Stable linkage disequilibrium owing to sexual antagonism
Francisco Úbeda1,*, David Haig2 and Manus M. Patten2

1Department of Ecology and Evolutionary Biology, University of Tennessee, Knoxville, TN 37996, USA
2Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA 02138, USA

Linkage disequilibrium (LD) is an association between genetic loci that is typically transient. Here, we identify a previously overlooked cause of stable LD that may be pervasive: sexual antagonism. This form of selection produces unequal allele frequencies in males and females each generation, which upon admixture at fertilization give rise to an excess of haplotypes that couple male-beneficial with male-beneficial and female-beneficial with female-beneficial alleles. Under sexual antagonism, LD is obtained for all recombination frequencies in the absence of epistasis. The extent of LD is highest at low recombination and for stronger selection. We provide a partition of the total LD into distinct components and compare our result for sexual antagonism with Li and Nee’s model of LD owing to population subdivision. Given the frequent observation of sexually antagonistic selection in natural populations and the number of traits that are often involved, these results suggest a major contribution of sexual antagonism to genomic structure.

Keywords: intralocus conflict; population genetics; sexual conflict; population structure; two-locus model

1. INTRODUCTION
Linkage disequilibrium (LD) is a covariance between genetic loci that develops for various reasons including epistatic selection, drift, mutation and population structure [1]. LD may be transient or permanent, depending on which of these causes is responsible. Most instances of LD are transient because recombination randomly assigns alleles into gametes each generation. Unless the cause of LD is persistent, the association between loci is expected to decline geometrically over successive generations [2].

There are three known causes of stable LD. The first is epistatic interaction among loci [2–4]. Under this scenario, the fitness of an allele or genotype at one locus depends on the state of the second locus. This favours particular allelic combinations, which generate a statistical association between loci. Provided that selection maintains allelic variation at both loci, epistasis ensures that coupling between loci persists. Second, if fitness is determined multiplicatively across two overdominant loci that are sufficiently tightly linked, then a polymorphic equilibrium is necessarily accompanied by LD [5]. Third, permanent LD is possible in a metapopulation owing to allele frequency differences between subpopulations [6]. In fact, LD is expected in subdivided populations because linkage equilibrium requires the restrictive condition of equal allele frequencies in all subpopulations at one or both loci. LD persists provided that genetic variation is maintained in at least one subpopulation (or different combinations of alleles are fixed in different subpopulations).

In this paper, we identify sexual antagonism as a previously overlooked cause of permanent LD. Sexually antagonistic selection occurs when males and females have different phenotypic optima [7,8] and causes genotype frequencies in the male and female population to differ from each other after selection. Theoretically, sexual antagonism can maintain polymorphism [9]; empirically, sexually antagonistic selection is common [10] and sexually antagonistic fitness variation is prevalent in natural and laboratory populations [11–17]. Thus, for sexual antagonism, not only are the elements to produce stable LD in place, they are in abundance.

2. MODEL AND RESULTS
We consider a model of two genetic loci with two alleles segregating at each: \(A_1\) and \(A_2\), \(B_1\) and \(B_2\). Let \(x_i\) and \(y_i\) be the frequencies of the \(i\)th and \(i\)th haplotypes in eggs and sperm, respectively, such that: \(x_1, y_1\) are the frequencies of the \(A_1B_1\) haplotype; \(x_2, y_2\) the frequencies of the \(A_1B_2\) haplotype; \(x_3, y_3\) the frequencies of the \(A_2B_1\) haplotype; and \(x_4, y_4\) the frequencies of the \(A_2B_2\) haplotype. We arrange the haplotype frequencies in eggs and sperm in vectors \(x\) and \(y\), respectively (henceforth, lower-case bold letters represent vectors). Let \(p_x\) and \(p_y\) be the frequency of the \(A_1\) allele in eggs and sperm and \(q_x\) and \(q_y\) be the frequency of the \(B_1\) allele in eggs and sperm, respectively, calculated as:

\[
p_x = \frac{1}{2} \left( x_1 + x_2 \right) \quad \text{and} \quad p_y = \frac{1}{2} \left( y_1 + y_2 \right) \quad \text{with} \quad \xi \in \{x, y\}.
\]

Let genotypes \(A_1A_1, A_1A_2\) and \(A_2A_2\) have fitness \(u_{11} = 1 - s, u_{12} = 1 - h_1 l_1, u_{22} = 1\) in females, and \(u_{11} = 1, u_{22} = 1 - h_2 m_2, u_{12} = 1 - s m_1\) in males, where \(0 \leq s \leq 1\). Fitness at the \(B\) locus, \(v_x\) with \(x \in \{f, m\}\), is parameterized in an analogous manner (table 1). For simplicity we assume that the allelic effects are additive at both loci \((h = 1/2, m = 1/2)\), which guarantees opposing directional selection in the two sexes. We assume that the fitness of a
zygote results from the product of the fitness values at each locus. Thus, \( w_{xy} = u_k v_{kl} \) where \( k \) and \( l \in \{1, 2, 3\} \), is the fitness of a zygote of sex \( \chi \) that results from the union of the \( \theta \)th egg and \( \eta \)th sperm haplotype (table 2). This assumption eliminates multiplicative epistasis within sexes by rendering the contributions of both loci independent from one another [18]. We arrange the fitness values for females and males in matrices \( W_f = (w_{i,j}) \) and \( W_m = (w_{im}) \). Given our assumptions, the number of different fitness values reduces to six (table 2b).

The frequency of haplotype \( i \) in eggs and sperm after one generation is

\[
\pi_i x_i' = \frac{1}{2} \left[ x_i(W_f y)_j + y_j(W_f^T x)_i \right] - e_i r \left( 1 - \frac{1}{2} r \right) D_t \quad (2.1a)
\]

and

\[
\pi_m y_i' = \frac{1}{2} \left[ x_i(W_m y)_j + y_j(W_m^T x)_i \right] - e_i r \left( 1 - \frac{1}{2} s_m \right) D_t \quad (2.1b)
\]

where \( \pi_f \) and \( \pi_m \) are the mean fitness of females and males, respectively (defined as \( w_{i} = x^T W_{i} y \)), \( r \) is the recombination rate between loci, \( e_i \) provides the sign of the LD factor (\( e_i \) is equal to 1 when \( i = 1, 4 \) and \(-1 \) when \( i = 2, 3 \)) and \( D_t \) is the LD in the population.

When the selection regime is the same in both sexes, gametic haplotype frequencies are sex-independent \( (x = y) \) and \( D_t \) takes the familiar form \( D_t = x_1x_4 - x_2x_3 \). However, with sex-specific selection, \( D_t \) must be calculated as

\[
D_t = \frac{1}{2} [x_1y_4 + x_4y_1 - x_2y_3 - x_3y_2]. \quad (2.2)
\]

When selection differs between the sexes, gametic haplotype frequencies depend on sex-specific allele frequencies at the two loci and the LD in eggs and sperm:

\[
x_1 = p_x q_x + D_x, \quad y_1 = p_y q_y + D_y;
\]

\[
x_2 = p_x (1 - q_x) - D_x, \quad y_2 = p_y (1 - q_y) - D_y;
\]

\[
x_3 = (1 - p_x) q_x - D_x, \quad y_3 = (1 - p_y) q_y - D_y;
\]

\[
x_4 = (1 - p_x) (1 - q_x) + D_x, \quad y_4 = (1 - p_y) (1 - q_y) + D_y.
\]

Substituting equation (2.3) into equation (2.2) and simplifying yield

\[
D_t = \frac{1}{2} (D_x + D_y) + \frac{1}{2} (p_x - p_y)(q_x - q_y), \quad (2.5)
\]

where the second term on the right is equal to twice the covariance between \( p \) and \( q \); \( 2 \text{ Cov}(p, q) \); thus,

\[
D_t = \frac{1}{2} (D_x + D_y) + 2 \text{ Cov}(p, q), \quad (2.6)
\]

which provides a neat partition of the total LD generated by sexually antagonistic selection into two components, namely the average LD in gametes and the covariance between the allele frequencies at each locus.

Equation (2.6) is similar in form to Nei & Li’s [19] eqn (5) \( (D_T = 1/2 \ (D_x + D_y + \text{Cov}(p, q)) \) for LD in a subdivided population. This formal similarity allows an interesting interpretation of our result. The two sexes, with their differing ecologies, physiologies and selection regimes [20], are effectively distinct subpopulations that freely exchange migrants from one generation to the next. The appendix explores the close analogy between our model and the model of Li & Nei ([6]; see also [19]).

We are interested in LD at a stable polymorphic equilibrium \( (x, y) \), i.e. \( D_t = 1/2 \left[ x_1y_4 + x_4y_1 - x_2y_3 - x_3y_2 \right] \). Calculating equilibrium haplotype frequencies analytically, even with our simplifying assumptions, proved too complex. Instead, we carried out numerical analysis of the recursions in equation (2.1) to determine the equilibrium points for different coefficients of selection \( s_t \) and \( s_m \). When the equilibrium was polymorphic, we calculated the LD value for each pair \( (s_t, s_m) \) and for different values of recombination, \( r \).

Figure 2 presents LD at polymorphic equilibria broken down into its components: \( 1/2 (D_x + D_y) \) and \( 2 \text{ Cov}(p, q) \). LD is present at equilibrium for any value of \( r \) including free recombination \( (r = 1/2) \) (figure 2c). While the LD in gametes is absent at equilibrium when there is free recombination (figure 2a), the covariance between allele frequencies remains positive regardless of the recombination rate (figure 2b). Therefore, \( D_t \) remains positive.

LD increases with decreasing \( r \) (figure 2c). The covariance between allele frequencies, however, is less sensitive to \( r \) than the LD in gametes (figure 2a,b). For low \( r \), the LD in gametes is the major contributor to \( D_t \), whereas the covariance between allele frequencies becomes the major contributor for high \( r \). The transition from \( 1/2 (D_x + D_y) \) being the major contributor to \( 2 \text{ Cov}(p, q) \) being the major contributor happens in the range of values between \( r = 0.1 \) and \( r = 0.3 \) (figure 3).

Finally, LD increases with strength of selection (figures 2 and 4c). This is not always the case for both of its components. It is always true for the covariance between allele frequencies (figure 4b). It is also true for

---

<table>
<thead>
<tr>
<th>( A_1A_1 )</th>
<th>( A_1A_2 )</th>
<th>( A_2A_2 )</th>
<th>( B_1B_1 )</th>
<th>( B_1B_2 )</th>
<th>( B_2B_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>males</td>
<td>1</td>
<td>1 - 1/2 ( s_m )</td>
<td>1 - ( s_m )</td>
<td>1</td>
<td>1 - 1/2 ( s_m )</td>
</tr>
<tr>
<td>females</td>
<td>1 - ( s_t )</td>
<td>1 - 1/2 ( s_t )</td>
<td>1</td>
<td>1 - ( s_t )</td>
<td>1 - 1/2 ( s_t )</td>
</tr>
</tbody>
</table>
the LD in gametes when the recombination rate is high (\( r > 0.1 \) in the symmetric example provided in figure 4a). When the recombination rate is low (\( r < 0.01 \) in the symmetric example provided in figure 4a), however, LD in gametes initially increases with strength of selection but then decreases at higher selection strengths (figure 4a).

### 3. DISCUSSION

The ultimate source of LD in our model is admixture between two gene pools with differing allele frequencies. Sexual antagonism results in a higher frequency of male beneficial alleles in sperm and of female beneficial alleles in eggs. Fertilization admixes these distinct gene pools and thereby generates LD in zygotes while homogenizing allele frequencies between the sexes (figure 1a). Whereas one-time admixture between geographically diverged populations produces transient LD that declines through subsequent generations, in our model, admixture between divergent subpopulations is recurrent because sexual antagonism alters the allele frequencies in opposing directions every generation. Therefore, permanent LD is maintained.

Equation (2.6) invites some further interpretation. This equation partitions \( \Delta \) into two components: 2 \( \text{Cov}(p_0q_0) \) represents the contribution to LD from sex differences in allele frequencies arising from sexually antagonistic selection in the immediately previous generation, whereas \( 1/2(D_h + D_r) \) represents accumulated LD from earlier generations. LD in gametes accumulates because of the correlated history of alleles at the two loci. When there is free recombination between maternally inherited and paternally inherited haplotypes, there is no correlation in the histories of the alleles at the two loci (whether they were present in the same or different sex) beyond the immediately preceding generation; further, the covariance between loci produced by admixture is erased by one round of free recombination; thus, \( D_h, D_r \) are expected to be zero at equilibrium. When recombination is less than 50 per cent, the histories of the alleles at the two loci in earlier generations are correlated and some of the LD produced by admixture in zygotes persists into the gametes, both of which make \( D_h, D_r \neq 0 \) at equilibrium.

Sex-specific selection does not directly produce LD within a sex. The multiplicative fitness assumption precludes this. For instance, a population of zygotes in linkage equilibrium that then undergoes sexually antagonistic selection produces gametes with zero LD. Sexual antagonism is, however, indirectly responsible for the LD that results after admixture by virtue of its effect on the second term of equation (2.6). This term results from admixing sperm and egg gene pools, which are made unequal by sexual antagonism.

Selection that does not protect polymorphism cannot stabilize LD. Other selective causes of permanent LD considered in §1 (epistatic interactions, multiplicative fitness at two overdominant loci) require special—perhaps extraordinary—circumstances. For instance, fitness variance owing to epistatic interactions exists, but whether epistasis is responsible for the maintenance of this variation is unknown. In order for two overdominant loci to maintain LD, the recombination rate must be less than the marginal segregation load of the population [5], which for selection coefficients under 50 per cent requires that the recombination rate be less than approximately 0.0625. Allele frequency differences among subpopulations are degraded through time by migration unless there is subpopulation-specific selection. The subdivision of populations into two sexes and sexually antagonistic selection are both common [10]. Therefore, the prerequisites for permanent LD under sexual antagonism are present in many natural populations.

However, two theoretical considerations may lessen the occurrence of LD caused by sexual antagonism. The first is the possibility for conflict resolution—i.e. sexual dimorphism [21]. Should a sexually antagonistic locus evolve sex-limited expression, the allele that is favoured in the expressing sex is expected to fix, thus eliminating variation and precluding LD. The time scale over which intralocus conflict operates—and therefore the likelihood of developing permanent LD—is currently unknown: if conflicts are constantly arising but then are quickly (on an evolutionary time scale) resolved by the evolution of dimorphism, the build-up of LD should be minor; if
polymorphisms are maintained for longer periods, then LD has more time to develop and should be more widespread. Dating the age of alleles at sexually antagonistic loci could potentially shed light on this question. Such loci have recently been identified by Innocenti & Morrow [22]. The second consideration draws on the work of...
Turelli & Barton [23]. They showed that at most two loci could remain polymorphic owing to sexually antagonistic selection. Their model makes assumptions that are relaxed in our model. Specifically, Turelli & Barton [23] considered a model in which loosely linked loci determined a single trait subject to weak sexually antagonistic selection. Our model is agnostic about the number of traits subject to sexually antagonistic selection, permits any value of recombination and considers strong as well as weak selection. Future analyses with more than two loci should explore the potential to maintain large amounts of LD.

LD will be created between any two loci that are polymorphic owing to sexually antagonistic selection. Their model makes assumptions that are relaxed in our model. Specifically, Turelli & Barton [23] considered a model in which loosely linked loci determined a single trait subject to weak sexually antagonistic selection. Our model is agnostic about the number of traits subject to sexually antagonistic selection, permits any value of recombination and considers strong as well as weak selection. Future analyses with more than two loci should explore the potential to maintain large amounts of LD.

The division of the population into two sexes that are subject to sex-specific selection is just one way that a gene pool could be subdivided into groups that undergo differential selection. Regular admixture between such subgroups will generate persistent LD [6]. As an example, suppose that a population consists of a mixture of vegetarians and meat-eaters but the two groups sometimes intermarry and children have some independence in the diet they choose to adopt. If the different diets result in differential selection and allele frequency differences between reproductive vegetarians and reproductive meat-eaters, then admixture of these gene pools in their offspring will be a source of LD. This LD arises because of a correlation in the selective forces acting at different loci in the two subpopulations. This hypothetical example suggests that the existence of differential selection at two loci can be a source of persistent LD whenever the selective ‘environments’ at two loci are correlated.

Figure 2. Linkage disequilibrium and its components for various recombination frequencies and selective strengths. LD is represented by the area of the circle at any given point \((s_m, s_f)\) in the parameter space. The axes of each plot are the selection coefficient in males \(s_m\) (x-axis) and females \(s_f\) (y-axis). Recombination rate, \(r\), differs for each column, increasing from left to right. (a) Average LD in gametes \(1/2 (D_x + D_y)\). (b) Twice covariance between allele frequencies, \(2 \text{Cov}(p, q)\). (c) Total LD, \(D_t\), in zygotes.
Figure 3. Contribution to the total LD of the covariance between allele frequencies component relative to the LD in gametes component. The area of each circle represents the coefficient $4 \text{Cov}(p,q)/(D_x + D_y)$. Filled circles represent the case in which $1/2 (D_x + D_y) > 2 \text{Cov}(p,q)$, while empty circles represent the case in which $1/2 (D_x + D_y) < 2 \text{Cov}(p,q)$. The axes of each plot are the selection coefficient in males $s_m$ (x-axis) and females $s_f$ (y-axis). Recombination rate, $r$, differs for each column, increasing from left to right. (a) $r = 0.1$; (b) $r = 0.2$; (c) $r = 0.3$.

Figure 4. Linkage disequilibrium and its components as a function of selection strength for the particular case in which $s_m = s_f$. Each line represents the stable LD for different recombination rate values, indicated next to the line to the right of each figure. The axes in each plot are the selection coefficient $s_m = s_f$ (x-axis) and the LD or its component (y-axis). The first row (a) corresponds to the average LD in gametes $1/2 (D_x + D_y)$. The second row (b) corresponds to the covariance between allele frequencies. The third row (c) corresponds to the total LD, $D_t$, in zygotes.
We thank R. C. Lewontin for comments on an early draft of this paper, and Richard Harrison and two anonymous referees for valuable comments.

**APPENDIX A**

The discrepancy between our theoretical value of total LD, \(D_L = 1/2 \left(D_x + D_y\right) + 2 \text{Cov}(p,q)\), and Nei & Li’s [19] value, \(D_T = 1/2 \left(D_x + D_y\right) + \text{Cov}(p,q)\), arises from the way we define LD. Standard two-locus models that lack sex differences obscure the fact that there are two possible choices. One interpretation is that LD is the quantity that must be added or subtracted empirically and to the definition set out by the originators of the term:

The two interpretations of LD, \(D_T\) and \(D_0\), differ by the magnitude of \(\text{Cov}(p,q)\) but essentially capture the same phenomenon. Nei & Li’s [19] interpretation sticks closer to what would probably be measured empirically and to the definition set out by the originators of the term:

The equations... imply that any time the gametic frequencies will deviate from the equilibrium frequencies by an amount \(D\) which is the product of the coupling gametic frequencies minus that of the repulsion gametic frequencies. \(D\), thus defined, may be considered as a measure of linkage disequilibrium (italics in original; 2, p. 459).

However, our interpretation better captures the quantity that matters to the evolution of the haplotype frequencies [4] and follows the definition used by Crow & Kimura ([25], p. 197), namely ‘the difference in frequency between the two types of heterozygotes’.

**REFERENCES**


