Theoretical and Applied Genetics 42, 335–345 (1972)

The Effects of Finite Population Size and Selection on the Correlation between Gene Frequency Changes at two Different Loci and on the Amount of Linkage Disequilibrium

SUSAN J. GALLEY and R. N. CURNOW

Department of Applied Statistics, University of Reading, Reading, Berkshire (England)

Summary. A potentially infinite random-mating population of monoecious diploid individuals is subjected to a single cycle of sampling and selection based on the values of a quantitative character. In the absence of epistatic interactions, expressions are obtained for the correlation between the gene frequencies at two linked loci and for the mean and variance of linkage disequilibrium after one cycle of selection. Numerical results are presented for a range of population sizes and for various values of the genetic parameters.

1. Introduction

This paper is concerned with the effects of a single cycle of selection based on the value of a quantitative character and applied to a finite population of monoecious diploid individuals. An infinite population, arising from the potentially infinite gametic output of the parents, is sampled to provide N individuals. These individuals are ranked according to the value of the character and the top n individuals are selected to become the parents of the next generation.

Hill (1969) has given formulae for the probability distribution of the genotypic constitution of the population after selection, when the value of the character for each genotype has a general distribution. Kojima (1961) has given approximate expressions for the mean and variance of the change in gene frequency at a single locus, following selection. He assumed that there were two alleles at the locus in question, and that the value of the character for the i^{th} genotype ($i = 1, 2, 3$) was distributed normally with mean μ_i and variance σ^2, the same variance for all three genotypes. Let the probability of an individual from the infinite population being of the i^{th} genotype be a_i, ($i = 1, 2, 3$). Then if $d_i = \mu_i - \sum_{r=1}^{3} a_r \mu_r$, is the deviation of the mean of the i^{th} genotype from the mean of the infinite population, Kojima's results, to order $\frac{k d_i}{\sigma}$, are the same as those of Hill for the same model. Here k is the selection intensity defined by Kojima and is the expected value of the best n values in a sample of size N taken from a normal distribution with zero mean and unit variance. Without loss of generality we shall assume that $\sigma^2 = 1$, and then make the assumption that $k\,d_i$ is sufficiently small that terms of order $(k\,d_i)^2$ and higher can be neglected.

Kojima (1961) showed that if n_i is the number of individuals selected belonging to the i^{th} genotype, $\left(\sum_{i=1}^{3} n_i = n\right)$, then to order $k\,d_i$,

$$\left.\begin{aligned} E(n_i) &= n\,a_i\,(1 + k\,d_i)\,, \\ var\,(n_i) &= n\,a_i\,(1 - a_i) + n\,k\,a_i\,(1 - 2\,a_i)\,d_i \end{aligned}\right\} \quad (1)$$

and

$$cov\,(n_i, n_j) = -\,n\,a_i\,a_j - n\,k\,a_i\,a_j\,(d_i + d_j)\,.$$

If q is the initial frequency of the allele A_1, and q' is the frequency after sampling and selection, then the results above imply that

$$E(q') = q + k\left(a_1 d_1 + \frac{1}{2}\,a_2 d_2\right),$$

and

$$var\,(q') = \frac{q(1-q)}{n} - \frac{2}{4\,n} + \\ + \frac{k}{n}\left(a_1 d_1 + \frac{1}{2}\,a_2 d_2\right)(1 - 2\,q) - \frac{k\,a_2 d_2}{4\,n},$$

where $i = 1$, 2, and 3 refer to the A_1A_1, A_1A_2 and A_2A_2 genotypes respectively.

We shall extend Kojima's results to two loci so that the correlation between the changes in gene frequency at different loci and the expected amount of linkage disequilibrium generated and its variance can be evaluated. The degree of correlation and the amount of disequilibrium generated can be important in evolutionary terms and also in plant and animal breeding programmes because they affect the variance of the change in the mean value of the character as the population undergoes sampling and selection. The variance of the change in mean value is important in designing breeding and selection programmes and in interpreting selection experiments. The dependence of the derived quantities on the initial gene frequencies, the initial linkage disequilibrium, the

selection intensity, the population size, the recombination factor, and the genetic effects at the two loci will be described. We shall assume random mating and no epistatic interactions. The random mating is complete in that selfing is allowed to occur.

2. The Theory for two Loci

Let the alleles at one locus be A_1 and A_2 and at the other B_1 and B_2. Let the gametic frequencies in the parental population be as follows: —

Gamete	A_1B_2	A_1B_2	A_2B_1	A_2B_2
Frequency	P_1	P_2	P_3	P_4

where $P_1 + P_2 + P_3 + P_4 = 1$.

The initial linkage disequilibrium is $D = P_1P_4 - P_2P_3$. Write the gene frequencies of A_1 and B_1 as $X = P_1 + P_2$ and $Y = P_1 + P_3$ respectively, and the recombination fraction as r. Table 1 shows the ten genotypes, separating the two double heterozygotes into the coupling and repulsion phases, and provides a notation for the associated frequencies, the numbers selected and the deviations of the genotypic means from the population mean.

Using primes to indicate frequencies after sampling and selection,

$$\left.\begin{aligned} 2nP_1' &= 2N_1 + N_2 + N_4 + (1-r)N_5 + rN_6 \\ 2nP_2' &= 2N_3 + N_2 + N_7 + rN_5 + (1-r)N_6 , \\ 2nP_3' &= 2N_8 + N_9 + N_4 + rN_5 + (1-r)N_6 , \\ 2nP_4' &= 2N_{10} + N_9 + N_7 + (1-r)N_5 + rN_6 , \end{aligned}\right\} \quad (2)$$

and

$$\begin{aligned} 2nX' &= 2n(P_1' + P_2') = 2(N_1 + N_2 + N_3) \\ &\quad + (N_4 + N_5 + N_6 + N_7) , \\ 2nY' &= 2n(P_1' + P_3') = 2(N_1 + N_4 + N_8) \\ &\quad + (N_2 + N_5 + N_6 + N_9) . \end{aligned}$$

With this notation we require formulae for the covariance of X' and Y' and for the expected value and variance of $D' = P_1'P_4' - P_2'P_3'$.

The set of equations (1) apply to the figures in Table 1 as an obvious extension of their validity for the one locus case. The variance-covariance matrix of the ten element random vector $(N_1, N_2, \ldots, N_{10})$ is therefore

$$n\boldsymbol{A} = n(a_{ij}), \quad i, j = 1, 2, \ldots, 10 ,$$

where

$$a_{ij} = \begin{cases} -[a_i a_j + k a_i a_j (d_i + d_j)] , & i \neq j \\ a_i(1 - a_i) + k a_i (1 - 2a_i) d_i , & i = j . \end{cases}$$

Unlabelled summations will always be over the ten genotypes. Letting $\boldsymbol{X}^T$ denote the transpose of a matrix $\boldsymbol{X}$, the covariance of any two linear forms in the N's is given by

$$\begin{aligned} &Cov(\Sigma\alpha_i N_i, \Sigma\beta_i N_i) = n(\alpha_1, \ldots, \alpha_{10}) \\ &\times \boldsymbol{A}(\beta_1, \beta_2, \ldots, \beta_{10})^T = n(a_1\alpha_1, a_2\alpha_2, \ldots, a_{10}\alpha_{10}) \times \\ &\begin{bmatrix} 1 + k d_1 - a_1(1 + 2k d_1) & -a_2[1 + k(d_1 + d_2)] & \cdots \\ -a_1[1 + k(d_1 + d_2)] & 1 + k d_2 - a_2(1 + 2k d_2) & \cdots \\ \vdots & \vdots & \cdots \\ \vdots & \vdots & \cdots \\ \vdots & \vdots & \cdots \end{bmatrix} \times \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \vdots \\ \beta_{10} \end{bmatrix} \\ &= n(a_1\alpha_1, a_2\alpha_2, \ldots, a_{10}\alpha_{10})\, \boldsymbol{B}(\beta_1, \beta_2, \ldots, \beta_{10})^T \\ &+ n(a_1\alpha_1, a_2\alpha_2, \ldots, a_{10}\alpha_{10})\, \boldsymbol{C}(a_1\beta_1, a_2\beta_2, \ldots, a_{10}\beta_{10})^T \end{aligned}$$

where $\boldsymbol{B} = (b_{ij})$ and $\boldsymbol{C} = (c_{ij})$, $i, j = 1, 2, \ldots 10$, with

$$b_{ij} = \begin{cases} 1 + k d_i & i = j , \\ 0 & i \neq j , \end{cases}$$

and

$$c_{ij} = -[1 + k(d_i + d_j)] .$$

Thus

$$\begin{aligned} Cov(\Sigma\alpha_i N_i, \Sigma\beta_i N_i) &= n\Sigma a_i\alpha_i\beta_i + nk\Sigma a_i\alpha_i\beta_i d_i \\ &\quad - n\Sigma a_i\alpha_i \Sigma a_j\beta_j - kn\Sigma a_i\alpha_i \Sigma a_j\beta_j d_j \\ &\quad - kn\Sigma a_j\beta_j \Sigma a_i\alpha_i d_i . \end{aligned} \quad (3)$$

Substituting $d_i = g_i - \mu$, where g_i is the mean of the i^{th} genotype and $\mu = \Sigma a_i g_i$, is the population mean

Table 1. *Notation*

Genotype a	Initial frequencies a_i	Number selected N_i	Deviation of genotypic mean from population mean $d_i = g_i - \mu$	α_i and β_i required for covariance calculation α_i	β_i
$A_1A_1B_1B_1$	P_1^2	N_1	$\theta_1 + \theta_2 - \mu$	1	1
$A_1A_1B_1B_2$	$2P_1P_2$	N_2	$\theta_1 + \Phi_2 - \mu$	1	1/2
$A_1A_1B_2B_2$	P_2^2	N_3	$\theta_1 - \theta_2 - \mu$	1	0
$A_1A_2B_1B_1$	$2P_1P_3$	N_4	$\Phi_1 + \theta_2 - \mu$	1/2	1
A_1B_1/A_2B_2	$2P_1P_4$	N_5	$\Phi_1 + \Phi_2 - \mu$	1/2	1/2
A_1B_2/A_2B_1	$2P_2P_3$	N_6		1/2	1/2
$A_1A_2B_2B_2$	$2P_2P_4$	N_7	$\Phi_1 - \theta_2 - \mu$	1/2	0
$A_2A_2B_1B_1$	P_3^2	N_8	$-\theta_1 + \theta_2 - \mu$	0	1
$A_2A_2B_1B_2$	$2P_3P_4$	N_9	$-\theta_1 + \Phi_2 - \mu$	0	1/2
$A_2A_2B_2B_2$	P_4^2	N_{10}	$-\theta_1 - \theta_2 - \mu$	0	0
	Total = 1	Total = n			

in (3) gives

$$Cov\,(\Sigma \alpha_i N_i, \Sigma \beta_i N_i) = n\,[\Sigma a_i \alpha_i \beta_i - \Sigma a_i \alpha_i \Sigma a_j \beta_j] + k\,n\,[\Sigma a_i \alpha_i \beta_i g_i - \Sigma a_i \alpha_i \Sigma a_j \beta_j g_j - \Sigma a_j \beta_j \Sigma a_i \alpha_i g_i] - k\,n\,\mu\,[2 \Sigma a_i \alpha_i \Sigma a_j \beta_j - \Sigma a_i \alpha_i \beta_i]\,.$$

Table 1 also gives the values of α_i and β_i, which, together with the a_i and g_i, are needed for calculating $cov\,(X', Y')$, the covariance of the changes in gene frequencies at the two loci. The initial population is assumed to have been formed by random mating and there are assumed to be no epistatic effects. The additive effects of two loci on the character are written θ_1 and θ_2 and the dominance effects are written Φ_1 and Φ_2.

After considerable algebra, substitution from Table 1 leads to

$$cov\,(X', Y') = \frac{D}{2n} + \frac{k\,D}{2n}\{\theta_1\,(1 - 2X) + \theta_2\,(1 - 2Y) + \Phi_1\,[1 - 6X\,(1 - X)] + \Phi_2\,[1 - 6Y\,(1 - Y)]\}\,. \quad (4)$$

Kojima (1961) has already given, in one locus form, the formulae for the variances but they can be more easily obtained here by substituting $\alpha_i = \beta_i$, $(i = 1, 2, \ldots, 10)$ to give

$$Var\,(X') = \frac{X\,(1 - X)}{2n} + \frac{k}{2n}\{\theta_1 X\,(1 - X)\,(1 - 2X) + \theta_2 D\,(1 - 2X) + \Phi_1 X\,(1 - X)\,[1 - 6X\,(1 - X)] + \Phi_2\,[D\,(1 - 2X)\,(1 - 2Y) - 2D^2]\}$$

and

$$Var\,(Y') = \frac{Y\,(1 - Y)}{2n} + \frac{k}{2n}\{\theta_1 D\,(1 - 2Y) + \theta_2 Y\,(1 - Y)\,(1 - 2Y) + \Phi_1\,[D\,(1 - 2X)\,(1 - 2Y) - 2D^2] + \Phi_2 Y\,(1 - Y) \times [1 - 6Y\,(1 - Y)]\}\,. \quad (5)$$

Formulae (4) and (5) can be used to obtain

$$\varrho = \text{correlation coefficient of } X' \text{ and } Y' = \frac{Cov\,(X', Y')}{\sqrt{Var\,(X')\,Var\,(Y')}}$$

Clearly ϱ is independent of n and, for obvious reasons, of r. Ignoring terms of order $(k\,d_i)^2$

$$\varrho = \frac{D}{[X\,Y\,(1 - X)\,(1 - Y)]^{1/2}}\left\{1 + k\,\theta_1\left[\frac{1}{2} - X - \frac{D\,(1 - 2Y)}{2Y\,(1 - Y)}\right] + k\,\theta_2\left[\frac{1}{2} - Y - \frac{D\,(1 - 2X)}{2X\,(1 - X)}\right] + k\,\Phi_1\left[\frac{1}{2} - 3X\,(1 - X) - \frac{D^2}{Y\,(1 - Y)} - \frac{D\,(1 - 2X)\,(1 - 2Y)}{2Y\,(1 - Y)}\right] + k\,\Phi_2\left[\frac{1}{2} - 3Y\,(1 - Y) - \frac{D^2}{X\,(1 - X)} - \frac{D\,(1 - 2X)\,(1 - 2Y)}{2X\,(1 - X)}\right]\right\}. \quad (6)$$

The variances and covariance of the changes in gene frequencies at the two loci are the expected values of only three of all the possible quadratic forms in the genotypic numbers $N_1, N_2, \ldots, N_{10}$. The linkage disequilibrium $D' = P_1' P_4' - P_2' P_3'$ is another. The covariances of P_1' and P_4' and of P_2' and P_3' can be evaluated by the methods used above and hence the expected values of $P_1' P_4'$ and $P_2' P_3'$ derived to give

$$E(D') = \frac{1}{2} D\,(1 - 2r)\,[1 + k\,(\Phi_1 + \Phi_2 - \mu)] + \frac{1}{2} D\left(1 - \frac{1}{n}\right)[1 + k\,\Phi_1\,(1 - 2X)^2 + k\,\Phi_2\,(1 - 2Y)^2 - k\,\mu]\,, \quad (7)$$

where μ is the population mean before sampling and selection, viz.

$$\mu = \theta_1\,(2X - 1) + \theta_2\,(2Y - 1) + \Phi_1\,2X\,(1 - X) + \Phi_2\,2Y\,(1 - Y)\,.$$

The effect of selection on $E(D')$ is again evident in the addition of terms with a factor k to the value of $E(D')$ due to sampling alone which is

$$E(D') = D\left(1 - r - \frac{1}{2n}\right).$$

Hill and Robertson (1966 and 1968) and Karlin and McGregor (1968) have both given the sampling expectation of D' as

$$E(D') = D\,(1 - r)\left(1 - \frac{1}{2n}\right).$$

The difference between our formula and theirs is due to a difference in the underlying models which have been termed "random union of zygotes" and "random union of gametes" respectively. Ohta (1968) and Watterson (1970) have both discussed the difference. Briefly, the model which we have used, random union of zygotes, takes into account the formation of individuals from the gametic output of the parents, and is more correct biologically. The second model only considers sampling from one gametic pool to form another in the next generation and does not allow selection to operate on the actual individuals formed between the two gametic productions.

3. The Dependence of ϱ on the Genetic Parameters

The limits of D for given values of the gene frequencies X and Y correspond to the situations when at least one gamete is missing from the population. Thus D is maximum, $D = P_1 P_4$ when

$$P_2 = X\,(1 - Y) - D = 0 \quad \text{or}$$
$$P_3 = Y\,(1 - X) - D = 0\,,$$

and D is minimum, $D = -P_2 P_3$ when

$$P_1 = X\,Y + D = 0 \quad \text{or}$$
$$P_4 = (1 - X)\,(1 - Y) + D = 0\,.$$

Table 2. *Values of ρ for the six genetic models, given x, y and D, with $k\theta_1 = 0$ and 0.25*

		$x = 1/4$				$x = 1/2$				$x = 3/4$			
	D	−0.063	−0.031	0	0.094	0.188	−0.125	−0.063	0	0.063	0.125	−0.188	−0.094	0	0.031	0.063
	−/−	−0.333	−0.167	0	0.500	1.000	−0.577	−0.289	0	0.289	0.577	−1.000	−0.500	0	0.167	0.333
	a/−	−0.361	−0.179	0	0.515	1.000	−0.603	−0.295	0	0.283	0.555	−1.000	−0.483	0	0.154	0.306
$y = 1/4$	d/−	−0.362	−0.177	0	0.505	1.000	−0.577	−0.277	0	0.266	0.529	−1.000	−0.472	0	0.153	0.305
	a/a	−0.385	−0.189	0	0.526	1.000	−0.636	−0.312	0	0.301	0.591	−1.000	−0.501	0	0.168	0.338
	d/a	−0.386	−0.188	0	0.518	1.000	−0.617	−0.298	0	0.287	0.571	−1.000	−0.491	0	0.167	0.338
	d/d	−0.388	−0.187	0	0.510	1.000	−0.619	−0.294	0	0.283	0.571	−1.000	−0.481	0	0.165	0.339
	D	−0.125	−0.063	0	0.063	0.125	−0.25	−0.125	0	0.125	0.25	−0.125	−0.063	0	0.063	0.125
	−/−	−0.577	−0.289	0	0.289	0.577	−1.000	−0.500	0	0.500	1.000	−0.577	−0.289	0	0.289	0.577
	a/−	−0.612	−0.306	0	0.306	0.612	−1.000	−0.500	0	0.500	1.000	−0.540	−0.270	0	0.270	0.540
$y = 1/2$	d/−	−0.613	−0.303	0	0.303	0.613	−1.000	−0.475	0	0.475	1.000	−0.539	−0.266	0	0.266	0.539
	a/a	−0.636	−0.312	0	0.306	0.591	−1.000	−0.500	0	0.500	1.000	−0.516	−0.264	0	0.277	0.568
	d/a	−0.638	−0.309	0	0.297	0.591	−1.000	−0.475	0	0.475	1.000	−0.514	−0.260	0	0.273	0.568
	d/d	−0.619	−0.294	0	0.283	0.571	−1.000	−0.444	0	0.444	1.000	−0.480	−0.238	0	0.250	0.533
	D	−0.188	−0.094	0	0.031	0.063	−0.125	−0.063	0	0.063	0.125	−0.063	−0.031	0	0.094	0.188
	−/−	−1.000	−0.500	0	0.167	0.333	−0.577	−0.289	0	0.289	0.577	−0.333	−0.167	0	0.500	1.000
	a/−	−1.000	−0.515	0	0.179	0.361	−0.555	−0.283	0	0.295	0.603	−0.306	−0.154	0	0.483	1.000
$y = 3/4$	d/−	−1.000	−0.505	0	0.177	0.362	−0.529	−0.266	0	0.277	0.577	−0.305	−0.153	0	0.472	1.000
	a/a	−1.000	−0.501	0	0.168	0.338	−0.516	−0.264	0	0.277	0.568	−0.273	−0.140	0	0.462	1.000
	d/a	−1.000	−0.491	0	0.167	0.338	−0.483	−0.243	0	0.255	0.535	−0.271	−0.137	0	0.449	1.000
	d/d	−1.000	−0.481	0	0.165	0.339	−0.480	−0.238	0	0.250	0.533	−0.268	−0.135	0	0.436	1.000

The correlation between X' and Y', the gene frequencies of A_1 and B_1 respectively in the generation after sampling and selection, can become quite large in amount (positive or negative in sign) if D is near one of its limits, i.e. when one of the gametic frequencies in the parent generation is near to zero.

Complete positive correlation is achieved when both P_1 and P_3 are zero and $D = P_1 P_4$. The correlation is large if $D > \frac{1}{2} P_1 P_4$. Complete negative correlation occurs when both P_1 and P_4 are zero and $D = -P_2 P_3$, $|\varrho|$ again becoming if $D < -\frac{1}{2} P_1 P_3$.

Hence correlations can be expected to approach ± 1 if either both coupling or both repulsion gametes have very small frequencies in the parent generation, and quite large values of $|\varrho|$ can be expected if any one gametic frequency is close to zero.

The formulae derived in this paper all assume that the $k\theta$ and $k\Phi$ terms are small so that terms of order $(k\theta)^2$ and $(k\Phi)^2$ can be neglected. Latter (1965) discusses this approximation. It would appear that our results are probably reasonably accurate for $k\theta$ and $k\Phi$ less than about 0.4. In Table 2 we quote results for $k\theta_1 = 0$ and 0.25. To our order of approximation all functions are linear in $k\theta$ and so intermediate results can be obtained by linear interpolation. Table 2 gives the values of ϱ calculated from formulae (4) and (5) for a range of values of X and Y and for a range of values of D, from its minimum to its maximum value. The genetic models considered and the abbreviations used in Table 2 are as follows.

1. −/−: no selection, (or, equivalently, genes of no effect); $k\theta_i = k\Phi_i = 0$, $i = 1, 2$.
2. a/−: additive/no effect; $\theta_2 = \Phi_1 = \Phi_2 = 0$
3. d/−: dominant/no effect; $\theta_1 = \Phi_1$, $\theta_2 = \Phi_2 = 0$
4. a/a: additive/additive; $\theta_1 = \theta_2$, $\Phi_1 = \Phi_2 = 0$
5. d/a: dominant/additive; $\theta_1 = \theta_2 = \Phi_1$, $\Phi_2 = 0$
6. d/d: dominant/dominant; $\theta_i = \Phi_i$, $i = 1, 2$.

The covariance of X and Y is clearly zero if X or Y is 0 or 1. The values of ϱ in this table are given in the order above. It is evident that for the different genetic models there is little change in the size of ϱ, and hence that its value depends mainly on the amount of sampling and selection and not on the genetic model. Except when the initial frequencies at one or both loci are small, the effect of selection is to decrease the amount of correlation below the value due to sampling alone.

Although, as mentioned above, different genetic models have little effect on the value of ϱ we shall now consider the different genetic models in more detail. We shall call the sampling correlation ϱ_0, i.e. $\varrho_0 = \frac{D}{[X\,Y\,(1-X)\,(1-Y)]^{1/2}}$. The signs of ϱ_0 and ϱ are both determined by the sign of D. From equation (6), the relative change in ϱ_0 due to selection can, to the

first order of $k\,\theta_1$, be considered as arising from

1. an additive effect $k\,\theta_1\left[\frac{1}{2} - X - \frac{D\,(1-2\,Y)}{2\,Y\,(1-Y)}\right]$

and

2. a dominance effect $k\,\Phi_1\left[\frac{1}{2} - 3\,X\,(1-X)\right.$

$$\left. - \frac{D^2}{Y\,(1-Y)} - \frac{D\,(1-2\,X)\,(1-2\,Y)}{2\,Y\,(1-Y)}\right],$$

with similar effects due to the second locus.

The additive effect consists, at the first locus, of i) the 'direct additive effect', $\frac{1}{2} - X$, depending on the gene frequency at the locus in question

and ii) the 'cross additive effect', $\frac{-D\,(1-2\,Y)}{2\,Y\,(1-Y)}$, depending on both D and the gene frequency at the other locus.

Considering the first locus, the direct additive effect will be largest when X is small and will cause the greatest decrease in ϱ_0 when X is large. The cross additive effect will cause the greatest increase in ϱ if X and Y are both large and D is near maximum, or if X is large, Y is small, and D is near minimum.

Thus the greatest increase in ϱ from the additive genetic effects at both loci will be when X and Y are both small, due to the direct additive effects. If D is then also small, the cross additive effects will also tend to increase ϱ. The greatest decrease in ϱ due to the additive genetic effects will occur when X and Y are both large. At neither locus will the cross additive effects be large, but if D is near its maximum these effects will reinforce the depression of ϱ.

The dominance effect at the first locus consists likewise of

i) a direct dominance effect, $\frac{1}{2} - 3\,X\,(1-X)$, depending on X alone,

and ii) a cross dominance effect,

$$\frac{-D^2}{Y\,(1-Y)} - \frac{D\,(1-2\,X)\,(1-2\,Y)}{2\,Y\,(1-Y)},$$

depending on X, Y and D. Similar results hold for the second locus.

The direct effect is greatest when X approaches 0 or 1 and has a minimum of $-\frac{1}{4}$ when $X = \frac{1}{2}$. The cross effect may become large (and negative) if the second term is positive in sign. If X and Y are both small and D is near its minimum the overall dominance effect due to both loci will be positive and give the greatest increase in ϱ. The decreasing effects will be greatest when D is near its maximum and X and Y are both close to but less than $\frac{1}{2}$, or when D is near minimum and X and Y are again close to, but on opposite sides, of $\frac{1}{2}$.

Situations where one locus has a large and the other a small gene frequency will cause least change from the value ϱ_0 when the loci are dominant/additive. When the gene frequencies are of the same order of size, this model behaves more as the dominant/dominant and additive/additive models.

Apart from the situations (i) $X = Y = \frac{1}{4}$, (ii) $X = \frac{1}{4}$, $Y = \frac{1}{4}$, $D < 0$ and (iii) $X = \frac{1}{2}$, $Y = \frac{1}{4}$, $D < 0$, it can be seen from Table 2 that selection causes a decrease below $|\varrho_0|$ in nearly every situation. Consideration of additive and dominance effects show that where the change is an increase above $|\varrho_0|$, the additive/additive model causes the greatest increase, and where the change is a decrease, the dominant/dominant model is responsible for the greatest depression.

4. The Effects of Selection on $E(D')$

When there is no selection, ($k = 0$), the expected value of linkage disequilibrium after sampling is

$$E(D') = D\left(1 - r - \frac{1}{2\,n}\right).$$

Clearly sampling will reduce, eventually to zero, the expected amount of disequilibrium, and this reduction will be fastest for small populations and when the loci are not tightly linked.

With selection following sampling, rearrangement of equation (7) gives, to the first order of $k\,\theta_1$,

$$\begin{aligned}\frac{E(D')}{D} = &\left(1 - r - \frac{1}{2\,n}\right)\{1 - k\,\theta_1\,(2\,X-1) \\ &- k\,\theta_2\,(2\,Y-1) + k\,\Phi_1[1 - 2\,X\,(1-X)] \\ &+ k\,\Phi_2\,[1 - 2\,Y\,(1-Y)]\} \\ &- \left(1 - \frac{1}{n}\right)\{k\,\Phi_1\,2\,X\,(1-X) \\ &+ k\,\Phi_2\,2\,Y\,(1-Y)\}. \qquad (8)\end{aligned}$$

The righthand side of this expression is independent of D and gives the factor by which D is expected to change. Values of this factor were calculated for different fractions, ($r = 0$, 0.1 and 0.5) and X and Y taking all values on the set $\left\{\frac{1}{4}, \frac{1}{2}, \frac{3}{4}\right\}$. $k\,\theta_1$ takes the values 0 and 0.25. Table 3 shows these values when $n = 4$ and $n = 32$ respectively, these being well representative of the effect of increasing n. D, and hence $E(D')$, will be zero when either X or Y is zero.

From the tables it is evident that the factor is increased by increasing the combined selection and genetic effects, $k\,\theta_1$, by increasing the population size and by tighter linkage.

The additive effect, $k\,\theta_1\left(1 - r - \frac{1}{2\,n}\right)(1 - 2\,X)$ at the A-locus, will decrease the factor when X is

Table 3. *Values of $E(D')/D$. Population sizes $n = 4$ and 32. $k\,\theta_1 = 0$ and 0.25*

			$x = 1/4$		$x = 1/2$		$x = 3/4$	
	r	Model	$n = 4$	$n = 32$	$n = 4$	$n = 32$	$n = 4$	$n = 32$
$y = 1/4$	0	−/−	0.875	0.984	0.875	0.984	0.875	0.984
		a/−	0.984	1.107	0.875	0.984	0.766	0.861
		d/−	1.051	1.170	0.891	0.986	0.832	0.924
		a/a	1.094	1.230	0.984	1.107	0.875	0.984
		d/a	1.160	1.293	1.000	1.109	0.941	1.047
		d/d	1.227	1.356	1.066	1.172	1.008	1.110
	0.1	−/−	0.775	0.884	0.775	0.884	0.775	0.884
		a/−	0.872	0.995	0.775	0.884	0.678	0.774
		d/−	0.923	1.042	0.778	0.874	0.729	0.821
		a/a	0.969	1.105	0.872	0.995	0.775	0.884
		d/a	1.020	1.153	0.875	0.984	0.826	0.932
		d/d	1.070	1.200	0.926	1.032	0.877	0.979
	0.5	−/−	0.375	0.484	0.375	0.484	0.375	0.484
		a/−	0.422	0.545	0.375	0.484	0.328	0.424
		d/−	0.410	0.530	0.328	0.424	0.316	0.409
		a/a	0.469	0.605	0.422	0.545	0.375	0.484
		d/a	0.457	0.590	0.375	0.484	0.363	0.469
		d/d	0.445	0.575	0.363	0.469	0.352	0.454
$y = 1/2$	0	−/−	0.875	0.984	0.875	0.984	0.875	0.984
		a/−	0.984	1.107	0.875	0.984	0.766	0.861
		d/−	1.051	1.170	0.891	0.986	0.832	0.924
		a/a	0.984	1.107	0.875	0.984	0.766	0.861
		d/a	1.051	1.170	0.891	0.986	0.832	0.924
		d/d	1.066	1.172	0.906	0.988	0.848	0.926
	0.1	−/−	0.775	0.884	0.775	0.884	0.775	0.884
		a/−	0.872	0.995	0.775	0.884	0.678	0.774
		d/−	0.923	1.042	0.778	0.874	0.729	0.821
		a/a	0.872	0.995	0.775	0.884	0.678	0.774
		d/a	0.923	1.042	0.778	0.874	0.729	0.821
		d/d	0.926	1.032	0.781	0.863	0.732	0.811
	0.5	−/−	0.375	0.484	0.375	0.484	0.375	0.484
		a/−	0.422	0.545	0.375	0.484	0.328	0.424
		d/−	0.410	0.528	0.328	0.424	0.316	0.409
		a/a	0.422	0.545	0.375	0.484	0.328	0.424
		d/a	0.410	0.528	0.328	0.424	0.316	0.409
		d/d	0.363	0.469	0.281	0.363	0.270	0.348
$y = 3/4$	0	−/−	0.875	0.984	0.875	0.984	0.875	0.984
		a/−	0.984	1.107	0.875	0.984	0.766	0.861
		d/−	1.051	1.170	0.891	0.986	0.832	0.924
		a/a	0.875	0.984	0.766	0.861	0.656	0.738
		d/a	0.941	1.047	0.781	0.863	0.723	0.801
		d/d	1.008	1.110	0.848	0.926	0.789	0.864
	0.1	−/−	0.775	0.884	0.775	0.884	0.775	0.384
		a/−	0.872	0.995	0.775	0.884	0.678	0.774
		d/−	0.923	1.042	0.778	0.874	0.729	0.821
		a/a	0.775	0.884	0.678	0.774	0.581	0.663
		d/a	0.826	0.932	0.681	0.763	0.632	0.711
		d/d	0.877	0.979	0.732	0.811	0.683	0.758
	0.5	−/−	0.375	0.484	0.375	0.484	0.375	0.484
		a/−	0.422	0.545	0.375	0.484	0.328	0.424
		d/−	0.410	0.530	0.328	0.424	0.316	0.409
		a/a	0.375	0.484	0.328	0.424	0.281	0.363
		d/a	0.363	0.469	0.231	0.363	0.270	0.348
		d/d	0.352	0.454	0.270	0.348	0.258	0.333

large, i.e. near 1, and increase it when X is small, having no effect when $X = 1/2$. The dominance effect, $k\,\Phi_1\left\{\left(1 - r - \frac{1}{2n}\right)[1 - 2X(1-X)] - \left(1 - \frac{1}{n}\right)2X(1-X)\right\}$ at the A-locus, has a minimum when $X = \frac{1}{2}$, the effect increasing as X

decreases to 0 or increases to 1. The dominance effect increases the factor, but the greatest increase occurs when X is large or small.

Comparing the factor $E(D')/D$ with that when $k = 0$, (-/- model), higher gene frequencies with selection cause a greater reduction and lower gene frequencies cause less reduction. When there is selection, and X and Y are small, linkage is tight and the population size is sufficiently large, $E(D')$ can be greater than D in size. This cannot happen without selection.

5. The Variance of Linkage Disequilibrium, *var (D')*

Knowing $E(D')$, the variance of D' can be derived if an expression for $E(D'^2)$ can be obtained. This quantity,

$$E(D'^2) = E\left[(P_1' P_4')^2 + (P_2' P_3')^2 - 2 P_1' P_2' P_3' P_4'\right],$$

involves the fourth order moments of $N_1, \ldots, N_{10}$.

Watterson (1970) obtained an expression for $E(D'^2)$ when there is sampling only, by means of the moment generating function of the distribution of (P_1', P_2', P_3', P_4'), given (P_1, P_2, P_3, P_4) in the parent generation. After some rearrangement, Watterson's expression becomes

$$\begin{aligned} E(D'^2) = {} & \frac{n(n-1)(n-2)(n-3)}{n^4}(1-r)^2 D^2 \\ & + \frac{n(n-1)(n-2)}{2n^4}[S - Z + 7D^2 \\ & + r(-2S + 2Z - 16D^2) \\ & + r^2(S + 10D^2)] + \frac{n(n-1)}{4n^4} \times \\ & \times [4S - 3Z + 6D^2 + r(-10S \\ & + 8Z - 16D^2) + r^2(6S + 16D^2)] \\ & + \frac{n}{8n^4}(1-2r)^2 S, \end{aligned}$$

where, following Watterson, $Z = X(1-X)Y(1-Y)$ and $S = P_1 P_4 + P_2 P_3 = 2Z + 2D^2 + D(1-2X)(1-2Y)$.

To obtain an expression for $E(D'^2)$ when selection follows sampling, the moment generating function of the distribution of (P_1', P_2', P_3', P_4') was obtained, the differentiation and subsequent algebra proving less formidable than the manipulation of the fourth order moments of $N_1, \ldots, N_{10}$, referred to above.

The numbers of the different genotypes selected are multinomially distributed, the probability associated with n_i being $a_i(1 + k d_i)$, $i = 1, \ldots, 10$, where a_i and d_i are given in Table 1. Thus

$$prob\,[N_1 \cdots N_{10} | P_1, P_2, P_3, P_4] = \frac{n!\,\pi_1^{N_1}\pi_2^{N_2}\cdots\pi_{10}^{N_{10}}}{N_1!\,N_2!\cdots N_{10}!}$$

where $\pi_i = a_i(1 + k d_i)$. The moment generating function for (P_1', P_2', P_3', P_4') is

$$M(t_1, t_2, t_3, t_4) = E\left[e^{\sum_{i=1}^{4} t_i P_i'} | P_1, P_2, P_3, P_4\right]$$

The P_i', $i = 1 \ldots 4$, are given in terms of $N_1, \ldots, N_{10}$ in equation (2). Substitution and rearrangement gives

$$M(t_1, t_2, t_3, t_4) = E\left[e^{\sum_{i=1}^{10} N_i u_i} | P_1, P_2, P_3, P_4\right]$$

$= (\pi_1 e^{u_1} + \cdots + \pi_{10} e^{u_{10}})^n$ the m.g.f. of $N_1, \ldots, N_{10}$, where

$n u_1 = t_1$, $2n u_2 = t_1 + t_2$, $n u_3 = t_2$, $2n u_4 = t_1 + t_3$,
$2n u_5 = t_1 + t_4 + r(t_2 + t_3 - t_1 - t_4)$,
$2n u_6 = t_2 + t_3 - r(t_2 + t_3 - t_1 - t_4)$
$2n u_7 = t_2 + t_4$, $n u_8 = t_3$, $2n u_9 = t_3 + t_4$ and
$n u_{10} = t_4$.

Writing $q_i = P_i e^{t_i/2n}$, $i = 1 \cdots 4$, and substituting for the d_i, further rearrangement gives

$$\begin{aligned} & M(t_1 \cdots t_4 | P_1 \cdots P_4) \\ & = \{[(1 - k\mu)(q_1 + q_2 + q_3 + q_4)^2 - 2(q_1 q_4 + q_2 q_3)] \\ & \quad + [1 - k\mu + k(\Phi_1 + \Phi_2)] \\ & \quad \times \left[2(q_1 q_4 + q_2 q_3)\cosh\frac{r}{2n}(t_2 + t_3 - t_1 - t_4)\right. \\ & \quad \left. + 2(q_1 q_4 - q_2 q_3)\sinh\frac{r}{2n}(t_2 + t_3 - t_1 - t_4)\right] \\ & \quad + k\theta_1[(q_1 + q_2)^2 - (q_3 + q_4)^2] \\ & \quad + k\theta_2[(q_1 + q_3)^2 - (q_2 + q_4)^2] \\ & \quad + 2k\Phi_1(q_1 q_3 + q_2 q_4) + 2k\Phi_2(q_1 q_2 + q_3 q_4)\}^n. \end{aligned}$$

When $k = 0$, this expression reduces to that given by Watterson. After differentiation and much manipulation, an expression is obtained for $E(D'^2)$.

This, together with $[E(D')]^2$ obtained by squaring equation (8) gives the following expression for var D', ignoring terms of order $(k\theta_1)^2$ and higher.

$$Var\,D' = (n^{-1}\, n^{-2}\, n^{-3})\, \boldsymbol{V}\, (1\; r\; r^2)^T, \qquad (9)$$

where $\boldsymbol{V} = (v_{ij})$ is a net italic 3×3 matrix with the following elements in which

$$\begin{aligned} W &= D(1-2X)(1-2Y), \\ A &= DX(1-X)(1-2Y), \\ B &= DY(1-Y)(1-2X). \end{aligned}$$

$$\begin{aligned} v_{11} = {} & \frac{1}{2}\{Z + W - D^2 + [k\theta_1(1-2X) \\ & + k\theta_2(1-2Y)][Z - W - 4D^2] \\ & - k\theta_1 A - k\theta_2 B \\ & + (k\Phi_1 + k\Phi_2)(Z + W - 6D^2) \\ & + [k\Phi_1 X(1-X) + k\Phi_2 Y(1-Y)] \\ & \times [24D^2 - 4Z - 5W]\}, \end{aligned}$$

$$\begin{aligned} v_{12} = {} & D^2 - Z - W + [k\theta_1(1-2X) \\ & + k\theta_2(1-2Y)][4D^2 - Z] + k\theta_1 A \\ & + k\theta_2 B + (k\Phi_1 + k\Phi_2)(4D^2 - Z - W) \\ & + [k\Phi_1 X(1-X) + k\Phi_2 Y(1-Y)] \\ & \times [Z + 3W - 14D^2], \end{aligned}$$

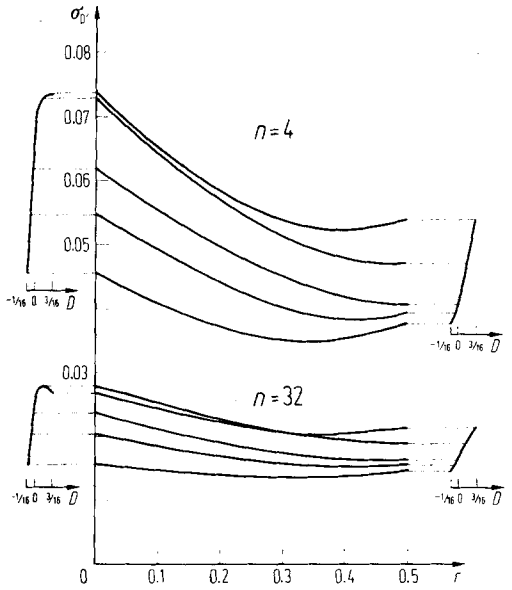

Fig. 1. The relationship between $\sigma_{D'}$, and r for different initial D; no selection model; initial gene frequencies, $X = 1/4$, $Y = 1/4$

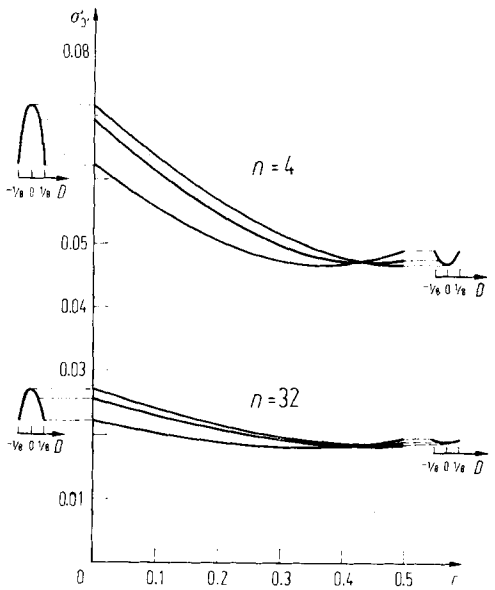

Fig. 2. The relationship between $\sigma_{D'}$, and r for different initial D; no selection model; initial gene frequencies, $X = 1/4$, $Y = 1/2$

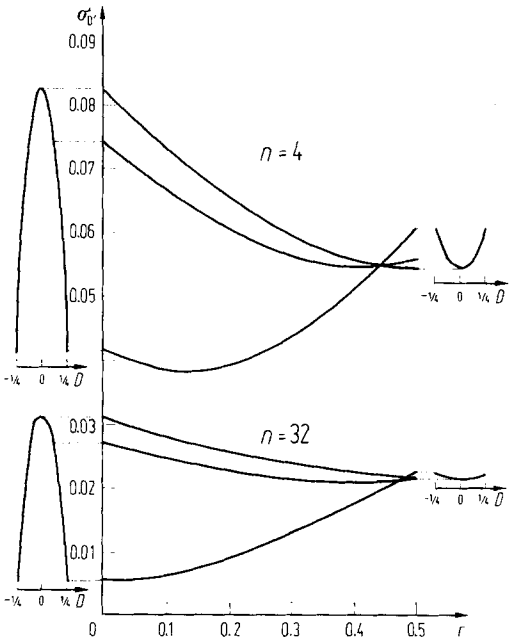

Fig. 3. The relationship between $\sigma_{D'}$, and r for different initial D; no selection model; initial gene frequencies $X = 1/2$, $Y = 1/2$

$$v_{13} = \frac{1}{2}\{2Z + W + [k\theta_1(1-2X) + k\theta_2(1-2Y)][2Z + W - 2D^2] + (k\Phi_1[1-2X(1-X)] + k\Phi_2[1-2Y(1-Y)]) \times (2Z + W - 2D^2)\},$$

$$v_{21} = \frac{1}{4}\{3D^2 - Z - 2W + [k\theta_1(1-2X) + k\theta_2(1-2Y)][10D^2 - Z - 2W] + 3(k\theta_1 A + k\theta_2 B) + (k\Phi_1 + k\Phi_2)(18D^2 - Z - 2W) + [k\Phi_1 X(1-X) + k\Phi_2 Y(1-Y)] \times [6Z + 15W - 82D^2]\},$$

$$v_{22} = \frac{1}{2}\{W - 2D^2 + [k\theta_1(1-2X) + k\theta_2(1-2Y)][W - 6D^2] - 2(k\theta_1 A + k\theta_2 B) + [k\Phi_1(1-2X)^2 + k\Phi_2(1-2Y)^2] \times [W - 6D^2]\}$$

$$v_{23} = -16D^2[k\theta_1(1-2X) + k\theta_2(1-2Y)],$$

$$v_{31} = \frac{1}{8}\{W - 2D^2 + [k\theta_1(1-2X) + k\theta_2(1-2Y)][W - 6D^2] - 2(k\theta_1 A + k\theta_2 B) + (k\Phi_1 + k\Phi_2)(W - 14D^2) + 12(k\Phi_1 X(1-X) + k\Phi_2 Y(1-Y)(6D^2 - W)\},$$

$$v_{32} = 0,$$

and

$$v_{33} = -v_{23}.$$

The expression (9) was used to calculate the approximate standard deviation of D', $\sigma_{D'}$, for the same genetic models as before, with $k\theta_1 = 0.25$. The population sizes used were $n = 4$ and 32 and the

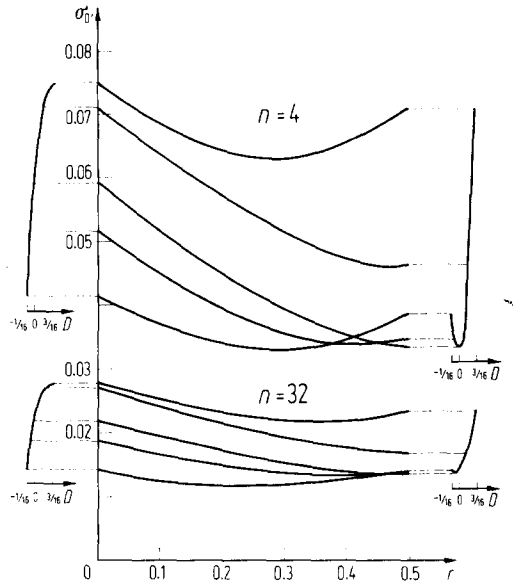

Fig. 4. The relationship between $\sigma_{D'}$, and r for different initial D; dominant/dominant model, $k\,\theta_1 = 0.25$; initial gene frequencies $X = 3/4$, $Y = 3/4$

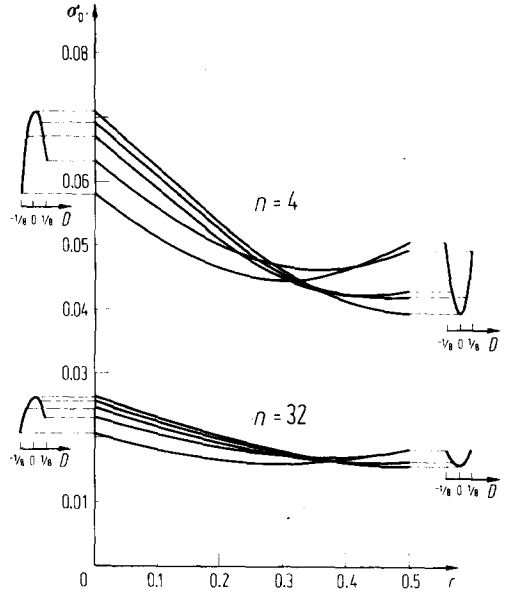

Fig. 5. The relationship between $\sigma_{D'}$, and r for different initial D; dominant/dominant model, $k\,\theta_1 = 0.25$; initial gene frequencies $X = 1/2$, $Y = 3/4$

recombination fraction was $r = 0$, 0.01, 0.1, 0.4 and 0.5. As before, the initial linkage disequilibrium took 5 values throughout the range possible for each pair of gene frequencies, and these took the values $\frac{1}{4}$, $\frac{1}{2}$ and $\frac{3}{4}$ in all combinations.

Figures 1 to 3 show the relationship between $\sigma_{D'}$, and r when there is no selection, (Model $-/-$), for the population sizes $n = 4$ and $n = 32$ and the five different initial values of D. These values could be obtained from the formulae given by Watterson (1970). The initial gene frequencies (X, Y) in the three figures are $\left(\frac{1}{4}, \frac{1}{4}\right)$, $\left(\frac{1}{4}, \frac{1}{2}\right)$ and $\left(\frac{1}{2}, \frac{1}{2}\right)$ respectively. The situation for $\left(\frac{3}{4}, \frac{3}{4}\right)$ if identical to that for $\left(\frac{1}{4}, \frac{1}{4}\right)$, as those for $\left(\frac{1}{2}, \frac{1}{4}\right)$, $\left(\frac{1}{2}\ \frac{3}{4}\right)$ and $\left(\frac{3}{4}, \frac{1}{2}\right)$ are to $\left(\frac{1}{4}, \frac{1}{2}\right)$. The diagrams for $\left(\frac{1}{4}, \frac{3}{4}\right)$ and $\left(\frac{3}{4}, \frac{1}{4}\right)$ are identical to $\left(\frac{1}{4}, \frac{1}{4}\right)$ if D takes the opposite sign. From Figures 1 to 3 the effect of sampling can be estimated. In every situation, except $X = Y = \frac{1}{2}$ with D near $\pm\frac{1}{4}$, tight linkage (r close to 0) gives a greater value of $\sigma_{D'}$, than $r = \frac{1}{2}$, although the minimum of $\sigma_{D'}$ does not generally occur at $r = \frac{1}{2}$. When $X = Y = \frac{1}{2}$, $D = \pm\frac{1}{4}$, the situation is one in which initially either both repulsion or both coupling gametes are

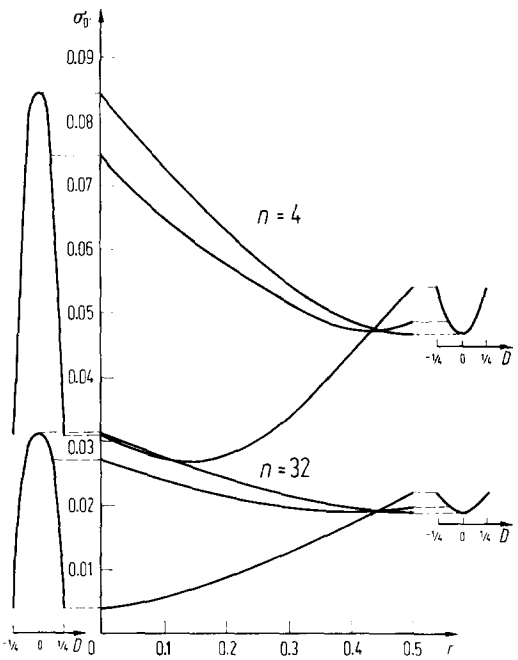

Fig. 6. The relationship between $\sigma_{D'}$, and r for different initial D; dominant/dominant model, $k\,\theta_1 = 0.25$; initial gene frequencies, $X = 1/2$, $Y = 1/2$

absent from the population. With no selection, in each case the differences in $\sigma_{D'}$ due to different values of D are reduced as r approaches $\frac{1}{2}$. The greatest changes in $\sigma_{D'}$ occur as D approaches its extremes, exceptionally for the initial frequencies $(X, Y) = \left(\frac{1}{4}, \frac{1}{4}\right)$, $\left(\frac{1}{4}, \frac{3}{4}\right)$, $\left(\frac{3}{4}, \frac{1}{4}\right)$, $\left(\frac{3}{4}, \frac{3}{4}\right)$, as may

Table 4. *Values of $\sigma_{D'}$ when $D = 0$ initially. Population sizes $n = 4$ and 32, recombination fractions $r = 0.1$, 0.5 and $k\theta_1 = 0$ and 0.25*

n	r	Y \ X	−/−			a/−			d/−			a/a			d/a			d/d		
			1/4	1/2	3/4	1/4	1/2	3/4	1/4	1/2	3/4	1/4	1/2	3/4	1/4	1/2	3/4	1/4	1/2	3/4
4	0.1	1/4	.0553	.0638	.0553	.0536	.0638	.0517	.0600	.0635	.0533	.0618	.0677	.0553	.0631	.0674	.0568	.0645	.0690	.0582
		1/2	.0638	.0737	.0638	.0677	.0737	.0597	.0693	.0733	.0615	.0677	.0737	.0597	.0693	.0733	.0615	.0690	.0730	.0612
		3/4	.0553	.0638	.0553	.0586	.0638	.0517	.0600	.0635	.0533	.0553	.0597	.0479	.0568	.0594	.0496	.0582	.0612	.0513
	0.5	1/4	.0406	.0469	.0406	.0431	.0469	.0380	.0425	.0438	.0373	.0454	.0497	.0406	.0443	.0469	.0400	.0442	.0461	.0393
		1/2	.0469	.0541	.0469	.0497	.0541	.0438	.0490	.0506	.0431	.0497	.0541	.0438	.0490	.0506	.0431	.0461	.0469	.0397
		3/4	.0406	.0469	.0406	.0431	.0469	.0380	.0425	.0438	.0373	.0406	.0438	.0352	.0400	.0406	.0344	.0393	.0397	.0337
32	0.1	1/4	.0210	.0243	.0210	.0223	.0243	.0197	.0227	.0240	.0201	.0235	.0257	.0210	.0239	.0255	.0215	.0243	.0259	.0219
		1/2	.0243	.0280	.0243	.0257	.0280	.0227	.0262	.0277	.0233	.0259	.0280	.0227	.0262	.0277	.0233	.0259	.0273	.0229
		3/4	.0210	.0243	.0210	.0223	.0243	.0197	.0227	.0240	.0201	.0210	.0227	.0182	.0215	.0224	.0187	.0219	.0229	.0192
	0.5	1/4	.0163	.0188	0.163	.0173	.0188	.0153	.0171	.0176	.0150	.0182	.0200	.0163	.0180	.0188	.0161	.0178	.0185	.0158
		1/2	.0188	.0217	0.188	.0200	.0217	.0176	.0197	.0203	.0173	.0200	.0217	.0176	.0197	.0203	.0173	.0185	.0188	.0160
		3/4	.0163	.0188	.0163	.0173	0.188	.0153	.0171	.0176	.0150	.0163	.0176	.0141	.0161	.0163	.0138	.0158	.0160	.0135

be seen in Figure 1, when $r = 0$ and $|D|$ is near $\frac{3}{16}$, and when $r = \frac{1}{2}$ and $|D|$ is near $\frac{1}{16}$. The greatest change in σ_D occurs when $X = Y = \frac{1}{2}$, for r close to 0, where $\sigma_{D'}$ rapidly increases as $|D| \to 0$. The situations represented in Figure 2 show that σ_D is not greatly affected by changes in D when the gene frequencies are of unequal but intermediate sizes.

Figures 4 to 6 show the relationships of figures 1 to 3 for the dominant/dominant model with $k\theta_1 = 0.25$ and initial frequencies $\left(\frac{3}{4}, \frac{3}{4}\right)$, $\left(\frac{1}{2}, \frac{3}{4}\right)$ and $\left(\frac{1}{2}, \frac{1}{2}\right)$ respectively. Curves for the selection models were very similar to the ones for the no selection situation. Figures 4 to 6 are representative of the most marked differences between the selection and no selection models.

Table 4 gives the value of $\sigma_{D'}$ for the different models when $D = 0$ initially, with $k\theta_1 = 0.25$, $n = 4$ and 32, $r = 0.1$ and 0.5. The table gives some idea of the response of $\sigma_{D'}$ to changes in the initial gene frequency at the two loci.

The greatest variation in the value of $\sigma_{D'}$ is due to the size of the population selected. As n increases, $\sigma_{D'}$ decreases, and the range of values taken by $\sigma_{D'}$ for different D correspondingly decreases. The initial values of X, Y and D contribute more to differences in $\sigma_{D'}$ than the underlying genetic model or the effect of selection. Thus, as before, sampling is of greater importance than selection.

6. Discussion

All the formulae and results given in this paper refer to a single cycle of sampling and selection. The use of results from a single cycle in understanding the consequences of evolutionary pressures is clearly limited. We hope to find ways of extending our studies, so that repeated cycles of sampling and selection can be studied. On the evolutionary time scale, other problems would then arise because of the probable variations from generation to generation of the population size, the selection intensity, and the genetic effects. The selection of a fixed number of individuals to be the parents of the next generation, rather than the selection of individuals with probabilities depending only on their own genetic values, may also be more relevant in plant and animal breeding programmes than in natural selection.

We are hoping to find ways of extending the results obtained, so that a small number of cycles of selection can be studied. These results would be of considerable importance in artificial selection programmes. In particular, we plan to study the way in which the variance of the mean performance of a small population develops under sampling and selection. A previous discussion of the between

population variance in mean performance (Baker and Curnow, 1969) ignored the correlations between gene frequency changes at different loci.

The results obtained in this paper do suggest that with a small number of cycles of sampling and selection, the effects of sampling can often dominate those of selection. Clearly the balance between the effects of sampling and the effects of selection depend on the population size, the selection intensity, and the size of the genetic effects. What we have shown in this paper is that if the genetic effects are of the size generally associated with quantitative characters, i.e. $d_i/\sigma < 0.2$, and the selection not too intense (i.e. $k < 2$), then sampling often has a larger effect than selection on the development of linkage disequilibrium and of correlations between gene frequencies at different loci.

Acknowledgements

Our colleague Mr. D. J. Pike has previously obtained the formulae for $E(D')$ and $Cov\,(X', Y')$ using similar methods. We are grateful to Mr. Pike for his assistance in the preparation of this paper. We are also grateful to the secretarial staff of the Department for their typing of a paper containing many complex mathematical expressions. We thank the Science Research Council for financial support in connection with this research.

Literature

1. Baker, L. H., Curnow, R. N.: Choice of population size and use of variation between replicate populations in plant breeding selection programs. Crop Science **9**, 555—560 (1969). — 2. Hill, W. G.: On the theory of artificial selection in finite populations. Genet. Research, Camb., **13**, 143—163 (1969). — 3. Hill, W. G., Robertson, A.: The effect of linkage on limits to artificial selection. Genet. Research., Camb., **8**, 269—294 (1966). — 4. Hill, W. G., Robertson, A.: Linkage disequilibrium in finite populations. Theoret. Appl. Genetics **38**, 226—231 (1968). — 5. Karlin, S., McGregor, J.: Rates and probabilities of fixation for two locus random mating populations without selection. Genetics **58**, 141—159 (1968). — 6. Kojima, K.: Effects of dominance and size of population on response to mass selection. Genet. Research, Cambridge, **2**, 177—188 (1961). — 7. Latter, B. D. H.: The response to artificial selection due to autosomal genes of large effect. I. Changes in gene frequency at an additive locus. Aus. J. Biol. Sci. **18**, 585—598 (1965). — 8. Ohta, T.: The effect of initial linkage disequilibrium and epistasis on fixation probability in a small population, with two segregating loci. Theoret. Appl. Genetics **38**, 243—248 (1968). — 9. Watterson, G. A.: The effect of linkage in a finite random mating population. Theoretical Population Biology **1**, 72—87 (1970).

Received February 25, 1972

Communicated by A. Robertson

Susan J. Galley
R. N. Curnow
Department of Applied Statistics
University of Reading
Reading, Berkshire (England)