SEX VERSUS ASEX: AN ANALYSIS OF THE ROLE OF VARIANCE CONVERSION

ANDREW E.M. LEWIS-PYE[1] & ANTONIO MONTALBÁN[2]

[1]Department of Mathematics, London School of Economics and Political Science, London, UK, WC2A 2AE.
[2]Department of Mathematics, University of California, Berkeley, CA 94720-3840, USA.

Abstract. The question as to why most complex organisms reproduce sexually remains a very active research area in evolutionary biology. Theories dating back to Weismann have suggested that the key may lie in the creation of increased variability in offspring, causing enhanced response to selection. Under appropriate conditions, selection is known to result in the generation of negative linkage disequilibrium, with the effect of recombination then being to increase genetic variance by reducing these negative associations between alleles. It has therefore been a matter of significant interest to understand precisely those conditions resulting in negative linkage disequilibrium, and to recognise also the conditions in which the corresponding increase in genetic variation will be advantageous. Here we prove rigorous results for the multi-locus case, detailing the build up of negative linkage disequilibrium, and describing the long term effect on population fitness for models with and without bounds on fitness contributions from individual alleles. Under the assumption of large but finite bounds on fitness contributions from alleles, the non-linear nature of the effect of recombination on a population presents serious obstacles in finding the genetic composition of populations at equilibrium, and in establishing convergence to those equilibria. We describe techniques for analysing the long term behaviour of sexual and asexual populations for such models, and use these techniques to establish conditions resulting in higher fitnesses for sexually reproducing populations.

1. Introduction

Sexual propagation must certainly confer immense benefits on those populations undergoing it, given that sex involves substantial costs such as the breaking down of favourable gene combinations established by past selection. There are many hypotheses as to the form these advantages take, and they fall naturally into two groups (Felselstein [1], Maynard-Smith [2], Kondrashov [3]). On the one hand a function of sexual reproduction and meiotic recombination may be in providing immediate and physiological benefits, such as allowing repair of double strand DNA damage (Bernstein [4], Michod [5]). Such mechanisms alone, however, are unlikely to account for the continued prevalence of sexual reproduction (Barton and Charlesworth [7],

This paper has no principal author. The ordering is alphabetical. Both authors contributed equally to the construction of proofs and simulations. The first author was supported by a Royal Society University Research Fellowship. The second author was supported by a Packard Fellowship.

Kondrashov [3], Maynard-Smith [6]), and so, on the other hand, decades of research have seen evolutionary biologists looking to develop explicit theoretical models which explain the advantages of sex in terms of the interaction between variation and selection. Many of these models (Barton [8], Otto and Barton [9], Hill and Robertson [10]) focus on ideas originally due to Morgan [11], Fisher [12] and Muller [13] which stress the ability of recombination to place beneficial mutations together on the same chromosome. In a similar vein one may consider the accumulation of deleterious mutations (Muller [14], Felselstein [1]). Since the effect of selection is dictated by levels of genetic variability in a population, one may also look more directly to understand the effect of recombination on genetic variance. The key observation here is that under appropriate conditions negative linkage disequilibria will build up, impeding the response of the population to directional selection (Mather [15], Felselstein [16]).

The mechanisms by which negative linkage equilibrium may be created in the first place, may be classified as either deterministic or stochastic. A key finding for deterministic models (Barton [18]) is that recombination may be favoured when weak negative epistasis (measured relative to the multiplicative contribution of individual gene fitnesses) exists between favourable alleles. There is strong evidence also that stochastic effects (Hill, Robertson, [10] and Barton, Otto [19]) may be substantial in the realistic setting of finite populations. The basic mechanism in this case may be seen as follows. In the rare event that particularly beneficial alleles at distinct loci combine in a single genome, selection acts quickly to achieve fixation for the coupled beneficial alleles, meaning that the associated positive disequilibrium disappears quickly. In the case of a strongly beneficial allele which initially appears on a genome with weaker alleles at other loci, however, selection is slowed down (when recombination is weak or non-existent), meaning that the negative disequilibrium persists for a much longer period of time. Any variance in disequilibrium thus ultimately leads to negative disequilibrium on average.

Here we shall consider a deterministic setting in which sex is seen to robustly outperform asex across a broad spectrum of models, and in which the fitness contributions of genes which can be attained via mutation may be bounded or unbounded. We shall make certain simplifying assumptions. It will be convenient to carry out most of our analysis, for example, relative to models in which individual genes contribute *additively* to the fitness of the genome, and relative to this assumption of additive contributions from individual genes we shall assume zero epistasis. It should be noted that seen relative to models in which genes contribute to fitness multiplicatively, our model therefore assumes negative epistasis, and so may be expected to display benefits to recombination (e.g. Barton [18]). We shall also assume that loci are unlinked, so that they either correspond to loci on distinct chromosomes (one may consider that we are choosing a 'representative' from each chromosome), or else lie at sufficient distances when they share a chromosome. As well as facilitating the mathematical analysis, these simplifications allow us to establish the most basic conditions under which certain mechanisms of variance conversion (described in detail in later sections) will operate with substantial effect. If a phenomenon is already observed in such a model, it is because no extra hypotheses are necessary to make it true - that a cause is already present within the few features of the simple

model. Moreover, analysing our proofs, we can extract key ideas that surely carry over to more general models. An added benefit of working with these simplified models is also a dramatic reduction in the computational complexity of running large simulations. Even before providing mathematical proofs of our results, we are able to run simulations modelling populations with many more loci and more alleles than would otherwise be possible. Simulations for these vast fitness landscapes robustly show sexual populations achieving more rapid increases in mean fitness. Figure 1 shows a small cross-section of the results of simulations for models with finite or infinite haploid populations and where fitness contributions from individual genes may be combined additively or multiplicatively (further examples are given in Figures 6-10 Appendix E). It is worth noting a fact first observed by Maynard-Smith [34] and illustrated in (e) of Figure 1, that in the multiplicative model with zero epistasis and infinite populations beginning in linkage equilibrium, the sexual and asexual populations remain identical. This holds because selection then preserves linkage equilibrium.

We then concentrate our mathematical analysis on the infinite populations additive model, since dealing with this case allows us to avoid some of the complexities inherent in the finite population models while illustrating basic principles which carry through to the finite population additive model. We are able to give a rigorous mathematical analysis of the manner in which, during the process of asexual propagation, a negative linkage disequilibrium will be created and maintained, meaning that an occurrence of recombination at any stage of the process will cause an immediate increase in fitness variance and a corresponding increase in the rate of growth in mean fitness. For contexts where there is a large but finite bound on allele fitnesses, it is not surprising that the long term behaviour differs qualitatively from the case where there is no a priori bound of the fitnesses of genes resulting from mutation. In this case, a standard application of the Perron-Frobenius Theorem suffices to establish that the asexual process converges to a fixed point of the corresponding dynamical system, but a deeper analysis is required in order to establish the mean fitness of the population at this fixed point and to relate this to the long term behaviour for sexual populations. We develop techniques which suffice to carry out such an analysis, and establish higher resulting mean fitnesses for sexual populations in these bounded models.

2. The model

We consider haploid populations with non-overlapping generations. In the absence of dominance between alleles at a single locus, our analysis could easily be extended to consider diploid populations. We describe here the additive infinite population variants of the model (other variants are described in Appendix D). We do not assume alleles come from a pre-existent pool, but consider a (form of random walk mutation) model in which alleles are created by mutation as time passes, possibly without any bound on attainable fitness. For certain aspects of the mathematical analysis it will be convenient to be able to assume that gene fitnesses occur in a discrete range rather than taking any real value. We ensure this by assuming that gene fitnesses take integer values. Appropriate scaling means this entails essentially no loss in

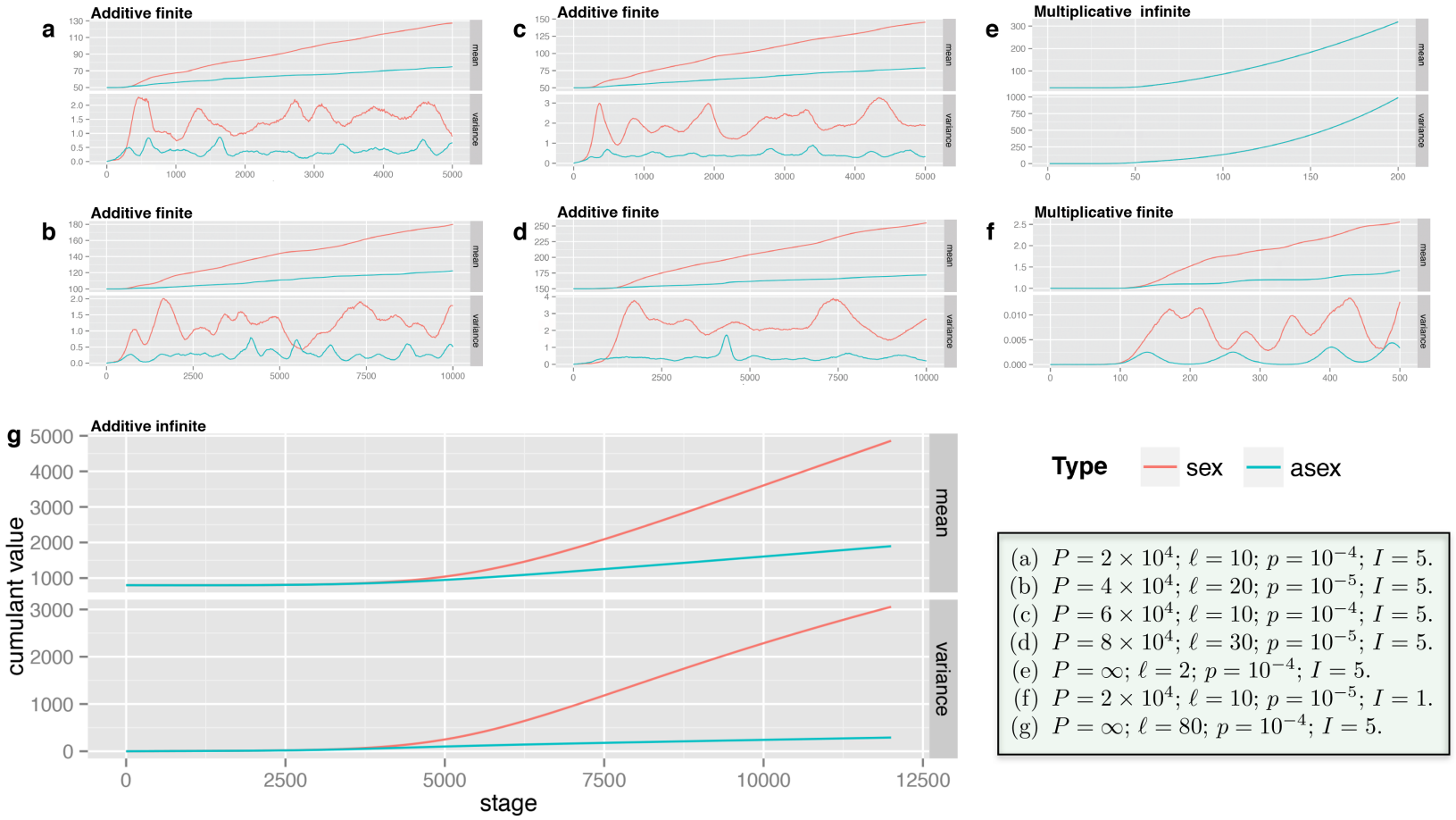

FIGURE 1. The dominance of sexual populations. Each plot corresponds to one simulation and shows the evolution of mean fitness (top) and fitness variance (bottom) over time for sexual and asexual populations beginning in identical states at linkage equilibrium. Finite models are described in Appendix D. Each simulation is specified by a tuple (P, ℓ, p, q, I): $P \in \{\{\infty\} \cup \mathbb{N}\}$ is population size; ℓ is number of loci; p is the probability of mutation; q is the probability a given mutation is beneficial (=0.1 in all simulations displayed here); I is the initial fitness contribution of alleles. For the infinite multiplicative model (e) the sexual and asexual populations stay identical. For the finite multiplicative model (f) the fitness values have been divided by 10^ℓ.

generality – in order to simulate a model in which fitnesses take values to d decimal places, we can simply multiply all fitness values by 10^d, apply the model as described here, and then finally divide by 10^d in order to correct fitness values at any stage of the process. One could consider a model in which fitnesses can take any real values, without substantial changes in the behaviour of the model. Most other features of the model, which we now describe in more detail, are essentially standard in the literature.

Each instance of the model is determined by three principal parameters: ℓ, D and μ. First, $\ell \in \mathbb{N}$ (> 1) specifies the number of loci. With each individual specified by ℓ genes, in the absence of epistasis we need only be concerned with the fitness contributions corresponding to those genes, and so each individual can be identified with a tuple $\boldsymbol{x} = (x_1, ..., x_\ell) \in \mathbb{Z}^\ell$. The *fitness* of $\boldsymbol{x}$ is $F(\boldsymbol{x}) = \sum_{i=1}^{\ell} x_i$. (For the multiplicative model, one would define $F(\boldsymbol{x}) = \prod_{i=1}^{\ell} x_i$ instead.) Second, the *domain* $D \subset \mathbb{Z}^\ell$ determines which individuals are allowed to exist. We will use three types of domains in this paper: The $\mathbb{N}$-*model* uses as domain $D = \mathbb{N}^\ell$, where $\mathbb{N} = \{1, 2, 3, ...\}$; the $\mathbb{Z}$-*model* uses $D = \{\boldsymbol{x} \in \mathbb{Z}^\ell : F(\boldsymbol{x}) > 0\}$; and the *bounded-model* uses $D = \{1, ..., N\}^\ell$ for some upper bound $N \in \mathbb{N}$ on gene fitness contributions. In practice there is almost no difference between the $\mathbb{N}$- and $\mathbb{Z}$-models, but there are situations when it is simpler to consider one or the other. Third, $\mu \colon \mathbb{Z} \to \mathbb{R}^{\geq 0}$, the *mutation probability function*, determines how mutation affects gene fitness contributions: $\mu(k)$ is the probability that the fitness contribution of a gene will increase by k. For the sake of simplicity we assume this distribution to be identical for all loci. While there is no clear canonical choice for μ, the behaviour of the model is robust to changes in this parameter so long as negative mutations are more likely than positive ones, both being possible. This is because any such choice of μ will approximate a Gaussian distribution over multiple generations. The simplest mutation distributions one may consider are those taking non-zero values only on $\{-1, 0, 1\}$. Unless stated otherwise, it should be assumed that from now on mutations are of this form and that $\mu(0) > \mu(-1) > \mu(1)$ (giving a form of stepwise-mutation model [26]).

By a *population* we mean a probability distribution $\boldsymbol{\phi} \colon \mathbb{Z}^\ell \to \mathbb{R}^{\geq 0}$, where $\boldsymbol{\phi}(\boldsymbol{x})$ is the proportion of individuals that have 'genotype' $\boldsymbol{x} \in \mathbb{Z}^\ell$. For a population $\boldsymbol{\phi}$, we shall also use $X = (X_1, ..., X_\ell)$, where the X_i's take values in $\mathbb{Z}$, to denote a random variable that picks an individual with gene fitness contributions X_1,...,X_ℓ according to the distribution given by $\boldsymbol{\phi}$. We let $M(\boldsymbol{\phi})$ denote the *mean fitness* of the population $\boldsymbol{\phi}$, namely $E(F(X))$. It should be assumed throughout that all populations considered have finite means, variances, and that all cumulants are finite (as is the case, for example, for distributions $\boldsymbol{\phi}$ with finite support, i.e. with finitely many $\boldsymbol{x} \in \mathbb{Z}^\ell$ such that $\boldsymbol{\phi}(\boldsymbol{x}) \neq 0$).

For a sexual population, the next generation is obtained by application of three operations: selection, mutation and recombination. We refer to the consecutive application of these operations over multiple generations as the *sex process*. For the *asex process*, the operations applied are selection and mutation, and the recombination phase is omitted. With a much less significant effect, at the end of each generation we will also apply a truncation operation that erases individuals falling outside the domain.

Selection. The fitness of a young zygote is proportional to the number of young zygotes that it will produce. If $\boldsymbol{\phi}$ is the population prior to selection then the resulting population, $\mathtt{Sel}(\boldsymbol{\phi})$, is given by:

$$\mathtt{Sel}(\boldsymbol{\phi})(\boldsymbol{x}) = \frac{F(\boldsymbol{x})}{M(\boldsymbol{\phi})}\boldsymbol{\phi}(\boldsymbol{x}), \quad \text{for } \boldsymbol{x} \in \mathbb{Z}^\ell.$$

The factor $1/M(\boldsymbol{\phi})$ normalises the probability distribution.

Mutation. Let C_i be i.i.d. random variables taking values in $\mathbb{Z}$ with distribution μ. If we apply mutation to a random variable $X = (X_1, ..., X_\ell)$ we get $(X_1 + C_1, ..., X_\ell + C_\ell)$. Equivalently, if $\boldsymbol{\phi}$ $(= \mathtt{Sel}(\boldsymbol{\phi}')$ where $\boldsymbol{\phi}'$ was the population prior to selection) is the population prior to mutation then, for $\boldsymbol{x} \in \mathbb{Z}^\ell$:

$$\mathtt{Mut}(\boldsymbol{\phi})(\boldsymbol{x}) = \sum_{\boldsymbol{y} \in D} \boldsymbol{\phi}(\boldsymbol{y}) \cdot \boldsymbol{\mu}(\boldsymbol{y} - \boldsymbol{x}),$$

where $\boldsymbol{\mu}$ is the extension of μ to a function on $\mathbb{Z}^\ell$ according to the assumption that mutations act independently on distinct loci (i.e., $\boldsymbol{\mu}(a_1, ..., a_\ell) = \prod_{i=1}^{\ell} \mu(a_i)$).

Recombination. For the sake of simplicity we assume that the ℓ loci are unlinked, so that they either correspond to loci on distinct chromosomes (one may consider that we are choosing a 'representative' from each chromosome), or else lie at sufficient distances when they share a chromosome. In general the effect of recombination is to leave the distributions at individual loci unchanged, while bringing the population towards linkage equilibrium. We make the simplifying assumption (for the infinite models) that the effect of a single application of recombination is to bring the population immediately to linkage equilibrium, i.e. to make the random variables X_i independent. If $\phi_i(x) \colon \mathbb{Z} \to \mathbb{R}^{\geq 0}$ is the distribution at locus i, (i.e. $\phi_i(x) = \sum_{\boldsymbol{y} \in D, y_i = x} \boldsymbol{\phi}(\boldsymbol{y})$ where $\boldsymbol{\phi} = \mathtt{Mut}(\mathtt{Sel}(\boldsymbol{\phi}'))$ if $\boldsymbol{\phi}'$ was the population at the start of the present generation) then the resulting population is given by:

$$\mathtt{Rec}(\boldsymbol{\phi})(\boldsymbol{x}) = \prod_{i=1}^{\ell} \phi_i(x_i), \quad \text{for } \boldsymbol{x} = (x_1, ..., x_\ell) \in \mathbb{Z}^\ell.$$

Recombination as we consider it here is thus equivalent to multiple applications of recombination in its standard form. Since in reality the effect of recombination for unlinked loci is to half the linkage disequilibrium (LD_2 as formally defined in the next section) in each generation, linkage disequilibrium is kept to very low levels – so our simplifying assumption is not too large an approximation.

Mutation and recombination may create individuals that fall outside the domain D. At the end of each generation, we therefore perform *truncation* to remove those outlying individuals. $\mathtt{Tru}(\boldsymbol{\phi})(\boldsymbol{x})$ is defined to be $\boldsymbol{\phi}(\boldsymbol{x})/s$ if $\boldsymbol{x} \in D$, and 0 otherwise, where s is the normalising factor $s = \sum_{\boldsymbol{x} \in D} \boldsymbol{\phi}(\boldsymbol{x})$. We will see (Tables 1-3, Appendix E) that under light conditions the proportion of the population moving outside the bounds of D in each generation is negligible, and that truncation along the lower bounds will have an insignificant effect on the whole process.

A final comment concerning notation before we analyse the model: it is standard practice in the population genetics literature to normalise so that the initial genotype ("the wildtype") has

fitness 1 and then have other than wildtype alleles each associated with a selection coefficient s_i. An individual would then have fitness $1 + \sum_{i=1}^{\ell} s_i$ in the additive model and $\prod_{i=1}^{\ell}(1 + s_i)$ in the multiplicative model. Of course normalisation does not really have an effect on the process, and one could use the same convention here. The reason that we do not is that this would severely complicate dealing with the mutation operation. As the model is defined here, the mutation distribution remains constant throughout the process. This means, for example, that if at a later stage of the evolutionary process individuals have higher average fitnesses, allele mutations are then likely to have *proportionately* less effect on the fitness of the individual.

3. Analysing the model

The objective now is to give a mathematical analysis establishing higher mean fitness for sexual populations (reduction to selection at the gene level can then be achieved in a standard fashion, by consideration of the effect of selection on genes which code for sexual rather than asexual reproduction). Proofs of all claims in this section appear in Appendix A, Appendix B and Appendix C.

First let us review the *direct* effect of selection, mutation and recombination on mean fitness. Here we shall simply state the facts, but proofs of all claims appear in Appendix A. Each generation sees two forces acting on the mean fitness $M = M(\boldsymbol{\phi})$. On the one hand, mutation causes a fixed decrease in M by an amount that depends only on μ – recall that deleterious mutations are more likely than beneficial ones. Selection, on the other hand, can be shown to increase mean fitness by V_F/M – a form of Fisher's fundamental theorem [12] – where $V_F = V_F(\boldsymbol{\phi}) = \mathrm{Var}(F(X))$ is the variance of the fitness of $\boldsymbol{\phi}$ (for a proof see Appendix A, Lemma 4). Recombination does not affect M directly. Thus, for fixed μ, the increase in mean fitness at each generation is determined by the variance. The difference between the sex and asex processes will be seen to stem from the effect of recombination on variance, which then results in an increase to the change in mean fitness for the sex process during the selection phase.

The effect of mutation on the variance is a fixed increase at each generation, again entirely determined by μ. The effect of selection on variance (Appendix A, Lemma 4) is given by:

$$V_F(\mathtt{Sel}(\boldsymbol{\phi})) - V_F(\boldsymbol{\phi}) = \frac{\kappa_3}{M} - \left(\frac{V_F}{M}\right)^2,$$

where κ_3 is the third cumulant of $F(X)$ – recall that (roughly speaking) the third cumulant can be seen as a measure of asymmetry in the distribution, with negative values indicating a longer left tail. Our first theorem, proved in Appendix A, shows that for the sex process, the effect of recombination on variance is positive.

Theorem 1. *If $\boldsymbol{\phi}^* = \mathtt{Sel}(\boldsymbol{\phi})$ was obtained by an application of selection to a population $\boldsymbol{\phi}$ at linkage equilibrium, then the effect of recombination on fitness variance is given by:*

$$V_F(\mathtt{Rec}(\boldsymbol{\phi}^*)) - V_F(\boldsymbol{\phi}^*) = \frac{\sum_{i \neq j} V_i V_j}{M^2},$$

where $V_i = \mathrm{Var}(\phi_i)$ and $M = M(\boldsymbol{\phi})$. This effect is therefore non-negative.

This theorem applies to the sex process because a previous application of recombination would bring the population $\boldsymbol{\phi}$ to linkage equilibrium. Linkage equilibrium is then preserved by mutation (see Lemma 5, Appendix A and surrounding comments).

While Theorem 1 describes the positive effect of recombination during the sex process, our second theorem (which requires considerably more work to prove) shows that recombination has a positive effect on variance in a much more general situation, as for instance, during a process which is asexual up until a given generation at which recombination occurs. This theorem establishes that for a population initially at linkage equilibrium, *any* subsequent applications of recombination during later generations always give an increase in variance and so a corresponding increase in the rate of change of mean fitness.

Theorem 2. *For the $\mathbb{Z}$-model, starting with a population at linkage equilibrium, suppose we iterate the operations of mutation, selection and recombination in any order (possibly applying only mutation and selection over multiple generations, and of course applying truncations when relevant). Then any non-trivial application of recombination has a positive effect on variance.*

By a *trivial* application of recombination we mean one acting on a population which is already at linkage equilibrium, and so which has no effect at all. This is the case, for instance, if one applies recombination twice in a row: the second application is trivial. The theorem is stated only for the $\mathbb{Z}$-model because truncation creates technical difficulties when producing a proof for the other models. With the effect of truncation being so small, however, the claim of the theorem is, in fact, verified in all simulations we have run for any of the models.

To explain what is behind Theorem 2, we need to consider two key terms: the linkage disequilibrium term LD_2 and the *flat variance*. We define $LD_2(\boldsymbol{\phi})$ to be the covariance term $\sum_{i\neq j} E(X_iX_j) - E(X_i)E(X_j)$, which is similar to the standard notion of *linkage disequilibrium coefficient*, but has terms weighted according to fitness values rather than just considering frequencies. The variance V_F can be expressed:

$$\begin{aligned} V_F(\boldsymbol{\phi}) &= E\Big(\big(\sum_{i=1}^{\ell} X_i\big)^2\Big) - \Big(E\big(\sum_{i=1}^{\ell} X_i\big)\Big)^2 \\ &= \sum_{i=1}^{\ell}\Big(E(X_i^2) - E(X_i)^2\Big) + \sum_{i\neq j}\Big(E(X_iX_j) - E(X_i)E(X_j)\Big). \\ &= \Big(\sum_{i=1}^{\ell} V_i\Big) + \sum_{i\neq j}\Big(E(X_iX_j) - E(X_i)E(X_j)\Big). \end{aligned}$$

We therefore have:

$$LD_2 = V_F - \sum_{i=1}^{\ell} V_i.$$

Since $\mathtt{Rec}(\boldsymbol{\phi})$ is at linkage equilibrium, and recombination leaves each ϕ_i unchanged, we conclude that $V_F(\mathtt{Rec}(\boldsymbol{\phi})) = \sum_i V_i$, where $V_i = \mathrm{Var}(\phi_i)$. Thus $LD_2(\boldsymbol{\phi})$ is precisely the decrease

in variance produced by recombination:

$$LD_2(\boldsymbol{\phi}) = V_F(\boldsymbol{\phi}) - V_F(\texttt{Rec}(\boldsymbol{\phi})).$$

So LD_2 is the decrease in the second central moment produced by recombination, and of course one could define (more complicated) analogous terms for higher moments or cumulants. Theorem 2 thus states that $LD_2(\boldsymbol{\phi})$ is negative at all stages of the process, unless the population is at linkage equilibrium, in which case $LD_2(\boldsymbol{\phi}) = 0$.

A more geometric way of understanding LD_2 is through the notion of flat variance. Let $\boldsymbol{M} = (E(X_1), E(X_2), ..., E(X_\ell)) \in \mathbb{R}^\ell$; this vector represents the average individual in the population. The term $GV(\boldsymbol{\phi}) = E(\|X - \boldsymbol{M}\|^2)$ is then the sum of the variances contributed by individual loci. Of course, recombination does not affect $GV(\boldsymbol{\phi})$ at all, but rather changes the shape of the population by increasing the variance in the direction that is useful for selection, namely $V_F(\boldsymbol{\phi})$. Consider the diagonal line $d = \{(x_1, ..., x_\ell) \in \mathbb{R}^\ell : x_1 = x_2 = \cdots = x_\ell\}$ and its $(\ell - 1)$-dimensional orthogonal complement $P = \{(x_1, ..., x_\ell) \in \mathbb{R}^\ell : x_1 + x_2 + \cdots + x_\ell = 0\}$, and let π_d and π_P be the projection functions onto d and P respectively. Using that $F(X)$ is the inner product of X and $(1, 1, ..., 1)$, one can show that:

$$V_F(\boldsymbol{\phi}) = \ell \cdot \mathrm{Var}(\|\pi_d(X)\|).$$

We define the *flat variance* of a population to be the variance of its projection onto P multiplied by a correcting factor:

$$FV(\boldsymbol{\phi}) = \frac{\ell}{\ell - 1} \cdot E(\|\pi_P(X - \boldsymbol{M})\|^2).$$

Informally, $V_F(\boldsymbol{\phi})$ measures how *tall* a population is along the vector $(1, 1, ...1)$, while $FV(\boldsymbol{\phi})$ measures how *fat* it is. Unlike the other standard terms considered here, the flat variance does not seem to have an exact counterpart in the existing literature, but can be useful in providing a clear way in which to visualise and understand the effect of selection on variance. For realistic values of ℓ, the flat variance may be very close to $GV(\boldsymbol{\phi})$, but with the qualitative difference that it is affected by selection in a manner which allows us to view this effect in terms of *conversion* from one form of variance to another.

The effect of recombination on variance and flat variance then satisfies a simple formula, which can be derived as follows. Using that $\pi_P(X) + \pi_d(X) = X$ and that, by Pythagoras, $\|\pi_P(X - \boldsymbol{M})\|^2 + \|\pi_d(X - \boldsymbol{M})\|^2 = \|X - \boldsymbol{M}\|^2$, we get:

$$GV = \sum_{i=1}^{\ell} V_i = E(\|X - \boldsymbol{M}\|^2) = \\ E(\|\pi_P(X - \boldsymbol{M})\|^2) + E(\|\pi_d(X - \boldsymbol{M})\|^2) = \frac{\ell - 1}{\ell} FV + \frac{1}{\ell} V_F.$$

Thus, since $\sum_{i=1}^{\ell} V_i$ is unaffected by recombination, so is $((\ell - 1)FV + V_F)/\ell$. We can also deduce that if $\boldsymbol{\phi}^*$ is at linkage equilibrium and $V_F^* = \sum_{i=1}^{\ell} V_i^*$, then $FV^* = V_F^*$. It follows that the effect of recombination on V_F and FV is to make them equal while leaving $((\ell - 1)FV +$

$V_F)/\ell$ unchanged, thus making them both equal to $((\ell-1)FV + V_F)/\ell$. We then have:

$$V_F(\mathtt{Rec}(\boldsymbol{\phi})) - V_F(\boldsymbol{\phi}) = \frac{\ell-1}{\ell}(FV - V_F) \quad \text{and} \quad FV(\mathtt{Rec}(\boldsymbol{\phi})) - FV(\boldsymbol{\phi}) = \frac{1}{\ell}(V_F - FV),$$

and

$$LD_2 = ((\ell-1)/\ell)(V_F - FV).$$

Thus, LD_2 being negative is equivalent to FV being greater than V_F, or, more informally, the population being fatter than it is tall along the diagonal d. The dynamics of this interaction are explained in Figure 2, and the effects for unbounded and bounded domains are illustrated in Figures 3 and 4 respectively. Proving Theorem 2 then amounts to using this framework in order to understand the evolving shape of sexual and asexual populations over multiple generations. While the technical details are fairly involved (and are described in Appendix B), the basic intuition can be described quite simply. At a particular locus, selection is stronger among individuals whose *other* genes have lower fitness values - the lower fitness values at other loci increase the significance of differing fitnesses between alleles at the locus in question. By analysing the effect of each of the basic operations of selection, mutation, recombination and truncation, we are able to establish structural conditions on the shape of the distribution defining a population, which ensure that in the absence of recombination and over multiple generations, a negative linkage disequilibrium will be created and maintained. Under appropriate conditions, selection increases the value $FV(\boldsymbol{\phi}) - V_F(\boldsymbol{\phi})$, and the combined effects of mutation and truncation are to preserve this imbalance, as well as structural conditions required in order to ensure that subsequent applications of selection preserve negative LD_2.

Theorems 1 and 2 show an important advantage that sex has over asex. In comparing sex and asex populations evolving independently, however, these theorems do not suffice to entirely specify how the variances of the two populations differ at any given generation. To make this comparison we would need to understand the evolution of the third cumulant, which behaves differently in each process. The evolution of the third cumulant depends on the fourth, which depends on the fifth, and so on.

Rather than analysing further the evolution of populations over time, we now study what happens to the sexual and asexual populations in the long term. We prove that, for the bounded model, whatever the initial populations are, sex outperforms asex in the long run.

We state the following theorem in terms of a mixed population containing both sexual and asexual individuals competing for resources. Thus the population distribution $\boldsymbol{\phi}$ now has domain $D \times \{\mathtt{s}, \mathtt{a}\}$, the second coordinate indicating whether the individual is sexual or asexual. Mutation acts exactly as before among each type of individual. Selection is also the same, now using $M(\boldsymbol{\phi}) = \sum_{\boldsymbol{x} \in D \times \{\mathtt{s},\mathtt{a}\}} F(\boldsymbol{x})\boldsymbol{\phi}(\boldsymbol{x})$ to normalise. Recombination acts only among the sexual individuals.

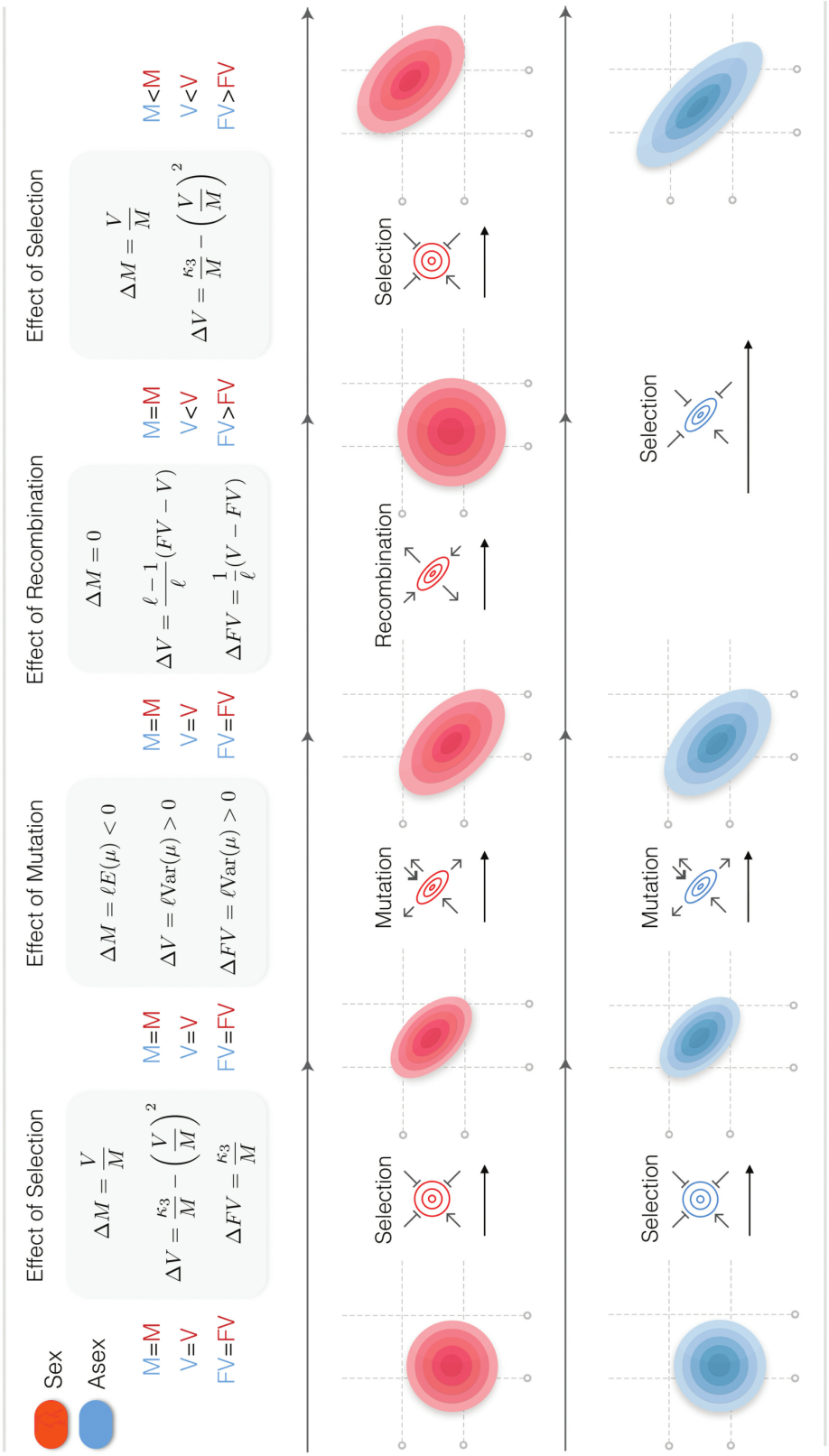

FIGURE 2. This illustration shows the level curves for 2-locus sex (red) and asex (blue) populations which begin with the same Gaussian type distribution (loci distributions are i.i.d. with $\kappa_3 = 0$). The x and y axes indicate fitnesses at the first and second loci respectively, fitness increasing along the up-right diagonal. The reason the last application of selection gives a greater increase in mean fitness for sex is that, *as opposed to asex, the sex process capitalises on the increase in flat variance due to mutation*. The first phase is selection which decreases the fitness variance (given $\kappa_3 = 0$), and does not interact with flat variance since fitness (measured along the $\diagup$-diagonal) is the only factor influencing the ability of an individual $\boldsymbol{x}$ to survive selection, while flat variance is measured along the planes where fitness is constant (the $\diagdown$-diagonal). Selection thus causes the flat variance to be greater than the variance, or equivalently, causes negative LD_2, as we show in Theorem 2. Mutation then increases both flat variance and variance by the same amount. Recombination, only occurring in the sex process, averages out the fitness variance and flat variance, increasing the variance and decreasing the flat variance, as seen by the rounded form of the level curves at that phase in the figure. Notice that recombination does not increase $GV(\phi)$; it just transforms the flat variance, which is useless for selection, into fitness variance. Finally, selection now has an increased effect on the mean fitness of the sexual population due to the larger fitness variance produced by recombination. For the asex population, mutation will keep on increasing the flat variance, but, in the absence of recombination, selection will not capitalise on this growth.

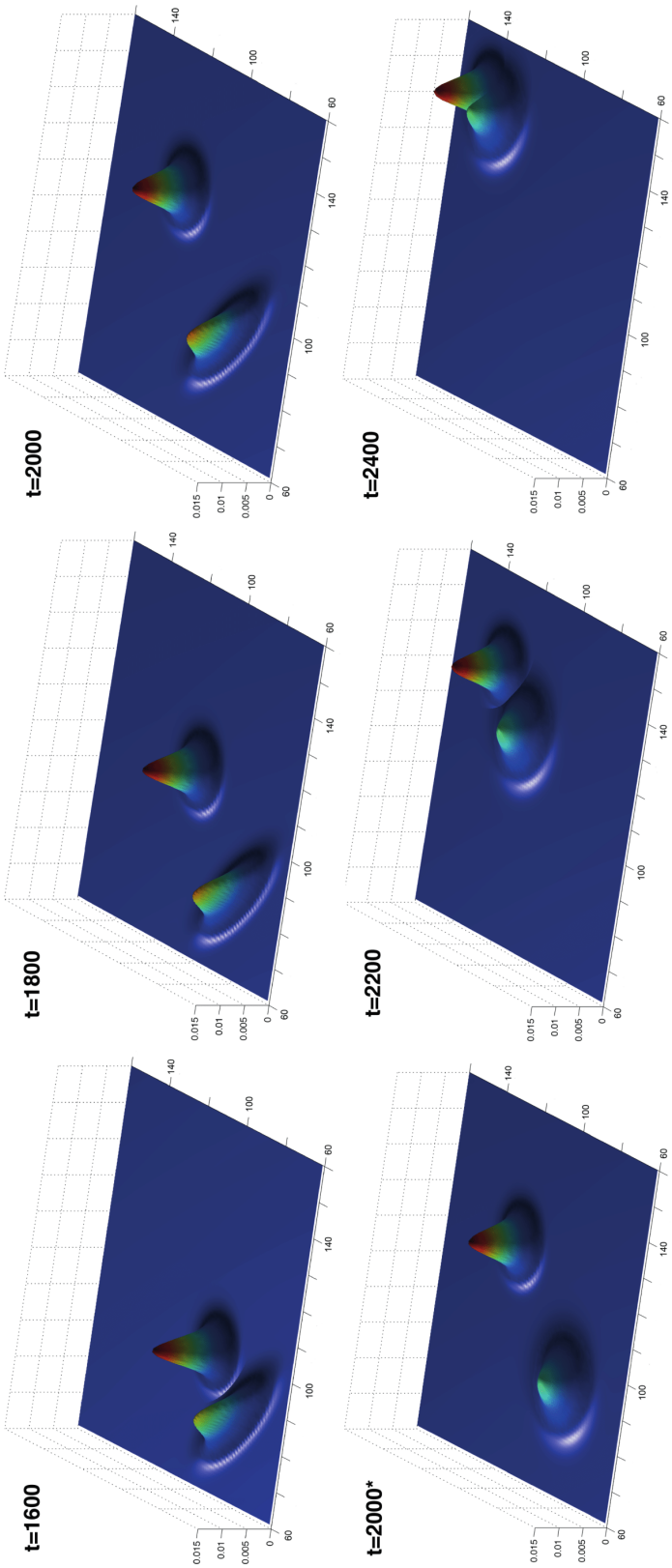

FIGURE 3. The benefit of recombination. This figure shows the progress of two 2-locus populations (for the $\mathbb{N}$ model), one of which is sexual and the other of which is *initially* asexual. The x and y axes correspond to fitness contributions at the first and second loci respectively, the z axis corresponding to probability density. Both populations begin with initial allele fitness contributions of 5, and progress with mutation rate 10^{-4}, the probability any given mutation is beneficial being 0.1. During the first 2000 generations the sexual population quickly achieves greater mean fitness, and one can clearly see the increased flat variance of the asexual population. At stage 2000, a single application of recombination is made to the previously asexual population, converting that flat variance into $V_F(\phi)$ and greatly increasing the rate of increase in mean fitness. Without any subsequent applications of recombination, however, the asexual population will eventually have smaller variance than the sexual one and will once again fall behind in mean fitness.

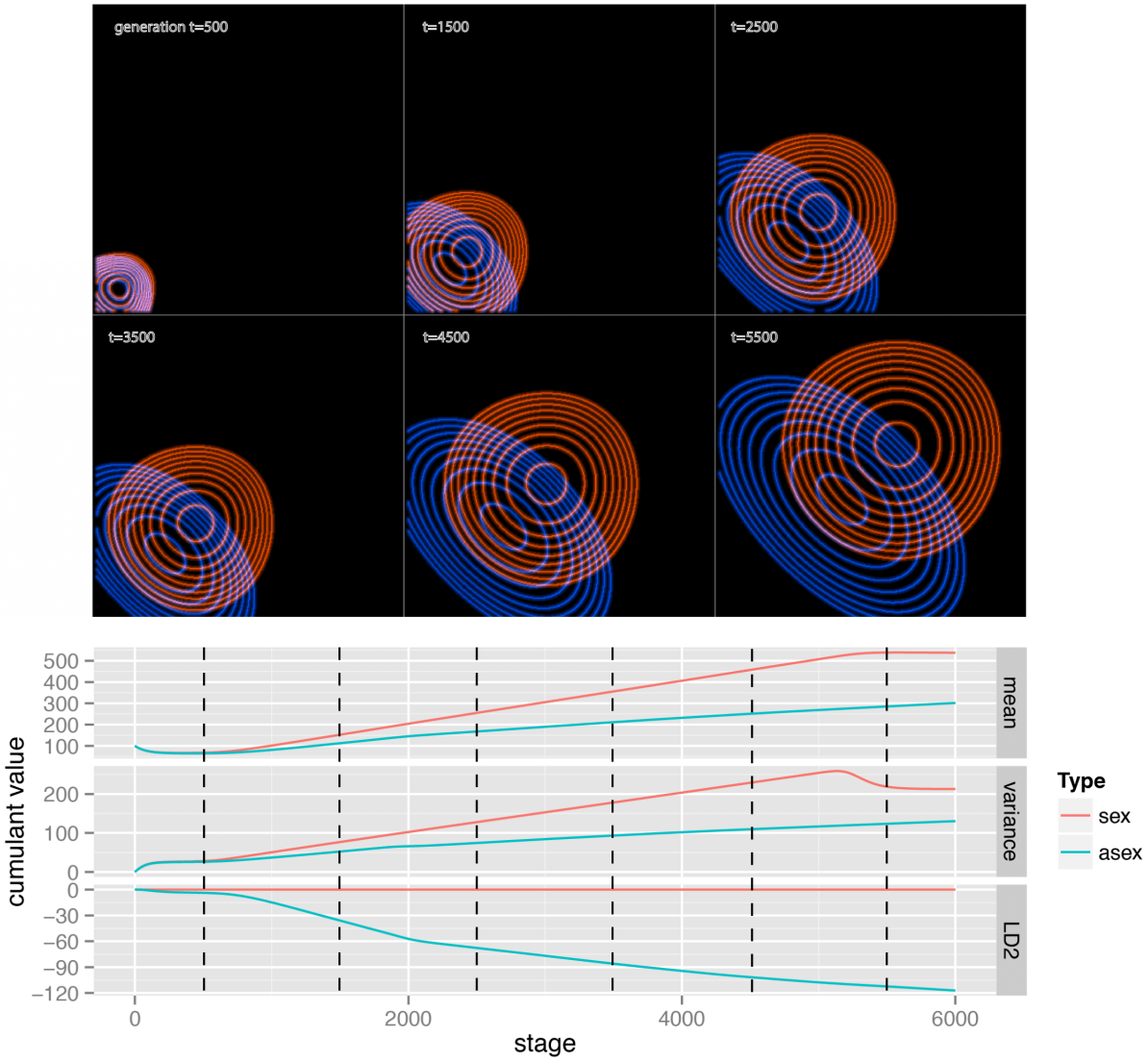

FIGURE 4. This figure shows the level curves for 2-locus populations proceeding according to the bounded model, with maximum allele fitness 400, mutation rate 0.2, and with the probability any given mutation is beneficial being 10^{-4}. All alleles initially have fitness 50. The probability density level curves are depicted at stages 500, 1500, 2500, 3500, 4500 and 5500. We can again observe the increase in flat variance and decrease in $V_F(\boldsymbol{\phi})$ for the asexual population, and also that the sexual population does not necessarily have a higher value $GV(\boldsymbol{\phi})$.

Theorem 3. *Given μ and ℓ, for all sufficiently large bounds N and for any initial population in which the proportion of sexual individuals is non-zero, the proportion of sexual individuals converges to 1 and the proportion of asexual ones converges to 0.*

This is the longest and most complicated proof of the paper: the proof appears in Appendix C. Standard techniques involving the application of the Perron-Frobenius Theorem

suffice to prove that the asex distribution (i.e. the distribution within the asexual part of the population in isolation) converges to a limit, since mutation and selection can be treated as linear operations with deferred normalisation incorporated appropriately into the analysis. New techniques are required, however, in order to establish a good approximation to the mean fitness at that limit. The key idea is to establish a translation between any given process and one in which negative and positive mutations are equally likely. Limit distributions are easily understood for processes of the latter kind, and so the established translation then allows us to describe the mean fitness of asex limit distributions more generally. Unfortunately, the nonlinear nature of recombination means that analysing the sex process is much more difficult. We do not prove that the sex process converges to a limit, but still get a good estimate of the geometrical average of the mean fitness over generations. Such ideas would not work for the $\mathbb{N}$- and $\mathbb{Z}$-models as in those cases the mean fitness diverges to infinity in both the sex and the asex processes. Figure 5 shows the manner in which sexual and asexual populations converge to their respective fixed points over time (while we do not prove that convergence to a fixed point always occurs for sexual populations, such convergence was observed in all simulations).

4. Discussion

In nature one must surely expect a variety of mechanisms to be of significance in determining the most efficient methods of reproduction. As well as those factors already mentioned, sex may provide advantages for species not subject to random mating by strengthening selection [29], for example, or may provide a straightforward advantage in providing two parents to care for young offspring [30]. Such arguments, however, do not suffice to explain the prevalence of sex in species for which random mating is a good approximation or without parental care. Our aim here has been to contribute to our understanding of a fundamental and underlying mechanism conferring strong advantages to sex, whereby the effect of recombination is to break down negative associations between loci. Of course a natural question, having considered the infinite populations case, is the extent to which this analysis carries over to the finite populations model. A crucial difference in moving to finite populations is that the process is no longer deterministic. The equations governing the change in mean fitness and variance due to selection and mutation for the infinite population model would now perfectly describe the *expected* effect of mutation and selection for finite populations, and the finite populations model could be seen simply as a stochastic approximation to the infinite case, were it not for the loss in variance and higher cumulants due to sampling (since picking n individuals from a distribution with variance v produces a population with expected variance $v(n-1)/n$). For sufficiently large populations this effect will be small on a stage by stage basis, and so our analysis for infinite populations can be seen as a good approximation over a number of generations which is not too large. Ultimately, however, sampling will have the effect that mean fitness for the population no longer increases without limit: once variance is sufficiently large the expected loss in variance due to sampling balances the increase that one would see for an infinite population with the same cumulants. Larger populations are thus able to

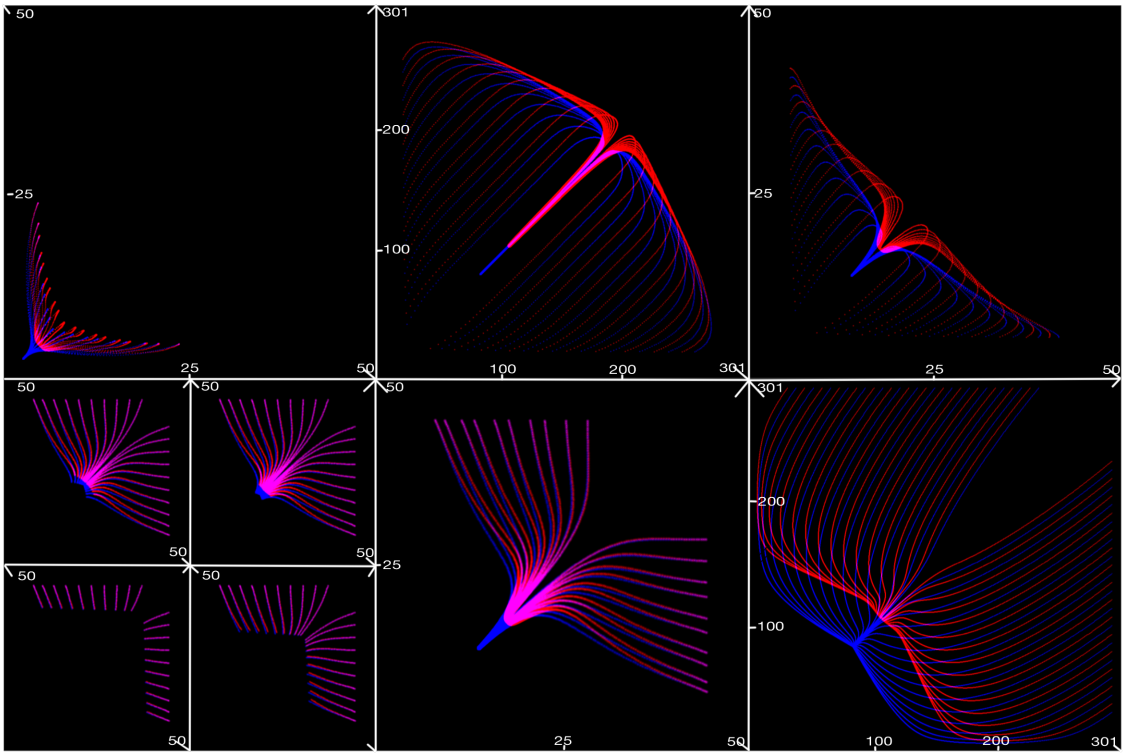

FIGURE 5. Each of the six plots shows the trajectory of the centre of mass for various sexual and asexual 2-locus populations over multiple generations, for a number of different initial populations and for the bounded model. Each point represents the centre of mass of a population at a single generation, and the populations were then allowed to evolve for sufficiently many generations that an equilibrium point was reached. In each case one can see convergence to an equilibrium, which corresponds to a higher fitness value for the sexual population. The bottom-left plot shows intermediate steps in the evolution towards the middle plot in the bottom row. For that plot, we have 40 different initial populations, half sexual (red), half asexual (blue). The bound, N, on gene fitness is 50 for all plots except for the top-centre and bottom-right, where $N = 301$. The probability of mutation is 0.5 except for the top-left plot, where the probability of mutation is 0.9. The probability that a mutation is beneficial is 0.001 in all cases. Starting from the top-left and moving clockwise, the original populations are Gaussian distributions with standard deviations 5, 25, 6, 8, 6 and 6 respectively.

sustain much higher mean fitnesses than small ones. As mentioned previously, there are also mechanisms which are only of relevance in the finite population setting, and which are believed to be a significant factor in the creation of negative linkage equilibrium [19].

While sexual reproduction has been seen here to confer strong advantages in the setting of simplistic and entirely modular fitness landscapes, we have said nothing about how this picture changes in the presence of complex epistasis. Assuredly, the task of efficiently navigating fitness landscapes (i.e. optimisation) is one that, beyond its relevance here, is of fundamental significance across large areas of applied mathematics and computer science. However large the role of epistasis in the biological context, it is certainly true that in most of these applications complicated forms of epistasis (in one guise or another) play a crucial role, and so the interesting question becomes that as to whether sexual reproduction continues to offer these substantial benefits in the face of more complex fitness landscapes. It may be the case that as well as capitalising more efficiently on existing modularity, sex plays a fundamental role in *finding* modularity[32]. One would expect a proper analysis to require classification of fitness landscapes in terms of their amenability to different forms of population based search (see, for example, the work of Prugel-Bennet[33]).

References

[1] Felsenstein J (1974) The evolutionary advantage of recombination. *Genetics* 78: 737-756.

[2] Maynard-Smith J (1978) *The evolution of sex*. Cambridge University Press, Cambridge.

[3] Kondrashov AS (1993) Classification of hypotheses on the advantage of amphimixis. *J. Hered.* 84: 372-387.

[4] Bernstein C, Bernstein H (1991) *Aging, sex and DNA repair*. San Diego Academic Press.

[5] Michod RE (1993) Genetic error, sex and diploidy. *J. Hered.* 84: 360-371.

[6] Maynard-Smith J (1988) The evolution of recombination. In *The evolution of sex: an examination of current ideas*, Michod, R.E. & Levin, B.R. eds. Sunderland, Massachusetts, Sinauer: 106-125.

[7] Barton NH, Charlesworth B (1998) Why sex and recombination? *Science* 281: 1986–1990.

[8] Barton NH (1995) Linkage and the limits to natural selection. *Genetics* 140: 821–841.

[9] Otto SP, Barton NH (1997) The evolution of recombination: removing the limits to natural selection. *Genetics* 147:879-906.

[10] Hill WG, Robertson A (1966) The effect of linkage on the limits to artificial selection. *Genetical Research* 8:269-294.

[11] Morgan TH (1913) *Heredity and sex*. Columbia University Press, New York.

[12] Fisher RA (1930) *The genetical theory of natural selection*. Oxford University Press, Oxford.

[13] Muller HJ (1932) Some genetic aspects of sex. *American Naturalist* 66:118-138.

[14] Muller HJ (1964) The relation of recombination to mutational advance. *Mutation Research* 1:2-9.

[15] Mather K (1943). Polygenic inheritance and natural selection, *Biological Reviews*, 18: 32–64.

[16] Felselstein J (1965) The effect of linkage on directional selection, *Genetics* 42: 349-363.

[17] Charlesworth B (1993) Directional selection and the evolution of sex and recombination, *Genet. Res. Camb*, 61: 205–224.

[18] Barton NH (1995), A general model for the evolution of recombination, *Genet. Res. Camb*, 65: 123–144.

[19] Barton NH, Otto, SP (2005) Evolution of recombination due to random drift. *Genetics* 169:2353-2370.

[20] Salathé M, Kouyos RD, Bonhoeffer S (2009) On the causes of selection for recombination underlying the Red Queen hypothesis. *American Naturalist* 174(suppl.):S31-S42.

[21] Peters AD, Lively CM (1999) The Red Queen and fluctuating epistasis: a population genetic analysis of antagonistic coevolution. *American Naturalist* 154:393-405.

[22] Gandon S, Otto SP (2007) The evolution of sex and recombination in response to abiotic or coevolutionary fluctuations in epistasis. *Genetics* 175:1835-1853.

[23] Pylkov KV, Zhivotovsky LA, Feldman MW (1998) Migration versus mutation in the evolution of recombination under multi-locus selection. *Genetical Research* 71:247-256.

[24] Lenormand T, Otto SP (2000) The evolution of recombination in a heterogeneous environment. *Genetics* 156:423-438.

[25] Otto SP (2009) The evolutionary enigma of sex. *American Naturalist*, 174: S1-S14.

[26] Ohta T, Kimura M (1973) A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. *Genet. Res.* 22, 201-204.

[27] Moran PAP (1976) Global stability of genetic systems governed by mutation and selection, *Math. Proc. Camb. Phil. Soc.*, 80, 331-336.

[28] Moran PAP (1977) Global stability of genetic systems governed by mutation and selection II, *Math. Proc. Camb. Phil. Soc.*, 81, 435-441.

[29] Geodakyan VA (1965) The role of sexes in the transmission and transformation of genetic information, *Probl. Peredachi Inf.*, 1, 105-112, (In Russian).

[30] Lloyd DL (1980) Benefits and handicaps of sexual reproduction, *Evol Biol*, 13, 69-111.

[31] Livnat A, Papadimitriou C, Dushoff J, Feldman MW (2008) A mixability theory for the role of sex in evolution, *PNAS*, vol. 105, no. 50, 19803-19808.

[32] Livnat A, Papadimitriou C, Pippenger N, Feldman MW (2009) Sex, mixability and modularity, *PNAS*, vol. 107, no. 4, 1452-1457.

[33] Prügel-Bennett A (2004) When a genetic algorithm outperforms hill-climbing, *Theoretical Computer Science*, 320, (1), 135-153.

[34] Maynard-Smith J (1968) Evolution in sexual and asexual populations, *American Naturalist*, 102, 469-678.

[35] Crow JF, Kimura M (1964) The theory of genetic loads, *Proc XI Int. Congr. Genetics.*, vol 2, 495-505.

5. Appendix A

5.1. **The structure of the proofs.** In Appendices A, B and C we provide a much deeper analysis of all the claims made in the main article. We start, here in Appendix A, by proving Theorem 1 and showing how the various operations affect different population properties such as mean and variance. Appendix B is dedicated to the proof of Theorem 2. Although Theorem 1 and Theorem 2 are both concerned with establishing negative values for LD_2, their proofs are completely different and give us alternative ways of understanding the model. Theorem 3 is proved in Appendix C. The proof of this theorem is much longer than those of the previous theorems, and, again, it is very different in style. Once we have proved our main theorems, we move on to discuss variants of our model in Appendix D. The variants we consider are the finite version and the multiplicative version. We do not have a full mathematical analysis for those models, but present the results of simulations. The outcomes of simulations are presented in Appendix E.

5.2. **The evolution of the key values.** In this subsection we review some well known facts and describe in more detail how mutation, selection and recombination affect mean fitness, variance, LD_2 and flat variance, proving the claims made in Section 3. The objective is to establish all of the results in the Table 1 below.

5.2.1. *The evolution of mean fitness and variance.* The impact of mutation on the mean, variance and all cumulants is simply described (recall that mean fitness and variance are the

Effect of:	selection	mutation	recombination
ΔM	V_F/M	$\ell\ E(\mu)$	0
ΔV_F	$\kappa_3/M - (V_F/M)^2$	$\ell\ \mathrm{Var}(\mu)$	$-LD_2$
ΔLD_2	$-\sum_{i\neq j} V_iV_j/M^2 \qquad (*)$	0	$-LD_2$
$\Delta\kappa_3$	$(V_F/M)((\kappa_4/V_F) - 3(\kappa_3/M) + 2(V_F/M)^2)$	$\ell\ \kappa_3(\mu)$	$-LD_3$

TABLE 1. By ΔM is meant the change in M produced by the relevant operation. All values (M, V_F, etc) inside the table are with respect to the population before the relevant operation is applied: the box stating that ΔM for selection is V_F/M should be read $M(\mathtt{Sel}(\boldsymbol{\phi})) - M(\boldsymbol{\phi}) = V_F(\boldsymbol{\phi})/M(\boldsymbol{\phi})$. $(*)$ The stated effect of selection on LD_2 is only valid in the case that selection is acting on a population at linkage equilibrium.

first and second cumulants of $F(X)$). If Y and C are independent random variables and κ_n is the nth cumulant, then $\kappa_n(Y+C) = \kappa_n(Y) + \kappa_n(C)$. Thus the effect of mutation on the mean fitness is to increase it by $\ell E(\mu)$ (which will be negative given our assumptions on μ). Similarly, the effect on variance is to increase it by $\ell\,\mathrm{Var}(\mu)$.

The effect of selection is given by the following lemma (while these claims are either well known or easily established, we include a proof for the sake of completeness). Here $M = M(\boldsymbol{\phi})$, $V_F = V_F(\boldsymbol{\phi})$, $\kappa_3 = \kappa_3(\boldsymbol{\phi})$, $\kappa_4 = \kappa_4(\boldsymbol{\phi})$ and M^*, V_F^*, κ_3^* are the corresponding values for $\boldsymbol{\phi}^* = \mathtt{Sel}(\boldsymbol{\phi})$.

Lemma 4. *The effect of selection on mean fitness, variance and κ_3 is given by:*

$$\begin{aligned} M^* - M &= V_F/M, \\ V_F^* - V_F &= \kappa_3/M - (V_F/M)^2, \\ \kappa_3^* - \kappa_3 &= (V_F/M)((\kappa_4/V_F) - 3(\kappa_3/M) + 2(V_F/M)^2). \end{aligned}$$

Proof. We prove the first two identities. The third then follows with a little more algebraic manipulation, by almost identical methods. In order to see the first identity, note that:

$$M^* = \sum_{\boldsymbol{x}} F(\boldsymbol{x})\ \mathtt{Sel}(\boldsymbol{\phi})(\boldsymbol{x}) = \frac{1}{M}\sum_{\boldsymbol{x}} F(\boldsymbol{x})^2\boldsymbol{\phi}(\boldsymbol{x}).$$

Now, using that the second moment about the origin, $\sum_{\boldsymbol{x}} F(\boldsymbol{x})^2\boldsymbol{\phi}(\boldsymbol{x})$, is equal to $V_F + M^2$ we get:

$$V_F = \Big(\sum_{\boldsymbol{x}} F(\boldsymbol{x})^2\boldsymbol{\phi}(\boldsymbol{x})\Big)\ - M^2 = M^*M - M^2.$$

This gives the identity $V_F/M = M^* - M$, as required. In order to derive the second identity, we recall the formula for the third central moment:

$$\begin{aligned}
\kappa_3 &= \sum_{\boldsymbol{x}} (F(\boldsymbol{x}) - M)^3 \boldsymbol{\phi}(\boldsymbol{x}) \\
&= \sum_{\boldsymbol{x}} (F(\boldsymbol{x})^3 - 3F(\boldsymbol{x})^2 M + 3F(\boldsymbol{x})M^2 - M^3)\ \boldsymbol{\phi}(\boldsymbol{x}) \\
&= \Big(\sum_{\boldsymbol{x}} F(\boldsymbol{x})^3 \boldsymbol{\phi}(\boldsymbol{x})\Big) - 3M\left(\Big(\sum_{\boldsymbol{x}} F(\boldsymbol{x})^2 \boldsymbol{\phi}(\boldsymbol{x})\Big) - M^2\right) - M^3 \\
&= \Big(\sum_{\boldsymbol{x}} F(\boldsymbol{x})^3 \boldsymbol{\phi}(\boldsymbol{x})\Big) - 3MV_F - M^3.
\end{aligned}$$

Then:

$$V_F^* = \Big(\sum_{\boldsymbol{x}} F(\boldsymbol{x})^2\ \boldsymbol{\phi}^*(\boldsymbol{x})\Big) - (M^*)^2 = \frac{1}{M}\Big(\sum_{\boldsymbol{x}} F(\boldsymbol{x})^3 \boldsymbol{\phi}(\boldsymbol{x})\Big)\ - (M^*)^2.$$

Substituting $V_F/M + M$ for M^*, we get:

$$V_F^* - V_F = \frac{1}{M}\Big(\sum_{\boldsymbol{x}} F(\boldsymbol{x})^3 \boldsymbol{\phi}(\boldsymbol{x})\Big) - \big(M^2 + 2V_F + V_F^2/M^2\big) - V_F = \kappa^3/M - V_F^2/M^2,$$

as required. □

Let us now consider recombination. Recall that X_i is a random variable taking values according to the distribution ϕ_i (specifying the distribution at the ith locus). At any point, the mean is given by $M(\boldsymbol{\phi}) = \sum_i E(\phi_i)$. Since recombination has no effect on each ϕ_i, it also has no impact on $M(\boldsymbol{\phi})$. As we showed previously, the change in variance due to recombination is $-LD_2$.

5.2.2. *The evolution of LD_2.* The most direct way in which recombination affects mean fitness is by changing the variance, which then affects the growth in mean fitness via selection. The change in variance due to recombination is given by $-LD_2$. Thus, to show that recombination has a positive effect on variance, one must show that LD_2 is negative. In this subsection we analyse the effect on LD_2 given by the different operations. As part of our analysis we get a proof of Theorem 1.

Since $LD_2 = 0$ when at linkage equilibrium, we have $LD_2(\texttt{Rec}(\boldsymbol{\phi})) = 0$.

Mutation has no effect at all on LD_2 as shown by the following lemma.

Lemma 5. *For any population $\boldsymbol{\phi}$, $LD_2(\texttt{Mut}(\boldsymbol{\phi})) = LD_2(\boldsymbol{\phi})$.*

Proof. Recall the definition of mutation in terms of the random variables C_i.

$$\begin{aligned}
LD_2(\texttt{Mut}(\boldsymbol{\phi})) &= \sum_{i \neq j} \Big(E((X_i + C_i)(X_j + C_j)) - E(X_i + C_i)E(X_j + C_j)\Big) \\
&= \sum_{i \neq j} \Big(E(X_iX_j) + E(X_iC_j) + E(C_iX_j) + E(C_iC_j) \\
&\qquad -E(X_i)E(X_j) - E(X_i)E(C_j) - E(C_i)E(X_j) - E(C_i)E(C_j)\Big).
\end{aligned}$$

Since C_i and C_j are independent, and are independent of X_i and X_j, most of these terms cancel, leaving $E(X_iX_j) - E(X_i)E(X_j)$ as required. □

So mutation has no effect at all on linkage equilibrium: This is because if the variables X_i are independent, so are the variables $X_i + C_i$. Also, since $FV = V_F - (\ell/(\ell-1))LD_2$, we conclude that the effect of mutation on flat variance is the same as that on variance: $FV(\texttt{Mut}(\boldsymbol{\phi})) - FV(\boldsymbol{\phi}) = \ell \operatorname{Var}(\mu)$.

The effect of selection on LD_2 is more complex and is given by Theorem 1 (restated below) in the case that the operation is applied to a population at linkage equilibrium. The rest of the subsection is dedicated to proving it. We define LD_3 to be the decrease in the third cumulant of $F(X)$ produced by recombination. Thus,

$$LD_3(\boldsymbol{\phi}) = \kappa_3(F(X)) - \sum_{i=1}^{\ell} \kappa_3(X_i).$$

As with the other values, we use κ_3 to denote $\kappa_3(F(X))$ and $\kappa_{3,i}$ to denote $\kappa_3(X_i)$.

Theorem 1. *If $\boldsymbol{\phi}^* = \texttt{Sel}(\boldsymbol{\phi})$ was obtained by an application of selection to a population $\boldsymbol{\phi}$ at linkage equilibrium, then the effect of recombination on variance is given by:*

$$V_F(\texttt{Rec}(\boldsymbol{\phi}^*)) - V_F(\boldsymbol{\phi}^*) = \frac{\sum_{i\neq j} V_iV_j}{M^2},$$

where $V_i = \operatorname{Var}(\phi_i)$ and $M = M(\boldsymbol{\phi})$.

Theorem 1 asserts, in other words, that $LD_2(\texttt{Sel}(\boldsymbol{\phi})) = -(\sum_{i\neq j} V_iV_j)/M^2$ if $\boldsymbol{\phi}$ is at linkage equilibrium. The key to the proof is to study the effect of selection on each locus separately, as given by the following lemma. Let us describe our notation. Let $\boldsymbol{\phi}^* = \texttt{Sel}(\boldsymbol{\phi})$. Recall that we use a boldface greek character, $\boldsymbol{\phi}$, to denote the distribution of a population in $\mathbb{Z}^\ell$, and the lightface version of that character, ϕ_i to denote the distribution at the ith locus. We denote the mean fitness at locus i by $W_i = E(\phi_i)$. By the linearity of expectation we have $M = \sum_{i=1}^{\ell} W_i$. We use $\hat{W}_i$ to denote the mean fitness of the loci other than i, i.e., $\hat{W}_i = M - W_i$. Use use V_i to denote the variance in fitness at the ith locus: $V_i = \operatorname{Var}(\phi_i)$. The notation is analogous for $\boldsymbol{\phi}^*$: $W_i^* = E(\phi_i^*)$, $V_i^* = \operatorname{Var}(\phi_i^*)$, etc.

Lemma 6. *If selection acts on a population at linkage equilibrium, the effect on the ith locus is given by:*

$$\phi_i^*(x) = \frac{1}{M}\left(x + \hat{W}_i\right)\phi_i(x).$$

Proof. First, let us observe that $E(F(X) \mid X_i = x) = x + \hat{W}_i$:

$$E(F(X) \mid X_i = x) \;=\; \sum_{j=1}^{\ell} E(X_j | X_i = x) \;=\; x + \sum_{j\neq i} E(X_j) \;=\; x + \sum_{j\neq i} W_j.$$

For $x \in \mathbb{Z}$ and $\boldsymbol{y} \in \mathbb{Z}^{\ell-1}$ let $x \,\widehat{}_i \boldsymbol{y}$ be the vector of length ℓ with x as the i-coordinate and with all other coordinates given by $\boldsymbol{y}$ in corresponding order. Second, we calculate $\phi_i^*(x)$:

$$\begin{aligned}\phi_i^*(x) &= \sum_{\boldsymbol{y}\in\mathbb{Z}^{\ell-1}} \boldsymbol{\phi}^*(x \,\widehat{}_i \boldsymbol{y}) = (1/M) \sum_{\boldsymbol{y}\in\mathbb{Z}^{\ell-1}} F(x \,\widehat{}_i \boldsymbol{y})\boldsymbol{\phi}(x \,\widehat{}_i \boldsymbol{y}) \\ &= (1/M)\ \phi_i(x)\ E(F(X) \mid X_i = x).\end{aligned}$$

Putting these equations together, we get the result of the lemma. □

The next lemma shows the effect of selection on the fitness and variance at a single locus.

Lemma 7. *If selection acts on a population at linkage equilibrium, the effect on fitness and variance at locus i is given by:*

$$\begin{aligned}W_i^* - W_i &= \frac{V_i}{M}, \\ V_i^* - V_i &= \frac{\kappa_{3,i}}{M} - \left(\frac{V_i}{M}\right)^2.\end{aligned}$$

Proof. For the first equation:

$$\begin{aligned}W_i^* &= \sum_x x\phi_i^*(x) \\ &= (1/M)\ \sum_x x(x + \hat{W}_i)\phi_i(x) \\ &= (1/M)\ \left(\sum_x x^2\phi_i(x) + \hat{W}_i \sum_x x\phi_i(x)\right) \\ &= (1/M)\ \left((V_i + W_i^2) + \hat{W}_i W_i\right) \\ &= (1/M)\ (V_i + MW_i).\end{aligned}$$

This establishes the first equation of the lemma.

For the second equation, let $\tilde{V}_i$ be the *second moment about the origin of* ϕ_i, that is, $\tilde{V}_i = \sum_x x^2\phi_i(x)$, and analogously for ϕ_i^*. Let $\tilde{\kappa}_{3,i}$ be the *third moment about the origin of* ϕ_i, that is, $\tilde{\kappa}_{3,i} = \sum_x x^3\phi_i(x)$. Then

$$\begin{aligned}\tilde{V}^* &= \sum_x x^2\phi_i^*(x) \\ &= (1/M)\ \sum_x x^2\left(x + \hat{W}_i\right)\phi_i(x) \\ &= (1/M)\ \left(\sum_x x^3\phi_i(x) + \hat{W}_i \sum_x x^2\phi_i(x)\right) \\ &= (1/M)\ \left(\tilde{\kappa}_{3,i} + \hat{W}_i\tilde{V}_i\right).\end{aligned}$$

Now, using the developments of the moments about the origin in terms of the central moments we get:

$$\begin{aligned}\tilde{V}_i^* &= V_i^* + W_i^{*2} = V_i^* + W_i^2 + 2V_iW_i/M + (V_i/M)^2,\\ \tilde{\kappa}_{3,i} &= \kappa_{3,i} + 3V_iW_i + W_i^3,\\ \tilde{V}_i &= V_i + W_i^2.\end{aligned}$$

The equation above then becomes

$$V_i^* + W_i^2 + 2V_i\frac{W_i}{M} + \left(\frac{V_i}{M}\right)^2 = \frac{\kappa_{3,i}}{M} + 3V_i\frac{W_i}{M} + \frac{W_i^3}{M} + \frac{\hat{W}_i}{M}(V_i + W_i^2),$$

which we can re-arrange as

$$V_i^* + \left(2V_i\frac{W_i}{M} - 3V_i\frac{W_i}{M} - V_i\frac{\hat{W}_i}{M}\right) = \frac{\kappa_{3,i}}{M} - \left(\frac{V_i}{M}\right)^2 + \left(\frac{W_i^3}{M} + \frac{\hat{W}_i}{M}W_i^2 - W_i^2\right).$$

To finish the proof of the lemma one only has to observe that $\left(2V_i\frac{W_i}{M} - 3V_i\frac{W_i}{M} - V_i\frac{\hat{W}_i}{M}\right) = -V_i$ and that $\left(\frac{W_i^3}{M} + \frac{\hat{W}_i}{M}W_i^2 - W_i^2\right) = 0$. □

We now continue with the proof of Theorem 1. Using that $LD_2 = V_F - \sum_i V_i$, we get:

$$\begin{aligned}LD_2(\boldsymbol{\phi}^*) - LD_2(\boldsymbol{\phi}) &= (V_F^* - V_F) - (\sum_i V_i^* - V_i)\\ &= \left(\frac{\kappa_3}{M} - \left(\frac{V_F}{M}\right)^2\right) - \sum_i\left(\frac{\kappa_{3,i}}{M} - \left(\frac{V_i}{M}\right)^2\right)\\ &= \frac{LD_3}{M} + \frac{(\sum_i V_i^2) - V_F^2}{M^2}\\ &= -\frac{\sum_{i\neq j} V_iV_j}{M^2}\end{aligned}$$

The last equality follows since $LD_3 = 0$ for a population at linkage equilibrium.

6. Appendix B

6.1. **The ordering on distributions.** This subsection is dedicated to proving some basic combinatorial lemmas which are required for the proof of Theorem 2. Our new key notion is the ordering $\preceq$ among probability distributions on $\mathbb{Z}$, which will be useful throughout the rest of the paper. We made the assumption earlier that all cumulants of populations are finite. It is similarly to be assumed that all cumulants of distributions discussed in this section are finite.

Definition 8. *Given two distributions ψ_1 and $\psi_2 \colon \mathbb{Z} \to \mathbb{R}^{\geq 0}$, we define:*

$$\psi_2 \preceq \psi_1 \quad \iff \quad (\forall b_1 < b_2 \in \mathbb{Z}) \;\; \psi_1(b_1)\psi_2(b_2) \leq \psi_1(b_2)\psi_2(b_1).$$

We let $\psi_2 \prec \psi_1$ if, in addition, there exist $b_1 < b_2 \in \mathbb{Z}$ with $\psi_1(b_1)\psi_2(b_2) < \psi_1(b_2)\psi_2(b_1)$.

To give some intuition for the meaning of $\preceq$, let us remark that if ψ_1 and ψ_2 are non-zero on an interval $[A, B]$, and zero elsewhere, then:

$$\psi_2 \preceq \psi_1 \quad \iff \quad (\forall b \in \mathbb{Z} \text{ with } A \leq b < B) \;\; \frac{\psi_2(b+1)}{\psi_2(b)} \leq \frac{\psi_1(b+1)}{\psi_1(b)}.$$

If $\psi_2 \preceq \psi_1$, this gives a lot of information about the supports of ψ_1 and ψ_2 (i.e. those x for which $\psi_1(x) \neq 0$ or $\psi_2(x) \neq 0$). If x is in the support of ψ_1, then for any $y > x$ in the support of ψ_2, y must also be in the support of ψ_1. Similarly, if x is in the support of ψ_2, then for any $y < x$ in the support of ψ_1, y must also be in the support of ψ_2. We can therefore find disjoint (possibly empty) sets Π_1, Π_2 and Π_3 such that the support of ψ_2 is $\Pi_1 \cup \Pi_2$, the support of ψ_1 is $\Pi_2 \cup \Pi_3$, and all the elements of Π_1 are below all the elements of Π_2 which are all below all the elements of Π_3.

The main three properties of the ordering $\preceq$ are that it is preserved by mutation, it is preserved by selection, and it preserves the ordering of expected values. The proof of Theorem 2 will use all of these lemmas to show that LD_2 becomes and remains negative during an asex process initially at linkage equilibrium.

Lemma 9. *The orderings $\prec$ and $\preceq$ are preserved by mutation. That is:*

$$\psi_2 \preceq \psi_1 \quad \Rightarrow \quad \texttt{Mut}(\psi_2) \preceq \texttt{Mut}(\psi_1).$$

The same holds for $\prec$. Here `Mut` *refers to the mutation operation for $\ell = 1$.*

Proof. We must show that for any values $b_2 > b_1$:

$$(1) \qquad \sum_d \psi_2(d)\mu(b_2 - d) \cdot \sum_c \psi_1(c)\mu(b_1 - c) \leq \sum_d \psi_1(d)\mu(b_2 - d) \cdot \sum_c \psi_2(c)\mu(b_1 - c).$$

The r.h.s. can be re-expressed:

$$\sum_c \Big(\psi_1(c)\psi_2(c)\mu(b_2 - c)\mu(b_1 - c)) + \sum_{d>c} \big((\psi_1(d)\psi_2(c)\mu(b_2 - d)\mu(b_1 - c) + \psi_1(c)\psi_2(d)\mu(b_2 - c)\mu(b_1 - d)\big)\Big).$$

The l.h.s. is:

$$\sum_c \Big(\psi_2(c)\psi_1(c)\mu(b_2 - c)\mu(b_1 - c) + \sum_{d>c} \big((\psi_2(d)\psi_1(c)\mu(b_2 - d)\mu(b_1 - c) + \psi_2(c)\psi_1(d)\mu(b_2 - c)\mu(b_1 - d)\big)\Big).$$

For any given pair (d, c) such that $d > c$ define:

$$\alpha_1 = \psi_1(c)\psi_2(d), \;\; \alpha_2 = \psi_1(d)\psi_2(c),$$

$$\beta_1 = \mu(b_2 - c)\mu(b_1 - d), \;\; \beta_2 = \mu(b_2 - d)\mu(b_1 - c).$$

Now for any values $d > c$ we have $\alpha_2 \geq \alpha_1$ because $\psi_2 \preceq \psi_1$. We claim that we also have $\beta_2 \geq \beta_1$: this holds because in order to have $\beta_1 > 0$ one requires $b_2 \leq c + 1$ and $d \leq b_1 + 1$,

which can only be the case if $b_1 + 1 = c + 1 = b_2 = d$. In that case $\beta_1 = \mu(1)\mu(-1)$ and $\beta_2 = \mu(0)\mu(0)$, and it follows that $\beta_2 > \beta_1$ from our assumption that $\mu(0) > \mu(-1) > \mu(1)$. Thus:

$$\alpha_2\beta_2 + \alpha_1\beta_1 \geq \alpha_1\beta_2 + \alpha_2\beta_1.$$

This establishes the inequality (1).

Now suppose that $\psi_2 \prec \psi_1$, and let $b_1 < b_2$ be such that $\psi_1(b_1)\psi_2(b_2) < \psi_1(b_2)\psi_2(b_1)$. Consider again the expansions of the l.h.s. and r.h.s. of (1). Since we have already shown that each term on the r.h.s. is greater than or equal to the corresponding term on the left, we need only identify one term on the right which is strictly greater than the corresponding term on the left. The reasoning above already suffices to give this strict inequality for the case $c = b_1, d = b_2$, since then $\alpha_2 > \alpha_1$ and $\beta_2 = \mu(0)^2 > \beta_1$. □

The next lemma shows that $\preceq$ is also preserved by selection. In fact we shall prove a stronger result. For $\ell = 1$ and $W \in \mathbb{R}$, we define a new form of selection, which, as we saw in Lemma 6, allows us to understand the effect of selection on a single locus under certain conditions. For ϕ a probability distribution on $\mathbb{Z}$ and $x \in \mathbb{Z}$, we define

$$\mathtt{Sel}_W(\phi)(x) = \left(\frac{1}{s}\right)(x + W)\phi(x),$$

where s is the normalising factor required to make $\mathtt{Sel}_W(\phi)$ a probability distribution: $s = \sum_{x\in\mathbb{Z}}(x + W)\phi(x) = E(\phi) + W$. We call a probability distribution on $\mathbb{Z}$ *non-trivial* if its support consists of more than one point.

Lemma 10. *If $W_1 \leq W_2$ and $\psi_2 \preceq \psi_1$, then $\mathtt{Sel}_{W_2}(\psi_2) \preceq \mathtt{Sel}_{W_1}(\psi_1)$. Furthermore, if $W_1 < W_2$, $\psi_2 \preceq \psi_1$ and at least one of ψ_1 and ψ_2 is non-trivial, then $\mathtt{Sel}_{W_2}(\psi_2) \prec \mathtt{Sel}_{W_1}(\psi_1)$.*

Proof. Let $\psi_1^* = \mathtt{Sel}_{W_1}(\psi_1)$, let $\psi_2^* = \mathtt{Sel}_{W_2}(\psi_2)$ and consider $b_1 < b_2$. On the one side we have

$$\psi_1^*(b_1)\psi_2^*(b_2) = \frac{1}{s_1 s_2}(b_1 + W_1)(b_2 + W_2)\psi_1(b_1)\psi_2(b_2),$$

which we need to show is less than or equal to

$$\psi_1^*(b_2)\psi_2^*(b_1) = \frac{1}{s_1 s_2}(b_2 + W_1)(b_1 + W_2)\psi_1(b_2)\psi_2(b_1),$$

where s_1 and s_2 are the normalising factors for ψ_1 and ψ_2. We know that $\psi_1(b_1)\psi_2(b_2) \leq \psi_1(b_2)\psi_2(b_1)$, so it is enough to show that $(b_1 + W_1)(b_2 + W_2) \leq (b_2 + W_1)(b_1 + W_2)$. For this, one just needs to observe that:

$$(b_2 + W_1)(b_1 + W_2) - (b_1 + W_1)(b_2 + W_2) = (b_2 - b_1)(W_2 - W_1) \geq 0.$$

Note that this actually suffices to show $(b_1 + W_1)(b_2 + W_2) < (b_2 + W_1)(b_1 + W_2)$ if $W_1 < W_2$.

Now suppose that we also have $W_1 < W_2$. The reasoning above actually suffices to show *for all pairs* $b_1 < b_2$ that $\psi_1^*(b_1)\psi_2^*(b_2) < \psi_1^*(b_2)\psi_2^*(b_1)$, *so long as* $\psi_1(b_2)\psi_2(b_1) > 0$. If at least one of ψ_1 and ψ_2 is non-trivial then there exists a pair $b_1 < b_2$ with $\psi_1(b_2)\psi_2(b_1) > 0$, giving $\mathtt{Sel}_{W_2}(\psi_2) \prec \mathtt{Sel}_{W_1}(\psi_1)$ as required. □

Lemma 11. *If $\psi_2 \preceq \psi_1$, then $E(\psi_2) \leq E(\psi_1)$. Furthermore, if $\psi_2 \prec \psi_1$ then $E(\psi_2) < E(\psi_1)$.*

Proof. The proof is divided into various cases depending on the supports of ψ_1 and ψ_2. Let Π_1, Π_2 and Π_3 be as defined subsequent to Definition 8.

Case 1: The support of both ψ_1 and ψ_2 is a finite interval $[A, B]$ (so $\Pi_1 = \Pi_3 = \emptyset$ and $\Pi_2 = [A, B]$). This is the simplest case, but gives the principal idea for the entire proof. We will define probability density functions φ_i for $i \in [A, B]$, with $\psi_2 = \varphi_A \preceq \varphi_{A+1} \preceq \cdots \preceq \varphi_B = \psi_1$, and $E(\varphi_i) \leq E(\varphi_{i+1})$ for all $i \in [A, B)$. Each φ_i will satisfy:

$$(\forall b \in [A, i)) \quad \frac{\varphi_i(b+1)}{\varphi_i(b)} = \frac{\psi_1(b+1)}{\psi_1(b)} \quad \text{and} \quad (\forall b \in [i, B)) \quad \frac{\varphi_i(b+1)}{\varphi_i(b)} = \frac{\psi_2(b+1)}{\psi_2(b)}.$$

Suppose we have already defined φ_i and we want to define φ_{i+1}. We need to change the value of $\frac{\varphi_i(i+1)}{\varphi_i(i)}$ from $\frac{\psi_2(i+1)}{\psi_2(i)}$ to $\frac{\psi_1(i+1)}{\psi_1(i)}$ without changing any of the other fractions. For this, we need to find values c, d such that defining $\varphi_{i+1}(b) = c\ \varphi_i(b)$ for $b \leq i$ and $\varphi_{i+1}(b) = d\ \varphi_i(b)$ for $b > i$ gives the required probability density function. To find such c and d all one needs to do is to solve the following equation:

$$\begin{aligned} cS + d(1-S) &= 1 \\ d\ \psi_1(i)\psi_2(i+1) &= c\ \psi_1(i+1)\psi_2(i), \end{aligned}$$

where $S = \sum_{j=A}^{i} \varphi_i(j)$. Since $\frac{\psi_2(i+1)}{\psi_2(i)} \leq \frac{\psi_1(i+1)}{\psi_1(i)}$, we have $c \leq 1 \leq d$, and if $\frac{\psi_2(i+1)}{\psi_2(i)} < \frac{\psi_1(i+1)}{\psi_1(i)}$ then $c < 1 < d$. Since we are increasing the values of $\varphi_i(b)$ for $b > i$ and decreasing them for $b \leq i$, it is not hard to see that $E(\varphi_i) \leq E(\varphi_{i+1})$, and that if $\frac{\psi_2(i+1)}{\psi_2(i)} < \frac{\psi_1(i+1)}{\psi_1(i)}$, then $E(\varphi_i) < E(\varphi_{i+1})$. This finishes the construction of the φ_is and the proof for the case where the support of ψ_1 and ψ_2 is $[A, B]$.

Case 2: The support of ψ_1 and ψ_2 is not an interval, but it still finite and equal for both functions. The proof above works almost the same way, except that one has to skip the values not in the support.

Case 3: The supports of ψ_1 and ψ_2 are equal, but while Π_2 is bounded below it is not bounded above. One runs the same proof, but now constructs an infinite sequence $\psi_2 = \varphi_A \preceq \varphi_{A+1} \preceq \cdots$. Ultimately ψ_1 is the limit of this sequence, i.e. for all b, $\psi_1(b) = \lim_i \varphi_i(b)$. Since we have assumed that ψ_1 and ψ_2 have finite means, it follows that $E(\psi_1) = \lim_{i\to\infty} E(\varphi_i)$.

Case 4: The supports of ψ_1 and ψ_2 are equal, but while Π_2 is bounded above it is not bounded below. One runs the same proof, but now constructs an infinite sequence $\psi_2 = \varphi_B \preceq \varphi_{B-1} \preceq \cdots$, such that $E(\varphi_i) \leq E(\varphi_{i-1})$ for all $i \leq B$. Again we have ψ_1 as the limit of this sequence and $E(\psi_1) = \lim_{i\to\infty} E(\varphi_i)$.

Case 5: The supports of ψ_1 and ψ_2 are equal, and Π_2 neither bounded above nor bounded below. One runs almost the same proof, but now in two stages. Choosing $A \in \Pi_2$, we first construct an infinite sequence $\psi_2 = \varphi_A \preceq \varphi_{A+1} \preceq \cdots$ which has the intermediate value ψ_3 as limit. One then constructs an infinite sequence $\psi_3 = \varphi'_A \preceq \varphi'_{A-1} \preceq \cdots$ with ψ_1 as limit.

Case 6: At least one of Π_1 or Π_3 is non-empty. If Π_2 is empty then it immediately follows that $E(\psi_2) < E(\psi_1)$, so suppose this does not hold. Let ψ_1^* be the probability density function

formed from ψ_1 by restricting the support to Π_2 (and normalising as appropriate), and form ψ_2^* similarly. If Π_1 is non-empty then we have:

$$E(\psi_2) < E(\psi_2^*) \leq E(\psi_1^*) \leq E(\psi_1).$$

If Π_3 is non-empty then we have:

$$E(\psi_2) \leq E(\psi_2^*) \leq E(\psi_1^*) < E(\psi_1).$$

□

6.2. **The properties** (†) **and** (††)**.** This section is dedicated to proving Theorem 2, which asserts that LD_2 stays negative throughout the process, independent of what operations are applied and in which order, except when recombination has just been applied in which case $LD_2 = 0$. This is for the $\mathbb{Z}$-model, and assuming truncation is applied after each application of mutation and recombination (or at least before any application of selection).

We previously described a distribution on $\mathbb{Z}$ as non-trivial if there exists more than more point in the support. We shall refer to a population $\boldsymbol{\phi}$ (on $\mathbb{Z}^\ell$ for $\ell > 1$) as non-trivial if at least two of the marginal distributions ϕ_i are non-trivial (and providing this remains the case when truncation is applied to $\boldsymbol{\phi}$).

The key idea behind the proof of Theorem 2 is to consider a property, (†), which suffices to ensure that LD_2 is non-positive. We will also consider a strengthening of (†), which we call (††) and which ensures that LD_2 is negative. To prove Theorem 2 we then use induction on the generations and show that the property (†) is satisfied at each step of the process, and that, in fact, the stronger property (††) is also satisfied from some early stage onwards (the first stage after which `Sel` is applied to a non-trivial population).

Definition 12. *Given a probability distribution $\psi\colon \mathbb{Z}^2 \to \mathbb{R}^{\geq 0}$ and $a \in \mathbb{Z}$, we shall say that ψ^a is defined if $\psi(a) = \sum_b \psi(a,b) \neq 0$. In this case ψ^a is the distribution given by $\psi^a(b) = \psi(a,b)/\psi(a)$.*

We say ψ satisfies (†) if for every $a_1 < a_2$ such that ψ^{a_1} and ψ^{a_2} are defined, $\psi^{a_2} \preceq \psi^{a_1}$. We say that ψ satisfies (††) if, in addition, there exist $a_1 < a_2$ such that ψ^{a_1} and ψ^{a_2} are defined and $\psi^{a_2} \prec \psi^{a_1}$ as witnessed by a pair $b_1 < b_2$ with $a_1 + b_1 > 0$.

To define (†) for a population with $\ell > 2$, we need to consider each locus compared to the rest of the loci altogether. For $i \neq j \in \{1, ..., \ell\}$, let $F_i(X) = X_1 + \cdots + X_{i-1} + X_{i+1} + \cdots + \mathcal{X}_\ell = F(X) - X_i$. Given a population $\boldsymbol{\phi}$, let $\hat{\phi}_i$ be the distribution corresponding to the random variable $(X_i, F_i(X))$. Equivalently:

$$\hat{\phi}_i(a,b) = \sum_{\substack{\boldsymbol{x} \in \mathbb{Z}^\ell, \\ x_i = a,\ F(\boldsymbol{x}) = a+b}} \boldsymbol{\phi}(\boldsymbol{x}).$$

Definition 13. *A population $\boldsymbol{\phi}$ satisfies (†) if $\hat{\phi}_i$ does for every $i = 1, ..., \ell$. A population $\boldsymbol{\phi}$ satisfies (††) if there exists i such that $\hat{\phi}_i$ satisfies (††).*

The next step is to prove that (†) and (††) are preserved through the operations. Recall that we are assuming the process starts at linkage equilibrium. Note that if $\boldsymbol{\phi}$ is at linkage equilibrium, then (†) holds – in that case we have equality between the left-hand side and the right-hand side in the definition of the $\preceq$ relation.

Lemma 14. *For $\ell \geq 2$, if $\boldsymbol{\phi}$ satisfies (†), then $\mathtt{Sel}(\boldsymbol{\phi})$ satisfies (†), and if $\boldsymbol{\phi}$ is non-trivial then $\mathtt{Sel}(\boldsymbol{\phi})$ satisfies (††).*

Proof. Let $\boldsymbol{\phi}^* = \mathtt{Sel}(\boldsymbol{\phi})$. Fix $i \in \{1, ..., \ell\}$. A similar argument to that of Lemma 6 shows that $\hat{\phi}_i^*(a,b) = (1/M)\,(a+b)\,\hat{\phi}_i(a,b)$. Thus, $(\hat{\phi}_i^*)^a = \mathtt{Sel}_a((\hat{\phi}_i)^a)$. Suppose $a_1 < a_2$ are such that $(\hat{\phi}_i^*)^{a_1}$ and $(\hat{\phi}_i^*)^{a_2}$ are both defined. It follows from Lemma 10 that, since $(\hat{\phi}_i)^{a_2} \preceq (\hat{\phi}_i)^{a_1}$, we have $(\hat{\phi}_i^*)^{a_2} \preceq (\hat{\phi}_i^*)^{a_1}$. If $\boldsymbol{\phi}$ is non-trivial it also follows from Lemma 10 that $(\hat{\phi}_i^*)^{a_2} \prec (\hat{\phi}_i^*)^{a_1}$. Thus $\hat{\phi}_i^*$ satisfies (†) and also satisfies (††) if $\boldsymbol{\phi}$ is non-trivial, as required. □

Lemma 15. *Both (†) and (††) are preserved by mutation.*

Proof. First, let us note that mutation can be broken down into a number of consecutive steps, by treating one locus at a time. Let $\mathtt{Mut}^k$ be the application of mutation on the kth locus, i.e., $\mathtt{Mut}^k(X_1, ..., X_k, ..., X_\ell) = (X_1, ..., X_k + C_k, ..., X_\ell)$. We will show that both (†) and (††) are preserved by each of the operations $\mathtt{Mut}^k$. Let $\boldsymbol{\phi}^* = \mathtt{Mut}^k(\boldsymbol{\phi})$. Fix i. Assuming (†) or (††) for $\hat{\phi}_i$, we establish that the same condition holds for $\hat{\phi}_i^*$.

Suppose first that $k \neq i$. Then for $a \in \mathbb{Z}$, $(\hat{\phi}_i^*)^a = \mathtt{Mut}((\hat{\phi}_i)^a)$, because the mutation happens in one of the loci included in the second coordinate, and has the same effect on $F_i(X)$. Consider $a_1 < a_2$ such that $(\hat{\phi}_i)^{a_1}$ and $(\hat{\phi}_i)^{a_2}$ are both defined. It then follows from Lemma 9 that since $(\hat{\phi}_i)^{a_2} \preceq (\hat{\phi}_i)^{a_1}$ we have $(\hat{\phi}_i^*)^{a_2} \preceq (\hat{\phi}_i^*)^{a_1}$, and hence that $\hat{\phi}_i^*$ satisfies (†). We get (††) similarly.

If $k = i$, then the proof is the same. One just needs to observe that $\psi\colon \mathbb{Z}^2 \to \mathbb{R}^{\geq 0}$ satisfies (†) (or (††)) if and only if $\psi'(a,b) = \psi(b,a)$ does. □

So far both lemmas hold for any of the models. The following lemma only holds for the $\mathbb{Z}$-model.

Lemma 16. *Both (†) and (††) are preserved by truncation for the $\mathbb{Z}$-model.*

Proof. Let $\boldsymbol{\phi}^*$ be the population which results from an application of truncation to $\boldsymbol{\phi}$. Let $s = \sum_{\boldsymbol{x} \in D} \boldsymbol{\phi}(\boldsymbol{x})$, where D is as in the $\mathbb{Z}$-model. Note that $\hat{\phi}_i^*(a,b) = 0$ if $a + b \leq 0$ and $\hat{\phi}_i^*(a,b) = \hat{\phi}_i(a,b)/s$ if $a + b > 0$.

Fix i, $a_1 < a_2$ and $b_1 < b_2$. Then if $\hat{\phi}_i^*(a_1, b_1) \neq 0$ we have $a_1 + b_1 > 0$, and hence both $a_1 + b_2$ and $a_2 + b_1$ are positive. Therefore, if $(\hat{\phi}_i)^{a_1}(b_2)(\hat{\phi}_i)^{a_2}(b_1) \geq (\hat{\phi}_i)^{a_2}(b_2)(\hat{\phi}_i)^{a_1}(b_1)$ then $(\hat{\phi}_i^*)^{a_1}(b_2)(\hat{\phi}_i^*)^{a_2}(b_1) \geq (\hat{\phi}_i^*)^{a_2}(b_2)(\hat{\phi}_i^*)^{a_1}(b_1)$. In the same way if (††) holds because $(\hat{\phi}_i)^{a_1}(b_2)(\hat{\phi}_i)^{a_2}(b_1) > (\hat{\phi}_i)^{a_2}(b_2)(\hat{\phi}_i)^{a_1}(b_1)$ holds and $a_1 + b_1 > 0$ (the latter condition being required by Definition 12), then this implies $(\hat{\phi}_i^*)^{a_1}(b_2)(\hat{\phi}_i^*)^{a_2}(b_1) > (\hat{\phi}_i^*)^{a_2}(b_2)(\hat{\phi}_i^*)^{a_1}(b_1)$. Thus both (†) and (††) are preserved, as required. □

The second part of the proof of Theorem 2 is to show the connection between the properties (†), (††) and LD_2. Recall that $\mathrm{Co}(X,Y)$ is the *covariance* of the random variables X and Y,

i.e., $\mathrm{Co}(X,Y) = E(XY) - E(X)E(Y)$. Using our calculations from §?? we obtain that:

$$LD_2(\boldsymbol{\phi}) = \sum_{i=1}^{\ell} Co(\hat{\phi}_i).$$

Lemma 17. *If $\varphi\colon \mathbb{Z}^2 \to \mathbb{R}^{\geq 0}$ satisfies (†), then $Co(\varphi) \leq 0$. Furthermore, if φ satisfies (††) then $Co(\varphi) < 0$.*

Proof. Let X, Y be random variables such that (X,Y) has probability distribution φ. The key idea is to use that (†) implies that $E(Y|X=a)$ is decreasing in a, which follows from Lemma 11.

Set $v = E(X)$. Now we would like to put $u = E(Y|X=v)$, but since we may have $v \notin \mathbb{Z}$ this presents a slight difficulty. If $v \notin \mathbb{Z}$, we let u be a number in between $E(Y|X=\lfloor v \rfloor)$ and $E(Y|X=\lceil v \rceil)$. Now let $W = X - v$ and $Z = Y - u$. Since v and u are constants we have:

$$\mathrm{Co}(X,Y) = \mathrm{Co}(W,Z) = E(WZ) - E(W)E(Z).$$

Let $\psi(W,Z)$ specify the distribution on the pair (W,Z). Notice that since φ satisfies (†), so does ψ. Let ψ_W and ψ_Z specify the corresponding marginal distributions, and let $\psi_Z(Z|W)$ specify the conditional distribution. We claim that $E(WZ)$ is non-positive. This is because:

$$\sum_a \sum_b \psi(a,b)\cdot ab \;=\; \sum_a a\psi_W(a) \sum_b b\psi_Z(b|a) \;=\; \sum_a a\psi_W(a) E(Z|W=a).$$

Now satisfaction of (†) and Lemma 11 imply that when a is positive $E(Z|W=a)$ is non-positive, and when a is negative $E(Z|W=a)$ is non-negative. Also $E(W)E(Z) = 0$ because $E(W) = 0$.

If φ satisfies (††) then $E(Z|W=a)$ is non-zero somewhere, and hence $E(WZ) < 0$, as needed to get $\mathrm{Co}(X,Y) < 0$. □

Corollary 18. *If $\boldsymbol{\phi}$ satisfies (†), then $LD_2(\boldsymbol{\phi}) \leq 0$. Furthermore, if $\boldsymbol{\phi}$ satisfies (††) then $LD_2(\boldsymbol{\phi}) < 0$.*

To finish the proof of Theorem 2, all we need to observe is that since the population starts at linkage equilibrium, it initially satisfies (†). By Lemmas 14, 15 and 16, the condition (†) is then satisfied throughout the process. Furthermore, the condition (††) is satisfied after an application of `Sel` to any non-trivial population (mutation will quickly produce non-trivial populations), and then remains satisfied until any such point as `Rec` is applied. Since satisfaction of (††) ensures negative LD_2, any non-trivial application of `Rec` therefore increases variance.

7. Appendix C: the bounded model

This section provides an analysis of the asymptotic behavior of the bounded model and is dedicated to giving the full proof of Theorem 3. Recall that in the bounded models the fitness values of the genes are restricted to $\{1,...,N\}$, which we denote $[1,N]$, and the domain of the process is $D = [1,N]^\ell$, which we sometimes denote N^ℓ. Also recall our assumption

that our mutation distributions only take non-zero values on -1, 0 and 1; we let $a = \mu(-1)$, $b = \mu(0)$ and $c = \mu(1)$. Even though that assumption seems to be inessential in simulations, it is important for the type of formal analysis we do in this section. Recall also that we assume that $c < a < b$.

Let us restate Theorem 3. In the statement of the theorem, we consider the case where both types of individuals, sexual and asexual, live together. Mutation and selection act the same way on both, but recombination acts only among the sexual individuals. One may consider a context, for example, in which a small proportion of individuals begin reproducing sexually in a population which has previously been entirely asexual. The theorem states that over time the proportion of the population which is sexual will then tend to 1.

Theorem 3. *For every a, b, c with $c < a$, for every $\ell > 1$, and for all sufficiently large N (i.e. there exists N_0 such that $\forall N \geq N_0$), whatever the initial population is, so long as the proportion of sexual individuals is non-zero, we have that the proportion of sexual individuals converges to 1 and the proportion of asexual ones converges to 0.*

Let us fix the values of a, b, c and ℓ throughout the rest of Section 7.

The proof of Theorem 3 proceeds by showing that the mean fitness of the sexual population is eventually higher than that of the asexual population, over a time average. We will be able to show that the asexual population in isolation converges to a limit distribution, and we will provide an upper bound for the mean fitness of that limit. While we strongly suspect that the sexual population in isolation also converges to a limit (as evidenced by simulations), we have not been able to prove it. Nevertheless, we can still provide a lower bound for the geometric mean of the mean fitness over generations, which is larger than the upper bound for asex. We will show this is enough to establish that sex outperforms asex. To obtain these upper and lower bounds, the key technique is to study the case when positive and negative mutations have the same probabilities. This case is much easier to analyse, and we then find a way of translating those results to the case we are interested in, where downward mutations are more likely. An issue that we have to be constantly aware of is how truncation affects the populations.

Let us start by showing that we can analyse the sexual and asexual populations separately, as we were doing earlier in the paper. Let $\boldsymbol{\phi}^t \in \mathbb{R}^{D \times \{\mathtt{s},\mathtt{a}\}}$ be the probability distribution for the entire population at stage t (no confusion should result from any conflict in notation with that given in Definition 12 – the latter notation may be considered only to apply to the proof of Theorem 2). We can split $\boldsymbol{\phi}^t$ into two functions $\boldsymbol{\phi}^t_{\mathtt{sex}}$ and $\boldsymbol{\phi}^t_{\mathtt{asex}}$, each with domain D. Let $\varphi^t_{\mathtt{sex}}$ and $\varphi^t_{\mathtt{asex}}$ be the proportions of individuals which are sexual and asexual respectively, i.e., $\varphi^t_{\mathtt{sex}} = |\boldsymbol{\phi}^t_{\mathtt{sex}}|$ and $\varphi^t_{\mathtt{asex}} = |\boldsymbol{\phi}^t_{\mathtt{asex}}|$, where $|g|$ is the *taxicab norm*: For $\boldsymbol{\psi} \in \mathbb{R}^D$,

$$|\boldsymbol{\psi}| = \sum_{\boldsymbol{x} \in D} |\boldsymbol{\psi}(\boldsymbol{x})|.$$

So far, we have always assumed that populations, $\boldsymbol{\psi}$, are probability distributions, and hence that $|\boldsymbol{\psi}| = 1$. This is not the case with $\boldsymbol{\phi}^t_{\mathtt{sex}}$ and $\boldsymbol{\phi}^t_{\mathtt{asex}}$, and we will not be making that

assumption anymore. If a probability distribution is really required, all we have to do is consider normalisation:

$$\texttt{Proj}(\boldsymbol{\psi}) = \frac{\boldsymbol{\psi}}{|\boldsymbol{\psi}|}.$$

From now on when we refer to the operations of truncation and selection, we will omit the normalisation: Thus, when we apply truncation, giving $\texttt{Trunc}(\boldsymbol{\psi})$, we just erase the individuals outside the domain, and applying selection, giving $\texttt{Sel}(\boldsymbol{\psi})$, simply involves multiplying $\boldsymbol{\psi}(\boldsymbol{x})$ by $F(\boldsymbol{x})$. We do not alter the definition of $\texttt{Mut}$ as this operation already preserves the norm. The definition of $\texttt{Rec}$ is now altered in the obvious manner, so that it remains norm preserving (requiring division by $|\boldsymbol{\psi}|^{\ell-1}$). Since the normalization operation, $\texttt{Proj}(\boldsymbol{\psi})$, commutes with all the other operations, it does not really matter when we apply it. The advantage of describing the same process in this new fashion, is that now mutation, truncation and selection act independently for the sexual and asexual populations, and it is only normalisation that involves interaction between them. Of the operations selection, recombination, mutation and truncation, only truncation and selection affect the values $\varphi_{\texttt{sex}}$ and $\varphi_{\texttt{asex}}$. Since recombination and mutation commute, we can and will assume in what follows that the operations cycle through in that order. In fact, we will assume that $\boldsymbol{\phi}^{t+1}$ is obtained from $\boldsymbol{\phi}^t$ by applying: mutation, truncation, selection and recombination in that order (recombination, of course, being applied for sex only). For each t, let:

$$\lambda^t_{\texttt{sex}} = \frac{|\texttt{Rec}(\texttt{Sel}(\texttt{Trunc}(\texttt{Mut}(\boldsymbol{\phi}^t_{\texttt{sex}}))))|}{|\boldsymbol{\phi}^t_{\texttt{sex}}|} \quad \text{and} \quad \lambda^t_{\texttt{asex}} = \frac{|\texttt{Sel}(\texttt{Trunc}(\texttt{Mut}(\boldsymbol{\phi}^t_{\texttt{asex}})))|}{|\boldsymbol{\phi}^t_{\texttt{asex}}|}.$$

For the quotients, we have:

$$\frac{\varphi^t_{\texttt{sex}}}{\varphi^t_{\texttt{asex}}} \;=\; \frac{\varphi^{t-1}_{\texttt{sex}}}{\varphi^{t-1}_{\texttt{asex}}}\frac{\lambda^{t-1}_{\texttt{sex}}}{\lambda^{t-1}_{\texttt{asex}}} \;=\; \frac{\varphi^0_{\texttt{sex}}}{\varphi^0_{\texttt{asex}}}\frac{\prod_{i=0}^{t-1}\lambda^i_{\texttt{sex}}}{\prod_{i=0}^{t-1}\lambda^i_{\texttt{asex}}}.$$

To establish Theorem 3, it will be enough to show that for all sufficiently large N:

$$\lim_t \frac{\prod_{i=0}^{t-1}\lambda^i_{\texttt{sex}}}{\prod_{i=0}^{t-1}\lambda^i_{\texttt{asex}}} = +\infty.$$

We may therefore consider the populations separately, because all we need be concerned with are the values $\prod_{i=0}^{t-1}\lambda^i_{\texttt{sex}}$ and $\prod_{i=0}^{t-1}\lambda^i_{\texttt{asex}}$, which evolve independently. We will establish the limit above, by showing that the geometric mean of $\lambda^i_{\texttt{sex}}$ is eventually always greater than that of $\lambda^i_{\texttt{asex}}$ by at least a fixed margin.

Let us fix the notation for dealing with geometric means. Given a finite sequence $a_1, ..., a_n$ of numbers, we let $\mathrm{GM}_{t\leq n}(\{a_t\}) = \sqrt[n]{\prod_{i=1}^n a_i}$. Given an infinite sequence $\{a_t\}_{t\in\mathbb{N}}$ we let

$$\overline{\underset{t}{\mathrm{GM}}}(a_t) = \limsup_n \sqrt[n]{\prod_{i=1}^n a_i}, \qquad \underline{\underset{t}{\mathrm{GM}}}(a_t) = \liminf_n \sqrt[n]{\prod_{i=1}^n a_i},$$

and if both limits are the same we call this common value $\mathrm{GM}_t(a_t)$.

The rest of this subsection is dedicated to proving the following theorems:

Theorem 19. *For every N, the limit of λ^t_{asex} exists and, for $\tau = b + 2\sqrt{ac}$:*

$$\lim_{t\to\infty} \lambda^t_{asex} < N\ell\tau^\ell.$$

Using facts observed from simulations, we are confident in claiming that in actual fact the following holds: $\lim_{N\to\infty} \lim_{t\to\infty} \lambda^t_{\mathtt{asex}}/N = \ell\tau^\ell$. We will not need this extra fact, however, and the result of the theorem will be enough for our purposes.

Theorem 20. *Let $\tau = b + 2\sqrt{ac}$ as above. For all sufficiently large N:*

$$\underline{\mathrm{GM}}_{t\to\infty}(\lambda^t_{sex})/ > N\ell\tau^\ell.$$

Using facts observed from simulations, we are confident in claiming that in actual fact we have $\lim_{N\to\infty} \lim_{t\to\infty} \lambda^t_{\mathtt{sex}}/N = \ell\tau/(\ell - \tau(\ell-1))$, which is greater than $\ell\tau^\ell$ for $\ell > 1$ and $\tau < 1$. Once again, we will not need this extra fact, however, and the result of the theorem will be enough for our purposes.

It then follows from the theorems above that $\underline{\mathrm{GM}}_{t\to\infty}(\lambda^t_{\mathtt{sex}}/\lambda^t_{\mathtt{asex}}) > 1$ for any large enough N, and hence that $\lim_t(\prod_{i=0}^{t-1} \lambda^i_{\mathtt{sex}})/(\prod_{i=0}^{t-1} \lambda^i_{\mathtt{asex}}) = +\infty$ as required.

7.0.1. *Understanding λ^t_{sex} and λ^t_{asex}.* In general, given a population $\boldsymbol{\psi} \in \mathbb{R}^{\mathbb{N}^\ell}$, we define

$$\lambda(\boldsymbol{\psi}) = \frac{|(\mathtt{Sel}(\mathtt{Trunc}(\mathtt{Mut}(\boldsymbol{\psi}))))|}{|\boldsymbol{\psi}|}.$$

Then $\lambda^t_{\mathtt{sex}} = \lambda(\boldsymbol{\phi}^t_{\mathtt{sex}})$, and similarly for asex.

Given a population $\boldsymbol{\psi} \in \mathbb{R}^{\mathbb{N}^\ell}$, let $\rho(\boldsymbol{\psi})$ be the proportion of individuals surviving mutation followed by truncation, i.e.:

$$\rho(\boldsymbol{\psi}) = \frac{|\mathtt{Trunc}(\mathtt{Mut}(\boldsymbol{\psi}))|}{|\boldsymbol{\psi}|}.$$

Let us remark that $\rho(\boldsymbol{\psi}) \leq 1$. Given a population $\boldsymbol{\psi} \in \mathbb{R}^{\mathbb{N}^\ell}$, we use $M(\boldsymbol{\psi})$ to denote its mean fitness, even in the case that $\boldsymbol{\psi}$ is not normalised:

$$M(\boldsymbol{\psi}) = \frac{\sum_{\boldsymbol{x}\in N^\ell} F(\boldsymbol{x})\boldsymbol{\psi}(\boldsymbol{x})}{|\boldsymbol{\psi}|}.$$

The increase in norm caused by an application of selection (ignoring normalization) is given by the mean fitness:

$$\frac{|\mathtt{Sel}(\boldsymbol{\psi})|}{|\boldsymbol{\psi}|} = \frac{\sum_{\boldsymbol{x}\in D} F(\boldsymbol{x})\boldsymbol{\psi}(\boldsymbol{x})}{|\boldsymbol{\psi}|} = M(\boldsymbol{\psi}).$$

Let us remark that $M(\boldsymbol{\psi}) \leq N\ell$ because $F(\boldsymbol{x}) \leq N\ell$ for every $\boldsymbol{x} \in D$.

Since mutation and recombination do not affect the norms, we have:

$$\lambda(\boldsymbol{\psi}) = \rho(\boldsymbol{\psi})M(\boldsymbol{\psi}'),$$

where $\boldsymbol{\psi}' = \mathtt{Trunc}(\mathtt{Mut}(\boldsymbol{\psi}))$.

7.0.2. *Changing the parameters.* A key idea here is to use the case when positive and negative mutations are equiprobable to get information about the case we are interested in, where $c < a$. In this subsection we show how we can change the mutation parameters from a, b, c to

a', b', c' satisfying $a' = c'$, in a manner which allows us to translate from one process to the other in a controlled way.

We define a', b', c' so that they satisfy the following equations:

$$a' + b' + c' = 1, \qquad \frac{b}{\sqrt{ac}} = \frac{b'}{\sqrt{a'c'}} \quad \text{and} \quad a' = c'.$$

The reason we require $b/\sqrt{ac} = b'/\sqrt{a'c'}$ will become clear later. These equations are enough to determine the values of a', b' and c' as follows. Since $a' = c' = (1 - b')/2$, for $\tau = b + 2\sqrt{ac}$ we get:

$$\frac{b}{\sqrt{ac}} = \frac{b'}{\frac{1}{2}(1 - b')} \tag{2}$$

$$b(1 - b') = 2b'\sqrt{ac} \tag{3}$$

$$b = b'(b + 2\sqrt{ac}) \tag{4}$$

$$\frac{\sqrt{ac}}{\sqrt{a'c'}} = \frac{b}{b'} = \tau. \tag{5}$$

So $b' = b/\tau$ and $a' = c' = \sqrt{ac}/\tau$.

To better visualize the translation from the abc-process to the $a'b'c'$-process, let us start by considering the case $\ell = 1$ first. Consider the diagonal square matrix C of size $N \times N$ given by:

$$C(x, x) = (a/c)^{x/2}.$$

Our goal now is to show that applying the abc-process to a population $\boldsymbol{\phi}$ is equivalent to applying the $a'b'c'$-process to $C \cdot \boldsymbol{\phi}$ up to a factor of τ. In other words, we will show that $\tau \cdot \mathtt{Sel}(\mathtt{Mut}_{a'a'}(C \cdot \boldsymbol{\phi})) = C \cdot \mathtt{Sel}(\mathtt{Mut}_{ac}(\boldsymbol{\phi}))$.

Let $\mathtt{Mut}_{ac}$ be the matrix corresponding to an application of mutation with probabilities $\mu(-1) = a$, $\mu(0) = b$ and $\mu(1) = c$, followed by truncation (but without normalisation). That is:

$$\mathtt{Mut}_{ac} = \begin{pmatrix} b & a & 0 & 0 & \dots & 0 & 0 \\ c & b & a & 0 & \dots & 0 & 0 \\ 0 & c & b & a & \dots & 0 & 0 \\ 0 & 0 & c & b & \ddots & & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & & \ddots & b & a \\ 0 & 0 & 0 & 0 & \dots & c & b \end{pmatrix}$$

From now on, we will assume truncation is part of mutation, and mutation refers to multiplication by $\mathtt{Mut}_{ac}$.

Lemma 21. *In the case $\ell = 1$:*

$$\tau \cdot \mathtt{Mut}_{a'a'} = C \cdot \mathtt{Mut}_{ac} \cdot C^{-1}$$

Proof. We carry out the matrix multiplications:

$$C\cdot \mathtt{Mut}_{ac}\cdot C^{-1}=\begin{pmatrix} b\sqrt{a/c}^{1}\sqrt{a/c}^{-1} & a\sqrt{a/c}^{1}\sqrt{a/c}^{-2} & 0 & 0 & ... \\ c\sqrt{a/c}^{2}\sqrt{a/c}^{-1} & b\sqrt{a/c}^{2}\sqrt{a/c}^{-2} & a\sqrt{a/c}^{2}\sqrt{a/c}^{-3} & 0 & ... \\ 0 & c\sqrt{a/c}^{3}\sqrt{a/c}^{-2} & b\sqrt{a/c}^{3}\sqrt{a/c}^{-3} & a\sqrt{a/c}^{3}\sqrt{a/c}^{-4} & ... \\ 0 & 0 & c\sqrt{a/c}^{4}\sqrt{a/c}^{-3} & b\sqrt{a/c}^{4}\sqrt{a/c}^{-4} & ... \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

$$=\begin{pmatrix} b & a\sqrt{a/c}^{-1} & 0 & 0 & ... \\ c\sqrt{a/c} & b & a\sqrt{a/c}^{-1} & 0 & ... \\ 0 & c\sqrt{a/c} & b & a\sqrt{a/c}^{-1} & ... \\ 0 & 0 & c\sqrt{a/c} & b & ... \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}=\begin{pmatrix} b & \sqrt{ac} & 0 & 0 & ... \\ \sqrt{ac} & b & \sqrt{ac} & 0 & ... \\ 0 & \sqrt{ac} & b & \sqrt{ac} & ... \\ 0 & 0 & \sqrt{ac} & b & ... \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

$$=\sqrt{\frac{ac}{a'c'}}\begin{pmatrix} b' & \sqrt{a'c'} & 0 & 0 & ... \\ \sqrt{a'c'} & b' & \sqrt{a'c'} & 0 & ... \\ 0 & \sqrt{a'c'} & b' & \sqrt{a'c'} & ... \\ 0 & 0 & \sqrt{a'c'} & b' & ... \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}=\tau\cdot \mathtt{Mut}_{a'a'}.$$

The last equality uses that $\sqrt{a'c'}=a'=c'$ and that $\sqrt{\frac{ac}{a'c'}}=\tau$. □

Notice that, in the 1-locus case, selection without normalization is given by a diagonal matrix $\mathtt{Sel}$ where $\mathtt{Sel}(i,i)=i$. Since diagonal matrices commute, we have:

$$\tau\cdot \mathtt{Sel}\cdot \mathtt{Mut}_{a'a'}\cdot C=C\cdot \mathtt{Sel}\cdot \mathtt{Mut}_{ac}.$$

The case when $\ell>1$ is not overly different, but the notation is now a little more cumbersome. Consider the diagonal square matrix C of size $N^\ell\times N^\ell$ given by

$$C((x_1,...,x_\ell),(x_1,...,x_\ell))=(a_{-1}/a_1)^{(\sum x_i)/2}$$

where a_{-1},a_0,a_1 are a,b,c respectively.

Let us use $\mathtt{Mut}_{ac}$ to denote the matrix corresponding to abc-mutation with ℓ genes. We should actually denote this matrix by $\mathtt{Mut}_{ac,\ell,N}$, but since there is no risk of confusion we prefer to simplify the notation.

Lemma 22. *For $\ell\geq 1$:*

$$\tau^\ell\cdot \mathtt{Mut}_{a'a'}=C\cdot \mathtt{Mut}_{ac}\cdot C^{-1}$$

Proof. Consider $\boldsymbol{\psi}\in\mathbb{R}^{N^\ell}$. Then, for $\boldsymbol{x}=(x_1,...,x_\ell)$,

$$C^{-1}\cdot\boldsymbol{\psi}(\boldsymbol{x})=\boldsymbol{\psi}(\boldsymbol{x})\sqrt{a_{-1}/a_1}^{-\sum x_i}$$

and

$$\mathtt{Mut}_{ac}\cdot C^{-1}\cdot\boldsymbol{\psi}(\boldsymbol{x})=\sum_{i_1=-1}^{1}\sum_{i_2=-1}^{1}...\sum_{i_\ell=-1}^{1}\left(\prod_{j=1}^{\ell}a_{i_j}\right)\boldsymbol{\psi}(x_1-i_i,x_2-i_2,...,x_\ell-i_\ell)\sqrt{a_{-1}/a_1}^{-\sum x_j-i_j}.$$

In the equation above, assume that if $(x_1 - i_i, x_2 - i_2, ..., x_\ell - i_\ell) \notin D$, then $\boldsymbol{\psi}(x_1 - i_i, x_2 - i_2, ..., x_\ell - i_\ell) = 0$. Replacing each a_0 by $a'_0\tau$, each a_{-1} by $a'_{-1}\tau\sqrt{a_{-1}/a_1}$ and each a_1 by $a'_1\tau\sqrt{a_1/a_{-1}}$ we get

$$= \sum_{i_1=-1}^{1}\sum_{i_2=-1}^{1}...\sum_{i_\ell=-1}^{1}\left(\prod_{j=1}^{\ell} a'_{i_j}\tau\sqrt{a_1/a_{-1}}^{i_j}\right)\boldsymbol{\psi}(x_1 - i_i, x_2 - i_2, ..., x_\ell - i_\ell)\sqrt{a_{-1}/a_1}^{-\sum_j x_j + \sum_j i_j}$$

$$= \tau^\ell\sqrt{a_{-1}/a_1}^{-\sum_j x_j}\sum_{i_1=-1}^{1}\sum_{i_2=-1}^{1}...\sum_{i_\ell=-1}^{1}\left(\prod_{j=1}^{\ell} a'_{i_j}\right)\boldsymbol{\psi}(x_1 - i_i, x_2 - i_2, ..., x_\ell - i_\ell)$$

$$= \tau^\ell \cdot C^{-1} \cdot \mathtt{Mut}_{a'a'} \cdot \boldsymbol{\psi}(\boldsymbol{x}). \qquad \square$$

7.0.3. *The fixed point for Asex.* In this subsection we prove Theorem 19 which states that the limit of $\lambda^t_{\mathtt{asex}}$ is less than or equal to $N\ell\tau^\ell$. The proof has two steps. First we show that if the asex process reaches a fixed point $\boldsymbol{\psi}$, then $\lambda(\boldsymbol{\psi}) \leq N\ell\tau^\ell$. Second, we show that, independent of the starting point, the asex population always converges to a fixed point.

As we mentioned before, mutation (which we now consider to incorporate truncation) acts on a population (considered as a vector) by multiplying this vector by the matrix $\mathtt{Mut}_{ac}$. We use $\mathtt{SelMut}_{ac}$ to denote the matrix $\mathtt{Sel} \cdot \mathtt{Mut}_{ac}$. In the case $\ell = 1$ we have

$$\mathtt{SelMut}_{ac} = \begin{pmatrix} b & a & 0 & 0 & ... \\ 2c & 2b & 2a & 0 & ... \\ 0 & 3c & 3b & 3a & ... \\ 0 & 0 & 4c & 4b & ... \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}.$$

The following lemma shows how useful is the translation developed in the previous subsection.

Lemma 23. *Suppose that $\boldsymbol{\psi}_{ac,\ell,N}$ is a fixed point for the asex process. Then $\lambda(\boldsymbol{\psi}_{ac,\ell,N}) < N\ell\tau^\ell$.*

Proof. Let $\boldsymbol{\psi} = \boldsymbol{\psi}_{ac,\ell,N}$. That $\boldsymbol{\psi}$ is a fixed point for the asex process means $\boldsymbol{\psi} = \mathtt{Proj}(\mathtt{SelMut}_{ac} \cdot \boldsymbol{\psi})$, or equivalently that $\boldsymbol{\psi}$ is an eigenvector for $\mathtt{SelMut}_{ac}$ with eigenvalue $\lambda(\boldsymbol{\psi})$, i.e., $\mathtt{SelMut}_{ac} \cdot \boldsymbol{\psi} = \lambda(\boldsymbol{\psi})\boldsymbol{\psi}$. From Lemma 22 we have that $\tau^\ell \cdot \mathtt{SelMut}_{a'a'} \cdot C = C \cdot \mathtt{SelMut}_{ac}$. It follows that $\boldsymbol{\vartheta} = C \cdot \boldsymbol{\psi}$ is an eigenvector of $\mathtt{SelMut}_{a'a'}$ with eigenvalue $\tau^{-\ell}\lambda(\boldsymbol{\psi})$. Thus

$$\tau^\ell\lambda(\boldsymbol{\vartheta}) = \lambda(\boldsymbol{\psi}),$$

where $\lambda(\boldsymbol{\vartheta})$ is calculated using the mutation $\mu'(-1) = a'$, $\mu'(0) = b'$, $\mu'(1) = a'$. (We should use the notation $\lambda_{a'a'}(\boldsymbol{\vartheta})$ and $\lambda_{ac}(\boldsymbol{\psi})$ to specify the mutation used, but it will be clear from context which definition we are using.) Since $\lambda(\boldsymbol{\vartheta}) = \rho(\boldsymbol{\vartheta})M(\mathtt{Mut}_{a'a'} \cdot \boldsymbol{\vartheta}) < N\ell$, we have $\lambda(\boldsymbol{\psi}) < N\ell\tau^\ell$ as required. $\square$

Theorem 19 now follows from the following lemma.

Lemma 24. *For every a, b, c, ℓ, N, there is a unique $\boldsymbol{\psi}_{ac,\ell,N} \in \mathbb{R}^{N^\ell}$ such that for any non-negative, non-zero $\boldsymbol{\phi} \in \mathbb{R}^{N^\ell}$:*

$$lim_{t\to\infty}(\mathtt{ProjSelMut}_{ac})^t \cdot \boldsymbol{\phi} = \boldsymbol{\psi}_{ac,\ell,N}.$$

Proof. We apply the Perron-Frobenius theorem, which states that a non-negative, irreducible and primitive matrix has a positive (real) eigenvalue λ whose absolute value is larger that that of any other eigenvalue, and that λ has a unique (up to scaling) associated eigenvector all whose coordinates are positive. The matrix $\mathtt{SelMut}_{ac}$ is *non-negative*, in the sense that all its entries are non-negative. It is also *irreducible and primitive* because all the entries of $(\mathtt{SelMut}_{ac})^N$ are positive. So we can apply the Perron-Frobenius theorem and get a positive eigenvector $\boldsymbol{\psi} = \boldsymbol{\psi}_{ac,\ell,N} \in \mathbb{R}^{N^\ell}$ which is a probability distribution with a positive eigenvalue λ that is the largest in absolute value. As a corollary of the Perron-Frobenius theorem we also get that $\lim_{t\to\infty} (\mathtt{SelMut}_{ac})^t/\lambda^t$ is the projection to the eigenspace given by $\boldsymbol{\psi}$, and that this projection is non-zero for any non-zero non-negative initial population. This implies that $\boldsymbol{\psi}$ is a universal attractor of the system defined by iterating $\mathtt{SelMut}_{ac}$ and normalisation.(For a similar application of the Perron-Frobenius Theorem in the previous literature, but which does not make use of the techniques established here to provide estimates for the mean of the resulting fixed point, see [27, 28].) □

Before we move on to consider the asymptotic behaviour for the sex process, we need to form a stronger version of Theorem 19 for the 1-locus case (where the sex and asex processes are identical). While we shall not establish for general ℓ that $\lim_{N\to\infty}\lim_{t\to\infty}\lambda^t_{\mathtt{asex}}/N = \ell\tau^\ell$, we shall now do so for the case $\ell = 1$ (since we shall later be able to apply this result in analysing the sex process).

Lemma 25. *Let $\psi_{ac,N}$ be the probability distribution which is the fixed point of the 1-locus process, and let $\vartheta_{a'a',N} = \mathtt{Proj}(C \cdot \psi_{ac,N})$. Then:*

(1) $lim_{N\to\infty}\lambda(\vartheta_{a'a',N})/N = 1$.
(2) $lim_{N\to\infty}\rho(\vartheta_{a'a',N}) = 1$.
(3) $lim_{N\to\infty}M(\mathtt{Mut}_{a'a'} \cdot \vartheta_{a'a',N})/N = 1$.

Proof. Since we consider a, b and c to be fixed, let $\psi_N = \psi_{ac,N}$ and $\vartheta_N = \vartheta_{a'a',N}$. We shall establish (1) and (2), and then (3) follows immediately from the definition of $\lambda(\vartheta_N)$. The key to understanding ϑ_N is to calculate the following quotients. For $k < N$, define:

$$\eta_N(k) = \frac{\vartheta_N(k+1)}{\vartheta_N(k)}.$$

Let $\lambda_N = \lambda(\vartheta_N)$. Since ϑ_N is a fixed point we have that $\vartheta_N(1) = (b'\vartheta_N(1) + a'\vartheta_N(2))/\lambda_N$ and $\vartheta_N(N) = (c'\vartheta_N(N-1) + b'\vartheta_N(N))N/\lambda_N$. It follows that:

$$\eta_N(1) = \frac{\lambda_N - b'}{a'} \quad \text{and} \quad \eta_N(N-1) = \frac{c'}{\lambda_N/N - b'}.$$

For $x \notin \{1, N\}$ we have $\vartheta_N(x) = (c'\vartheta_N(x-1) + b'\vartheta_N(x) + a'\vartheta_N(x+1))x/\lambda_N$. Using that $\vartheta_N(x+1) = \eta_N(x)\vartheta_N(x)$, we get:

$$c'\eta_N(x-1)^{-1} + a'\eta_N(x) = \lambda_N/x - b'.$$

Now suppose that (1) does not hold. In this case there exists an infinite set $\Pi \subseteq \mathbb{N}$, such that $\lim_{N\in\Pi} \lambda_N/N = \kappa < 1$ (note that $\lambda_N \leq N$). For each $x \in \mathbb{N}$, define:

$$R(x) = \lim_{N\in\Pi} \eta_N(N-x)^{-1}.$$

From the formulas for $\eta(N-x)$ above (and using that $a' = c'$), we deduce that R satisfies the following inductive definition:

$$R(1) = \frac{\kappa - b'}{a'} \quad \text{and} \quad R(k+1) = \frac{\kappa - b'}{a'} - R(k)^{-1}.$$

All values of R are non-negative, because so are the corresponding values of $\eta_N(k)$. Notice that $R(2) < R(1)$, and that $R(k) < R(k-1)$ implies $R(k+1) < R(k)$, from which we may conclude that R is decreasing. R must then have a limit, α say. This limit must satisfy $\alpha + \alpha^{-1} = (\kappa - b')/a'$. Since for every $\alpha \in \mathbb{R}^+$, $\alpha + \alpha^{-1} \geq 2$, $2 \leq (\kappa - b')/a'$. From the fact that $b' = 1 - 2a'$, it follows that $\kappa \geq 1$, which gives the required contradiction.

In order to establish (2), we show first of all that $\lim_{N\to\infty}\vartheta_N(1) = 0$. This now follows easily, however, from the fact that $\lim_{N\to\infty}\lambda_N = \infty$ and $\eta_N(1) = (\lambda_N - b')/a'$.

The final step is to show that $\lim_{N\to\infty}\vartheta_N(N) = 0$. Once again, consider the sequence $R(x)$ as defined above. We have:

$$R(1) = \frac{1 - b'}{a'} = 2 \quad \text{and} \quad R(x+1) = 2 - R(x)^{-1}.$$

We conclude that $R(x) > 1$ for all x. From this it follows that for each x and all sufficiently large N, $\eta_N(N-x) < 1$. This suffices to ensure that $\lim_{N\to\infty}\vartheta_N(N) = 0$, as required. □

Lemma 26. *Let $\psi_{ac,N}$ be the probability distribution which is the fixed point of the 1-locus process. Then:*

(1) $lim_{N\to\infty}\lambda(\psi_{ac,N})/N = \tau$.
(2) $lim_{N\to\infty}\rho(\psi_{ac,N}) = 1$.
(3) $lim_{N\to\infty}M(\mathtt{Mut}_{ac} \cdot \psi_{ac,N})/N = \tau$.

Proof. Again, let $\psi_N = \psi_{ac,N}$ and let $\vartheta_N = \vartheta_{a'a',N}$ be as defined in the statement of Lemma 25. Given Lemma 25, and the fact that $\tau\lambda(\vartheta_N) = \lambda(\psi_N)$ (as established in the proof of Lemma 23), it suffices to establish (2). That $\lim_{N\to\infty}\psi_N(N) = 0$, follows from the corresponding fact for ϑ_N, however, since $\psi_N \prec \vartheta_N$. It remains then, to show that $\lim_{N\to\infty}\psi_N(1) = 0$. We use a similar method to the proof of Lemma 25. This time for $k < N$, define:

$$\eta_N(k) = \frac{\psi_N(k+1)}{\psi_N(k)}.$$

Now let $\lambda_N = \lambda(\psi_N)$. We have that:

$$\eta_N(1) = \frac{\lambda_N - b}{a}.$$

The result then follows from the fact that $\lim_{N\to\infty}\lambda_N = \infty$. □

7.0.4. *The asymptotic behavior for sex.* The rest of Section 7 is dedicated to proving Theorem 20, which gives a lower bound for the geometric mean of $\lambda^t_{\mathtt{sex}}$. For all sufficiently large N:

$$\underset{t\to\infty}{\mathrm{GM}}(\lambda^t_{\mathtt{sex}}) > N\ell\tau^\ell.$$

The proof requires a sequence of lemmas, some of which we will state now and prove in later subsections. Before describing the general architecture of the argument, let us consider how to analyse the ℓ-locus sex process by looking at the different loci individually. The reason we can do this is that, since the sex population stays at linkage equilibrium, its probability distribution is determined by the product of the distributions of the individual loci.

Let $\boldsymbol{\psi} \in \mathbb{R}^{N^\ell}$ be a population where all the loci are independent, as for instance after an application of recombination. Assume $\boldsymbol{\psi}$ has been normalised. Let $\psi_i \in \mathbb{R}^N$ be the probability distribution for the ith locus. We would like to analyse the abc-sex-process on $\boldsymbol{\psi}$ by analysing its process on ψ_i. It is not hard to see that if a population is at linkage equilibrium then the effect of mutation and truncation on the whole population is equivalent to considering the effect of mutation on each single locus independently as we did in the proof of Lemma 15. Let $\psi'_i = \mathtt{Mut}_{ac,1}\cdot\psi_i$, where $\mathtt{Mut}_{ac,1}$ is the 1-locus mutation, and $\rho_i = \rho(\psi_i) = |\psi'_i|/|\psi_i|$. If we let $\boldsymbol{\psi}' = \mathtt{Mut}_{ac}\cdot\boldsymbol{\psi}$, then we have that $\boldsymbol{\psi}'(\boldsymbol{x}) = \prod_{i=1}^\ell \psi'(x_i)$ and $\rho(\boldsymbol{\psi}) = \prod_{i=1}^\ell \rho(\psi_i)$. Finally, we let $W_i = M(\psi'_i)$ and $\hat{W}_i = \sum_{j\neq i} W_j$; let us recall that $M(\boldsymbol{\psi}') = \sum_{i=1}^\ell W_i = W_i + \hat{W}_i$. From Lemma 6 we have that the effect of selection on a single locus is given by $\psi^*_i = \mathtt{Sel}_{\hat{W}_i}\cdot\psi'_i$, where $\mathtt{Sel}_K$ is the diagonal matrix with $\mathtt{Sel}_K(j,j) = (j+K)/M$. Since this matrix actually depends also on M, from now on, to simplify the notation we let

$$\mathtt{Sel}_K(j,j) = j + K$$

and leave the normalisation for later if necessary. The increase in norm produced by $\mathtt{Sel}_K$ applied to $\psi \in \mathbb{R}^N$ is then $M(\psi) + K$. Hence for $\psi'_i = \mathtt{Mut}_{ac,1}\cdot\psi_i$ we have:

$$\lambda(\psi_i) := \frac{|\mathtt{Sel}_{\hat{W}_i}\cdot\mathtt{Mut}_{ac,1}\cdot\psi_i|}{|\psi_i|} = \rho(\psi_i)(M(\psi'_i) + \hat{W}_i).$$

The next step is to describe how we are going to use the translation to the $a' = c'$ case for sex populations. Consider the following setting. Let $\{K^t\}_{t\in\mathbb{N}}$ be a sequence of real numbers in $[0,(\ell-1)N]$ (which will later represent the sequence $\hat{W}^t_i$ for some fixed i). Let $\psi^0 \in \mathbb{R}^N$ (later this will represent the initial sexual population at some locus), and let $\vartheta^0 = C\cdot\psi^0$. Assume that ψ^0 is not the zero vector. For every $t\in\mathbb{N}$, define:

$$\psi^{t+1} = \mathtt{Sel}_{K^t}\cdot\mathtt{Mut}_{ac,1}\cdot\psi^t \quad \text{and} \quad \vartheta^{t+1} = \mathtt{Sel}_{K^t}\cdot\mathtt{Mut}_{a'a',1}\cdot\vartheta^t.$$

From Lemma 21 we have $\tau^t\cdot\vartheta^t = C\cdot\psi^t$ for every t. Define:

$$\lambda^t_\psi = \frac{|\psi^{t+1}|}{|\psi^t|} = \rho(\psi^t)\cdot(M((\psi^t)') + K^t) \quad \text{and} \quad \lambda^t_\vartheta = \frac{|\vartheta^{t+1}|}{|\vartheta^t|} = \rho(\vartheta^t)\cdot(M((\vartheta^t)') + K^t),$$

where $(\psi^t)' = \mathtt{Mut}_{ac,1} \cdot \psi^t$ and $(\vartheta^t)' = \mathtt{Mut}_{a'a',1} \cdot \vartheta^t$.

If ψ^t was a fixed point, then using that $\tau^t \cdot \vartheta^t = C \cdot \psi^t$ we could conclude that ϑ^t is also a fixed point (all given the appropriate normalisations), and that $\lambda^t_\psi = \tau\lambda^t_\vartheta$ as in the proof of Lemma 23. Even without assuming that the process converges to a limit distribution, we still get that these values have the same geometric means:

Lemma 27. $\mathrm{GM}_{t\to\infty}(\lambda^t_\psi/\lambda^t_\vartheta) = \tau$.

Proof. For every k, we have that:

$$\frac{\tau^k|\vartheta^k|}{|\vartheta^0|} = \frac{|C\cdot\psi^k|}{|C\cdot\psi^0|} = \frac{|\psi^k|}{|\psi^0|}\frac{|C\cdot\mathtt{Proj}(\psi^k)|}{|C\cdot\mathtt{Proj}(\psi^0)|}.$$

The set $\{\phi \in (\mathbb{R}^{\geq 0})^N : |\phi| = 1\}$ is compact and hence the image of the continuous map $\psi \mapsto |C\cdot\mathtt{Proj}(\psi)|$ is a closed interval of the form $[\alpha, \alpha\beta]$ for $0 < \alpha$ and $1 \leq \beta$ (we get that $\alpha > 0$ because $|C\cdot\mathtt{Proj}(\psi))|$ is always positive). We then have:

$$\operatorname*{GM}_{t<k}(\lambda^t_\psi) = \sqrt[k]{\frac{|\psi^k|}{|\psi^0|}} \leq \sqrt[k]{\beta\frac{\tau^k|\vartheta^k|}{|\vartheta^0|}} = \tau\sqrt[k]{\beta}\ \operatorname*{GM}_{t<k}(\lambda^t_\vartheta).$$

Symmetrically $\mathrm{GM}_{t<k}(\lambda^t_\psi) \geq \tau\sqrt[k]{\beta^{-1}}\,\mathrm{GM}_{t<k}(\lambda^t_\vartheta)$. The lemma then follows from the fact that both $\sqrt[k]{\beta^{-1}}$ and $\sqrt[k]{\beta}$ converge to 1 as $k \to \infty$. □

The next step is to give an approximate calculation for λ^t_ϑ, which holds irrespective of the choice for ψ^0. The next lemma shows that λ^t_ϑ is eventually always close to $N + K^t$. The reason this holds is that $\rho(\vartheta^t)$ is eventually always close to 1, and $M((\vartheta^t)')$ is eventually always close to N because positive mutations are as likely as negative ones.

Lemma 28. *For any choice of non-negative $\psi^0 \in \mathbb{R}^N$ and $\{K^t\}_{t\in\mathbb{N}}$, let $\{\lambda^t_\vartheta\}_{t\in\mathbb{N}}$ be defined as above. There is a sequence $\{\epsilon_N\}_{N\in\mathbb{N}}$ converging to 0 such that, for every N, every sequence $\{K^t\}_{t\in\mathbb{N}}$ and every non-negative ψ^0, the following holds for all sufficiently large t:*

$$1 - \epsilon_N < \frac{\lambda^t_\vartheta}{N + K^t} < 1.$$

The proof of this lemma is a little technical, so we delay it until Subsection 7.0.5. The following is a small lemma concerning geometric means, which will allow us to compare the geometric means of λ^t_ϑ and K^t.

Lemma 29. *Let $\{a^t\}_{t<k}$ be a sequence of positive real numbers and let b be a positive number. Then $b + \mathrm{GM}_{t<k}(a^t) \leq \mathrm{GM}_{t<k}(b + a^t)$.*

Proof. This is a corollary of Jensen's inequality which states that $\varphi(k^{-1}\sum_{i=1}^k \gamma_i) \leq k^{-1}\sum_{i=1}^k \varphi(\gamma_i)$ for $\varphi\colon \mathbb{R} \to \mathbb{R}$ convex. One has to apply it to the values $\gamma_i = \log(a^i)$ and the function $\varphi(x) = \log(b + e^x)$ which is convex because $\varphi''(x) = be^x/(b + e^x)^2 > 0$. □

For a given choice of ψ^0, we can use what we have so far to get a lower bound for $\mathrm{GM}(\rho(\psi^t))$. For all sufficiently large k:

$$\underset{t<k}{\mathrm{GM}}\,\rho(\psi^t) = \underset{t<k}{\mathrm{GM}}\,\frac{\lambda^t_\psi}{M((\psi^t)') + K^t} \tag{6}$$

$$\geq \underset{t<k}{\mathrm{GM}}\,\frac{\tau\lambda^t_\vartheta}{M((\psi^t)') + K^t}(1-\epsilon_N) \tag{7}$$

$$\geq \tau\left(\underset{t<k}{\mathrm{GM}}\,\frac{N+K^t}{M((\psi^t)') + K^t}\right)(1-\epsilon_N)^2. \tag{8}$$

The value inside the large parentheses is greater than 1 for large N and sufficiently large t because there exists a sequence $\{\epsilon_N\}_{N\in\mathbb{N}}$ with limit 0 such that, for large t, $M((\psi^t)') < \tau N + \epsilon_N < N$, as proved in the next lemma. The lemma also shows that, for large N and for sufficiently large t, the value of ψ^t at the upper boundary N is very close to 0.

Lemma 30. *For any choice of $\psi^0 \in \mathbb{R}^N - \{\mathbf{0}\}$ and $\{K^t\}_{t\in\mathbb{N}}$, let $\{\psi^t\}_{t\in\mathbb{N}}$ be defined as above, and let $(\psi^t)' = \mathtt{Mut}_{ac,1}\cdot\psi^t$. There exists a sequence $\{\epsilon_N\}_{N\in\mathbb{N}}$ converging to 0 such that, for every N, every sequence $\{K^t\}_{t\in\mathbb{N}}$ and every ψ^0, the following holds for all sufficiently large t:*

$$M((\psi^t)') < \tau N + \epsilon_N \quad \textit{and} \quad \psi^t(N)/|\psi^t| < \epsilon_N.$$

Proof. Let $\phi^0 \in \mathbb{R}^N$ be the probability distribution with $\phi^0(N) = 1$. For each t define $\phi^{t+1} = \mathtt{Sel}_0\mathtt{Mut}_{ac,1}\cdot\phi^t$. By induction on t and using Lemmas 9 and 10 we conclude that $\psi^t \preceq \phi^t$ and $(\psi^t)' \preceq (\phi^t)'$ for all t. Using Lemma 11 we then get that $M((\psi^t)') \leq M((\phi^t)')$. Applying Lemma 26, we conclude that there exists a sequence $\{\epsilon_N\}_{N\in\mathbb{N}}$ with limit 0 such that $M((\phi^t)')$ (and so also $M((\psi^t)')$) remains below $\tau N+\epsilon_N$ for all sufficiently large t. The second claim of the lemma also follows from Lemma 26 and the fact that $\psi^t \preceq \phi^t$. □

For the last stretch of the proof we need to be more concrete about the sexual population we are analysing. Let $\boldsymbol{\psi}^0 \in \mathbb{R}^{N^\ell}$ be the initial sexual population, and $\{\boldsymbol{\psi}^t\}_{t\in\mathbb{N}}$ be the sequence obtained by iterating *ac*-mutation, selection and recombination. For each locus i and generation t, let ψ^t_i be the distribution at locus i at stage t, but ignoring normalisation. We use the same notation we have been using so far:

- $(\psi^t_i)' = \mathtt{Mut}_{ac,1}\cdot\psi^t_i$;
- $(\boldsymbol{\psi}^t)' = \mathtt{Mut}_{ac}\cdot\boldsymbol{\psi}^t$;
- $W^t_i = M((\psi^t_i)')$;
- $M^t = M((\boldsymbol{\psi}^t)') = \sum_{i=1}^\ell W^t_i$;
- $\hat{W}^t_i = M^t_i - W^t_i$;
- $\psi^{t+1}_i = \mathtt{Sel}_{\hat{W}^t_i}\cdot\mathtt{Mut}_{ac,1}\cdot\psi^t_i$;
- $\rho(\psi^t_i) = |(\psi^t_i)'|/|\psi^t_i|$,
- $\rho(\boldsymbol{\psi}^t) = |(\boldsymbol{\psi}^t)'|/|\boldsymbol{\psi}^t| = \prod_{i=1}^\ell \rho(\psi^t_i)$.

The objective now is to show that the geometric mean of $\lambda(\boldsymbol{\psi}^t) = \rho(\boldsymbol{\psi}^t)M^t = (\prod_{i=1}^\ell \rho(\psi^t_i))M^t$ is above $N\ell\tau^\ell$. We will apply the results we have obtained thus far for $K^t = \hat{W}^t_i$. In order to be able to do this, however, we need to be able to compare $\hat{W}^t_i$ and M^t. If all loci were

identical, we would have $\hat{W}_i^t = M^t(\ell-1)/\ell$. When the loci are not identical, the following lemma gives us an approximation to $N+\hat{W}_i^t$, and tells us that it is close – at least in geometric mean – to $N + M^t(\ell-1)/\ell$, just as it would be if the loci were identical.

Lemma 31. *There is a sequence $\{\epsilon_N : N \in \mathbb{N}\}$ converging to 0 such that for every N and every initial population $\boldsymbol{\psi}^0 \in \mathbb{R}^{N^\ell}$ as above, the following holds for all sufficiently large k:*

$$1-\epsilon_N < \underset{t<k}{\mathrm{GM}} \left(\frac{N+\hat{W}_i^t}{N+M^t((\ell-1)/\ell)} \right) < 1+\epsilon_N.$$

We will prove this lemma in Subsection 7.0.6. Lemmas 28, 30 and 31 all assert the existence of certain sequences $\{\epsilon_N\}_{N\in\mathbb{N}}$ with limit 0. We now let $\{\epsilon_N\}_{N\in\mathbb{N}}$ be a sequence with limit 0, which majorises each of the sequences provided by these lemmas.

We are now ready to finish the proof of Theorem 20. Let k be large. We start by cleaning up equation (8) using what we now know from Lemmas 29 and 31. Fix $i \leq \ell$.

$$\begin{aligned}
\underset{t<k}{\mathrm{GM}}\, \rho_i^t &\geq \tau \left(\underset{t\leq k}{\mathrm{GM}} \frac{N+\hat{W}_i^t}{W_i^t+\hat{W}_i^t} \right) (1-\epsilon_N)^2 \\
&\geq \tau \left(\underset{t\leq k}{\mathrm{GM}} \frac{N+M^t(\ell-1)/\ell}{M^t} \right) (1-\epsilon_N)^3 \\
&= \tau \left(\underset{t\leq k}{\mathrm{GM}} \left(\frac{\ell-1}{\ell} + \frac{N}{M^t} \right) \right) (1-\epsilon_N)^3 \\
&\geq \tau \left(\frac{\ell-1}{\ell} + \underset{t\leq k}{\mathrm{GM}} \left(\frac{N}{M^t} \right) \right) (1-\epsilon_N)^3 \\
&= \tau\,(1+\xi_k/\ell)\,(1-\epsilon_N)^3,
\end{aligned}$$

where $\xi_k = \mathrm{GM}_{t\leq k}(\ell N/M^t) - 1$. Notice that $\mathrm{GM}_{t\leq k}(\ell N/M^t) > 1$. Furthermore, by Lemma 30, $M^t < \ell(\tau N + \epsilon_N)$ for sufficiently large t. Adjusting the sequence ϵ_N as necessary (but maintaining the fact that it has limit 0), we then have that for sufficiently large k, $\xi_k \geq ((1-\tau)/\tau) - \epsilon_N$.

Finally,

$$\underset{t<k}{\mathrm{GM}}(\lambda_{\mathtt{sex}}^t) = \underset{t<k}{\mathrm{GM}}(M^t) \prod_{i=1}^{\ell} \underset{t<k}{\mathrm{GM}}(\rho_i^t) \tag{9}$$

$$\geq \frac{N\ell}{1+\xi_k} \left(\tau\,(1+\xi_k/\ell)\,(1-\epsilon_N)^3 \right)^\ell \tag{10}$$

$$= \left(N\ell\tau^\ell \right) \left(\frac{(1+\xi_k/\ell)^\ell}{1+\xi_k} \right) (1-\epsilon_N)^{3\ell}, \tag{11}$$

for all large enough k. The last observation to make is that there exists $\epsilon > 0$, independent of N and k, such that the factor

$$\left(\frac{(1+\xi_k/\ell)^\ell}{1+\xi_k} \right) (1-\epsilon_N)^{3\ell}$$

is greater than $1+\epsilon$ for large N and sufficiently large k. To see this, note that the function $x \mapsto (1+x/\ell)^\ell/(1+x)$ is always greater than 1 for $x > 0$ and tends to $+\infty$ as $x \to +\infty$.

(It is actually increasing for $x > 0$.) Let N^* be large enough that $\epsilon_N < (1-\tau)/\tau$ for all $N > N^*$. Among all the x's with $x \geq ((1-\tau)/\tau) - \epsilon_{N^*}$, there is a minimum possible value for $(1+x/\ell)^\ell/(1+x)$, call it ζ, which is greater than 1. Let ϵ be such that $\zeta = 1+2\epsilon$. Then for $N \geq N^*$ for which ϵ_N is sufficiently small, we have $\left(\frac{(1+\xi_k/\ell)^\ell}{1+\xi_k}\right)(1-\epsilon_N)^{3\ell} > 1+\epsilon$ for all large enough k.

It remains to prove Lemmas 28 and 31.

7.0.5. *The proof of Lemma 28.* Roughly speaking, we need to show that λ^t_ϑ gets close to $N+K^t$ as t becomes large. Recall that $\lambda^t_\vartheta = \rho(\vartheta^t)(M((\vartheta^t)') + K^t)$, and that $\rho(\vartheta^t) = 1 - a\vartheta^t(1)/|\vartheta^t| - c\vartheta^t(N)/|\vartheta^t|$. The proof will have three parts which are: showing that $M((\vartheta^t)')$ gets close to N, showing that $\vartheta^t(1)/|\vartheta^t|$ gets close to 0, and showing that $\vartheta^t(N)/|\vartheta^t|$ gets close to 0. We remark that it is not surprising that $M(\vartheta^t)$ gets close to N: since positive and negative mutations are equiprobable, mutation without truncation does not bring the mean fitness down, while selection only ever increases mean fitness. Therefore the mean fitness can be expected to rise, this rise being halted only by effect of truncation at the upper boundary.

The first idea for the proof is to consider an alternative population which evolves according to a different sequence $\{K^t\}_{t\in\mathbb{N}}$; one that is constant, and that is either always larger or else always smaller than the original one. The fact that the sequence is constant allows us to apply the Perron-Frobenius theorem and establish a limit population, which we can later analyse. Choosing a sequence K^t with larger (smaller) values will guarantee that the new sequence is $\prec$-below (-above) the original. This allows us to compare the mean fitnesses of the two populations, as well as their values at the boundaries 1 and N.

Let us begin with the analysis of the limit populations. Fix a value of K, for which we will later substitute either 0 or $(\ell-1)N$. Define

$$\phi^0 = \vartheta^0 \quad \text{and} \quad \phi^{t+1} = \mathtt{Sel}_K\mathtt{Mut}_{a'a',1}\cdot\phi^t.$$

Since ϕ^t is defined by iterating a linear system which is non-negative, primitive and irreducible, we can apply the Perron-Frobenius theorem, exactly as we did in Lemma 24, to deduce that the populations ϕ^t must converge to a limit population ϕ_N which is independent of the starting population (and depends only on N, K and a'). In order to analyse ϕ_N, we proceed much as in the proof of Lemma 25. Once again, the key idea is to consider quotients between consecutive values in the distribution. For $k < N$, define:

$$\eta_N(k) = \frac{\phi_N(k+1)}{\phi_N(k)}.$$

Let $\lambda_N = \lambda(\phi_N) = \rho(\phi_N)(M((\phi_N)') + K)$. Since $\phi_N(1) = (b'\phi_N(1) + a'\phi_N(2))(1+K)/\lambda_N$ and $\phi_N(N) = (c'\phi_N(N-1) + b'\phi_N(N))(N+K)/\lambda_N$ we have:

$$\eta_N(1) = \frac{\lambda_N/(1+K) - b'}{a'} \quad \text{and} \quad \eta_N(N-1) = \frac{c'}{\lambda_N/(N+K) - b'}.$$

For $x \notin \{1, N\}$ we have $\phi_N(x) = (c'\phi_N(x-1) + b'\phi_N(x) + a'\phi_N(x+1))(x+K)/\lambda_N$. Since $\phi_N(x+1) = \eta_N(x)\phi_N(x)$ this gives:

$$c'\eta_N(x-1)^{-1} + a'\eta_N(x) = \lambda_N/(x+K) - b'.$$

Let us now move into the proof that the mean fitness grows close to N. Consider $K = (\ell-1)N$, and define ϕ_N as above for that K. Since $K^t \leq (\ell-1)N$ (where K^t is the sequence given in the statement of the lemma), it follows by induction using Lemmas 9 and 10 that for every t, $\phi^t \prec \vartheta^t$. This means that $\phi^t(1)/|\phi^t| \geq \vartheta^t(1)/|\vartheta^t|$, and (by Lemma 11) that $M(\phi^t) \leq M(\vartheta^t)$. In order to establish that $\lim_N \lambda_N/(\ell N) = 1$, suppose otherwise. Then there must exist an infinite set Π, such that $\lim_{N\in\Pi}\lambda_N/(\ell N) = \kappa < 1$ (note that $\lambda_N \leq \ell N$). For each $x \in \mathbb{N}$, define:

$$R(x) = \lim_{N\in\Pi} \eta_N(N-x)^{-1}.$$

From the formulas for $\eta(N-x)$ above (and using that $a' = c'$), it follows that each $R(x)$ is defined and satisfies the following inductive definition:

$$R(1) = \frac{\kappa - b'}{a'} \quad \text{and} \quad R(k+1) = \frac{\kappa - b'}{a'} - R(k)^{-1}.$$

All the values of R are non-negative, because so are the corresponding values of $\eta_N(k)$. Note that $R(2) < R(1)$, and that $R(k) < R(k-1)$ implies $R(k+1) < R(k)$, from which we conclude that R is decreasing. R must then have a limit, α say. This limit must satisfy $\alpha + \alpha^{-1} = (\kappa - b')/a'$. Since for every $\alpha \in \mathbb{R}^+$, $\alpha + \alpha^{-1} \geq 2$, we have that $2 \leq (\kappa - b')/a'$. Since $b' = 1 - 2a'$, it follows that $\kappa \geq 1$, which gives the required contradiction.

So far we have concluded that $\lim_{N\to\infty}\lambda_N/N = \ell$. Since $\lambda_N = \rho(\phi_N)(M(\phi'_N) + (\ell-1)N)$ and $\rho(\phi_N) \leq 1$, it follows that $\lim_{N\to\infty} M(\phi'_N)/N = 1$. For now, let $\epsilon_N = 1 - M(\phi'_N)/N$. Since $(\phi^t)' \preceq (\vartheta^t)'$, we also know that $\liminf_t M((\vartheta^t)')/N \geq 1 - \epsilon_N$.

The second step is to show that $\vartheta^t(1)/|\vartheta^t|$ is small for large t. Since $\phi^t \preceq \vartheta^t$, we know that $\phi^t(1)/|\phi^t| \geq \vartheta^t(1)/|\vartheta^t|$, so it is enough to show that once normalised $\phi_N(1)$ is small for large N. This time we define:

$$R(k) = \lim_{N\to\infty} \eta_N(1).$$

We have that:

$$R(1) = \frac{\ell/(\ell-1) - b'}{a'} \quad \text{and} \quad R(k+1) = \frac{\ell/(\ell-1) - b'}{a'} - R(k)^{-1}.$$

Since $(\ell/(\ell-1) - b')/a' > (1-b')/a' = 2$ it follows inductively that $R(k) > 1$ for all k. This means that for every k, there exists N large enough such that $\eta_{N'}(x) > 1$ for all $N' \geq N$ and all $x \leq k$. Redefine ϵ_N to be the maximum between the value ϵ_N specified in the above and $1/k$ for the largest k such that $\eta_N(x) > 1$ for all $x \leq k$. It follows that for that for all N, $\phi_N(1) \leq \epsilon_N$ and that the sequence ϵ_N converges to 0.

The third step is to consider $\vartheta^t(N)$. This time we set $K = 0$ and consider the new corresponding sequence ϕ^t, with the new limit ϕ_N. We now have that $\vartheta^t \preceq \phi^t$, and hence that

$\vartheta^t(N)/|\vartheta^t| \leq \phi^t(N)/|\phi^t|$. This time, for each x we define:

$$R(x) = \lim_{N\in\Pi} \eta_N(N-x)^{-1}.$$

By the same argument as above we get that $\lim_{N\to\infty}\lambda_N/(N+K) = 1$, and in this case this means that $\lim_{N\to\infty}\lambda_N/N = 1$. R now satisfies:

$$R(1) = \frac{1-b'}{a'} = 2 \quad \text{and} \quad R(x+1) = 2 - R(x)^{-1}.$$

Again we have that $R(k) > 1$ for all k, which means that for sufficiently large N, $\eta_N(N-x) < 1$ for all $x \leq k$. We can therefore redefine ϵ_N so that this sequence still converges to 0 and:

$$\liminf_t \frac{\rho(\vartheta^t)(M((\vartheta^t)') + K^t)}{N + K^t} \geq 1 - \epsilon_N,$$

as needed for Lemma 28.

7.0.6. *The proof of Lemma 31.* In this section we prove the last lemma required to complete the proof of Theorem 20. Roughly speaking, Lemma 31 asserts that the various W_i's (for varying i) eventually stay relatively close to each other, even if they are initially quite different. In simulations we have observed that in fact all of the W_i's converge to the same value M/ℓ (see for instance Figure 5), but this seems to be hard to prove. Instead, we prove that $N + \hat{W}_i^t$ becomes close to $N + M^t(\ell-1)/\ell$ in geometric mean, which is enough for our purposes.

Let us begin by looking at a 2-locus *ac*-sex process where selection acts with an additive value K^t at stage t. More formally, let $\{K^t : t \in \omega\}$ be a sequence of numbers in $[0, (\ell-2)N]$, let $\upsilon_0^0, \upsilon_1^0 \in \mathbb{R}^N$ be the initial distributions corresponding to each of those two loci, and define:

$$\upsilon_0^{t+1} = \mathtt{Sel}_{M((\upsilon_1^t)')+K^t}\mathtt{Mut}_{ac,1} \cdot \upsilon_0^t \quad \text{and} \quad \upsilon_1^{t+1} = \mathtt{Sel}_{M((\upsilon_0^t)')+K^t}\mathtt{Mut}_{ac,1} \cdot \upsilon_1^t.$$

Notice how the the M-value used in defining selection at a given locus is the mean corresponding to the other locus (as it should be for the 2-locus sex process).

Lemma 32. *There is a sequence $\{\delta^t : t \in \omega\}$ such that each $\delta^t > 1$, with $\overline{\mathrm{GM}}_{t\to\infty}\,\delta^t < 1 + \epsilon_N$ and such that for all t:*

$$\frac{1}{\delta^t} \leq \frac{N + M((\upsilon_1^t)') + K^t}{N + M((\upsilon_0^t)') + K^t} \leq \delta^t.$$

Proof. Let $\phi_0^0, \phi_1^0 \in \mathbb{R}^N$ be new initial populations, such that ϕ_0^0 is the probability distribution with $\phi_0^0(1) = 1$ and ϕ_1^0 is the probability distribution with $\phi_1^0(N) = 1$. Let the ϕ_0^t and ϕ_1^t processes evolve as follows:

$$\phi_0^{t+1} = \mathtt{Sel}_{M((\phi_1^t)')+K^t}\mathtt{Mut}_{ac,1} \cdot \phi_0^t \quad \text{and} \quad \phi_1^{t+1} = \mathtt{Sel}_{M((\phi_0^t)')+K^t}\mathtt{Mut}_{ac,1} \cdot \phi_1^t.$$

Consider the translations of ϕ_0 and ϕ_1 to the $a'a'$-process: i.e., let $\vartheta_0^t = \tau^{-t}C \cdot \phi_0^t$ and $\vartheta_1^t = \tau^{-t}C \cdot \phi_1^t$. From Lemmas 27 and 28 we get that:

$$1 - \epsilon_N < \frac{\mathrm{GM}(\rho(\phi_1^t)(M((\phi_1^t)') + M((\phi_0^t)') + K^t)}{\tau\,\mathrm{GM}(N + M((\phi_0^t)') + K^t)} < 1 + \epsilon_N,$$

and

$$1-\epsilon_N < \frac{\mathrm{GM}(\rho(\phi_0^t)(M((\phi_0^t)')+M((\phi_1^t)')+K^t)}{\tau\,\mathrm{GM}(N+M((\phi_1^t)')+K^t)} < 1+\epsilon_N.$$

Taking the quotient of these equations we conclude that we can redefine the sequence ϵ_N so that it still converges to 0, and so that:

$$(1-\epsilon_N) < \mathrm{GM}\left(\frac{\rho(\phi_1^t)}{\rho(\phi_0^t)}\right)\mathrm{GM}\left(\frac{N+M((\phi_1^t)')+K^t}{N+M((\phi_0^t)')+K^t}\right) < (1+\epsilon_N).$$

From Lemmas 9 and 10, it follows inductively that $\phi_0^t \preceq \upsilon_0^t \preceq \phi_1^t$ and $\phi_0^t \preceq \upsilon_1^t \preceq \phi_1^t$ for every t. We will now use the fact that $\phi_0^t \preceq \phi_1^t$ to establish that the numerators above are essentially greater than the denominators. We know from Lemma 11 that $(\phi_0^t)' \preceq (\phi_1^t)'$ implies $M((\phi_1^t)') \geq M((\phi_0^t)')$. Also, $\phi_0^t \preceq \phi_1^t$ implies that $\phi_1^t(1)/|\phi_1^t| \leq \phi_0^t(1)/|\phi_0^t|$. We know that $\phi_1^t(N)/|\phi_1^t| < \epsilon_N$ from Lemma 30 (with ϵ_N as specified there). We therefore have that:

$$\rho(\phi_1^t) = 1 - a\phi_1^t(1)/|\phi_1^t| - c\phi_1^t(N)/|\phi_N^t| \geq 1 - a\phi_0^t(1)/|\phi_0^t| - c\phi_0^t(N)/|\phi_N^t| - c\epsilon_N = \rho(\phi_0^t) - c\epsilon_N.$$

Since $\rho(\phi_0^t) > 1-a-c = b > c$, it follows that $\rho(\phi_0^t) - c\epsilon_N \geq \rho(\phi_0^t)(1-\epsilon_N)$. We can therefore redefine the ϵ_N so that the sequence still converges to 0 and:

$$1 \leq \mathrm{GM}\left(\frac{N+M((\phi_1^t)')+K^t}{N+M((\phi_0^t)')+K^t}\right) < (1+\epsilon_N).$$

Let δ^t be the term inside the large parentheses, i.e., $\delta^t = \frac{N+M((\phi_1^t)')+K^t}{N+M((\phi_0^t)')+K^t} \geq 1$ and $\mathrm{GM}(\delta_t) < 1+\epsilon_N$. Since $M((\phi_0^t)') \leq M((\upsilon_0^t)') \leq M((\phi_1^t)')$ and $M((\phi_0^t)') \leq M((\upsilon_1^t)') \leq M((\phi_1^t)')$, we have:

$$\frac{1}{\delta^t} \leq \frac{N+M((\upsilon_1^t)')+K^t}{N+M((\upsilon_0^t)')+K^t} \leq \delta^t. \qquad \square$$

Now let us return to the proof of Lemma 31. For each $i<j$, let $\delta_{i,j}^t$ be the δ^t whose existence is ensured by Lemma 32 for the case $K^t = \hat{W}_{i,j}^t = M^t - W_i^t - W_j^t$. Let $\delta^t = \prod_{i<j\leq\ell}\delta_{i,j}^t$. We have that $\mathrm{GM}(\delta^t) < (1+\epsilon_N)^{\ell(\ell-1)}$ and that for every t and $i \neq j$:

$$\frac{N+\hat{W}_i^t}{N+\hat{W}_j^t} < \delta^t.$$

Since $\sum_{i=1}^{\ell}(N+\hat{W}_i^t) = \ell N + (\ell-1)M^t$, it follows that for some j, $N+\hat{W}_j^t > N + M^t(\ell-1)/\ell$. Therefore, for every i, $N+\hat{W}_i^t > (N+M^t(\ell-1)/\ell)/\delta^t$. A similar argument shows that $N+\hat{W}_i^t < (N+M^t(\ell-1)/\ell)\delta^t$, which completes the proof of Lemma 31.

8. Appendix D: Variants of the model

In this subsection we briefly consider variants of the model for which populations may be finite or infinite and fitnesses contributions may be additive or multiplicative. For the most part, our analysis here will rely on the results of simulations, although we shall also be able to draw some concrete conclusions concerning fundamental similarities and differences between the models. First, let us describe these variants.

8.0.1. *The finite model.* For the finite model we consider an extra parameter $P \in \mathbb{N}$, which determines the size of the population. This size is then fixed through the generations, so that a population always consists of P vectors, $\boldsymbol{x}_1, ..., \boldsymbol{x}_P$, in $\mathbb{Z}^\ell$. Let us consider the sex process first. In order to apply selection one chooses $2P$ individuals, by sampling independently from the population with replacement: if M is the mean fitness of the population and $F(\boldsymbol{x})$ is the fitness of individual $\boldsymbol{x}$, then the probability that individual $\boldsymbol{x}$ is chosen for the nth sample $(1 \leq n \leq 2P)$ is $F(\boldsymbol{x})/M$. One may consider the parent generation as forming a pool of gametes. The probability that a gamete chosen uniformly at random from this pool comes from a given individual $\boldsymbol{x}$, is proportional to the fitness of $\boldsymbol{x}$. During the selection phase we are choosing P many pairs of individuals from which gametes are taken (recombination later being applied to each of these pairs). To apply the mutation operation, we take in turn each individual from the P-many pairs chosen during selection, and for each locus we change its fitness value by -1, 0 or 1 with probabilities $\mu(-1)$, $\mu(0)$ and $\mu(1)$ respectively. To apply recombination, we take each of the P pairs resulting from mutation in turn. Suppose that the nth pair is $\boldsymbol{x}_n^1 = (x_{n,1}^1, ..., x_{n,\ell}^1)$ and $\boldsymbol{x}_n^2 = (x_{n,1}^2, ..., x_{n,\ell}^2)$. Then we form $\boldsymbol{x}_n^*$ which is the nth member of the next generation by taking each locus i in turn and defining either $x_{n,i}^* = x_{n,i}^1$ or $x_{n,i}^* = x_{n,i}^2$, each with equal probability. The assumption of maximum recombination rates might be justified by considering that one is choosing a representative gene from each chromosome, meaning that the monitored genes lie on distinct chromosomes. For the asex process, one proceeds similarly, except that P many individuals rather than pairs are chosen during the selection phase, and the recombination phase is omitted.

The finite model is clearly the most important to understand, and the analysis we have provided for the infinite model provides a good approximation for large populations and over a number of generations which is not too large. As mentioned previously, the equations governing the change in mean fitness and variance due to selection and mutation for the infinite population model would now perfectly describe the *expected* effect of mutation and selection for finite populations, and the finite populations model could be seen simply as a stochastic approximation to the infinite case, were it not for the loss in variance and higher cumulants due to sampling. While the effect of sampling may not be too significant for large populations on a stage by stage basis, long term it will have the effect that the mean fitness no longer tends to infinity over stages. Without providing a rigorous proof, one may reason that this is perhaps unsurprising as follows. Mutation will still have a fixed expected effect on the mean and variance at each stage. For a population $\boldsymbol{\phi}$ with $M = M(\boldsymbol{\phi})$, $V_F = V_F(\boldsymbol{\phi})$ and $\kappa_3 = \kappa_3(\boldsymbol{\phi})$, however, while the expected effect of selection on the mean is just as for the infinite population model, the expected effect of selection on variance is now:

$$\left(\frac{\kappa_3}{M} - \left(\frac{V_F}{M}\right)^2\right)\frac{P-1}{P} - \frac{V_F}{P}.$$

Now the ratios between cumulants will not tend to increase without limit (in the infinite populations model simulations show these ratios converging to fixed values over time, and such behaviour is also approximated for the finite model). Thus, if variance was to increase

without limit, selection would soon produce decreases in variance outweighing any increases given by mutation. A similar analysis can be made including the effect of recombination, establishing that for sufficiently large variances, the effect of sampling will outweigh any other increases in variance.

8.0.2. *The multiplicative model.* This model is defined exactly like the additive one with the sole difference that the fitness of an individual is calculated multiplicatively, i.e., $F(X) = \prod_{i=1}^{\ell} X_i$. For the infinite case, the sex and asex processes now behave identically, given populations initially at linkage equilibrium. This was initially observed by Maynard-Smith [34].

Lemma 33. *Multiplicative selection preserves linkage equilibrium.*

Proof. Suppose that $\boldsymbol{\phi}$ is a population at linkage equilibrium. Let $X_1, ..., X_\ell$ be random variables distributed according to $\boldsymbol{\phi}$, and let $X_1^*, ..., X_\ell^*$ be distributed according to $\boldsymbol{\phi}^* = \mathtt{Sel}(\boldsymbol{\phi})$. We show that selection maintains independence between the first two loci, as the general result is very similar. We must show that whenever $\mathbf{P}(X_1^* = m_1) \neq 0$ and $\mathbf{P}(X_1^* = m_2) \neq 0$:

$$\mathbf{P}(X_2^* = n |\ X_1^* = m_1) = \mathbf{P}(X_2^* = n |\ X_1^* = m_2).$$

This is equivalent to:

$$\frac{\mathbf{P}(X_1^* = m_1\ \wedge X_2^* = n)}{\mathbf{P}(X_1^* = m_1)} = \frac{\mathbf{P}(X_1^* = m_2\ \wedge X_2^* = n)}{\mathbf{P}(X_1^* = m_2)}.$$

Now, since selection acts according to multiplicative fitnesses:

$$\frac{\mathbf{P}(X_1^* = m_1)}{\mathbf{P}(X_1^* = m_2)} = \frac{m_1}{m_2}\frac{\mathbf{P}(X_1 = m_1)}{\mathbf{P}(X_1 = m_2)}.$$

Also:

$$\frac{\mathbf{P}(X_1^* = m_1\ \wedge X_2^* = n)}{\mathbf{P}(X_1^* = m_2\ \wedge X_2^* = n)} = \frac{nm_1}{nm_2}\frac{\mathbf{P}(X_1 = m_1\ \wedge X_2 = n)}{\mathbf{P}(X_1 = m_2\ \wedge X_2 = n)},$$

so the result follows from linkage equilibrium for $\boldsymbol{\phi}$. □

Thus, if a population begins at linkage equilibrium, this linkage equilibrium will be preserved throughout all the stages (for the $\mathbb{N}$- and bounded-models). Each application of recombination now has no effect on the population.

For the finite multiplicative model, however, sampling will produce linkage disequilibrium and sex now robustly outperforms asex (as seen in the outcomes of simulations presented in §9). For an insightful analysis of mechanisms which may allow negative LD_2 to build up in this context see [19].

8.1. **The simulations.** For a small number of loci ℓ one can implement the algorithms described directly. If one wishes to deal with a larger number of loci for the infinite population sex process then one can achieve more efficient simulations (for the $\mathbb{N}$ and bounded models,

and which will give only tiny margins of error due to truncation issues for the $\mathbb{Z}$ model), by making use of Lemma 6, which allows one to track the entire population by monitoring each locus separately. Similarly, one can achieve more efficient simulations for the asex infinite population process (for the $\mathbb{Z}$-model, and with only tiny margins of error for the other domains) by monitoring only the distribution on the total fitness of individuals. For finite populations, such mechanisms for improving efficiency are not generally necessary (or indeed possible). In considering the unbounded infinite populations models, of course one can only deal with a bounded domain in practice. One is therefore limited in the number of generations which can be simulated. To make the computations more precise for the infinite bounded model, we represented real numbers by their logarithms, as the values of the probability distribution at the upper and lower bounds are extremely small.

9. Appendix E: Extended Data

Figures 6 and 7 display the outcome of simulations for the additive finite populations model. Figures 8 and 9 display the outcome of simulations for the additive infinite populations model. Figure 10 displays the outcome of simulations for the finite populations multiplicative model. Where required for our proofs, we have shown that the proportion of a population at the boundaries will generally be small after sufficiently many generations have passed. The three tables show the proportion of the population at the boundaries for the additive infinite populations $\mathbb{N}$-model and also for the bounded model. All variants of the model are described in Appendix D.

	initial gene fitness =1	10	20
0.2	3.5×10^{-57} / 1.0×10^{-31}	1.1×10^{-57} / 3.2×10^{-32}	3×10^{-58} / 5.6×10^{-33}
0.1	1.6×10^{-70} / 1.0×10^{-39}	3.7×10^{-71} / 1.5×10^{-40}	8.2×10^{-72} / 6.6×10^{-42}
0.05	2.9×10^{-82} / 6.6×10^{-45}	5.7×10^{-83} / 3.6×10^{-46}	1.1×10^{-83} / 2.4×10^{-48}
0.025	2.7×10^{-92} / 1.2×10^{-47}	4.6×10^{-93} / 1.7×10^{-49}	1.0×10^{-93} / 9.4×10^{-53}
0.0125	9.8×10^{-101} / 1.5×10^{-48}	1.5×10^{-101} / 4.2×10^{-51}	4.1×10^{-102} / 8.0×10^{-56}

Table 2. The table concerns the infinite additive $\mathbb{N}$-model, and shows the proportion of a 2-locus population which has fitness contribution 1 at either locus for sex/asex, after 1000 generations, for varying initial gene fitness contributions, and for varying mutation rates. In all cases the probability that a given mutation is beneficial is 10^{-1}.

	$N = 200$	300	400
0.2	3.4×10^{-57} / 3.8×10^{-49}	1.5×10^{-85} / 2.1×10^{-73}	6.6×10^{-114} / 1.2×10^{-97}
0.1	3.5×10^{-100} / 5.0×10^{-95}	3.8×10^{-150} / 2.1×10^{-142}	3.9×10^{-200} / 8.9×10^{-190}
0.05	1.4×10^{-150} / 1.1×10^{-147}	7.5×10^{-226} / 1.8×10^{-221}	3.9×10^{-301} / 2.8×10^{-295}

TABLE 3. The table concerns the infinite additive bounded model, and shows the proportion of a 2-locus population which has fitness contribution 1 at either locus for sex/asex, after 25000 generations, for varying N (maximum allele fitness), and for varying mutation rates. In all cases the probability that a given mutation is beneficial is 10^{-3} and all alleles initially have fitness 50.

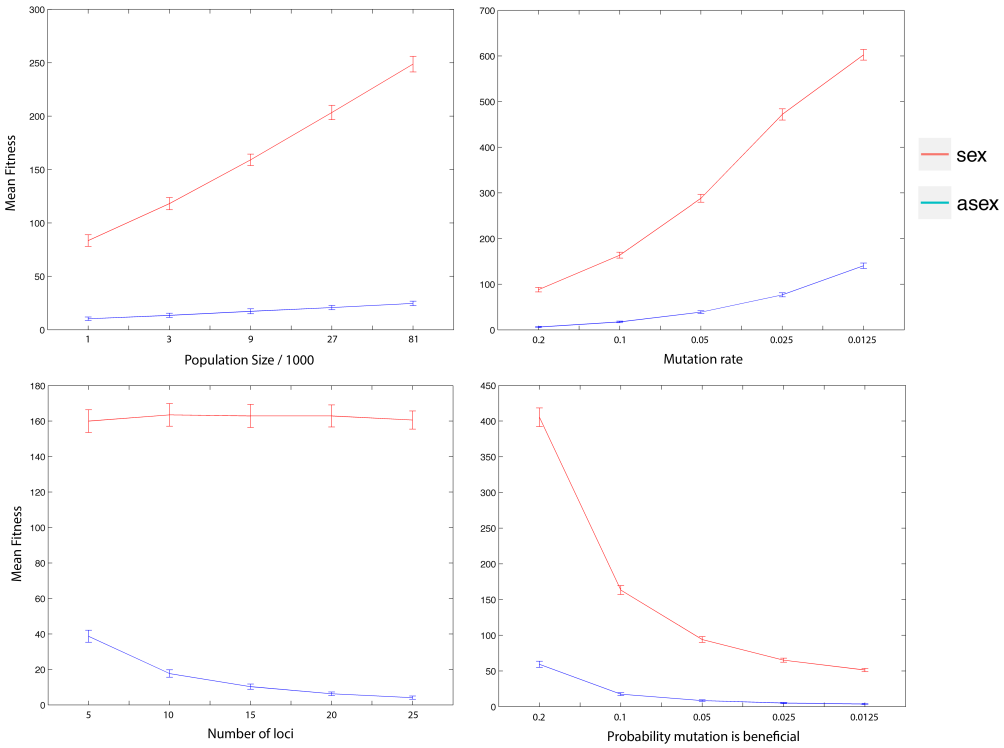

FIGURE 6. Simulations for the finite additive model. In these simulations the 'standard' input parameters were: population size 10000; mutation rate 0.1; probability mutation is positive 0.1; 10 loci, initial gene fitness contribution 5. In each graph one parameter is varied, while the other parameters take the standard values. 100 simulations were run for each parameter set, and the mean fitnesses as well as the standard deviations for these mean fitnesses are depicted, after a number of generations which is sufficient for the mean fitness to stabilise. This number of generations was taken to be 4000, except for the case of varying mutation rate where 20000 generations were run for each simulation.

	$N = 200$	300	400
0.2	6.5×10^{-29} / 1.2×10^{-30}	3.4×10^{-42} / 9.3×10^{-45}	2.6×10^{-55} / 1.1×10^{-58}
0.1	9.2×10^{-16} / 2.5×10^{-16}	7.1×10^{-23} / 1.1×10^{-23}	7.5×10^{-30} / 6.3×10^{-31}
0.05	1.5×10^{-8} / 1.0×10^{-8}	2.3×10^{-12} / 1.3×10^{-12}	4.5×10^{-16} / 2.2×10^{-16}

TABLE 4. The table concerns the infinite additive bounded model, and shows the proportion of a 2-locus population which has fitness contribution N (maximum allele fitness contribution) at either locus for sex/asex, after 25000 generations, for varying N and for varying mutation rates. In all cases the probability that a given mutation is beneficial is 10^{-3} and all alleles initially have fitness 50.

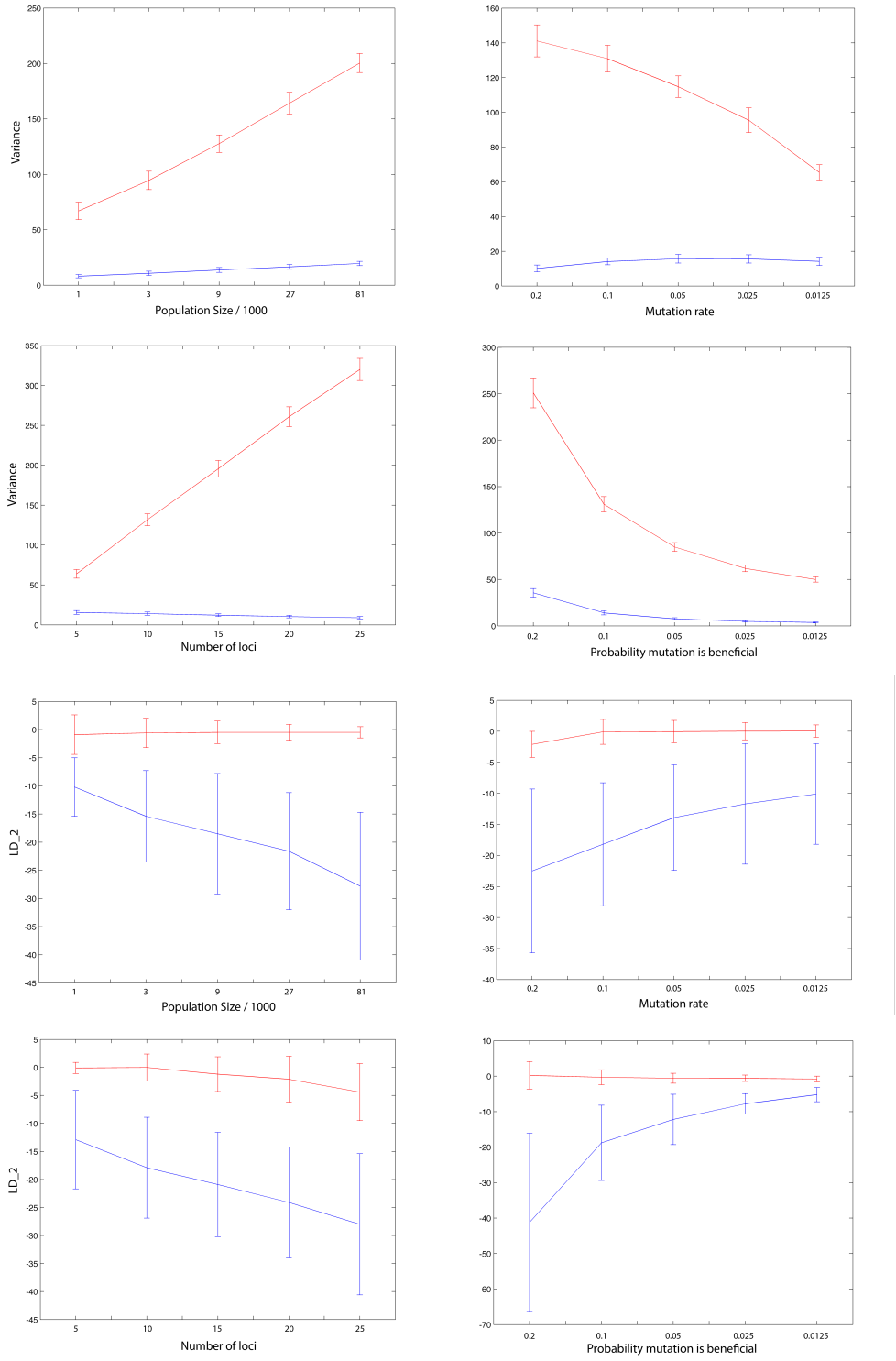

FIGURE 7. These graphs display variance and LD_2 for the same simulations which have their mean fitnesses displayed in Figure 6.

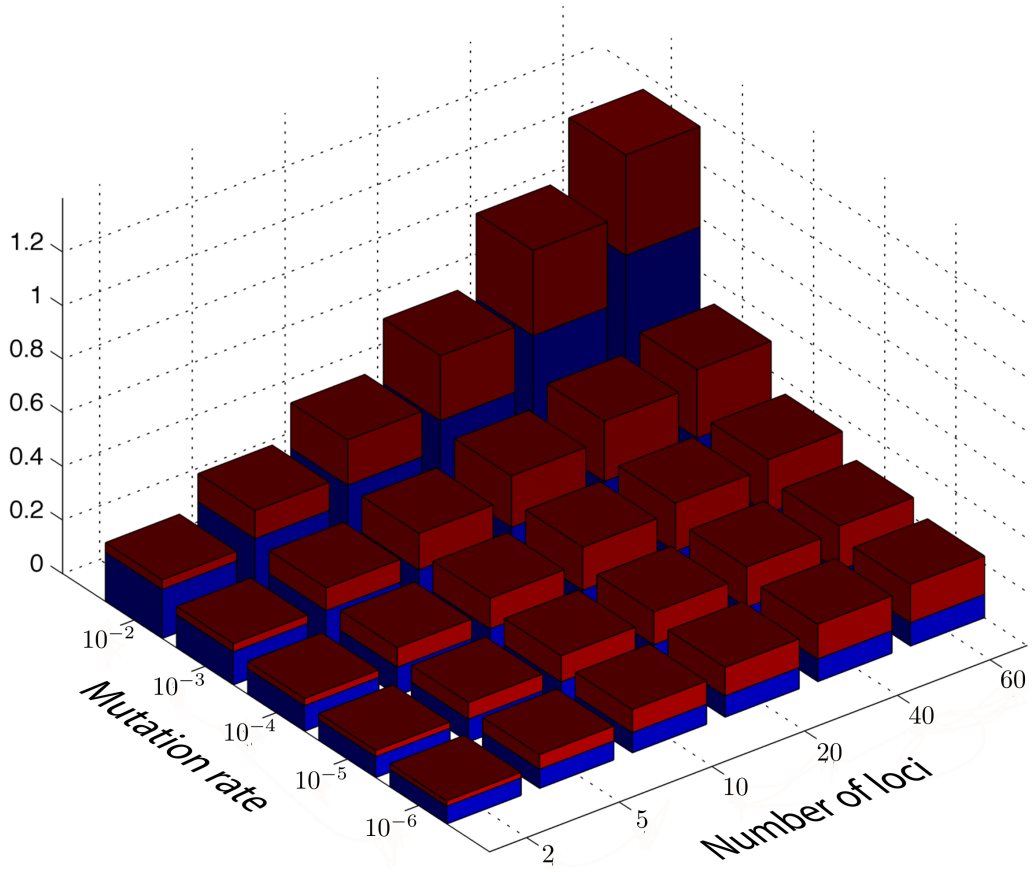

FIGURE 8. Simulations for the infinite additive model appear to show V_F/M reaching a limit value over time. The figure shows approximate values for these limits, for sex (red) and asex (blue). In all these simulations the probability that a given mutation is beneficial was fixed at 0.1, and gene fitness contributions were initially 5 (although the latter parameter has no effect on the limit values found).

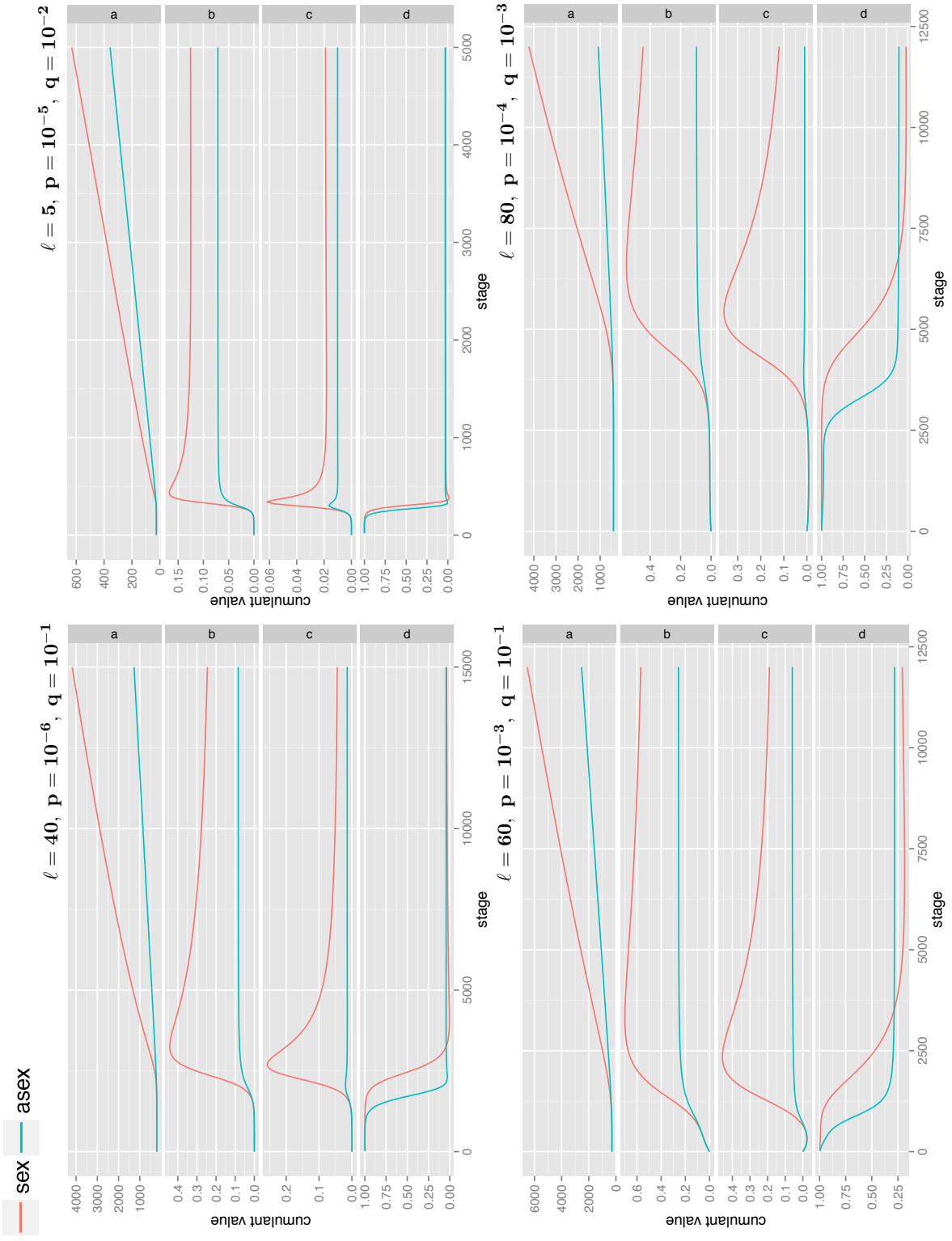

FIGURE 9. Simulations for the additive infinite model. Each set of four graphs shows (a) M, (b) V_F/M, (c) κ_3/M, (d) κ_4/V_F for one simulation. In all simulations initial gene fitness contributions are 5, and each simulation is then specified by a triple: ℓ specifies the number of loci, p is the probability of mutation, q is the probability a given mutation is beneficial.

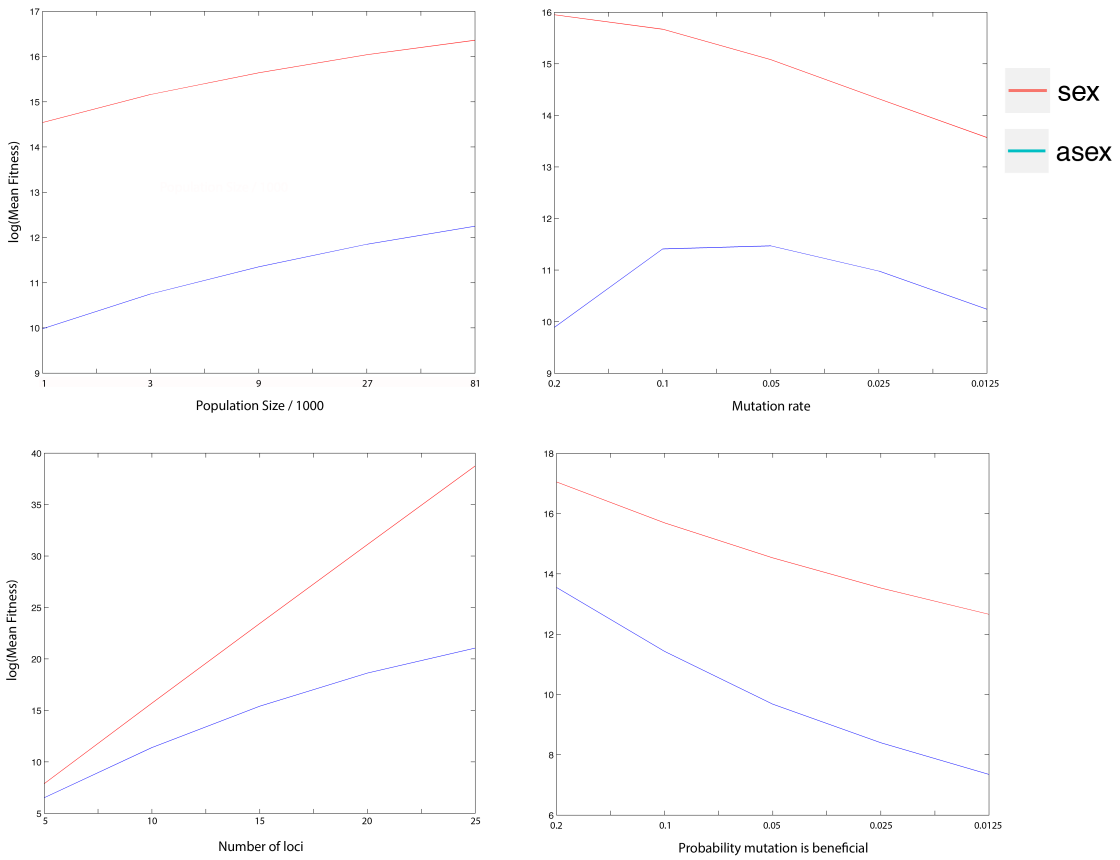

FIGURE 10. Simulations for the finite multiplicative model. In these simulations the 'standard' input parameters were: population size 10000; mutation rate 0.1; probability mutation is positive 0.1; 10 loci, initial gene fitness contribution 1. In each graph one parameter is varied, while the other parameters take the standard values. 100 simulations were run for each parameter set, and the logarithms (base 10) of the mean fitnesses are depicted after 500 generations (without any suggestion that the mean fitness has stabilised by this point).