay, J. C. Evidence for hitchhiking of deleterious mutations within the human genome. *PLoS Genet.* **7**, e1002240 (2011).
95. Hartfield, M. & Otto, S. P. Recombination and hitchhiking of deleterious alleles. *Evolution* **65**, 2421–2434 (2011).
96. Hadany, L. & Feldman, M. W. Evolutionary traction: the cost of adaptation and the evolution of sex. *J. Evol. Biol.* **18**, 309–314 (2005).
97. Boyko, A. R. *et al.* Assessing the evolutionary impact of amino acid mutations in the human genome. *PLoS Genet.* **4**, e1000083 (2008).
98. Zeng, K. & Charlesworth, B. The joint effects of background selection and genetic recombination on local gene genealogies. *Genetics* **189**, 251–266 (2011).
99. Gerstein, A. C. & Otto, S. P. Ploidy and the causes of genomic evolution. *J. Hered.* **100**, 571–581 (2009).
100. Stephan, W. Genetic hitchhiking versus background selection: the controversy and its implications. *Phil. Trans. R. Soc.* **365**, 1245–1253 (2010).
101. Hudson, R. R. 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* **91**, 6815–6818 (1994).
102. Fiston-Lavier, A. S., Singh, N. D., Lipatov, M. & Petrov, D. A. *Drosophila melanogaster* recombination rate calculator. *Gene* **463**, 18–20 (2010).
103. True, J. R., Mercer, J. M. & Laurie, C. C. Differences in crossover frequency and distribution among three sibling species of *Drosophila*. *Genetics* **142**, 507–523 (1996).
104. Hillier, L. W. *et al.* Comparison of *C. elegans* and *C. briggsae* genome sequences reveals extensive conservation of chromosome organization and synteny. *PLoS Biol.* **5**, e167 (2007).
105. Mancera, E., Bourgon, R., Brozzi, A., Huber, W. & Steinmetz, L. M. High-resolution mapping of meiotic crossovers and non-crossovers in yeast. *Nature* **454**, 479–485 (2008).
106. McGaugh, S. E. *et al.* Recombination modulates how selection affects linked sites in *Drosophila*. *PLoS Biol.* **10**, e1001422 (2012).
107. Kong, A. *et al.* Fine-scale recombination rate differences between sexes, populations and individuals. *Nature* **467**, 1099–1103 (2010).
108. Dumont, B. L., White, M. A., Steffy, B., Wiltshire, T. & Payseur, B. A. Extensive recombination rate variation in the house mouse species complex inferred from genetic linkage maps. *Genome Res.* **21**, 114–125 (2011).
109. Auton, A. *et al.* A fine-scale chimpanzee genetic map from population sequencing. *Science* **336**, 193–198 (2012).
110. Coop, G., Wen, X., Ober, C., Pritchard, J. K. & Przeworski, M. High-resolution mapping of crossovers reveals extensive variation in fine-scale recombination patterns among humans. *Science* **319**, 1395–1398 (2008).
111. Sandor, C. *et al.* Genetic variants in *REC8*, *RNF212*, and *PRDM9* influence male recombination in cattle. *PLoS Genet.* **8**, e1002854 (2012).
112. Ross, J. A. *et al. Caenorhabditis briggsae* recombinant inbred line genotypes reveal inter-strain incompatibility and the evolution of recombination. *PLoS Genet.* **7**, e1002174 (2011).
113. Comeron, J. M., Ratnappan, R. & Bailin, S. The many landscapes of recombination in *Drosophila melanogaster*. *PLoS Genet.* **8**, e1002905 (2012).
This is an important experimental demonstration of how different genetic backgrounds yield different profiles of recombination along their genomes.
114. Neher, R. A. & Shraiman, B. I. Genetic draft and quasi-neutrality in large facultatively sexual populations. *Genetics* **188**, 975–996 (2011).
115. Charlesworth, D. & Wright, S. I. Breeding systems and genome evolution. *Curr. Opin. Genet. Dev.* **11**, 685–690 (2001).
116. Innan, H. & Stephan, W. Distinguishing the hitchhiking and background selection models. *Genetics* **165**, 2307–2312 (2003).
117. Charlesworth, B. Fundamental concepts in genetics: effective population size and patterns of molecular evolution and variation. *Nature Rev. Genet.* **10**, 195–205 (2009).
118. Siol, M., Wright, S. I. & Barrett, S. C. H. The population genomics of plant adaptation. *New Phytol.* **188**, 313–332 (2010).
119. Wakeley, J. Nonequilibrium migration in human history. *Genetics* **153**, 1863–1871 (1999).
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* **182**, 205–216 (2009).
121. Cutter, A. D., Wang, G.-X., 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.* **21**, 1345–1359 (2012).
122. Chen, Y., Marsh, B. J. & Stephan, W. Joint effects of natural selection and recombination on gene flow between *Drosophila ananassae* populations. *Genetics* **155**, 1185–1194 (2000).
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.* **70**, 155–174 (1997).
124. Eldon, B. & Wakeley, J. Coalescent processes when the distribution of offspring number among individuals is highly skewed. *Genetics* **172**, 2621–2633 (2006).
125. Harvey, P. H. & Pagel, M. D. *The Comparative Method in Evolutionary Biology* (Oxford Univ. Press, 1991).
126. O'Brien, S. J. *et al.* The promise of comparative genomics in mammals. *Science* **286**, 458–481 (1999).
127. Steppan, S. J., Phillips, P. C. & Houle, D. Comparative quantitative genetics: evolution of the G matrix. *Trends Ecol. Evol.* **17**, 320–327 (2002).
128. Kimura, M. & Crow, J. F. The number of alleles that can be maintained in a finite population. *Genetics* **49**, 725–738 (1964).
129. Strathern, J. N., Shafer, B. K. & McGill, C. B. DNA-synthesis errors associated with double-strand-break repair. *Genetics* **140**, 965–972 (1995).
130. Holbeck, S. L. & Strathern, J. N. A role for REV3 in mutagenesis during double-strand break repair in *Saccharomyces cerevisiae*. *Genetics* **147**, 1017–1024 (1997).
131. Kimura, M. Evolutionary rate at molecular level. *Nature* **217**, 624–626 (1968).
132. Stevison, L. & Noor, M. Genetic and evolutionary correlates of fine-scale recombination rate variation in *Drosophila persimilis*. *J. Mol. Evol.* **71**, 332–345 (2010).
133. Hellmann, I., Ebersberger, I., Ptak, S. E., Paabo, S. & Przeworski, M. A neutral explanation for the correlation of diversity with recombination rates in humans. *Am. J. Hum. Genet.* **72**, 1527–1535 (2003).
134. Lercher, M. J. & Hurst, L. D. Human SNP variability and mutation rate are higher in regions of high recombination. *Trends Genet.* **18**, 337–340 (2002).
135. Kulathinal, R. J., Stevison, L. S. & Noor, M. A. F. The genomics of speciation in *Drosophila*: diversity, divergence, and introgression estimated using low-coverage genome sequencing. *PLoS Genet.* **5**, e1000550 (2009).
136. Huang, S. W., Friedman, R., Yu, N., Yu, A. & Li, W. H. How strong is the mutagenicity of recombination in mammals? *Mol. Biol. Evol.* **22**, 426–431 (2005).
137. Pal, C., Papp, B. & Hurst, L. D. Does the recombination rate affect the efficiency of purifying selection? The yeast genome provides a partial answer. *Mol. Biol. Evol.* **18**, 2323–2326 (2001).
138. McVean, G. A. T. & Charlesworth, B. A population genetic model for the evolution of synonymous codon usage: patterns and predictions. *Genet. Res.* **74**, 145–158 (1999).
139. Duret, L. Evolution of synonymous codon usage in metazoans. *Curr. Opin. Genet. Dev.* **12**, 640–649 (2002).
140. Hershberg, R. & Petrov, D. A. Selection on codon bias. *Annu. Rev. Genet.* **42**, 287–299 (2008).
141. Kliman, R. M. & Hey, J. Reduced natural selection associated with low recombination in *Drosophila melanogaster*. *Mol. Biol. Evol.* **10**, 1239–1258 (1993).
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* **98**, 5688–5692 (2001).
143. Comeron, J. M. & Kreitman, M. Population, evolutionary and genomic consequences of interference selection. *Genetics* **161**, 389–410 (2002).
144. Loewe, L. & Charlesworth, B. Background selection in single genes may explain patterns of codon bias. *Genetics* **175**, 1381–1393 (2007).
145. Strasburg, J. L. *et al.* Effective population size is positively correlated with levels of adaptive divergence among annual sunflowers. *Mol. Biol. Evol.* **28**, 1569–1580 (2011).
146. Lartillot, N. & Poujol, R. A phylogenetic model for investigating correlated evolution of substitution rates and continuous phenotypic characters. *Mol. Biol. Evol.* **28**, 729–744 (2011).
147. Nachman, M. W., Bauer, V. L., Crowell, S. L. & Aquadro, C. F. DNA variability and recombination rates at X-linked loci in humans. *Genetics* **150**, 1133–1141 (1998).
148. Bullaughey, K., Przeworski, M. & Coop, G. No effect of recombination on the efficacy of natural selection in primates. *Genome Res.* **18**, 544–554 (2008).
149. Nachman, M. W. Patterns of DNA variability at X-linked loci in *Mus domesticus*. *Genetics* **147**, 1303–1316 (1997).
150. Takahashi, A., Liu, Y. H. & Saitou, N. Genetic variation versus recombination rate in a structured population of mice. *Mol. Biol. Evol.* **21**, 404–409 (2004).
151. Geraldes, A., Basset, P., Smith, K. L. & Nachman, M. W. Higher differentiation among subspecies of the house mouse (*Mus musculus*) in genomic regions with low recombination. *Mol. Ecol.* **20**, 4722–4736 (2011).
152. Rao, Y., Sun, L., Nie, Q. & Zhang, X. The influence of recombination on SNP diversity in chickens. *Hereditas* **148**, 63–69 (2011).
153. Fang, L. *et al.* Positive correlation between recombination rate and nucleotide diversity is shown under domestication selection in the chicken genome. *Chinese Sci. Bull.* **53**, 746–750 (2008).
154. Stump, A. D. *et al.* Centromere-proximal differentiation and speciation in *Anopheles gambiae*. *Proc. Natl Acad. Sci. USA* **102**, 15930–15935 (2005).
155. Andolfatto, P. & Przeworski, M. Regions of lower crossing over harbor more rare variants in African populations of *Drosophila melanogaster*. *Genetics* **158**, 657–665 (2001).
This empirical study demonstrates how population genetic features other than simply the amount of diversity are affected by selection at linked sites.
156. Hey, J. & Kliman, R. M. Interactions between natural selection, recombination and gene density in the genes of *Drosophila*. *Genetics* **160**, 595–608 (2002).
157. Charlesworth, B. Background selection and patterns of genetic diversity in *Drosophila melanogaster*. *Genet. Res.* **68**, 131–149 (1996).
158. Kulathinal, R. J., Bennett, S. M., Fitzpatrick, C. L. & Noor, M. A. Fine-scale mapping of recombination rate in *Drosophila* refines its correlation to diversity and divergence. *Proc. Natl Acad. Sci. USA* **105**, 10051–10056 (2008).
159. 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.* **19**, 1399–1406 (2002).
160. Denver, D. R. *et al.* A genome-wide view of *Caenorhabditis elegans* base-substitution mutation processes. *Proc. Natl Acad. Sci. USA* **106**, 16310–16324 (2009).
161. Cutter, A. D. Multilocus patterns of polymorphism and selection across the X-chromosome of *Caenorhabditis remanei*. *Genetics* **178**, 1661–1672 (2008).
162. Noor, M. A. F. Mutagenesis from meiotic recombination is not a primary driver of sequence divergence between *Saccharomyces* species. *Mol. Biol. Evol.* **25**, 2439–2444 (2008).

© 2013 Macmillan Publishers Limited. All rights reserved

163. Connallon, T. & Knowles, L. L. Recombination rate and protein evolution in yeast. *BMC Evol. Biol.* **7**, 235 (2007).
164. Weber, C. C. & Hurst, L. D. Protein rates of evolution are predicted by double-strand break events, independent of crossing-over rates. *Genome Biol. Evol.* **1**, 340–349 (2009).
165. Kliman, R. M., Irving, N. & Santiago, M. Selection conflicts, gene expression, and codon usage trends in yeast. *J. Mol. Evol.* **57**, 98–109 (2003).
166. Dvorak, J., Luo, M. C. & Yang, Z. L. Restriction fragment length polymorphism and divergence in the genomic regions of high and low recombination in self-fertilizing and cross-fertilizing *Aegilops* species. *Genetics* **148**, 423–434 (1998).
167. Schmid, K. J., 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* **169**, 1601–1615 (2005).
168. Slotte, T. *et al.* Genomic determinants of protein evolution and polymorphism in *Arabidopsis*. *Genome Biol. Evol.* **3**, 1210–1219 (2011).
169. Marais, G., Charlesworth, B. & Wright, S. I. Recombination and base composition: the case of the highly self-fertilizing plant *Arabidopsis thaliana*. *Genome Biol.* **5**, R45 (2004).
170. Kraft, T., Sall, T., Magnusson-Rading, I., Nilsson, N. O. & Hallden, C. Positive correlation between recombination rates and levels of genetic variation in natural populations of sea beet (*Beta vulgaris* subsp. *maritima*). *Genetics* **150**, 1239–1244 (1998).
171. Roselius, K., Stephan, W. & Stadler, T. The relationship of nucleotide polymorphism, recombination rate and selection in wild tomato species. *Genetics* **171**, 753–763 (2005).
172. Stephan, W. & Langley, C. H. DNA polymorphism in *Lycopersicon* and crossing-over per physical length. *Genetics* **150**, 1585–1593 (1998).
173. Branca, A. *et al.* Whole-genome nucleotide diversity, recombination, and linkage disequilibrium in the model legume *Medicago truncatula*. *Proc. Natl Acad. Sci. USA* **108**, E864–E870 (2011).
174. Molina, J. *et al.* Molecular evidence for a single evolutionary origin of domesticated rice. *Proc. Natl Acad. Sci. USA* **108**, 8351–8356 (2011).
175. Tenaillon, M. I., U'Ren, J., Tenaillon, O. & Gaut, B. S. Selection versus demography: a multilocus investigation of the domestication process in maize. *Mol. Biol. Evol.* **21**, 1214–1225 (2004).
176. Thuillet, A. C. *et al.* A weak effect of background selection on trinucleotide microsatellites in maize. *J. Hered.* **99**, 45–55 (2008).

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

Competing interests statement
The authors declare no competing financial interests.

FURTHER INFORMATION

Asher D. Cutter's homepage: http://labs.eeb.utoronto.ca/cutter
Bret A. Payseur's homepage: http://payseur.genetics.wisc.edu

ALL LINKS ARE ACTIVE IN THE ONLINE PDF

© 2013 Macmillan Publishers Limited. All rights reserved