Disequilibrium in Two-Locus Mutation-Selection Balance Models

Alan Hastings

Department of Mathematics and Division of Environmental Studies, University of California, Davis, California 95616

Manuscript received August 5, 1987
Accepted December 2, 1987

ABSTRACT

Equilibrium behavior of two-locus mutation-selection balance models is analyzed using perturbation techniques. The classical result of Haldane for one locus is shown to carry over to two loci, if fitnesses are replaced by marginal fitnesses. If the fitness of the double heterozygote is smaller than would be produced by a multiplicative model, as in additive or quantitative fitness models, the disequilibrium is negative—an excess of gametes with one rare allele. In this case the disequilibrium can be as large as one-half its maximum value possible, if the recombination rate is small, not greater than the strength of selection. If the fitness of the double heterozygote is larger than would be produced by a multiplicative model, the disequilibrium is positive, and is very small relative to its maximum value possible, even if the recombination rate is zero.

Among the classic results achieved in population genetics during the 1920s was Haldane's (1927) computation of the equilibrium allele frequency at a single locus where there is a balance between selection and mutation. Haldane considered the case where, without mutation, one of the alleles would be fixed, and where the mutation rate was small. More recently, Karlin and McGregor showed in (1971) that in a two-locus setting with similar relationships between fitness and mutation rates that again there would be a balance between mutation and selection with both loci close to fixation at equilibrium. Their approach, based on the techniques elaborated further in Karlin and McGregor (1972), also applies directly to the multilocus multialle case. However, they did not calculate what the equilibrium frequencies would be. In particular, the magnitude of disequilibrium, or correlation between alleles at the two loci, was not calculated. Some special cases of mutation selection balance for two locus two allele models were discussed in Christiansen and Frydenberg (1977). The goal of the current paper is to extend Haldane's approach to calculate (approximate) equilibrium frequencies under mutation selection balance in a two locus model. Extensions of the results here to multiple loci will be treated in a future paper.

There are a number of important biological questions in which multilocus mutation-selection balance models play an important role. A number of authors, and Mukai et al. (1974) in particular, have discussed the role of mutation selection balance in maintaining variability for viability polygenes. How are these experimental analyses affected by interactions among loci? The role of mutation selection balance in maintaining variability has also been stressed in the context of quantitative characters, particularly by Lande (1976) and a number of other authors as discussed in Nagylaki (1984) and Turelli (1984). When it is valid to ignore disequilibrium in these analyses, and what is the effect of recombination? A final question of biological interest is the sign of the disequilibrium, $D$ (Langley and Crow, 1974). Langley, Tobari and Kojima (1974) showed in a survey that in the majority of cases in a survey of Drosophila melanogaster, the linkage disequilibria were negative, meaning an excess of gametes with one rare and one common allele. What are the signs of disequilibria generated by mutation-selection balance models? Feldman, Christiansen and Brooks (1980) have discussed the evolution of recombination modifiers in diploids when both selection and mutation are acting. The outcome is determined by the sign of the disequilibrium coefficient, $D$.

Thus, there are two sets of questions of biological interest that I will answer. The first is, in a two-locus model, how large is the disequilibrium describing the association between alleles at different loci under mutation selection balance? What effect does recombination have on allele frequencies and disequilibrium? The second is, what is the sign of linkage disequilibrium?

I will make reasonable assumptions about the form of mutation and selection below. First, I will assume that mutation rates are much smaller than the strength of selection. This would imply the existence of a "common" allele at each locus, with other alleles at low frequency. This is supported by the majority of biological evidence, as reviewed in Turelli (1984). Second, I assume that there is a homozygote of higher
fitness than any other genotype with two or more of the common alleles.

MODEL

The model I investigate here is the standard two-locus, two-allele model describing the effects of mutation and selection. Let there be two loci with alleles $A$ and $a$ at the first and alleles $B$ and $b$ at the second. Let the frequencies of the four gametic types $AB, Ab, aB, ab$ be $x$, after viability selection and before mutation, and $y_i$ after mutation and before viability selection, for $i$ from one to four, respectively. Let $w_{ij} = w_{j'i'}$, with $w_{14} = w_{23}$, be the fitness (viability) of an individual with gametes whose frequencies are given by $y_i$ and $y_j$. Let the linkage disequilibrium, 

$$D = y_1y_4 - y_2y_3,$$

and let $r$ be the recombination rate between the two loci. Denote the mutation rate from $a$ to $A$ by $\mu_{Aa}$, from $A$ to $a$ by $\mu_{aA}$, from $b$ to $B$ by $\mu_{Bb}$ and from $B$ to $b$ by $\mu_{bb}$. Here the $\mu$'s are roughly 1, and $\epsilon$ is a small parameter—a typical value is $10^{-5}$. Finally, let the marginal mean fitness of gamete $i$ be

$$w_i = \frac{4}{\epsilon} \sum_{j=1}^6 y_j w_{ij},$$

and the mean fitness is

$$\bar{w} = \frac{4}{\epsilon} \sum_{i=1}^6 y_i w_i.$$ (2)

Then the dynamics of the mutation selection model are described by the pairs of equations (where the sign is plus for $i = 2$ or 3 and minus otherwise):

$$x'_i = (y_i w_i \pm r w_{14} D) / \bar{w},$$

and

$$y_i = x_i + \sum_{j=i}^6 u_{ij} x_j - \sum_{j=i}^6 u_{ji} x_i, \quad (5)$$

where $u_{ij}$ is the mutation rate from gamete $j$ to gamete $i$. Note that under the usual assumption that mutation is independent at the two loci, there is a simple relationship between the $u_{ij}$'s and the $\mu$'s, e.g.:

$$u_{14} = \epsilon^2 \mu_{Aa} \mu_{Bb},$$

$$u_{44} = \epsilon \mu_{aA} \mu_{bb} - \epsilon^2 \mu_{Aa} \mu_{Bb}. \quad (6)$$

Finally, in line with the biological assumptions discussed in the introduction, assume

$$w_{44} = 1 > w_{4a}, \quad (7)$$

where $i$ is 1, 2 or 3. A stronger assumption than (7) will in fact be used below, namely that mutation is weak relative to selection:

$$1 - w_{4a} > \epsilon. \quad (7')$$

This assumption is adequate for determining mutation-selection balance to first order. However, the second order results below depend on the further assumption that:

$$1 - w_{ij} > \epsilon$$

if the pair $(i, j)$ is not $(4, 4)$.

RESULTS

The major biological results follow, and are demonstrated below. Here mutation is weak relative to selection and $y$ and $x$ refer to equilibrium gamete frequencies.

RESULT 1. Haldane’s classic result for one locus mutation selection balance carries over to two loci, if fitnesses are replaced by induced fitnesses (Ewens and Thompson 1977). Thus, if $\mu_A$ is the frequency of allele $A$ after mutation, before selection, at equilibrium:

$$\mu_A = \epsilon \mu_{Aa} / (1 - w_{4a}) \approx \epsilon \mu_{Aa} / (1 - w_{24}). \quad (9)$$

where $w_{4a}$ is the induced fitness of the genotype $Aa$. Similarly, one finds:

$$\mu_B = \epsilon \mu_{Bb} / (1 - w_{4b}) \approx \epsilon \mu_{Bb} / (1 - w_{34}). \quad (10)$$

RESULT 2. The sign of the linkage disequilibrium to lowest order is determined by the form of selection. In particular, the disequilibrium $D$ is given by:

$$D = \epsilon^{2} \mu_{Aa} \mu_{Bb} (w_{14} - w_{24} w_{34}) / [(1 - w_{24}) \times (1 - w_{34}) (1 - w_{14} (1 - r))]. \quad (11)$$

A number of conclusions can be drawn from this formula. The sign of the disequilibrium is determined by the sign of $w_{14} - w_{24} w_{34}$ (as noted in Feldman, Christiansen, and Brooks 1980), so it is the fitness of the double heterozygote relative to the fitness of a double homozygote in a multiplicative model which determines the sign of disequilibrium. Second, let $1 - w_{ij} = s_0$, where $s$ is a small parameter, and $w_{ij}$ is order one. Then, consider fitnesses which are not multiplicative. In this case, if $r$ is the same magnitude as $s$, or smaller (even zero), $D$ is $O((s)2)$. If $r$ is much larger than $s$, $D$ is $O(\epsilon^{2} / s)$, which is very small. The interpretation of these magnitudes should be made relative to the largest (absolute) values possible for $D$. If $D$ is positive its maximum value is the minimum of $\mu_A (1 - p_B)$ and $\mu_B (1 - p_A)$, both of which are $O(\epsilon/s)$, so in this case the disequilibrium is very small.

However, if the disequilibrium is negative, as would be the case under models of additive or of quadratic or gaussian stabilizing selection, $D$ is large relative to its largest absolute value possible. The most negative $D$ can be is the maximum of $-p_A p_B$ and $-(1 - p_A)(1 - p_B)$, which in these models is $-p_A p_B$. Thus, to lowest order, if $D$ is negative, $D$ divided by its largest
negative value possible is:
\[-(w_{14} - w_{24}w_{34})/(1 - w_{14}(1 - r)).\] (12)

If, for example \(w_{14} = 1 - 4k\) and \(w_{24} = w_{34} = 1 - k\) as in a model of quadratic stabilizing selection, formula (12) becomes approximately:
\[2k/(4k + r).\] (13)

Thus the disequilibrium can be large—as large as one half its maximum value—but only if the recombination rate is small. For large recombination rates the disequilibrium varies as the inverse of \(r\).

**ANALYSIS**

The equilibrium is found as a power series in \(\epsilon\), using standard perturbation techniques. (The accuracy of the approximations was checked by numerically iterating the equations, and the approximations were accurate for selection coefficients as weak as \(10^{-2}\), when mutation rates were roughly \(10^{-5}\).) Thus all the terms in the model will be written as in the following example:

\[x_i = x_{i,0} + \epsilon x_{i,1} + \epsilon^2 x_{i,2} + \ldots,\] (14)

where the subscript after the comma indicates which power of \(\epsilon\) the coefficient multiplies. Note that for \(j > 1\),
\[\sum_{i=1}^{4} y_{i,j} = \sum_{i=1}^{4} x_{i,j} = 0.\] (15)

Consequently, for \(j > 1\), I find \(y_{4,j}\) and \(x_{4,j}\) by finding \(y_{i,j}\) and \(x_{i,j}\) for \(i = 1\) to 3.

First set \(x'_i = x_i\) to obtain equations for the equilibrium. Begin the analysis with Equation 4 and express all the relevant variables and parameters as power series. Equate terms of order \(\epsilon\) raised to the powers zero, one, and two respectively to obtain the following equations:

\[
\begin{align*}
\bar{w}_0 x_{i,0} &= y_{i,0} w_{i,0} \pm r w_{14} D_{1,0} \\
\bar{w}_0 x_{i,1} + \bar{w}_1 x_{i,0} &= y_{i,0} w_{i,1} + y_{i,1} w_{i,0} \\
&\pm r w_{14} D_{1,1} \\
\bar{w}_0 x_{i,2} + \bar{w}_1 x_{i,1} + \bar{w}_2 x_{i,0} &= y_{i,0} w_{i,2} + y_{i,1} w_{i,1} + y_{i,2} w_{i,0} \\
&\pm r w_{14} D_{1,2}
\end{align*}
\] (16a, 16b, 16c)

where, for example,
\[D_{2} = y_{1,0} y_{4,1} + y_{1,1} y_{4,4} + y_{1,2} y_{4,0} - y_{2,0} y_{4,2} - y_{2,1} y_{4,3} - y_{2,2} y_{4,0}.\] (17)

The equations describing mutation will be written out below, as needed.

Since there is no mutation to lowest order, the zero-th order approximation is determined only by Equation 16a. Moreover, Equation 8 implies that the (stable) solution of the zero-th order Equation 16a is
\[y_{i,0} = x_{i,0} = 0\] (18a)

for \(i = 1\) to 3 and
\[y_{4,0} = x_{4,0} = 1.\] (18b)

This solution merely reflects the assumptions about the form of the fitnesses and the lack of mutation to this order. Note that (18) also implies that:
\[\bar{w}_0 = w_{44} = 1\] (19)

and
\[w_{i,0} = w_{i4} = \bar{w}_i.\] (20)

**Solution to first order:** Making use of the zero-th order solution (18), the equations describing mutation to first order become:

\[
\begin{align*}
y_{4,1} &= x_{4,1} - (\mu_A + \mu_B) \\
y_{3,1} &= x_{3,1} + \mu_B \\
y_{2,1} &= x_{2,1} + \mu_A \\
y_{1,1} &= x_{1,1}
\end{align*}
\] (21a, 21b, 21c, 21d)

At this point, using (18b) and (21d), one can deduce that
\[y_{1,1} = x_{1,1} = 0.\] (22)

A small amount of algebra then yields:

\[
\begin{align*}
x_{2,1} &= \mu_A w_{24}/(1 - w_{24}) \\
x_{2,1} &= \mu_A/(1 - w_{24}) \\
x_{3,1} &= \mu_B w_{34}/(1 - w_{34}) \\
x_{3,1} &= \mu_B/(1 - w_{34})
\end{align*}
\] (23a, 23b, 23c, 23d)

As the fitness of the single heterozygote, \(w_{24}\) or \(w_{34}\), determines the induced fitness (EWENS and THOMSON 1977) of a heterozygote at one locus, to this order, (22) and (23) are identical to Haldane’s classic result for mutation-selection balance at a single locus. This is Equations 9 and 10. From (22) and (23), one can then deduce that the disequilibrium \(D\) satisfies:
\[D_{1} = 0.\] (24)

**Solution to second order:** Since no information about the disequilibrium or the role of recombination shows up at this point, I proceed to calculate \(D\) to second order to determine information about linkage disequilibrium. Using the first order results, (17) simplifies to:
\[D_{2} = y_{1,2} y_{4,0} - y_{2,1} y_{3,1}.\] (25)

Use (22)–(24) in (13c) for \(i = 1\) to 3 to obtain:
\[x_{i,2} + \bar{w}_1 x_{i,1} = y_{i,1} w_{i,1} + y_{i,2} w_{i4} \pm r w_{14} D_{2}\] (26)
Note that
\[ \bar{w}_i = -2[\mu_A + \mu_B] \]  
(27)

For \( i = 1, \) (26) simplifies to:
\[ x_{1,2} = y_{1,2}w_{14} - rw_{14}D_{2}. \]  
(28)

The equations describing mutation to second order are:
\[ y_{1,2} = x_{1,2} + \mu_B x_{2,1} + \mu_A x_{3,1} + \mu_A \mu_B \]
\[ y_{2,2} = x_{2,2} - (\mu_A + \mu_B)x_{2,1} + \mu_A x_{4,1} - \mu_A \mu_B \]
\[ y_{3,2} = x_{3,2} - (\mu_A + \mu_B)x_{3,1} + \mu_B x_{4,1} - \mu_A \mu_B \]  
(29)
\[ y_{4,2} = x_{4,2} + \mu_A x_{2,1} + \mu_B x_{3,1} + \mu_A \mu_B \]
\[ - \mu_A x_{4,1} - \mu_B x_{4,1} \]

Thus,
\[ x_{1,2} = \mu_A \mu_B [x_{14}(1 - r)[1 + w_{24}]/(1 - w_{24}) \]
\[ + w_{34}/(1 - w_{34})] + rw_{14}/[(1 - w_{24}) \]
\[ \times (1 - w_{34})]/(1 - w_{14}(1 - r)) \]
\[ y_{1,2} = \mu_A \mu_B [(1 + w_{24}/(1 - w_{24}) + w_{34}/(1 - w_{34}) \]
\[ + rw_{14}/[(1 - w_{24}))(1 - w_{34})]/(1 - w_{14}(1 - r))]. \]  
(31)

Combining (31) with (23) and (26) and simplifying yields:
\[ D_{2} = \mu_A \mu_B (w_{14} - w_{24}w_{34})/[(1 - w_{24}) \]
\[ \times (1 - w_{34})/(1 - w_{14}(1 - r))]. \]

This is Equation 11, which is discussed extensively above.

Using (26), (27), and (29)–(32) one can easily determine the second order terms in the approximations for the other gametes. As I do not wish to draw any biological conclusions from these results, I do not report them here.

**DISCUSSION**

The results here fall into two categories. First, I have shown that the single locus mutation selection balance results carry over to multiple loci in a natural way, if the overall mutation rate is small enough.

The second results concern linkage disequilibrium. The sign of the linkage disequilibrium is determined by the fitnesses relative to those of a multiplicative model. FELDMAN, CHRISTIANSEN, and BROOKS (1980) show that if disequilibrium generated by mutation selection balance is positive, modifiers which increase recombination rate will increase in frequency if they are tightly linked to the loci for which disequilibrium is positive. Conversely, they show that if disequilibrium is negative, modifiers which decrease recombination rate will increase in frequency if they are loosely linked to the loci for which disequilibrium is negative. Thus the importance of this effect depends on the form of selection. However, both additive selection, or selection schemes based on quadratic stabilizing selection lead to negative linkage disequilibrium, as also shown by FELDMAN, CHRISTIANSEN, and BROOKS (1980).

Several conclusions can be drawn concerning the magnitude of linkage disequilibrium. All of these conclusions drawn from the deterministic model presented here apply only to large populations. First, the assumption that linkage disequilibrium can be ignored at equilibrium in mutation-selection balance models is strongly supported here, if the magnitude of selection is less than here, if the magnitude of selection is less than the recombination coefficient.

However, for those loci so closely linked that recombination is approximately equal to selection in magnitude, disequilibrium can have a large effect. Second, the fact that even as the recombination rate goes to zero, large positive disequilibrium (a deficiency of the gametes with one rare allele) implies that mutation selection balance can be ruled out as the source of the maintenance of variability between two closely linked loci (or sites) with large positive disequilibrium. Among the possibilities maintaining large positive disequilibrium at equilibrium would be epistatic selection or population subdivision. Another likely cause of large observed values of disequilibrium for tightly linked loci would be historical effects, which would take a long time to be eliminated for tightly linked loci.

I thank John Gillespie, Chuck Langley and Mike Turelli for helpful discussions, Thomas Nagylaki for very helpful comments, and Lisa Brooks and Freddy Bugge Christiansen for helpful comments on the manuscript. Supported by U.S. Public Health Service grant GM32130.

**LITERATURE CITED**


Communicating editor: W. J. EWENS