A gene-based model of fitness and its implications for genetic variation

Parul Johri[1,2,3,*] and Brian Charlesworth[4,*]

[1]Department of Biology, University of North Carolina at Chapel Hill, NC 27599, United States
[2]Department of Genetics, University of North Carolina at Chapel Hill, NC 27599, United States
[3]Integrative Program for Biological and Genome Sciences, University of North Carolina at Chapel Hill, NC 27599, United States
[4]Institute of Evolutionary Biology, School of Biological Sciences, University of Edinburgh, EH9 3FL, United Kingdom

*correspondence may be addressed to these authors: pjohri@unc.edu; brian.charlesworth@ed.ac.uk

Abstract

A widely used model of the effects of mutations on fitness (the "sites" model) assumes that heterozygous recessive or partially recessive deleterious mutations at different sites in a gene complement each other, similarly to mutations in different genes. However, the general lack of complementation between major effect allelic mutations suggests an alternative possibility, the "gene" model. This assumes that a pair of heterozygous deleterious mutations in *trans* behave effectively as homozygotes, so that the fitnesses of *trans* heterozygotes are lower than those of *cis* heterozygotes. We examine the properties of the two different models, using both analytical and simulation methods. We show that the gene model results in a slightly lower mutational load, but a much smaller inbreeding load, than the sites model, implying that standard predictions of mutational contributions to inbreeding depression may be overestimates. The gene model also predicts positive linkage disequilibrium (LD) between derived variants within the coding sequence under conditions when the sites model predicts zero or slightly negative LD. We also show that focussing on rare variants when examining patterns of LD, especially with Lewontin's D' measure, is likely to produce misleading results with respect to inferences concerning the causes of the sign of LD. Synergistic epistasis between pairs of mutations was also modeled; it is less likely to produce negative LD under the gene model than the sites model. The theoretical results are discussed in relation to data on inbreeding load in *Drosophila melanogaster* and patterns of LD in natural populations of several species.

Keywords: fitness model, genetic load, inbreeding load, linkage disequilibrium, epistasis

Running title: Fitness models and genetic variation

Introduction

There is much interest in the effects of deleterious mutations on properties of populations such as their genetic load, inbreeding depression, and variance in fitness, tracing back to the pioneering work of Haldane (1937), Muller (1950), Morton et al. (1956) and Crow (1958), and to more recent studies of the genetics of *Drosophila* fitness components (reviewed by Crow 1993; Charlesworth 2015). A large volume of recent work on this topic has been stimulated by concerns about the implications of deleterious mutations for the survival of small, endangered populations, and the availability of population genetic data on non-model organisms that shed light on the prevalence of deleterious mutations in natural populations (e.g., Teixeira and Huber 2021; Kardos et al. 2021; Bertorelli et al. 2023; Kyriazis, Robinson and Lohmueller 2023). There are, however, sharp disagreements about the interpretation of these data (Kyriazis et al., 2022, 2023; García-Dorado and Hedrick 2023), which indicate the need for a secure theoretical basis for data analyses.

The theoretical predictions of these properties of populations, which we will refer to as "load statistics", are mostly based on population genetics models of diploid organisms that tacitly assume that nucleotide sites subject to mutation and selection can be treated as independent of each other with respect to their fitness effects (reviewed by Kyriazis, Robinson and Lohmueller 2023). Models of this type were first introduced by Haldane (1937), who assumed that fitnesses were multiplicative across loci. This assumption implies that individuals carrying two haplotypes that each carry a deleterious mutation at two different sites within the coding sequence can be treated as being heterozygous at each site (Figure 1A). With the same selection coefficients against homozygotes (s) and dominance coefficients (h) at each site, the fitness of these individuals is $1 - 2hs$ if effects are additive across sites, regardless of whether or not the two mutations are carried on the same haplotype (*cis*), or on the maternal and paternal haplotypes (*trans*). In contrast, if the gene is treated as a single entity that is subject to mutation, and there is no allelic complementation between mutations within the same gene (more accurately, a cistron as defined by Benzer [1957]), the fitness of an individual that is heterozygous for two mutations depends on whether or not they are in *cis* or *trans* (Figure 1B).

The difference between these two possibilities can strongly affect the effective level of recessivity experienced by deleterious mutations, as can be seen from the following toy model, where we describe the two cases as the "sites" and "gene" models, respectively. Consider a single coding sequence that is segregating for mutations at two different nucleotide sites (1 and 2) with the same selection coefficient s against homozygotes at each site. Assume that the double mutant haplotypes are sufficiently rare that they can be ignored in what follows. Let the frequency of mutant variants at site i be q_i. Then haplotypes carrying a mutation at site 1 will meet a haplotype that is wild-type at site 1 and mutant at site 2 with an approximate frequency of p_1q_2; the fitness of this genotype is $1 - s$ if there is no allelic complementation (*i.e.*, under the gene model), but the sites model would assign it a fitness of $1 - 2hs$ (assuming additive effects across sites). It will meet haplotypes that are mutant at site 1 but wild-type at site 2 with frequency q_1p_2. Both models would assign fitness $1 - s$ to this genotype. A similar argument applies to a haplotype carrying a mutation at site 2, with the appropriate switch between subscripts 1 and 2.

All of the double mutant genotypes are classed as homozygotes under the gene model, which is biologically realistic for the majority of genes where sites with allelic complementation are relatively rare (Fincham et al. 1979, pp.348-353; Hawley and Gilliland 2006). The net difference between the gene and sites models in the fitness reduction assigned to haplotypes that each carry a mutation at one or other sites is thus:

$$\delta w = (1 - 2h)s\{\frac{p_1q_2}{q_1} + \frac{p_2q_1}{q_2}\} \qquad (1)$$

If the frequencies of the deleterious mutations at the two sites are comparable, so that the subscripts can be dropped, $\delta w \approx 2(1 - 2h)sp$. If $h << 0.5$, this quantity could be quite substantial compared to s; In the case of full recessivity ($h = 0$), $\delta w \approx 2sp$. This implies that the sites model can substantially underestimate the true homozygous fitness effects of deleterious mutations. It follows that deleterious mutations are more likely to segregate at higher allele frequencies under the sites model than the gene model, leading to a lower mean fitness of the population, a higher genetic load, and a larger effect of inbreeding on fitness.

In addition, the fact that double heterozygotes with mutations in *cis* under the gene model have higher fitnesses than those with mutations in *trans* implies an intragenic position effect with respect to fitness when $h < ½$. This is likely to favor positive linkage disequilibrium (LD) relative to the sites model, if the sign of the coefficient of linkage

disequilibrium (D) is defined as positive when the frequencies of haplotypes carrying pairs of deleterious mutant alleles and pairs of wild-type alleles are in excess of random expectation. Position effects of this kind have rarely been included in population genetics models, with the exception of a study of two-locus polymorphisms by Nordborg et al. (1995), and the models of mutation and selection at multiple sites within a gene by Thornton et al. (2013) and Sanjak et al. (2015) and at two sites by Ragsdale (2022), who used somewhat different fitness models from ours. Such gene-based fitness models are likely to result in patterns of LD between deleterious mutations that are different from those expected under the sites model. Moreover, a number of recent theoretical studies of several different fitness models have yielded conflicting results on measures of LD between deleterious mutations (e.g., Garcia and Lohmueller 2021; Roze 2021; Good 2022; Ragsdale 2022).

The purpose of the present study is to examine the consequences of the differences between the gene and sites models for the load statistics and for patterns of linkage disequilibrium (LD) among variants at sites under selection, using a model of a single coding sequence subject to deleterious mutations. An analytical treatment of a two-site sequence in an infinite, randomly mating population is given first, followed by stochastic simulations of a multi-site sequence, intended to provide an approximate description of a coding sequence. We show that the difference between the two models has significant consequences for the level of inbreeding depression caused by deleterious mutations, and for the patterns of LD among such mutations. Interpretations of patterns of LD observed in natural populations need to take account of the possibility that the gene model is a more accurate representation of the dominance properties of deleterious mutations.

Methods

Multi-locus simulations

Forward simulations were performed using SLiM 4.0 (Haller and Messer 2023) with a Wright-Fisher population of N =1000 diploid individuals, a mutation rate $u = 0.5 \times 10^{-5}$ per nucleotide site/generation, and a rate of crossing over of 10^{-5} per nucleotide site/generation. A gene with 1000 selected sites was simulated, assuming a dinucleotide model, *i.e.*, there were two alleles at each site: a wildtype (denoted by +) and a selectively deleterious mutant (denoted by –) allele. Note that the two alleles were equally likely to mutate to each other, i.e., the + allele could mutate to the – allele at rate u and vice versa. This model contrasts with similar previous studies (Garcia and Lohmueller 2021; Good 2022; Ragsdale 2022), where

mutation was unidirectional from + to –. This procedure has the advantage that it allows a true stationary distribution of allele frequencies (Wright 1931), in contrast to models with one-way mutation, where all sites will eventually become fixed for mutant alleles.

Simulations were run either assuming the same selection coefficient at each site or a distribution of fitness effects (DFE). In the first case, the selective disadvantage of mutant homozygotes scaled by twice the population size was fixed at $\gamma = 2Ns = 2$, 20, 100, or 1000. When simulating a distribution of fitness effects (DFE), a gamma distribution was assumed for most of the results presented here, with shape parameter α = 0.3 and means of $\bar{\gamma} = 2N\bar{s}$ = 2, 20, 100, or 1000. The homozygous selective disadvantage of a – allele at a particular site remained constant throughout a given simulation, although it varied across simulation replicates. The dominance coefficient (h) of the mutant allele was fixed at 0.0, 0.2, or 0.5. The fitness of a diploid individual was calculated using either the gene or the sites model, as described below. All sites (polymorphic and fixed) were included when calculating the fitness of individuals. Simulations were run for $10N$ generations. For each scenario, 1000 replicates were simulated.

Relating the simulations to Drosophila populations

These simulation parameters can be placed in the context of the *D. melanogaster* genome, there are on average about 5 exons per gene with a mean length of ~300 bp (Misra et al. 2002). As only about two-thirds of coding sites are nonsynonymous, there are expected to be ~1000 (= 300 × 5× 2/3) selected sites per gene, corresponding well to our simulations. Intron lengths, and hence, gene lengths vary across the genome, with a mean and median length of 5 kb and 2 kb respectively. Assuming $N_e = 10^6$ for *D. melanogaster* and a sex-averaged rate of cross-over of 10^{-8} per site/generation (Johri et al. 2020), $NrL \approx 20 - 50$ over a genic region, where L is the mean length of a gene. With our simulation parameters, the 1000 selected sites have $NrL = 10$ when rescaled to a population size of 10^6, and therefore approximately represent a genic region of *Drosophila*. Silent site polymorphism data imply that $Nu \approx 0.003 - 0.005$ in *Drosophila* populations (Johri et al. 2020), consistent with the $Nu = 0.005$ used in our simulations.

Simulations of human populations

A Wright-Fisher population of N = 1000 diploid individuals was simulated with a mutation rate $u = 2.5 \times 10^{-7}$per nucleotide site/generation and crossing over rate (r) of 2×10^{-7}per

nucleotide site/generation. This resulted in Nu = 0.00025 and Nr = 0.0002 in our simulations, which are consistent with the estimates of these parameters for human populations: N_e of 20,000, mutation rate of 1.25×10^{-8} per site/generation (Kong et al., 2012), and recombination rate of 1.0×10^{-8} per site/generation (Altshuler et al., 2005). Assuming that there are 8 exons per gene on average (Johri et al. 2021), that each exon is about 350 bp, and that 70% of all coding sites in humans are nonsynonymous (Kryukov et al. 2007), there should be about 1960 selected sites in a gene. We therefore simulated 2000 selected sites. The DFE was assumed to be a gamma distribution with the shape parameter α = 0.23 and $\bar{\gamma} = 850$, as estimated by Eyre-Walker et al. (2006) and Kim et al. (2017).

Modeling the fitnesses of individuals

As mentioned in the Introduction, two different models of main effects on fitness (the sites and gene models) were used to calculate the fitnesses of diploid genotypes. The sites model assumes that heterozygous mutations at all sites complement each other fully. The gene model assumes a complete absence of complementation between heterozygous mutations present in *trans*. In both cases, additivity of the main effects across sites was assumed, in contrast to the multiplicative fitness model that is often used in multi-locus studies. For a single gene with a modest number of segregating sites, as simulated here, there should be little difference between additive and multiplicative models, unless selection is much stronger than is generally assumed here. In addition to the main effects of each site, a simple model of pairwise epistasis was used, in which a single parameter (ϵ) describes a pairwise interaction effect on fitness of mutations. The total number of sites within a gene is denoted by n. The parameter ϵ is positive if there is synergistic epistasis, so that double mutants reduce fitness by more than is expected from the effects of each mutation on its own, and negative if there is antagonistic (diminishing returns) epistasis, where double mutants behave in the opposite fashion (Crow 1970; Kondrashov 1988).

The analytical results derived in the Appendix are for the case of n = 2, with the same selection coefficient s at each site. In order to avoid multiple epistatic parameters, we assume that the magnitude of epistasis scales linearly with the sizes of the selective effects at each site with a constant of proportionality ϵ, so that the fitness of a mutant double homozygote is $1 - 2s - 2\epsilon s$. The fitnesses of all possible genotypes under the sites model are defined in the upper section of Table 1, where the assumption of linear scaling and complete complementation of mutations at different sites in double heterozygotes gives $e_3 = 2\epsilon s$, $e_1 =$

$2\epsilon hs$ and $e_2 = \epsilon(1+h)s$. For the gene model, a double heterozygote with mutations in *cis* has fitness $1 - 2hs - 2\epsilon hs$, whereas the absence of complementation in a *trans* double heterozygote implies that its fitness is $1 - s$. It is less obvious how to assign the fitness of a genotype with one site homozygous for a mutation and the other heterozygous; we simply assume that a lack of complementation implies that these can be treated like homozygotes, with the heterozygous site contributing a fitness reduction of $\frac{1}{2}s$ and the homozygous site a reduction of s, giving a net additive effect of $1.5s$ and an epistatic effect of $1.5\epsilon s$. With $e_3 = 2\epsilon s$, we thus have $e_1 = \epsilon hs$ and $e_2 = 1.5\epsilon s$ in the lower sections of Table 1 and Figure 1.

Modeling fitness with multiple sites and no epistasis

For describing fitness (w) when there are n sites within a gene, as in the simulations, consider the two haplotypes (1 and 2) of a diploid individual, where $w = 1$ when all sites are wild-type. With the same selection coefficient at all sites, fitness for the gene model is calculated as follows.

If there is at least one – allele in 1 or 2, but not both, we have:

$$w = 1 - Vhs \qquad (2a)$$

where V is the total number of – alleles over all n sites.

If there is at least one – allele in both 1 and 2, we have

$$w = 1 - 0.5Ws \qquad (2b)$$

where W is the total number of – alleles present.

In the case of a distribution of fitness effects (DFE), with selection coefficient si, these equations become:

$$w = 1 - h\sum_{i=1}^{V} s_i \qquad (3a)$$

$$w = 1 - 0.5\sum_{i=1}^{W} s_i \qquad (3b)$$

For the sites model with a fixed selection coefficient, we have:

$$w = 1 - Mhs - Ns \qquad (4)$$

where M is the number of heterozygous (+/–) sites and N is the number of mutant homozygous (–/–) sites. With a DFE, Equation (4) becomes:

$$w = 1 - h\sum_{i=1}^{M} s_i - \sum_{j=1}^{N} s_j \qquad (5)$$

Modeling epistasis with multiple sites

All epistatic effects are assumed to be between pairs of sites that are either homozygous or heterozygous for mutant alleles, *i.e.*, there are no higher-order epistatic effects. Each pairwise interaction involves the sum of a joint measure of the fitness effects of the alleles at the pair of sites in question. The simplest model, which applies to both the sites and gene models, is to assume that the epistatic contribution from a pair of homozygous sites i and j is $\epsilon(s_i + s_j)$; this is summed over all pairs of homozygous sites, and the net fitness reduction of the genotype where all sites are homozygotes for mutations is given by this term and the sum of the s_i over all homozygous mutant sites.

For the gene model, a heterozygous individual whose mutations are all in *cis* would have an epistatic component given by the sum over all pairs of i and j of $\epsilon h(s_i + s_j)$. The net fitness reduction is the sum of this quantity plus the sum of the hs_i over all the heterozygous sites. For an individual who is homozygous for a mutation, or heterozygous in *trans*, at more than one site, the sum of the epistatic contributions over all pairs of sites is obtained by weighting ϵs_i for heterozygous sites by ½ instead of h, and ϵs_i for homozygous sites by 1. The net fitness reduction is giving by the sum of this quantity and the sum of the main effects across sites, weighting s_i at homozygous sites by 1 and heterozygous sites by ½. With a fixed selection coefficient and two sites, this procedure reduces to the model described above.

For the sites model, a heterozygous site i can be considered as contributing a main effect of hs_i to the fitness reduction of a genotype, and ϵhs_i to each of the interactions with other sites. Its interaction with a homozygous site j thus contributes $\epsilon(hs_i + s_j)$ and its interaction with another heterozygous site k contributes $\epsilon h(s_i + s_k)$. Similarly, a pair of homozygous sites i and j contributes a fitness reduction of $\epsilon(s_i + s_j)$. The net epistatic contribution is obtained by summing over all pairs of sites, and the net fitness reduction is given by this sum plus the sum of the fitness reductions from each heterozygous and homozygous mutant site.

Estimating the load and linkage disequilibrium statistics from the simulations

The state of the population with respect to the load statistics was summarized using four quantities –the genetic load (L), the inbreeding load (B), the variance in fitness between individuals (V), and the mean frequency of mutant alleles per site ($\bar{q}$). The genetic load was calculated as $L = 1 - \bar{w}$, where $\bar{w}$ was the mean fitness of the population calculated for all individuals in the population, relative to a value of 1 for individuals homozygous for the + allele at all sites. Similarly, V was calculated from all individuals sampled after the burn-in for each simulation replicate. To calculate B, which is defined here as the mean difference in fitness between completely outbred and completely inbred individuals, 100 genomes were sampled randomly from the population, and 100 diploid inbred individuals were artificially created (*i.e.*, the haploid genomes were duplicated to form a diploid individual). The inbreeding load was calculated as $B = \bar{w} - \bar{w}_I$, where $\bar{w}_I$ is the mean fitness of the subsampled inbred population. As each simulation replicate yields a single data point, the distributions shown in the figures and table are across the 1000 replicates.

Summaries of linkage disequilibrium (LD) between sites from simulated data were calculated using the mean of the coefficient of linkage disequilibrium D across all pairs of segregating sites:

$$D = q_a q_b - p_{ab} \qquad (6)$$

and the normalised measures of LD:

$$\bar{\sigma}_d = \frac{\bar{D}}{\sqrt{\overline{p_a p_b q_a q_b}}} \qquad (7a)$$

$$\bar{\sigma}_d^1 = \frac{\bar{D}}{\overline{p_a p_b q_a q_b}} \qquad (7b)$$

where q_a is the selected or minor allele frequency at the first site, q_b is the selected/minor allele frequency at the second site, $p_a = 1 - q_a$, $p_b = 1 - q_b$, and p_{ab} is the frequency of the haplotype "*a b*" which corresponds to either the selected or the minor allele at both sites. The overbars represent the mean of the respective quantities over all SNP pairs and simulation replicates.

$\bar{\sigma}_d$ corresponds approximately to the mean correlation coefficient between the two loci (Ohta and Kimura 1969); the denominator of the statistic $\bar{\sigma}_d^1$, used by Good (2022), and

Ragsdale (2022) greatly magnifies the apparent magnitude of LD compared with $\bar{\sigma}_d$. However, the denominator of $\bar{\sigma}_d^1$ (the mean crossproduct of the four allele frequencies at a pair of segregating sites, rather than its square root) means that it is dimensionally somewhat unsatisfactory. Note that, because the mean values ($\bar{\sigma}_d$ and $\bar{\sigma}_d^1$) were calculated for segregating sites across the replicates (and not across all sites), $\bar{\sigma}_d$ and $\bar{\sigma}_d^1$ did not always have the same sign as $\bar{D}$ when $\bar{D}$ was close to 0 However, this did not greatly affect our results, because it only happened when $\bar{D}$ was not significantly different from 0.

Note that in our case, a derived allele may not always be equivalent to a *deleterious* allele, since back mutations from – to + are possible. We have used the *selected* alleles as the basis for calculations of the direction of *D*; with sufficiently strong selection ($Ns > 4$), there will be little difference between these two classifications, since deleterious alleles are almost always derived. This is shown clearly in Table S3 of Supplementary File S3, where the *D* values calculated using derived and minor alleles for $Ns = 100$ are exactly the same.

Two-locus simulations

We also investigated the effect of weighting LD statistics toward low frequency derived variants, a procedure suggested by Garcia and Lohmueller (2021) and Good (2022). We used the approach of Good (2022) for simulating a two-site haploid model without recombination. This procedure assumes that a mutation from A_1 to A_2 at the first site (A) is present at a frequency q, whose probability distribution is given by the standard formula for drift and selection (Equation S2.1 of Supplementary File S1). A second mutation, from B_1 to B_2, was introduced at the second site, either in *cis* with A_1 (case 1, probability $1 - q$) or in *cis* with A_2 (case 2, probability q). The fitness model for determining the post-selection frequencies of the haplotypes in a given generation assumed that the relative fitnesses of A_1B_1, A_1B_2, A_2B_1 and A_2B_2 are 1, $1 - s$, $1 - s$ and $1 - 2s(1 + \epsilon)$, respectively, where ϵ is a measure of epistasis of the same form as that used in the two-site diploid model. As described in section 2 of Supplementary File S1, multinomial sampling of *N* haplotypes was then carried out each generation, until one of the two mutations was fixed or lost, and the mean values over all generations and replicates of *D*, D^2 and the allele frequency cross-product $P = p_A q_A p_B q_B$ (where subscripts *A* and *B* denote sites A and B) were obtained. The means of *D*, D^2 and *P*, denoted by $\bar{D}$, $\overline{D^2}$ and $\bar{P}$, were found by dividing the respective sums of these statistics by the sum of the durations of the replicate simulations.

Data availability

The scripts used to determine the properties of two-site equilibrium populations, perform all the simulations, and calculate statistics are provided at https://github.com/paruljohri/Gene_vs_sites_model/tree/main.

Results

Deterministic results for two selected sites

Here we examine the properties of equilibria under mutation/selection balance at two selected sites within a coding sequence, with a recombination frequency r between them. The fitness matrices in Table 1 are used for this purpose. The mutation rate from + to – at both sites is u. The symmetry of the fitness models means that the equilibrium allele frequencies are the same at both sites, so that the system can be completely described by the frequency q of the – allele at each site and the coefficient of linkage disequilibrium D between them. The correlation coefficient between the two selected sites is then given by $R = D/q(1-q)$. A full algebraic treatment is given in the Appendix. We focus on the case when $r = 0$, for which the equilibrium properties depend only on the ratio u/s, denoted by $\tilde{u}$.

The complexity of the equilibrium equations, even when $r = 0$, means that exact results can only be obtained by numerical solutions, as described in the Appendix. Several approximations can, however, be derived, which shed light on the behavior of the system. The approximations are reasonably accurate when $h \geq 0.2$ (see Supplementary File S2), but fail badly for small values of h. Results for $h = 0$ can, however, be derived, as shown in section 3 of Supplementary File S1.

The approximations for non-zero h show that positive values of D and R are expected for both the gene and sites models when q is sufficiently small and there is no synergistic epistasis – see Equations (A.6c) and (A.12c). In this case, the most extreme approximation shows that $R \approx \tilde{u}^2/(2h^3)$ for the gene model and $R \approx \tilde{u}^2/h^2$ for the sites model when $h > 0$, so that considerably larger LD is expected for the gene model than the sites model with partially recessive mutations. In both cases, however, R is second-order in $\tilde{u}$, consistent with the fact that the change in D under selection is second-order in s. In addition, it can be shown that a modest degree of synergistic epistasis ($\epsilon > 0$) is sufficient to cause negative values of D and R, with a lighter condition on ϵ for the sites model than for the gene model – compare Equations (A.15d) and (A.20).

The population genetics statistics q, L, B, and V (the equilibrium frequency of mutant alleles, the genetic load, the inbreeding load, and the genetic variance) can also easily be determined once the equilibrium haplotype frequencies are known (see Equations S1.1-S1.6 in Supplementary File S1). For the purpose of illustrating the results for the two-site model, L and B are represented by the ratios of their values to those for an additive two-site model with no LD, $h > 0$ and $u << hs$, for which $L \approx 4u$ and $B \approx 4u(0.5h - 1)$. To avoid large values when h is small, the variance is represented by the ratio of hV to $4uhs$, the equilibrium value of V for the purely additive case with no LD and $u << hs$. This means that results for the variances with different h values are not directly comparable, but the sites and gene models with the same h can be compared.

Table 2 and Figure 2 show the results for the case of $\tilde{u} = 0.01$. This is clearly not a realistic value, but produces relatively large values of D and R, enabling the gene and sites models to be compared, as well as illustrating the effects of synergistic epistasis ($\epsilon > 0$). Since antagonistic epistasis ($\epsilon < 0$) is well known to induce positive D values (Shnol and Kondrashov 1993), we have focussed here on synergistic epistasis. As shown below, the general properties of this case mirror those of the multi-site simulations. Several points emerge.

First, over the range of values of ϵ considered here, its magnitude has relatively little effect on L and B, for both the sites and gene models. It effect on q is more complex. For both models, q decreases with ϵ, but faster for the sites model when $h = ½$, with the result that in this case q is smaller for the sites than the gene model when ϵ is large despite their having the same q with no epistasis. With $h < ½$, q is always smaller for the gene model than the site model, especially for small values of h. The gene model, as might be expected, is associated with lower values of L and B than the sites model when $h <½$, but has higher values of V, with the differences becoming more marked with smaller h decreases. With $h = ½$, $B = 0$, as expected, L is the same for the two models, and V increases slightly with ϵ for the sites model compared with the gene model, the reverse of the pattern for q, resulting in a slightly higher V for the sites than the gene model with large ϵ, probably reflecting the larger contributions of terms in ϵhs to the sites model, as can be seen in the fitness matrices above.

Seemingly paradoxically, Table 2 shows that the load is higher for the gene model than the sites model when $h = 0.01$ and ϵ is small, despite its producing a lower q. This comes about because of the considerably lower fitness of $+-/-+$ heterozygotes with the gene model compared with the sites model when h is small (see Table 1). With our

symmetric fitness model, the frequency of this genotype is proportional to the square of the frequency of + – or – +, and so the effect of its presence on mean fitness can outweigh the effect of a lower value of q under the gene model. A more formal argument for the extreme case of $h = 0$ is given in section 1 of Supplementary File S1 (Equations S1.7 and S1.8).

As can be seen in Figure 2, the effect of synergistic epistasis on D and R is strikingly different from its effect on the other statistics. For $\epsilon = 0$ and for sufficiently small $\epsilon > 0$, and D and R are positive, but decline almost linearly with ϵ, becoming negative when ϵ is sufficiently large. The gene model is associated with larger values of D and R compared with the sites model when $h < ½$, and the difference is more marked, the smaller h, with small h being associated with larger magnitudes of D and R. For some combinations of h and ϵ, D and R can be positive for the gene model but negative for the sites model. These properties of LD partly explain the behavior of the variance, since a positive covariance in fitness effects associated with $D > 0$ inflates the genetic variance over that with $D = 0$, and vice-versa for the negative covariance associated with $D < 0$ (Bulmer 1974).

Sets of results for other values of $\tilde{u}$ are shown in Supplementary File S2. The general features described above can be seen in these cases. However, as might be expected, the smaller the value of $\tilde{u}$, and hence q, the smaller the differences between the gene and sites models, and the smaller the magnitudes of D and R. While D and R are still positive in the absence of epistasis, they become negative at much smaller ϵ values than when $\tilde{u}$ is large. Thus, in order to determine whether there are biologically important consequences of the features just described, it is necessary to use computer simulations to examine what happens with multiple selected sites, where larger effects are to be expected.

Simulation results for multiple sites

The analytical results presented above were supplemented with forward simulations of a single gene with multiple selected sites, with an equal probability of mutations between the + and – alleles and vice versa (see the section *Methods*, *Multi-locus simulations*). We first describe the results with no epistasis.

Load statistics with no epistasis

The analytical results described above suggest that, relative to the sites model, the gene model is expected to result in smaller mean frequencies of the deleterious alleles, causing a higher mean fitness of the population and a lower genetic load. The multiple site simulations

show that the equilibrium frequencies of deleterious alleles are much smaller for the gene than the sites model, except when selection is weak relative to drift, i.e., when $\bar{\gamma} = 2$ for the case of a gamma distribution of selection coefficients (Table 3) and $\gamma = 2$ for the case of a fixed selection coefficient (Table S1 in Supplementary File S3). There is also a drastic difference between the estimated inbreeding load (B) between the two models; B is much smaller for the gene than the sites model with $h < ½$. There is a much smaller effect on the genetic load (L), which is noticeable only when mutations are fully recessive and selection is of intermediate strength.

As discussed in the section *Deterministic results for two selected sites*, deleterious alleles are more likely to be associated on the same haplotypes in the gene model than in the sites model, resulting in larger positive mean D values for the gene model, and inflating the variance (V). This effect can be counteracted by lower allele frequencies under the gene model, an effect that is especially strong for fully recessive mutations, leading to a much lower V under the gene model when $h = 0$. However, V is significantly higher for the gene model when $h = 0.2$. With $h = ½$, there is no difference between the two models, as expected. The difference in V between the two models thus has a non-linear relation with h.

Because the results discussed above are most relevant to *D. melanogaster*-like population parameters, we also conducted simulations with human-like parameters, with much smaller values of Nu and Nr. The overall patterns are very similar to those for the *Drosophila*-like populations (Table S2 in Supplementary File S3). The mean frequencies of deleterious alleles are significantly smaller with the gene than the sites model; genetic load was slightly lower, while inbreeding load is again drastically smaller for the gene than the sites model.

Linkage disequilibrium with no epistasis

In simulations without epistasis in which s was fixed, and LD statistics were calculated by assigning positive values to cases when + + and – – haplotypes were in excess of random expectation, the gene model results in higher mean LD than the sites model for closely linked sites whenever there are significantly non-zero values (Table S3 in Supplementary File S3). The difference between the two models increases with increasing recessivity and strength of selection, with mean positive values of $\bar{D}$ and $\bar{\sigma}_d$ for the gene model versus negative values with the sites model when $h < 0.5$ and $\gamma \geq 20$ (Table S3 in Supplementary File S3).

Similar results were found with a DFE, with the gene model yielding positive values of $\bar{D}$ and $\bar{\sigma}_d$ with $h < 0.5$ and $\bar{\gamma} \geq 100$, as opposed to slightly negative (but not significantly non-zero) values for the sites model (Table 4). Note that, while the two-sites theoretical results predict that $\bar{D}$ values are expected to be slightly positive for both models, we consistently observed small negative values for the sites model with $h < ½$. This is probably due to Hill-Robertson interference (HRI) effects generated by multiple linked selected sites (Hill and Robertson 1966; Garcia and Lohmueller 2021; Roze 2021), which are not allowed for in the two-site deterministic model. To test this, we analyzed the constant selection coefficient results for variants separated by a greater-than-average distance (800-100 base pairs; Table S4 in Supplementary File S3). The $\bar{D}$ values for selected variants under the sites model then approach zero, suggesting reduced effects of HRI with higher frequencies of recombination. However, with the gamma distribution with $\bar{\gamma} = 1000$ and $h < ½$ under the gene model, $\bar{D}$ and $\bar{\sigma}_d$ are still substantially greater than zero, suggesting significantly positive LD even at larger distances (Table S4 in Supplementary File S3).

Interestingly, when an excess frequency of haplotypes with pairs of minor alleles was used as the criterion for positive *D*, there is not much difference in $\bar{D}$ and $\bar{\sigma}_d$ between the gene and sites models when selection was weak, with mainly positive values for both (Table 4 and Tables S2 and S3). However, with stronger selection against – alleles and $h < ½$, the gene model yield slightly lower values of $\bar{D}$ and $\bar{\sigma}_d$. The sign of $\bar{D}$ is expected to be positive for minor alleles in cases where the overall value of $\bar{D}$ for derived alleles is zero; this is because, in cases where two derived alleles are in negative LD, and one is at a frequency greater than 0.5 while the other is at a frequency lower than 0.5, LD between their corresponding minor alleles is positive (Potapova and Kondrashov 2023). The difference between the gene and sites models probably reflects the lower allele frequencies with the gene model. It should also be noted that we estimated $\bar{D}$ with respect to *selected* alleles, which were not necessarily derived (due to back mutations). However, as was pointed out in the section *Estimating the load and linkage disequilibrium statistics from the simulations,* this should only affect our results when selection is weak (i.e., when $\gamma \leq 2$); selected alleles are equivalent to derived alleles in most other cases.

In contrast to the simulations with *Drosophila*-like parameters, which we have just discussed above, simulations of human-like populations yielded high variances between replicates and somewhat unclear patterns for the LD statistics (Table 5). Overall, however, LD between selected sites is higher with the gene model than the sites model when $h < ½$. In

this case, we show $\bar{\sigma}_d^1$ as well as $\bar{\sigma}_d$, for comparison with the analyses of human data by Ragsdale (2022) (see *Discussion, Patterns of linkage disequilibrium in population genetic data*). The difference in LD between the two models is much clearer when using $\bar{\sigma}_d$ or $\bar{\sigma}_d^1$ as opposed to $\bar{D}$, because $\bar{\sigma}_d$ and $\bar{\sigma}_d^1$ (partially) correct for differences in allele frequencies.

Load statistics and LD with epistasis

With synergistic epistasis between deleterious mutations, a decrease in their mean equilibrium frequencies is expected compared with the case with no epistasis, due to the reduced fitness of genotypes with multiple mutations caused by synergistic epistasis. In the simulations with multiple linked sites under a DFE with a relatively large mean scaled selection coefficient ($\bar{\gamma} = 2N\bar{s} = 100$) and varying degrees of epistasis (with ϵ ranging from 0.02 to 0.32), a severe decrease in the mean equilibrium frequency of deleterious mutations occurs with increasing ϵ, for both the gene and sites models (Figure S1 of Supplementary File S3). Even with the lowest strength of epistasis ($\epsilon = 0.02$), there is no significant difference in mean equilibrium frequencies of deleterious alleles between the gene and sites models. There is a corresponding decrease in the inbreeding load (B) with increasing ϵ for both the gene and sites model (Figure 4) when $h < ½$.

B remains significantly and drastically higher with the sites model than the gene model for both $h = 0$ and $h = 0.2$. There is no difference between the two models with $h = 0.5$, as expected; unexpectedly, however, B becomes increasingly slightly negative as ϵ increases. One possible explanation is that this is an artefact of the way B in which was estimated; the fitness of inbred individuals was calculated by sampling 50 diploid individuals, where the mean fitness was calculated using the entire population. However, even if we calculate the fitness of outbred individuals by sampling 100 diploid individuals, B remains negative for these parameter sets (see Figure S4 in Supplementary File S3), so that this explanation can be ruled out. The probable cause of this effect is as follows. With the parameter set in question, the frequencies of deleterious alleles are sufficiently high that essentially all haplotypes carry at least one deleterious mutation. In the two-site case with $h = 0.5$, a – +/+ – individual has fitness reduction $s(1 + \epsilon)$, whereas – + or + – homozygotes have a fitness reduction s (there is no effect epistasis). Thus, the heterozygote actually has a lower fitness than the homozygote. In other words, the outbred population has more opportunities for pairwise fitness interactions that reduce fitness than the inbred population, so we get slightly negative inbreeding depression. Epistatic interactions of this kind have

been studied theoretically and empirically in connection with the biometrical genetics of crosses between inbred lines (e.g., Wright 1968, Chap. 15, pp. 391-403; Jinks 1983). A lower value of a fitness component in the F_1 compared with the mean of the two parent of the kind expected under this model appears to be uncommon in such crosses, with a prevalence of the opposite pattern (heterosis), consistent with the effects of largely recessive effects of deleterious alleles or heterozygote advantage outweighing epistatic effects of this kind.

There is a much smaller effect of epistasis on the genetic load (L in Figure 4), with a minor and steady decrease in L with an increase in ϵ for both fitness models. As before, the difference in the genetic load between the two models is relatively small and only significant when mutations are completely recessive. The variance in fitness (V) also increases with an increase in the epistasis coefficient when h = 0.2 or 0.5 (and only very slightly when mutations are fully recessive). However, as was also found without epistasis, the variance is much larger for the gene model compared to the sites model, especially when mutations have increased recessivity and intermediate/low levels of epistasis (Figure S1 of Supplementary File S3). With extremely high levels of epistasis ($\epsilon = 0.32$), there is no difference in V between the two fitness models. In summary, if there were widespread epistasis between deleterious mutations, there would be no substantial difference in mean equilibrium allele frequencies between the two models, but the inbreeding and genetic loads would be lower with the gene model than the sites model if mutations have moderate to high recessivity.

Similarly to the analytical results discussed earlier, there is a substantial decrease in $\overline{D}$ and $\bar{\sigma}_d$ (using the selected/derived allele criterion) with increasing strength of synergistic epistasis for both models when h is close to 0.0, but only small changes when h is 0.2 or 0.5 (Figures 5 and S2). This is somewhat surprising, as V tends to increase with ϵ even when h is 0.5 (Figure S1). However, $\overline{D}$ and $\bar{\sigma}_d$ between selected alleles remain significantly higher with the gene model than the sites model (for which both statistics become negative) with large ϵ . There seems to be no significant difference between the gene and sites model with these parameters when h = 0.2 and ϵ is very small, consistent with their large SEs relative to the means when $\bar{\gamma}$ = 100 (see also Table 4 for results with no epistasis and Figure S3 for a comparison between results in Table 4 and Figure 5); $\overline{D}$ and $\bar{\sigma}_d$ are both not significantly different from zero in this case. Importantly, however, with h = 0.2 and ϵ, $\bar{\sigma}_d$is negative with the sites model, but is close to zero with the gene model, so that the two models display differences in LD even with epistasis.

When the same LD statistics are calculated with minor/major alleles as the sign criterion, there is a substantial decrease in LD with an increase in the epistasis coefficient; however, the values always remain positive and there is no significant difference in $\bar{\sigma}_d$ between the two fitness models (Figure 5), although $\bar{D}$ obtained using minor alleles is consistently and slightly smaller for the gene than the sites model, possibly due to lower mean allele frequencies at equilibrium (Figure S2). In summary, if synergistic epistasis were prevalent amongst deleterious alleles, LD between selected/derived alleles will probably still be higher and positive with the gene model than the sites model when $h < ½$, whereas no difference is expected between LD statistics calculated using minor alleles.

With much lower rates of crossing over (0.1 × the standard rate used here), the behavior of the load statistics remains the same as described above (Figures S4 and S5 in Supplementary File S3). However, when mutations are fully recessive, LD calculated using minor alleles is significantly higher with the sites model than the gene model (Figures S6 and S7). The difference between the two models is largest when there is no epistasis, but persists even when the epistasis coefficient is 0.32. The predicted differences in LD between the gene and the sites models using the derived and minor allele sign criteria can thus be entirely opposite, depending on the degree of dominance. HRI in the absence of epistasis seems to produce only a weak tendency towards negative LD when $h = ½$ and for the sites but not the gene model when $h = 0$ or 0.2.

Effects of weighting towards low frequencies on LD statistics

Garcia and Lohmueller (2021) suggested that the effects of HRI on LD between deleterious mutations would be more easily detected in samples from populations if doubleton variants were used for estimating LD. This procedure would help to correct for the effect of allele frequency on LD statistics; in addition, because rare variants are more likely to be deleterious than common ones, larger negative D values might be expected for nonsynonymous than synonymous variants when conditioning on doubletons. In their multi-locus simulations, they indeed found negative LD between derived variants at sites subject to purifying selection, both with $h = 0$ and $h = ½$, with LD being measured by Lewontin's D' statistic (Lewontin 1964). In contrast, Good (2022) proposed that multiplying allele frequencies by a function that gives greater weights to low frequency variants should lead to positive expected D values, for both neutral and deleterious variants. He validated this expectation by a combination of simulations and analytical formulae, using a model of a haploid population;

this is equivalent to a diploid population with h = ½ in the absence of epistasis. However, while simulations performed by Garcia and Lohmueller (2021) involved multiple selected sites, those of Good (2022) involved only two sites.

In order to reconcile these seemingly contradictory findings, we used the approach described in the section *Methods: Two-locus simulations*. This enables the means of D, D^2 and the crossproduct P of the four allele frequencies at sites segregating for two mutations, $\overline{D}$, $\overline{D^2}$ and $\overline{P}$ to be found, yielding estimates of the standardized LD statistics $\bar{\sigma}_d = \overline{D}/\sqrt{\overline{P}}$, $\bar{\sigma}_d^1 = \overline{D}/\overline{P}$ and $\sigma_d^2 = \overline{D^2}/\overline{P}$. σ_d^2 corresponds approximately to the mean squared correlation coefficient between the two loci (Ohta and Kimura 1969); the other two statistic are defined by Equations (7). These statistics were estimated either with or without weighting; in the latter case, the population frequencies of haplotypes carrying the derived, deleterious variants at each site (frequencies q_A and q_B, respectively) were weighted by exp($-fq_A$) and the frequencies of haplotypes carrying B_2 by exp($-fq_B$). The factor f is 0 if no weighting is applied, and $f > 1$ if low frequency mutant alleles receive preferential weights (f is the inverse of the factor used by Good [2002)]). In addition, the effect of calculating LD statistics from variants present only at a pre-set frequency in a sample of k haplotypes (Garcia and Lohmueller 2021) was determined by trinomial sampling of k haplotypes each generation, recording statistics only for samples where the numbers of alleles at each site in the sample matched the pre-set allele frequency (usually 2 out of k).

For neutral variants, the unweighted case gave values of $\overline{D}$ and $\bar{\sigma}_d$ that were non-significantly negative and close to zero: $\overline{D} = 0.75 \times 10^{-4} \pm 5.14 \times 10^{-4}$ and $\bar{\sigma}_d = 0.00144 \pm 0.00713$ over 5 repeated simulations of 10^6 introductions of B_2. The value of σ_d^2 was close to the theoretical value of 0.45 for the case of no recombination under the infinites sites model (Ohta and Kimura 1971). In contrast, with $f = 10$ and a single set of 10^6 introductions of B_2, $\bar{\sigma}_d$ and $\bar{\sigma}_d^1$ were 0.0107 and 0.915, respectively. The values when doubletons were sampled were $\bar{\sigma}_d = 0.133$ and $\sigma_d^1 = 13.5$.

Some representative results for $\bar{\sigma}_d$ with selection (one set of 10^6 introductions of B_2; N = 1000) are shown in Figure 3, for two different strengths of selection ($\gamma = 2Ns = 10$ and $\gamma = 2Ns = 50$) and varying amounts of synergistic epistasis. With no epistasis and no weighting, $\overline{D}$ and $\bar{\sigma}_d$ were slightly negative, but became increasingly negative as ϵ increased. 20 repetitions of the case of $\gamma = 50$ without epistasis gave no evidence for significantly non-zero LD ($\overline{D} = -5.35 \times 10^{-7} \pm 4.16 \times 10^{-7}$; $\bar{\sigma}_d = -8.88 \times 10^{-5} \pm 6.93 \times 10^{-5}$), implying that HRI effects with two sites are not sufficiently strong to overcome any tendency for positive LD to

be produced by purifying selection. Weighting causes the LD statistics to take larger values, which are positive with $f = 100$ even for the largest value of ϵ (0.64). They remain positive over the whole range of ϵ if doubletons are sampled.

These results show that focussing on mutations at low frequencies results in positive LD between derived variants, which can persist even if there is strong synergistic epistasis, especially if doubletons are sampled. A heuristic argument that explains this effect is provided in section 4 of Supplementary File S1. These considerations suggest that the use of LD statistics estimated from low frequency variants is not helpful in diagnosing HRI effects caused by purifying selection. On the contrary, it is likely to conceal the effects of HRI or synergistic epistasis on LD between derived variants. This raises the question of why Garcia and Lohmueller (2021) found consistently negative LD in their multi-locus simulations of deleterious mutations when sampling doubletons.

It turns out that this was because they applied the D' statistic (Lewontin 1964) to estimates of D for derived variants. D' is defined as $D/|D_{max}|$, where $|D_{max}|$ is the maximum value of the absolute value of D conditioned on the allele frequencies. Using our notation, when defining D as positive for associations between derived variants we have:

$$|D_{max}| = \min(p_A\, p_B, q_A\, q_B\ ), \quad D < 0 \qquad (8a)$$

$$|D_{max}| = \min(p_A\, q_B, q_A\, p_B\ ), \quad D > 0 \qquad (8b)$$

First, consider the situation if unweighted haplotype frequencies from the whole population are used for the LD statistics. As described in the section *Methods: two-locus simulations*, the LD statistics were calculated after a second mutation at site B, was introduced at a site segregating for a mutation at site A, either in *cis* with the ancestral A variant (case 1) or in *cis* with the derived variant (case 2). For case 1 in the neutral model described above, $D < 0$ for all generations, and $D' = -q_A q_B / q_A\, q_B = -1$ if $q_A\, q_B < p_A\, p_B$ or $D' = -q_A q_B / p_A\, p_B > -1$ if $q_A\, q_B > p_A\, p_B$. Given that q_A and q_B are the frequencies of derived variants, whose probability distribution is skewed towards low frequencies, the large majority of cases will involve the first situation, and the expectation of D' for case 1 will be –1 or close to – 1. For case 2, $D' = p_A q_B / p_A q_B = 1$ if $p_A\, q_B < q_A\, p_B$ and $D' = p_A q_B / q_A p_B > 1$ if $p_A\, q_B > q_A\, p_B$. Again, the majority of cases will satisfy the first condition, since q_B is likely to be much smaller than q_A, due to B_2 arriving in the population as the second

mutation. The expectation of D' for case 2 will thus be 1 or close to 1. Case 1 has a higher probability of occurring than case 2 (their probabilities are proportional to the integrals between $1/N_H$ and $1 - 1/N_H$ of pq^{-1} and 1, respective), so that the overall expectation of D' is negative. If purifying selection is acting, the probability of case 1 is increased over case 2, due to the increased skew of the probability distribution of allele frequencies towards low values, so the magnitude of D' will increase with the strength of selection.

This shows that unweighted estimates of mean D' that take into account the sign of the LD statistic defined by Equations (4) are biased towards negative values. It seems likely that this will also apply to estimates based on doubleton frequencies, although the expectation of D itself in this case is biased towards positivity unless there is strong synergistic epistasis (see Figure 3). This expectation was tested by using the simulations described above to estimate mean D' from doubletons. For the cases of neutrality, $\gamma = 10$ and $\gamma = 50$ (with no epistasis), the mean D' values for cases 1 and 2 were – 1 and 1, respectively; the overall mean D' values were – 0.705, – 0.857, and – 0.922, respectively. Consistent with the findings of Garcia and Lohmueller (2021), neutral variants have a smaller magnitude of negative LD as measured by D'. However, this effect observed here is entirely due to the fact that the probability of case 2 increases as selection becomes stronger. In other words, purifying selection itself can bias D' values at selected alleles to be lower than those of neutral alleles, even if their frequencies are similar, so that this effect should be taken into account when testing for HRI effects. However, the magnitude of this effect is currently unclear, especially for a multi-site model. Alternatively, one could use unweighted values of D or $\bar{\sigma}_d$. However, it is unclear how nonequilibrium scenarios like population size changes might affect the sign of these statistics and whether unweighted values of D would be the best to use in all cases.

Discussion

The role of complementation

A major finding of this work is the effect on load and LD statistics of the distinction between the gene and sites models, which we discuss in detail below. This raises the question of how widely the gene model might apply to real organisms. The distinction between the two models is based on the well-established finding of classical and molecular genetics that two recessive loss-of-function mutations affecting the amino-acid sequence of the same polypeptide usually fail to complement each other when present in the same cell in *trans* in a double heterozygote, due to the fact that both copies of the polypeptide are non-functional. In

contrast, the same mutations present in *cis* in a double heterozygote result in the cell containing one functional and one non-functional polypeptide, with near wild-type performance (see Figure 1). This finding is of great historical importance, since underlies the recognition of the gene as a coding sequence whose sites could be separated by recombination, and to Seymour Benzer's concept of the cistron as the functional genetic unit (reviewed by Benzer 1957; Pontercorvo 1958; Hawley and Gilliland 2006).

If mutations within the same coding sequence fail to complement each other when in *trans*, the assumptions of the gene model are satisfied. The sites model must usually apply to mutations affecting different coding sequences, which necessarily complement each other unless there are unusual types of epistasis between them, e.g., due to interactions between mutations in different members of a multi-protein complex. Examples of such a failure of complementation between non-allelic loss of function mutations are reviewed in Hawley and Gilliland (2006) and Navarro-Quiles et al. (2024); these appear to be very rare compared with complementation.

It should be noted, however, that there are exceptions to a lack of allelic complementation (reviewed by Fincham et al. 1979, Chap. 4; Hawley and Gilliland 2006). In such cases, loss-of-function mutations affecting the same polypeptide complement each other, so that *trans* double heterozygotes show wild-type or near wild-type performance. In many cases, some mutations within the same gene complement each other, whereas others do not, allowing the construction of complementation maps (Fincham et al. 1979, Chap. 4). Some mutations within the same gene may thus obey the gene model and others the sites model. Allelic complementation has been best studied in fungi like *Neurospora*, and often involves cases where multimers of the same polypeptide form part of the same protein molecule (Fincham et al. 1979, pp. 348-353). In other cases, spatial interactions between mutations in distinct functional subunits of the same polypeptide are involved (Fincham et al. 1979, p. 352; Eng et al. 2015). It is often found that interallelic complementation is incomplete, when phenotypes such as enzyme activity are measured (Fincham et al. 1979, pp. 348-351), so that the distinction between the two models is somewhat blurred. For simplicity, we have not attempted to include any of these complexities in our models.

It is, of course, less clear what happens to the heterozygous fitness effects of more minor effect mutations in the same coding sequence when configured in *cis* or *trans*, and there appears to be little relevant data. The simplest assumption is that the effects parallel those of loss-of-function mutations and are recessive or partially recessive, but are additive in their effects when in *trans*, as we have assumed here. There are, of course, many other

possibilities. Thornton et al. (2013) and Sanjak et al. (2015) assumed that the fitness of an individual is determined by a quantitative trait subject to stabilizing selection, where the trait value is a multiplicative function of the total number of mutations present in a given individual. With small effects, this is close to the additive model without epistasis used here, with h = ½ as far as trait value (as distinct from fitness) is concerned; there is then no distinction between the gene and sites model, so that complementation does not play a role. Ragsdale (2022, Table S1) describes two versions of a two-site model. His "simple dominance" model is identical with our sites model (allowing for the difference in notation). His "gene-based" model is similar to our gene model, in that it assigns different fitnesses to *cis* and *trans* double heterozygotes. However, it differs in several other respects, e.g., by assigning the same fitness to single homozygotes, single heterozygotes, double homozygotes, and *cis* double heterozygotes.

The situation with respect to mutations affecting the same non-coding regulatory sequence such as a UTR, an enhancer or an insulator is somewhat unclear, although it seems likely that mutations within the same cis-regulatory sequence should show similar *cis*/*trans* behavior to mutations in the same polypeptide, fitting the gene model, while mutations in different cis-regulators of the same gene should complement each other, and hence fit the sites model. The phenomenon of transvection involves complementation between mutations in different cis-regulatory sequences of the same gene when the two homologous chromosomes or regions of homologs are paired in a cell nucleus, but failure of complementation when they are unpaired (Lewis 1954; Duncan 2002; Mellert and Truman 2012; Galouzis and Prud'homme 2021); the normal situation is thus for them to fit the sites model.

Load statistics with no epistasis

Differences between the gene and sites models.

The main conclusion from the analytical and simulation results for the case of no epistasis is that partially and fully recessive deleterious mutations are likely to segregate at higher allele frequencies in a randomly mating population when individual fitnesses follow the sites model compared with the gene model. The results in Table 3 show that the gene model has a slightly lower variance in fitness and genetic load with recessive or partially recessive mutations, when there is a relatively large mean scaled selection coefficient ($\bar{\gamma} \geq 100$) combined with a wide distribution of selection coefficients of the type suggested by population genetic data

(e.g., Charlesworth 2015; Robinson et al. 2023). For weak selection ($\bar{\gamma} \leq 20$), the difference between the two models is either negligible or even slightly in the opposite direction. In contrast, the inbreeding load *B* is always much smaller under the gene model than the sites model, being effectively zero for $\bar{\gamma} \leq 20$, unless the dominance coefficient *h* is close to 0.5, in which case *B* is close to zero for both models, as is expected from basic theory (Morton, Crow and Muller 1956). A similar pattern is seen with synergistic epistasis, but *B* becomes very small for both gene and sites models when epistasis is strong (Figure 4).

The mean frequency of deleterious alleles ($\bar{q}$) is smaller for the gene model than the sites model when $\bar{\gamma} \geq 100$, but its relation to $\bar{\gamma}$ does not track the behavior of *B*; for example, with $\bar{\gamma} = 100$ and $h = 0.2$, *B* and $\bar{q}$ for the gene model are approximately 6% and 30% of the values for the sites model, respectively; when $\bar{\gamma} = 1000$ and $h = 0.2$, *B* and $\bar{q}$ for the gene model are 30% and 12% of the sites model values. These results raise the question of why *B* should be so small under the gene model when selection is relatively weak and *h* is small. The probable cause of this phenomenon is the fact that haplotypes carrying – alleles are relatively common in these cases; for example, with $\bar{\gamma} = 100$ and $h = 0.2$ the mean number of – alleles per haploid genome is 4.67. When drift causes haplotypes carrying – alleles to reach high frequencies in the population, heterozygous genotypes with – alleles in *trans* will contribute substantially to the mean fitness of the population. Under the gene model, these genotypes have fitnesses that are much closer to homozygotes for the same haplotypes compared with the sites model, so that there will be a much closer correlation between the mean fitnesses of the outbred population and the inbred population than for the sites model, thereby reducing *B*. This is consistent with the fact that the relative difference in *B* between the two models diminishes with the strength of selection, and is much smaller for the two-locus deterministic case (Table 2) than for the multi-locus simulations.

Implications for data on inbreeding load

Predictions of load statistics have universally been based on the sites model, including recent modeling studies that have used population genetic data to inform parameter values (Kyriazis, Robinson and Lohmueller 2023). Multiplicative fitnesses across selected sites have usually been assumed in these studies – but see Charlesworth (1998) and García-Dominguez et al. (2018) for exceptions. Our results are for a single gene, so it is important to consider what happens if they are extrapolated to the whole genome. For convenience, we will assume purely autosomal inheritance, leading to a probable overestimation of the load statistics for

the entire genome, given that the X chromosome is associated with lower loads than autosomes due to selection against hemizygous mutations in males (Wilton and Sved 1979).

To scale up to multiple genes, we assume that the sites model applies to between-gene effects for both models of single gene fitness effects, due to complementation. As in previous studies, starting with Haldane (1937), we assume multiplicative fitness effects across different genes; with small selective effects the differences from additive effects will be small for most of the load statistics. Our simulation results were based on a scaling of population size from 10^6 to 10^3; to generate realistic results for a *Drosophila*-like system, we would need to rescale back to the $N = 10^6$ values; this means dividing linear quantities in s by 10^3 and quadratic quantities such as the variance by 10^6. This assumes that the probability distribution of haplotype frequencies is controlled by the product of N_e and the magnitudes of the deterministic forces, *i.e.*, all evolutionary forces are weak, and diffusion equations are adequate to describe the evolutionary process (Ewens 2004, Chap. 5). If we assume 14000 genes in the *Drosophila* genome (Misra et al. 2002), the individual gene linear quantities (the loads and inbreeding loads) in the tables should thus be multiplied by 14 and the variance by 0.014. This ignores any contributions from mutations in non-coding sequences, which will be considered below.

In general, let the linear and quadratic multipliers due to rescaling be f and f^2, and the number of genes be n. For the inbreeding load, defined as the difference between the natural logarithms of outbred mean fitness and fully inbred mean fitness, the B values in the tables need only be multiplied by nf to obtain the total inbreeding load, B_T. If the mean load L for each gene is small, the total genetic load with multiplicative effects across n genes is given by:

$$L_T \approx 1 - \exp\,(-nfL) \qquad (9)$$

Similarly, if there is no LD between different genes, using the rule that the expectation of a product of a set of independent variables is the product of their expectations, the total variance with multiplicative effects across gene is given by:

$$V_T \approx \exp(-2nfL)[\exp(-nf^2V_g) - 1] \qquad (10)$$

Given n and f, the values of the relevant statistics in the tables can easily be converted into values for the whole genome by using these relations. For example, using the mean values of the load statistic in Table 3 for $\bar{\gamma}$ = 1000, Table 6 shows the values obtained with n = 14000 and f = 14; σ_T is the standard deviation corresponding to V_T. These results show that the difference between the two models has only a small effect on the expected genome-wide genetic load and genetic variance/standard deviation from variants with such small effects on fitness (mean s is only 5×10^{-4} after rescaling N), but has a drastically smaller total inbreeding load due to deleterious mutations at nonsynonymous sites than the sites model with h = 0.2 (0.050 versus 0.165), despite the fact that the sites model applies to between-locus fitness effects.

These findings suggest that, for the same distribution of selection and dominance coefficients, an equilibrium population is likely to harbor a much smaller amount of inbreeding load than expected under previous theoretical calculations, if fitness effects of recessive or partially recessive mutations at individual genes obey the gene model rather than the usual sites model and the mean scaled strength of selection is of the order of 1000 or less. This conclusion is important, because reconciling observed and predicted values of B has been the subject of much debate, with some workers asserting that the observed values can be explained by mutation-selection balance with multiplicative fitnesses (e.g., Pérez-Pereira et al. 2021; Robinson et al. 2023), while others have proposed that additional factors such as variation due to balancing selection are likely to be involved (e.g., Charlesworth 2015).

Our results suggest that the sites model of mutation and selection is likely to considerably *overpredict* the value of B, if multiplicative fitnesses between genes are assumed, unless the mean strength of selection is so large and the distribution of selection coefficients is so tight that the population can be treated as deterministic. If it can be shown that empirical estimates of B are incompatible with values predicted on this basis, then contributions from balancing selection or synergistic epistasis of the type modeled by Charlesworth (1998) must be involved. One problem is that reliable estimates of B for net fitness in natural populations are hard to obtain, with substantial disagreements among different workers concerning the typical magnitude of B (Robinson et al. 2023). This makes it difficult to compare the data on inbreeding load with theoretical predictions.

Inbreeding load in Drosophila

Here we take a somewhat different approach. An ingenious method for estimating B for net fitness under laboratory conditions for a single major chromosome of some *Drosophila* species was devised by John Sved, which is known as the balancer equilibration (BE) technique (Sved and Ayala 1970). This is based on using population cages to compete wild-type chromosomes extracted from natural populations (chosen to be homozygous viable and fertile) against homozygous lethal balancer chromosomes that suppress crossing over when heterozygous. The equilibrium frequencies of wild-type chromosomes can be used to estimate the ratio R of the mean fitness of chromosomal homozygotes to the mean fitness of chromosomal heterozygotes (reviewed by Frankham 2023). The mean value of R for chromosome 2 of *D. melanogaster* over two well-replicated experiments was 0.18±0.03, giving an estimate of $B = -\ln(R)$ of 1.7 (Frankham 2023). The BE results using the data on all three major chromosomes imply that B for the whole genome of *D. melanogaster* (including the contribution from homozygous lethal mutations) is approximately 5, so that the fitness of completely homozygous flies (exp –5) is effectively zero (Frankham 2023).

The analysis of population genetic data for *D. melanogaster* presented in sections 1-3 of Supplementary File S6 yields an prediction of $R = 0.66$ and $B = 0.42$ for chromosome 2 under the sites model. Here, possible effects of purifying selection at non-coding sites as well nonsynonymous sites and strongly selected (but non-lethal) mutations are taken into account, using fairly liberal assumptions with respect to the magnitude of B. This result is based on the deterministic expressions for R and B given by Equations (S3.1) and (S3.2), a very low h value of 0.125, and an approximate correction for the effects of drift obtained by multiplying u by the proportion of mutations with $2N_ehs < 2.5$ under a gamma distribution with shape parameter 0.3, thus removing the contribution of nearly neutral mutations (see section 2 of Supplementary File S6). Given that the gene model predicts a substantially lower value of B than the sites model, $B = 0.42$ for chromosome 2 is almost certainly an overestimate, but is less than 25% of the observed value. There thus appears to be much more inbreeding load due to deleterious mutations under laboratory conditions in *D. melanogaster* than can be accounted for purifying selection under a multiplicative fitness model.

Why is the inbreeding load reduced when N_es is small?

Numerical methods for obtaining an exact single-locus prediction for B in a finite population have been developed by Charlesworth (2018), and extensive multi-locus simulation results using a variety of parameter values have been described in the literature (Kyriazis et al. 2023). Both of these approaches show that the level of inbreeding load caused by deleterious

mutations is negatively related to $N_e s$ under a wide range of model assumptions. As has been widely discussed in the literature (see review by Robinson et al. 2023), there are two possible reasons for this effect – the "purging" of deleterious, strongly recessive variants from small populations as a result of their exposure to selection against homozygotes (Nei 1968), and the weakening of selection against deleterious mutations relative to mutation pressure, resulting in increased mean frequencies of deleterious mutations, including their fixation (Kimura et al. 1963).

For the results presented in Table 3 and Table S1 in Supplementary File S3, the second factor must be involved in the reduction in B with smaller $N_e s$, since we are comparing different strengths of selection while keeping N_e and the mutation rate constant. As expected from the effects of reduced s, $\bar{q}$ changes from 0.0092 ± 0.0028 with $\bar{\gamma}$ = 1000 to 0.0383 ± 0.0053 with $\bar{\gamma}$ = 2, the opposite of what is expected with purging; the ratio of $\bar{q}$ values (equal to 4.16) is, however, much less than the 50-fold difference in $\bar{\gamma}$. A similar pattern of decrease in $\bar{q}$ with increasing $\bar{\gamma}$ is seen with h = 0.5. The explanation for the lack of proportionality between the two variables is explored in section 4 of Supplementary File S6.

In contrast to B, the genetic load and variance in fitness are little affected by the effects of drift when selection is weak. This is illustrated by the results on the genetic loads under the sites model for different $\bar{\gamma}$ values with h = 0.2 in Table 3; the mean loads for $\bar{\gamma}$ = 1000 and $\bar{\gamma}$ = 2 are 0.0101 ± 0.001 and 0.0095 ± 0.001, respectively, and the approximate deterministic value of $2000u$ (Haldane 1937) is 0.01. Even $\bar{\gamma}$ = 2, the load is close to the deterministic value, in contrast to the sizeable reduction in B. The variances in fitness under the sites model for h = 0.2 and 0.5 in Table 3 are also well approximated by the deterministic value of $4000uh\bar{s}$.

The weak effects of drift on the load and variance in fitness are at first sight surprising, given that with $\bar{\gamma}$ = 2 a substantial proportion of the distribution of selection coefficients falls below the nearly neutral threshold of $h\gamma < 2.5$ described above, where large departures from the deterministic values are expected. However, this part of the distribution involves very small selection coefficients, and thus makes an extremely small contribution to the genetic load and the variance in fitness.

The role of epistasis

The extent to which deleterious mutations with minor effects on fitness have epistatic effects rather than obeying an additive or multiplicative model for their combined effects, and the

typical nature of such interactions, is still an unsettled issue with respect to both theoretical (Diaz-Colunga et al. 2023) and empirical (Bakerlee et al. 2022) questions. Attempts to characterize epistasis for fitness components among deleterious mutations in a variety of organisms have used analyses of the change in the means over generations of mutation accumulation lines or populations subject to inbreeding. These experiments have yielded somewhat conflicting results, although overall there seems to be a predominance of weak synergistic epistasis (reviewed by Charlesworth 1998; Fry 2004; de Visser et al. 2012; García Dominguez et al. 2019). A problem with intepreting these results is that data on fitness, or its components such as viability or fecundity, are inherently noisy and there may be statistical biases (due to regression towards the mean) that lead to an underestimation of synergistic epistasis among deleterious mutations (Berger and Postma 2014).

Microbial systems have allowed estimates of the fitness consequences of combining deletions of two or more genes, providing extensive information about epistasis among drastically deleterious mutations (e.g., Costanzo et al. 2016; Kuzmin et al. 2018). New technologies that allow the introduction of multiple mutations into the same coding sequence have opened a path towards characterizing epistasis between mutations at the intragenic level. This can been done either at the level of a specific biochemical function, such as resistance to an antibiotic (e.g., Bershtein et al. 2006) or fluorescence intensity of the green fluorescent protein (e.g., Sarkisyan et al. 2016; Somermeyer et al. 2022), or by measurements of the fitness effects of mutant strains in microbial competition experiments (e.g., Bank et al. 2015, 2016; Puchta et al. 2016; Domingo et al. 2018; Bakerlee et al. 2022; Chen et al. 2023; Papkou et al. 2023). Interestingly, although the stability of proteins is nearly linearly reduced by increasing numbers of mutations (De Pristo et al. 2005; Bershtein et al. 2006), the effect of mutations on function or fitness in the experiments just cited seem to show a predominance of synergistic epistasis among deleterious mutations, despite complexities such as the dependence of patterns of interaction between the same mutations in different environments (Bank et al. 2016) or between-species differences in fitness landscapes (Somermeyer et al. 2022). However, there appear to be no reliable estimates of the parameters involved in quantifying epistasis (such as our ϵ), making it difficult to evaluate the relevance of our model of epistasis.

Epistasis and inbreeding load

One important aspect of our simulation results with synergistic epistasis is that B for the gene and sites models is strongly reduced by even a small amount of epistasis when $h < ½$, despite

only a small reduction in load; there is a much smaller effect on B under the gene model, where it is always small (Figure 4). This reduction in B with increasing epistasis is in sharp contrast to what is found with the model of Charlesworth (1998), which predicts that synergistic epistasis causes an *increase* in B when the logarithm of fitness is a quadratic, decreasing function of the number of deleterious mutations. As shown in section 5 of Supplementary File S1, when only homozygous genotypes are considered, the pairwise epistasis model used here is identical to the quadratic model, so the opposite behavior of the two models is surprising at first site.

The explanation lies in the fact that the model of Charlesworth (1998) assumes that $h > 0$. The quadratic term for heterozygous mutations is proportional to the negative of the product of h^2 and the square of the number of heterozygous mutations (n^2), whereas under the pairwise epistasis model it is proportional to $-hn^2$. This means that the fitness of the outbred population is relatively insensitive to the quadratic fitness term under the first model, since most mutations are present in heterozygotes, but the fully inbred genotypes have strongly reduced fitnesses due to the quadratic term, resulting in an increase in B. Under the pairwise model, there is a sharp decline in $\bar{q}$ with increasing ϵ, while the load is relatively unaffected (compare Figure S1 of Supplementary File S1 with Figure 4). Since B is strongly determined by $\bar{q}$, the net result is that synergistic epistasis reduces B, even when $h = 0$.

The consequences of epistasis for inbreeding depression are thus highly model dependent, and it cannot be assumed that synergistic epistasis necessarily leads to increased inbreeding depression compared with the multiplicative fitness model, as has been done in the past (Crow 1970; Charlesworth 1998; García-Dominguez et al. 2018). The pairwise model has the advantage of greater flexibility in modeling the fitnesses of heterozygotes and differences in selection coefficients between sites, whereas the quadratic model is forced to make somewhat arbitrary assumptions about the effects of dominance (for details, see Charlesworth 1998). However, it is important to note that we have only developed this model for within-gene interactions; the nature of epistatic interactions between deleterious mutations in different genes and their consequences for the load statistics remain to be explored.

Linkage disequilibrium: theory

It has generally been accepted that multiplicative fitness models of multi-locus mutation and selection in randomly mating populations predict an absence of linkage disequilibrium when

equilibrium has been reached in an infinite population (Haldane 1937; Kimura and Maruyama 1966). Synergistic epistasis on a multiplicative scale results in negative LD when a positive sign of D is assigned to haplotypes with pairs of mutant or wild-type alleles, whereas antagonistic epistasis causes positive LD (Kondrashov 1982, 1988; Charlesworth 1990; Shnol and Kondrashov 1993). In discrete generation finite populations, there is a general expectation that negative LD will be generated by Hill-Robertson interference in the absence of epistasis on the multiplicative fitness scale, if linkage is sufficiently tight (Hill and Robertson 1966; Felsenstein 1974; Garcia and Lohmueller 2021; Roze 2021).

Since an absence of epistasis on a linear scale results in weakly synergistic epistasis on a multiplicative scale, it might be expected that our two-site model would result in small negative values of D in the deterministic limit when the epistatic parameter $\epsilon = 0$, and that the finite population simulations without epistasis or with synergistic epistasis would also exhibit negative LD. The results presented here do not entirely conform to these expectations. Figure 2 shows that there are small positive deterministic equilibrium values of the correlation coefficient R under both models even with $h = ½$, for sufficiently small values of ϵ, although the gene model consistently gives larger values of R when $h < ½$. As explained above (in the section *The role of complementation*), the intragenic position effect that characterizes the gene model is expected to favor positive LD, due to the higher fitness of *trans* versus *cis* double heterozygotes, so that this difference between the two models is not surprising.

Similarly, the finding of positive LD in simulations of the gene model without epistasis, at least with $h < ½$ and when selection is sufficiently strong (see *Simulation results for multiple sites*; *linkage disequilibrum with no epistasis*), makes sense given the intragenic position effect on fitness. Furthermore, the negative LD sometimes found with the sites model without epistasis is indicative of the effects of HRI, consistent with the approach to zero of such LD with increasing distance between pairs of sites (Table S3 in Supplementary File S3).

The discrepancy with the standard expectation for mutation-selection balance with the sites model arises from the fact that the standard results tacitly assume that the effect of selection on homozygotes can be ignored, which is reasonable when mutant alleles are sufficiently rare; this assumption is not made here. Consistent with this interpretation, it can be seen from the results in Supplementary File S2 that, for sufficiently small values of u/s, the amount of LD becomes negligible, in accordance with the standard expectation. However, this does not explain why positive LD is generated by the sites model; we will now consider

this question. Positive LD also been found in other recent theoretical studies. Roze (2021) studied a diploid finite population model with two loci under selection according to a multiplicative version of the sites model (so that the fitness of a double heterozygote was $(1 - hs)^2$ rather than $1 - 2hs$), assuming that $\gamma = 2N_e s >> 1$ but $s << 1$. He found small positive mean D and σ_d values with sufficiently low recombination rates and $h < 0.25$ in both his approximate analytical treatment and in multi-locus simulations (see his Figure 1). Similarly, the analysis of the sites model for a two-locus system in a finite population by Ragsdale (2022) found positive mean values of D and σ_d^1 for sufficiently large values of h and r, provided that γ was sufficiently large. His model of an intragenic position effect also generated positive LD.

These findings with relatively large γ values suggest that the deterministic case should shed light on why positive values of D and related statistics are found with the sites model when recombination is rare, in situations where standard theory predicts an absence of LD; as already indicated, small positive values of D and R are indeed found in this case. Some insight into the causes of positive LD in this situation is provided by the analysis of the completely recessive case ($h = 0$) with fixed s in section 3 of Supplementary File S1. It is shown there that the reduction in frequency of the mutant allele due to selection at either site is less than expected in the absence of LD, consistent with positive LD among mutant alleles. This is because of selection against homozygous mutant alleles, either when homozygous singly or when present together in *cis*. If the term $-2sy(x + y)$ in Equation (S3.1b), which represents the effect of such selection, were replaced by $-sy(x + y)$ this effect would disappear. It therefore seems that the standard models, which ignore this effect of selection against homozygous mutations, fail to fully capture the small amount of positive LD that can be generated in the absence of HRI effects when recombination is rare.

The existence of positive LD among derived mutations discussed here should not be confused with the much more frequently observed positive D when an excess frequency of haplotypes with pairs of minor alleles was used (e.g., Table 4). As explained in the section *Simulation results for multiple sites: linkage disequilibrium with no epistasis*, this effect is expected even for neutral variants (Potapova and Kondrashov 2023).

Patterns of linkage disequilibrium in population genetic data

Previous theoretical and simulation studies have suggested that, while HRI and synergistic epistasis create negative LD between selected variants, a number of other processes, such as

antagonistic epistasis, admixture, and population bottlenecks could result in positive LD (Sohail et al. 2017; Sandler et al. 2021; Garcia and Lohmueller 2021; Ragsdale 2022). Several recent studies have obtained estimates of signed LD among variants in coding sequences in natural populations of *D. melanogaster*, humans, *Capsella grandiflora* and *Schizophyllum commune* (Pool et al. 2012; Sohail et al. 2017; Sandler et al. 2021; Garcia and Lohmueller 2021; Ragsdale 2022; Stolyarova et al. 2022). Garcia and Lohmueller (2021) found a larger magnitude of negative LD among doubleton nonsynonymous variants compared with synonymous variants in humans when using Lewontin's D' statistic (Lewontin 1964). As described in the subsection *Effects of weighting towards low frequencies on LD statistics, D'* is biased towards negative values when low frequency variants are used, and differences in D' between neutral and selected sites do not necessarily reflect the effects of epistasis or HRI alone. Corrections for this bias are thus needed to disentangle the relative contribution of the causes of signed LD at selected sites, if this method is to be used .

Sohail et al. (2017) and Sandler et al. (2021) used estimates of LD based on the excess or deficiency over the Poisson expectation of the variance in the numbers of minor alleles over a set of defined genomic regions, which is proportional to the sum of pairwise D values over the regions in question, whereas Stolyarova et al. (2022) used estimates of the correlation coefficients between pairs of variants, polarizing with respect to minor versus major alleles. Both of these methods are likely to give higher magnitudes of the LD statistic for variants at higher frequencies. Ragsdale (2022) polarized LD using the ancestral/derived criterion; he employed the σ_d^1 measure, whose denominator involves the crossproducts of the four variant frequencies at a pair of diallelic sites (see Equation 7b). This will tend to give *higher* values for sites with lower frequencies, for a given correlation coefficient.

These methodological differences across studies make it hard to compare the results, but nevertheless there are some patterns in common. Notably, in most cases signed LD between synonymous sites was found to be significantly higher than that between nonsynonymous variants, with the exception of the analysis of the hypervariable mushroom species *S. commune* by Stolyarova et al. (2022), where the opposite pattern was observed – see their Appendix 2, Figure 1.

There seems to be a preponderance of slightly negative values of LD between loss-of-function (LOF) mutations (presumed to be the most harmful), but positive values between pairs of synonymous and nonsynonymous variants. Because these patterns applied to loosely linked variants in humans and *Drosophila*, and the extent of under-dispersion of loosely

linked loss-of-function alleles was found to be much higher than explained by simulations that included the demographic history of humans, Sohail *et al*. (2017), proposed that synergistic epistasis was responsible.

However, a plausible explanation for a predominance of negative signed LD between loosely linked LOF mutations is that these are kept at very low frequencies in the population. This means that a finite sample from the population has a high probability of not containing the – – haplotype at a pair of diallelic sites. This automatically generates $D < 0$, even if D is zero in the population as a whole, as would be expected for most pairs of sites that are not tightly linked. With two sites with $D = 0$ in the population and a minor allele frequency q at each site, a sample of size k has a probability of approximately $[1 – \exp(–kq)]^2$ of segregating at both sites, and a probability of approximately kq^2 of containing the – – haplotype, giving a net probability of $1 – kq^2/[1 – \exp(–kq)]^2$ that a segregating sample contains no – – haplotypes. With $k = 1000$ and $q = 0.001$, this probability is equal to 0.9975. But with $q = 0.01$, it is only 0.9. Despite the fact that the expected value of D is zero in both cases, with $q = 0.001$ a finite set of pairs of segregating sites is likely to have a high probability of containing none or only a few – – haplotypes, resulting in a large preponderance of cases with $D < 0$. This is illustrated in section 5 of Supplementary File S6. Observations of negative LD between loosely linked or independent LOF mutations, such as those shown in Figure S6 of Sohail et al. (2017), seem likely to be the result of this sampling effect. It follows that it cannot safely be inferred that synergistic epistasis is the cause of this pattern; our analytical and simulation results confirm that it does indeed generate negative LD for the sites model but not necessarily for the gene model (see Figures 3 and 5).

Our modeling work adds a new perspective on findings of positive LD among closely linked pairs of putatively selected variants. If fitness is governed by the gene model, then LD between variants at non-synonymous sites within the same gene should be relatively larger than under the sites model, and the sign should often be positive, providing the effects of sampling just discussed do not overwhelm the effects of selection. However, as we have discussed in the section *The role of complementation*, the gene model is unlikely to apply between different genes, but is likely to be applicable within a gene. Our models thus make the following predictions about LD patterns in population genomics data. (*a*) signed LD between alleles at nonsynonymous sites within genes should be positive or nearly zero (assuming that the population is panmictic and at equilibrium). (*b*) LD between nonsynonymous variants in the same gene should be higher than between variants in different

genes. (*c*) There should be higher LD between nonsynonymous vs synonymous variants within the same gene.

However, there are several caveats that need to be borne in mind when using these finding to interpret data. First, the sign of LD statistics when calculated using minor/major alleles (as opposed to using derived/ancestral) is almost always positive, unless the minor alleles are very rare (Potapova and Kondrashov 2023). This applies to neutral sites (Potapova and Kondrashov 2023) and to deleterious variants that obey the sites model as well as the gene model – see Table 3. It is thus not surprising that the studies which polarize their variants by the minor/major allele criterion (Pool *et al*. 2012; Sohail et al. 2017; Sandler et al. 2021; Stolyarova et al. 2022) often found positive LD. Second, weighting towards low allele frequencies (Good 2022) also tends to promote positive LD, as described in the subsection *Effects of weighting towards low frequencies on LD statistics*, and should thus be avoided if the sign of LD is used for inferences about selection. It follows that patterns of signed LD using this criterion do not necessarily imply effects of admixture, non-equilibrium demography, or antagonistic/complementary epistasis, which have all been proposed as possible explanations in these previous studies.

This leaves the study of humans by Ragsdale (2022), who found that LD between derived missense variants was found to be significantly positive within the same domain of a gene, as opposed to near-zero values between SNPs in different domains, even after matching the distance between SNPs found within compared between domains. This observation could be explained by a gene model operating at the level of domains within genes, perhaps because there is higher probability of complete failure of complementation within domains vs across the entire gene. Ragsdale's estimates of $\bar{\sigma}_d^1$ from the Yoruba population (which is closest to having a stable size) between missense variants within genes is ~0.02 (his Figure 8). As the average length of a gene in humans is quite large (~ 60 kb), it is difficult to compare this estimate to that obtained in our simulations. Other estimates of $\bar{\sigma}_d^1$ within 100 base pairs provided by Ragsdale (2002) range from 0.25-0.3 (within the same gene; his Figure S17), ~0.5 (within the same domain; his Figure S18), and 0.2 (outside domains; his Figure S19). While estimates of $\bar{\sigma}_d^1$ within 100 bases in our simulations are within this range (Table S5), the large SEs make it difficult to make any reliable conclusion. In general, however, both demographic processes (like admixture and bottlenecks) as well as selective processes such as positive epistasis can result in positive LD between derived nonsynonymous variants, and joint modeling of these processes in a specific population

would be required to distinguish between these scenarios, taking into account important features such as recombination hotspots, which strongly influence the rates of recombination within genes in taxa such as mammals and flowering plant (e.g., Choi et al. 2013; Baudat et al. 2014; Sandler et al. 2021; Joseph et al. 2024).

There are multiple reasons why within-gene LD should be higher than between-gene LD: (a) On average, there is a much higher frequency of recombination between pairs of sites within different genes than within the same gene (b) fitness follows the gene model within but not between genes (c) there is antagonistic epistasis within genes but not between genes. While it might be possible to differentiate between (a) and (b) by comparing the patterns of LD between species with compact genes with little spacing and/or low recombination rates between genes with patterns in species with high recombination rates between genes, it will be harder to distinguish between (a) and (c).

In summary, it will probably be quite difficult to distinguish the different possible explanations for patterns of signed LD from population genomic data, but it is first essential to use the ancestral/derived rather than the minor/major allele criterion for the sign of LD . The problem is further complicated by the fact that, if average gene lengths and rates of recombination are sufficiently large, even patterns of LD within genes may mostly reflect genome-wide patterns generated by demographic processes rather than the effects of selection. Analyses of large-scale resequencing data in the light of population genetic models of the type investigated here, using suitable choices of species with respect to their genomic properties and demographic histories, offer the best way to approach this problem.

Acknowledgements

We thank Kirk Lohmueller for his helpful comments on the manuscript. This research was conducted using computational resources provided by ITS Research Computing at the University of North Carolina at Chapel Hill. PJ was funded by the National Institutes of Health grant 1R35GM154969-01.

Literature Cited

Altshuler D, Donnelly P, The International HapMap Consortium. 2005. A haplotype map of the human genome. Nature. 437(7063):1299–1320.

Bakerlee CW, Nguyen Ba AN, Shulgina Y, Rojas Echenique JI, Desai MM. 2022. Idiosyncratic epistasis leads to global fitness-correlated trends. Science 376:630-635.

Bank C, Hietpas RT, Jensen JD, Bolon DN. 2015. A systematic survey of an intragenic epistatic landscape. Mol. Biol. Evol. 31:229-238.

Bank C, Matuszewski S, Hietpas RT, Jensen JD. 2016. On the (un)predictability of a large intragenic fitness landscape. Proc. Natl. Acad. Sci. USA 113:14085-14090.

Barton NH, Otto SP. 2005. Evolution of recombination due to random drift. Genetics 65:123-144.

Baudat F, Imai Y, De Massy B. 2014. Meiotic recombination in mammals: localization and regulation. Nat. Rev. Genet. 14:794-806.

Benzer S. 1957. The elementary units of heredity. In: McElroy WD, Glass B, editors. The Chemical Basis of Heredity. Baltimore, MD: Johns Hopkins Press. p. 70-93.

Berger D, Postma E. 2015. Biased estimates of diminishing-returns epistasis: Empirical evidence revisited. Genetics 198:1417-1420.

Bershtein S, Segal M, Bekerman S, Tokuriki N, Tawfik DS. 2006. Robustness–epistasis link shapes the fitness landscape of a randomly drifting protein. Nature 444:929-931.

Bertorelli G, Raffini M, Bortoluzzi C, Ianucci A, Trucchi E, Morales HE, Van Oosterhout C. 2022. Genetic load: genomic estimates and applications in non- model animals. Nat. Rev. Genet. 23:492-503.

Bulmer MG. 1974. Linkage disequilibrium and genetic variability. Genet. Res. 23:281-289.

Charlesworth B. 1998. The effect of synergistic epistasis on the inbreeding load. Genet. Res. 71:85-89.

Charlesworth B. 2015. Causes of natural variation in fitness: evidence from studies of Drosophila populations. Proc. Natl. Acad. Sci. USA 12:1662-1669.

Charlesworth B. 2018. Mutational load, inbreeding depression and heterosis in subdivided populations. Mol. Ecol. 24:4991-5003.

Chen Y, Hu R, LI K, Zhang Y, Fu L, Zhang J, Si T. 2023. Deep mutational scanning of an oxygen-Independent fluorescent protein CreiLOV for comprehensive profiling of mutational and epistatic effects. ACS Synth. Biol. 12:1461-1473.

Choi K, Zhao X, Kelly KA, Venn O, Higgins JD, Yelina NE, Hardcastle TJ, Ziolkowski PA, Copenhaver GP, Franklin FCH, et al. 2013. Arabidopsis meiotic crossover hot spots overlap with H2A.Z nucleosomes at gene promoters. Nat Genet. 45(11):1327–1336.

Costanzo M, VanderSluis B, Koch EN, Baryshnikova A, Pons C, Tan G, et al. 2016. A global genetic interaction network maps a wiring diagram of cellular function. Science 353:aaf1420-1421-1414.

Crow JF. 1958. Some possibilities for measuring selection intensities in man. Hum. Biol. 30:1-13.

Crow JF. 1970. Genetic loads and the cost of natural selection. In: Kojima K, editor. Mathematical Topics in Population Genetics. Berlin: Springer-Verlag. p. 128-177.

Crow JF. 1993. Mutation, mean fitness, and genetic load. Oxf. Surv. Evol. Biol. 9:3-42.

DePristo MA, Weinreich DM, Hartl DL. 2005. Missense meandering in sequence space: a biophysical view of protein evolution. Nature Rev. Genet. 6:678-687.

Domingo J, Diss G, Lehner B. 2018. Pairwise and higher-order genetic interactions during the evolution of a tRNA. Nature 558:117-121.

De Visser JAGM, Cooper TF, Elena SF. 2011. The causes of epistasis. Proc. R. Soc. B. 278:3617-3624.

Diaz-Colunga J, Skwara A, Gowda K, Diaz-Uriarte R, Tikhonov M, Bajic D, Sanchez A. 2023. Global epistasis on fitness landscapes. Phil. Trans. R. Soc. B. 378:20220053.

Duncan IW. 2002. Transvection effects in Drosophila. Annu. Rev. Genet. 36:521-556.

Eng T, Guacci V, Koshland D. 2015. Interallelic complementation provides functional evidence for cohesin–cohesin interactions on DNA. Mol. Biol. Cell 26:4224-4235.

Ewens WJ. 2004. Mathematical Population Genetics. 1. Theoretical Introduction. New York: Springer.

Eyre-Walker A, Woolfit M, Phelps T. 2006. The distribution of fitness effects of new deleterious amino acid mutations in humans. Genetics. 173:891–900.

Felsenstein J. 1974. The evolutionary advantage of recombination. Genetics 78:737-756.

Fincham JRS, Day PR, Radford A. 1979. Fungal Genetics. Berkeley, CA: University of California Press.

Fisher RA. 1930. The distribution of gene ratios for rare mutations. Proc. R. Soc. Ed. 50:205-220.

Frankham R. 2023. Effects of genomic homozygosity on total fitness in an invertebrate: lethal equivalent estimates for *Drosophila melanogaster*. Cons. Genet. 24:193-201.

Fry JD. 2004. On the rate and linearity of viability declines in Drosophila mutation-accumulation experiments. Genetics 166:797-806.

Galouzis CC, Prud'homme B. 2021. Transvection regulates the sex-biased expression of a fly X-linked gene. Science 371:396-400.

Garcia JA, Lohmueller KE. 2021. Negative linkage disequilibrium between amino acid changing variants reveals interference among deleterious mutations in the human genome. PloS Genet. 17:e1009676.

García-Dominguez S, García C, Quesada H, Caballero A. 2018. Accelerated inbreeding depression suggests synergistic epistasis for deleterious mutations in *Drosophila melanogaster*. Heredity 123:709-722.

García-Dorado A, Hedrick PW. 2023. Some hope and many concerns on the future of the vaquita. Heredity 130:179-182.

Good BH. 2022. Linkage disequilibrium between rare mutations. Genetics 220:iyac004.

Haldane JBS. 1937. The effect of variation on fitness. Am. Nat. 71:337-349.

Haller, BC, Messer, PW. 2023 SLiM 4: Multispecies eco-evolutionary modeling. Am. Nat. 201: E127–E139.

Hawley RS, Gilliland WD. 2006. Sometimes the result Is not the answer: The truths and the lies that come from using the complementation test. Genetics 174:5-15.

Hill WG, Robertson A. 1966. The effect of linkage on limits to artificial selection. Genet. Res. 8:269-294.

Hill WG, Robertson A. 1968. Linkage disequilibrium in finite populations. Theor. Appl. Genet. 38:226-231.

Jinks JL. 1983. Biometrical genetics of heterosis. In: Frankel R, editor. Heterosis: Reappraisal of Theory and Practice. Berlin: Springer-Verlag. p. 1-46.

Johri P, Charlesworth B, Jensen JD. 2020. Toward an evolutionarily appropriate null model: Jointly inferrring demography and purifying selection. Genetics 215:173-192.

Johri P, Riall K, Becher H, Excoffier L, Charlesworth B, Jensen JD. 2021. The impact of purifying and background selection on the inference of population history: problems and prospects. Mol Biol Evol. 38(7):2986–3003.

Joseph J, Prentout D, Laverré A, Duret L. 2024. High prevalence of PRDM9-independent recombination hotspots in placental mammals. Proc. Natl. Acad. Sci. USA 121:e2401973121.

Kardos M, Armstrong EE, Ftzpatrick SW, S. H, Hedrick PW, Miller DE, Tallmon DA, Funk WC. 2021. The crucial role of genome-wide genetic variation in conservation. Proc. Natl. Acad. Sci. USA 118:e2104642118.

Kim BY, Huber CD, Lohmueller KE. 2017. Inference of the distribution of selection coefficients for new nonsynonymous mutations using large samples. Genetics. 206:345–361.

Kimura M, Maruyama T, Crow JF. 1963. The mutation load in small populations. Genetics 48:1303-1312.

Kimura M, Maruyama T. 1966. The mutational load with epistatic gene interactions in fitness. Genetics 54:1303-1312.

Kondrashov AS. 1982. Selection against harmful mutations in large sexual and asexual populations. Genet. Res 44:325-332.

Kondrashov AS. 1988. Deleterious mutations and the evolution of sexual reproduction. Nature 336:435-440.

Kong A, Frigge ML, Masson G, Besenbacher S, Sulem P, Magnusson G, Gudjonsson SA, Sigurdsson A, Jonasdottir Aslaug, Jonasdottir Adalbjorg, et al. 2012. Rate of *de novo* mutations and the importance of father's age to disease risk. Nature. 488(7412):471–475.

Kryukov GV, Pennachio LA, Sunyaev S. 2007. Most rare missense alleles are deleterious in humans: implications for complex disease and association studies. Am. J. Hum. Genet. 80:727-739.

Kuzmin E, VanderSluis B, Wang W, Tan G, Despande R, Cheng Y, et al. 2018. Systematic analysis of complex genetic interactions. Science 360:eaao1729.

Kyriazis C. C., R. K. Wayne, and K. E. Lohmueller, 2021 Strongly deleterious mutations are a primary determinant of extinction risk due to inbreeding depression. Evolution Letters 5: 33–47.

Kyriazis CC, Robinson JA, Nigenda-Morales SF, Beichman AC, Rojas-Bracho L, Robertson KM, et al. 2023. Models based on best-available information support a low inbreeding load and potential for recovery in the vaquita. Heredity 130:183-187.

Kyriazis CC, Robinson JA, Lohmueller KE. 2023. Using computational simulations to model deleterious variation and genetic load in natural populations. Am. Nat. 302:737-752.

Langley CH, Crow JF. 1974. The direction of linkage disequilibrium. Genetics 63:229-238.

Lewis EB. 1954. The theory and application of a new method of detecting chromosomal rearrangements in *Drosophila melanogaste*r. Am. Nat. 88:225-239.

Lewontin RC. 1964. The interaction of selection and linkage. I. General considerations, heterotic models. Genetics 49:49-67.

Mellert DJ, Truman JW. 2012. Transvection Is common throughout the Drosophila genome. Genetics 191:1121-1141.

Misra S, Crosby MA, Mungall CJ, Matthews BB, Campbell KS, et al. 2002. Annotation of the *Drosophila melanogaster* euchromatic genome: a systematic review. Gen. Biol. 3:Research0083.

Morton NE, Crow JF, Muller HJ. 1956. An estimate of the mutational damage in man from data on consanguineous marriages. Proc. Natl. Acad. Sci. USA 42:855-863.

Muller HJ. 1950. Our load of mutations. Am. J. Human Genet. 2:111-176.

Navarro-Quiles C, Lup SD, Munoz-Nortes T, Candela H, Micol JL. 2024. The genetic and molecular basis of haploinsuf!ciency in flowering plants. Trnds. Plant Sci. 29:72-95.

Nei M. 1968. The frequency distribution of lethal chromosomes in finite populations. Proc. Natl. Acad. Sci. USA 60:517-524.

Nordborg M, Franklin IR, Feldman MW. 1995. Effects of cis-trans viability selection on some two-locus models. Theor. Pop. Biol. 47:365-392.

Ohta T, Kimura M. 1969. Linkage disequilibrium due to random genetic drift. Genet. Res. 13:47-55.

Ohta T, Kimura M. 1971. Linkage disequilibrium between two segregating nucleotide sites under steady flux of mutations in a finite population. Genetics 68:571-580.

Papkou A, Garcia-Pastor L, Escudero JA, Wagner A. 2023. A rugged yet easily navigable fitness landscape. Science 382:eadh3860.

Pérez-Pereira N, Pouso R, Rus A, López-Cortegano E, García-Dorado A, Quesada H, Caballero A. 2021. Long-term exhaustion of the inbreeding load in *Drosophila melanogaster*. Heredity 127:373-383.

Pontecorvo G. 1958. Trends in Genetic Analysis. New York, NY: Columbia University Press.

Pool JE, Corbett-Detig RB, Sugino RP, Stevens KA, Cardeno CM, Crepeau MW, et al. 2012. Population genomics of sub-Saharan *Drosophila melanogaster*: African diversity and non-African admixture. PloS Genet 8:e1003080.

Potapova NA, Kondrashov AS. 2023. Positive association between alleles at selectively neutral loci. Biorxiv 2023.06.14.544962:1-12.

Puchta O, Czeke B, Czaja H, D. T, Sanguinetti G, Kudla G. 2016. Network of epistatic interactions within a yeast snoRNA. Science 352:840-844.

Ragsdale AP. 2022. Local fitness and epistatic effects lead to distinct patterns of linkage disequilibrium in protein-coding genes. Genetics 221:iyac097.

Robinson JA, Kyriazis CC, Yuan SC, Lohmueller KE. 2023. Deleterious variation in natural populations and implications for conservation genetics. Annu. Rev. Animal Biosci. 11:93-114.

Robinson JA, Kyriazis CC, Nigenda-Morales SF, Beichman AC, Rojas-Bracho L, R, Robertson KM, et al. 2022. The critically endangered vaquita is not doomed to extinction by inbreeding depression. Science 376:635-639.

Roze D. 2021. A simple expression for the strength of selection on recombination generated by interference among mutations. Proc. Natl. Acad. Sci. USA 118:e2022805118.

Sandler G, Wright SI, Agrawal AF. 2021. Patterns and causes of signed linkage disequilibria in flies and plants. Mol. Biol. Evol. 38:4310-4321.

Sanjak JS, Long AD, Thornton KR. 2015. A Model of compound heterozygous, loss-of-function alleles Is broadly consistent with observations from complex-disease GWAS datasets. PloS Genet. 13:e1006573.

Sarkisyan KS, Bolotin DA, Meer MV, Usmanova DR, Mishin AS, Sharanov GV, et. al. 2016. Local fitness landscape of the green fluorescent protein. Nature 533:397-401.

Shnol EE, Kondrashov AS. 1993. Effect of selection on the phenotypic variance. Genetics 134:995-996.

Simmons MJ, Crow JF. 1977. Mutations affecting fitness in *Drosophila* populations. Ann. Rev. Genet. 11:49-78.

Sohail M, Vakhrusheva OA, Sul JH, Pulit SA, Francioli LC, Genome of the Netherlands Consortium, et al. 2017. Negative selection in humans and fruit flies involves synergistic epistasis. Science 356:539-542.

Somermeyer LG, Fleiss A, Mishin AS, Bozhanova NG, Igolkina AA, Meiler J, Pujol M-EA, Putintseva EV, Sarkisyan KS, Kondrashov FA. 2022. Heterogeneity of the GFP fitness landscape and data-driven protein design. Elife 11:e75842.

Stolyarova AV, Neretina TV, Zvyagina EA, Fedotova AV, Kondrashov AS, Bazykin GA. 2022. Complex fitness landscape shapes variation in a hyperpolymorphic species. Elife 11:e76073.

Sved JA, Ayala FJ. 1970. A population cage test for heterosis in *Drosophila pseudoobscura*. Genetics 66:97-113.

Takahasi KR, Innan H. 2007. The direction of linkage disequilibrium: A new measure based on the ancestral-derived status of segregating alleles. Genetics 179:1705-1712.

Teixeira JC, Huber CD. 2021. The inflated significance of neutral genetic diversity in conservation genetics. Proc. Natl. Acad. Sci. USA 118:e2015096118.

Thornton KR, Foran AJ, Long AD. 2013. Properties and modeling of GWAS when complex disease risk is due to non-complementing, deleterious mutations in genes of large effect. PloS Genet. 9:e1003258.

Wilton AN, Sved JA. 1979. X-chromosomal heterosis in *Drosophila melanogaster*. Genet. Res. 34:303-315.

Wright S. 1931. Evolution in Mendelian populations. Genetics 16:97-159.

Wright S. 1968. Evolution and the Genetics of Populations. Vol.1. Genetic and Biometric Foundations. Chicago, IL: University of Chicago Press.

Appendix

Deterministic treatment of the gene model: no epistasis

Here we use the the two-locus fitness matrix for the additive gene model presented in the Methods section. Let the mutation rate per locus from + to – be u; if $u << hs$, the frequency q of the – allele at each locus is << 1. Let the recombination rate between the two loci be r and the coefficient of linkage disequilibrium be D. Use of these assumptions generates a set of recurrence relations for the frequencies of the four haplotypes, x_1 (+ +), x_2 (– +), x_3 (+ –) and x_4 (– –).

Given the symmetry of the fitness matrix, we can write $x_2 = x_3 = x = q(1 - q) - D$, $x_4 = y = q^2 + D$, $x_1 = 1 - x - y$. It follows that only the recursions for x and y (or q and D) need be considered. We have:

$$\bar{w}x' = x[1 - (1 - 2x - y)hs - 2xs - 1.5ys] + r[y(1 - x - y)(1 - 2hs) - x^2(1 - s)] + u(1 - 2x - y) - ux \quad \text{(A.1}a\text{)}$$

$$\bar{w}y' = y[1 - (1 - 2x - y)2hs - 3xs - 2ys] - r[y(1 - x - y)(1 - 2hs) - x^2(1 - s)] + 2ux + u^2(1 - 2x - y) \quad \text{(A.1}b\text{)}$$

At equilibrium, single-locus theory implies that $\bar{w} = 1 - O(u)$ and x is $O(u/hs)$, so that $\bar{w}$ can be replaced by 1 in Equations (A.1) as an accurate approximation. In addition, u^2 can be neglected for biologically realistic values of the per nucleotide site mutation rates. The equilibrium values of x and y are then given by:

$$0 = x[(1 - 2x - y)hs + 2xs + 1.5ys] - r[y(1 - x - y)(1 - 2hs) - x^2(1 - s)] - u(1 - 3x - y) \quad \text{(A.2}a\text{)}$$

$$0 = y[(1 - 2x - y)2hs + 3xs + 2ys] + r[y(1 - x - y)(1 - 2hs) - x^2(1 - s)] - 2ux \quad \text{(A.2}b\text{)}$$

This is a pair of non-linear equations in x and y, which can only be solved exactly by numerical methods. A simple procedure for the case of $r = 0$ is to note that s can be eliminated by replacing u with $\tilde{u} = u/s$. This yields the following equations:

$$2(1-h)x^2 + [h + 3\,\tilde{u} + (1.5-h)y]x - \tilde{u}(1-y) = 0 \qquad \text{(A.3}a\text{)}$$

$$2(1-h)y^2 + [2h + (3-4h)x]y - 2\tilde{u}x = 0 \qquad \text{(A.3}b\text{)}$$

Since y is $O(x^2)$, an initial solution to Equation (A.3a) is obtained by setting $y = 0$ and solving the resulting quadratic equation for x. This can be substituted into Equation (A.3b), and solving for y. This value of y is then substituted into Equation (A.3a), and the whole procedure repeated until the values of x and y stabilize. The allele frequency q of deleterious alleles is given by $x + y$, and $D = q(1 - q) - x$. The correlation coefficient in allele frequency between the two loci is $R = D/q(1 - q) = 1 - x/\,q(1 - q)$.

An approximate solution for x can be obtained as follows, by neglecting the terms in y in Equation (A.2a), thereby converting it into a simple quadratic equation in x, with coefficients $a = 2(1 - h)$ for x^2, $b = h + 3\tilde{u}$ for x and $c = -\,\tilde{u}$ for the constant term. The discriminant $\Delta = (b^2 - 4ac)^{\frac{1}{2}}$ can be approximated by using the first two terms in its Taylor series expansion as follows:

$$\Delta = b[1 - 4acb^{-2}]^{\frac{1}{2}} \approx b - 2acb^{-1} - (2a^2c^2b^{-3})$$

so that:

$$x \approx \tilde{u}(h + 3\tilde{u})^{-1}[1 - 2(1-h)\tilde{u}(h + 3\tilde{u})^{-2}] \qquad \text{(A.4}a\text{)}$$

In addition, y can be approximated by ignoring the term in y^2 in Equation (3b), which should be accurate when q is small:

$$y \approx 2\tilde{u}x[2h + (3-4h)x]^{-1} \qquad \text{(A.4}b\text{)}$$

With $h > 0$, these expressions can be further approximated by ignoring the terms in $3\tilde{u}$ in Equation (A.4a) and second-order terms in x in Equation (A.4b), giving:

$$x \approx \tilde{u}h^{-1}[1 - 2(1-h)\tilde{u}h^{-2}] \qquad \text{(A.5}a\text{)}$$

and

$$y \approx xz \qquad \text{(A.5}b\text{)}$$

$$q = x(1 + z) \qquad \text{(A.5}c\text{)}$$

where $$z = \left[1 - \tfrac{1}{2}(3h^{-1} - 4)x\right]\tilde{u}h^{-1} \qquad \text{(A.5}d\text{)}$$

so that $$D = q(1-q) - x = xz - [x(1+z)]^2 \qquad \text{(A.5}e\text{)}$$

If zx^2 can be ignored, we have:

$$D \approx x(z - x) \qquad \text{(A.6}a\text{)}$$

$$R \approx 1 - (1+z)(1-x)^{-1} \qquad \text{(A.6}b\text{)}$$

To this level of approximation, the sign of D is determined by that of $z - x$. After some algebra, we obtain:

$$z - x = (\tilde{u}h^{-1})^2\,[(2h)^{-1} + (3h^{-1} - 4)(1-h)\tilde{u}h^{-2}] \qquad \text{(A.6}c\text{)}$$

If $0 < h < ¾$, this expression implies that positive LD is always expected when zx^2 is sufficiently small relative to zx, which requires low frequencies of the deleterious alleles.

$$D \approx \tfrac{1}{2}\tilde{u}^3h^{-4} \qquad \text{(A.7}a\text{)}$$

$$R \approx \tfrac{1}{2}\tilde{u}^2h^{-3} \qquad \text{(A.7}b\text{)}$$

Deterministic treatment of the sites model: no epistasis

Using the fitness matrix for the additive sites model, Equations (A.2a) and (A2b) are replaced with:

$$0 = x[(1-2x-y)hs + x(1+2h)s + y(1+h)s] - r[y(1-2x-y) - y^2](1-2hs) - u(1-3x-y) \qquad \text{(A.8}a\text{)}$$

$$0 = y[(1-2x-y)2hs + 2x(1+h)s + 2ys] + r[y(1-2x-y) - y^2](1-2hs) - 2ux \qquad \text{(A.8}b\text{)}$$

With $r = 0$, these equations simplify to:

$$x^2 + (h + 3\,\tilde{u} + y)x - \tilde{u}(1-y) = 0 \qquad \text{(A.9}a\text{)}$$

$$(1-h)y^2 + [h + (1-h)x]y - \tilde{u}x = 0 \qquad \text{(A.9}b\text{)}$$

These equations can be solved numerically by the method used for the gene model. Similarly, approximate expressions for x and y can be obtained by neglecting the terms in y in Equation (A.8a) and in y^2 in Equation (A.8b). These yield the following equations:

$$x \approx \tilde{u}(h + 3\tilde{u})^{-1}[1 - \tilde{u}(h + 3\tilde{u})^{-2}] \qquad \text{(A.10}a\text{)}$$

$$y \approx \tilde{u}x[h + (1 - h)x]^{-1} \qquad \text{(A.10}b\text{)}$$

A further approximation can for $h > 0$ be obtained in the same way as for the gene model:

$$x \approx \tilde{u}h^{-1}(1 - \tilde{u}h^{-2}) \qquad \text{(A.11}a\text{)}$$

$$y \approx xw \qquad \text{(A.11}b\text{)}$$

where

$$w = [1 - (h^{-1} - 1)x]\tilde{u}h^{-1} \qquad \text{(A.11}c\text{)}$$

$$q = x + y = x(1 + w) \qquad \text{(A.11}d\text{)}$$

If wx^2 can be ignored relative to wx, we have:

$$D \approx x(w - x) \qquad \text{(A.12}a\text{)}$$

$$R \approx 1 - (1 + w)(1 - x)^{-1} \qquad \text{(A.12}b\text{)}$$

Similarly to the gene model, the sign of D is then determined by that of $w - x$:

$$w - x = (\tilde{u}h)^{-2}\,[1 + (h^{-1} - 1)\tilde{u}h^{-2}] \qquad \text{(A.12}c\text{)}$$

If $h \leq 1$, this expression implies that positive LD is always expected when wx^2 is sufficiently small. A further level of approximation can be obtained in the same way as for the gene model:

$$D \approx \tilde{u}^3h^{-3} \qquad \text{(A.13}a\text{)}$$

$$R \approx \tilde{u}^2h^{-2} \qquad \text{(A.13}b\text{)}$$

It is easily seen that identical values of x and y for the two models are obtained from both the exact and approximate expressions when $h = ½$, as expected. If $h < ½$, it is also easy

to see from the approximate expressions that both x and y are smaller for the gene model than the site model, as might be expected intuitively. Since $q = x + y$, this also implies that the deleterious allele frequencies are lower under the gene model, again as expected.

Deterministic gene model with epistasis

The equilibrium equations with $r = 0$ for when epistasis is included are as follows:

$$2(1-h)x^2 + [h + 3\,\tilde{u} + (1.5(1+\epsilon) - h)y]x - \tilde{u}(1-y) = 0 \qquad \text{(A.14}a\text{)}$$

$$(1-h)(1+\epsilon)y^2 + (1+\epsilon)\{h + (1.5 - 2h)x]\}y - \tilde{u}x = 0 \qquad \text{(A.14}b\text{)}$$

These expressions can be used for the exact numerical solutions. Approximate solutions for $h > 0$ can be obtained on the same lines as before. No change is needed to Equation (A.4a), the first approximate solution for x. However, Equation (A.4b) is replaced by:

$$y \approx 2\tilde{u}x(1+\epsilon)^{-1}[2h + (3-4h)x]^{-1} \qquad \text{(A.15}a\text{)}$$

Equation (A.5d) is replaced by:

$$z = \left[1 - \tfrac{1}{2}(3h^{-1} - 4)x\right]\tilde{u}h^{-1}(1+\epsilon)^{-1} \qquad \text{(A.15}b\text{)}$$

If $x\epsilon \ll 1$, this can be approximated by:

$$z \approx \left[1 - \tfrac{1}{2}(3h^{-1} - 4)x - \epsilon\right]\tilde{u}h^{-1} \qquad \text{(A.15}c\text{)}$$

Using Equation (6c), we have:

$$z - x = (\tilde{u}h^{-1})^2\,[(2h)^{-1} + (3h^{-1} - 4)(1-h)\tilde{u}h^{-2} - \epsilon h\tilde{u}^{-1}] \qquad \text{(A.15}d\text{)}$$

If $\epsilon > 0$ (i.e., there is synergistic epistasis), this shows that a modest degree of epistasis can result in negative LD when $h\varepsilon \gg \tilde{u}$; a necessary condition is $1 < 2\epsilon\tilde{u}^{-1}$.

The further approximations used in Equations (A.7) become:

$$D \approx \tilde{u}^3 h^{-3}[(2h)^{-1} - \epsilon h \tilde{u}^{-1}] \qquad \text{(A.16}a\text{)}$$

$$R \approx \tilde{u}^2 h^{-2}[(2h)^{-1} - \epsilon h \tilde{u}^{-1}] \qquad \text{(A.16}b\text{)}$$

Deterministic site model with epistasis

The fitness matrix for the two-site model described in the Methods section is used here. The equilibrium equations with $r = 0$ become:

$$(1 + 2h\epsilon)x^2 + \{h + 3\,\tilde{u} + [1 + (1 + h)\epsilon]y\}x - \tilde{u}(1 - y) = 0 \qquad \text{(A.17}a\text{)}$$

$$(1 - h)(1 + \epsilon)y^2 + [h + (1 - h)x](1 + \epsilon)y - \tilde{u}x = 0 \qquad \text{(A.17}b\text{)}$$

These equations can be solved numerically, as before. The first level of approximation with $h > 0$ gives:

$$x \approx \tilde{u}(h + 3\tilde{u})^{-1}[1 - (1 + 2h\epsilon)\tilde{u}(h + 3\tilde{u})^{-2}] \qquad \text{(A.18}a\text{)}$$

$$y \approx \tilde{u}x(1 + \epsilon)^{-1}[h + (1 - h)x]^{-1} \qquad \text{(A.18}b\text{)}$$

The second level of approximation gives:

$$x \approx \tilde{u}h^{-1}[1 - (1 + 2h\epsilon)\tilde{u}h^{-2}] \qquad \text{(A.19}a\text{)}$$

$$y \approx xw \qquad \text{(A.19}b\text{)}$$

where:

$$w = (1 + \epsilon)^{-1}\,[1 - (h^{-1} - 1)x] \qquad \text{(A.19}c\text{)}$$

$$w - x \approx \tilde{u}h^{-1}\,\{(1 + \epsilon)^{-1}\,[1 - (h^{-1} - 1)x] - [1 - (1 + 2h\epsilon)\tilde{u}h^{-2}]\} \qquad \text{(A.19}d\text{)}$$

If $\epsilon \ll 1, (1 + \epsilon)^{-1} \approx 1 - \epsilon$, and $\epsilon\tilde{u}h^{-1}$ is the dominant term in the above expression, so that we obtain the following approximation:

$$w - x \approx \tilde{u}^2 h^{-2}[(1 + \tilde{u}h^{-2}) - \epsilon h \tilde{u}^{-1}] \qquad \text{(A.20)}$$

If $h\epsilon \gg \tilde{u}$, $w - x < 0$, implying that D and R are also negative, as was also the case with the gene model.

The third level of approximation gives:

$$D \approx \tilde{u}^3 h^{-3}(1 - h\tilde{u}^{-1}\,\epsilon) \qquad \text{(A.21}a\text{)}$$

$$R \approx \tilde{u}^2 h^{-2}(1 - h\tilde{u}^{-1}\,\epsilon) \qquad \text{(A.21}b\text{)}$$

FIGURES AND TABLES

Figure 1: A schematic of the additive gene versus additive sites models. Note that the additive sites model is approximately equivalent to the standard multiplicative fitness model.

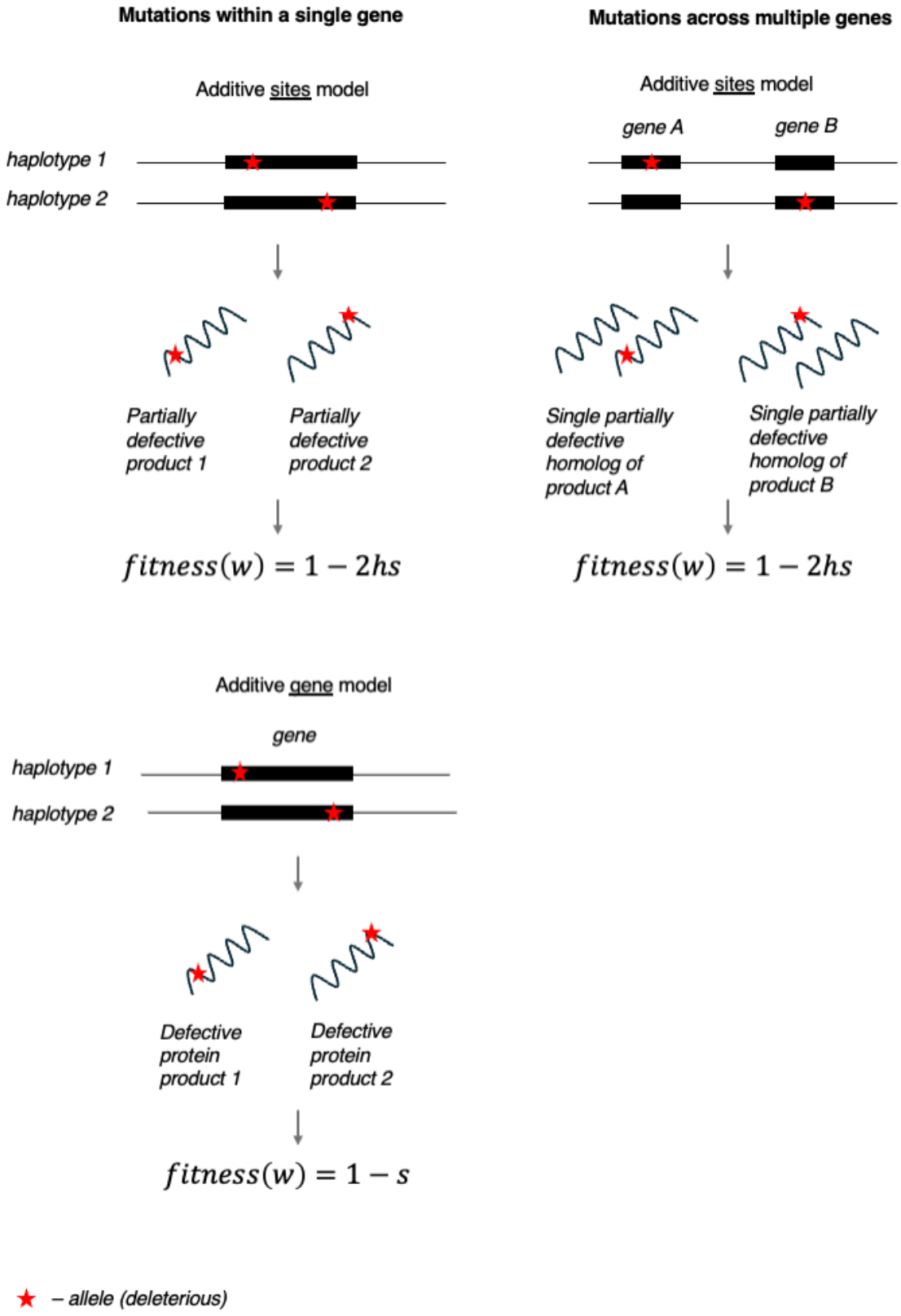

Figure 2. Linkage disequilibrium (LD) as a function of the epistasis coefficient (ϵ) for the deterministic equilibrium model with two sites, with $u/s = 0.01$ and $r = 0$. LD is measured by the correlation coefficient (R) multiplied by 1000. The results for three different dominance coefficients are shown.

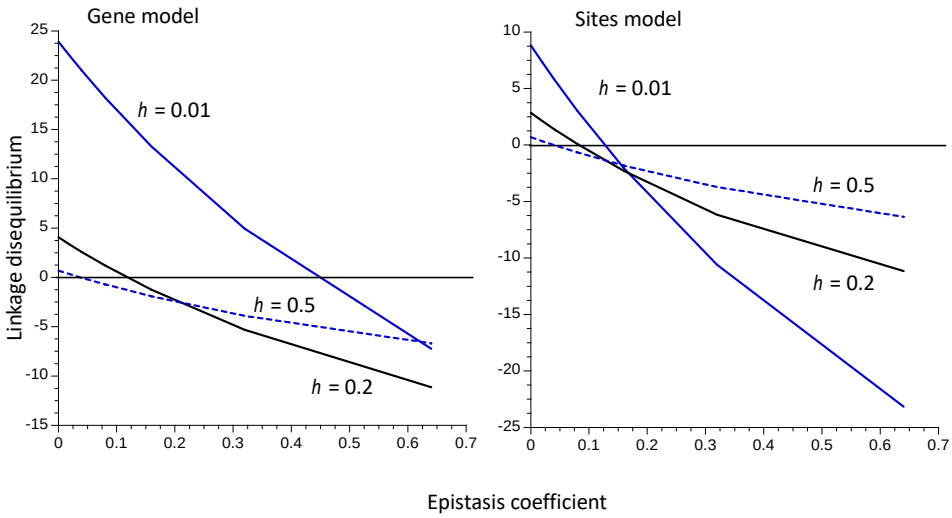

Figure 3. Mean linkage disequilibrium (LD) as a function of the epistasis coefficient (ϵ), from stochastic simulations of the two site haploid model with $r = 0$. LD is measured by the $\bar{\sigma}_d$ statistic described in the text (multiplied by 100). The results for 10^6 replicate simulations with two different values of $\gamma = 2Ns$ are shown, with $N = 1000$. The dashed curve with filled markers shows the population values of LD with no weighting towards low allele frequencies; the full curve with open markers is for doubleton variants in a sample of size 100. The dashed curve with open markers shows population values with a weight of 10 toward low frequency variants; the full curve with filled markers is for a weight of 100. See the main text for further details.

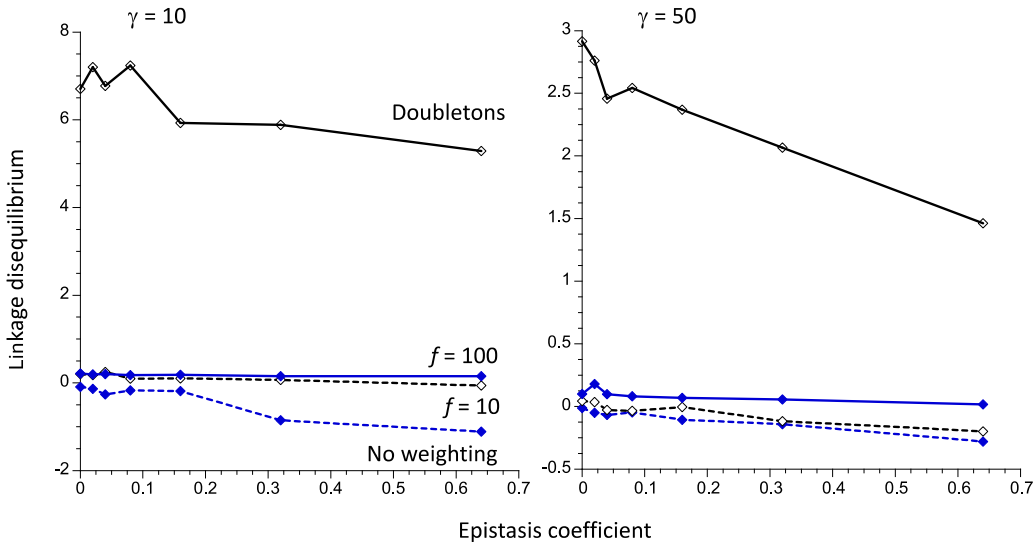

Figure 4: The means across replicate simulations of the genetic load (L) and inbreeding load (B) for the gene and sites fitness models, with varying values of the epistasis coefficient (ϵ). The selection coefficients follow a gamma distribution with $\bar{\gamma} = 100$ and shape parameter 0.3. 1000 selected sites were simulated, with $Nu = 0.005$ and $Nr = 0.01$, where N is the population size ($N = 1000$), u is the mutation rate per site/generation, and r is the rate of crossing over between adjacent sites. The error bars represent the SEs across 1000 replicate simulations.

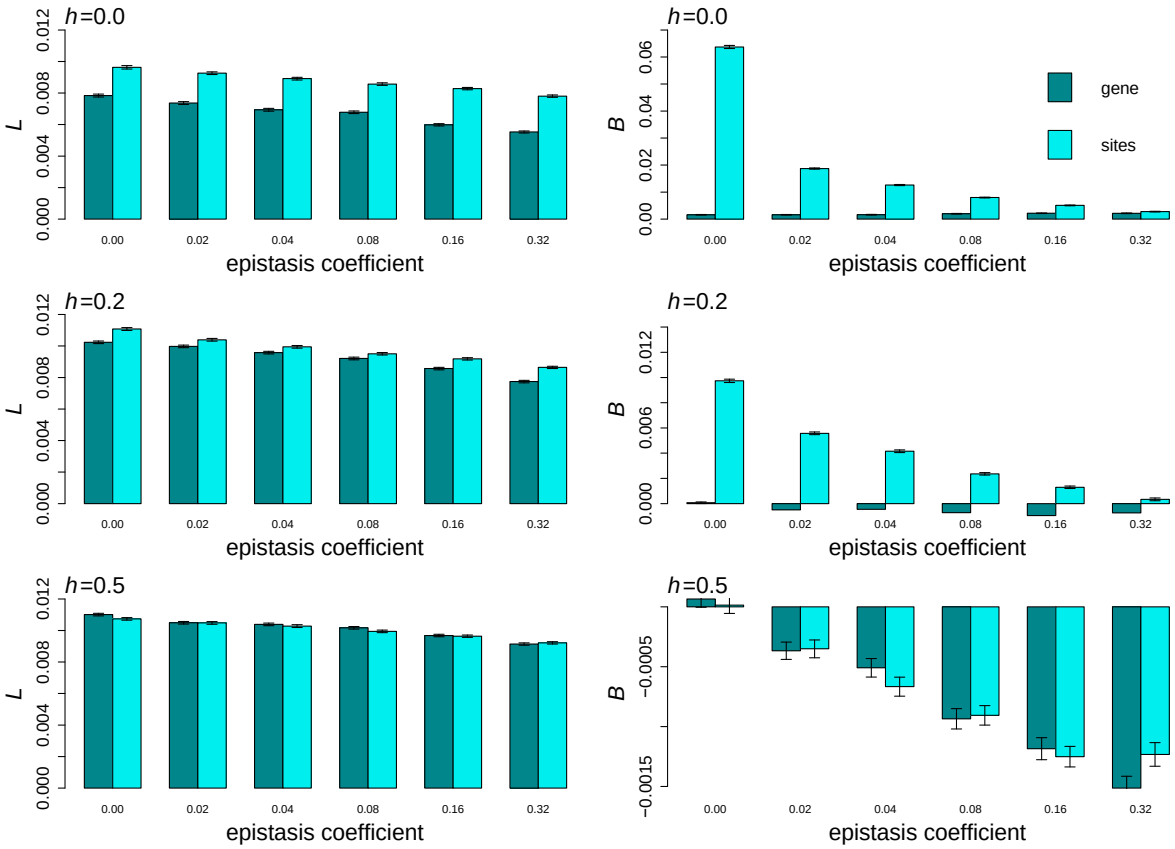

Figure 5: Statistical summaries of the $\bar{\sigma}_d$ measure of linkage disequilibrium between all pairs of sites segregating for deleterious mutations for gene and sites model with epistasis, for closely linked pairs of alleles (1-100 basepairs apart). The simulation parameters were the same as in Figure 4. The sign of the LD statisic $\bar{\sigma}_d$ was assigned either by giving a positive sign to cases with excesses over random combinations of pairs of selectively deleterious alleles or to cases with excesses of combinations of minor alleles. The error bars represent the SEs across replicates.

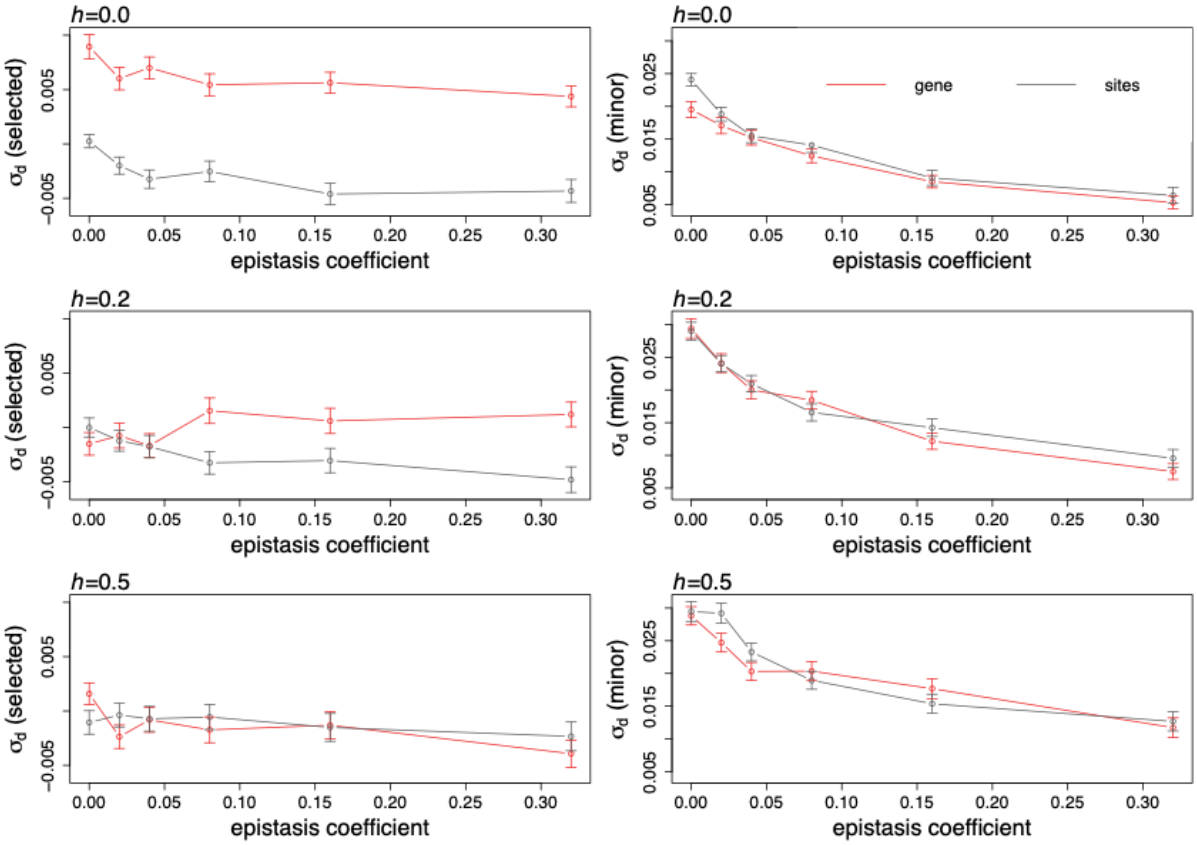

Table 1. Models of the fitness effects of two sites with a fixed selection coefficient

The "sites" model

	++	+ -	- +	- -
+ +	1	$1 - hs$	$1 - hs$	$1 - 2hs - e_1$
- +	$1 - hs$	$1 - 2hs - e_1$	$1 - s$	$1 - s - hs - e_2$
+ -	$1 - hs$	$1 - s$	$1 - 2hs - e_1$	$1 - s - hs - e_2$
- -	$1 - 2hs - e_1$	$1 - s - hs - e_2$	$1 - s - h\,s - e_2$	$1 - 2s - e_3$

$e_1 = 2\epsilon hs$; $e_2 = \epsilon(1 + h)s$; $e_3 = 2\epsilon s$.

The "gene" model

	+ +	+ -	- +	- -
+ +	1	$1 - hs$	$1 - hs$	$1 - 2hs - e_1$
- +	$1 - hs$	$1 - s$	$1 - s$	$1 - 1.5s - e_2$
+ -	$1 - hs$	$1 - s$	$1 - s$	$1 - 1.5s - e_2$
- -	$1 - 2hs - e_1$	$1 - 1.5s - e_2$	$1 - 1.5s - e_2$	$1 - 2s - e_3$

$e_1 = 2\epsilon hs$; $e_2 = 1.5\epsilon s$; $e_3 = 2\epsilon s$.

Table 2. Population genetic statistics for the gene and sites models

h=0.01

	Gene model				Sites model			
ϵ	$\bar{q}$	L	Var	B	$\bar{q}$	L	Var	B
0	0.0654	0.466	0.428	2.80	0.0866	0.454	0.372	3.87
0.02	0.0653	0.466	0.430	2.80	0.0864	0.454	0.373	3.87
0.04	0.0652	0.466	0.431	2.80	0.0862	0.454	0.375	3.87
0.08	0.0650	0.466	0.435	2.80	0.0860	0.454	0.377	3.87
0.16	0.0646	0.466	0.441	2.81	0.0854	0.454	0.383	3.87
0.32	0.0641	0.467	0.454	2.81	0.0845	0.455	0.393	3.87
0.64	0.0633	0.467	0.481	2.81	0.0832	0.456	0.415	3.87

h=0.2

ϵ	$\bar{q}$	L	Var	B	$\bar{q}$	L	Var	B
0	0.724	0.801	0.241	1.01	0.775	0.820	0.205	1.12
0.02	0.723	0.801	0.241	1.01	0.774	0.820	0.206	1.12
0.04	0.723	0.801	0.242	1.01	0.772	0.820	0.206	1.11
0.08	0.722	0.801	0.242	1.01	0.770	0.819	0.207	1.11
0.16	0.720	0.801	0.244	1.01	0.765	0.818	0.209	1.11
0.32	0.717	0.801	0.246	1.01	0.756	0.816	0.214	1.09
0.64	0.712	0.801	0.252	1.01	0.741	0.812	0.224	1.07

h=0.5

ϵ	$\bar{q}$	L	Var	B	$\bar{q}$	L	Var	B
0	0.929	0.929	0.448	0.00	0.929	0.929	0.448	0.00
0.02	0.929	0.929	0.448	0.00	0.928	0.929	0.448	0.00
0.04	0.928	0.929	0.448	0.00	0.927	0.928	0.449	0.00
0.08	0.928	0.929	0.448	0.00	0.925	0.928	0.450	0.00
0.16	0.926	0.929	0.449	0.00	0.922	0.927	0.452	0.00
0.32	0.925	0.929	0.451	0.00	0.915	0.929	0.458	0.00
0.64	0.922	0.929	0.454	0.00	0.904	0.921	0.472	0.00

The ratio of mutation rate to selection coefficient is u/s = 0.01.

The mean allele frequency $\bar{q}$, genetic load L and inbreeding load B are expressed as their ratios with respect to their single-locus deterministic values with $u >> hs$, i.e. when the frequency of A_2 is $u/(hs)$; the genetic variance (Var) is expressed as the ratio of hV/s^2 to $4\tilde{u}$.

Table 3. The load statistics under the gene and sites models for a gamma distribution of scaled selection coefficients with shape parameter 0.3 and mean $\bar{\gamma} = 2N\bar{s}$. 1000 selected sites were simulated, with Nu = 0.005 and Nr = 0.01, where N is the population size (N = 1000), u is the mutation rate per site/generation, and r is the rate of crossing over between adjacent sites. Details of how to estimate the load statistics are given in the Methods section. The means and standard errors (SEs) for 1000 replicate simulations are shown.

$\bar{\gamma} = 2$		***h*=0.0**		***h*=0.2**		***h*=0.5**	
		Gene model	**Sites model**	**Gene model**	**Sites model**	**Gene model**	**Sites model**
genetic load	mean	0.00976	0.00860	0.00978	0.00951	0.00980	0.00978
	SE	0.00008	0.00008	0.00008	0.00008	0.00008	0.00008
inbreeding load	mean	0.00001	0.00758	0.00001	0.00314	0.00000	0.00002
	SE	0.00001	0.00007	0.00001	0.00003	0.00001	0.00001
variance in fitness	mean	4.33×10^{-6}	1.15×10^{-5}	4.26×10^{-6}	4.21×10^{-6}	4.20×10^{-6}	4.15×10^{-6}
	SE	7.64×10^{-8}	2.31×10^{-7}	7.55×10^{-8}	8.27×10^{-8}	8.02×10^{-8}	6.79×10^{-8}
allele frequency	mean	0.02878	0.02997	0.03837	0.03826	0.03832	0.02772
	SE	0.00468	0.00427	0.00523	0.00530	0.00552	0.00430
$\bar{\gamma} = 20$							
genetic load	mean	0.01098	0.01063	0.01109	0.01157	0.01141	0.01121
	SE	0.00009	0.00010	0.00009	0.00010	0.00009	0.00009
inbreeding load	mean	0.00002	0.02960	-0.00003	0.00714	0.00003	-0.00002
	SE	0.00003	0.00026	0.00003	0.00007	0.00003	0.00003
variance in fitness	mean	4.89×10^{-5}	1.05×10^{-4}	4.85×10^{-5}	2.75×10^{-5}	4.93×10^{-5}	4.84×10^{-5}
	SE	7.22×10^{-7}	2.02×10^{-6}	7.24×10^{-7}	4.53×10^{-7}	7.88×10^{-7}	7.43×10^{-7}
allele frequency	mean	0.01897	0.02501	0.01200	0.01408	0.01662	0.01944
	SE	0.00386	0.00405	0.00287	0.00326	0.00354	0.00386

$\bar{\gamma} = 100$							
genetic load	mean						
		0.00768	0.00968	0.01026	0.01110	0.01089	0.01087
	SE	0.00009	0.00010	0.00008	0.00009	0.00009	0.00009
inbreeding load	mean						
		0.00158	0.06449	0.00006	0.00977	-0.00008	-0.00008
	SE	0.00010	0.00062	0.00007	0.00013	0.00006	0.00006
variance in fitness	mean						
		2.57×10^{-4}	3.78×10^{-4}	2.48×10^{-4}	1.13×10^{-4}	2.43×10^{-4}	2.45×10^{-4}
	SE	4.28×10^{-6}	8.76×10^{-6}	3.96×10^{-6}	1.87×10^{-6}	4.01×10^{-6}	3.85×10^{-6}
allele frequency	mean						
		0.00457	0.01339	0.00467	0.01580	0.01145	0.00852
	SE	0.00144	0.00311	0.00164	0.00351	0.00301	0.00252
$\bar{\gamma} = 1000$							
genetic load	mean						
		0.00563	0.00851	0.00828	0.01014	0.01061	0.01066
	SE	0.00008	0.00013	0.00008	0.00008	0.00009	0.00009
inbreeding load	mean						
		0.01992	0.25320	0.00359	0.01176	-0.00005	0.00035
	SE	0.00059	0.00332	0.00031	0.00035	0.00027	0.00026
variance in fitness	mean						
		0.00398	0.00754	0.00264	0.00102	0.00345	0.00347
	SE	0.00020	0.00045	0.00011	0.00002	0.00013	0.00012
allele frequency	mean						
		0.00018	0.01089	0.00108	0.00917	0.00684	0.00283
	SE	0.00004	0.00289	0.00053	0.00278	0.00234	0.00145

Table 4. Statistical summaries of linkage disequilibrium under the gene and sites models for a gamma distribution of scaled selection coefficients with shape parameter 0.3 and a mean of $\bar{\gamma} = 2N\bar{s}$. 1000 selected sites were simulated, with Nu = 0.005 and Nr = 0.01, where N is the population size (N = 1000), u is the mutation rate per site/generation, and r is the rate of crossing over between adjacent sites. $\bar{D}$ and $\bar{\sigma}_d$ were calculated for alleles that were less than 100 basepairs apart. The signs of $\bar{D}$ and $\bar{\sigma}_d$ were assigned either by giving a positive sign to cases with excesses over random combinations of pairs of selectively deleterious alleles, or to cases with excesses of combinations of minor alleles (see the Methods section). The means and standard errors (SEs) for 1000 replicate simulations are shown.

		***h*=0.0**		***h*=0.2**		***h*=0.5**	
selected alleles		**gene**	**sites**	**gene**	**sites**	**gene**	**sites**
$\bar{\gamma} = \mathbf{2}$							
$\bar{D}$	mean	-0.00003	-0.00001	-0.00009	-0.00009	0.00002	-0.00008
	SE	0.00005	0.00005	0.00005	0.00005	0.00005	0.00005
$\bar{\sigma}_d$	mean	-0.00012	-0.00009	-0.00099	-0.00089	0.00030	-0.00069
	SE	0.00056	0.00049	0.00056	0.00051	0.00058	0.00053
$\bar{\gamma} = \mathbf{20}$							
$\bar{D}$	mean	-0.00006	0.00001	0.00002	-0.00006	-0.00012	-0.00011
	SE	0.00005	0.00004	0.00006	0.00005	0.00005	0.00005
$\bar{\sigma}_d$	mean	-0.00061	0.00013	0.00025	-0.00046	-0.00145	-0.00135
	SE	0.00076	0.00052	0.00077	0.00065	0.00077	0.00075
$\bar{\gamma} = \mathbf{100}$							
$\bar{D}$	mean	0.00014	-0.00007	0.00015	-0.00007	0.00004	-0.00005
	SE	0.00005	0.00004	0.00007	0.00006	0.00007	0.00007
$\bar{\sigma}_d$	mean	0.00540	-0.00082	0.00241	-0.00088	0.00089	-0.00054
	SE	0.00100	0.00059	0.00103	0.00084	0.00103	0.00107
$\bar{\gamma} = \mathbf{1000}$							
$\bar{D}$	mean	0.00017	-0.00003	0.00058	-0.00012	0.00009	-0.00015
	SE	0.00003	0.00004	0.00009	0.00010	0.00013	0.00012
$\bar{\sigma}_d$	mean	0.00941	0.00017	0.01376	-0.00129	0.00197	-0.00217
	SE	0.00146	0.00075	0.00172	0.00147	0.00191	0.00176

minor alleles							
$\bar{\gamma} = \mathbf{2}$							
$\bar{D}$	mean	0.00404	0.00557	0.00378	0.00478	0.00390	0.00387
	SE	0.00011	0.00014	0.00011	0.00012	0.00011	0.00010
$\bar{\sigma}_d$	mean	0.04795	0.05656	0.04448	0.05202	0.04559	0.04609
	SE	0.00124	0.00130	0.00125	0.00124	0.00119	0.00119
$\bar{\gamma} = \mathbf{20}$							
$\bar{D}$	mean	0.00235	0.00295	0.00247	0.00284	0.00252	0.00239
	SE	0.00009	0.00009	0.00009	0.00010	0.00010	0.00010
$\bar{\sigma}_d$	mean	0.03295	0.03633	0.03455	0.03771	0.03490	0.03328
	SE	0.00122	0.00103	0.00119	0.00124	0.00126	0.00126
$\bar{\gamma} = \mathbf{100}$							
$\bar{D}$	mean	0.00106	0.00175	0.00185	0.00206	0.00208	0.00194
	SE	0.00008	0.00007	0.00010	0.00010	0.00011	0.00011
$\bar{\sigma}_d$	mean	0.01911	0.02636	0.02730	0.03011	0.03042	0.02872
	SE	0.00120	0.00094	0.00138	0.00129	0.00148	0.00145
$\bar{\gamma} = \mathbf{1000}$							
$\bar{D}$	mean	0.00017	0.00086	0.00077	0.00164	0.00187	0.00182
	SE	0.00003	0.00005	0.00009	0.00013	0.00015	0.00016
$\bar{\sigma}_d$	mean	0.00937	0.01643	0.01690	0.02483	0.02885	0.02630
	SE	0.00146	0.00088	0.00173	0.00178	0.00209	0.00210

Table 5: Statistical summaries of linkage disequilibrium under the gene and sites models in human-like populations, using a gamma distribution of scaled selection coefficients with shape parameter 0.23 and mean $\bar{\gamma} = 2N\bar{s}$. 2000 selected sites were simulated, with Nu = 0.00025 and Nr = 0.0002, where N is the population size (N = 1000), u is the mutation rate per site/generation, and r is the rate of crossing over between adjacent sites. The signs of $\bar{D}$ and $\bar{\sigma}_d$ were assigned either by giving a positive sign to cases with excesses over random combinations of pairs of selectively deleterious alleles, or to cases with excesses of combinations of minor alleles. The means and standard errors (SEs) for 1000 replicate simulations are shown.

		***h*=0.0**		***h*=0.2**		***h*=0.5**	
selected alleles		**gene**	**sites**	**gene**	**sites**	**gene**	**sites**
$\bar{\gamma} = \mathbf{850}$							
$\bar{D}$	mean	0.00015	0.00031	0.00101	0.00030	-0.00096	-0.00105
	SE	0.00045	0.00035	0.00062	0.00066	0.00063	0.00074
$\bar{\sigma}_d$	mean	0.01130	0.00650	0.01587	0.00345	-0.00706	-0.00345
	SE	0.00472	0.00353	0.00591	0.00547	0.00525	0.00600
$\bar{\sigma}_d^1$	mean	0.46777	0.25534	0.61505	0.14143	-0.12457	0.20649
	SE	0.17992	0.08900	0.21642	0.11249	0.09229	0.20202
minor alleles $\bar{\gamma} = \mathbf{850}$							
$\bar{D}$	mean	0.00123	0.00137	0.00319	0.00319	0.00244	0.00285
	SE	0.00046	0.00036	0.00063	0.00065	0.00071	0.00078
$\bar{\sigma}_d$	mean	0.01924	0.02183	0.03972	0.03556	0.02356	0.03175
	SE	0.00473	0.00373	0.00601	0.00548	0.00560	0.00625
$\bar{\sigma}_d^1$	mean	0.54411	0.51751	0.94606	0.55386	0.20582	0.60968
	SE	0.17966	0.09122	0.21699	0.11278	0.09449	0.20269

Table 6. The values of the load statistics obtained from Table 3 when scaled up to the whole *D. melanogaster* genome.

	***h* = 0**		***h* = 0.2**		***h* = 0.5**	
Variable	**Gene model**	**Sites model**	**Gene model**	**Sites model**	**Gene model**	**Sites model**
L_T	0.0752	0.112	0.109	0.132	0.138	0.139
B_T	0.279	3.54	0.0503	0.165	0	0
V_T	4.76×10^{-5}	8.34×10^{-5}	2.93×10^{-5}	1.08×10^{-5}	3.59×10^{-5}	3.60×10^{-5}
σ_T	0.00690	0.00913	0.00541	0.00329	0.00628	0.00603