Theoretical and Applied Genetics 38, 226–231 (1968)

Linkage Disequilibrium in Finite Populations

W. G. HILL and ALAN ROBERTSON*

Institute of Animal Genetics, Edinburgh, 9.

Summary. A theoretical investigation has been made of the influence of population size (N) and recombination fraction (c) on linkage disequilibrium (D) between a pair of loci. Two situations were studied: (i) where both loci had no effect on fitness and (ii) where they showed heterozygote superiority, but no epistacy.

If the populations are initially in linkage equilibrium, then the mean value of D remains zero with inbreeding, but the mean of D^2 increases to a maximum value and decreases until fixation is reached at both loci. The tighter the linkage and the greater the selection, then the later is the maximum in the mean of D^2 reached, and the larger its value. The correlation of gene frequencies, r, in the population of gametes within segregating lines was also studied. It was found that, for a range of selection intensities and initial gene frequencies, the mean value of r^2 was determined almost entirely by $N\,c$ and time, measured proportional to N.

The implication of these results on observations of linkage disequilibrium in natural populations is discussed.

Most of the mathematical theory of linkage has been developed for populations which are sufficiently large that a deterministic model can be used. In these large populations, which are not undergoing selection, the theory of the rate of approach to linkage equilibrium is well worked out, and it is known that populations in equilibrium remain in that state (GEIRINGER, 1944; BENNETT, 1954). More recent work has been devoted to the effect of selection on linkage disequilibrium in very large populations. LEWONTIN and KOJIMA (1960) showed that epistacy was necessary for linkage disequilibrium to be maintained in a selected population in which gene frequencies are at equilibrium. However in a population in which there are directional changes of gene frequency resulting from artificial selection, some linkage disequilibrium may be observed if there is no epistacy of gene action on the selected character (NEI, 1963; FELSENSTEIN, 1965), but not if the selective values at each loci combine in a multiplicative manner (FELSENSTEIN, 1965).

For finite unselected random mating populations, expressions for changes in linkage disequilibrium have been given by KIMURA (1963) and by HILL and ROBERTSON (1966). WRIGHT (1933) had previously derived formulae for the proportion of recombinants at final fixation. In this paper we shall mainly consider the fate of a pair of linked loci, which we may observe in a number of replicate lines drawn from a large population initially in linkage equilibrium. If the loci have no effect on fitness, then over the average of all replicates these loci will remain in equilibrium, but as a result of genetic sampling the disequilibrium will not be zero in each line. In other words, the variance of the linkage disequilibrium, D, will not be zero, though the mean will be. We shall evaluate this variance, and show that it can be of an order of magnitude similar to that of the variance of gene frequencies after some generations of inbreeding. We shall study the case of neutral genes in greatest detail, and then extend the results to include heterozygote advantage at each locus, but with no epistacy. The results may therefore have some bearing on the estimation and interpretation of linkage disequilibrium in natural populations.

* Member of the A.R.C. Unit of Animal Genetics.

Disequilibrium Between Neutral Loci

We consider two loci with alternative alleles A_1, A_2 and B_1, B_2 which have no effect on fitness, and we let p and q be the frequencies of A_1 and B_1 respectively. Linkage disequilibrium is commonly measured by the determinant D, given by

$$D = f(A_1\,B_1)\,f(A_2\,B_2) - f(A_1\,B_2)\,f(A_2\,B_1)$$

where f denotes the appropriate gametic frequency.

Using E to denote expectation, the recurrence equation for the mean of D after t generations of random mating with no selection is

$$E(D_t) = (1 - c)\,(1 - 1/2\,N)\,E(D_{t-1})\,, \qquad (1)$$

where N is the effective population size and c the cross-over distance (HILL and ROBERTSON, 1966). If c and $1/2\,N$ are sufficiently small that their product can be ignored

$$\begin{aligned} E(D_t) &= (1 - c - 1/2\,N)\,E(D_{t-1}) \\ &= D_0\,e^{-(2\,N\,c+1)\,t/2\,N}\,, \quad \text{approximately} \qquad (2) \end{aligned}$$

if the population size is constant. In general, if N is large and c is of order $1/N$ or less, changes in the distribution of gametic frequencies can be approximated in a continuous model using a diffusion equation. Under these assumptions it can be shown that the pattern of change in gametic frequencies is a function of only the initial conditions p_0, q_0 and D_0 and of the product $N\,c$, if time is expressed on a scale proportional to N (HILL and ROBERTSON, 1966). Equation (2) is clearly of this form.

Changes in the average value of D^2 can be obtained using a moment generating matrix (ROBERTSON, 1952). Let y be a column vector of moments with dimension three, and elements

$$\begin{aligned} y_1 &= E\,[p\,(1 - p)\,q\,(1 - q)]\,, \\ y_2 &= E\,[D\,(1 - 2\,p)\,(1 - 2\,q)]\,, \\ y_3 &= E[D^2]\,. \end{aligned}$$

If there is no crossing over, changes in these moments in successive generations can be obtained by taking expectations over the multinomial distribution of gametic frequencies with index n, where $n = 2\,N$, and rearranging the results in terms of y_1, y_2, and y_3. Denoting by $y_{(t)} = (y_{i(t)})$ the vector of moments at

some generation t, it can be shown that

$$y_{1(t+1)} = \left(1-\frac{1}{n}\right)^2 y_{1(t)} + \frac{1}{n}\left(1-\frac{1}{n}\right)^2 y_{2(t)} + \frac{2}{n^2}\left(1-\frac{1}{n}\right) y_{3(t)}$$

$$y_{2(t+1)} = \left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)^2 y_{2(t)} + \frac{4}{n}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right) y_{3(t)}$$

$$y_{3(t+1)} = \frac{1}{n}\left(1-\frac{1}{n}\right) y_{1(t)} + \frac{1}{n}\left(1-\frac{1}{n}\right)^2 y_{2(t)} + \left(1-\frac{1}{n}\right)\left[\frac{1}{n^2} + \left(1-\frac{1}{n}\right)^2\right] y_{3(t)} .$$

When crossing over occurs, the average values of gene frequencies are unchanged, terms in D are multiplied by a factor $(1-c)$ (c. f., equation (1)), and terms in D^2 by a factor $(1-c)^2$. Thus with crossing over followed by sampling, we obtain the following transition relationship:

$$y_{(t+1)} = \begin{pmatrix} \left(1-\frac{1}{n}\right)^2 & \frac{1}{n}\left(1-\frac{1}{n}\right)^2 (1-c) & \frac{2}{n^2}\left(1-\frac{1}{n}\right)(1-c)^2 \\ 0 & \left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)(1-c) & \frac{4}{n}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)(1-c)^2 \\ \frac{1}{n}\left(1-\frac{1}{n}\right) & \frac{1}{n}\left(1-\frac{1}{n}\right)^2 (1-c) & \left(1-\frac{1}{n}\right)\left[\frac{1}{n^2}+\left(1-\frac{1}{n}\right)^2\right](1-c)^2 \end{pmatrix} y_{(t)} \qquad (3)$$

$$= \boldsymbol{M}\, y_{(t)} ,$$

where $\boldsymbol{M}$ is termed the moment generating matrix and is independent of t, hence $y_{(t)} = \boldsymbol{M}^t y_{(0)}$.

We shall discuss in detail the case of initial linkage equilibrium, where $y'_{(0)} = (p_0(1-p_0)\, q_0(1-q_0)\, 0\, 0)$. Then $E(D^2)/[p_0(1-p_0)\, q_0(1-q_0)]$ is independent of the initial frequencies at the two loci, and under the continuous model assumptions will be a function only of $N\,c$ and time expressed proportional to N. Since $E(D) = 0$ if $D_0 = 0$, $E(D^2)$ measures the variance of D.

With complete linkage ($c = 0$) and large N an explicit solution for $E(D^2)$ can be obtained. This involves diagonalization of the moment generating matrix, and the derivation and general solution are given in the appendix. With initial equilibrium, the solution is

$$E(D^2) = \frac{1}{15} p_0(1-p_0)\, q_0(1-q_0)\,[6(1-F) - 5(1-F)^3 - (1-F)^6] ,$$

where F is the inbreeding coefficient.

Some results for $E(D^2)/[p_0(1-p_0)\, q_0(1-q_0)]$ for initial equilibrium are plotted in Figure 1, with time measured as $F = 1 - e^{-t/2N}$. The graphs were computed by repeated iteration of the matrix, $\boldsymbol{M}$, on to the vector $y_{(t)}$ using a population size of $N = 16$. When $c = 0$, $E(D^2)$ reaches a maximum of 0.165 $p_0(1-p_0)\, q_0(1-q_0)$ when $F = 0.4$, or $t = N$ generations approximately. With recombination, the maximum value of $E(D^2)$ is lower and is attained earlier. For example, for $N\,c = 1/4$, $E(D^2)$ reaches 0.14 $p_0(1-p_0)\, q_0(1-q_0)$ when $F = 0.31$ or $t = 0.75\,N$ generations.

We have made use of the generalisation derived from the continuous model that c only enters into the results in the form of $N\,c$ and that the time scale in generations is proportional to N. This was checked in the calculations and in Table 1 we present some of the results referring to the maximum values of D^2 reached. The table gives the observed maximum and the time in generations when it occurred.

Except for the smallest values of N and $N\,c = 4$ there appears to be sufficiently good agreement between the results obtained with different values of N for us to use this generalisation.

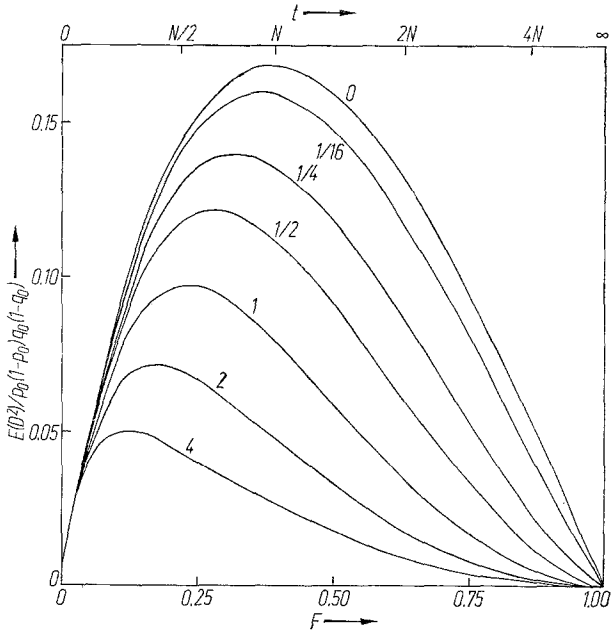

Fig. 1. The mean value of $D^2/[p_0(1-p_0)\, q_0](1-q_0)$ over segregating and non segregating lines for several values of $N\,c$ and no selection

Table 1. *The maximum value of $E(D^2)$ and the time in generations to reach it for different combinations of N and $N\,c$*

	N					$N\,c$
	8	16	32	64	$\to\infty$	
$D^2_{max}/[p_0(1-p_0)\, q_0(1-q_0)]$	.1708	.1678	.1663	.1656	.1649	0
t	8	16	32	64	1.0016 N	
$D^2_{max}/[p_0(1-p_0)\, q_0(1-q_0)]$	.1054	.0969	.0931	.0913	—	1
t	4	8	17	34	—	
$D^2_{max}/[p_0(1-p_0)\, q_0(1-q_0)]$	.0636	.0503	.0451	.0428	—	4
t	2	4	8	17	—	

The product of the variances in gene frequencies at the two loci, $p\,(1-p)\,q\,(1-q)$, is also affected by the degree of linkage. Starting with equilibrium, we have for $c = 0$ from the appendix that

$$E\,[p\,(1-p)\,q\,(1-q)\,] = \frac{1}{15}\,p_0\,(1-p_0)\,q_0\,(1-q_0) \times [6\,(1-F) + 10\,(1-F)^3 - (1-F)^6]\,.$$

With independent loci, the variance at each locus is proportional to $1-F$, and their product is proportional to $(1-F)^2$. However, with complete linkage, the product of the variances is

$$\frac{1}{15}\left[\frac{6}{1-F} + 10\,(1-F) - (1-F)^4\right]$$

times that with independence for all starting frequencies. This ratio is 1 at $F = 0$, rises to 1.13 at $F = 0.5$, 4.07 at $F = 0.9$ and becomes infinitely large as F approaches one.

Disequilibrium in Populations Segregating at Both Loci

We have developed the analysis so far in terms of the average values of D^2 computed over all replicate lines. But in many replicates one or other locus will become fixed after a few generations and in these D is zero. If we observe linkage disequilibrium between a pair of loci in natural populations, it can only be among those still segregating at both loci. We therefore need to describe the behaviour of the linkage disequilibrium within such lines. When $c = 0$, the average value of D^2 within lines still segregating at both loci, denoted D_s^2, can be obtained by dividing $E(D^2)$ from equation (3) by the proportion of lines still segregating. The latter can be calculated by series summation from formulae by KIMURA (1955), regarding the four gamete types as four alleles at a single locus. In the limiting case with $c = 0$, as F approaches 1 only lines in which two gametes segregate will remain. In those still segregating for both loci, gametes must either be entirely in the repulsion or entirely in the coupling phase. Therefore if we assume a final uniform continuous distribution of gametic frequencies ($A_1 B_2$, $A_2 B_1$) or ($A_1 B_1$, $A_2 B_2$) as will be true if N is large, it can be shown that D_s^2 approaches 1/30 for complete linkage as the inbreeding coefficient approaches one. This final value is independent of the initial conditions p_0, q_0 and D_0.

The values of both D^2 and D_s^2 depend during inbreeding on the initial frequencies, and we have found that a more useful statistic for lines segregating at both loci is the square of the correlation, r, of gene frequencies in the population of gametes, where $r = D/[p\,(1-p)\,q\,(1-q)]^{1/2}$. The expectation of r or r^2 is computed by averaging only over such lines. When there is initial equilibrium $E(r) = 0$. Changes in $E(r^2)$ with level of inbreeding were obtained by Monte Carlo simulation, using the same procedure as in our earlier work, but excluding selection (HILL and ROBERTSON, 1966). A population size of $N = 8$ was used and 10000 replicates were run for each level of recombination. Results are shown in Figure 2 for $p_0 = q_0 = 0.5$ and $D_0 = 0$. As replicate lines become fixed in the later generations, the sampling variances of the estimates of $E(r^2)$ increase, so results are plotted for only 48 generations ($F = 0.95$).

When there is complete linkage ($N\,c = 0$), $E(r^2)$ eventually reaches unity as all lines approach fixation. It is interesting that $E(r^2)$ and F are approximately equal to each other when $N\,c = 0$. When there is recombination, $E(r^2)$ approaches a limiting value dependent on $N\,c$ as F approaches one, the limit being reached earlier and at a lower level, the less tight the linkage. It is difficult to estimate the limit of $E(r^2)$ accurately when $N\,c$ is small because few lines are still segregating when $E(r^2)$ has reached a stable value.

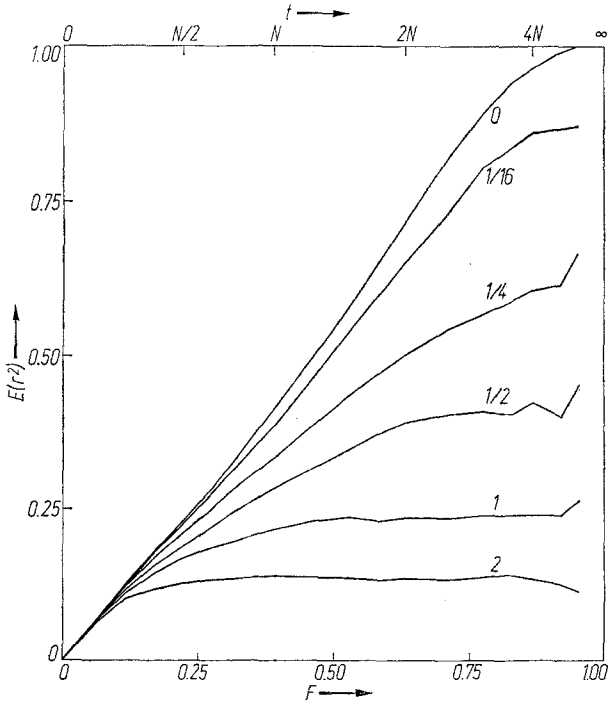

Fig. 2. The mean value of r^2 among segregating lines, $E(r^2)$, for several values of $N\,c$ with no selection

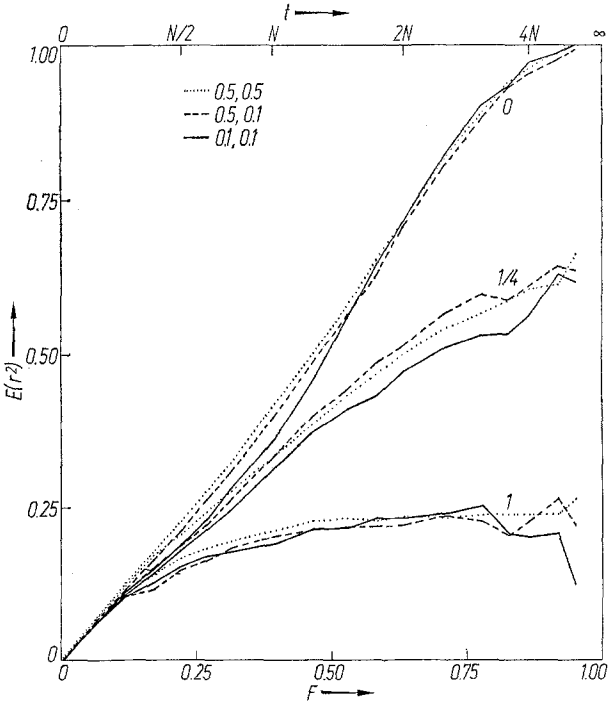

Fig. 3. The effect of initial frequency on $E(r^2)$ with no selection

The influence of initial frequency on $E(r^2)$ when there is no selection is shown in Figure 3. Three sets of initial frequencies are compared: (i) $p_0 = q_0 = 0.5$, (ii) $p_0 = 0.1$, $q_0 = 0.5$ and (iii) $p_0 = q_0 = 0.1$, each with $D_0 = 0$. It appears from Figure 3 that $E(r^2)$ is not very sensitive to changes in the initial frequency, and is mostly determined by $N\,c$ and time (measured as a function of N). As the inbreeding coefficient approaches unity, $E(r^2)$ depends only on the steady state distribution of gamete frequencies within the segregating lines, and is independent of the initial conditions.

Disequilibrium Between Loci Having Heterozygote Superiority

Our discussion has been restricted so far to the situation where there is no selection maintaining segregation. But genetic variation will be maintained for longer periods of time in small populations at loci in which the heterozygote has superior fitness to either homozygote, unless the homozygotes differ widely from each other in fitness (ROBERTSON, 1962; ROBERTSON and HILL, 1968). If there is no epistacy between these loci, selection will not cause linkage disequilibrium directly. But we must expect to find some disequilibrium in small populations as a result of genetic sampling. We can therefore predict that for pairs of loci each having heterozygote advantage, but not interacting with each other, we will have $E(D) = 0$, but $E(D^2) \neq 0$, just as for neutral genes.[1]

Let us assume that the relative selective advantages are as follows:

	B_1B_1	B_1B_2	B_2B_2
A_1A_1	$1 - r_1 - s_1$	$1 - r_1$	$1 - r_1 - s_2$
A_1A_2	$1 - s_1$	1	$1 - s_2$
A_2A_2	$1 - r_2 - s_1$	$1 - r_2$	$1 - r_2 - s_2$

The equilibrium gene frequencies for large populations are given by $\bar{p} = r_2/(r_1 + r_2)$ at the A locus, and $\bar{q} = s_2/(s_1 + s_2)$ at the B locus. The change in gene frequency at the A locus in one generation is given by $\delta p = -(r_1 + r_2)\, p\,(1 - p)\,(p - \bar{p})$, with a similar equation for locus B, where squared terms in selective values are ignored. On a continuous model it can be shown that, on a time scale proportional to N, the inbreeding and selection process is a function of only $\bar{p}$, $\bar{q}$, $N\,c$, $N\,(r_1 + r_2)$ and $N\,(s_1 + s_2)$ for a given set of initial conditions p_0, q_0 and D_0. No explicit solutions for this model could be obtained, so our Monte Carlo programme was modified to include selection for heterozygotes. The number of replicates used for each set of parameters depended on the rate of fixation observed, and was chosen so that roughly the same number of replicates were segregating at both loci after $4N$ generations as for the case of no selection with $p_0 = q_0 = 0.5$, and 10000 replicates. All simulation was done with $N = 8$, except for one example with $N\,(r_1 + r_2) = N\,(s_1 + s_2) = 4$ and $\bar{p} = \bar{q} = 0.5$ which was also run with $N = 16$ (Figure 4).

Selection is most effective in maintaining heterozygosity when the equilibrium gene frequency is one-half (ROBERTSON 1962; ROBERTSON and HILL, 1968).

[1] $E(D) = 0$ with heterozygote superiority at both loci only if $\bar{p} = 0.5$ and/or $\bar{q} = 0.5$.

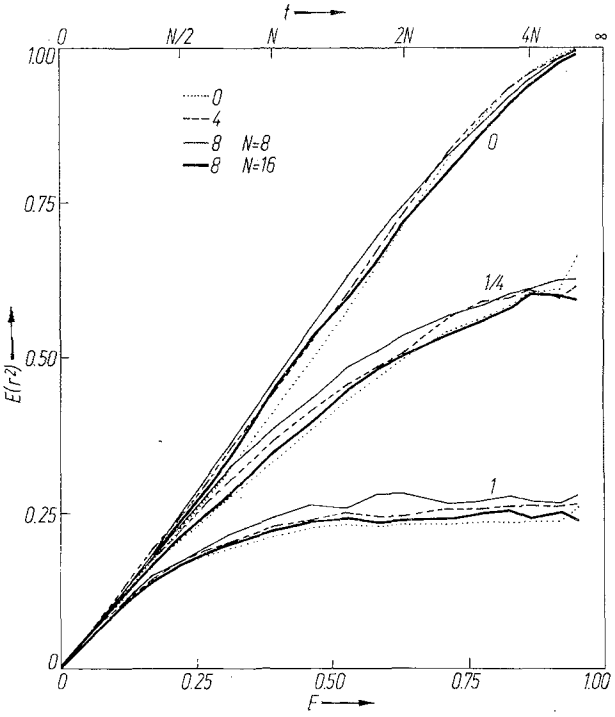

Fig. 4. The effect of selection for heterozygotes on $E(r^2)$ for $N\,(r_1 + r_2) = N\,(s_1 + s_2) = 0$, 4 and 8 and $p_0 = q_0 = \bar{p} = \bar{q} = 0.5$. Populations were simulated with both $N = 8$ and $N = 16$ for $N\,(r_1 + r_2) = 8$, otherwise $N = 8$

We shall therefore discuss this situation ($\bar{p} = \bar{q} = 0.5$) in most detail, and for simplicity assume that $N\,(r_1 + r_2) = N\,(s_1 + s_2)$.

Such selection has two related consequences which are relevant here. Firstly, the rate of fixation may be greatly reduced and secondly, the gene frequency distribution amongst unfixed lines becomes more concentrated around the equilibrium frequencies as selection becomes more intense. We found that $E(D^2)$ over all populations was increased by selection for heterozygotes. This appears to be mainly due to retardation of fixation as the effect on $E(D_s^2)$ in segregating lines, though present, is small.

However, on examining $E(r^2)$, we found that this reaches a limiting value which is little influenced by the intensity of selection (Figure 4). Further it appears that about the same level of $E(r^2)$ is reached when the equilibrium frequency in large populations is not 0.5 (Figures 5 and 6). The curves of $E(r^2)$ against F or t/N, are then dependent almost only on $N\,c$. The limiting value of $E(r^2)$ appears to approach $1/4\,N\,c$ as $N\,c$ increases. A crude derivation can be obtained by equating the loss in $E(D^2)$ each generation ($2\,c\,E(D^2)$ approximately) with the gain due to sampling ($p\,(1 - p)\,q\,(1 - q)/2N$). The second term in the vector y is then small because gene frequencies are close to 0.5.

Discussion

Many workers are now investigating polymorphisms in natural populations, and opportunities will no doubt arise for measuring the linkage disequilibrium between the segregating loci observed. We have used the square of the correlation of gene frequencies, r^2, as our statistic, which has a known sampling distribution when the true value is zero,

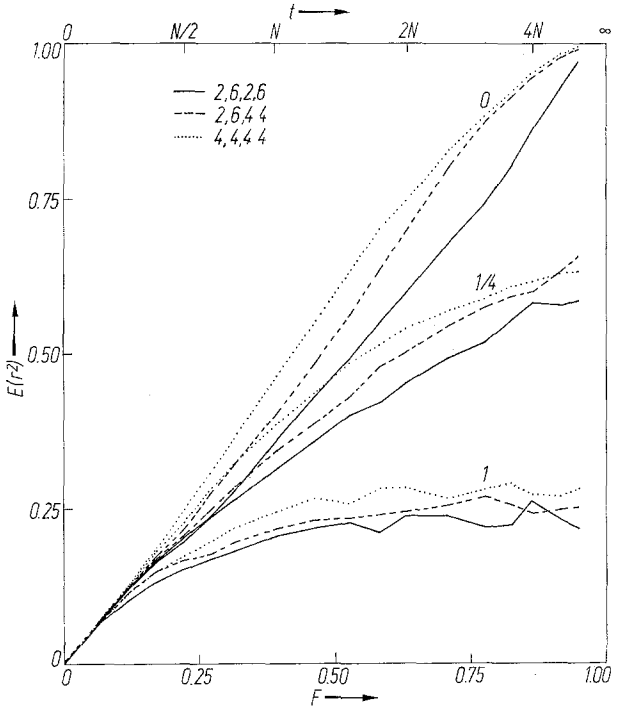

Fig. 5. The effect of equilibrium frequency on $E(r^2)$ with $N(r_1 + r_2) = N(s_1 + s_2) = 4$ and $p_0 = \bar{p}$, $q_0 = \bar{q}$

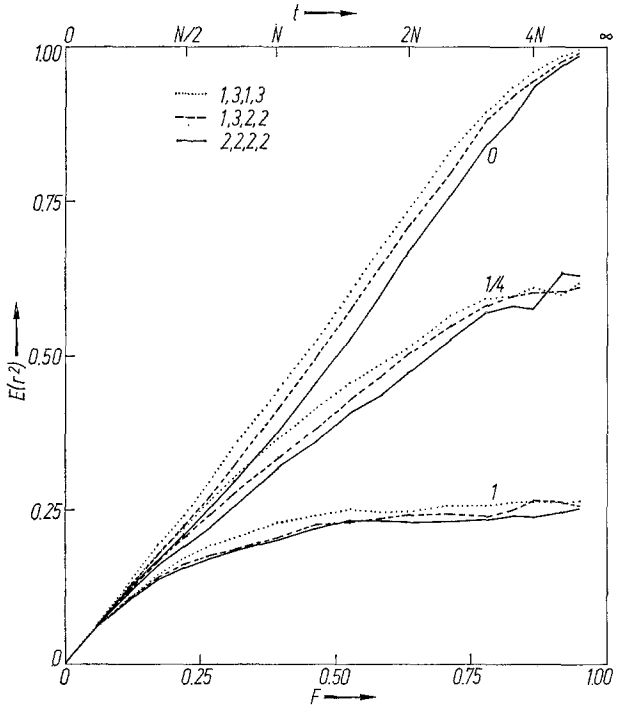

Fig. 6. As Figure 5, but $N(r_1 + r_2) = N(s_1 + s_2) = 8$

If a sample of T individuals are taken from the population, $T r^2$ is then distributed as χ^2 with one degree of freedom.

However our results show that when a significant departure from equilibrium is observed in a small population, we must be cautious about concluding that this is due to natural selection. Several models with interaction of selective advantage between the loci have been investigated in infinite populations. For example, LEWONTIN (1964) studied a two locus model with heterozygote advantage and epistacy, which had relative fitnesses as follows:

	A_1A_1	A_1A_2	A_2A_2
B_1B_1	.4	.6	.3
B_1B_2	.6	1.0	.5
B_2B_2	.5	.7	.4

From his Table 4 we can compute the values of r^2 reached at equilibrium. Two stable situations were possible, in which there was a final excess of either coupling or repulsion phases; we shall use only the latter. The results for the model were:

c	0	.01	.02	.04	.08
r^2	1.000	.799	.601	.221	.002

With no selection, or selection for heterozygotes with no epistacy, the mean value of r^2 within the segregating populations would reach the following approximate values, assuming a population size of $N = 25$ was maintained for many generations:

c	0	.01	.02	.04	.08
$E(r^2)$	1.00	.62	.41	.25	.12

These results correspond to $N c$ values of 0, .25, .5, 1 and 2. Thus two completely different processes lead to superficially similar results. It can be argued that $N = 25$ is much too small to represent a natural population. However, LEWONTIN's selective advantages with differences of factors of two at a single locus may be considered unrealistically large.

The model we have used may also be criticised because of the assumption of constant population size. However this does not effect the qualitative aspects of our results. Any restriction of population size may cause disequilibrium as a result of genetic sampling, and the return to equilibrium will be slow if the loci are tightly linked.

Zusammenfassung

Es wurde eine theoretische Untersuchung über den Einfluß der Populationsgröße (N) und der Rekombinationsfraktion (c) auf das Koppelungs-Ungleichgewicht (D) zwischen einem Paar von Loci angestellt. Die nachfolgenden zwei Situationen wurden studiert:

1. Beide Loci haben keinen Effekt auf die Fitness.

2. Die Heterozygoten zeigen Überlegenheit, jedoch keine Epistasie.

Befinden sich die Populationen in einem ursprünglichen Koppelungsgleichgewicht, so bleibt der mittlere Wert von D bei Inzucht gleich null, jedoch steigt das Mittel von D^2 bis zu einem Maximalwert und fällt, bis die Fixierung an beiden Loci erreicht worden ist. Je enger die Koppelung und je stärker die Selektion ist, desto später wird das Maximum im Mittel von D^2 erreicht und desto größer ist sein Wert. Ferner wurde die Korrelation von Genfrequenzen, r, in der Population von Gameten innerhalb spaltender Linien untersucht. Es wurde gefunden, daß der mittlere Wert von r^2 für einen Bereich von Selek-

tionsintensitäten und ursprünglichen Genfrequenzen praktisch vollkommen bestimmt wurde durch Nc und die Zeit, gemessen proportional zu N. Abschließend wird die Bedeutung dieser Ergebnisse für Beobachtungen von Koppelungs-Ungleichgewichten in natürlichen Populationen diskutiert.

References

1. Bennett, H. J.: On the theory of random mating. Ann. Eugenics 18, 311–317 (1954). — 2. Felsenstein, J.: The effect of linkage on directional selection. Genetics 52, 349–363 (1965). — 3. Geiringer, H.: On the probability theory of linkage in Mendelian heredity. Ann. Math. Statist. 15, 25–57 (1944). — 4. Hill, W. G., and A. Robertson: The effect of linkage on limits to artificial selection. Genet. Res. 8, 269–294 (1966). — 5. Kimura, M.: Random drift in a multi-allelic locus. Evolution 9, 419–435 (1955). — 6. Kimura, M.: A probability method for treating inbreeding systems, especially with linked genes. Biometrics 19, 1–17 (1963). — 7. Lewontin, R. C.: The interaction of selection and linkage. I. General considerations; heterotic models. Genetics 49, 49–67 (1964). — 8. Lewontin, R. C., and K. Kojima: The evolutionary dynamics of complex polymorphisms. Evolution 14, 458–472 (1960). — 9. Nei, M.: Effect of selection on the components of genetic variance. In: Statistical Genetics and Plant Breeding, ed. by W. D. Hanson and H. F. Robinson. Publ. 982, National Academy of Sciences, National Research Council, Washington, D.C., pp. 501–515 (1963). — 10. Robertson, A.: The effect of inbreeding on the variation due to recessive genes. Genetics 37, 189–207 (1952). — 11. Robertson, A.: Selection for heterozygotes in small populations. Genetics 47, 1291–1300 (1962). — 12. Robertson, A., and W. G. Hill: The effects of inbreeding at loci with heterozygote advantage. Genetics (submitted for publication 1968). — 13. Wright, S.: Inbreeding and recombination. Proc. Nat. Acad. Sci., Wash. 19, 420–433 (1933).

Appendix: Diagonalisation of the Moment Generating Matrix with Complete Linkage

We use well known theory to find for the matrix M the scalar latent roots λ_1, λ_2 and λ_3 and their associated latent vectors v_1, v_2, and v_3 of dimension 3 such that

$$M v_i = v_i \lambda_i , \qquad i = 1, 2, 3 .$$

Thus, if we let $\mathbf{\Lambda}$ be a 3×3 diagonal matrix of the latent roots λ_i and let $V = (v_1\, v_2\, v_3)$ be the 3×3 matrix of latent vectors, we have

$$M V = V \mathbf{\Lambda}$$

and

$$M = V \mathbf{\Lambda} V^{-1}$$

also

$$M^2 = V \mathbf{\Lambda} V^{-1} \cdot V \mathbf{\Lambda} V^{-1} = V \mathbf{\Lambda}^2 V^{-1}$$

and so on.

To obtain the moments $y_{(t)}$ we require

$$y_{(t)} = M^t y_{(0)}$$

which is given by

$$y_{(t)} = V \mathbf{\Lambda}^t V^{-1} y_{(0)} \tag{1A}$$

and needs only scalar multiplication to evaluate $\mathbf{\Lambda}^t$.

With complete linkage, M is given by setting $c = 0$ in equation (3) of the text and its latent roots and vectors are easily obtained. These are

$$\lambda_1 = 1 - \frac{1}{n}, \qquad \lambda_2 = \left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right),$$

$$\lambda_3 = \left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right)\left(1 - \frac{3}{n}\right)$$

and

$$V = \begin{pmatrix} 1 & -2 & 1 \\ \frac{n-2}{n-1} & 2 & -4 \\ 1 & 1 & 1 \end{pmatrix}.$$

Since the inbreeding coefficient, F, equals $1 - \left(1 - \frac{1}{n}\right)^t$, it follows that for large n

$$\lambda_1^t = 1 - F , \quad \lambda_2^t = (1 - F)^3 , \quad \lambda_3^t = (1 - F)^6 ,$$

approximately,

$$\text{and } V = \begin{pmatrix} 1 & -2 & 1 \\ 1 & 2 & -4 \\ 1 & 1 & 1 \end{pmatrix}, \text{ approximately.}$$

The inverse of V is then

$$V^{-1} = \frac{1}{15}\begin{pmatrix} 6 & 3 & 6 \\ -5 & 0 & 5 \\ -1 & -3 & 4 \end{pmatrix}$$

If there is initial equilibrium (a restriction not required by the preceding theory)

$$y'_{(0)} = p_0 (1 - p_0)\, q_0 (1 - q_0)\, (1\ 0\ 0)$$

and substitution into (1A) gives the result

$$y_{(t)} = \frac{1}{15} p_0 (1 - p_0)\, q_0 (1 - q_0) \times$$

$$\times \begin{pmatrix} 6(1 - F) + 10(1 - F)^3 - (1 - F)^6 \\ 6(1 - F) - 10(1 - F)^3 + 4(1 - F)^6 \\ 6(1 - F) - 5(1 - F)^3 - (1 - F)^6 \end{pmatrix}.$$