Sign epistasis extends the effects of balancing selection on genetic diversity

Ksenia A. Khudiakova*[1], Nicholas H. Barton[1] and Goran Arnqvist[2]

[1]Institute of Science and Technology Austria, Am Campus 1, 3400, Klosterneuburg, Austria
[2]Animal Ecology, Department of Ecology and Genetics, Evolutionary Biology Centre, Uppsala University, Uppsala, Sweden
*Corresponding author. Email: khudyakovaks@gmail.com

Abstract

Balancing selection, a form of selection that maintains genetic diversity, is difficult to detect, and the importance of balancing selection for the maintenance of genetic variation may be larger than often assumed. We model the possibility that the diversity-promoting effects of balancing selection extend to other loci that show sign epistasis with a locus under balancing selection. Rather than focusing on overdominance, as was done in previous efforts, we explore the effects of negative frequency dependence and show that this has important effects on the conditions under which the diversity-promoting effect of epistasis can occur in diploids. Our results show that not only recombination rate but also the dominance of sign epistasis are key parameters that determine the maintenance of polymorphism beyond the locus under direct balancing selection. We suggest that the effect we explore may play a significant role, especially when balancing selection acts on major effect loci.

Keywords

sign epistasis; balancing selection; within-population genetic variation

1 Introduction

Efforts to intersect DNA sequencing data and population genetics theory with estimates of genetic variance suggest that the observed variability in fitness components cannot be maintained solely by mutation-selection balance (Barton & Keightley, 2002; Bonnet et al., 2022; B. Charlesworth, 2015). Genetic variation in life history traits is often sizeable (Burch et al., 2024; Mousseau & Roff, 1987) and within-population polymorphism in major-effect variants such as inversions are common (Faria et al., 2019). These observations call for a better understanding of the processes that act to maintain substantial levels of non-neutral genetic diversity. One possibility is that the genome-wide effects of balancing selection are larger than typically assumed (B. Charlesworth, 2015; Kroymann & Mitchell-Olds, 2005). Balancing selection (BLS) is selection that acts to maintain variation, including both overdominance (i.e., heterozygote advantage) and negative frequency-dependent selection (NFDS).

The conditions under which we expect balancing selection in a relatively restricted number of loci to affect genetic diversity more widely have been studied previously. For example, it is commonly recognized that balancing selection increases diversity across linked neutral sites (D. Charlesworth, 2006). Furher, functional dependencies between loci under balancing selection and other sites in the genome, such that different alleles are favored on different genetic backgrounds, called sign epistasis, may play a significant role in this context. Indeed, models

of haploid systems confirm that epistasis can strengthen the diversifying effects of balancing selection (Kelly, 2000; Kelly & Wade, 2000).

Sign epistasis represents inter-locus interactions between alleles such that the fitness effect of a given allele in one locus changes from beneficial to deleterious depending on the allelic state at one or more other loci (fig. 1). Such epistatic interactions tend to generate complex fitness valleys, i.e., clusters of polygenic genotypes of low fitness, and fitness ridges or even distinct fitness peaks (Gavrilets, 2004; Saona et al., 2022; Poelwijk et al., 2011; Wolf et al., 2000). Epistasis has a central role in our understanding of incipient speciation, where evolving epistatic interactions within populations (Corbett-Detig et al., 2013; Ortiz-Barrientos et al., 2016) may result in reproductive isolation between populations through the build-up of epistatic Dobzhansky-Muller incompatibilities (Orr & Turelli, 2001). Yet, most evolutionary genetic models ignore interactions between loci (Hansen, 2013) and links between the role of epistasis in adaptation and its role in speciation, potentially connecting micro- and macroevolution, are not fully understood (Johnson, 2000; Wade, 2000). Although the contribution of physiological epistasis to additive genetic variance can be very substantial, epistasis is typically regarded as inconsequential for the short-term response to selection (Mäki-Tanila & Hill, 2014). Contrary to this view, epistasis can contribute to the long-term response to selection (Paixão & Barton, 2016; Bourg et al., 2024) and several experimental studies have unveiled an important role for interactions between loci in adaptation (Bloom et al., 2013; Good et al., 2017; Pettersson et al., 2011; Wolf et al., 2000). Further, it has long been recognized that epistatic variance can be "converted" into additive genetic variance following, e.g., population bottlenecks (Barton & Turelli, 2004), and line-cross analyses have shown that the epistatic contribution to genetic variance in life history traits is often substantial (Burch et al., 2024; Roff & Emerson, 2006). However, our understanding of whether and how epistasis affects those processes that maintain allelic diversity in interacting loci within populations is relatively restricted (Hansen, 2013; Hemani et al., 2013).

Since recombination tends to disassociate favorable allelic combinations, the potential of epistasis to maintain genetic variation depends on the balance between the strength of selection and the rate of recombination. Hence, we expect two related effects: 1) the efficacy of balancing and epistatic selection in maintaining variation should decrease with physical distance, and 2) epistatic selection will lead to evolution of tighter linkage. Already Fisher (1930, p. 103) argued that sign epistasis can lead to decreased recombination. A formal analysis was provided by Kimura (1956), who investigated the equilibria and stability of a deterministic two-locus model in which one locus exhibits heterozygote advantage and interacts with a second locus through sign epistasis. He showed that the maintenance of polymorphism at both loci is possible when linkage is sufficiently tight. Later work generalised Kimura's fitness table in a class of fitness tables called "symmetric viability models", i.e. fitness tables that are symmetric with respect to the fitness of the double heterozygote upon relabeling of loci and alleles (Bodmer & Felsenstein, 1967; Karlin & Feldman, 1970; Christiansen, 1999). Another inherent feature of the symmetric viability models is constant fitnesses, with balancing selection being modeled by heterozygote advantage. For all symmetric viability models, the conclusions remained similar to Kimura's: the diversity-promoting effects of sign epistasis are restricted to the cases when loci are tightly linked, which predicts that this effect may be of limited evolutionary significance.

Here, we provide an exploration beyond symmetric viability models where we relax the assumption of constant fitnesses: instead of overdominance, we consider negative frequency dependence as the cause of balancing selection. We demonstrate that when epistatic effects show dominance, even free recombination allows for stable coexistence of polymorphic loci, challenging the long-standing view that tight linkage is required. Our results help clarify the conditions under which sign epistasis can contribute to the maintenance of within-population genetic diversity.

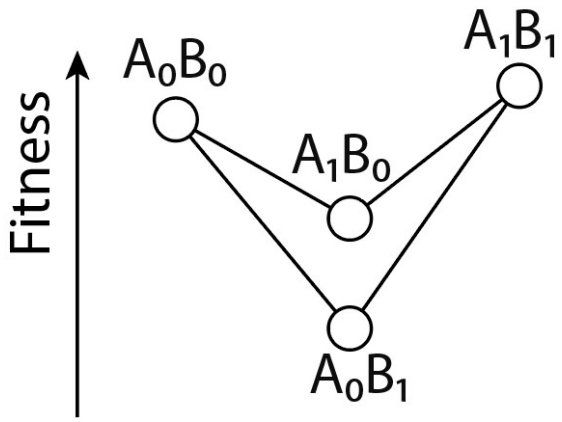


Figure 1: Sign epistasis between two haploid biallelic loci.

2 Materials and methods

We consider the deterministic dynamics of two biallelic loci under selection and recombination in an effectively infinite population. Locus A with alleles A_0 and A_1 is under symmetric negative frequency-dependent selection with selection coefficient t. Locus B with alleles B_0 and B_1 is in sign epistasis with locus A, which means that the genotypes containing a combination of conflicting alleles A_0 and B_1 or A_1 and B_0 experience an additional disadvantage proportional to the epistatic selection coefficient s. We denote the frequency of allele A_0 (resp. B_0) by q_A (q_B) and the frequency of allele A_1 (resp. B_1) by p_A (p_B).

For haploids, this results in the following fitness table:

$$\begin{array}{cccc} & \mathbf{B_0} & \mathbf{B_1} \\ \mathbf{A_0} & 1-tq_A & 1-tq_A-s \\ \mathbf{A_1} & 1-tp_A-s & 1-tp_A. \end{array}$$

For diploids, two additional parameters that describe the dominance relationship within the two loci are possible. Here, we assume that allelic effects at the A locus are additive, such that the fitness of the heterozygote A_0A_1 is intermediate to the fitnesses of the two homozygotes. The dominance relationship at locus B is arbitrary, and we model it with the parameter α. When $\alpha = 0$, selection at locus B is additive. When $\alpha > 0$, this generates dominance such that the fitness of each B heterozygote is closer to the fitter two-locus homozygote (B_0 shows dominance in one background and B_1 in the other). We term this dominance of the epistatic effect. Note that we consider $\alpha < s/2$, such that there is no direct overdominance at locus B. Nevertheless, as we will see later, $0 < \alpha < s/2$ can further stabilize double polymorphism. Thus, we consider the following fitness table:

$$\begin{array}{cccc} & \mathbf{B_0B_0} & \mathbf{B_1B_0} & \mathbf{B_1B_1} \\ \mathbf{A_0A_0} & 1-tq_A+s/2 & 1-tq_A+\alpha & 1-tq_A-s/2 \\ \mathbf{A_1A_0} & 1-t/2 & 1-t/2+\alpha & 1-t/2 \\ \mathbf{A_1A_1} & 1-tp_A-s/2 & 1-tp_A+\alpha & 1-tp_A+s/2. \end{array}$$

We assume random mating and no sex differences in fitness. Recombination rate between two loci is denoted by r, which means that proportion r of gametes in the new generation will contain combinations of alleles different from those of the parents.

We study the dynamics analytically using the multilocus algebra introduced in Barton & Turelli (1991) and Kirkpatrick et al. (2002).

3 Results

3.1 The haploid case

The haploid dynamics are described by the following set of equations:

$$\begin{aligned}
&\Delta p_A = a_A p_A q_A + a_B D + a_{AB}\left(q_A - p_A\right) D \\
&\Delta p_B = a_A D + a_B p_B q_B + a_{AB}\left(q_B - p_B\right) D \\
&\Delta D = (1-r)\Bigg(a_A\left(q_A - p_A\right) D + a_B\left(q_B - p_B\right) D + \\
&\quad a_{AB}\left(p_A q_A p_B q_B + \left(q_A - p_A\right)\left(q_B - p_B\right) D - D^2\right) - \Delta p_A \Delta p_B\Bigg) - rD,
\end{aligned} \tag{1}$$

where

$$\begin{aligned}
&\bar{W} a_A = t(q_A - p_A) - s(q_B - p_B) \\
&\bar{W} a_B = -s(q_A - p_A) \\
&\bar{W} a_{AB} = 2s \\
&\bar{W} = 1 + 2Ds - p_B s - p_A(q_B - p_B)s - t(1 - 2p_A q_A).
\end{aligned} \tag{2}$$

An example of the dynamics is shown on fig. 4a, SM. The system 1 admits a double-polymorphic equilibrium with $p_A = p_B = 1/2$ and the value of linkage disequilibrium

$$D_{eq} = \frac{1}{4}\left(\sqrt{\left(\frac{rB}{2s}\right)^2 + (1-r)} - \frac{rB}{2s}\right), \tag{3}$$

where $B = (2 - s - t)$.

This equilibrium is stable when

$$r < 2t\left(\frac{2 - s - \sqrt{s^2 + t^2}}{4 - 4s - t^2}\right). \tag{4}$$

Assuming weak selection, the stability condition reduces to $r \lesssim t$.

Numerical analyses of the selection equations confirmed these results (figs. 4b and 5, SM).

3.2 The diploid case

The diploid dynamics are described by the following set of equations.

$$\begin{aligned}
\Delta p_A &= a_A p_A q_A + a_B D + a_{AB}(q_A - p_A) D \\
\Delta p_B &= a_A D + a_B p_B q_B + a_{AB}(q_B - p_B) D,
\end{aligned} \tag{5}$$

where

$$\begin{aligned}
&\bar{W} a_A = \frac{t}{2}(q_A - p_A) - \frac{s}{2}(q_B - p_B) \\
&\bar{W} a_B = -\frac{s}{2}(q_A - p_A) + \alpha(q_B - p_B) \\
&\bar{W} a_{A,B} = \bar{W} a_{AB,\varnothing} = \frac{s}{2} \\
&\bar{W} a_{BB} = -2\alpha \\
&\bar{W} = 1 + Ds + \frac{s}{2}(q_A - p_A)(q_B - p_B) - t(1 - 2p_A q_A) + 2\alpha p_B q_B.
\end{aligned} \tag{6}$$

Note that in the selection coefficients, the terms involving s and t are half of those in the haploid case.

For linkage disequilibrium in diploids, we distinguish two types of statistical associations: those on the same chromosome will be denoted $D_{AB,\varnothing}$, and those on different chromosomes will be denoted $D_{A,B}$. Epistatic selection builds up $D_{A,B}$, while each round of free recombination brings $D_{A,B}$ towards zero. The total change in D is given by

$$\Delta D = (1-r)D^{sel}_{AB,\varnothing} + rD^{sel}_{A,B} - \Delta p_A \Delta p_B - D, \tag{7}$$

where

$$\begin{aligned} D^{sel}_{AB,\varnothing} &= D + a_A(q_A - p_A)D + a_B(q_B - p_B)D + a_{AB,\varnothing}\left(p_A q_A p_B q_B + D(q_A - p_A)(q_B - p_B) - D^2\right) \\ D^{sel}_{A,B} &= a_{A,B} p_A q_A p_B q_B + a_{A,B} D^2 + a_{BB} p_B q_B D. \end{aligned} \tag{8}$$

The system (5, 7) admits double-polymorphic equilibrium ($p_A = p_B = 1/2$) with

$$D_{eq} = \frac{1}{4}\left(\sqrt{\left(\frac{2rB}{s}\right)^2 + 1} - \frac{2rB}{s}\right), \tag{9}$$

where $B = (2 - t + 2\alpha)$.

With $s < t$ and $\alpha > 0$, the double-polymorphic equilibrium is stable whenever

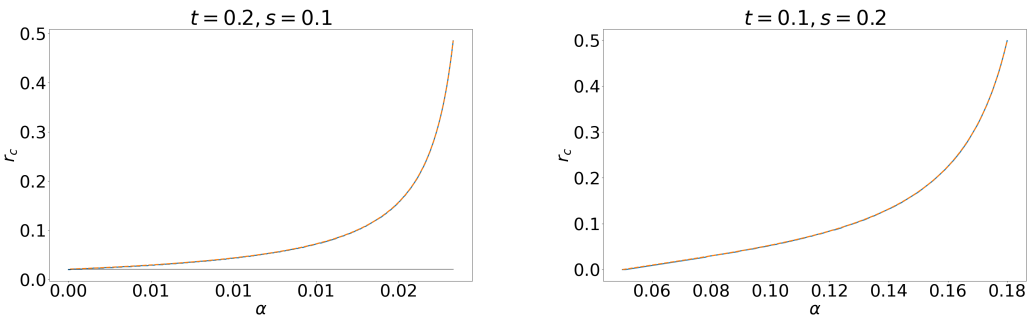
$$r < \frac{s^2}{16\alpha t} \frac{\left(s\sqrt{(4t\alpha)^2 + (s(t-2\alpha))^2} - (t+2\alpha)(s^2 - 4t\alpha)\right)}{(2 - t + 2\alpha)(s^2 - 2t\alpha)}, \tag{10}$$

and

$$r < \frac{t^2 - s^2}{4t(2-t)}, \tag{11}$$

when $\alpha = 0$. The analytical results match the numerical analysis of selection equations exactly (figs. 6 and 7, SM). As in the haploid case, thus, the condition for a double-polymorphic equilibrium in diploids under $\alpha = 0$ requires that recombination is low relative to selection. When $s > t$ such that epistatic selection is stronger than balancing selection, the double-polymorphic equilibrium is stable under condition 10 and only if α is large enough (fig. 2b). Increasing α increases the amount of recombination that can be tolerated without destabilizing the double-polymorphic equilibrium (fig. 2).

The increase of the parameter range with a stable double-polymorphic equilibria with an increasing α has an intuitive interpretation. When locus A is at equilibrium ($p_A = q_A = 0.5$), the genotypes $A_0A_0B_0B_0$ and $A_1A_1B_1B_1$ are the most-fit ones. However, due to their coexistence in a large population and random mating, less fit heterozygote A_1A_0 genotypes will be generated. Among those, $A_1A_0B_1B_0$ are the fittest whenever $\alpha > 0$. Therefore, the reason for the stabilizing effect of α is essentially selection in favor of double heterozygotes that manifests itself when locus A is at or close to equilibrium.



(a) $s < t$ (b) $s > t$

Figure 2: The recombination rate that can be tolerated without destabilizing the double-polymorphic equilibrium increases with the dominance of epistatic effects α. The figure shows critical r, below which the double polymorphic equilibrium is stable, against α in cases where (a) balancing selection is stronger than epistatic selection and (b) epistatic selection is stronger than balancing selection. Solid line represents numerical analyses and dashed lines analytical result. When balancing selection is stronger than epistatic selection and $\alpha = 0$, $r_c = (t^2 - s^2)/(4t(2-t))$ (black line). When epistatic selection is stronger than balancing selection, double-polymorphic equilibrium is not stable under low values of α.

4 Discussion

Whether sign epistasis, a form of interaction between alleles across loci where certain combinations confer high fitness, can promote genetic diversity within populations is a long-standing question. A necessary condition for this effect is the existence of stable polymorphism in a region under balancing selection. Because recombination tends to break up associations between favorable combinations of alleles (Hansen, 2013), recombination tends to decrease the stability of double-polymorphic equilibria and the general conclusion from the analysis of symmetric viability models is that linkage must be tight for double polymorphic equilibrium to be stable. Our model is not a symmetric viability model due to negative frequency-dependence. We can, nevertheless, compare our findings with previous results at equilibrium for the diploid case (the focus in previous work) by setting $p_A = q_A = 1/2$. Indeed, we recover our results for D_{eq} (eq. 9) in Kimura (1956) if we set $\alpha = 0$, or in Karlin & Feldman (1970) and Bodmer & Felsenstein (1967) for any α.

As for the stability conditions, Kimura (1956), Karlin & Feldman (1970) and Bodmer & Felsenstein (1967) found that a stable coexistence of two highly adapted genotypes is only possible if linkage is tighter than balancing selection ($r \lesssim t$). We recovered this conclusion for the haploid case, but this differs dramatically from what we observe in diploids where the parameter α tunes between additive×additive (when $\alpha = 0$) and additive×dominant (when $\alpha > 0$) effects in the fitness table. An intuitive way to illustrate the effect of α is to note that when $\alpha > 0$, the fitness of the heterozygote for the B locus is closer to B_1B_1 in a A_1A_1 homozygote background and closer to B_0B_0 in a A_0A_0 homozygote background (fig. 3). This illustrates why we expect $\alpha > 0$ to stabilize double polymorphism. The same stabilizing effect can also occur in general symmetric viability models (Bodmer & Felsenstein, 1967; Karlin & Feldman, 1970), where there is enough flexibility in the parameters to allow for arbitrary dominance relationships within loci. However, in symmetric viability models, the dominance of epistasis does not increase the critical recombination rate as much as in the model with frequency-dependent selection, and it seems that the critical importance of epipstatic dominance was therefore overlooked in previous contributions.

By analogy with a general discussion of context-specific dominance effects (Connallon & Chenoweth, 2019; Curtsinger et al., 1994), we may refer to the case $\alpha > 0$ as a beneficial reversal

of dominance of epistasis because the fitness of the B heterozygote will be similar to the most fit homozygote across different genetic backgrounds. We note that this is in spirit analogous to dominance reversals across other contexts (e.g., environments, sexes, etc), which are known to promote stable polymorphism in single locus scenarios (Connallon & Chenoweth, 2019). Drawing on parallels in single locus scenarios (Siljestam et al., 2024; Spencer & Priest, 2016), we suggest that selection should favor dominance of sign epistasis since it effectively reduces segregation load. Because the process we describe here is predicted to extend the life span of alleles across interacting loci, it may provide sufficient time to allow the evolution of elevated dominance of epistasis which, in turn, will tend to further stabilize double-polymorphism. We note that context-specific dominance effects can emerge from a variety of non-linearities at any level of the genotype-to-phenotype map (Billiard et al., 2021).

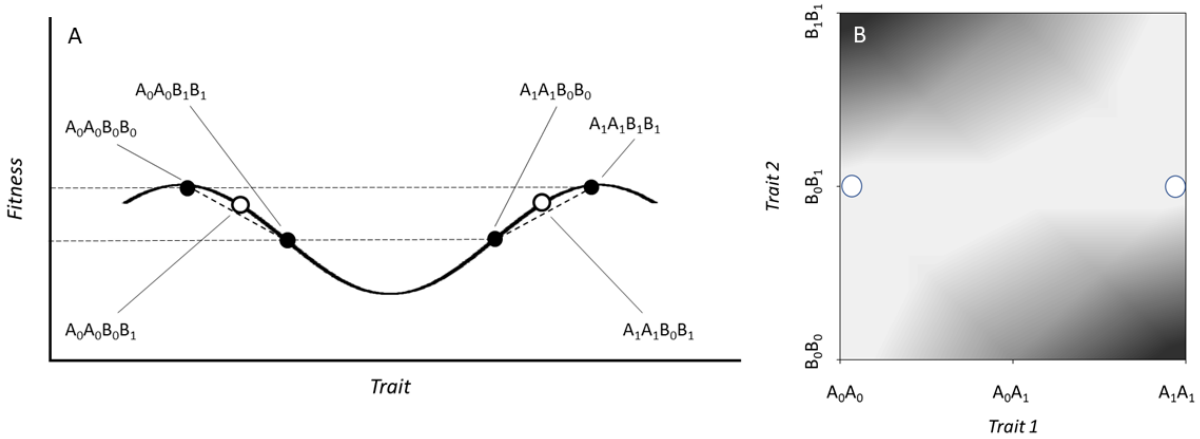


Figure 3: For polygenic traits under non-linear selection, epistatic interactions between loci are inevitable, and curved fitness landscapes will then lead to complex allelic effects across loci. Shown are two schematic and simplified illustrations of ways in which dominance of sign epistasis could emerge. In these, genotype determines individual trait values which in turn translate into individual fitness. Under balancing selection, the relation between fitness and trait depends on allele frequencies. To simplify illustration, the frequencies at locus A are here fixed at $p_A = q_A = 1/2$. (A) When two loci have additive effects on a single phenotypic trait with two distinct fitness peaks (i.e., non-linear disruptive selection), the fitness of the heterozygotes for the B locus (open circles) may be closer to the homozygote B_0B_0 in the A_0A_0 background and closer to B_1B_1 in the A_1A_1 background, such that the dominance of the two B alleles is reversed across A genetic backgrounds (A_0A_1 heterozygotes not included here, for simplicity). (B) Alternatively, for two loci coding for distinct traits, correlational selection will also result in epistasis for fitness (light - high relative fitness, dark - low relative fitness). This may also reverse the dominance of the two B alleles across A genetic backgrounds as a result of the curvature of the fitness landscape. In this example, the fitness of the heterozygotes for the B locus (open circles) is closer to the homozygote B_0B_0 in the A_0A_0 background and closer to B_1B_1 in the A_1A_1 background.

Sources of epistasis. The scenario studied here can be seen as representing a variety of mechanisms for the emergence of sign epistasis. This can appear on a genetic level in the form of co-adapted amino acids or nucleotide substitutions (Kondrashov et al., 2002) or more generally as co-adapted gene complexes (Dobzhansky & Pavlovsky, 1958). A general possibility is the emergence of sign epistasis for fitness as a result of non-linear fitness functions for polygenic traits (fig. 3). For example, it has been suggested that major effect life history genes may frequently experience balancing selection (Arnqvist & Rowe, 2023; Christie & McNickle, 2023) and the processes we describe may act to maintain variation in loci that show epistasis with such

genes. There are many potential examples of the scenario we explore (Arnqvist & Rowe, 2023; Christie & McNickle, 2023). A common case may be inversion polymorphism maintained by balancing selection (Berdan et al., 2022), where linkage disequilibrium between the inversion and other loci are commonly observed and actually constituted a very early demonstration of epistatic selection (Prakash & Lewontin, 1968). More specific examples include a polymorphic major effect life history locus located on the sex-chromosomes in *Drosophila mercatorum*, which interacts with several autosomal modifiers all of which are highly polymorphic in natural population (Templeton, 2000). A more familiar case is color pattern in male guppies, which shows high levels of standing genetic variation maintained by negative frequency-dependent selection. New work suggests that although a major effect locus is located on the Y-chromosome, color pattern is determined by an epistatic interaction between Y and an autosomal region in LD with the Y (Paris et al., 2022). Similarly, major effect variants that affect coloration in butterfly mimicry systems maintained by balancing selection have epistatic modifiers that apparently fine-tune coloration (Nijhout, 1991), and balancing selection on inter-locus allele combinations is likely important in maintaining polymorphism in flower morphology in plants (Kelly, 2000). A distinct but classic example may in fact be sickle-cell anemia in humans, which is controlled by the major effect HBB locus well known to be under balancing selection. It is perhaps less well known that the phenotypic manifestation of HBB genotype is determined by epistatic interactions with several other polymorphic loci (Templeton, 2000). Another illustrative example comes from supergenes (chromosomal inversions) that control social behavior in some ants and are maintained as balanced polymorphisms (Purcell & Brelsford, 2025). In the ant *Formica cinerea*, a second supergene on a different chromosome influences body size and is likewise polymorphic. Despite being unlinked, the two supergenes show strong linkage disequilibrium, which appears to be maintained by strong epistatic selection (Scarparo et al., 2026). Finally, an analogue to the scenario we explore is loci under sexually antagonistic selection, where different alleles are favored in males and females (Bonduriansky & Chenoweth, 2009). Such loci essentially exhibit sign epistasis for fitness with sex-determining loci under balancing selection, which can act to maintain variation in sexually antagonistic loci given permissive patterns of dominance of epistasis (Siljestam et al., 2024).

Possible scenarios for the accumulation of epistatic polymorphisms. Suppose that we start with locus A under balancing selection. Alleles at other loci favored in one genetic background but not the other may then accumulate if linkage is tight enough. This will generate indirect selection for reduced recombination rate (Otto & Feldman, 1997), a scenario analogous to the canonical model of sex chromosome evolution. Indeed, our analysis of the haploid case shows that the stability of a double polymorphism requires $r \lesssim t$.

However, we show that in diploids, the situation is significantly different because of possible dominance effects. Given balancing selection on locus A, there are at least two scenarios where these sorts of effects could be important. First, in cases where epistatic balancing selection maintains two-locus polymorphism for long enough, selection will favor the evolution of "beneficial" dominance effects (i.e., selection will favor the evolution of larger α), with effects similar to reduced recombination. This will, in turn, further stabilize the polymorphism. Second, as new mutations with epistatic effects enter into a population, those with permissive epistatic dominance effects (i.e., $\alpha > 0$) will tend to be retained by epistatic balancing selection. This way, selection may act as a sieve favoring the maintenance of, or increased lifespan of, mutations showing overall $\alpha > 0$. In both cases, we would expect the build-up of genetic polymorphism and $\alpha > 0$ to be common. Needless to say, the build-up of genetic polymorphism will be generally promoted by linkage, but as we show here, dominance effects make close linkage a less critical requirement.

How to detect the effect in real data? Uncovering the effects we discuss here using genome scans or population genomic data alone will clearly be very challenging (Arnqvist & Rowe, 2023). Co-adaptation of alleles at different loci should lead to positive linkage disequi-

librium between them and the tools available for identifying regions under balancing selection are improving (Bitarello et al., 2023). Scanning for loci under balancing selection and inspecting regions or sites in positive LD with those may therefore be rewarding. For example, Stolyarova et al. (2022) found pervasive elevated short-range LD between nonsynonymous sites within divergent intra-genic haploblocks in the fungus *Schizophyllum communae*. This finding suggests that this may be due to a combination of balancing and epistatic selection, potentially similar to the scenario we model here. Another possibility may be to empirically estimate the dominance of epistasis (i.e., α), through for example advanced line-cross analyses (Burch et al., 2024), as this is expected to be positive under stable double-polymorphism. Genomic time-series data may also offer interesting routes to assess epistatic selection, as manifested by synchronous frequency shifts in populations.

In summary, we delineate the conditions under which the diversity-promoting effect of balancing selection is expected to extend to other loci through epistasis. We find that the dominance of epistasis is central in this context, and a diversifying effect is predicted even under fairly high rates of recombination if this dominance is sizeable.

Data availability

The Python code used for the numerical analyzes is available online at `https://github.com/khudyakovaks/Sign-epistasis-extends-the-effects-of-balancing-selection-on-genetic-diversity`.

Author contributions

GA, NB, and K.K. conceived and designed the analysis; NB and KK performed the analysis; GA and KK wrote the paper;

Funding

This work was funded by grants from the Swedish Research Council (2023-03730 to G.A.) and the DOC fellowship from the Austrian Academy of Science (26293 to K.K.).

Conflict of interest

The authors declare no conflicts of interest.

Acknowledgments

We thank L. Rowe for valuable discussions.

References

Arnqvist, G., & Rowe, L. (2023). Ecology, the pace-of-life, epistatic selection and the maintenance of genetic variation in life-history genes. *Molecular Ecology*, *32*(17), 4713–4724. doi: 10.1111/mec.17062

Barton, N. H., & Keightley, P. D. (2002). Understanding quantitative genetic variation. *Nature Reviews Genetics*, *3*(1), 11–21.

Barton, N. H., & Turelli, M. (1991). Natural and sexual selection on many loci. *Genetics*, *127*(1), 229–255. Retrieved from `http://dx.doi.org/10.1093/genetics/127.1.229` doi: 10.1093/genetics/127.1.229

Barton, N. H., & Turelli, M. (2004). Effects of genetic drift on variance components under a general model of epistasis. *Evolution*, *58*(10), 2111–2132. doi: 10.1111/j.0014-3820.2004.tb01591.x

Berdan, E. L., Flatt, T., Kozak, G. M., Lotterhos, K. E., & Wielstra, B. (2022). Genomic architecture of supergenes: connecting form and function. *Philosophical Transactions of the Royal Society B*, *377*(1856), 20210192.

Billiard, S., Castric, V., & Llaurens, V. (2021). The integrative biology of genetic dominance. *Biological Reviews*, *96*(6), 2925–2942.

Bitarello, B. D., Brandt, D. Y. C., Meyer, D., & Andrés, A. M. (2023). Inferring balancing selection from genome-scale data. *Genome Biology and Evolution*, *15*(3). doi: 10.1093/gbe/evad032

Bloom, J. S., Ehrenreich, I. M., Loo, W. T., Lite, T.-L. V., & Kruglyak, L. (2013). Finding the sources of missing heritability in a yeast cross. *Nature*, *494*(7436), 234–237.

Bodmer, W. F., & Felsenstein, J. (1967). Linkage and selection: Theoretical analysis of the deterministic two locus random mating model. *Genetics*, *57*(2), 237–265. Retrieved from `http://dx.doi.org/10.1093/genetics/57.2.237` doi: 10.1093/genetics/57.2.237

Bonduriansky, R., & Chenoweth, S. F. (2009). Intralocus sexual conflict. *Trends in Ecology & Evolution*, *24*(5), 280–288. doi: 10.1016/j.tree.2008.12.005

Bonnet, T., Morrissey, M. B., de Villemereuil, P., Alberts, S. C., Arcese, P., Bailey, L. D., ... Kruuk, L. E. B. (2022). Genetic variance in fitness indicates rapid contemporary adaptive evolution in wild animals. *Science*, *376*(6596), 1012-1016. doi: 10.1126/science.abk0853

Bourg, S., Bolstad, G. H., Griffin, D. V., Pélabon, C., & Hansen, T. F. (2024). Directional epistasis is common in morphological divergence. *Evolution*, *78*(5), 934–950. doi: 10.1093/evolut/qpae029

Burch, J., Chin, M., Fontenot, B. E., Mandal, S., McKnight, T. D., Demuth, J. P., & Blackmon, H. (2024). Wright was right: leveraging old data and new methods to illustrate the critical role of epistasis in genetics and evolution. *Evolution*, *78*(4), 624–634. doi: 10.1093/evolut/qpae003

Charlesworth, B. (2015). Causes of natural variation in fitness: evidence from studies of drosophila populations. *Proceedings of the National Academy of Sciences*, *112*(6), 1662–1669.

Charlesworth, D. (2006). Balancing selection and its effects on sequences in nearby genome regions. *PLoS Genetics*, *2*(4), e64.

Christiansen, F. B. (1999). *Population genetics of multiple loci.* Chichester, UK; New York, NY: John Wiley & Sons.

Christie, M. R., & McNickle, G. G. (2023). Negative frequency dependent selection unites ecology and evolution. *Ecology and Evolution*, *13*(7). doi: 10.1002/ece3.10327

Connallon, T., & Chenoweth, S. F. (2019). Dominance reversals and the maintenance of genetic variation for fitness. *PLOS Biology*, *17*(1), e3000118. doi: 10.1371/journal.pbio.3000118

Corbett-Detig, R. B., Zhou, J., Clark, A. G., Hartl, D. L., & Ayroles, J. F. (2013). Genetic incompatibilities are widespread within species. *Nature*, *504*(7478), 135–137.

Curtsinger, J. W., Service, P. M., & Prout, T. (1994). Antagonistic pleiotropy, reversal of dominance, and genetic polymorphism. *The American Naturalist*, *144*(2), 210–228.

Dobzhansky, T., & Pavlovsky, O. (1958). Interracial hybridization and breakdown of coadapted gene complexes in Drosophila Paulistorum and Drosophila Willistoni. *Proceedings of the National Academy of Sciences*, *44*(6), 622–629. doi: 10.1073/pnas.44.6.622

Faria, R., Johannesson, K., Butlin, R. K., & Westram, A. M. (2019). Evolving inversions. *Trends in ecology & evolution*, *34*(3), 239–248.

Gavrilets, S. (2004). *Fitness landscapes and the origin of species.* Princeton University Press.

Good, B. H., McDonald, M. J., Barrick, J. E., Lenski, R. E., & Desai, M. M. (2017). The dynamics of molecular evolution over 60,000 generations. *Nature*, *551*(7678), 45–50.

Hansen, T. F. (2013). Why epistasis is important for selection and adaptation: Perspective. *Evolution*, *67*(12), 3501–3511. doi: 10.1111/evo.12214

Hemani, G., Knott, S., & Haley, C. (2013). An evolutionary perspective on epistasis and the missing heritability. *PLoS Genetics*, *9*(2), e1003295.

Johnson, N. A. (2000). Gene interactions and the origin of species. In *Epistasis and the evolutionary process* (p. 197-221). Oxford University Press.

Karlin, S., & Feldman, M. W. (1970). Linkage and selection: Two locus symmetric viability model. *Theoretical Population Biology*, *1*(1), 39–71. Retrieved from `http://dx.doi.org/10.1016/0040-5809(70)90041-9` doi: 10.1016/0040-5809(70)90041-9

Kelly, J. K. (2000). Epistasis, linkage and balancing selection. In *Epistasis and the evolutionary process* (p. 146-157). Oxford University Press.

Kelly, J. K., & Wade, M. J. (2000). Molecular evolution near a two-locus balanced polymorphism. *Journal of Theoretical Biology*, *204*(1), 83–101.

Kimura, M. (1956). A model of a genetic system which leads to closer linkage by natural selection. *Evolution*, *10*(3), 278–287. Retrieved from `http://dx.doi.org/10.1111/j.1558-5646.1956.tb02852.x` doi: 10.1111/j.1558-5646.1956.tb02852.x

Kirkpatrick, M., Johnson, T., & Barton, N. H. (2002). General models of multilocus evolution. *Genetics*, *161*(4), 1727–1750. doi: 10.1093/genetics/161.4.1727

Kondrashov, A. S., Sunyaev, S., & Kondrashov, F. A. (2002). Dobzhansky–Muller incompatibilities in protein evolution. *Proceedings of the National Academy of Sciences*, *99*(23), 14878–14883. doi: 10.1073/pnas.232565499

Kroymann, J., & Mitchell-Olds, T. (2005). Epistasis and balanced polymorphism influencing complex trait variation. *Nature*, *435*(7038), 95–98. doi: 10.1038/nature03480

Mäki-Tanila, A., & Hill, W. G. (2014). Influence of gene interaction on complex trait variation with multilocus models. *Genetics*, *198*(1), 355–367. doi: 10.1534/genetics.114.165282

Mousseau, T. A., & Roff, D. A. (1987). Natural selection and the heritability of fitness components. *Heredity*, *59*(2), 181–197. doi: 10.1038/hdy.1987.113

Nijhout, H. F. (1991). *The development and evolution of butterfly wing patterns.* Smithsonian Institution.

Orr, H. A., & Turelli, M. (2001). The evolution of postzygotic isolation: Accumulating Dobzhansky-Muller incompatibilities. *Evolution*, *55*(6), 1085–1094. doi: 10.1111/j.0014-3820.2001.tb00628.x

Ortiz-Barrientos, D., Engelstädter, J., & Rieseberg, L. H. (2016). Recombination rate evolution and the origin of species. *Trends in Ecology & Evolution*, *31*(3), 226–236. doi: 10.1016/j.tree.2015.12.016

Otto, S. P., & Feldman, M. W. (1997). Deleterious mutations, variable epistatic interactions, and the evolution of recombination. *Theoretical Population Biology*, *51*(2), 134-147. doi: https://doi.org/10.1006/tpbi.1997.1301

Paixão, T., & Barton, N. H. (2016). The effect of gene interactions on the long-term response to selection. *Proceedings of the National Academy of Sciences*, *113*(16), 4422–4427.

Paris, J. R., Whiting, J. R., Daniel, M. J., Ferrer Obiol, J., Parsons, P. J., van der Zee, M. J., ... Fraser, B. A. (2022). A large and diverse autosomal haplotype is associated with sex-linked colour polymorphism in the guppy. *Nature Communications*, *13*(1), 1233.

Pettersson, M., Besnier, F., Siegel, P. B., & Carlborg, Ö. (2011). Replication and explorations of high-order epistasis using a large advanced intercross line pedigree. *PLoS Genetics*, *7*(7), e1002180.

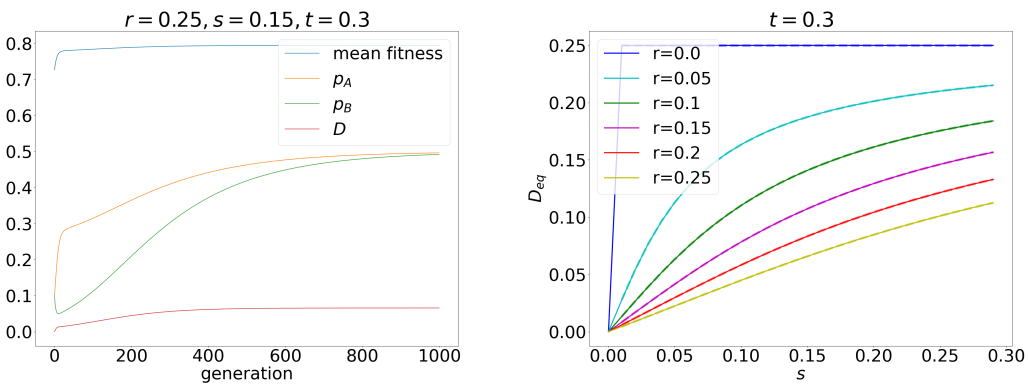
Poelwijk, F. J., Tănase-Nicola, S., Kiviet, D. J., & Tans, S. J. (2011). Reciprocal sign epistasis is a necessary condition for multi-peaked fitness landscapes. *Journal of Theoretical Biology*, *272*(1), 141–144. doi: 10.1016/j.jtbi.2010.12.015

Prakash, S., & Lewontin, R. (1968). A molecular approach to the study of genic heterozygosity in natural populations. 3. Direct evidence of coadaptation in gene arrangements of Drosophila.

Proceedings of the National Academy of Sciences, *59*(2), 398–405.
Purcell, J., & Brelsford, A. (2025). Supergenes in organismal and social development of insects: ideas and opportunities. *Current Opinion in Insect Science*, *68*, 101303.
Roff, D. A., & Emerson, K. (2006). Epistasis and dominance: Evidence for differential effects in life-history versus morphological traits. *Evolution*, *60*(10), 1981–1990. doi: 10.1111/j.0014-3820.2006.tb01836.x
Saona, R., Kondrashov, F. A., & Khudiakova, K. A. (2022). Relation between the number of peaks and the number of reciprocal sign epistatic interactions. *Bulletin of Mathematical Biology*, *84*(8). doi: 10.1007/s11538-022-01029-z
Scarparo, G., Lopez, A., Muro, E., Brelsford, A., & Purcell, J. (2026). Lethal epistasis maintains strong linkage disequilibrium between unlinked supergenes. *bioRxiv*, 2026–01.
Siljestam, M., Rueffler, C., & Arnqvist, G. (2024). Sex-specific dominance and its effects on allelic diversity in sexually antagonistic loci. *bioRxiv preprint*. doi: 10.1101/2024.05.06.592709v2
Spencer, H. G., & Priest, N. K. (2016). The evolution of sex-specific dominance in response to sexually antagonistic selection. *The American Naturalist*, *187*(5), 658–666.
Stolyarova, A. V., Neretina, T. V., Zvyagina, E. A., Fedotova, A. V., Kondrashov, A. S., & Bazykin, G. A. (2022). Complex fitness landscape shapes variation in a hyperpolymorphic species. *eLife*, *11*, e76073. Retrieved from `https://doi.org/10.7554/eLife.76073` doi: 10.7554/eLife.76073
Templeton, A. R. (2000). Epistasis and complex traits. In *Epistasis and the evolutionary process* (p. 41-57). Oxford University Press.
Wade, M. J. (2000). Epistasis as a genetic constraint within populations and an accelerant of adaptive divergence among them. In *Epistasis and the evolutionary process* (p. 213-231). Oxford University Press.
Wolf, J. B., Brodie, E. D., & Wade, M. J. (2000). *Epistasis and the evolutionary process*. Oxford University Press.

A Supplementary Material

A.1 Haploids



(a) (b)

Figure 4: (a) Change of p_A, p_B, D and mean fitness over time and (b) the relationship between the strength of epistatic selection s and linkage disequilibrium D. Solid lines represent numerical iteration of selection equations and dashed lines analytical prediction (3).

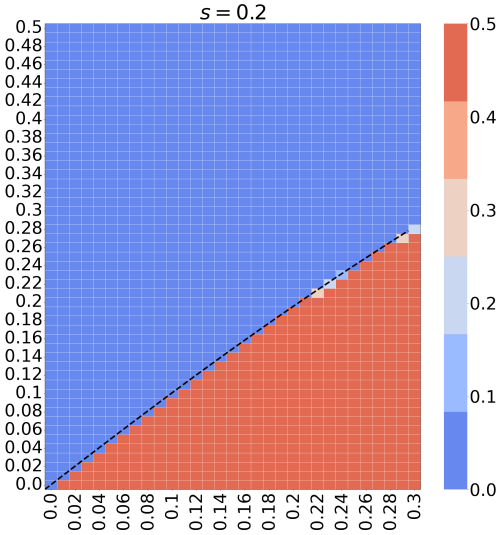
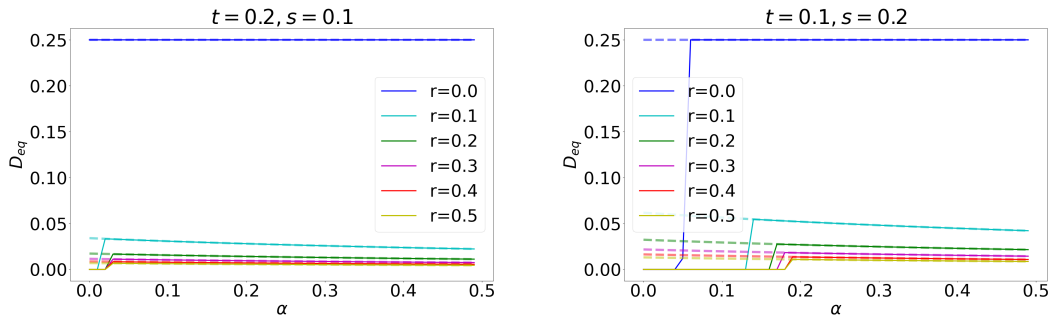


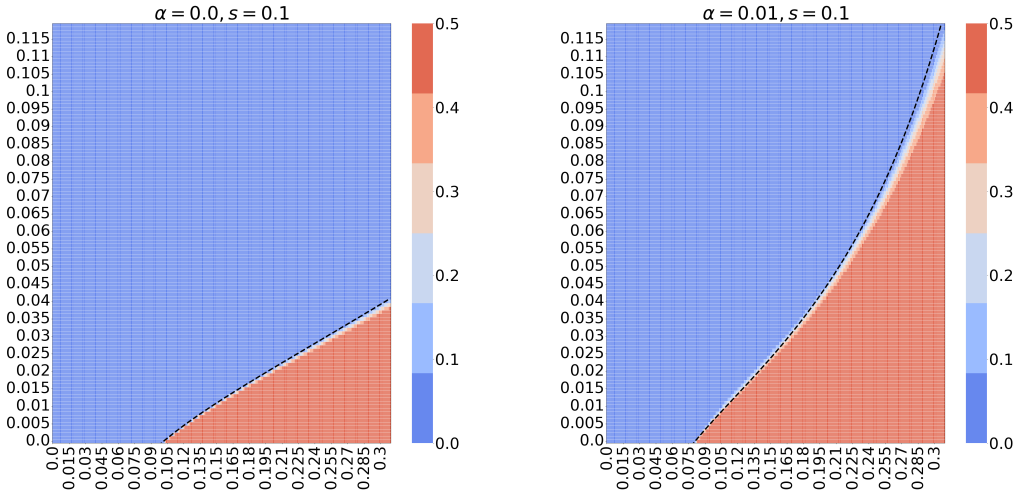
Figure 5: Equilibrium frequency of allele B_1 in the numerical iteration of selection equations on the (r, t) space. Dashed line represents stability condition (4).

A.2 Diploids



(a) $s < t$ (b) $s > t$

Figure 6: Linkage disequilibrium D at double-polymorphic equilibria decreases with increasing recombination rate, but is contingent upon the strength of selection and epistatic dominance α. Conditions where (a) balancing selection is stronger than epistatic selection and (b) epistatic selection is stronger than balancing selection. Solid lines represent numerical analysis and dashed lines are analytical prediction (9).



(a) $\alpha = 0$

(b) $\alpha > 0$

Figure 7: Equilibrium p_B in the iteration of selection equations on the (r, t) space. The dashed lines are critical r below which the double polymorphism is predicted to be stable: $r_c = (t^2 - s^2)/(8t - 4t^2)$ when $\alpha = 0$ and as in (10) when $\alpha > 0$.