

THE DIRECTION OF LINKAGE DISEQUILIBRIUM^{1,2}

CHARLES H. LANGLEY³ AND JAMES F. CROW

Genetics Laboratory, University of Wisconsin, Madison, Wisconsin 53706

Manuscript received May 9, 1973

Revised copy received May 3, 1974

ABSTRACT

The previous paper (LANGLEY, TOBARI and KOJIMA 1974) reports that the directional linkage disequilibria, $D_\omega = P_{AB}P_{ab} - P_{Ab}P_{aB}$, tend to be negative for data between allozymes and linked to inversions. A and B stand for the two alleles with the greatest frequency in the population. In this paper we show that linkage disequilibrium in this direction is produced at equilibrium when double homozygotes have fitnesses that are a constant fraction of the product of the two component single homozygote fitnesses, a pattern that is frequently observed in experimental data.

IN the preceding report LANGLEY, TOBARI and KOJIMA (1974) pointed out a method of orienting the direction of linkage disequilibrium with respect to gene frequencies. They noted that the majority of linkage disequilibria between allozymes linked to inversions in *D. melanogaster* populations indicated an excess of gametes composed of one frequent allele and one rarer allele. Although this observation is not statistically significant, it did lead us to the following theoretical findings.

Let A , a , B , and b represent alleles at two linked loci and designate their frequencies by p^A , p^a , p^B , and p^b . The symbols can stand for allozymes, gene arrangements, or any other entity that is inherited as a unit. If we designate A and B as the more frequent types ($p^A > p^a$ and $p^B > p^b$), then the observed direction of linkage disequilibrium tends to be such that

$$D_\omega = p_{AB}p_{ab} - p_{Ab}p_{aB} < 0. \quad (1)$$

The product of the frequency of the gametes containing the two most frequent and the two least frequent alleles is less than the product of the frequencies of the other two. The symbol D_ω is used instead of D to emphasize that gene frequencies are considered and therefore the sign of D_ω is meaningful.

It will be shown that very plausible systems of fitness interactions lead to $D_\omega < 0$ at equilibrium. We assume that the polymorphism is maintained by selective advantage of heterozygotes or some other form of balancing selection. We further assume that double homozygotes are less fit than if the two loci were independent in their effects on survival and fertility. There is evidence for

¹ This is paper No. 1809 of the Laboratory of Genetics, University of Wisconsin, Madison, Wisconsin.

² Supported by NIH grant GM 15422.

³ Present address: Mutagenesis Branch, National Institute of Environmental Health Sciences, P.O. Box 12233, Research Triangle Park, North Carolina 27709.

departures from linearity in this direction in the decline of fitness with inbreeding (SPASSKY, DOBZHANSKY and ANDERSON 1965; TEMIN *et al.* 1969; KOSUDA 1971).

The Two-Locus Model

We assume a continuous-time model with random mating. Let $x_1 = p_{AB}$, $x_2 = p_{Ab}$, $x_3 = p_{aB}$, and $x_4 = p_{ab}$ be the frequencies of the four chromosomes. Let m_{ij} be the fitness of the corresponding genotype; the fitness is being measured in Malthusian parameters. Then the rate of change of the i^{th} chromosome type is given by

$$\dot{x}_i = x_i(m_i - \bar{m}) - \varepsilon_i r b D \tag{2}$$

where

$$m_i = \sum_j x_j m_{ij}$$

$$\bar{m} = \sum_i x_i m_i$$

r = recombination fraction between the loci
 b = birth rate per generation of the double heterozygote, AB/ab or Ab/aB

$$D = x_1 x_4 - x_2 x_3$$

$$\varepsilon_1 = -\varepsilon_2 = -\varepsilon_3 = \varepsilon_4 = 1$$

For a derivation of these equations and a discussion of the circumstances where they are appropriate, see KIMURA (1956, 1965), FELSENSTEIN (1965), CROW and KIMURA (1970, p. 196ff), and NAGYLAKI and CROW (1974).

We measure epistasis following FISHER (1918) by

$$E_i = m_{i1} - m_{i2} - m_{i3} + m_{i4}$$

and

$$\bar{E} = \sum_i x_i E_i \tag{3}$$

and define Z as

$$Z = x_1 x_4 / x_2 x_3 .$$

Then (see CROW and KIMURA 1970, p. 201)

$$\dot{Z} = Z(\bar{E} - r b D P) \tag{4}$$

where

$$P = \sum_i \frac{1}{x_i} .$$

We are interested in polymorphic equilibrium, in which case $\dot{Z} = 0$, and

$$\bar{E} = r b D P. \tag{5}$$

Note that $x_1 = p_A p_B + D$, $x_2 = p_A p_b - D$, $x_3 = p_a p_B - D$, and $x_4 = p_a p_b + D$. Substituting these into (3) yields

$$\begin{aligned} \bar{E} &= p_A p_b E_1 + p_A p_b E_2 + p_a p_B E_3 + p_a p_b E_4 + D \sum_i \varepsilon_i E_i \\ &= \bar{E} + D E \end{aligned} \tag{6}$$

where

$$\bar{E} = p_A p_B E_1 + p_A p_b E_2 + p_a p_B E_3 + p_a p_b E_4 \quad (7)$$

and

$$E = E_1 - E_2 - E_3 + E_4. \quad (8)$$

\bar{E} is the average additive \times additive epistatic deviation; E is the same quantity for a population with the same gene frequencies at linkage equilibrium. These two quantities differ by DE , where D is the measure of linkage equilibrium and E is the sum of dominance \times dominance epistatic deviations (COCKERHAM 1954; KEMPTHORNE 1957).

If we assume equilibrium and equate the right-hand side of (5) and (6), this leads to

$$\bar{E} = D(rbP - E). \quad (9)$$

We are interested in inquiring as to when $D_\omega < 0$, the situation observed in natural populations. From (9) it is apparent that sufficient conditions for this are

$$E < rbP \quad (10)$$

and

$$\bar{E} < 0. \quad (11)$$

The alternative sufficient condition, $E > rbP$ and $\bar{E} > 0$, seems much less likely to be found in a polymorphic situation.

Application of (10) and (11) to test a specific model for the direction of linkage disequilibrium when the population reaches equilibrium would depend on knowing the equilibrium values of p_A and p_B . This is usually not easy to do. However, as DR. THOMAS NAGYLAKI has pointed out to us equation (7) is linear in p_A for constant p_B and in p_B for constant p_A ; therefore there is no maximum nor minimum in the region of interest, $1/2 < p_A, p_B < 1$. This means that if $D_\omega < 0$ for the four values of (p_A, p_B) , corresponding to (1,1), (1,1/2), (1/2,1), and (1/2,1/2) it must be for all intermediate values. Applying this to equation (7), $\bar{E} < 0$ if

are

$$\begin{aligned} (E_1 + E_2 + E_3 + E_4) &< 0 \\ (E_1 + E_2) &< 0 \\ (E_1 + E_3) &< 0, \text{ and} \\ E_1 &< 0. \end{aligned} \quad (12)$$

These, along with (10), are sufficient conditions for D_ω to be negative. If $E > rbP$, then all the inequalities in (12) must be reversed.

Although these are strong conditions, and sufficient, they are not necessary. There may well be models where one or more of the corners do not meet the criterion, although an equilibrium point does.

General Two-Locus Heterotic Model

Assume that the fitnesses, measured in Malthusian parameters, of the genotypes are

	<i>AA</i>	<i>Aa</i>	<i>aa</i>
<i>BB</i>	$m_{11} = -a_1 - b_1 - k_1$	$m_{13} = -b_1$	$m_{33} = -a_2 - b_1 - k_3$
<i>Bb</i>	$m_{12} = -a_1$	$m_{14} = m_{23} = 0$	$m_{34} = -a_2$
<i>bb</i>	$m_{22} = -a_1 - b_2 - k_2$	$m_{24} = -b_2$	$m_{44} = -a_2 - b_2 - k_4$

$$E_1 = -k_1, E_2 = k_2, E_3 = k_3, E_4 = -k_4 \quad (13)$$

$$E = -(k_1 + k_2 + k_3 + k_4). \quad (14)$$

If $k_i = 0$, this corresponds to additivity between loci (no epistasis). There is no linkage disequilibrium ($E = 0, D = 0$).

If $k_i = k$, a positive constant, and the a 's and b 's are positive, then substitution into (7) shows that $E < 0$; and, since $E < 0 < rbP$, $D_\omega < 0$. Thus a constant reduction in fitness of the double homozygotes in comparison to their non-epistatic expectations leads to a negative value of D_ω . This corresponds to a discrete generation model in which the fitness of the double homozygotes is a constant fraction (less than 1) of the product of the two component single homozygotes. More generally, using (12) it is apparent that any situation where k_1 is larger than k_2 or k_3 and where $k_1 + k_4 > k_2 + k_3$ satisfies the conditions for negative D_ω . Also note from (7) that the more nearly p_A and p_B approach 1 the more the first term dominates the expression and the more likely it is for D_ω to be negative.

Another model that has often been used is quadratic. If a^2 is the decrease in fitness of one single homozygote and b^2 is the decrease for the other single homozygote, then the double homozygote is assumed to have a fitness decrease $(a + b)^2$, all measured with respect to the double heterozygote. This model does not conform to the sufficient conditions given above. However, several computer trials with a variety of numerical values and starting points ended up with equilibria corresponding to $D_\omega < 0$, and we conjecture that this may be the case in general although we are not able to prove it.

It is perhaps of interest to mention an example that leads to $D_\omega > 0$. KIMURA (1956, p. 279) gives a model where one locus is heterotic and the other is not, but where the polymorphism is maintained by linkage and epistasis. In this case $E_1 = E_2 = E_3 = E_4 = s$, $E = 0$, $\bar{E} > 0$, and $D_\omega > 0$.

DISCUSSION

The data so far analyzed suggest that when there is linkage disequilibrium, such as for linked allozymes and inversions in *Drosophila melanogaster*, $D_\omega < 0$ (LANGLEY, TOBARI and KOJIMA 1974). There are no comparable observations in other species. The most accessible material appears to be in *Drosophila pseudoobscura* and studies on this species seem to offer the best possibility of substantiating or refuting the data so far summarized. Since linked allozymes show no pattern of linkage disequilibrium, this suggests that the negative value of D_ω

for the association of allozymes and inversions is not sampling bias or some peculiar properties of two-locus random drift.

This paper relates two quite different types of data. The decrease in viability under inbreeding shows a consistent nonlinear component (SPASSKY, DOBZHANSKY and ANDERSON 1965; TEMIN *et al.* 1969; KOSUDA 1971). This nonlinearity can be interpreted as negative dominance by dominance epistasis. The model that we have discussed in this paper ($k_i = k$) has a negative dominance by dominance epistasis, and this leads to a negative value of D_{ω} .

That we can suggest models for fitness interaction which lead to linkage disequilibria in the direction observed in nature does not, of course, prove that these are the mechanisms by which the disequilibria are generated. Further information is needed, both as to the reality and generality of the experimental observation, and of the variety of models that lead to this expectation.

LITERATURE CITED

- COCKERHAM, C. D., 1954 An extension of the concept of partitioning hereditary variance for analysis of covariance among relatives when epistasis is present. *Genetics* **39**: 859-882.
- CROW, J. F. and M. KIMURA, 1970 *An Introduction to Population Genetics Theory*. Harper and Row, New York.
- FELSENSTEIN, J., 1965 The effect of linkage on directional selection. *Genetics* **52**: 349-363.
- FISHER, R. A., 1918 The correlation between relatives on the supposition of mendelian inheritance. *Trans. Roy. Soc. Edinburgh* **52**: 399-433.
- KEMPTHORNE, O., 1957 *An Introduction to Genetic Statistics*. John Wiley and Sons, New York.
- KIMURA, M., 1956 A genetic system which leads to closer linkage by natural selection. *Evolution* **10**: 278-287. —, 1965 Attainment of quasilinkage equilibrium when gene frequencies are changing by natural selection. *Genetics* **52**: 875-890.
- KOSUDA, K., 1971 Synergistic interaction between second and third chromosomes on viability of *Drosophila melanogaster*. *Japan. J. Genet.* **46**: 41-52.
- LANGLEY, C. H., Y. N. TOBARI and K. KOJIMA, 1974 Linkage disequilibrium in natural populations of *Drosophila melanogaster*. *Genetics* (this issue).
- NAGYLAKI, T. and J. F. CROW, 1974 Continuous selective models. *Theoret. Pop. Biol.* **52**: 257-283.
- SPASSKY, B., Th. DOBZHANSKY and W. W. ANDERSON, 1965 Genetic studies of natural populations. XXXVI. Epistatic interactions of the components of the genetic load in *Drosophila pseudoobscura*.
- TEMIN, R. G., H. U. MEYER, P. S. DAWSON and J. F. CROW, 1969 The influence of epistasis on homozygous viability depression in *Drosophila melanogaster*. *Genetics* **61**: 497-519.

Corresponding editor: R. C. LEWONTIN