Natural selection and the advantage of recombination*

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

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

(Dated: 25 July 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, to the contrary, that on average: 1) natural selection amplifies poorly-matched gene combinations and 2) creates time-averaged negative associations in the process. Recombination breaks up these poorly-matched combinations, neutralizes the negative associations, and should thus be passively and universally favored.

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.

To study how indirect selection affects recombination rate, a common approach is to model two or more fitness-related genes (or *loci*) – among which recombination may occur – as well as an additional locus, called a *modifier* locus, that determines the recombination rate. The action of natural selection on the fitness-related loci can indirectly affect the selective value and fate of different gene variants, or *alleles*, that confer different recombination rates at the modifier locus, thereby causing the mean recombination rate of the population to increase or decrease.

An allele at the modifier locus can have short-term and long-term effects [22] that can, in theory, complement or oppose each other. In the short term, modifiers that increase the recombination rate, or *up-modifiers*, will be indirectly favored if the population harbors 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. Assembling selectively matched combinations increases the efficiency of natural selection: putting high-fitness alleles together can expedite their fixation [22–26], and putting low-fitness alleles together can expedite their elimination [27, 28]. In the long term, up-modifiers of recombination rate, can be indirectly favored because of the fitness variation they augment [22]. This long-term advantage was identified in the earliest speculations as to why recombination (and sex) might have evolved [29] which, curiously, predates the rediscovery of Mendel's work and was thus written under

the paradigm of Darwinian blending inheritance.

A modifier can itself be subject to the very recombination it modulates and can thus have limited-term linkage to the fitness loci whose recombination rate it modifies. Whether the selective value of recombination is determined by short-term or long-term effects depends on how long a modifier will typically remain linked to the fitness loci whose recombination rate it modifies; a loosely-linked modifier will be affected by short-term effects whereas a tightly-linked modifier will be affected by both short- and long-term effects. We derive the selective value and dynamics of a recombination-competent (rec^+) modifier under loose and tight modifier linkage.

To address the evolution of sex and recombination, we have taken a reductionist approach. Our aim is restricted to studying the effects of one very key process, namely natural selection, in isolation (no mutation, no drift, etc), and we distill this problem to what we believe is its most essential form: we ask, how does the action of natural selection, by itself, affect the selective value and fate of recombination? Details of our analyses, proofs, simulation descriptions, and generalizations to m loci and n genotypes are found in companion publications EV1 [30] and EV2 [31] as well as the Supplemental Materials (SM).

We consider a large population consisting of an organism with two loci and a number of distinct alleles at each locus. An allele at the first locus contributes an amount X to total organismal fitness; an allele at the second locus contributes an amount Y to total organismal fitness; X and Y are random variables. We let σ_{XY} denote covariance between genic fitness contributions X and Y. We find that when linkage between modifier and fitness loci is incomplete, the selective advantage of a recombination-competent (rec^+) modifier in an otherwise non-recombining population is $\bar{s}_r \approx -\tau\sigma_{XY}$, where τ is the characteristic duration of linkage between modifier and fitness loci. Because we are interested primarily in the sign of the selective value of recombination, we can let $\tau = 1$ without loss of generality. To make our language precise, we define "loose linkage" to mean $\tau = 1$, and we will let $\hat{s}_r$ denote selective advantage under loose linkage; we note that the selective value of recombination is equal to the selective value of recombinants in this case. We define "incomplete linkage" to mean $\tau > 1$, and we will let $\bar{s}_r$ denote selective advantage under incomplete linkage.

The selective value of recombination under loose or incomplete linkage is naturally partitioned in an illuminating and biologically meaningful way by the two terms of the *total covariance*:

$$\sigma_{XY} \;=\; \langle\sigma_{XY}\rangle \;+\; \sigma_{\langle X\rangle\langle Y\rangle}$$

where angular brackets denote expectation in a finite population (or ensemble average). The two terms on the right-hand side correspond to the two parts of our study. When natural selection acts in isolation, the first term is transient and we are interested in its time integral as an indication of whether recombination is promoted or suppressed on average during the *process* of natural selection within local populations. The second term is not transient and we are interested in its temporal limit as an indication of whether recombination is promoted or suppressed on average between different *products* of natural selection coming from different local populations. Mathematically, the two prongs of our study thus focus on the quantities $\int_0^\infty \langle\sigma_{XY}\rangle dt$ (EV2 [31]) and $\lim_{t\to\infty} \sigma_{\langle X\rangle\langle Y\rangle}$ (EV1 [30]).

We note that covariance is related to a commonly-used metric for selective imbalance affecting recombination, namely linkage disequilibrium (LD) [7, 22, 24, 32–35], which measures bias in allelic frequencies across loci but does not retain information about the selective value of those alleles. Covariance, on the other hand, retains information about both the frequencies and selective value of alleles. Negative LD, like negative covariance, is indicative of selective conditions that favor recombination; however, negative LD does not always favor recombination whereas negative covariance does.

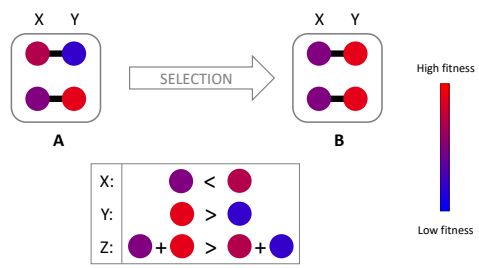

Figure 1. Two loci, two alleles. Here, a large (infinite) population consists of individuals whose genome has only two loci x and y, each of which carries one of two alleles: genotype 1 encodes quantified phenotype X_1 at the x locus and Y_1 at the y locus, and genotype 2 carries quantified phenotype 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 σ_{XY}: 1) across populations (among different B), given by $\sigma_{\langle X\rangle\langle Y\rangle} \xrightarrow{t} \sigma_{X_{(2)}Y_{(2)}}$, and 2) within populations (going from A to B), given by $\int_0^\infty \langle\sigma_{XY}\rangle dt$.

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 fitness-related loci. This simple setting provides a connection to foundational evolution-of-sex studies: Fisher [36] 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 [37] 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.

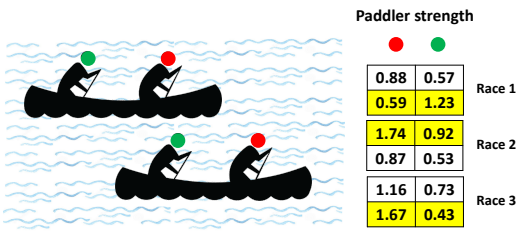

Figure 2. Canoe race analogy. Each canoe contains two paddlers. The strength of each paddler is measured and reported in the table. In any given canoe race, there is no correlation between paddler strengths A (green) and B (red). In each race, paddler strengths are recorded (tables on right), and the 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.

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 $1 - p$. The question we ask is this: Does the action of natural selection, by itself, affect covariance between X and Y 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, 25, 38–41]. 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: $\sigma_{\langle X \rangle \langle Y \rangle} = -\sigma_X \sigma_Y \leq 0$. 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 nth order statistic (maxima) of Z with genic components $X_{(n)}$ and $Y_{(n)}$ (called *concomitants* in the probability literature [42, 43]). 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 pop-

ulations 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 (EV1 [30]).

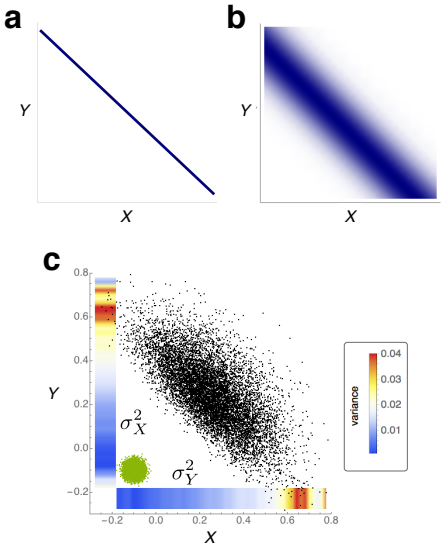

Figure 3. Natural selection promotes negative associations. In the absence of recombination, selection does not act independently on X and Y but organismal fitness which, for simplicity, we here assume to be their sum, $Z = \phi(X, Y) = X+Y$. Perhaps counterintuitively, this fact alone creates negative 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, 28, 44–47]: larger genic fitness at one locus relaxes selection on the other locus allowing for larger fitness variance at the that locus.

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: $\lim_{t\to\infty} \sigma_{\langle X\rangle\langle Y\rangle} < 0$. This effect provides novel insight into established observations that population structure can favor recombination [41, 44, 48–50] and speaks to notions that out-crossing can create hybrid vigor (heterosis) by providing a general theoretical basis for pseudo-overdominance [51–56] (explained in EV1 [30]).

Much of evolution indeed takes place in structured populations providing ample opportunity for cross-subpopulation recombination. It is thought, for example, that primordial life forms evolved primarily on surfaces that provided spatial structure [57, 58] which can enhance the advantage of recombination [5, 40, 41, 50, 59]. 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 [60]. We now turn to the question of how the process of evolution by natural selection affects the selective value and fate of recombination within such unstructured populations (EV2 [31]).

We begin with the case of a loosely-linked modifier. Here again, Fig. 1 shows how the problem is posed analytically. Natural selection will cause the two competing genotypes to change in frequency, causing covariance to change over time. Our measure of the net effect of natural selection on recombination under loose modifier linkage is the quantity $\int_0^\infty \langle\sigma_{XY}\rangle dt$; if this quantity is positive (negative), we conclude that natural selection opposes (favors) recombination on average.

In our companion paper (EV2 [31]), we show that, in expectation, time-averaged covariance is unconditionally non-positive, $\int_0^\infty \langle\sigma_{XY}\rangle dt \leq 0$, implying that *the process of natural selection always creates conditions that favor recombination, in expectation, even when modifier and fitness loci are loosely linked.*

This 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 (EV2 [31]). Remarkably, 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, *recombination is advantageous, in expectation, regardless 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, our result holds unconditionally when genic fitness contributions are additive, in contrast with equilibrium-based studies that require non-additivity [22, 32, 34], and it holds when the non-additive component (epistasis) is only loosely constrained to a wide interval about zero (EV2 [31]). Lastly, temporal fluctuations in fitness and/or epistasis, as invoked by some previous studies [32, 34, 61–64], are not required.

Our analyses further show that natural selection promotes recombination in expectation even when recombinants are present in the initial variation upon which natural selection acts, an immediate consequence of the independence of recombinant advantage on the initial fitness distribution (corroborated in EV2 [31] and SM). 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

recombination when linkage between modifier and fitness loci is incomplete (and loose). We now turn to the case of complete linkage between modifier and fitness loci.

Our analyses (EV2 [31]) show that, under complete linkage, the asymptotic selective advantage of the rec^+ allele is unconditionally non-negative. This finding is again 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 (EV1 [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* (EV2 [31]). This observation may be seen as countering, in a sense, prevailing concerns in the literature that strong selection is required [3, 21].

Otto and Barton [5, 65] have argued that negative associations build up within a population because positive associations, 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 (EV1 [30]) diverge slightly from this *Otto & Barton effect*: we find that even genotypes that are ultimately fixed carry selectively mismatched alleles on average. On the other hand, however (EV2 [31]), our findings are entirely consistent with this effect; indeed, our proof that $\int_0^\infty \langle \sigma_{XY}(u) \rangle du \leq 0$, unconditionally (Prop 3 in EV2 [31]), may be interpreted as a proof of the universality of the Otto & Barton effect.

Our study identifies phenomena that are inherent consequences of natural selection and give rise to negative fitness associations across loci. Generally speaking, this pervasive phenomenon is an example of counter-intuitive effects caused by probabilistic conditioning. For example, "Berkson's paradox" [66, 67] arises when a biased observational procedure produces spurious negative correlations. In the original context, among those admitted to hospital, it is disproportionately more likely for a patient to have, say, diabetes or, say, hypertension than to have both. Observed negative correlation between diabetes and hypertension is therefore often an artefact created by hospital admission and not representative of the general population. The selective agent in this case is hospital admission. The negative correlation may also be understood by considering that people with neither diabetes nor hypertension are less likely to be admitted to the hospital than people chosen at random from the general population.

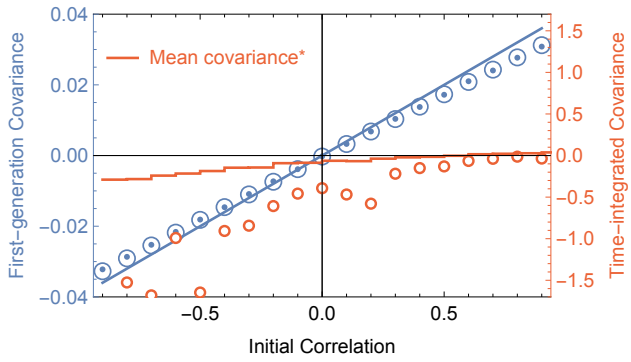

Figure 4. Comparison of the immediate effect (Bulmer effect) vs time-integrated effect of selection on covariance. Blue points plot simulation results after one round of selection starting with 10 genotypes whose two fitness components X and Y were pulled from a bivariate normal distribution with zero means, standard deviations $\sigma_X = 0.2$ and $\sigma_Y = 0.2$, and pre-selection correlation indicated by the horizontal axis and whose pre-selection covariance is the straight blue line. Open blue circles plot the theoretical expectation given by Eq S31 (SM). Our one-generation model effectively mimics the Bulmer effect although from a different theoretical starting point. Orange circles plot Montecarlo expectations of time-integrated covariance as given by Prop 3 in EV2 [31]. Orange line plots per-generation mean covariance *[multiplied by a factor of 30 to make the trend visible]. As there was no way to obtain these means analytically, plotted are the means of 500 simulations of populations of size 10^5 initially consisting of two genotypes at equal frequency.

Similarly, negative correlations arise across genic fitnesses in part because genotypes in which both loci have low genic fitness are purged by selection; in this case, however, the bias is not observational but actual, as these low-fitness genotypes are selectively removed from the population. In this genetic context, where selection is imposed artificially through breeding, this biasing effect has been contemplated and modeled; it is known as the "Bulmer effect" [68–73]. When a desired (artificially selected) trait is determined by several genetic components, one generation of breeding for the desired trait will result in negative associations among the trait's components.

On the surface, the Bulmer effect appears similar to the phenomena we describe. On closer examination, however, it differs in several critical ways that are listed and discussed in EV1 [30] and the SM. In the context of the evolution of recombination, the most consequential difference is depicted in Fig 4. The Bulmer effect is only non-negligible under strong selection in small populations and assumes the initial distribution of fitness components is multivariate normal. In small populations, it can slightly reduce covariance from the starting covariance; if the starting covariance is positive, the Bulmer effect can reduce covariance to a slightly less-positive value, but it does not guarantee that the resulting covariance is negative. In contrast, the phenomenon we describe operates in infinite populations and results in negative time-integrated covariance independent of the initial covariance and, more generally, of the initial distribution of fitness components. In light of the above, it is clear that our findings cannot be dismissed as a simple extrapolation of the Bulmer effect.

Our rationale for isolating natural selection was that this approach provides an in-depth understanding of one process – arguably evolution's most influential process – by itself (an approach taken by much of classical population genetics). 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 tendency to create selective conditions that favor recombination, both across the *products* of selection and during the *process* of selection.

Previous studies have shown that when populations are at mutation-selection-recombination equilibrium (e.g., in the absence of adaptive evolution), decreased recombination rates are always favored, dubbed a "general reduction principle" for recombination [9, 11, 74]. Increasingly, however, empirical evidence suggests that evolution is best described as a non-equilibrium process [75–78] in which adaptive evolution is ongoing [79–83]. The setting we study is one of a population not at equilibrium. The ubiquitous recombination-augmenting tendency we describe in this paper (and companion papers EV1 [30] and EV2 [31]) may perhaps be seen as providing a "general inflation principle" of sorts, for the non-equilibrium case. Indeed, in light of the clean contrast between non-equilibrium and equilibrium findings, the ubiquity of sex and recombination in nature might be interpreted as evidence for the non-equilibrium quality of evolution generally.

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.

We have found that natural selection, left to its own devices, has an encompassing tendency to promote the evolution recombination. Other more stochastic process such as mutation, migration, and drift are not explicitly modeled but are, in sense, modeled implicitly by our assumption that "something produces the heritable variation upon which natural selection acts". In the present study, we are not concerned what that "something" is. Accounting for these other processes in the present framework is an obvious next step.

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 evolution 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.

Acknowledgements

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. P.G. received further support from the National Institute Of General Medical Sciences of the National Institutes of Health under Award Number R35GM137919 (awarded to Gideon Bradburd). 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 EV1 [30] and EV2 [31] 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, Theor. Popul. Biol. **30**, 125 (1986).

[11] L. Altenberg, U. Liberman, and M. W. Feldman, Unified reduction principle for the evolution of mutation, migration, and recombination, Proceedings of the National Academy of Sciences **114**, E2392 (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] N. H. Barton, Linkage and the limits to natural selection, Genetics **140**, 821 (1995).

[23] P. J. Gerrish and R. E. Lenski, The fate of competing beneficial mutations in an asexual population, Genetica **102-103**, 127 (1998).

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

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

[26] J. A. Arjan, M. Visser, C. W. Zeyl, P. J. Gerrish, J. L. Blanchard, and R. E. Lenski, Diminishing returns from mutation supply rate in asexual populations, Science **283**, 404 (1999).

[27] A. S. Kondrashov, Deleterious mutations and the evolution of sexual reproduction, Nature **336**, 435 (1988).

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

[29] A. Weismann, Significance of sexual reproduction, in *Essays upon heredity and kindred biological problems* (Clarendon Press, Oxford, 1889).

[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] 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).

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

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

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

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

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

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

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

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

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

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

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

[43] 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.

[44] G. Martin, S. P. Otto, and T. Lenormand, Selection for recombination in structured populations, Genetics **172**,

593 (2006).
[45] D. Roze and N. H. Barton, The Hill-Robertson effect and the evolution of recombination, Genetics **173**, 1793 (2006).
[46] 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, .
[47] W. G. Hill and A. Robertson, The effect of linkage on limits to artificial selection, Genet. Res. **8**, 269 (1966).
[48] 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).
[49] 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).
[50] T. Lenormand and S. P. Otto, The evolution of recombination in a heterogeneous environment, Genetics **156**, 423 (2000).
[51] D. M. Waller, Addressing darwin's dilemma: Can pseudo-overdominance explain persistent inbreeding depression and load?, Evolution **75**, 779 (2021).
[52] 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).
[53] J. A. Birchler, H. Yao, S. Chudalayandi, D. Vaiman, and R. A. Veitia, Heterosis, Plant Cell **22**, 2105 (2010).
[54] 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.
[55] J. A. Birchler, D. L. Auger, and N. C. Riddle, In search of the molecular basis of heterosis, Plant Cell **15**, 2236 (2003).
[56] 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).
[57] 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).
[58] H.-C. Flemming and S. Wuertz, Bacteria and archaea on earth and their abundance in biofilms, Nat. Rev. Microbiol. **17**, 247 (2019).
[59] S. P. Otto and T. Lenormand, Resolving the paradox of sex and recombination, Nat. Rev. Genet. **3**, 252 (2002).
[60] S. L. Miller, J. W. Schopf, and A. Lazcano, Oparin's "origin of life": Sixty years later, J. Mol. Evol. **44**, 351 (1997).
[61] R. D. Kouyos, S. P. Otto, and S. Bonhoeffer, Effect of varying epistasis on the evolution of recombination, Genetics **173**, 589 (2006).
[62] S. P. Otto and Y. Michalakis, The evolution of recombination in changing environments, Trends Ecol. Evol. **13**, 145 (1998).
[63] 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).
[64] 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).
[65] N. H. Barton and S. P. Otto, Evolution of recombination due to random drift, Genetics **169**, 2353 (2005).
[66] 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).
[67] J. Berkson, Limitations of the application of fourfold table analysis to hospital data, Biometrics **2**, 47 (1946).
[68] G. Gorjanc, P. Bijma, and J. M. Hickey, Reliability of pedigree-based and genomic evaluations in selected populations, Genet. Sel. Evol. **47**, 65 (2015).
[69] E. M. Van Grevenhof, J. A. M. Van Arendonk, and P. Bijma, Response to genomic selection: the bulmer effect and the potential of genomic selection when the number of phenotypic records is limiting, Genet. Sel. Evol. **44**, 26 (2012).
[70] G. M. Tallis, Ancestral covariance and the bulmer effect, Theor. Appl. Genet. **73**, 815 (1987).
[71] M. Dupont-Nivet, J. Mallard, J. C. Bonnet, and J. M. Blanc, Evolution of genetic variability in a population of the edible snail, helix aspersa müller, undergoing domestication and short-term selection, Heredity **87**, 129 (2001).
[72] M. G. Bulmer, Linkage disequilibrium and genetic variability, Genet. Res. **23**, 281 (1974).
[73] M. G. Bulmer, The effect of selection on genetic variability, Am. Nat. **105**, 201 (1971).
[74] U. Liberman and M. W. Feldman, A general reduction principle for genetic modifiers of recombination, Theor. Popul. Biol. **30**, 341 (1986).
[75] V. Mustonen and M. Lässig, Fitness flux and ubiquity of adaptive evolution, Proc. Natl. Acad. Sci. U. S. A. **107**, 4248 (2010).
[76] A. D. Kern and M. W. Hahn, The neutral theory in light of natural selection, Mol. Biol. Evol. **35**, 1366 (2018).
[77] V. Mustonen and M. Lässig, From fitness landscapes to seascapes: non-equilibrium dynamics of selection and adaptation, Trends Genet. **25**, 111 (2009).
[78] E. Kussell and M. Vucelja, Non-equilibrium physics and evolution—adaptation, extinction, and ecology: a key issues review, Rep. Prog. Phys. **77**, 102602 (2014).
[79] I. Gordo, Evolutionary change in the human gut microbiome: From a static to a dynamic view, PLoS Biol. **17**, e3000126 (2019).
[80] L. Perfeito, L. Fernandes, C. Mota, and I. Gordo, Adaptive mutations in bacteria: high rate and small effects, Science **317**, 813 (2007).
[81] D. W. Hall and S. B. Joseph, A high frequency of beneficial mutations across multiple fitness components in saccharomyces cerevisiae, Genetics **185**, 1397 (2010).
[82] D. W. Hall, R. Mahmoudizad, A. W. Hurd, and S. B. Joseph, Spontaneous mutations in diploid saccharomyces cerevisiae: another thousand cell generations, Genet. Res. **90**, 229 (2008).
[83] C. Payen, A. B. Sunshine, G. T. Ong, J. L. Pogachar, W. Zhao, and M. J. Dunham, High-Throughput identification of adaptive mutations in experimentally evolved yeast populations, PLoS Genet. **12**, e1006339 (2016).