# Natural selection and the advantage of recombination*

Philip J Gerrish,[1,2,3] Benjamin Galeota-Sprung,[4] Fernando Cordero,[5] Paul Sniegowski,[4] Alexandre Colato,[6] Nicholas Hengartner,[7] Varun Vejalla,[8] Julien Chevallier,[9] and Bernard Ycart[9]

[1] *Department of Biology, University of New Mexico, Albuquerque, New Mexico, USA*†
[2] *Theoretical Biology & Biophysics, Los Alamos National Lab, Los Alamos, New Mexico, USA*
[3] *Instituto de Ciencias Biomédicas, Universidad Autónoma de Ciudad Juárez, México*
[4] *Department of Biology, University of Pennsylvania, Philadelphia, Pennsylvania, USA*
[5] *Biomathematics and Theoretical Bioinformatics, Technische Fakultät, Universität Bielefeld, Germany*
[6] *Departamento de Ciências da Natureza, Matemática e Educação, Univ Fed de São Carlos, Araras SP, Brazil*
[7] *Theoretical Biology & Biophysics, Los Alamos National Lab, Los Alamos, New Mexico, USA*
[8] *Thomas Jefferson High School for Science and Technology, Alexandria, Virginia, USA*
[9] *Mathematique Apliquée, Laboratoire Jean Kuntzmann, Université Grenoble Alpes, France*

(Dated: 5 June 2021)

The ubiquity of sex and recombination in nature is widely viewed as enigmatic, despite an abundance of limited-scope explanations. Natural selection, it seems, should amplify well-matched combinations of genes. Recombination would break up these well-matched combinations and should thus be suppressed. We show that this common intuition is flawed: rather, natural selection 1) amplifies poorly-matched gene combinations and 2) creates negative associations in the process. The genetic shuffling effected by recombination should thus be advantageous quite generally.

The ability to exchange genetic material through recombination (and sex) is a heritable trait [1, 2] that is influenced by many different evolutionary and ecological factors, both direct and indirect, both positive and negative. Evidence from nature clearly indicates that the net effect of these factors must be positive: recombination across all levels of organismal size and complexity is undeniably the rule rather than the exception [3–6]. Theoretical studies, on the other hand, have revealed a variety of different mechanisms and circumstances that can promote the evolution of recombination, but each one by itself is of limited scope [5–7]. These studies would thus predict that the absence of recombination is the rule and its presence an exception [8–13]. The sheer abundance of these exceptions, however, can be seen as amounting to a rule in its own right – a "pluralist" view that has been adopted by some authors to explain the ubiquity of recombination [3, 4, 14]. The necessity of this pluralist view, however, may be seen as pointing toward a fundamental shortcoming in existing theory: perhaps some very general factor that would favor recombination has been missing [3, 6, 7, 15].

Existing theories of the evolution and maintenance of sex and recombination can be divided into those that invoke *direct* vs *indirect* selection on recombination. Theories invoking direct selection propose that recombination evolved and is maintained by some physiological effect that mechanisms of recombination themselves have on survival or on replication efficiency [16–20]. Such theories might speak to the origins of sex and recombination but they falter when applied to their maintenance [21]. Most theories invoke indirect selection: they assume that any direct effect of recombination mechanisms is small compared to the trans-generational consequences of recombination.

While differing on the causal factors involved, the various theoretical approaches are unanimous in their identification of the fundamental selective environment required for sex and recombination to be indirectly favored: parents of recombinant offspring must, on average, harbor an excess of selectively mismatched combinations of alleles across loci and a deficit of selectively matched combinations. Recombination is favored under these conditions because on average it breaks up the mismatched combinations and assembles matched combinations.

Studies that invoke indirect selection can be divided into those that rest on optimality arguments (group selection) and those that study the evolutionary fate of modifiers of recombination. Competition among recombination-rate variants at a modifier locus under the above-mentioned selective conditions will tend to increase recombination rate [2, 22]. A modifier locus can be affected by short- and long-term indirect selection [23]. In the long term, modifiers that increase the recombination rate, or *up-modifiers*, can in theory be favored because of the fitness variation they augment [23]. In the short term, recombinants formed from randomly-chosen parents harboring the aforementioned imbalance have an advantage that is relatively immediate, thereby creating indirect selection for up-modifiers. Our analyses focus on: 1) the selective value of recombinants, thereby quantifying the short-term advantage of recombination, and 2) the selective value of recombination and the dynamics of a $rec^+$ modifier, which implicitly account for both short- and long-term effects. These two approaches may be thought of as modeling two extremes where the modifier is loosely (1) and tightly (2) linked to the fitness loci whose recombination rate is modified.

The great challenge in explaining the evolution of recombination has been to identify an evolutionary source of the aforementioned imbalance whose prevalence in nature is comparable to the prevalence of sex and recombination in nature. One feature of living things whose prevalence approximates that of sex and recombination is evolution by natural selection. In what follows, we assess the effects of natural selection by itself on selective imbalance and hence on the evolution of recombination.

We preface our developments with an essential technical point. In much of the relevant literature, the measure of selective mismatch across loci affecting the evolution of recombination is linkage disequilibrium (LD) [7, 23–28], which measures bias in allelic frequencies across loci but does not retain information about the selective value of those alleles. Here, the objectives of our study require a slight departure from tradition: our measure of selective mismatch will be covariance between genic fitnesses. This departure is necessary because covariance retains information about both the frequencies and selective value of alleles, and it is convenient because the mean selective advantage accrued by recombinants over the course of a single generation is equal to minus the covariance [29]. Many of our results will thus be given in terms of covariance and we recall: negative covariance means positive selection for recombinants.

In the present study, we reduce the problem to what we believe is its most essential form: we ask how the selective value of haploid recombinants is affected when natural selection simply acts on standing heritable variation. We ask this in two different ways: 1) when recombination occurs across the *products* of natural selection (PRE1 [30]), i.e., across genotypes that have, through the action of natural selection, become locally prevalent in different populations, sub-populations, demes, or competing clones, and 2) when recombination occurs across individuals within the same population during the *process* of natural selection (PRE2 [29]). Most, if not all, of the indirect influences shaping the evolution of recombination can be categorized as deriving from one or the other of these two contexts in which recombination may occur.

To isolate the effects of natural selection, we consider large (effectively infinite) populations. For compactness of presentation, we here describe the simplest scenario in which each population consists of just two competing genotypes that differ in both of two genes (or two *loci*). This simple setting provides a connection to foundational evolution-of-sex studies: Fisher [31] considered the case of a single beneficial mutation arising on a variable background, thereby effectively giving rise to two competing genotypes – wildtype and beneficial mutant – that differ in both the gene with the beneficial mutation (call it the $x$ gene) and its genetic background (call it the $y$ gene); Muller [32] considered the case of two competing genotypes, one carrying a beneficial mutation in the $x$ gene and the other in the $y$ gene. Both of these approaches consider two competing genotypes that differ in both of two loci, and our qualitative findings thus apply to these foundational models and others. Treatment of more loci and more alleles is found in companion papers [30] and [29].

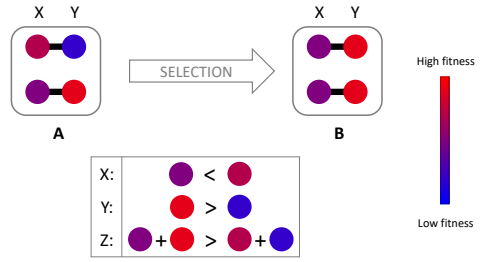


Figure 1. **Two loci, two alleles.** Here, a large (infinite) population consists of individuals whose genome has only two loci $x$ and $y$, each of which carries one of two alleles: genotype 1 encodes quantified phenotype $X_1$ at the $x$ locus and $Y_1$ at the $y$ locus, and genotype 2 carries quantified phenotype $X_2$ at the $x$ locus and $Y_2$ at the $y$ locus. Fitness is indicated by color. An individual's fitness is a function of the two phenotypes: $Z = \phi(X, Y)$; here we make the simplifying assumption that $\phi(X, Y) = X+Y$, so that the fitnesses of genotypes 1 and 2 are $Z_1 = X_1 + Y_1$ and $Z_2 = X_2 + Y_2$, respectively. The fitter of these two genotypes has total fitness denoted $Z^{[2]}$ (i.e., $Z^{[2]} = \text{Max}\{Z_1, Z_2\}$) and genic fitnesses $X_{(2)}$ and $Y_{(2)}$ (i.e., $Z^{[2]} = X_{(2)} + Y_{(2)}$). Similarly, the less-fit of these two genotypes has total fitness $Z^{[1]} = X_{(1)}+Y_{(1)}$. We note: $Z^{[2]} > Z^{[1]}$ by definition, but this does *not* guarantee that $X_{(2)} > X_{(1)}$ or that $Y_{(2)} > Y_{(1)}$, as illustrated in the lower box. The population labeled $A$ consists of two distinct genotypes but selection acts to remove the inferior genotype leaving a homogeneous population in which individuals are all genetically identical (with fitness $Z^{[2]}$) as illustrated in the population labeled $B$. We derive selective mismatch measured by covariance $\sigma_{XY}$: 1) across populations (among different $B$), given by $\sigma_{X_{(2)}Y_{(2)}}$, and 2) within populations (going from $A$ to $B$), given by $\int_0^\infty \sigma_{XY}(t)dt$.

Figure 1 illustrates how the simplest version of the problem is posed analytically. We consider a clonal haploid organism whose genome consists of just two fitness-related loci labeled $x$ and $y$. Genetically-encoded phenotypes at these two loci are quantified by random variables $X$ and $Y$, both of which are positively correlated with fitness. In each large population of such organisms, two genotypes exist: one encodes the phenotype $(X_1, Y_1)$, has fitness $Z_1 = \phi(X_1, Y_1)$ and exists at some arbitrary initial frequency $p$; the other encodes phenotype $(X_2, Y_2)$, has fitness $Z_2 = \phi(X_2, Y_2)$ and exists at initial frequency $1 - p$. We note that, in the absence of epistasis or dominance, the scenario we describe is formally equivalent to considering a diploid organism whose genome consists of one locus and two alleles available to each haploid copy. The question we ask is this: Does the action of natural selection, by itself, affect covariance between $X$ and $Y$,

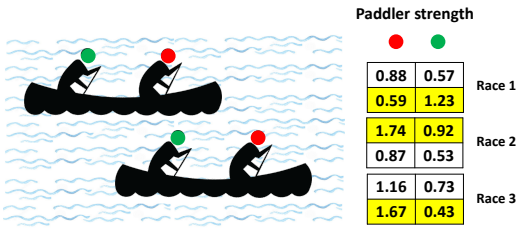


Figure 2. **Canoe race analogy.** Each canoe contains two paddlers. The strength of each paddler is measured and reported in the table. In any given canoe race, there is no correlation between paddler strengths $A$ (green) and $B$ (red). In each race, paddler strengths are recorded (tables on right), and the winning canoe is that in which the sum of the strengths of the two paddlers is the greatest (highlighted). Three such canoe races are conducted. We ask: what is the covariance between the strengths of paddlers $A$ and $B$ among winning canoes only? While it seems reasonable to suppose that winning canoes would carry two strong paddlers thereby resulting in positive covariance, the counter-intuitive answer we find is that the covariance is, for all practical purposes, unconditionally negative in expectation. By analogy, paddlers are genes, paddler strength is genic fitness, and canoes are genotypes. Natural selection picks the winner.

denoted $\sigma_{XY}$, and if so, how?

Figure 2 illustrates the problem by analogy to a set of canoe races. On the surface, one might suspect that natural selection would promote well-matched combinations in which large values of $X$ are linked to large values of $Y$, thereby creating a positive association between $X$ and $Y$. In fact, this notion is so intuitive that it is considered self-evident, explicitly or implicitly, in much of the literature [3–5, 14, 21, 33–37]. If this notion were true, recombination would break up good allelic combinations, on average, and should thus be selectively suppressed. Such allele shuffling has been called "genome dilution", a label that betrays its assumed costliness. We find, however, that the foregoing intuition is flawed. To the contrary, we find that natural selection will, on average, promote an excess of mismatched combinations in which large values of $X$ are linked to small values of $Y$, or vice versa, thereby creating a negative association between $X$ and $Y$. Recombination will on average break up the mismatched combinations amplified by natural selection, assemble well-matched combinations, and should thus be favored.

Figure 3 illustrates why our initial intuition was wrong and why natural selection instead tends to create negative fitness associations among genes. For simplicity of presentation, we assume here that an individual's fitness is $Z = \phi(X, Y) = X + Y$, i.e., that $X$ and $Y$ are simply additive genic fitness contributions, and that $X$ and $Y$ are independent. In the absence of recombination, selection does not act independently on $X$ and $Y$ but on their sum, $Z = X + Y$. Perhaps counter-intuitively, this fact alone creates negative associations. To illustrate, we suppose that we know the fitness of successful genotypes to be some constant, $z$, such that $X + Y = z$; here, we have the situation illustrated in Fig. 3a and we see that $X$ and $Y$ are negatively associated; indeed, covariance is immediate: $\sigma_{XY} = -\sigma_X \sigma_Y \leq 0$. Of course, in reality the fitnesses of successful genotypes will not be known *a priori* nor will they be equal to a constant; instead, they will follow a distribution of maxima of $Z$ as illustrated in Fig. 3b. This is because, in large populations, the successful genotype will practically always be the genotype of maximum fitness. If populations consist of $n$ contending genotypes, then the successful genotype will have fitness $Z^{[n]} = X_{(n)} + Y_{(n)}$, the $n^{th}$ order statistic (maxima) of $Z$ with genic components $X_{(n)}$ and $Y_{(n)}$ (called *concomitants* in the probability literature [38, 39]). In general, $Z^{[n]}$ will have smaller variance than $Z$. Components $X_{(n)}$ and $Y_{(n)}$, therefore, while not exactly following a line as in Fig. 3a, will instead be constrained to a comparatively narrow distribution about that straight line, illustrated by Fig. 3b, thereby creating a negative association. Figure 3c plots ten thousand simulated populations evolving from their initial (green dots) to final (black dots) mean fitness components; this panel confirms the predicted negative association. More rigorous confirmation that selected genotypes will tend to carry selectively mismatched alleles across loci is found in the general mathematical proofs and further simulations of our companion paper (PRE1 [30]).

What we have shown so far is that, if recombination occurs across different *products* of natural selection, the resulting offspring should be more fit than their parents, on average. This effect provides novel insight into established observations that population structure can favor recombination [37, 40, 45–47] and speaks to notions that out-crossing can create hybrid vigor (heterosis) by providing a general theoretical basis for pseudo-overdominance [48–53] (explained in PRE1 [30]).

Much of evolution indeed takes place in structured meta-populations providing ample opportunity for cross-population (or cross-deme) recombination. It is thought, for example, that primordial life forms evolved primarily on surfaces that provided spatial structure [54, 55] which can enhance the advantage of recombination [5, 36, 37, 47, 56]. It is also true, however, that much of evolution takes place within unstructured (or "well-mixed") populations; primitive life forms, for example, also existed in planktonic form [57]. We now turn to the question of how the process of evolution by natural selection affects the selective value of recombinants and the evolutionary fate of recombination within such unstructured populations (PRE2 [29]).

We begin by studying the selective value of recombi-

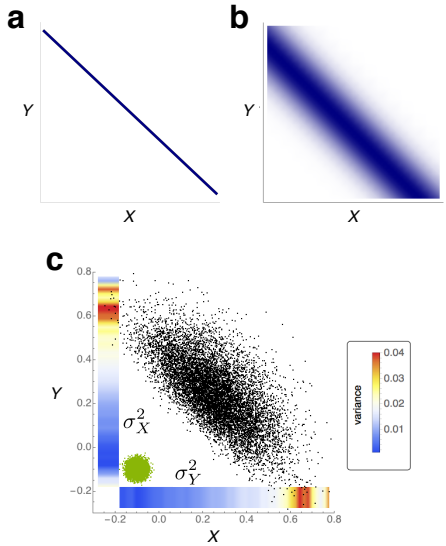


Figure 3. **Natural selection promotes negative associations.** In the absence of recombination, selection does not act independently on $X$ and $Y$ but organismal fitness which, for simplicity, we here assume to be their sum, $Z = \phi(X, Y) = X + Y$. Perhaps counterintuitively, this fact alone creates negative associations. As discussed in the main text, this fact gives rise to a correlation of exactly negative one when the sum is a constant (**a**) and something intuitively negative when the sum is distributed as expected (**b**), i.e., as an order statistic. (**c**), Ten thousand simulated populations move from their initial (green dots) to final (black dots) mean fitnesses. Here, the predicted negative covariance in the final state is apparent. The heatmap bars indicate variance in $Y$ along the $x$-axis and variance in $X$ along the $y$-axis, a manifestation of Hill-Robertson interference [7, 40–44]: larger genic fitness at one locus relaxes selection on the other locus allowing for larger fitness variance at the that locus.

nants. Here again, Fig. 1 shows how the problem is posed analytically. Natural selection will cause the two competing genotypes to change in frequency, causing $\sigma_{XY}$ to change over time ($\sigma_{XY} = \sigma_{XY}(t)$). Our measure of the net effect of natural selection on recombinants is the quantity $\int_0^\infty \sigma_{XY}(t)dt$; if this quantity is positive (negative), we conclude that natural selection opposes (favors) recombinants on average.

In our companion paper (PRE2 [29]), we show that, in expectation, time-averaged covariance is unconditionally non-positive, $\mathbb{E}[\int_0^\infty \sigma_{XY}(t)dt] \leq 0$, implying that *the process of natural selection always creates conditions that favor recombinants, in expectation.*

This remarkable finding requires no assumptions about the bivariate distribution of genic fitness contributions $(X, Y)$ in the initial population; in fact, a smooth density is not required (PRE2 [29]). This distribution can even have strongly positive covariance, and yet the net effect of natural selection is still to create negative time-integrated covariance. Put differently, *recombinants are advantageous, in expectation, completely independently of the source of heritable variation upon which natural selection acts* – whether it be drift, migration, mutation, etc, or what the specific parameters, dynamics or interactions of these processes might be. Furthermore, this result holds unconditionally when genic fitness contributions are additive (countering equilibrium-based studies [23, 24, 26]), and it holds when the non-additive component (epistasis) is only loosely constrained to a wide interval about zero (PRE2 [29]). Lastly, temporal fluctuations in fitness and/or epistasis, as invoked by some previous studies [24, 26, 58–61], are not required.

Our analyses further show that natural selection creates recombinant advantage in expectation even when recombinants are present in the initial variation upon which natural selection acts (PRE2 [29]). Put differently, even in the presence of recombination, the effect of natural selection is to promote increased recombination. The implication is that natural selection not only promotes the evolution of recombination but also its maintenance.

Until now, our focus has been on the selective value of recombinants. Our findings that recombinants are selectively favored in expectation suggest, but do not prove, that recombination should also be selectively favored.

To examine the selective value and fate of *recombination*, we shift our focus from recombinants themselves to a separate locus – a modifier locus – with two alleles: one allele that confers the ability to recombine ($rec^+$) and the wildtype allele that does not ($rec^-$). We now ask: what is the selective value and fate of the $rec^+$ allele?

Our analyses [29] show that the asymptotic selective advantage of the $rec^+$ allele is unconditionally non-negative. This finding is again completely independent of the bivariate fitness distribution governing the initial variation. Our analyses further show that the expected asymptotic frequency of the $rec^+$ allele is effectively equal to the probability that the fittest possible genotype is a virtual (or potential) recombinant. When covariance between $X$ and $Y$ in the initial variation is non-positive, as would be the case for example if the initial variants are themselves products of previous selection [30], this finding implies that the expected asymptotic modifier frequency is $\geq 1 - n^{-(m-1)}$, where $m$ is the number of loci and $n$ the number of alleles per locus. From this expression it is apparent that expected asymptotic modifier frequency can be very close to one under reasonable conditions. Asymptotic modifier frequency is only *not* close to one in expectation in the very unrealistic case in which the correlation coefficient of the initial fitness distribution is extremely close to +1. Remarkably, expected asymptotic modifier frequency is independent of the strength of selection [29]. This observation runs counter to prevailing concerns in the literature that strong selection is required.

Some authors [5, 62] have argued that negative associations build up within a population because positive as-

sociations, in which alleles at different loci are selectively well-matched, are either removed efficiently (when they are both similarly deleterious), or fixed efficiently (when they are both similarly beneficial), thereby contributing little to overall within-population associations. Genotypes that are selectively mismatched, on the other hand, have longer sojourn times, as the less-fit loci effectively shield linked higher-fitness loci from selection. The net effect, it is argued, should be that alleles across loci will on average be selectively mismatched within a population. On one hand, our findings (PRE1 [30]) differ slightly from these arguments: we find that even genotypes that are ultimately fixed carry selectively mismatched alleles. On the other hand, however (PRE2 [29]), our findings are entirely consistent with these arguments; indeed, these arguments provide an intuitive way to understand our remarkable Proposition 2 in PRE2 [29].

The first part of our study identifies a phenomenon that is an inherent consequence of natural selection and gives rise to selectively mismatched combinations of alleles across loci (Fig 2). Generally speaking, this pervasive phenomenon is an example of counter-intuitive effects caused by probabilistic conditioning. For example, "Berkson's paradox" [63, 64] arises when a biased observational procedure produces spurious negative correlations. In the original context, among those admitted to hospital due to illness, a negative correlation among potentially causative factors was observed because those with no illness (who tended to have no causative factors) were not admitted to the hospital and hence not observed. Similarly, negative correlations arise across genic fitnesses in part because genotypes in which both loci have low genic fitness are purged by selection; here, however, the bias is not observational but actual, as these low-fitness genotypes no longer exist in the population.

Our rationale for isolating natural selection was that this approach provides an in-depth understanding of one process – arguably the most influential process – by itself. Most previous studies incorporate several processes simultaneously; these processes may include natural selection, mutation, drift, migration and other processes. While including several processes at once is more realistic, such an approach can make it difficult to decipher which processes are doing what. Taken together, the findings of our study clearly show that natural selection, by itself, has a remarkably encompassing – bordering on universal – tendency to create selective conditions that favor recombination, both across the *products* of selection and during the *process* of selection.

The present study is limited by time: natural selection in isolation is a transient process that stops when one genotype is fixed. In order to study the long-term evolution of recombination, it will be necessary to incorporate a process that continually feeds a population with new heritable variation – a process such as migration or mutation. Incorporating mutation into our analyses is the subject of a forthcoming study. The dynamical sufficiency afforded by incorporating mutation will allow this forthcoming study to explore an intriguing connection between the within- (PRE2 [29]) and between- (PRE1 [30]) population forces acting on recombination. This connection derives from statistical physics – more specifically, from ergodic theory [65–67] – and will complement the present study in offering a qualitatively new way to understand the ubiquity of sex and recombination in nature.

Because natural selection in isolation is a transient process, the particular flavor of selection invoked (directional, stabilizing, truncation, etc) is relatively immaterial to the present study. As long as the rate of environmental change, for example, is less than the rate at which selective sorting results in fixation, our results still apply. The initial bivariate fitness distribution may reflect a population close to, at, or far from a fitness peak, for example, or it may reflect a fitness peak that is moving randomly or not. The facts that our results derive from a transient process and are independent of the initial distribution indicate that our qualitative results are robust to these different scenarios and forms of selection.

Many previous studies, in one way or another, point to the increase in agility and efficiency of adaptation that recombination confers as the primary cause of its evolution. Here, we have inverted the perspective of those earlier studies, asking not whether recombination speeds adaptation, but whether adaptation via natural selection generally creates selective conditions that promote the emergence of recombination. If so, as our findings indicate, then: 1) the ubiquity of recombination in nature might be less enigmatic than previously thought, and 2) perhaps recombination arose and is maintained more as an unavoidable byproduct than as a catalyst of natural selection.

Much of this work was performed during a CNRS-funded visit (P.G.) to the Laboratoire Jean Kuntzmann, University of Grenoble Alpes, France, and during a visit to Bielefeld University (P.G.) funded by Deutsche Forschungsgemeinschaft (German Research Foundation, DFG) via Priority Programme SPP 1590 Probabilistic Structures in Evolution, grants BA 2469/5-2 and WA 967/4-2. P.G. and A.C. received financial support from the USA/Brazil Fulbright scholar program. P.G. and P.S. received financial support from National Aeronautics and Space Administration grant NNA15BB04A. The authors thank S. Otto and N. Barton for their thoughts on early stages of this work. Special thanks go to E. Baake for her thoughts on later stages of this work and help with key mathematical aspects. The authors thank D. Chencha, J. Streelman, R. Rosenzweig and the Biology Department at Georgia Institute of Technology for critical infrastructure and computational support.

* This article is published in concert with two companion papers referenced as PRE1 [30]and PRE2 [29]and Supplemental Materials referenced by the adding the prefix "S".

† pgerrish@unm.edu


[1] W. F. Bodmer and P. A. Parsons, Linkage and recombination in evolution, in *Advances in Genetics*, Vol. 11, edited by E. W. Caspari and J. M. Thoday (Academic Press, 1963) pp. 1–100.

[2] M. Nei, Modification of linkage intensity by natural selection, Genetics **57**, 625 (1967).

[3] S. P. Otto and T. Lenormand, Resolving the paradox of sex and recombination, Nat. Rev. Genet. **3**, 252 (2002).

[4] M. Hartfield and P. D. Keightley, Current hypotheses for the evolution of sex and recombination, Integr. Zool. **7**, 192 (2012).

[5] S. P. Otto, The evolutionary enigma of sex, Am. Nat. **174 Suppl 1**, S1 (2009).

[6] N. H. Barton and B. Charlesworth, Why sex and recombination?, Science **281**, 1986 (1998).

[7] J. Felsenstein, The evolutionary advantage of recombination, Genetics **78**, 737 (1974).

[8] M. W. Feldman, Equilibrium studies of two locus haploid populations with recombination, Theor. Popul. Biol. **2**, 299 (1971).

[9] M. W. Feldman and U. Liberman, An evolutionary reduction principle for genetic modifiers, Proceedings of the National Academy of Sciences **83**, 4824 (1986).

[10] U. Liberman and M. W. Feldman, Modifiers of mutation rate: A general reduction principle (1986).

[11] L. Altenberg, U. Liberman, and M. W. Feldman, Unified reduction principle for the evolution of mutation, migration, and recombination (2017).

[12] M. W. Feldman, Selection for linkage modification. i. random mating populations, Theor. Popul. Biol. **3**, 324 (1972).

[13] I. Eshel and M. W. Feldman, On the evolutionary effect of recombination, Theor. Popul. Biol. **1**, 88 (1970).

[14] S. A. West, C. M. Lively, and A. F. Read, A pluralist approach to sex and recombination, J. Evol. Biol. **12**, 1003 (1999).

[15] S. P. Otto and N. H. Barton, Selection for recombination in small populations, Evolution **55**, 1921 (2001).

[16] H. Bernstein, H. C. Byerly, F. A. Hopf, and R. E. Michod, Origin of sex, J. Theor. Biol. **110**, 323 (1984).

[17] H. Bernstein, Germ line recombination may be primarily a manifestation of DNA repair processes, J. Theor. Biol. **69**, 371 (1977).

[18] H. Bernstein, F. A. Hopf, and R. E. Michod, The molecular basis of the evolution of sex, Adv. Genet. **24**, 323 (1987).

[19] H. Bernstein and C. Bernstein, Evolutionary origin of recombination during meiosis, Bioscience **60**, 498 (2010).

[20] H. Bernstein, H. C. Byerly, F. A. Hopf, and R. E. Michod, Genetic damage, mutation, and the evolution of sex, Science **229**, 1277 (1985).

[21] J. A. G. M. de Visser and S. F. Elena, The evolution of sex: empirical insights into the roles of epistasis and drift, Nat. Rev. Genet. **8**, 139 (2007).

[22] J. Felsenstein and S. Yokoyama, The evolutionary advantage of recombination. II. individual selection for recombination, Genetics **83**, 845 (1976).

[23] N. H. Barton, Linkage and the limits to natural selection, Genetics **140**, 821 (1995).

[24] N. H. Barton, A general model for the evolution of recombination, Genet. Res. **65**, 123 (1995).

[25] N. H. Barton, Genetic linkage and natural selection, Philos. Trans. R. Soc. Lond. B Biol. Sci. **365**, 2559 (2010).

[26] S. P. Otto and M. W. Feldman, Deleterious mutations, variable epistatic interactions, and the evolution of recombination, Theor. Popul. Biol. **51**, 134 (1997).

[27] S. P. Otto and N. H. Barton, The evolution of recombination: removing the limits to natural selection, Genetics **147**, 879 (1997).

[28] M. Slatkin, Linkage disequilibrium—understanding the evolutionary past and mapping the medical future, Nat. Rev. Genet. **9**, 477 (2008).

[29] P. J. Gerrish, F. Cordero, B. Galeota-Sprung, A. Colato, V. Vejalla, and P. Sniegowski, Natural selection promotes the evolution of recombination 2: during the selective process, Physical Review E **In Review** (2021).

[30] P. J. Gerrish, B. Galeota-Sprung, P. Sniegowski, J. Chevallier, and B. Ycart, Natural selection promotes the evolution of recombination 1: among selected genotypes, Physical Review E **In Review** (2021).

[31] R. A. Fisher, *The genetical theory of natural selection* (Oxford Clarendon Press, 1930) p. 302.

[32] H. J. Muller, Some genetic aspects of sex, Am. Nat. **66**, 118 (1932).

[33] K. Jaffe, Emergence and maintenance of sex among diploid organisms aided by assortative mating, Acta Biotheor. **48**, 137 (2000).

[34] Redfield, A truly pluralistic view of sex and recombination, J. Evol. Biol. **12**, 1043 (1999).

[35] A. F. Agrawal, Evolution of sex: Why do organisms shuffle their genotypes?, Curr. Biol. **16**, R696 (2006).

[36] A. F. Agrawal, Spatial heterogeneity and the evolution of sex in diploids, Am. Nat. **174 Suppl 1**, S54 (2009).

[37] L. Becks and A. F. Agrawal, Higher rates of sex evolve in spatially heterogeneous environments, Nature **468**, 89 (2010).

[38] S. S. Yang, General distribution theory of the concomitants of order statistics, Ann. Stat. **5**, 996 (1977).

[39] H. A. David, Concomitants of extreme order statistics, in *Extreme Value Theory and Applications: Proceedings of the Conference on Extreme Value Theory and Applications, Volume 1 Gaithersburg Maryland 1993*, edited by J. Galambos, J. Lechner, and E. Simiu (Springer US, Boston, MA, 1994) pp. 211–224.

[40] G. Martin, S. P. Otto, and T. Lenormand, Selection for recombination in structured populations, Genetics **172**, 593 (2006).

[41] P. D. Keightley and S. P. Otto, Interference among deleterious mutations favours sex and recombination in finite populations, Nature **443**, 89 (2006).

[42] D. Roze and N. H. Barton, The Hill-Robertson effect and the evolution of recombination, Genetics **173**, 1793 (2006).

[43] A. O. B. Whitlock, K. M. Peck, R. B. R. Azevedo, and C. L. Burch, An evolving genetic architecture interacts with Hill–Robertson interference to determine the benefit of sex, .

[44] W. G. Hill and A. Robertson, The effect of linkage on limits to artificial selection, Genet. Res. **8**, 269 (1966).

[45] M. Hartfield, S. P. Otto, and P. D. Keightley, The maintenance of obligate sex in finite, structured populations

subject to recurrent beneficial and deleterious mutation, Evolution **66**, 3658 (2012).

[46] A. O. B. Whitlock, R. B. R. Azevedo, and C. L. Burch, Population structure promotes the evolution of costly sex in artificial gene networks, Evolution **73**, 1089 (2019).

[47] T. Lenormand and S. P. Otto, The evolution of recombination in a heterogeneous environment, Genetics **156**, 423 (2000).

[48] D. M. Waller, Addressing darwin's dilemma: Can pseudo-overdominance explain persistent inbreeding depression and load?, Evolution **75**, 779 (2021).

[49] X. Li, X. Li, E. Fridman, T. T. Tesso, and J. Yu, Dissecting repulsion linkage in the dwarfing gene dw3 region for sorghum plant height provides insights into heterosis, Proc. Natl. Acad. Sci. U. S. A. **112**, 11823 (2015).

[50] J. A. Birchler, H. Yao, S. Chudalayandi, D. Vaiman, and R. A. Veitia, Heterosis, Plant Cell **22**, 2105 (2010).

[51] E. T. Bingham, Role of chromosome blocks in heterosis and estimates of dominance and overdominance, in *Concepts and Breeding of Heterosis in Crop Plants* (Crop Science Society of America, Madison, WI, USA, 2015) pp. 71–87.

[52] J. A. Birchler, D. L. Auger, and N. C. Riddle, In search of the molecular basis of heterosis, Plant Cell **15**, 2236 (2003).

[53] J. A. Birchler, H. Yao, and S. Chudalayandi, Unraveling the genetic basis of hybrid vigor, Proc. Natl. Acad. Sci. U. S. A. **103**, 12957 (2006).

[54] J. T. Trevors, Hypothesized origin of microbial life in a prebiotic gel and the transition to a living biofilm and microbial mats, C. R. Biol. **334**, 269 (2011).

[55] H.-C. Flemming and S. Wuertz, Bacteria and archaea on earth and their abundance in biofilms, Nat. Rev. Microbiol. **17**, 247 (2019).

[56] S. P. Otto and T. Lenormand, Resolving the paradox of sex and recombination, Nat. Rev. Genet. **3**, 252 (2002).

[57] S. L. Miller, J. W. Schopf, and A. Lazcano, Oparin's "origin of life": Sixty years later, J. Mol. Evol. **44**, 351 (1997).

[58] R. D. Kouyos, S. P. Otto, and S. Bonhoeffer, Effect of varying epistasis on the evolution of recombination, Genetics **173**, 589 (2006).

[59] S. P. Otto and Y. Michalakis, The evolution of recombination in changing environments, Trends Ecol. Evol. **13**, 145 (1998).

[60] S. Gandon and S. P. Otto, The evolution of sex and recombination in response to abiotic or coevolutionary fluctuations in epistasis, Genetics **175**, 1835 (2007).

[61] A. D. Peters and C. M. Lively, The red queen and fluctuating epistasis: A population genetic analysis of antagonistic coevolution, Am. Nat. **154**, 393 (1999).

[62] N. H. Barton and S. P. Otto, Evolution of recombination due to random drift, Genetics **169**, 2353 (2005).

[63] J. B. Miller and A. Sanjurjo, A bridge from monty hall to the hot hand: The principle of restricted choice, J. Econ. Perspect. **33**, 144 (2019).

[64] J. Berkson, Limitations of the application of fourfold table analysis to hospital data, Biometrics **2**, 47 (1946).

[65] E. Baake and M. Baake, Ancestral lines under recombination, in *Probabilistic structures in evolution* (2021) pp. 365–382.

[66] C. C. Moore, Ergodic theorem, ergodic theory, and statistical mechanics, Proc. Natl. Acad. Sci. U. S. A. **112**, 1907 (2015).

[67] T. C. B. McLeish, Are there ergodic limits to evolution? ergodic exploration of genome space and convergence, Interface Focus **5**, 20150041 (2015).