When does frequency-independent selection maintain genetic variation?

Sebastian Novak[*,1] and Nicholas H. Barton[*]

[*] *Institute of Science and Technology Austria, Am Campus 1, 3400 Klosterneuburg, Austria*

Abstract

Frequency-independent selection is generally considered as a force that acts to reduce the genetic variation in evolving populations, yet rigorous arguments for this idea are scarce. When selection fluctuates in time, it is unclear whether frequency-independent selection may maintain genetic polymorphism without invoking additional mechanisms. We show that constant frequency-independent selection with arbitrary epistasis on a well-mixed haploid population eliminates genetic variation if we assume linkage equilibrium between alleles. To this end, we introduce the notion of frequency-independent selection at the level of alleles, which is sufficient to prove our claim and contains the notion of frequency-independent selection on haploids. When selection and recombination are weak but of the same order, there may be strong linkage disequilibrium; numerical calculations show that stable equilibria are highly unlikely. Using the example of a diallelic two-locus model, we then demonstrate that frequency-independent selection that fluctuates in time can maintain stable polymorphism if linkage disequilibrium changes its sign periodically. We put our findings in the context of results from the existing literature and point out those scenarios in which the possible role of frequency-independent selection in maintaining genetic variation remains unclear.

1 Introduction

There is a general understanding in the population genetics community that constant frequency-independent selection on a well-mixed haploid population eliminates genetic

[1] Corresponding author address: Institute of Science and Technology Austria, Am Campus 1, 3400 Klosterneuburg, Austria. Email: sebastian.novak@ist.ac.at.

variation. Selection is called frequency-independent if the fitness of a genotype does not depend on the frequencies of the other genotypes in the population. If constant selection acts on a haploid population, each haplotype (haploid genotype) has an externally determined fitness value. Hence, in the absence of recombination and mutation, it is clear that the haplotype with the highest fitness replaces the other haplotypes (Bürger, 2000, Chapter I.9). With multiple recombining loci and epistasis between alleles, it is much less evident that frequency-independent selection must reduce the population to a single haplotype. Even more, it remains an open task to identify and characterize the conditions under which frequency-independent selection that fluctuates in time may maintain genetic variation.

The matter of debate concerns an idealized population, under the sole influence of selection and recombination. Understanding the action of natural selection is crucial and justifies the analysis of pure selection models (c.f. Bürger, 2000, and references therein). We therefore neglect direct sources of genetic variation (mutation), spatial heterogeneities in selection that may protect alleles in spatial refuges (Strobeck, 1979), and factors generating frequency-dependent fitness, e.g., fluctuating selection in combination with overlapping generations (Ellner and Sasaki, 1996), or with explicit population regulation (Dean, 2005) and other ecological mechanisms. Similarly, a recent study (Gulisija *et al.*, 2016) showed that the presence of a modifier of phenotypic plasticity may shelter alleles at selected loci from adverse conditions (termed a "genomic storage effect") and thus lead to a protected polymorphism under temporally varying conditions. Here, however, we focus directly on fitness, as the only trait being modulated. Furthermore, we consider haploid populations, since the random combination of gametes into diploid genotypes may induce frequency-dependent selection on alleles, even if the fitnesses of the diploid genotypes are frequency-independent: with heterozygote advantage, even constant selection on diploids is capable of maintaining genetic variation. Similar mechanisms apply for diploid systems under fluctuating selection, which can lead to effective heterozygote advantage in terms of geometric mean fitnesses over time (Haldane and Jayakar, 1963; Hoekstra, 1975; Nagylaki, 1975).

Mathematically precise arguments for the erosion of genetic variation under constant frequency-independent selection are scarce. Various publications cover special cases of fixed numbers of loci and alleles (e.g., Appendix S.5 by Bank *et al.*, 2012). For arbitrary numbers of loci and alleles, Kirzhner and Lyubich (1997) showed that constant selection on haploids leads to genetic monomorphism in two extreme cases: strong linkage between

loci allowing for epistasis between genes, and arbitrary linkage between loci, but assuming that contributions of individual alleles to fitness are additive. More recently, MEHTA *et al.* (2015) showed that the discrete replicator dynamics of coordination games always converges to pure Nash equilibria. It follows as a special case that constant haploid selection does not support stable polymorphism under the assumption of linkage equilibrium between loci. However, their intricate proof is cast in the framework of game theory and computer science, and thus may not be accessible to readers with a background in evolutionary biology.

Under temporally fluctuating frequency-independent selection on haploids, KIRZHNER *et al.* (1994) showed numerically that stable polymorphism can be maintained with two di-allelic loci, at least for a very narrow range of parameters. Their exemplary gene-frequency dynamics with periodic fluctuations exhibit cyclically converging trajectories, whose period is typically much longer than the period of the selection coefficients. Their examples require strong selection such that gene frequencies change drastically between successive generations. Thus, the parameters needed for maintaining polymorphism may not be representative for natural populations, where frequency changes due to selection are typically small and selection is weaker than recombination, such that the latter efficiently breaks up associations between loci (NAGYLAKI, 1974, 1993). In particular, it has not yet been shown if fluctuating selection may maintain genetic variation in the continuous-time limit, where selection and recombination are both weak.

We consider a well-mixed haploid population with non-overlapping generations and frequency-independent selection on multiple loci, each with any number of alleles. We allow arbitrary epistasis between alleles. This includes non-linear selection on polygenic traits, where epistasis in fitness may emerge even if the underlying genes contribute additively to the trait. First, we present a simple yet rigorous proof that at linkage equilibrium, no genetic variation can be maintained by constant frequency-independent selection alone. We note that the result does not follow from frequency-independent selection at the level of genotypes as defined above, since frequency-dependence in the fitnesses of alleles may be induced, e.g., in diploid organisms. Instead, we provide a formal criterion for frequency-independent selection at the allelic level that is sufficient for proving the loss of genetic polymorphism. Our argument applies to haploid populations when the alleles are at or close to linkage equilibrium, which is the case when selection is weak relative to recombination (NAGYLAKI, 1993). Next, we show numerically that stable equilibria are highly unlikely when selection and recombination are weak but of the same order, so that there may

be strong linkage. Finally, we provide an example of weak and periodically fluctuating frequency-independent selection and recombination between two loci that maintains genetic variation at a stable equilibrium. We find that linkage disequilibrium is required for stable polymorphism, though its magnitude can be small. We conclude by summarizing and discussing our results, and by laying out the remaining open questions concerning the conundrum of frequency-independent selection and genetic variation.

2 Notation

We assume that there are L genetic loci. Locus i has K_i possible allelic variants and we write $\mathcal{K} = \sum_{i=1}^{L} K_i$ for the total number of alleles across all loci. Denote the k-th allele at locus i by P_k^i and the frequency of allele P_k^i by p_k^i. Furthermore, we write $\mathbf{p}^i$ for the vector of allele frequencies at locus i and collect the allele frequencies at all loci in a vector $\mathbf{p} = (\mathbf{p}^1, ..., \mathbf{p}^L)$.

The mean fitness $\bar{w}$ of the population is obtained by averaging the fitnesses of the genotypes over their frequencies. Similarly, the marginal fitness w_k^i of an allele P_k^i is defined as the mean fitness of all genotypes containing that allele. Consequently, the mean fitness $\bar{w}$ of the population can be seen as an average over the marginal fitnesses of alleles at any given locus,

$$\bar{w} = \sum_{k=1}^{K_i} p_k^i w_k^i \qquad \text{(for all } i = 1, ..., L\text{)}. \tag{1a}$$

Note that the choice of locus i to average across does not play a role here, i.e., the sum evaluates to the same value for any locus i.

We say that selection on haploids is frequency-independent (at the level of alleles) if the marginal fitness of an allele does not depend on the frequencies of alleles at the same locus. That is, for every choice of alleles j and k on any locus i, we have

$$\frac{\partial w_k^i}{\partial p_j^i} = 0. \tag{1b}$$

This definition encompasses the case of frequency-independent selection on haploids, where every haplotype has a fixed fitness value (Appendix A.1). However, its scope of application is broader, since it does not require a certain ploidy level and allows for certain kinds of frequency-dependent interactions between loci.

Importantly, the marginal fitnesses w_k^i of alleles at locus i generally depend on the

allele frequencies at the other loci, leading to non-trivial interactions between alleles. From equations (1a) and (1b), we may derive the familiar identity

$$\frac{\partial \bar{w}}{\partial p_k^i} = w_k^i \tag{2}$$

(c.f. WRIGHT, 1937). Hence, the derivative of mean fitness with respect to an allele yields the marginal fitness of that allele under frequency-independent selection.

3 Results

3.1 Constant frequency-independent selection at linkage equilibrium eliminates genetic variation

Assume that the L loci are at linkage equilibrium at all times. This is a restrictive assumption, but is valid in the limit where recombination breaks up linkage disequilibria between loci quickly, relative to the rate of selection. At linkage equilibrium, the frequency of each haplotype is the product of the frequencies of its constituent alleles. Consequently, it suffices to follow the dynamics of allele frequencies instead of those of the genotypes. We consider the standard selection dynamics from population genetics in discrete time (generations), neglecting all other evolutionary forces (e.g., mutation) and only describing the effect of selection. In particular, there is no genetic drift, i.e., the population is practically infinite and the dynamics are deterministic. The frequency $(p_k^i)'$ of allele k on locus i in the next generation is calculated from the frequencies $\mathbf{p}$ in the previous generation as

$$(p_k^i)' = p_k^i \, \frac{w_k^i}{\bar{w}} = f_k^i(\mathbf{p}) \qquad (i = 1, ..., L;\ k = 1, ..., K_i)\,. \tag{3}$$

To prove that constant frequency-independent selection at linkage equilibrium eliminates genetic variation, we consider the dynamics (3) and assume frequency-independent selection at the level of alleles as defined in equation (1b). Our argument consists of three parts. First, the selection dynamics converge to the set of its equilibrium points, i.e., there is no complicated dynamical behaviour. This result has been established for a wide range of (mostly diploid) selection models, e.g., for additivity (LYUBICH, 1992), weak epistasis, and weak selection (NAGYLAKI *et al.*, 1999). It is a consequence of the mean fitness increasing steadily under selection, which (assuming linkage equilibrium) is implied by Fisher's Fun-

damental Theorem of Natural Selection (Fisher, 1930) and by Wright's selection gradient formula (Wright, 1937). Note that fitness may decrease when recombination breaks up linkage disequilibria (Bürger, 2000, Chapter II.1). For our specific setting with linkage equilibrium, however, a commonly used argument that applies the inequality of Baum and Eagon (1967) (a method noted by Nagylaki, 1977) shows that all trajectories of the dynamics (3) converge to the set of their equilibria (Appendix A.2). Since the haploid selection dynamics can be seen as a special case of the diploid dynamics (c.f. Bürger, 2000, p.30), the proof may be extended to loose linkage, analogous to Nagylaki *et al.* (1999).

Second, we establish that every polymorphic equilibrium of the selection dynamics is generically unstable. A polymorphic equilibrium of equation (3) is a vector $\hat{\mathbf{p}} = (\hat{\mathbf{p}}^1, ..., \hat{\mathbf{p}}^L)$ of only non-zero entries that satisfy $f_k^i(\hat{\mathbf{p}}) = \hat{p}_k^i > 0$. From standard linear stability analysis, the stability of a polymorphic equilibrium $\hat{\mathbf{p}}$ is determined by the eigenvalues of the Jacobian matrix $J(\hat{\mathbf{p}})$ (see Appendix A.3 for details). If all eigenvalues of $J(\hat{\mathbf{p}})$ have a modulus less than one, the equilibrium $\hat{\mathbf{p}}$ is asymptotically stable; if the modulus of a single eigenvalue is greater than one, it is unstable. In the case when the modulus of the largest eigenvalue equals one, further analysis would be needed. Neglecting this degenerate case, a simple dimensionality argument shows that any polymorphic equilibria of the dynamics (3) under frequency-independent selection, equation (1b), is unstable (Theorem A.1 in Appendix A.3).

Due to the first two parts, all trajectories converge to the boundary of the state space of the dynamics (3). Hence, the selection dynamics cannot maintain all alleles in the population; some of them are lost and the set of possible genotypes is reduced. Thus third, we repeat the argument for the remaining genotypes by restricting the system to the loci that are still polymorphic, thereby iteratively eliminating alleles from the population. This procedure ends when all loci have become monomorphic, i.e., when only a single genotype is left. Therefore, in the absence of mechanisms creating variation, constant frequency-independent selection eventually deprives the population of all its genetic variation.

3.2 Constant frequency-independent selection with weak selection and recombination rates eliminates variation

We have shown that when selection is much weaker than recombination, so that the population is close to linkage equilibrium, constant frequency-independent selection cannot maintain variation. At the opposite extreme, when selection is much stronger than re-

# loci	λ_{min}	λ_{mean}	λ_{max}	ν_{min}	ν_{mean}	ν_{max}	# replicates
2	0.0000035	0.189	0.497	1	1	1	10^5
3	0.0338	0.374	0.886	1	1.61	2	10^4
4	0.130	0.507	1.108	1	2.36	4	10^4
5	0.207	0.593	1.267	2	3.21	5	10^4
6	0.365	0.638	1.049	3	4.1	5	10^3

Table 1: The minimum, mean, and maximum of the leading eigenvalue, λ, and of the number of positive eigenvalues, ν. Values are scaled relative to the rate of recombination between adjacent loci, which is assumed the same between all adjacent genes.

combination, we know that the fittest haplotype will fix. In the intermediate case, where epistasis and recombination are of the same order, can constant frequency independent selection maintain variation? To simplify matters, we assume that selection and recombination are weak, but of the same order, so that we can assume continuous time.

Appendix A.4 gives an expression for the Jacobian of the continuous-time dynamics in terms of the equilibrium genotype frequencies and the selection coefficients for each haplotype. Crucially, the equilibrium frequencies uniquely determine the selection coefficients, which are proportional to the recombination rates. Necessarily, linkage equilibrium implies neutrality: both epistasis and directional selection on alleles must be zero. Linkage disequilibrium requires epistasis, and directional selection must be set to the value that keeps allele frequencies constant.

We have not found a way to show that the Jacobian necessarily and in general implies instability. However, we have calculated eigenvalues numerically for random genotype frequencies, for up to 6 loci. Table A.1 summarises the leading eigenvalue, and the number of positive eigenvalues, for $L = 2, ..., 6$ loci; calculations become prohibitively slow for more than 6 loci, since they involve calculating a $2^L \times 2^L$ matrix. Results are based on drawing genotype frequencies uniformly from the space of genotype frequencies. This typically implies strong linkage disequilibrium, and hence epistasis of the same order as the recombination rate. Moreover, if some genotype is by chance very rare, the corresponding selection coefficient is very negative. The leading eigenvalue λ is typically of the same order as the recombination rate between adjacent loci, implying that variation is lost over the same time scale as recombination. However, with two loci the leading eigenvalue may by chance be extremely small, in which case variation would be lost extremely slowly. These results show that stable equilibria are at best rare.

3.3 Weak periodically fluctuating frequency-independent selection may maintain stable polymorphism in a two-locus model

3.3.1 Specification and intuition

We now drop the assumption of linkage equilibrium and consider two linked loci with two alleles each. The recombination rate between the two loci is $r > 0$. For convenience, we denote the alleles at the first locus by $P_1^1 = A$ and $P_2^1 = a$, and at the second locus by $P_1^2 = B$ and $P_2^2 = b$. Thus, we have four possible genotypes with frequencies p_{AB}, p_{Ab}, p_{aB}, and p_{ab}. We assume a symmetric fitness model, in which the alleles A and B provide an additive selective advantage $s > 0$ and there is an epistatic coefficient ϵ in genotypes consisting of an upper and a lower case allele, see Table 2.

Instead of following the four genotype frequencies, we may describe the temporal dynamics of selection by the frequencies p_A and p_B of the alleles A and B, and the linkage disequilibrium

$$D = p_{AB}\, p_{ab} - p_{Ab}\, p_{aB}.$$

Due to the symmetry of the parameters, however, it is useful to consider the mean P and difference δ of the upper case allele frequencies, i.e., we write $p_A = P - \delta/2$ and $p_B = P + \delta/2$. We assume that selection and recombination are both weak, i.e., we consider the limit where s, ϵ and r go to zero simultaneously. Then, the dynamics can be approximated by a set of differential equations in continuous time that may be written as (c.f. Appendix A.5.1)

$$\dot{P} = P(1-P)\left(s - \epsilon(1-2P)\right) + \left(D - \frac{\delta^2}{4}\right)\left(s + \epsilon(1-2P)\right), \tag{4a}$$

$$\dot{\delta} = \delta\left[s(1-2P) - \epsilon\left(1 - 2P(1-P) + 2D - \frac{\delta^2}{2}\right)\right], \tag{4b}$$

$$\dot{D} = 2\epsilon\left(P^2 - \frac{\delta^2}{4}\right)\left((1-P)^2 - \frac{\delta^2}{4}\right) - D\left(r - 2s(1-2P) + 2\epsilon D\right). \tag{4c}$$

Genotype	AB	Ab	aB	ab
Frequency	p_{AB}	p_{Ab}	p_{aB}	p_{ab}
Fitness	$w_{AB} = 1 + 2s$	$w_{Ab} = 1 + s - \epsilon$	$w_{aB} = 1 + s - \epsilon$	$w_{ab} = 1$

Table 2: A genotype fitness configuration with an additive fitness increment s and an epistatic coefficient ϵ. Fitness values are normalized with respect to the ab genotype.

Note that the use of all three parameters, s, ϵ and r, is redundant in this limit, since we may always absorb any one of them into the time variable t by proper rescaling. For clarity, however, we explicitly denominate all parameters. Consequently, the dynamics remain unchanged if we replace $s \mapsto \alpha s$, $\epsilon \mapsto \alpha\epsilon$, $r \mapsto \alpha r$, and $t \mapsto \alpha t$, for any $\alpha > 0$.

To further exploit the symmetry in the fitness coefficients we imposed above, we look for equilibria around $\delta \approx 0$. If additionally the alleles are at linkage equilibrium ($D = 0$), the dynamics of P is inverted in time if we replace s and ϵ by $-s$ and $-\epsilon$. Thus, fixing s and ϵ and periodically applying equation (4a) with parameters (s, ϵ) and $(-s, -\epsilon)$ for τ time units each leads to a trajectory that traces itself back to its initial point. This trivially maintains polymorphism, yet in a degenerate way: perturbations tip the balance of forces and lead to the fixation of one of the alleles at each locus. However, as we see below, including the dynamics of linkage disequilibrium D may stabilize trajectories and thus lead to protected polymorphism.

To obtain an intuitive understanding of how stability may come about, assume that it suffices to consider stability in the direction of P only. Then, the stability of a trajectory of P is given by the sign (negative sign implies stability) of

$$\mathbb{E}\left[\frac{\partial \dot{P}}{\partial P}\right] = \mathbb{E}\left[s\left(1-2P\right)\right] - \mathbb{E}\left[\epsilon\left(1-6P(1-P)\right)\right] - 2\mathbb{E}\left[\epsilon D\right] + \mathcal{O}(\delta^2), \tag{5}$$

where the expectation $\mathbb{E}[\cdot]$ is taken across a period of selection of length 2τ. Note that besides the variables P, D, and δ, also the parameters s and ϵ are functions of time. We conclude from this expression that including the dynamics of D explicitly may stabilize trajectories of P if D and ϵ correlate positively ($\mathbb{E}[\varepsilon D] > 0$). Equation (4c) indicates that the latter may be the case at least for small D, when $\dot{D} \propto \epsilon$.

3.3.2 Stable polymorphism in the limit of rapid fluctuations

We assume that selection fluctuates through the course of two seasons of equal length τ. In the first season, the intensities of selection and epistasis are given by constant s and ϵ, see Table 2. In the second season, selection and epistasis simultaneously change their signs, yet maintain their moduli. We further assume $\delta \approx 0$; the consistency of this assumption with our results will be affirmed below. As a consequence, we may neglect terms containing δ^2 in equations (4). Thus, the equations for P and D become independent of δ.

Finally, we assume quasi-linkage equilibrium, i.e., that $D = \hat{D}(P)$ is at its equilibrium given P at all times. This quasi-steady state assumption is only approximate, but the

absolute error in equations (4a)–(4b) is likely to be small, at least if D is small itself. By letting the recombination rate r be large relative to s and ϵ, the latter is satisfied and we may neglect terms containing D^2. From equation (4c), we then have

$$\hat{D}(P) = \frac{2\epsilon P^2(1-P)^2}{r - 2s + 2sP} \approx \frac{2\epsilon P^2(1-P)^2}{r}, \tag{6}$$

for $r \gg s, \epsilon$. Hence in particular, D correlates positively with ϵ.

We now iterate the two seasons as described in Table 3 with a very short duration, $\tau = \Delta t/2$ each, so that we may approximate the solution to the differential equations (4) by a single step of the Euler method. For season 1 and the equation for P, this means

$$P\left(t + \frac{\Delta t}{2}\right) = F_{(s,\epsilon)}(P(t)) = P(t) + \frac{\Delta t}{2}\,\tilde{F}_{(s,\epsilon)}\left(P(t)\right),$$

where $\tilde{F}_{(s,\epsilon)}(P) = P\,(1-P)\,(s + \epsilon(1-2P)) - \hat{D}(P)\,(s + \epsilon\,(1-2P))$. Analogously, the expression for season 2 is $P(t + \Delta t/2) = F_{(-s,-\epsilon)}(P(t))$. To apply the two seasons in sequence, we thus need to calculate $P(t+\Delta t) = \left(F_{(-s,-\epsilon)} \circ F_{(s,\epsilon)}\right)(P(t))$. Neglecting terms of order $\mathcal{O}(\Delta t^2)$, this yields

$$\Delta P = P(t+\Delta t) - P(t) = \frac{2\epsilon}{r}\,P(t)^2\,(1-P(t))^2\,[\epsilon\,(1-2P(t)) + s]\,\Delta t. \tag{7}$$

Equating this to zero produces a polymorphic equilibrium at $\hat{P} = (\epsilon + s)/(2\epsilon)$, given that $|s/\epsilon| \leq 1$.

Assuming that D is determined by equation (6), we thus have an equilibrium at $P = \hat{P}$ and $\delta = \hat{\delta} = 0$ (the latter is evident from equation (4b)). Since for small δ the equation for P is independent of δ (see above), the Jacobian matrix at this equilibrium becomes triangular, hence its eigenvalues that determine stability can be read from its diagonal. Thus, stability of P and δ can be assessed separately. For P, linearising the right-hand side of equation (7) yields a negative eigenvalue

$$\frac{\partial \Delta P}{\partial P}\Big|_{(\hat{P},\hat{\delta})} = -\frac{\left(\epsilon^2 - s^2\right)^2}{2\epsilon^2 r}\,\Delta t < 0. \tag{8}$$

Deriving a dynamics for $\Delta\delta$ similarly to equation (7) and taking the corresponding derivative shows that the second eigenvalue is negative as well, $(\partial\Delta\delta/\partial\delta)\,|_{(\hat{P},\hat{\delta})} < 0$. Thus, we have a stable polymorphism at $P = \hat{P}$ and $\delta = \hat{\delta} = 0$. In particular, the above assumption

$\delta \approx 0$ is justified.

3.3.3 Stable oscillations in the symmetric case

The previous section shows that genetic variation can be maintained in the limit of rapid alternations of the two seasons, assuming quasi-linkage equilibrium. In this section, we numerically demonstrate that protected polymorphism is possible for the full dynamics, equation (4), with a finite duration of seasons.

Assume that both seasons last for $\tau = 50$ time units, and that s, ϵ, and r in the two seasons are given by Table 3. This conforms to the symmetric situation of the previous section. We iterate the dynamics (4) through multiple seasons for various initial conditions. Initially, δ rapidly decays and converges to $\delta = 0$. The variables P and D enter periodic oscillations around the point $\hat{P} = 0.375$ and $D = 0$ (see Figure 1) as predicted by the above analysis. Their limit may be a global attractor of the system, since it is approached from all initial conditions we tested, see Appendix A.5.2. The correlation between ϵ and D across one cycle of selection can be calculated to be about $\mathbb{E}[\epsilon D] \approx 0.0045$. This is found to be the main contribution to the negativity of expression (5), indicating that these correlations secure the stability of the trajectory.

The stability of the equilibrium trajectory $(P(t), \delta(t), D(t))$ can be confirmed numerically by linearising a suitable Poincaré map, see Appendix A.5.2 for details. Figure 2 shows the leading eigenvalue of this linearisation as a function of the recombination rate r. Its value is a measure of the stability of the limit trajectory; the further it is below one, the larger the perturbations the system can sustain without the trajectory losing its stability. Interestingly, there is an intermediate value of r for which stability is greatest. Since the variation in linkage disequilibrium D scales with r (larger r keeps D closer to zero) this implies that intermediate linkage disequilibrium is most likely to lead to stable polymorphism. When recombination dominates, $r \to \infty$, the alleles remain at linkage equilibrium and polymorphism is maintained in the degenerate way discussed above for $D = 0$ (the leading eigenvalue equals 1). Without recombination ($r = 0$), we observe that the

	Selection	Epistasis	Recombination	Duration
Season 1:	$s = 0.005$	$\epsilon = -0.02$	$r = 0.01$	$\tau = 50$
Season 2:	$s = -0.005$	$\epsilon = 0.02$	$r = 0.01$	$\tau = 50$

Table 3: A symmetric parameter configuration affording a stable trajectory under equations (4) with seasonal fluctuations.

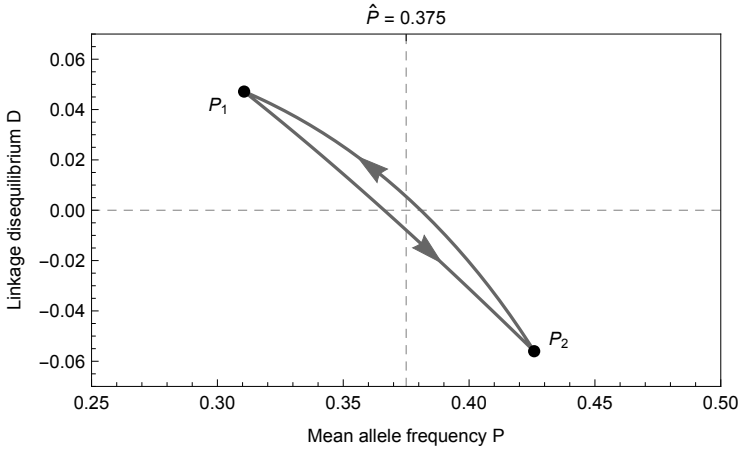

Figure 1: A stable trajectory in the P-D plane under equations (4). Seasonally fluctuating parameters are specified in Table 3. From all tested initial conditions, the dynamics converge to the shown periodic trajectory oscillating around $\hat{P} = 0.375$ with $\delta \approx 0$. In season 1, the dynamics evolve from P_1 to P_2; season 2 maps P_2 back to P_1.

dynamics (4) similarly become symmetric with respect to replacing $(s, \epsilon) \mapsto (-s, -\epsilon)$ and the same logic applies. The latter is in concordance with the literature about fluctuating selection on a single haploid locus that predicts the fixation of the allele with the highest geometric mean fitness over time (Felsenstein, 1976). Thus, intermediate linkage disequilibrium is required to maintain stable polymorphism in our example.

3.3.4 Stable oscillations in asymmetric cases

Our analysis of the leading eigenvalue confirms stability in two ways. First, the periodic trajectory is dynamically stable with respect to perturbations of the variables, e.g. due to genetic drift. Second, the model is structurally stable because of its continuous dependence on the model parameters. This implies that a stable periodic trajectory continues to exist under deviations from our reference parameters, even if they break the symmetry between the two seasons we imposed earlier. In particular, this concerns variation in the selection or epistatic coefficients of one or both seasons (e.g. a systematic bias such that the mean selection coefficient is non-zero), random perturbations to s and ϵ between seasons, and variation in the durations of seasons.

For example, consider two seasons as described in Table 4. Note that we chose a higher recombination rate of $r = 0.05$ to afford reasonably large perturbations of the parameters relative to the previous example (compare Table 3). For initial conditions close to $P = 0.5$, $\delta = 0$, and $D = 0$, the dynamics converge to a stable trajectory shown in Figure 3.

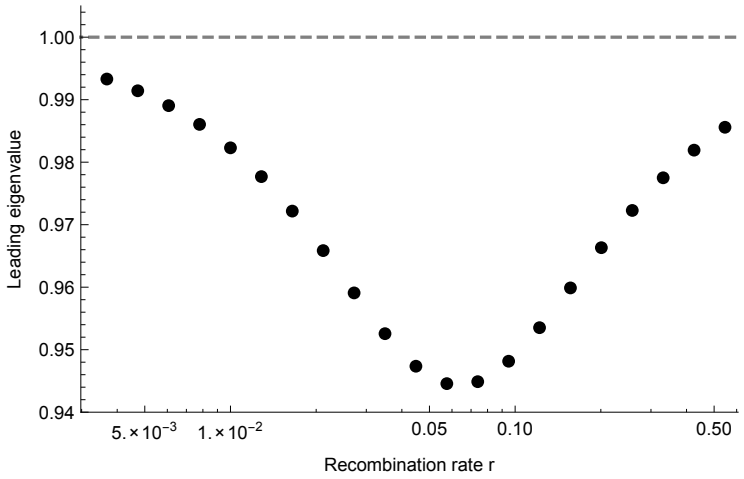

Figure 2: Stability of periodic trajectories under equations (4) with seasonal fluctuations. A periodic orbit corresponds to a fixed point of a suitable Poincaré map (c.f. Appendix A.5.2); we plot the leading eigenvalue of its linearisation as a function of the recombination rate r. Values below one (dashed line) indicate stability, which is greatest around $r \approx 0.06$.

This trajectory can be shown to be locally stable by the same techniques we used in the symmetric case. However, stability is not global; initial conditions with appreciable $\delta \neq 0$ converge towards the monomorphic equilibria given by $P = 0.5$, $\delta = \pm 1$, and $D = 0$. Note that rescaling time during one of the seasons corresponds to rescaling s, ε, and r in this season. Thus, this example also shows that stable polymorphism is possible if the two seasons do not have the same duration.

Stable polymorphism can also be maintained if the symmetry between the two loci is perturbed, i.e., if the two genotypes Ab and aB have different fitness values. This can be parametrized by assuming two epistatic coefficients, $\epsilon_1 \neq \epsilon_2$, for these genotypes. In Appendix A.5.3, we provide a numerical example for stable polymorphism under this asymmetry with substantial differences between ϵ_1 and ϵ_2.

	Selection	Epistasis	Recombination	Duration
Season 1:	$s = 0.00525$	$\epsilon = -0.021$	$r = 0.05$	$\tau = 50$
Season 2:	$s = -0.00475$	$\epsilon = 0.019$	$r = 0.05$	$\tau = 50$

Table 4: An asymmetric parameter configuration affording a stable trajectory under equations (4) with seasonal fluctuations.

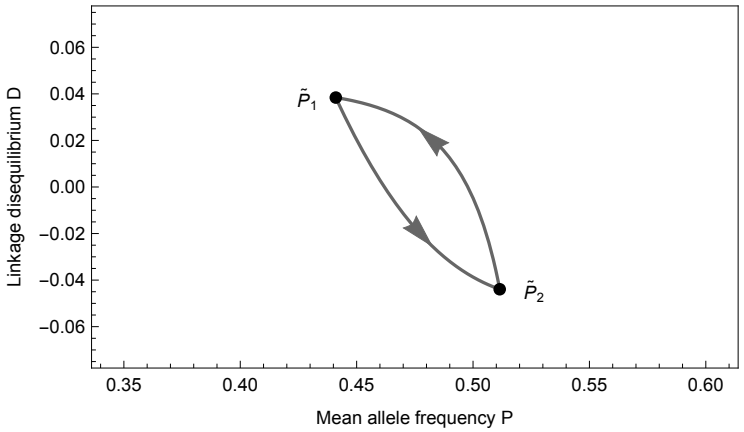

Figure 3: A stable trajectory in the P-D plane with the asymmetric parameter configuration specified in Table 4. We confirmed local stability numerically (c.f. the main text), yet the trajectory is approached only by a subset of tested initial conditions. In season 1, the dynamics evolve from $\tilde{P}_1$ to $\tilde{P}_2$; season 2 maps $\tilde{P}_2$ back to $\tilde{P}_1$.

4 Discussion

When discussing the inability of frequency-independent selection to maintain genetic variation, population geneticists implicitly refer to constant frequency-independent selection at the allelic level or on haploid populations, rather than to the broader definition of frequency-independent selection at the level of genotypes. For example, consider constant overdominance (heterozygote advantage) at a single diploid locus with alleles A and a. In this selection regime, the fitnesses of the three possible genotypes AA, Aa, and aa are typically scaled as 1, $1+hs$, and $1+s$, respectively, where $s>0$ and $h>1$. These values are independent of the genotype frequencies, so that selection is frequency-independent at the level of genotypes. Nevertheless, there is genetic polymorphism at equilibrium under selection; under the dynamics (3), the A allele is maintained at a frequency of $p_A=(h-1)/(2h-1)>0$. The fact that genetic variation is maintained in this example is not surprising, since with random mating it is simply not possible to have a population that consists solely of heterozygotes (unless both homozygotes are non-viable). One may thus argue that it is not selection that maintains genetic variation, but rather, the random pairing of gametes to form diploid organisms. However, since the latter cannot be separated from the dynamics of diploid genotypes, such a discussion is futile. In contrast, the dynamics of allele frequencies, equation (3), does not involve a pairing process. Accordingly, the notion of frequency-independent selection at the allelic level, as introduced by equation (1b), is sufficient to prove the erosion of genetic variation, and contains the

notion of frequency-independent selection on haploids.

In Section 3.1, we show that constant frequency-independent selection on a set of haploid loci at linkage equilibrium eliminates genetic variation. However, it is remains unclear if genetic variation may be stably maintained when recombination and selection are comparable. The case of complete linkage between loci is trivial, and Kirzhner and Lyubich (1997) showed that no stable polymorphism is possible with recombination and additive selection. Our argument covers the complementary case of linkage equilibrium, yet allowing for arbitrary epistasis between alleles. For two diallelic loci under constant selection, a complete analysis was performed by Bank *et al.* (2012), yet the interaction of recombination and epistasis in multi-locus systems is still not well understood.

Trivially, fluctuating selection may maintain polymorphism indefinitely if the direction of selection on two alleles at a single locus changes whenever either allele frequency falls below a threshold value. However, this example is artificial, since the pattern of fluctuations depends on the frequencies of the alleles. In the time-varying framework, frequency-independent selection requires that the change of the fitness functions be independent of the allele frequencies. Temporally fluctuating frequency-independent selection is capable of maintaining genetic polymorphism, as demonstrated by Kirzhner *et al.* (1994). Their examples require strong selection such that gene frequencies undergo large changes between successive generations, which may be considered unrealistic in practice. In Section 3.3, we provide an example of two diallelic loci under weak periodically fluctuating selection and recombination that maintains polymorphism in a stable periodic trajectory. Crucially, including linkage disequilibrium D explicitly is necessary to stabilise the trajectory; under complete linkage or at linkage equilibrium, the effect vanishes.

Heuristically, we may conclude from equation (5) that the dynamics of D may lead to a stable trajectory if they correlate positively with epistasis. This is the case if we approximate D by assuming quasi-linkage equilibrium at all times, when D has the sign of ϵ (see equation (6)). Fluctuating epistasis is thus a key element in the maintenance of genetic polymorphism in our example. Patterns of positive correlation between linkage disequilibrium and epistasis emerge naturally under quasi-linkage equilibrium. Thus, they may be of general importance in maintaining stable polymorphism under frequency-independent selection.

Cycling linkage disequilibrium is also present in a study by Gulisija *et al.* (2016), who demonstrate that stable polymorphism can be maintained in a haploid system with a locus subject to periodic frequency-independent selection and a modifier locus for phenotypic

plasticity. They argue that an allele that promotes phenotypic plasticity reduces the effect of selection on linked alleles, hence may cause a "genomic storage effect" by sheltering deleterious alleles. Their model reduces to a dialellic two-locus selection model with epistasis, in which the genotype fitnesses fluctuate periodically – its fitness structure, however, differs from the example we present above. Their study provides further indication that the explicit process of recombination between loci is vital for maintaining genetic variation; in the limit of high recombination rate (i.e., at linkage equilibrium), their proposed mechanism fails. In some cases, dynamically fluctuating linkage disequilibria may indeed give rise to effective negative frequency dependence over time, and thus to stable polymorphism. Further analyses will be necessary to understand the dynamics of linkage disequilibria and their effect on the persistence of genetic variation.

We have shown that constant frequency-independent selection on haploids, or on diploids in the absence of dominance, cannot maintain variation at linkage equilibrium; our numerical results show that constant and weak selection and recombination very likely do not maintain variation, even when linkage disequilibrium is strong. However, we do show, using a two-locus example, that fluctuating epistasis can maintain variation when recombination is of the same order as the strength of epistasis. This works through a general mechanism, whereby a positive correlation between linkage disequilibrium and epistasis helps stabilise the polymorphism. Since epistasis amongst a set of loci generates linkage disequilibrium amongst those loci, we expect this mechanism to act as a general stabilising process.

However, this process is likely to be weak, especially when epistasis is weak ($\varepsilon \ll 1$). The stabilising force is proportional to $\mathbb{E}[\varepsilon D]$, and since D is itself proportional to epistasis, the stabilising force is of order ε^2. This makes polymorphisms sensitive to biases in directional selection, $\mathbb{E}[s]$; these must be weaker than ε^2 for the polymorphism to be stable. Moreover, if such biases push the equilibrium away from the centre of allele frequency space, the stabilising force becomes much weaker, since $D \sim \varepsilon\, p_A\, q_A\, p_B\, q_B$. A caveat here is that it is not clear how these arguments scale with the numbers of loci: epistatic coefficients become weaker relative to directional, but there are many more of them. An extension to many loci would be desirable, though not easy.

Is fluctuating frequency-independent selection likely to maintain variation in nature? We know that selection on traits is typically strong and typically fluctuates across generations (Kingsolver *et al.*, 2001); classic examples of selection on discrete loci also involve strong selection that changes through time (Lewontin *et al.*, 1981; Cook *et al.*, 1986; Lenormand *et al.*, 1999). Recent genome-wide surveys of polymorphism in *D.*

melanogaster show that large numbers of SNP rapidly change frequency through the seasons (Bergland *et al.*, 2014), though it is not clear just how many causal loci drive this pattern. Over the longer time scale of the fossil record, we see that phenotypes may change rapidly in the short term, and yet much more slowly in the long term (Gingerich, 2009). This suggests that selection may typically fluctuate strongly, yet largely cancel in the long term.

Nevertheless, we believe that the mechanism that we identify is unlikely to be significant in explaining the bulk of variation across the genome. Although selection on the organism can be strong, selection on nucleotide sites must typically be weak; indeed, indirect estimates from patterns of molecular variation suggest selection coefficients that are typically small, even if often strong relative to drift ($s < 10^{-2}$, say; Charlesworth, 2015). Most sets of loci are loosely linked ($r \sim 1/2$), and so weak epistasis is unlikely to provide much stability. More important, as discussed above, systematic bias in directional selection is likely to be much stronger than the stabilising interaction between epistasis and linkage disequilibrium. In principle, frequency dependent selection gives a much more robust mechanism that can give an advantage to rare alleles, and thereby maintain variation. Even here, though, many of the proposed mechanisms require either strong selection, or a delicate balancing of parameters, e.g. the Levene Model (Levene, 1953) or more generally Maynard Smith and Hoekstra (1980). Thus, after half a century of debate, it remains unclear that balancing selection can account for widespread variation across the genome.

Acknowledgements

We would like to thank Katarína Bodďová, Reinhard Bürger, Josef Hofbauer, Meike Wittmann, and two anonymous reviewers for valuable suggestions and stimulating ideas. This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under Grant Agreement 618091 Speed of Adaptation in Population Genetics and Evolutionary Computation (SAGE).

References

Bank, C., *et al.*, 2012 The limits to parapatric speciation: Dobzhansky–Muller incompatibilities in a continent–island model. Genetics **191**: 845–863.

Baum, L. E., and J. A. Eagon, 1967 An inequality with applications to statistical estimation for probabilistic functions of markov processes and to a model for ecology. Bull. Amer. Math. Soc **73**: 360–363.

Bergland, A. O., *et al.*, 2014 Genomic evidence of rapid and stable adaptive oscillations over seasonal time scales in drosophila. PLoS Genetics **10**: e1004775.

Bürger, R., 2000 *The mathematical theory of selection, recombination, and mutation*, volume 228. Wiley Chichester.

Charlesworth, B., 2015 Causes of natural variation in fitness: evidence from studies of drosophila populations. Proceedings of the National Academy of Sciences **112**: 1662–1669.

Cook, L. M., *et al.*, 1986 Postindustrial melanism in the peppered moth. Science(Washington) **231**: 611–613.

Dean, A. M., 2005 Protecting haploid polymorphisms in temporally variable environments. Genetics **169**: 1147–1156.

Ellner, S., and A. Sasaki, 1996 Patterns of genetic polymorphism maintained by fluctuating selection with overlapping generations. Theoretical Population Biology **50**: 31–65.

Felsenstein, J., 1976 The theoretical population genetics of variable selection and migration. Annual Review of Genetics **10**: 253–280.

Fisher, R. A., 1930 *The Genetical Theory of Natural Selection*. Clarendon Press, Oxford.

Gingerich, P. D., 2009 Rates of evolution. Annual Review of Ecology, Evolution, and Systematics **40**: 657–675.

Gulisija, D., *et al.*, 2016 Phenotypic plasticity promotes balanced polymorphism in periodic environments by a genomic storage effect. Genetics **202**: 1437–1448.

Haldane, J. B. S., and S. D. Jayakar, 1963 Polymorphism due to selection of varying direction. Journal of Genetics **58**: 237–242.

Hoekstra, R. F., 1975 A deterministic model of cyclical selection. Genetical Research **25**: 1–15.

454 Kimura, M., 1956 A model of a genetic system which leads to closer linkage by natural selection. Evolution **10**: 278–287.

456 Kingsolver, J. G., *et al.*, 2001 The strength of phenotypic selection in natural populations. The American Naturalist **157**: 245–261.

458 Kirzhner, V. M., *et al.*, 1994 Cyclical behavior of genotype frequencies in a two-locus population under fluctuating haploid selection. Proceedings of the National Academy of 460 Sciences **91**: 11432–11436.

Kirzhner, V. M., and Y. Lyubich, 1997 Multilocus dynamics under haploid selection. 462 Journal of Mathematical Biology **35**: 391–408.

Lenormand, T., *et al.*, 1999 Tracking the evolution of insecticide resistance in the 464 mosquito culex pipiens. Nature **400**: 861–864.

Levene, H., 1953 Genetic equilibrium when more than one ecological niche is available. 466 The American Naturalist **87**: 331–333.

Lewontin, R. C., *et al.*, 1981 *Dobzhansky's genetics of natural populations I-XLIII*. 468 Number 575.1 D6.

Lyubich, Y. I., 1992 *Mathematical Structures in Population Genetics*. Springer, Berlin, 470 Heidelberg, New York.

Maynard Smith, J., and R. Hoekstra, 1980 Polymorphism in a varied environment: 472 how robust are the models? Genetics Research **35**: 45–57.

Mehta, R., *et al.*, 2015 Natural selection as an inhibitor of genetic diversity: Multi-474 plicative weights updates algorithm and a conjecture of haploid genetics [working paper abstract]. In *Proceedings of the 2015 Conference on Innovations in Theoretical Computer* 476 *Science*. ACM, 73–73.

Nagylaki, T., 1974 Quasilinkage equilibrium and the evolution of two-locus systems. 478 Proceedings of the National Academy of Sciences **71**: 526–530.

Nagylaki, T., 1975 Polymorphisms in cyclically-varying environments. Heredity **35**: 480 67–74.

NAGYLAKI, T., 1977 *Selection in one-and two-locus systems*, volume 15 of *Lecture Notes in Biomathematics*. Springer, Berlin, Heidelberg, New York.

NAGYLAKI, T., 1993 The evolution of multilocus systems under weak selection. Genetics **134**: 627–647.

NAGYLAKI, T., *et al.*, 1999 Convergence of multilocus systems under weak epistasis or weak selection. Journal of Mathematical Biology **38**: 103–133.

STROBECK, C., 1979 Haploid selection with n alleles in m niches. American Naturalist **113**: 439–444.

WRIGHT, S., 1937 The distribution of gene frequencies in populations. Proceedings of the National Academy of Sciences **23**: 307–320.

Appendices

A.1 Frequency-independence at the allelic level

A haplotype (haploid genotype) $\{P^1_{k_1}, ..., P^L_{k_L}\}$ (where $k_i \in \{1, ..., K_i\}$) is a list of alleles that sit at their respective loci. Under the assumption of linkage equilibrium between loci, the frequency of any haplotype $\{P^1_{k_1}, ..., P^L_{k_L}\}$ is given by the product of the frequencies of the alleles it consists of,

$$p_{\{P^1_{k_1},...,P^L_{k_L}\}} = \prod_{i=1}^{L} p^i_{k_i}.$$

We denote the fitness of the haplotype $\{P^1_{k_1}, ..., P^L_{k_L}\}$ by the constant $w_{\{P^1_{k_1},...,P^L_{k_L}\}} > 0$. Note that this does not imply anything about epistatic interactions between the individual alleles.

The mean fitness $\bar{w}$ of the population is obtained by averaging the haplotype fitnesses over the haplotype frequencies. Then, the mean fitness is a homogeneous polynomial of degree L, whose positive coefficients are the fitness values of the haplotypes,

$$\bar{w} = \sum_{k_1=1}^{K_1} \cdots \sum_{k_L=1}^{K_L} \left(w_{\{P^1_{k_1},...,P^L_{k_L}\}} \prod_{i=1}^{L} p^i_{k_i} \right). \tag{A.1a}$$

Similarly, the marginal fitness of an allele $P^{i_\lambda}_{k_\kappa}$ can be written as

$$w^{i_\lambda}_{k_\kappa} = \sum_{k_1=1}^{K_1} \cdots \sum_{k_{i_\lambda-1}=1}^{K_{i_\lambda-1}} \sum_{k_{i_\lambda+1}=1}^{K_{i_\lambda+1}} \cdots \sum_{k_L=1}^{K_L} \left(w_{\{P^1_{k_1},...,P^{i_\lambda}_{k_\kappa},...,P^L_{k_L}\}} \prod_{i\neq i_\lambda} p^i_{k_i} \right). \tag{A.1b}$$

From the mean and marginal fitnesses under constant haploid selection, equations (A.1a) and (A.1b), it is straightforward to see that the marginal fitness of an allele is independent of the allele frequency at its locus, equation (1b).

Here, we show that frequency-independent selection at the allelic level, as defined by equation (1b), implies that the mean and marginal fitnesses can be written by equations (A.1a) and (A.1b) under the assumption of linkage equilibrium. In this sense, constant haploid selection is contained in the notion of frequency-independent selection at the allelic

level. Due to its (verbal) definition as the average over genotypes (see the main text), the mean fitness can generally be written as a linear combination of the marginal fitnesses, i.e., equation (1a) holds in general. It thus suffices to show that the mean fitness $\bar{w}(\mathbf{p})$ can be written as a homogeneous polynomial of degree L with positive coefficients under the additional assumption of frequency-independent selection at the allelic level as given by equation (1b).

From equation (1a) and (1b), it follows that

$$\frac{\partial^2 \bar{w}}{\partial p_j^i \partial p_k^i} = \frac{\partial w_j^i}{\partial p_k^i} = 0 \qquad \text{for all } \mathbf{p} \tag{A.2}$$

at any locus $i \in \{1, ..., L\}$ and for any choice of alleles $j, k \in \{1, ..., K_i\}$. Furthermore, from equation (2) we have

$$\frac{\partial \bar{w}}{\partial p_j^i} = w_j^i \geq 0 \qquad \text{for all } \mathbf{p}, \tag{A.3}$$

since marginal fitnesses are always non-negative by definition. Thus, in terms of any given locus, the mean fitness $\bar{w}$ is a linear function of the allele frequencies at the locus. Its coefficients w_j^i are non-negative functions only of the frequencies of alleles at other loci. For each pair of loci, $i, j \in \{1, ..., L\}$, $i \neq j$, we may write the mean fitness in two ways as

$$\bar{w} = \sum_{k=1}^{K_l} p_k^l w_k^l = \sum_{j=1}^{K_i} p_j^i w_j^i, \tag{A.4}$$

c.f. equation (1a). Taking the derivative with respect to the frequency of any allele k at locus l yields

$$\frac{\partial \bar{w}}{\partial p_k^l} = w_k^l = \sum_{j=1}^{K_i} p_j^i \, \frac{\partial w_j^i}{\partial p_k^l}. \tag{A.5}$$

The function $w_k^l(\mathbf{p})$ is constant in the allele frequencies p_j^l at locus l. Choosing $p_j^i = 1$ (which implies $p_m^i = 0$ for $m \neq j$) shows that also the derivative $\partial w_j^i / \partial p_k^l$ is constant in p_j^l. Hence, the marginal fitnesses w_j^i are linear functions in the allele frequencies at the loci different from locus i.

Overall, it follows that $\bar{w}$ is a polynomial with positive coefficients, whose degree is at most L. Furthermore, each term of this polynomial contains at most one allele frequency from each locus. If a locus i is not represented in any one of the terms, we may multiply it with $\sum_{j=1}^{K_i} p_j^i = 1$ to homogenize the polynomial. Hence, $\bar{w}$ is completed to degree L

534 homogeneously and our claim is proved.

A.2 Convergence to the set of equilibrium points

536 From equation (2), it follows that the dynamics (3) for allele k on locus i may be written as

$$(p_k^i)' = \frac{p_k^i \frac{\partial \bar{w}}{\partial p_k^i}}{\sum_{j=1}^{K_i} p_j^i \frac{\partial \bar{w}}{\partial p_j^i}}. \tag{A.6}$$

538 Furthermore, $\bar{w}$ is a homogeneous polynomial with positive coefficients, see Appendix A.1. Thus, the conditions for the inequality of Baum and Eagon (1967) are met. It follows
540 that the mean fitness $\bar{w} = \bar{w}(\mathbf{p})$ is strictly increasing along trajectories of $\mathbf{p}$ under the dynamics (3), remaining constant only at equilibrium,

$$\bar{w}(\mathbf{p}') \geq \bar{w}(\mathbf{p}) \qquad \text{and} \qquad \bar{w}(\mathbf{p}') = \bar{w}(\mathbf{p}) \Leftrightarrow \mathbf{p}' = \mathbf{p}.$$

542 The space of admissible $\mathbf{p}$ is a compact set, hence the values of $\bar{w}(\mathbf{p})$ are bounded from above. Thus, every trajectory of the dynamics (3) converges to the set of its equilibrium
544 points. This set may be complicated; nevertheless, the existence of other attractors (e.g. periodic orbits, chaotic attractors) can be excluded (e.g. Lyubich, 1992, Ch.9). Further
546 assuming that all equilibria are hyperbolic implies that there are only finitely many equilibria (Nagylaki *et al.*, 1999), hence each trajectory of the dynamics converges to an
548 equilibrium; in particular, there is no cycling behaviour. The hyperbolicity of equilibria is a generic property (Nagylaki *et al.*, 1999, Appendix B) that implies the non-degeneracy
550 of equilibria we point out in Theorem A.1 below.

A.3 Unstable polymorphism under constant selection

552 Because the p_k^i are frequencies, we have $\sum_{k=1}^{K_i} p_k^i = 1$ for every $i \in \{1, ..., L\}$. The dynamics of allele frequencies at locus i with K_i alleles evolve on the $(K_i - 1)$-dimensional simplex

$$\Delta_{K_i - 1} = \left\{ (p_1^i, ..., p_{K_i}^i) \in \mathbb{R}^{K_i} : \sum_{k=1}^{K_i} p_k^i = 1 \right\}. \tag{A.7}$$

The joint equations (3) for the change of allele frequencies are thus a dynamical system on the product of simplices

$$\mathbf{\Delta} = \Delta_{K_1-1} \otimes \Delta_{K_2-1} \otimes \cdots \otimes \Delta_{K_L-1},$$

which has $\mathcal{K} - L$ degrees of freedom due to the constraints on the p_k^i being frequencies.

The linearisation of the dynamics (3) around an equilibrium $\hat{\mathbf{p}}$ is given by the Jacobian $J(\hat{\mathbf{p}})$. This matrix has dimensions $\mathcal{K} \times \mathcal{K}$, and its entry at position (m, n) is obtained from taking the derivative of the m-th function f_k^i with respect to the n-th variable p_j^l (counting contiguously across loci), and evaluating at $\hat{\mathbf{p}}$. For example, the diagonal entries of $J(\hat{\mathbf{p}})$ are given by

$$(J(\hat{\mathbf{p}}))_{m,m} = \frac{\partial f_k^i(\hat{\mathbf{p}})}{\partial p_k^i}, \tag{A.8}$$

where $m = \sum_{j=1}^{i-1} K_j + k$.

According to Appendix A.2, the trajectories of allele frequencies under the dynamics (3) converge to the set of their equilibrium points. These equilibria may be fully polymorphic or on the boundary of the state space $\mathbf{\Delta}$. Theorem A.1 below states that any polymorphic equilibrium is unstable, hence trajectories converge to the latter type of equilibria and at least one allele is lost from the population.

Theorem A.1 (Instability of polymorphic equilibria)**.** *Under the assumptions of linkage equilibrium and constant frequency-independent selection, any fully polymorphic equilibrium* $\hat{\mathbf{p}}$ *of the dynamics* (3) *is unstable. This holds for all but a set of degenerate cases of measure zero, where all eigenvalues of the Jacobian* $J(\hat{\mathbf{p}})$ *are equal to* 1*.*

For the proof of the theorem, we note two simple results:

Lemma A.1. *The Jacobian* $J(\hat{\mathbf{p}})$ *at a polymorphic equilibrium* $\hat{\mathbf{p}}$ *of the dynamics* (3) *has at most* $\mathcal{K} - L$ *non-zero eigenvalues.*

Proof. This is the case because the dynamics (3) are a projection on the space

$$\mathbf{\Delta} = \Delta_{K_1-1} \otimes \Delta_{K_2-1} \otimes \cdots \otimes \Delta_{K_L-1},$$

which has dimension $\mathcal{K} - L$. □

Lemma A.2. *Consider a polymorphic equilibrium* $\hat{\mathbf{p}}$ *of the dynamics* (3) *under the same assumptions as in Theorem A.1. Then, the trace of the Jacobian* $J(\hat{\mathbf{p}})$ *(i.e., the sum of its diagonal entries) is*

$$tr(J(\hat{\mathbf{p}})) = \mathcal{K} - L.$$

Proof. At any polymorphic equilibrium of the dynamics (3) we have

$$w_k^i = \bar{w} \qquad \text{for all } i \in \{1, ..., L\} \text{ and } k \in \{1, ..., K_i\}. \tag{A.9}$$

Using this identity together with equations (1a) and (1b), we calculate the diagonal entries of $J(\hat{\mathbf{p}})$ (equation (A.8)) as

$$\frac{\partial f_k^i(\hat{\mathbf{p}})}{\partial p_k^i} = 1 - \hat{p}_k^i.$$

It follows that the trace of $J(\hat{\mathbf{p}})$ is

$$\begin{aligned} \mathrm{tr}(J(\hat{\mathbf{p}})) &= \sum_{i=1}^{L} \sum_{k_i=1}^{K_i} \frac{\partial f_{k_i}^i(\hat{\mathbf{p}})}{\partial p_{k_i}^i} = \sum_{i=1}^{L} \sum_{k_i=1}^{K_i} \left(1 - \hat{p}_{k_i}^i\right) = \\ &= \sum_{i=1}^{L} (K_i - 1) = \mathcal{K} - L. \end{aligned}$$

□

Proof of the Theorem. Consider a polymorphic equilibrium $\hat{\mathbf{p}}$ of equation (3) and its Jacobian $J(\hat{\mathbf{p}})$. Because $J(\hat{\mathbf{p}})$ is a matrix with dimensions $\mathcal{K} \times \mathcal{K}$, it has $\mathcal{K}$ eigenvalues (counting multiplicities). Due to Lemma A.1, at most $\mathcal{K} - L$ are non-zero. According to Lemma A.2, the trace and hence the sum of eigenvalues of $J(\hat{\mathbf{p}})$ is also $\mathcal{K} - L$. Therefore, unless all eigenvalues equal one (the degenerate case), at least one of them has a modulus greater than one. Thus, the polymorphic equilibrium $\hat{\mathbf{p}}$ is unstable. □

A.4 Stability of equilibria in the continuous-time limit

We assume that selection and recombination are weak, so that we can approximate time to be continuous. Selection and recombination are of the same order, so that there may be strong linkage disequilibria. Since only the ratio of the two processes matters, selection scales relative to the recombination rate. We define the genotype by the list of alleles, X. There may be multiple alleles per locus; gamete frequencies are $g[X]$, and the corresponding

fitnesses are $s[X]$. Because we assume continuous time, $s[X]$ can be positive or negative, and adding a constant to s would make no difference.

The rates of change of genotype frequencies are

$$\partial_t\, g[X] = (s[X] - \bar{s} - R^*[X])\, g[X] + \sum_{Y,Z} g[Y]\, g[Z]\, R[X|Y,Z], \tag{A.10}$$

where $\bar{s} = \sum_X s[X]\, g[X]$ and $R[X|Y,Z]$ is the rate at which a pair $\{Y,Z\}$ generates X. If we ignore multiple crossovers, this is a sum over the $n-1$ single recombination events, each with rate $r_1, ..., r_{n-1}$. $R^*[X]$ is the rate at which X is broken down, which equals the sum of the recombination rates, $r_T = \sum_{i=1}^{n-1} r_i$. Note that this includes recombination events that generate X itself, but these cancel with corresponding terms in the sum over Y, Z. Writing this explicitly:

$$\partial_t\, g[X] = (s[X] - \bar{s} - r_T)\, g[X] + \sum_{i=1}^{n-1} r_i\, g[X_{i,\leftarrow}]\, g[X_{i,\hookrightarrow}], \tag{A.11}$$

where $g[X_{i,\leftarrow}]$ denotes the frequency of all gametes that match X to the left of breakpoint i, and similarly for $g[X_{i,\hookrightarrow}]$. Because $\sum_X g[X] = 1$, the rates of change $\partial_t\, g[X]$ sum to zero over X.

At equilibrium, $\partial_t\, g[X] = 0$ for all X, and the stability is determined by the matrix

$$\frac{\partial\, \partial_t\, g[X]}{\partial\, g[Y]} = (s[X] - \bar{s} - r_T)\, \delta_{X,Y} + \sum_{i=1}^{n-1} r_i \left(\delta_{X_{i,\leftarrow},Y_{i,\leftarrow}}\, g[X_{i,\hookrightarrow}] + g[X_{i,\leftarrow}]\delta_{X_{i,\hookrightarrow},Y_{i,\hookrightarrow}}\right) - g[X]\, s[Y], \tag{A.12}$$

where the last term arises from differentiating $\bar{s}$, and $\delta_{X,Y}$ is 1 if $X = Y$ and 0 otherwise. If we specify the equilibrium frequencies and recombination rates, that uniquely determines the $s[X]$:

$$\begin{aligned} (s[X] - \bar{s} - r_T)\, g[X] &= -\textstyle\sum_{i=1}^{n-1} r_i\, g[X_{i,\leftarrow}]\, g[X_{i,\hookrightarrow}] \\ \Rightarrow\ \ s[X] &= -\textstyle\sum_{i=1}^{n-1} r_i\, \frac{g[X_{i,\leftarrow}]\, g[X_{i,\hookrightarrow}]}{g[X]} + (\bar{s} + r_T)\,. \end{aligned} \tag{A.13}$$

Substituting into Equation (A.12) yields

$$\begin{aligned} &\tfrac{\partial\, \partial_t\, g[X]}{\partial\, g[Y]} = \\ &= \textstyle\sum_{i=1}^{n-1} r_i \left(\delta_{X_{i,\leftarrow},Y_{i,\leftarrow}}\, g[X_{i,\hookrightarrow}] + g[X_{i,\leftarrow}]\delta_{X_{i,\hookrightarrow},Y_{i,\hookrightarrow}} - \tfrac{g[X_{i,\leftarrow}]\, g[X_{i,\hookrightarrow}]}{g[X]}\, \delta_{X,Y}\right) - g[X]\, s[Y]. \end{aligned} \tag{A.14}$$

eigenvalue	$-r_1-r_2-r_3$	$-r_2-r_3$	$-r_1-r_2$	$-r_3$	$-r_2$	$-r_1$	0	$r_1+r_2+r_3$
multiplicity	4	2	2	1	1	1	4	1

Table A.1: Eigenvalues for the symmetric and neutral case, with 4 loci. The eigenvalue $r_T = r_1+r_2+r_3$ corresponds to the trivial constant eigenvector, and should be disregarded.

Note that the sum over X is a constant,

$$\sum_X \frac{\partial\, \partial_t\, g[X]}{\partial\, g[Y]} = r_T - \bar{s}, \tag{A.15}$$

which implies that there is a constant eigenvector with eigenvalue $r_T - \bar{s}$. This corresponds to perturbations that violate the constraint $\sum_X g[X] = 1$, and therefore can be ignored. The stability of the equilibrium is determined by whether the largest real part of the remaining eigenvalues is positive.

We have not found a way to show analytically that this matrix implies instability, even in the three-locus case. However, it is instructive to examine the completely symmetric case, with two alleles per locus, where all haplotypes are equally common. This corresponds to linkage equilibrium, and allele frequencies all equal to 1/2. With complete symmetry, the eigenvalues can be calculated for up to 4 loci, and depend in a simple way on the recombination rates; Table A.1 shows the 16 eigenvalues for four loci, and suggests the obvious generalisation to n loci. There are n zero eigenvalues, corresponding to neutral change in allele frequency; the question is how these change when the frequencies are perturbed away from linkage equilibrium, and hence away from neutrality. Numerically, one or more of these eigenvalues becomes positive, consistent with our previous result that when selection is much weaker than recombination, so that linkage equilibrium is approached, polymorphism is unstable. However, this does not tell us whether stable equilibria exist when epistasis and linkage equilibria are strong. To investigate that, we calculate eigenvalues of the stability matrix numerically (see the main text).

A.5 Stable polymorphism at two diallelic loci under fluctuating selection

A.5.1 Formulation of the dynamics

Consider two diallelic loci with genotype fitnesses as given in Table 2. Under weak selection and recombination (formally: replacing $(s, \epsilon, r) \mapsto (\alpha s, \alpha\epsilon, \alpha r)$ and letting $\alpha \to 0$), the dynamics of genotype frequencies is

$$\dot{p}_{AB} = p_{AB}\,(2s - \bar{m}) - rD, \tag{A.16a}$$
$$\dot{p}_{Ab} = p_{Ab}\,(s - \epsilon - \bar{m}) + rD, \tag{A.16b}$$
$$\dot{p}_{bB} = p_{bB}\,(s - \epsilon - \bar{m}) + rD, \tag{A.16c}$$
$$\dot{p}_{ab} = -p_{ab}\,\bar{m} - rD, \tag{A.16d}$$

as first derived by Kimura (1956) (see also Bürger, 2000, Ch.II.1). The measure of linkage disequilibrium D is given by $D = p_{AB}p_{ab} - p_{Ab}p_{aB}$ as in the main text and $\bar{m} = 2sp_{AB} + (s - \epsilon)(p_{Ab} + p_{aB})$. Due to the constraint $p_{AB} + p_{Ab} + p_{aB} + p_{ab} = 1$, one of these equations is redundant. Rewriting the system (A.16) in terms of the allele frequencies $p_A = p_{AB} + p_{Ab}$ and $p_B = p_{AB} + p_{aB}$ yields

$$\dot{p}_A = p_A\,q_A\,(s - \epsilon(q_B - p_B)) + D\,(s + \epsilon(q_A - p_A))\,, \tag{A.17a}$$
$$\dot{p}_B = p_B\,q_B\,(s - \epsilon(q_A - p_A)) + D\,(s + \epsilon(q_B - p_B))\,, \tag{A.17b}$$
$$\dot{D} = 2\epsilon\,p_A\,p_B\,q_A\,q_A - D\,[r + 2s\,(p_A\,p_B - q_A\,q_B) + 2\epsilon D]\,, \tag{A.17c}$$

where $q_A = 1 - p_A$ and $q_B = 1 - p_B$. Introducing $P = (p_A + p_B)/2$ and $\delta = p_B - p_A$ then leads to the system (4).

A.5.2 Stability of a periodic trajectory under fluctuating selection

Using numerical routines of the computer algebra software *Mathematica*, we simulate fluctuating selection on two diallelic loci by iterating two seasons as described in the main text, Section 3.3. For all tested initial conditions (two sets of initial conditions were tested: (i) 10^3 randomly drawn starting points, and (ii) p_A and p_B distributed on a regular grid in the p_A-p_B plane as displayed in Figure A.1, and $D = \pm 0.1$), the dynamics converge to

the periodic trajectory displayed in Figure 1. Through the course of the two seasons, the periodic trajectory interpolates between the approximate values

$$(P, \delta, D) \approx (0.3106, 0, 0.0471) \qquad \text{and} \qquad (P, \delta, D) \approx (0.4259, 0, -0.0560)\,,$$

see the bold points in Figure 1. These values were calculated numerically as the fixed points of two Poincaré maps, $\mathcal{P}_1$ and $\mathcal{P}_2$, defined by the values of P, δ and D at the transition between season 1 and 2, and season 2 and 1 (i.e., at the time points $(2k-1)\tau$ and $2k\tau$ for $k = 1, 2, 3, ...$ respectively),

$$\begin{aligned} \mathcal{P}_1(k) &= \big(P((2k-1)\tau), \delta((2k-1)\tau), D((2k-1)\tau)\big) \\ \mathcal{P}_2(k) &= \big(P(2k\tau), \delta(2k\tau), D(2k\tau)\big) \end{aligned} \qquad (k = 1, 2, 3, ...). \tag{A.18}$$

Numerically calculating the eigenvalues λ_1, λ_2, λ_3 of the linearisation of either Poincaré map at its equilibrium yields

$$\lambda_1 \approx 0.9823, \qquad \lambda_2 \approx 0.9822, \qquad \text{and} \qquad \lambda_3 \approx 0.3549,$$

hence the periodic trajectory is asymptotically stable. Repeating the procedure for different recombination rates r produces the data presented in Figure 2.

A.5.3 An asymmetric example of stable polymorphism

We simulate the dynamics given by equation (4) with different epistatic coefficients $\epsilon_1 \neq \epsilon_2$ in the fitness values of the two genotypes with mixed upper and lower case letters, Ab and aB. The fitness configurations are given by Table A.2. We assume the two seasons to be of equal length $\tau = 50$ and the recombination rate to be $r = 0.05$. The specific values of s, ϵ_1, and ϵ_2 in each of the seasons are given in Table A.3. For a set of initial conditions (e.g. $P = 0.5$, $\delta = 0.1$, $D = 0$), the temporal dynamics of the system converge to a periodic orbit shown in Figure A.2. The asymptotic stability of this orbit can be established by the methods outlined above.

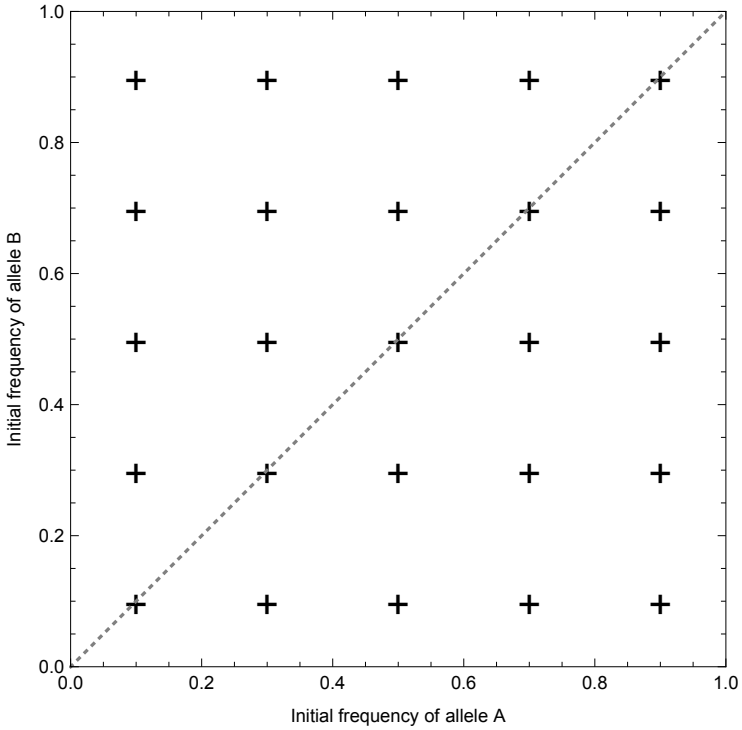

Figure A.1: Each of these 25 points with both $D = 0.1$ and $D = -0.1$ were used as initial conditions for the dynamics given by equations (4) and Table 3; each led to the convergence of the simulation to the periodic trajectory displayed in Figure 1.

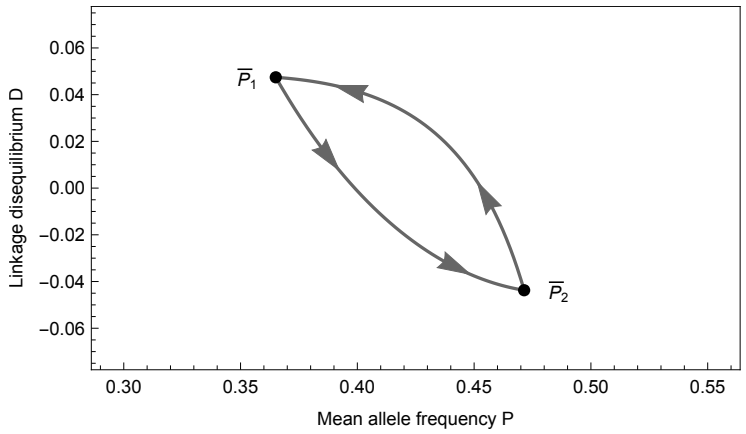

Figure A.2: The stable trajectory in the P-D plane under equations (4), based on the parameters given in Table A.3. The trajectory is locally stable. In season 1 the dynamics evolve from $\bar{P}_1$ to $\bar{P}_2$, season 2 maps $\bar{P}_2$ back to $\bar{P}_1$.

Genotype	AB	Ab	aB	ab
Frequency	p_{AB}	p_{Ab}	p_{aB}	p_{ab}
Fitness	$w_{AB} = 1 + 2s$	$w_{Ab} = 1 + s - \epsilon_1$	$w_{aB} = 1 + s - \epsilon_2$	$w_{ab} = 1$

Table A.2: A genotype fitness configuration generalizing Table 2. Using two epistatic coefficients ϵ_1 and ϵ_2 disrupts the symmetry between the two loci.

	Selection	Epistasis	Recombination	Duration
Season 1:	$s = 0.00525$	$\epsilon_1 = -0.021$, $\epsilon_2 = -0.031$	$r = 0.05$	$\tau = 50$
Season 2:	$s = -0.00475$	$\epsilon = 0.019$, $\epsilon_2 = 0.029$	$r = 0.05$	$\tau = 50$

Table A.3: A parameter configuration that disrupts the symmetry between the two loci and leads to a stable trajectory under equations (4) with seasonal fluctuations.