Direct and indirect consequences of meiotic recombination: implications for genome evolution

Matthew T. Webster1 and Laurence D. Hurst2

1 Science for Life Laboratory, Department of Medical Biochemistry and Microbiology, Uppsala University, Uppsala, Sweden
2 Department of Biology and Biochemistry, University of Bath, Bath, UK

There is considerable variation within eukaryotic genomes in the local rate of crossing over. Why is this and what effect does it have on genome evolution? On the genome scale, it is known that by shuffling alleles, recombination increases the efficacy of selection. By contrast, the extent to which differences in the recombination rate modulate the efficacy of selection between genomic regions is unclear. Recombination also has direct consequences on the origin and fate of mutations: biased gene conversion and other forms of meiotic drive promote the fixation of mutations in a similar way to selection, and recombination itself may be mutagenic. Consideration of both the direct and indirect effects of recombination is necessary to understand why its rate is so variable and for correct interpretation of patterns of genome evolution.

The evolutionary consequences of meiotic recombination

Meiosis involves the formation of DNA double-stranded breaks (DSBs), which are subsequently repaired by the process of homologous recombination. The functions of recombination are twofold: crossovers (COs) are necessary for accurate chromosomal disjunction in most eukaryotes [1] and the shuffling of alleles by recombination has a beneficial role for evolution [2,3]. However, recombination also has costs. For example, DSBs may aberrantly pair with non-homologous loci, leading to structural rearrangements via non-allelic homologous recombination (NAHR) [4]. Such exchanges are often lethal or highly deleterious [5]. These considerations suggest that the rate of crossing over should be tightly controlled. However, there is extensive variation in the frequency of crossing over both within and between genomes. What factors influence this variation and what other benefits and consequences does recombination have?

The major evolutionary advantage of recombination in eukaryotes is thought to be that it breaks up associations that occur between genetically linked variants, which allows natural selection to promote the fixation of beneficial mutations and removal of deleterious ones with greater efficacy [2]. This we term the indirect consequence of recombination. In theory, however, there are many conditions under which recombination is not advantageous. This is because associations between neighbouring loci may be either positive (when either beneficial variants or deleterious variants are linked) or negative (when beneficial and deleterious mutations are linked to each other) and there is only a net advantage to recombination if most associations are negative. In finite populations, negative associations accumulate between neighbouring loci owing to a process

---

Glossary

**Background selection**: reduction in linked genetic variation due to purifying selection at neighbouring loci.

**Codon bias**: most amino acids can be encoded by multiple codons. Codon bias is the unequal use of codons for a particular amino acid.

**Crossover (CO)**: exchange of genetic material between homologous chromosomes involving chromosomal breakage followed by rejoining to its homolog.

**Direct effects of recombination**: we define the direct effects of recombination as the transmission and mutational biases that may occur during meiosis.

**dH/dS**: the rate of nonsynonymous substitution (dH) relative to the rate of synonymous substitution (dS). This ratio is typically used to make inferences on the type and strength of selection on amino acid mutations acting on a gene. Assuming synonymous changes to be mostly neutral, positive selection on amino acid changes is expected to increase dH/dS, whereas negative selection decreases it.

**Effective population size (Ne)**: a fundamental concept in population genetics defined as the size of an ideal population that would experience an equal magnitude of genetic drift as the observed population.

**Fitness**: average reproductive success of an individual, or the effect a genetic variant has on this value. Individuals with greater fitness are likely to contribute more offspring to the next generation than are individuals with lesser fitness.

**Genetic drift**: random changes in allele frequency due to random sampling from the previous generation.

**Genetic hitchhiking**: process by which an evolutionarily neutral or deleterious allele spreads through the gene pool by virtue of being linked to a positively selected one.

**Hill–Robertson interference (HRI)**: build-up of negative linkage disequilibrium between linked loci due to selection. This association between variants with differing fitness effects reduces the efficacy of selection.

**Indirect effects of recombination**: we define the indirect effects of recombination as the breaking down of linkage disequilibrium between neighbouring loci owing to crossing over.

**Linkage disequilibrium (LD)**: nonrandom association of alleles at two or more loci.

**Meiotic drive**: unequal transmission of alleles from a heterozygote into its gametes.

**Positive selection**: tendency for selection to increase the frequency of beneficial variants towards fixation in a population.

**Purifying (negative) selection**: tendency for selection to reduce the population frequency of deleterious variants towards their loss from a population.

**Recombination hotspot**: localised regions of the genome (typically 1–2 kb) that exhibit elevated rates of meiotic recombination.

**Selective sweep**: process whereby an allele becomes more common in a population owing to positive selection. This typically results in a reduction in levels of linked variation by genetic hitchhiking.
known as the Hill–Robertson effect [6], which results from the interaction between genetic drift and selection. Selection works in opposite directions on negatively associated variants, which results in interference (known as Hill–Robertson interference; HRI; see Glossary). This reduces the efficacy of selection and provides an advantage to recombination [7].

The evolutionary advantage of recombination is supported by the observation that species reproducing completely asexually (and thus without recombination) are both extremely rare and typically short lived [8] and, similarly, that non-recombining genomic regions of sexually reproducing organisms accumulate deleterious mutations and deteriorate [9]. Experimental studies in nematodes and unicellular eukaryotes have demonstrated that strains undergoing recombination experience greater rates of adaptation when selection is strong, for example owing to new or fluctuating environmental conditions [10–12]. If recombination increases the rate of adaptation, then the recombination rate itself is predicted to be adaptive. Several studies have demonstrated that rates of recombination evolve in response to strong selection on traits unrelated to meiosis, such as when rotifers are forced to adapt in heterogeneous environments [13], or fruit flies are selected for geotaxis or DDT resistance [14]. It has also been suggested that domestic species exhibit increased recombination rates owing to the strong artificial selection they have experienced [15–17], although this awaits confirmation. However, the extent to which the rate of recombination limits adaptation in natural populations is unclear. Another important question, which we discuss here, is whether the local rate of recombination in a genomic region is adaptive.

More recently, evidence has accumulated that other recombination-associated processes can significantly influence genome evolution and play an important role in determining variation in the recombination rate. These include at least three forms of meiotic drive. The first, GC-biased gene conversion (gBGC) [18], results from biased incorporation of G and C nucleotides during repair of mismatches formed during homologous recombination. The second, which we term ‘hotspot drive’, results from the biased transmission of non-recombinogenic alleles over recombinogenic ones in hotspots of recombination [19]. The third, which we term ‘indel drive’, refers to the biased transmission of either the shorter or longer allele of an indel during meiosis. In addition to this, the contentious possibility that recombination generates point mutations has also been discussed [20]. We refer to these processes as the direct effects of recombination (further details are presented in Box 1 and Figure 1).

Although one of the advantages of the indirect effects of recombination is the efficient removal of deleterious alleles from a population, its direct effects may sometimes themselves be harmful. In particular, recent research has implicated gBGC in promoting fixation of weakly deleterious mutations [21]. In addition, the direct effects of recombination may play an important role in how recombination rates vary across the genome. For example, hotspot drive is believed to result in rapid shifts in the location of recombination hotspots [19,22].

Here, we propose that an understanding of both the direct and indirect effects of recombination is crucial for the interpretation of patterns of genome evolution and to explain variation in the recombination rate across the genome. There is now much evidence to suggest that selection influences the level of linked genetic variation across the genomes of eukaryotes, particularly in regions of low recombination rate [23–27]. This could also indicate that the efficacy of selection varies across the genome owing to variation in the degree of HRI mediated by recombination rate. If recombination rate places limits on the efficacy of selection and the rate of adaptation, then the local recombination rate could also be adaptive, leading to selection on local modifiers of recombination rate. However, what evidence is there that variation in the

Box 1. Direct effects of recombination

<table>
<thead>
<tr>
<th>GC-biased gene conversion (gBGC)</th>
<th>Indel drive</th>
</tr>
</thead>
<tbody>
<tr>
<td>During meiosis, gBGC causes heterozygotes with one AT and one GC allele to produce slightly more gametes containing the GC allele. This is hypothesised to result from a bias in the repair of mismatches caused by heterozygous sites occurring in heteroduplex DNA during recombination [18]. There is experimental evidence in yeast that gBGC is associated with meiotic gene conversion events [59] and indirect evidence that gBGC affects genome evolution and the segregation of alleles in many other taxa [18,59,73,102]. gBGC is thought to be responsible for the correlation between GC content and recombination [18] and the large-scale variation in GC content observed in many eukaryotic genomes [73]. It is also implicated in episodes of accelerated evolution in localised genomic regions [18,21,72,74].</td>
<td>Evidence of transmission bias at indel polymorphisms is limited. In Caenorhabditis elegans, chromosomes with insertions have been observed to segregate preferentially away from the X chromosome during meiosis, whereas chromosomes with deletions tend to be inherited together with the X chromosome [103]. In Drosophila melanogaster, preferential transmission of an allele containing a P-element insertion over a non-insertion allele has been observed [104]. Indel drive has also been observed in yeast and fungi, although its direction varies between loci and crosses [105,106].</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Hotspot drive</th>
<th>Mutagenic recombination</th>
</tr>
</thead>
<tbody>
<tr>
<td>Molecular genetic analysis of human recombination hotspots has shown that when an individual is heterozygous for a recombinogenic and non-recombinogenic allele, there is a bias in favour of transmitting the non-recombinogenic allele [63], a process inferred to be widespread using evolutionary genetic analyses [19]. Along with the rapid evolution of PRDM9 (see main text), this form of biased gene conversion is likely to be partly responsible for rapid shifts in the location of recombination hotspots [19,22,65].</td>
<td>In yeast, there is experimental evidence that mitotic recombination causes point mutations [83,84]. If meiotic recombination was also mutagenic, it could explain why levels of genetic variation correlate with the CO rate in some species [23,25,30,31,85]. However, whether this is true in yeast or any other species is far from resolved (see main text). It is clear, however, that recombination causes structural mutations via NAHR [4], including those involved in human disease [5]. In support of this, domains of high recombination tend to be more highly rearranged (e.g. [88]).</td>
</tr>
</tbody>
</table>
recombination rate leads to variation in the efficacy of selection across the genome? Could the direct effects of recombination affect measures of the efficacy of selection? What role do physical constraints have on determining recombination rate variation? In addition to its effect of breaking down genetic associations by shuffling alleles, we outline several other important ways in which recombination can potentially contribute to phenotypic evolution by altering the rate at which functional mutations occur and become fixed in populations.

Evidence of intragenomic variation in the efficacy of selection
Genomic regions that do not undergo meiotic recombination tend to accumulate deleterious mutations and lose functional genes [9], a process that is well documented in the neo-sex chromosomes of Drosophila miranda [28]. This supports the notion that recombination aids the efficient removal of deleterious mutations by selection. Both positive and negative selection are predicted to reduce the effective population size ($N_e$) at neighbouring loci, which reduces the local efficacy of selection owing to HRI [29]. Variation in the CO rate and density of sites under selection across the genome is therefore expected to generate variation in locus-specific $N_e$.

What evidence is there that variation in the CO rate across the genome outside non-recombining regions leads to variation in the efficacy of selection? Regions of low crossing over have reduced levels of neutral genetic variation in humans, Drosophila and some plants [23,26,30,31] (but not rice [32]). This is at least partly an effect of selection reducing levels of variation at linked sites, indicating that the evolutionary fates of mutations are affected by selection at linked loci, but whether this leads to substantial variation in the efficacy of selection is unclear. A study of variation in genome-wide patterns of neutral genetic diversity across the genomes of a variety of eukaryotes indicated that there is highly significant variation in $N_e$, although this variation showed no strong relationship with either the CO rate or the density of selected sites, leaving most of the intragenomic variation in $N_e$ unexplained [27].

Several other studies have specifically investigated whether the efficacy of selection varies according to the CO rate. Two measures have been used as proxies for the efficacy of selection: $d_N/d_S$ and codon usage bias. The $d_N/d_S$ ratio is a measure of the rate of protein adaptation. If most nonsynonymous mutations are weakly deleterious, then $d_N/d_S$ is likely to be reduced when selection is more efficient, because purifying selection would prevent more of these mutations rising to fixation. Conversely, if most nonsynonymous mutations are advantageous, then $d_N/d_S$ is likely to be increased when selection is more efficient, owing to a higher rate of adaptive nonsynonymous fixations. Codon usage bias is a measure of the proportion of ‘preferred’ codons in a gene, which is commonly used as an estimate of the intensity of selection in species with large population sizes (i.e. not mammals). Genes containing a higher proportion of preferred codons, which correspond to
the most abundant tRNAs, are often expressed more efficiently. There may therefore be a selective advantage for a gene to contain preferred codons. This leads to the prediction that codon usage bias should be greater in regions of the genome where selection is more efficient.

Studies in *Drosophila* have come to conflicting conclusions about the effect of the CO rate on the \( d_N/d_S \) ratio [33–36]. One genome-wide analysis found little evidence of differences in \( d_N \) or \( d_S \) between genomic regions with high and low CO rates, except in non-recombining domains, which have elevated \( d_N/d_S \) [35]. This may indicate that the efficacy of selection is reduced in non-recombining domains, but that low CO rates are sufficient to overcome the negative effects of HRI. A further genome-wide study [36] observed a weak but significant positive correlation between \( d_N/d_S \) and the recombination rate in a set of genes inferred to be under positive selection, and an opposite weak but significant negative correlation in all other genes. This was interpreted as an effect of HRI: if highly recombining regions experience more efficient selection, then genes under predominantly positive selection would experience elevated \( d_N/d_S \) in these regions, whereas those under predominantly negative selection would experience reduced \( d_N/d_S \). It was argued that these results were consistent with this scenario [36]. However, even in positively selected genes, most nonsynonymous variants could still be under purifying selection, so the net effect of differences in the efficacy of selection on \( d_N/d_S \) is difficult to predict.

In humans, it has been found that the average \( d_N/d_S \) is constant across CO rate categories, which may be evidence of an absence of variation in the efficacy of selection according to the genomic recombination environment [37]. Notably, in contrast to the studies in *Drosophila*, there is no evidence of elevated \( d_N/d_S \) in regions of no recombination in humans. In yeasts, there is a strong negative correlation between the local recombination rate and \( d_N \) and \( d_S \)[38]. However, this is largely explained as a covariate to the expression rate of the genes concerned, with slow-evolving highly expressed genes tending to be found disproportionately in regions of high recombination rate [38–40]. Indeed, the rate of pre-meiotic DSB events (which should have no effect on HRI) is also a strong predictor of rates of evolution, in part because this too is correlated with the expression rate [39]. There is no evidence of substantial variation in the efficacy of selection across the yeast genome owing to recombination when these factors are corrected for [38,39].

There is a slight positive correlation between the proportion of preferred codons and the CO rate in *Drosophila* [18]. One interpretation of this is that selection for preferred codons is more efficient in regions of elevated recombination [41]. However, preferred codons in *Drosophila* all end in G or C, and it has been demonstrated that the GC content of noncoding DNA is higher in regions of elevated recombination [42]. It is therefore possible that silent sites in coding regions and surrounding noncoding DNA both become GC-rich through a neutral effect of recombination on GC content owing to gGC. Variation in codon usage patterns across *Caenorhabditis elegans* and *Drosophila* could both be explained by this effect [43]. It has also been suggested that high rates of adaptive protein evolution on strongly selected genes could actually reduce levels of codon bias. This is because strong selection on nonsynonymous sites could interfere with weak selection on nearby synonymous sites owing to HRI [44].

There are several reasons why interpretation of the results presented above is problematic. First, without knowledge of the distribution of positive and negative fitness effects on each gene, it is impossible to know a priori whether a positive or negative correlation between \( d_N/d_S \) and the CO rate is expected. Any interpretation of a correlation, be it positive or negative, therefore comes with after-the-fact supposition of the form of the distribution. Unfortunately then, many results are in principle consistent with HRI, rendering the test weak by necessity. Second, the analyses are based on currently observed recombination rates, but these need not reflect the recombinational history of the gene if the local recombination rate is fast evolving [45]. Third, there are many other genomic features that correlate with recombination, which may also influence \( d_N \), \( d_S \) and codon usage bias. How and when a gene is expressed appears to correlate with the recombination rate [39,40,46] and the expression parameters of a gene are one of the best predictors of its evolutionary rate [47]. Furthermore, GC content correlates with the recombination rate, and GC content across regions of the genome is not always at equilibrium. In particular, GC-rich regions in the human genome and many other mammalian genomes are decaying, which may result in increased mutation rates in these regions [48,49]. The density of genes, which also correlates with both GC content and the recombination rate, may also play a role in the intensity of HRI in a particular genomic region, owing to the density of sites under selection. Direct effects of recombination can also affect \( d_N/d_S \). For example, increased \( d_N/d_S \) can occur in regions of high recombination owing to fixation of weakly deleterious nonsynonymous substitutions by gGC (see below; [50,51]).

Although there is good evidence that the rate of allele shuffling by recombination is adaptive and can be modified by selection in various eukaryotes [10–14], we find no compelling evidence that the recombination rate modulates the efficiency of selection across the genome in humans, *Drosophila* or yeast (save for in some non-recombining domains). This is consistent with theory suggesting that only a small amount of crossing over is necessary to achieve most of the benefits associated with recombination [52]. If so, it is unlikely that selection to minimise HRI is a strong force determining intragenomic variation in recombination rate.

**Molecular control of intragenomic variation in the recombination rate**

To understand the forces constraining the distribution of recombination and its effects on genome evolution, we need to understand the molecular mechanisms controlling it. On the large scale, it is known that COs are required for the accurate disjunction of chromosomes, placing a minimum requirement of one CO per chromosome arm [53] (or possibly chromosome [54]). This physical constraint ensures a minimum level of crossing over across the autosomes, although some other parts of the genome have intrinsically

---

104

*Trends in Genetics* March 2012, Vol. 28, No. 3
high CO rates (e.g. pseudoautosomal regions), whereas it is absent in others (e.g. the mitochondrial genome and mammalian Y chromosome).

Several factors have been shown to correlate with recombination on the large scale. Levels of germline DNA methylation appear to correlate with CO rates in human [55], which may suggest that it has a role in recombination. In addition, imprinted genes (and genes with monoallelic expression) exhibit high recombination rates [46], whereas genes transcribed in the germ line exhibit reduced rates of recombination [46,56] (but see [57]). Consequently, genes expressed in many tissues (i.e. housekeeping genes) tend to have low recombination rates in humans [46] and possibly also in Drosophila [58]. One explanation for these findings is that germline transcription somehow inhibits CO formation, which places a constraint on their location.

On the fine scale, recombination events in many eukaryotic genomes are largely confined to short (1–2 kb) hotspots [59,60]. A common mechanism of initiation of recombination in hotspots is the binding of the histone methyltransferase PR domain containing 9 (PRDM9) to a specific DNA sequence motif. This results in a histone methylation followed by the formation of a DSB. The precise locations of hotspots evolve rapidly in many species [22,61,62]. In humans, comparative genomic and sperm typing studies have both demonstrated the action of a form of biased genome conversion that we term ‘hotspot drive’ (Box 1, Figure 1), which promotes the spread of non-recombinogenic alleles at hotspots and, thus, their extinction [19,63]. The DNA-binding zinc finger domain of PRDM9 also appears to have been evolving rapidly in a wide range of animal genomes, which leads to regular shifts in the motif that it recognises [19,64]. Together, these processes suggest that a rapidly shifting recombination landscape on the local scale is a feature of many genomes containing an active PRDM9.

In some other species, PRDM9 evolves more slowly, which predicts a slower turnover of hotspot locations, whereas other species lack an active copy of PRDM9, and in these species hotspots are predicted either to be absent or controlled by another mechanism [65]. In dog, one of the only mammals missing an active PRDM9, recombination hotspots appear to be more evolutionarily stable [66]. On the fine scale, therefore, the genomic landscape of recombination and its rate of evolution both show large variability between species, but it is unclear whether these observations have any adaptive significance.

In humans, hotspots are biased against occurring in genes and a possible explanation for this is that hotspot locations are also subject to selective constraint to mediate deleterious effects, such as NAHR or other direct effects [60]. However, the factors that constrain recombination events on the fine scale are generally unknown. Nevertheless, it is clear that the recombination rate is governed by several physical and molecular constraints on both the large and local scale, which are unrelated to its evolutionary advantages.

Direct effects of recombination on evolutionary inference

One way in which consideration of the direct effects of recombination is important is for the inference of natural selection acting on genomic sequences. A popular approach to identify genes under positive selection is to scan the genome for acceleration in $d_N/d_S$ on a certain lineage (e.g. [67]). However, analyses of genome scans for genes with elevated $d_N/d_S$ in the human and other primate branches reveal that many such genes show GC-biased patterns of evolution on the branch on which they were identified, characteristic of the action of gBGC [50,51,68]. Mathematical modelling suggests that gBGC interferes with purifying selection, and can even cause $d_N/d_S$ to increase to above one under certain realistic parameter combinations, providing a false signal of positive selection [50,51]. Up to 20% of genes with elevated $d_N/d_S$ in humans may be better explained by gBGC than by positive selection [68].

An example of such a gene is the human adenylyl cyclase activating polypeptide 1 (ADCYAP1) gene, which was identified in a genome scan as having significantly elevated $d_N/d_S$ on the human lineage [67], and additionally exhibits all the hallmarks of gBGC [68,69]. It is found close to a telomere in a region of highly elevated male recombination, overlapping a recombination hotspot. Its two most diverged exons have accumulated 20 substitutions, which are all AT → GC, and this pattern extends into noncoding regions (Figure 2). Of the 20 substitutions, 17 change an amino acid, and it has been suggested that the high rate of evolution at nonsynonymous sites could not have been produced by gBGC alone [69]. However, mathematical modelling of the predicted effect of gBGC on this gene under a range of plausible scenarios, incorporating the population size-scaled coefficients of selection and gBGC and the distribution of negative fitness effects, suggests that a model of gBGC alone is sufficient to explain the acceleration in $d_N/d_S$ [68].

Careful evaluation of patterns of molecular evolution in genes with elevated $d_N/d_S$ ratios is warranted before concluding that positive selection is responsible. For instance, the brain-expressed microcephalin 1 (MCPH1) and abnormal spindle-like microcephaly-associated protein (ASPM) genes, both suggested to be under selection on the lineages leading to humans for increased cognitive function [70,71], also overlap human recombination hotspots and exhibit GC-biased patterns of evolution on the human lineage. Microcephalin has 20 AT → GC and seven GC → AT changes, whereas ASPM has ten AT → GC and four GC → AT, which are both different from the genome average, where GC → AT changes predominate (data from [67]). This may indicate that, contra to the prior assumption [70,71], gBGC is a more probable driver of elevated $d_N/d_S$ on the human lineage in these genes than is positive selection.

Direct effects of recombination on fitness

The evolutionary trajectory of a genetic variant depends not only on its phenotypic effect, but also on biases affecting its probability of transmission into the next generation. Because most new mutations with an effect on fitness are expected to be deleterious, by increasing the probability of fixation of GC alleles regardless of their fitness effects, the process of gBGC may incur a substantial ‘fixation load’ [18]. In regions of extremely elevated recombination, gBGC may generate bursts of GC allele fixation, which in species
comparisons appear as clusters of nucleotide substitutions [21]. Clusters of substitutions on the human lineage tend to be GC-biased and occur near telomeres and in regions of elevated recombination, with a stronger association with recombination in males [72,73]. It has subsequently been shown that GC bias in highly diverged sequences is a common feature in metazoans [74]. It is thought that these clusters are associated with recombination hotspots (which need not be currently active), which generate GC-biased fixations due to gBGC.

Several studies have identified human accelerated regions (HARs), which are short elements with a high degree of sequence conservation in vertebrates but that have experienced accelerated evolutionary rates in the human lineage (e.g. [75,76]). Although such a pattern of evolution might be expected owing to the action of positive selection, many of these elements are found in regions of elevated male recombination, and have an excess of GC-biased substitutions on the human lineage, which also extends into the flanking sequence, which is more indicative of gBGC. Hence, the action of gBGC in some HARs may have influenced the emergence of human-specific characteristics. Although elevated substitution rates in most HARs may be because of positive selection, the patterns of evolution in the two most highly accelerated regions identified so far, HAR1 and HAR2 (also called HACNS1), are strongly indicative of the action of gBGC [21].

Functional analysis of the human and chimpanzee versions of HACNS1 show that it acts as an enhancer of gene expression. A study using a mouse construct showed that the human-specific nucleotide substitutions result in increased expression in the developing limb [75]. On the basis of this finding, it was hypothesised that these substitutions resulted in a gain of function in this tissue, potentially contributing to the evolution of a human-specific characteristic. It has been argued that gBGC is unlikely to result in such a gain of function, because it is generally expected to result in the fixation of neutral or deleterious changes [69]. However, a recent analysis has provided evidence that the human-specific substitutions actually disrupt the function of a repressor module [77], which is potentially more consistent with the fixation of deleterious mutations by gBGC.

It has also been shown that gBGC can result in an elevated rate of fixation of weakly deleterious amino acid changes [50,51]. This can also have effects on the genome scale. A significantly elevated $d_N/d_S$ ratio has been observed in outcrossing wheat species compared with selfers [78]. This contrasts with classical predictions, which suggest that the efficacy of selection should be greater in outcrossing species owing to a reduction in HRI, which should lead to more efficient purifying selection and, hence, reduced $d_N/d_S$. Paradoxically, therefore, outcrossing species may suffer an additional load of deleterious fixations because gBGC represents an additional cost of recombination [79]. There is also evidence that gBGC contributes to the spread of deleterious mutations in human populations, because the site frequency spectrum of nonsynonymous mutations and those implicated in human genetic disease also show the hallmarks of a transmission bias towards GC [80].

Indel drive may yet prove to be equally important for genome evolution, although so far there is little evidence of a general transmission bias toward insertions or deletions (Box 1). If shorter indels have a transmission advantage, for example, this might explain the observation [81] that short introns are observed in domains of high recombination, rendering it unnecessary to evolve (again) differences
in the strength of selection, the conventional model for interpreting this finding. Similarly, a bias favouring the transmission of insertions over deletions could contribute to the observation that deletions segregate at lower frequencies in humans and are enriched in fixed differences since the human–chimpanzee ancestor compared with human polymorphism [82].

**Is recombination mutagenic?**

Besides its role in generating allelic transmission biases, recombination may also lead to new mutations. Although it is clear that recombination generates structural changes via NAHR, early suggestions that it also leads to point mutations [20] have not been conclusively proven. There is strong evidence in yeast that mitotic DSB repair is mutagenic [83,84], but the repair of DSBs formed during meiosis is mechanistically different and it is unclear whether this is also mutagenic. The extent to which these findings can be applied to other species is also not clear.

A correlation between genetic variation and the CO rate exists in several species [23,25,30,31,85], which has been suggested to reflect a larger effect of linked selection (selective sweeps or background selection) reducing variation in regions of low recombination. However, at least in humans, there is also a correlation between putatively neutral divergence and the recombination rate [85,86], which suggests that variation in the mutation rate could account for the correlation between recombination and genetic variation on a large scale. One interpretation of this correlation is that recombination is mutagenic. However, because human recombination hotspots are very fast evolving, any mutagenic effect of recombination need not necessarily result in higher divergence [19]. There is also no significant increase in the levels of human genetic variation around the recombinogenic hotspot motif in data from the 1000 Genomes Project [87], which would be expected if recombination were mutagenic.

It is possible that variation in the fixation of different nucleotides owing to the process of gBGC may generate a correlation between CO rate and divergence [73]. However, in the highly recombining human pseudoautosomal region, there is a significant elevation in substitution rate including A → T and G → C substitutions [88] and these should not be affected by gBGC. Another possibility is that linked selection has affected levels of divergence between human and chimpanzee [24]. Forward population genetic simulations suggest that both positive and negative selection in the ancestral population of humans and chimpanzees could generate a correlation between CO rate and divergence [26].

There is a lack of robust evidence of an association between divergence and CO rate in *Drosophila* [89] or yeast [90]. However, in *Drosophila*, this may be because of the low resolution of genetic maps, and there is evidence of a correlation when the CO rate is measured on a fine scale [91]. Furthermore, there is little power to detect a signal of mutagenic recombination in yeast [90] because modern yeast recombinates so very rarely that any faint signal would easily be drowned [92]. Moreover, whether *Drosophila* or yeast has a class of sites that are consistently neutraly evolving is unclear [25].

A mutagenic effect of meiotic recombination would have significance for a wide range of evolutionary and medical genetic analyses. However, although some of the observations presented above are compatible with such an effect, direct evidence is lacking.

**Concluding remarks and future perspectives**

Although it is commonly assumed that variation in the recombination rate across the genome leads to variation in the efficacy of selection owing to its indirect effects of shuffling alleles, there is no strong evidence of this, with the possible exception of non-recombining domains. A lack of such variation in the efficacy of selection would imply that variation in the local recombination rate is unlikely to be subject to strong selection. However, the direct effects of recombination appear to have strong demonstrable effects on the genome that are important to consider for correct evolutionary inference, and also have effects on the distribution of recombination hotspots and the evolution of functional genomic elements.

Why is the theory on the efficacy of selection not of greater relevance? One possibility is that positive selection has typically been assumed to result from recurrent ‘hard’ selective sweeps in which new beneficial mutations are rapidly driven to fixation by positive selection. However, selection can also entail more gradual or partial fixation of genetic variants that may be only weakly beneficial (a ‘soft sweep’ [93,94]). In both human [95] and *Drosophila* [96,97] recent evidence suggests that this mode of adaptation is common. Furthermore, much of the effect of selection on linked variation may be because of the elimination of rare deleterious mutations (background selection). Hard sweeps have a major effect on linked loci, potentially eliminating all genetic variation, making crossing over essential to preserve beneficial mutations and prevent fixation of deleterious ones in genetic linkage with those under selection. When variants do not rapidly sweep through the population (as in soft sweeps or background selection), it is less important to unlink them from neighbouring loci to prevent interference. This could potentially explain the lack of evidence for the efficacy of selection being limited by recombination in human and *Drosophila*, except in regions of extremely low recombination.

The interpretation of patterns of genome evolution is hampered by a poor understanding of the direct effects of recombination. We suggest that a thorough mechanistic analysis of the direct impact of recombination is necessary. What is the strength of gBGC? Why does it correlate with male, but not female, recombination [98]? Is meiotic recombination mutagenic and, if so, at what rate? These questions require experimental quantification using techniques such as sperm typing [99], experimental mapping of COs [100] and analysis of transmission in pedigrees [101]. High-throughput sequencing technologies will be useful for these goals. There is also a need for a better understanding of the determinants of the local recombination rate. Why do germline expressed genes appear to avoid recombination in mammals? Is this a necessary mechanistic constraint or the result of selection? What effect does DNA methylation have on promoting recombination? The use of more detailed post-genomic data sets will be useful to answer these
questions. To understand the evolutionary importance of recombination owing to its indirect effects, there is a need to clarify the direct effects of recombination on genome evolution.

Acknowledgements
We thank Jonas Berglund for help with Figure 2, and three reviewers for helpful comments.

References
27 Gossmann, T.I. et al. (2011) Quantifying the variation in the effective population size within a genome. Genetcs DOI: 10.1534/genetics.111.132654
35 Haddrill, P.R. et al. (2007) Reduced efficacy of selection in regions of the Drosophila genome that lack crossing-over. Genome Biol. 8, R18
49 Smith, N.G. et al. (2002) Deterministic mutation rate variation in the human genome. Genome Res. 12, 1350–1356
57 Weber, C.C. and Hurst, L.D. (2010) Intrinsic AT skew is a defendable proxy for germline transcription but does not predict crossing-over or
60 Myers, S. et al. (2005) A fine-scale map of recombination rates and hotspots across the human genome. Science 310, 321–324
66 Axelsson, E. et al. (2011) Death of PRDM9 coincides with stabilization of the recombination landscape in the dog genome. Genome Res. DOI: 10.1101/gr.124123.111
69 Prabhakar, S. et al. (2009) Response to comment on ‘Human-specific gain of function in a developmental enhancer’. Science 323, 714
95 Hernandez, R.D. et al. (2011) Classic selective sweeps were rare in recent human evolution. Science 331, 920–924
97 Karasov, T. et al. (2010) Evidence that adaptation in Drosophila is not limited by mutation at single sites. PLoS Genet. 6, e1000924
98 Pink, C.J. and Hurst, L.D. (2011) Late replicating domains are highly recombining in females but have low male recombination rates: implications for isochore evolution. PLoS ONE 6, e24480
105 Lamb, B.C. (1986) Gene conversion disparity: factors influencing its direction and extent, with tests of assumptions and predictions in its evolutionary effects. Genetics 114, 611–632