Selective Interference and the Evolution of Sex

Sarah P. Otto

From the Department of Zoology & Biodiversity Research Centre, University of British Columbia, 6270 University Boulevard, Vancouver, V6T 1Z4 Canada.

Address correspondence to S. P. Otto at the address above, or e-mail: otto@zoology.ubc.ca.

Received February 3, 2020; First decision April 22, 2020; Accepted July 27, 2020.

Sarah (Sally) Otto is a Killam Professor at the University of British Columbia. She is known for her theoretical studies investigating how biological systems evolve, using models to investigate the selective forces acting on genetic systems (recombination, ploidy level, and gene duplications) and mating systems (sexual vs. asexual reproduction, sexual selection, floral reproductive strategies). Complementing this approach, Dr. Otto’s group tracks yeast as they evolve to test evolutionary theories. Dr. Otto helped found the Canadian Society of Ecology and Evolution and has served as President of the Society for the Study of Evolution, and Council member of the American Genetic Association, among other Councils and Editorial Boards. Awards include a MacArthur Fellowship, a Guggenheim Fellowship, the Sewall Wright Award, and membership in the National Academy of Science, the Royal Society of Canada, and the American Academy of Arts and Sciences.

Abstract

Selection acts upon genes linked together on chromosomes. This physical connection reduces the efficiency by which selection can act because, in the absence of sex, alleles must rise and fall together in frequency with the genome in which they are found. This selective interference underlies such phenomena as clonal interference and Muller’s Ratchet and is broadly termed Hill-Robertson interference. In this review, I examine the potential for selective interference to account for the evolution and maintenance of sex, discussing the positive and negative evidence from both theoretical and empirical studies, and highlight the gaps that remain.

Keywords: Hill-Robertson effects, Muller’s ratchet, clonal interference, recombination, linkage disequilibrium, modifier, evolution of sex

Selection acts not on individual alleles but on the combinations of them present within the genomes of a species. The dynamics of an allele—and its very fate—is thus coupled to the genetic background(s) in which it is found. While recombination and the independent segregation of chromosomes mix up those backgrounds through sexual reproduction, associations among loci do not decay instantaneously.
and are rebuilt by the joint processes of mutation, drift, selection, and migration. Relative to the rate of evolution that could occur if selection acted upon each locus independently, evolution acting upon genomic combinations of alleles is often hampered, severely so when there is substantial genetic variation in fitness and little recombination (Barton 1995b; Weissman and Barton 2012; Good et al. 2014). This selective interference limits the pace of evolutionary change and is thought to be a major factor favoring the evolution of sex and recombination (see reviews by Otto and Lenormand 2002; Hartfield and Keightley 2012). In this perspective, I review the concept of selective interference and discuss what we know and what remains unknown about the role that selective interference plays in shaping evolution.

Selective Interference

The foundations of population genetics theory were constructed using models that describe the dynamics of single genes, independently of their genomic context. While this approach of “beanbag genetics” was lambasted by Mayr (1963), Haldane (1964) provided a strong defense, describing how the approach had provided insights into the pace of evolutionary change, the rate of mutation, and the strength of selection in species as diverse as the moth Biston betularia to humans. Recognizing the many simplifying assumptions made in early population genetics theory, Haldane argued that such models nevertheless provided a “kind of scaffolding within which a reasonably secure theory expressible in words may be built up…without such a scaffolding verbal arguments are insecure.” One of the key assumptions typically made in the early decades of population genetics theory was that allele frequency changes could be described at any one locus without accounting for genetic changes at other loci in the genome (Ewens 2008). With the explosion of genetic information during the last half of the 20th century—culminating in the 2000 announcement of the first draft sequence of the human genome (White House Press release 2000), population genetics theory rose to the challenge, adding to that early scaffolding by grappling with the interactions among sites within a genome.

One of the important early building blocks added to the scaffold was the concept of linkage disequilibrium. Early models assumed that alleles were independently inherited, implicitly assuming no linkage disequilibrium between loci even in models of 2 loci where such associations naturally arise. For example, in exploring the evolution of dominance, Fisher (1928) and Wright (1929) followed the ps and qs at 2 loci—one a target of selection and a second that modified the dominance properties of the target—rather than tracking the inheritance of the various chromosomal combinations, which is needed to account for the linkage disequilibrium between the loci (Ewens 1965).

Mathematically, linkage disequilibrium (D) quantifies the covariance in the state of the alleles at different sites. For 2 loci, D was not defined until 1960 by Lewontin and Kojima (1960), who examined the impact of associations among loci on evolutionary dynamics. In brief, Lewontin and Kojima defined linkage disequilibrium based on the difference between the frequency of a haplotype and its expected frequency, given the component allele frequencies. For example, at loci A and B with alleles A/a and B/b:

\[
\begin{align*}
freq(AB) &= freq(A) \times freq(B) + D \\
freq(Ab) &= freq(A) \times freq(b) - D \\
freq(aB) &= freq(a) \times freq(B) - D \\
freq(ab) &= freq(a) \times freq(b) + D
\end{align*}
\]

(1)

where D can also be calculated from

\[D = freq(AB) \times freq(ab) - freq(Ab) \times freq(aB)\]

Whenever there is linkage disequilibrium, the states of the 2 loci are correlated, such that an individual carrying A is more (positive D) or less (negative D) likely to carry allele B than expected. For clarity, I will define the alleles such that both A and B are the fittest, so that positive values of D imply that the favorable alleles are more commonly found together than expected by chance (a coupling association), while negative values of D imply that a favorable allele at one locus tends to be found with a deleterious allele at the other locus (an interfering association). In Box 1, I highlight some key results about linkage disequilibrium.

Accounting for linkage disequilibrium fundamentally changes how we think about and model evolution. With linkage disequilibrium, we can understand how selection at one locus impacts surrounding loci, because the increase in a selectively favored allele carries along with it any correlated alleles (Maynard Smith and Haigh 1974; Hermisson and Pennings 2005). With linkage disequilibrium, we can understand why evolution can lead fitness to decline, whenever recombination breaks apart favorable combinations more than selection increases their frequency (Moran 1964; Karlin and Carmelli 1975). With linkage disequilibrium, we can understand how genes that modify the rate of mutation, recombination, or sex would themselves evolve within a population, because modifier alleles hitchhike along with the allele combinations that they generate (see review by Otto 2013).

The initial models that incorporated linkage disequilibrium focused on how selection acts on 2 loci, using deterministic models of the type described in Box 1. While these models made it clear that the evolutionary response to selection at any locus typically depends on selection acting on all other loci (see Point 6 in Box 1), Maynard Smith (1968) pointed out that a gene would evolve independently of its genetic background if epistasis and genetic associations were absent (i.e., with multiplicative fitness interactions and D initially zero). Kimura (1965) broadened the scope of this result, showing that genetic associations rapidly reach a balance, which he called “quasi linkage equilibrium” (QLE), between epistasis building and recombination breaking apart associations (assuming weak epistasis and frequent recombination). Furthermore, when selection is weak and the system has reached QLE, the linkage disequilibrium between 2 loci has a negligible impact on the response to selection of each locus (see also Otto and Day 2007, section 9.3.2). Such deterministic models suggest that, as long as genes are loosely linked and interact weakly, evolution can be regarded as acting nearly independently on each locus.

Early in the evolutionary literature, several authors considered the evolutionary challenges faced by asexuals at the opposite extreme, without any sex and recombination. Fisher (1930) and Muller (1932) both noted that beneficial mutations that appear in different individuals cannot be combined within the same genome in the absence of sex, limiting the response of the population to selection, a phenomenon referred to as “clonal interference” (Gerrish and Lenski 1998). The Fisher–Muller hypothesis for sex highlights the advantage to sex that arises by bringing together beneficial alleles from different individuals into the same offspring. Muller (1964) and Crow and Kimura (1965) quantified the Fisher–Muller hypothesis by assuming each mutation appears in a single copy and, without recombination, can only fix if it arises in an individual carrying the previous mutation destined to fix. Their analyses suggested that sexual populations might incorporate beneficial mutations at rates that are orders of magnitude higher than asexual populations, especially in large populations. Asexuals also face a related problem of being
unable to bring together beneficial alleles if the best combination happens to be lost by chance (Muller 1964). With deleterious mutations occurring at high rates, few individuals carry the best combination, and the repeated loss of this best class by chance causes a “ratchet” in the absence of recombination. As a result, the fate of an allele in asexuals may often depend less on its own selective advantage/disadvantage and more on its genomic background (reviewed by Neher 2013), which can “greatly handicap a species” without recombination (Muller 1964).

There is, however, a large gray zone between genomes that are completely tied together within separate clonal lineages and fully randomized genomes lacking any genetic associations, a gray zone in which the vast majority of species exist. Within this gray zone, the key question is how strongly does selection at surrounding loci interfere with selection at a focal site? Hill and Robertson (1966) explored this question using finite population simulations of 2 loci with little or no epistasis. Deterministic models of infinite populations predict that the 2 loci should evolve independently (Point 3, Box 1), but the simulations revealed that selection at one locus reduced the fixation probability and slowed the speed of a beneficial allele at a second locus. That is, selection at a locus limited the response to selection at other loci, with the strength of interference declining as the rate of recombination \((r)\) increased between the loci.

Mathematically, Hill and Robertson interpreted their result in a couple of ways. First, selection at one locus reduces the effective population size at other loci to those individuals that either now or soon will (through recombination) carry the high fitness allele at that locus, reducing the efficacy of selection. Second, they noted that random genetic drift, along with selection, tends to generate negative genetic associations on average. To see this, consider the recursions in Box 1 with no epistasis (multiplicative selection). Disequilibrium then changes from generation to generation, according to:

\[
D' = \frac{D \{(1-r)w_Aw_Bw_Bw_B\}}{((w_Aw_B) + D \{(1-r)(w_A - w_B)(w_B - w_B)\})^2} \tag{2}
\]

where the denominator is the squared mean fitness, and the terms in curly braces are all positive by the convention that I have used to define the alleles \((A\) and \(B\) being fitter). Genetic drift causes stochastic variation in the amount of disequilibrium, but the effect is nonlinear because \(D\) influences both the numerator and denominator of \((2)\). The mean fitness is higher (larger denominator) when positive disequilibrium develops, by chance, causing positive disequilibrium to dissipate faster than negative disequilibrium. As a result, the average level of disequilibrium becomes negative over time (Figure 1A).

Essentially, positive associations among beneficial alleles increase the genetic variation and cause evolution to speed up, eliminating the associations faster, while negative associations impede evolutionary change. Thus, over time with drift and selection, genetic variation becomes increasingly hidden (negative \(D\)), with extant chromosomes bearing a mixture of beneficial and deleterious mutations, impeding the response to selection.

While the focus here has been on genetic associations between loci (measured by linkage disequilibrium), similar results hold for genetic associations within a locus (departures from Hardy–Weinberg), as measured by the intra-locus association measure, \(D_{AA} = \text{freq}(AA) - p_A^2\). While intra-locus associations dissipate immediately in a fully sexual population that randomly mates, drift in partially sexual populations leaves, on average, negative associations, with diploids more likely to be heterozygous for a good and bad allele than expected by chance (Figure 1B).

The slower response to selection found in models that account for chance events, either due to random genetic drift or due to the random appearance of mutations on specific genetic backgrounds, was coined the “Hill–Robertson effect” by Felsenstein (1974). As a synonym for the “Hill–Robertson effect,” I will use the term “selective interference,” which I have found to be more easily understood and remembered. This umbrella concept covers a wide variety of phenomena in the evolutionary literature, including the Fisher–Muller hypothesis, Muller’s ratchet, and clonal interference. It also includes the “ruby-in-the-rubbish” (Peck 1994) and “evolutionary traction” (Hadany and Feldman 2005) theories, which focus more on selective interference between beneficial mutations (the ruby) when surrounding sites are subject to deleterious mutations (the rubbish). Selective interference also reduces mean fitness in models of stabilizing selection due to the combined action of drift and selection (Wagner and Gabriel 1990), as well as in gene network models that allow the genomic architecture to evolve (Whitlock et al. 2016). Selective interference can be so strong that Gillespie (2001) considered the possibility that it (“genetic draft”), rather than population size (“genetic drift”), may determine the level of stochasticiy experienced at a locus.

Importantly, selective interference arises whether the population is strictly clonal or sexual, although the strength of interference is greatest when the linkage is tighter and/or rates of sex lower. Theory over the past few decades has greatly improved our mathematical understanding of selective interference, with analytical descriptions of how the probability of fixation is affected by selection at linked sites (Barton 1995b) and how the combination of selection and drift generates negative disequilibrium (Barton and Otto 2003) and negative intra-locus associations (Roze and Michod 2010), leaving a signature where beneficial alleles are disproportionately found on less fit genetic backgrounds.

A counterintuitive aspect of selective interference is how it depends on the population size. Given that interference depends on random drift to generate genetic associations, one might think that selective interference would only be relevant in populations of small size. Both analytical and simulation studies have, however, found increasing levels of interference in larger populations (Otto and Barton 2001; Gillespie 2001; Keightley and Otto 2006; Weissman and Barton 2012; Good et al. 2014). This counterintuitive result occurs because loci remain polymorphic for longer in larger populations, increasing the density of selected alleles segregating per megabase and increasing interference among selected sites. As the input rate of mutations increases (i.e., the product of population size and mutation rate), the rate of adaptation in sexual populations undergoes a phase shift, rising linearly with the input rate at first (nearly independent adaptation at different loci) but then transitioning to an interference regime, where the rate of adaptation depends mostly on the rate of recombination, rather than the population size or mutation rate (Weissman and Barton 2012). As a rule of thumb, Weissman and Barton conclude that selective interference prevents the rate of adaptive substitution from rising much above one per Morgan every 2 generations. The increasing strength of selective interference with increasing population size is also seen with deleterious mutations (Santiago and Caballero 1998; Gillespie 2001; Weissman and Barton 2012; Good et al. 2014).

Ultimately, one would expect selective interference to disappear as the population size rises to infinity (approaching the
model of Maynard Smith 1968), but this raises a subtle issue. With a finite number of sites under selection, the strength of selective interference does rise and then fall as the population size increases (Otto and Barton 2001), but the peak effect occurs at larger populations with more loci or with weaker selection per locus (Iles et al. 2003). Indeed, with many loci, the peak is not reached over the range of population sizes simulated (up to 100 000 in Iles et al. 2003 for beneficial alleles and Keightley and Otto 2006 for deleterious mutations). Thus, if the density of polymorphic sites under selection continues to increase as the population size rises, it is unclear what would happen in the limit of infinite population size. Furthermore, with spatial structure, it has been shown that selective interference can remain a potent force driving genetic associations in very large populations, even infinitely large, because of drift acting within local populations (Martin et al. 2006).

Evidence for Selective Interference

A number of experimental evolution studies have found that asexual lines do indeed adapt at a slower rate than recombining lines in a variety of organisms (e.g., *Chlamydomonas*: Colegrave 2002; Kaltz and Bell 2002; Lachapelle and Bell 2012, yeast: Goddard et al. 2005; Gray and Goddard 2012; McDonald et al. 2016, *E. coli*: Cooper 2007, q6 RNA virus: Poon and Chao 2004). While consistent with selective interference, whereby adaptive alleles in different clones interfere with one another, these studies do not prove that the advantage to sex comes from relieving selective interference. It could alternatively be that fitness interactions limit adaptation in asexual lines (i.e., epistasis is predominantly negative, generating negative genetic associations that impede selection; Eshel and Feldman 1970). Or there may be pleiotropic benefits to sex (arising from, e.g., different gene expression profiles, life histories, behaviours, etc.). For example, meiosis in yeast reverses the accumulation of age-induced damage by forming an extracellular compartment in which abnormalities are eliminated (extrachromosomal ribosomal DNA circles, abnormal nucleolar material, and protein aggregates with oxidatively damaged proteins) (Unal et al. 2011; King et al. 2019). To reduce the influence of such pleiotropic differences, Goddard et al. (2005) introduced genetic mutations (deleting both *SPO11* and *SPO13*) so that asexual yeast could also be passaged through sporulation and mating, but without recombination or independent segregation. Because all lines went through similar stages and environmental conditions, the study by Goddard et al. (2005) is particularly compelling evidence that sexuals can adapt faster, although it is unknown whether the experimental design equalized other aspects of sex, such as the elimination of accumulated damage.

Additional clues about selective interference can be obtained by deep genomic sequencing during the course of experimental evolution. Such studies have found extensive genetic hitchhiking in clonally adapting yeast and demonstrated that beneficial mutations have a lower probability of fixation in larger populations, where more clones compete, a hallmark of selective interference (Lang et al. 2013). Tracking the genetic trajectories of sexual and asexual lines, McDonald et al. (2016) found that deleterious mutations often rose in frequency and fixed within asexual lines but were eliminated within sexual lines following a round of sporulation and mating. As a consequence, cohorts of mutations, including a mixture of deleterious and beneficial mutations, often fixed in asexual populations, whereas beneficial mutations more often spread and fixed on their own in the sexual populations.

While extensively studied in the lab, a recent review of studies in nature comparing sexual and asexual lineages highlighted the “absence of field studies focused on testing whether an increased rate of adaptive evolution might help favor sexual over asexual organisms” (Neiman et al. 2018). While the relative rate of adaptive evolution has rarely been studied in the field, the authors identified several studies that have compared levels of deleterious mutations, finding that they accumulate more often in asexuals than sexuals (14 out of 19 studies surveyed). This evidence is

**Figure 1.** Selective interference results from drift with selection. One generation of drift in the production of offspring causes variability in the initial genetic associations (x axis), which when acted upon by selection causes genetic associations to become negative, on average (y axis). Degree of shading indicates the extent of negative associations, whereby “good” alleles are found disproportionately on “bad” genetic backgrounds. (A) Between-locus linkage disequilibrium, $D$, based on Equation 2. (B) Within-locus genetic associations, $D_{AA}$, illustrated for three different rates of sexual reproduction, $\sigma$ (see Supplementary File for analogous equation to (2)). Figure assumes that the initial associations created by drift are normally distributed around 0 with a standard deviation given by the x axis, that multiplicative selection favors beneficial alleles by 10%, and that the initial frequency of beneficial alleles is 0.1. In addition, drift by itself generates negative within-locus genetic associations, which impede selection, because chance oversampling of heterozygotes generates a strong negative $D_{AA}$, while chance oversampling of homozygotes has a weaker effect unless both homozygotes are oversampled (Roze and Michod 2010).
consistent with selective interference reducing the efficacy of purging in asexual populations, but other mechanisms could also account for this difference, including negative epistasis or pleiotropic differences. For example, mutators are expected to spread more readily in asexual populations because of the tighter linkage between alleles, increasing the mutation rate and adaptive mutations (Sniegowski et al. 2000), as has been found in experimental evolution studies in yeast (Raynes et al. 2012). Consequently, asexual lines might evolve a higher burden of deleterious mutations because mutators are more often favored.

Genomic scans for selection provide another source of evidence for selective interference. Several studies have found signals of selective interference in regions of low recombination, with fewer adaptive substitutions and higher frequencies/substitution rates of deleterious mutations (Drosophila: Mackay et al. 2012; Langley et al. 2012; Campos et al. 2014; birds: Gossmann et al. 2014; the fungal pathogen Zymoseptoria tritici: Grandaubert et al. 2019; influenza A: Strelkowa and Lässig 2012). Castellano et al. (2016) estimated that 27% of adaptive substitutions were “missing” due to selective interference across the genome of Drosophila melanogaster. Interference was most severe, as expected, in regions with both high gene density and high local mutation rates, where ~60% of adaptive substitutions were missing. Similarly, Uricchio et al. (2019) estimated that the rate of adaptive substitution is reduced by ~25% across the genomes of humans due to linkage.

**Selective Interference and the Evolution of Sex and Recombination**

When negative genetic associations have accumulated due to selective interference, sex and recombination can bring together favorable alleles from different genetic backgrounds into the same individual. Of course, deleterious alleles are also brought together (recombining bad alleles onto bad genetic backgrounds), but the increase in fitness variation generated by sex and recombination when negative associations predominate is one of the oldest explanations for sex, dating back to Weismann (1889).

Weismann did not, however, account for why variation would be missing in the first place (i.e., why genetic associations would be negative rather than absent or positive). It was Fisher (1930) and Muller (1932) who first provided a plausible scenario for why favorable combinations might be lacking: because beneficial mutations often arise in different individuals within a finite population and can only be brought together by sex and recombination. It is instructive to quote Fisher’s argument, made before linkage disequilibrium was even defined:

“The evolutionary progress of an asexual group thus presents the dilemma that it can only utilize all those beneficial mutations which occur, and survive the dangers of the initial period, if the rate of occurrence of mutations is so low that the population of competing organisms is normally in a state of genetic uniformity, and in such a state evolutionary progress will necessarily be slow. In such a state adequate variability will be present only by an extremely small mutation rate, both of beneficial and of deleterious mutations, are high enough to maintain any considerable genetic diversity, it will only be the best adapted genotypes which can become the ancestors of future generations, and the beneficial mutations which occur will have only the minutest chance of not appearing in types of organisms so inferior to some of their competitors, that their offspring will certainly be supplanted by those of the latter.”Fisher (1930), p. 122

As presented in this quote, the advantage to sex is a selective advantage to the species: by breaking down negative genetic associations, sex increases variability and the rate of evolution. Indeed, in his 1958 book revision, Fisher added a section arguing that species-level selection would generally be less effective than individual-level selection because the latter involves more events and acts more often (through individual births and deaths). But Fisher singled out sex as “the possible exception...which could be interpreted as evolved for the specific rather than for the individual advantage” (p. 49–50 in Fisher 1958).

Despite Fisher’s doubts, sex and recombination can evolve within a single population in response to individual-level selection and selective interference. Again, it is linkage disequilibrium that helps account for the evolution of sex at an individual level of selection. When negative genetic associations predominate among selected alleles, modifier genes that increase the frequency of sex or recombination give rise to more variable offspring, with a higher probability of having high fitness offspring (as well as low fitness offspring). As high fitness descendants are more likely to survive and reproduce, they pass along both the favorable gene combinations and the modifier alleles that generated them (Figure 2). Typically, selective interference most strongly favors increased sex and recombination at intermediate selection coefficients (e.g., Otto and Barton 1997; Iles et al. 2003; Roze 2014): very weakly selected loci have little impact while strongly selected loci are too transient to substantially interfere with one another.

The story is not so simple, however, because modifiers that increase the frequency of sex and recombination tend to switch genetic backgrounds more often, reducing the “ride” that they get while hitchhiking along with selectively favorable alleles. Furthermore, breaking apart genetic associations among or within loci can cause fitness to decline, creating a recombination load or segregation load among offspring, respectively (Crow 1970, see review by Otto and Lenormand 2002). The nature of this load depends on the form of fitness interactions. If selective interference has generated negative genetic associations, then breaking these down by recombination will cause an immediate reduction in the fitness of offspring if fitness surfaces are negatively curved (i.e., when good-plus-bad allele combinations are fitter than the average of good-good and bad-bad combinations, “negative epistasis”). On the other hand, if fitness surfaces are positively curved, then breaking apart negative genetic associations generated by selective interference gives an extra boost to modifiers that increase the amount of sex and recombination (because the good-good and bad-bad combinations are fitter on average with positive epistasis).

Mathematically, modifier models of selection within a single population have found that increased frequencies of sex (Roze and Michod 2010) and recombination (Otto and Barton 1997; Barton and Otto 2005) can evolve in the presence of drift and selection, especially when sex and recombination rates are not too high (so that modifiers stay linked for longer with beneficial alleles) and when fitness surfaces are weakly or positively curved (so that the intermediate load is mild or reversed; see Figure 2). Furthermore, selection of many forms, including fluctuating selection (Hodgson and Otto, 2005)
2012; da Silva and Galbraith 2017), purifying selection (Keightley and Otto 2006), and directional selection (Otto and Barton 1997; Barton and Otto 2005; Roze and Michod 2010), contribute to the amount of selective interference within the genome and to the negative genetic associations that tend to favor the evolution of sex and recombination.

Sex may thus have originally evolved—and may currently be maintained—to eliminate the selective interference that limits the ability of a population to adapt and to purge deleterious mutations. This explanation for sex and recombination is particularly appealing, given that all forms of selection potentially contribute to the strength of interference, it works on genetic associations both among and within loci (e.g., Barton and Otto 2005; Roze and Michod 2010, respectively), and it applies whether mutations are unique (as in the Fisher–Muller argument) or recurrent but subject to random genetic drift (as explored by Barton and Otto 2005).

The Challenges
While I would argue that reducing selective interference is the explanation for sex, there are a number of outstanding questions that remain:

- Is the density of selection across genomes high enough?
- Do other sources of genetic associations overwhelm selective interference?
- Are sex and recombination too frequent?

Addressing these issues will require a combination of empirical and theoretical work.

Empirically, we need to know more about the landscape of selection across the genome to determine if selective sites are densely distributed to generate substantial interference (i.e., so that species are often in the interference regime that selects for sex rather than in the regime whereby sites evolve independently). To answer this empirical question requires that we focus less on identifying specific sites under selection and more on estimating the density of selection across the genome (e.g., Tataru et al. 2017).

Theoretically, we also need models that integrate the various forms of selection to estimate the net strength of interference. [I am intentionally vague about how the strength of selective interference is assessed, as that depends on the question at hand and whether it is the effect on linkage disequilibrium, fixation probability, mean fitness, or modifier dynamics that is of interest.] Not all selection acts equally on selective interference, so models are needed to calculate the net interference given a variety of forms of selection and question of interest. In particular, rare alleles develop genetic associations that are tiny in absolute terms but large relative to their frequency. For deleterious alleles, these tiny associations remain tiny, and interference accumulates only when many such deleterious mutations occur. By contrast, if a previously rare allele is temporarily or permanently favored, these initially tiny associations grow into large associations as the allele frequency rises, contributing disproportionately to selective interference (Barton 1995b).

The fraction of beneficial substitutions, \( \alpha \), thus helps determine the likely strength of selective interference and can be estimated from genomic data on polymorphic and fixed differences (Smith and Eyre-Walker 2002). For example, Weissman and Barton estimated that beneficial substitutions occur in *Drosophila* at a rate of 0.001 substitutions per Morgan per generation (Weissman and Barton 2012), much lower than the rates at which selective interference strongly limits adaptation (~0.5). As they noted, however, averaging the rates of beneficial substitutions over time and over the genome fails to capture times and locations where interference is substantial. For example, Tataru et al. (2017) found that the majority of substitutions in chimpanzees were beneficial on the X chromosomes, but far fewer (<25%) on the autosomes. This clustering of mutations within the genome greatly increases selective interference among those beneficial mutations that do occur. Estimating substitution rates also fails to capture the interference that occurs among mutations that are only transiently beneficial and that rise in frequency and contribute to selective interference but do not ultimately substitute. Alleles involved in local adaptation, disease resistance, and other biotic interactions may often fall into this category. For example, acquired host immunity may reverse selection on pathogens over time and contribute to the rise and fall of transiently beneficial mutations (potentially contributing to the loss of beneficial mutations observed in *Drosophila*; Strelkowa and Lässig 2012). Currently, we do not know whether selective interference primarily limits adaptation on a population-wide scale or whether it mainly limits local adaptation across the variety of environments faced by a species. At this point, we simply do not have enough data to estimate the flux of such transient and/or locally beneficial mutations and the interference they may cause.

Another major unknown is whether the genetic associations among loci generated by other processes overwhelm those generated by drift and selection. In particular, fitness interactions between
(epistasis) and within (dominance) loci generate associations. That is, we cannot just assume that associations are built only by selective interference. If fitness interactions are strongly positive, good allele combinations are much fitter than expected based on the fitness effects of each allele. These combinations thus tend to accumulate over time, which can generate positive genetic associations (Point 4a in Box 1), even in the presence of drift and selection. Positive genetic associations speed up the response to selection and select against sex and recombination (top right region in Figure 2; Barton 1995a; Otto and Barton 2001; Roze and Michod 2010). Conversely, if fitness interactions are very strongly negative, good allele combinations are less fit than expected and so rise less rapidly, generating negative genetic associations (Point 4b in Box 1), akin to those built by selective interference. That said, breaking apart these negative genetic associations is a double-edged sword: it reveals hidden variation and speeds up the response to selection, but it also reduces the mean fitness of offspring because of the negative fitness curvature (recombination load and segregation load; bottom left region in Figure 2; Barton 1995a; Roze and Michod 2010). As an example, overdominance is a strongly negative form of dominance interaction, which causes heterozygotes to become overabundant and selects against sex and the segregation load it entails, both theoretically (in randomly mating populations, Dolgin and Otto 2003) and empirically (in diploid yeast, Leu et al. 2020). While the amount of epistasis does not appear to be predominantly positive or negative (Rice 2002; De Visser and Elena 2007), dominance of the wildtype allele over deleterious variants is common. Roze and Michod (2010) found that deleterious mutations limit the response to selection and favor the evolution of sex when mutations are slightly recessive to dominant, but highly recessive deleterious mutations select against sex due to the strong segregation load. This may be problematic for selective interference as an explanation for sex in diploids, given data suggesting that deleterious mutations often span the range from highly recessive to slightly recessive (Simmons and Crow 1977, Agrawal and Whitlock 2011, Manna et al. 2011, e.g., the latter estimate a weighted average $b = 0.27$ [95% confidence interval CI: 0.18–0.36] across studies), depending on the exact shape of the distribution of dominance and the extent to which interference is primarily among deleterious mutations.

Similarly, migration between patches generates genetic associations, and it is unclear if and when these associations dominate those generated by selective interference. If migrants create associations that are predominantly positive (as when locally adapted alleles are more often found together), then genetic associations might speed up the response to selection, rather than limit it, and would often select against sex and recombination (Lenormand and Otto 2000). On the other hand, migration can be an important source of genetic variance in fitness, boosting the number of selected loci that are polymorphic within a local population and increasing the potential for selective interference. Indeed, Martin et al. (2006) showed that migration, by replenishing polymorphism, increases the extent to which selective interference favors the evolution of recombination in spatially structured populations facing similar selective pressures, relative to the case of no migration.

Empirically, recent efforts have examined the evolution of sex within experimental evolution studies, tracking the frequency of sex within a species and pinpointing the sources of selection on sex. Increased rates of sex have evolved in Brachionus calyciflorus rotifers in a number of studies. In one study, the source of selection for sex was likely spatially heterogeneous selection given migration between 2 patches with different food sources (Becks and Agrawal 2010). In another, predator-prey interactions led to the evolution of higher rates of sex (Haake et al. 2016), but the mechanism is unknown and may or may not have involved selective interference. Finally, higher rates of sex evolved when the rotifers were placed in a new environment (Becks and Agrawal 2012). In this last study, crosses were performed to show that sex caused a fitness reduction (segregation load; in regions without red dashing in Figure 2) and an increase in fitness variation (blue dashing in Figure 2), but whether selective interference or epistasis or both were the primary drivers of sex is unknown.

Experiments like these on rotifers are exciting because we can watch the evolution of sex in action, tracking whether sex becomes more or less frequent over the course of the experiment and dissecting the reasons why. More such studies, coupled with genomic sequencing, such as performed by McDonald et al. (2016), promise to reveal the targets of selection and to determine whether beneficial alleles interfere with each other's spread and tend to drag along deleterious alleles and whether such interference was alleviated by the evolution of higher rates of sex.

Also needed are methods to assess the strength of selective interference. A variety of methods have been suggested, but these methods are often confounded by other processes. One can, for example, determine if negative genetic associations predominate, as expected under selective interference, by measuring if the variance in fitness is higher among recombinant offspring than clonal offspring (as conducted by Becks and Agrawal 2012). This signature could, however, reflect epistasis rather than interference (Figure 2). One can also look for evidence of clustered substitution events, especially in regions of lower recombination (Castellano et al. 2016) or among lineages with less sex (McDonald et al. 2016). Building upon this signal, Strelkowa and Lässig (2012) develop an extension of the McDonald-Kreitman test to look for evidence that alleles reach high frequency but are subsequently lost, as expected under selective interference. This signal could, however, also be generated by fluctuating selection or sign epistasis (i.e., when a previously beneficial allele becomes deleterious in the presence of subsequent mutations). Nevertheless, these are promising directions. Just as measures of selection have taken decades to develop and refine, measures of selective interference deserve development, alongside simulation testing to determine which measures of selective interference are most powerful and robust to confounding factors.

Finally, a major open question is whether selective interference can explain the high levels of sex observed in many species (Hadany and Comeron 2008), particularly in the face of the many costs of sex. Costs of sex have been incorporated into several theoretical studies, which find that sex is less favored and evolves to a lower level in the presence of its many potential costs (Keightley and Otto 2006; Roze and Michod 2010; Roze 2014), including the cost of dividing resources between male and female function, the cost of finding and securing a mate, and the costs associated with switching from mitotic to meiotic reproduction. Selective interference is strongest in fully clonal populations, readily explaining the evolution of some sex over no sex. Obligatory sex is, however, harder to explain. Potential resolutions include the possibility that organisms are often enough in a zone where the current genetic associations mismatch what is locally favored (reversing the segregation load, red zone in Figure 2), the difficulties of maintaining two modes of reproduction when the loss of one (sex) increases selective interference and the risk of extinction, and other features such as sexual selection that can favor high rates of sex (Hadany and Beker 2007).
Selective interference represents a major evolutionary challenge, reducing the rate of adaptation, the chance of local adaptation, and the efficacy by which deleterious mutations are purged, simply because alleles are selected, not individually, but within the genomic background in which they are found. Sex and recombination lift the cause alleles are selected, not individually, but within the genomic environment combining theoretical, experimental, and genomic analyses of selective interference over the past few decades, and future work combining theoretical, experimental, and genomic analyses promises to reveal whether selective interference is the primary explanation for sex.

Point 2: If there is no selection (all \( w_i = 1 \)), then linkage disequilibrium in the next generation, \( D' = x_{AB}' x_{ab}' - x_{Ab}' x_{aB}' \), can be shown to equal \((1 - r)D\), implying that genetic associations decline at a rate \( r \) per generation in the absence of any other process. Note, however, that even with random mating and unlinked loci \( (r = \frac{1}{2}) \), linkage disequilibrium does not disappear immediately but declines over time, \( t \), from its initial value according to \((1 - r)^tD\).

Point 3: If the loci are initially uncorrelated \((D = 0)\), then selection builds linkage disequilibrium and \( D' = p_{A}p_{A}p_{B}p_{B} (w_{AB}w_{ab} - w_{Ab}w_{aB}) / w^2 \). \( D \) will remain zero if there is multiplicative selection such that haploid fitness is the product of the fitness effects of each allele \((e.g., w_{ab} = w_{a}w_{b})\), because then \( w_{AB}w_{ab} - w_{Ab}w_{aB} = w_{A}w_{B}w_{i}w_{b} - w_{A}w_{B}w_{i}w_{b} = 0 \) (Maynard Smith 1968).

Point 4a: If selection causes the extreme genotypes \( AB \) and \( ab \) to be more fit than expected based on the fitness of \( Ab \) and \( aB \) combinations \((w_{AB}w_{ab} > w_{Ab}w_{aB})\), epistasis for fitness is defined to be positive; positive epistasis builds positive genetic associations over time \((D \) becomes positive; see Point 3), where \( AB \) and \( ab \) become more common than expected within the population (Eshel and Feldman 1970).

Point 4b: If selection causes the extreme genotypes \( AB \) and \( ab \) to be less fit than expected based on the fitness of \( Ab \) and \( aB \) combinations \((w_{AB}w_{ab} < w_{Ab}w_{aB})\), epistasis for fitness is negative; negative epistasis builds negative genetic associations over time \((D \) becomes negative; see Point 3), so that \( AB \) and \( ab \) are relatively uncommon (Eshel and Feldman 1970).

Point 5: Genetic associations and selection at other loci alter the evolutionary dynamics at each locus. For 2 loci, these effects can be captured by defining the marginal fitness of each allele averaged over the chromosomes in which it will be found. For example, with a life cycle where selection follows recombination, at locus \( A \), \( w_{A} = w_{AB}p_{A}p_{B} + (1 - r)D \) \( /w_{A} + w_{Ab}(p_{A}p_{B} - (1 - r)D) \) \( /w_{A} + w_{aB}(p_{A}p_{B} - (1 - r)D) \) \( /w_{A} \), where according to Equation 1 terms like \( p_{A}p_{B} + (1 - r)D \) \( /w_{A} \) define the frequency of offspring carrying allele \( A \) that also carry allele \( B \) (i.e., freq(\(AB\))\(freq(A) \) after one round of recombination). Selection then changes the frequency of allele \( A \) according to \( p'_{A} = p_{A}w_{A} / (p_{A}w_{A} + p_{a}w_{a}) \). Locus \( A \) will evolve independently of locus \( B \) only if linkage disequilibrium is absent and epistasis is multiplicative (the marginal fitness of allele \( A \) relative to allele \( a \), \( w_{A}/w_{a} \), then simplifies to \( w_{A}/w_{a} \)).

**Box 1**

As defined in Equation 1, linkage disequilibrium \((D)\) measures whether an allele at one locus tends to be found more often with a particular allele at another locus (a genetic association). By contrast, recombination rate \((r)\) is a measure of the distance between 2 loci and equals the probability that a gamete contains chromosomal combinations not found in the parent. These are coupled, as can be seen most easily in a deterministic model of haploid selection at 2 loci (ignoring drift). Assuming random mating followed by meiosis and then selection, the frequencies of each chromosome in the next generation are:

\[
x'_{AB} = \frac{x_{AB}^2 + x_{AB}x_{Ab} + x_{AB}x_{aB} + x_{AB}x_{ab}(1 - r) + x_{Ab}x_{aB}(r)}{w'_{AB}}
\]

\[
x'_{Ab} = \frac{x_{Ab} + rD}{w'_{Ab}}
\]

\[
x'_{aB} = \frac{x_{aB} + rD}{w'_{aB}}
\]

\[
x'_{ab} = \frac{x_{ab} - rD}{w'_{ab}}
\]

where \( x_{i} \) \( x'_{i} \) and \( w_{i} \) are the current frequency, the next generation frequency, and the fitness of haploid genotype \( i (AB, Ab, aB, ab) \), and \( w \) is the mean fitness calculated at the time of selection (the sum of the numerators). The simplification made on the first line (and implicit in the following lines) uses the fact that all of the potential partner gametes sum to one in frequency \((x_{AB} + x_{Ab} + x_{aB} + x_{ab} = 1)\).

Several key points about linkage disequilibrium can be gleaned from the above equations with a little algebra (Supplementary File).

**Point 1:** If there are no genetic associations in the current generation \((D = 0)\), then the rate of recombination does not affect the offspring frequencies. Recombination only breaks apart genetic associations that are present between the loci.

**Supplementary Material**

Supplementary material is available at *Journal of Heredity* online.

**Funding**

This work was supported by a Discovery grant from the Natural Sciences and Engineering Research Council of Canada (NSERC RGPIN-2016-03711).
Acknowledgments

I am grateful to the American Genetic Association for the invitation to give the Key Distinguished Lecture and to Maria Orive for organizing the “Sex & Asex: The Genetics of Complex Life Cycles” conference. I would also like to thank Alirio Rosales, Maria Orive, Bryn Wiley, and 2 anonymous reviewers for many helpful suggestions.

References