

The interpretation of selection coefficients

N. H. Barton and M. R. Servedio

Supplementary Information

Here we further describe details of the mutliocus notation, based upon its development in Barton and Turelli 1991 and Kirkpatrick et al. (2002).

Defining associations

Having chosen the reference point, we define the association amongst the set of genes \mathbb{U} by:

$$D_{\mathbb{U}} = E [\zeta_{\mathbb{U}}] \text{ where } \zeta_{\mathbb{U}} \equiv \prod_{i \in \mathbb{U}} \zeta_i \text{ and } \zeta_i \equiv (X_i - \wp_i) \quad (1)$$

where the expectation is over the population (which is assumed large). Suppose that the set of all positions is Ω , with $|\Omega| = n$ members, each of which can have value $X_i = 0$ or 1 . Then, there are $2^n - 1$ degrees of freedom, which can be represented by associations amongst all possible subsets of Ω , $D_{\mathbb{U}}$ for $\mathbb{U} \subseteq \Omega$. By convention, $D_{\emptyset} = 1$, where \emptyset is the empty set.

In the recursion for the effects of selection (Eqs. XXX), sets with repeated elements appear. Assuming two alleles per locus, labelled $X_i = 0$ or 1 , these can be expressed in terms of lower-order moments with distinct elements, so that we have a closed set of $2^n - 1$ equations:

$$D_{i \cup \mathbb{U}} = \wp_i (1 - \wp_i) D_{\mathbb{U}} - (2\wp_i - 1) D_{i \cup \mathbb{U}} \quad (2)$$

Changes in reference point

Associations depend on the reference point, \wp_i , which will usually be an allele frequency. However, in all but the simplest cases, allele frequencies will differ between contexts (for example, between sexes or locations), and will change through time. Thus, we must choose the most appropriate reference point, and can change from reference point \wp_i to \wp'_i using the following formula:

$$D'_{\mathbb{U}} = \prod_{i \in \mathbb{U}} \left(X_i - \wp_i - \left(\wp'_i - \wp_i \right) \right) = \sum_{V \subseteq \mathbb{U}} D_V (-\Delta \wp)_{\mathbb{U} \setminus V} \text{ where } (-\Delta \wp)_{\mathbb{U} \setminus V} \equiv \prod_{i \in \mathbb{U} \setminus V} (-\Delta_i) \text{ and } \Delta_i \equiv \wp'_i - \wp_i \quad (3)$$

and $U \setminus V$ denotes U , with the elements of V removed. For example, $D'_{ij} = D_{ij} - \Delta_i D_j - D_i \Delta_j + \Delta_i \Delta_j$. If the original reference point was equal to the original allele frequencies, $\varphi_i = p_i = E[X_i]$, then $D_i = D_j = 0$, and $D'_{ij} = D_{ij} + \Delta_i \Delta_j$; the change to a new reference point has generated an association $\Delta_i \Delta_j$.

Recombination

To complete the sexual life cycle, we represent recombination as a linear mixture of moments, using the BT notation. Defining $r_{S,T}$ as the probability that a set $U = ST$ is made up of S from the female parent, and T from the male parent, we have:

$$D_U = \sum_{ST=U} r_{S,T} D_{S,T} \quad (4)$$

This is a special case of transmission of genes from one context to another, without any genes being created or destroyed; conservative migration (i.e., migration that does not change overall allele frequencies) can be treated in a similar way. If transmission depends on genotype, then we can write $r_{S,T}$ as a polynomial function of the ζ ; this can be seen as a combination of selection with transmission, and includes both as special cases (Barton, 1995; Kirkpatrick et al., 2002).

Multiplicative selection on haploids

Multiplicative selection does not generate linkage disequilibria; yet, because we define fitness as a sum rather than a product, it does involve coefficients a_U for all sets of selected loci. Here, we show that the associations generated by multiplicative selection, relative to the initial allele frequencies, vanish when measured against the new allele frequencies. Let relative fitness be:

$$\frac{W}{\bar{W}} = 1 + \prod_{i \in W} (1 + \beta_i \zeta_i) - E \left[\prod_{i \in W} (1 + \beta_i \zeta_i) \right] = 1 + \sum_{U \subseteq W} \beta_U (\zeta_U - D_U) \quad (5)$$

This can be written in the form of Eq. XXX by identifying $a_U = \beta_U$, where we define $\beta_U \equiv \prod_{i \in U} \beta_i$. At linkage equilibrium, $D'_U = a_U D_U = \beta_U p q_U$. However, there will also be changes in allele frequency, $\Delta p_i = \beta_i p q_i$, and so we must change the reference point to match the new allele frequencies. Substituting into Eq. XXX, and assuming linkage equilibrium, we see that all the terms cancel, so that the population stays at linkage equilibrium:

$$D''_U = \sum_{ST=U} (-\Delta p)_S D'_T = \sum_{ST=U} (\prod_{i \in S} (-\beta_i p q_i)) (\prod_{i \in T} \beta_i p q_i) = 0 \quad (6)$$

This proliferation of terms that all disappear suggests that the polynomial form of Eq. 1 is not the best way to represent the population, when selection is

multiplicative. Bürger (1991) and Turelli and Barton (1994) develop an alternative representation, in which linkage disequilibria are represented by multilocus cumulants, rather than moments, and selection is represented by the gradient of mean fitness with respect to these cumulants. Thus, we define $\kappa_i = D_i$, $\kappa_{ij} = D_{ij} - D_i D_j$, $\kappa_{ijk} = D_{ijk} - D_{ij} D_k - D_{ik} D_j - D_{jk} D_i - D_i D_j D_k$, and so on; the selection gradients are defined as $\mathcal{L}_U = \partial \log(\bar{W}) / \partial \kappa_U$. Now, provided the population is at linkage equilibrium, multiplicative selection only generates first-order selection gradients, \mathcal{L}_i , and only changes the first-order cumulants, κ_i .

This derivation assumes that all haploid individuals are subject to the same form of multiplicative selection. If selection acts differently on the two sexes (or more generally, acts differently on subpopulations that will eventually combine to mate at random), then linkage disequilibria will be generated by mixing.

One can find the effect of small deviations from multiplicative fitness, using the QLE approximation, by letting $a_U = \beta_U + \alpha_U$, and finding D_U to leading order in α . This leads to complex expressions; in particular, weak epistasis α_U generates both higher- and lower-order associations, so that there is no longer a direct relationship between the epistatic coefficients, and the associations that they produce.

Quasi-linkage equilibrium with strong directional selection

We can extend Hastings' (1985) and Barton's (1986) argument, by allowing strong directional selection. Now, we must take care of the distinction between additive and multiplicative selection. The null model must be multiplicative selection, since that generates no linkage disequilibrium; this contrasts with the case when selection is weak, when there is no significant distinction between additive and multiplicative selection. This is at first glance puzzling: because the a_U for $|U| > 1$ measure deviations from *additivity*, there are non-zero higher order coefficients even though fitnesses are multiplicative, and so should generate linkage disequilibria. This is the case when associations are measured with reference point at the *initial* allele frequencies. However, these associations cancel precisely when we change the reference point to the new allele frequencies after selection (the terms that cause the cancellation are trivially small with weak selection), so that, provided that it does not differ between the sexes, multiplicative selection indeed generates no linkage disequilibria (Appendix, Eq. 6). Thus, neither the coefficients a_U nor \hat{a}_U can be interpreted as generating associations between the set of genes U . If we now superimpose weak deviations from multiplicative fitness, it is no longer true that selection for a set of genes U only generates associations D_U amongst that set, as would be the case under weak selection: additional associations are generated, and there is no one-to-one mapping between the coefficients of epistatic selection and the linkage disequilibria.