Theor. Appl. Genet. 54, 133-139 (1979)

Two Locus Models of Selection and Mutation within and among Full-Sib Lines

F.M. Stewart
Department of Mathematics, Brown University, Providence, Rhode Island (USA)

M.T. Clegg and J.F. Kidwell
Division of Biology and Medicine, Brown University, Providence, Rhode Island (USA)

Summary. General models for continued full-sib mating with two diallelic autosomal loci taking account of linkage, mutation and selection within and among lines are considered. The problems are first approached by deriving the full probability transition matrix, taking account of linkage, mutation and within-line selection. Exact solutions to the equilibrium system are possible, but the computational effort is prohibitive, and this is exacerbated by the introduction of between-line selection. A second approach is based on decomposing the transition matrix into blocks whose properties suggest approximations that lead to a rapid iterative solution of the equilibrium system. Extensive numerical analysis of models of within-line selection and of combined within- and between-line selection were made. The results show that equilibrium values are essentially independent of the degree of linkage under models of within-line selection. This is because mutation plays a dominant role in determining equilibrium structure. Results from models of combined within- and between-line selection show that between-line selection has the dominant influence on gene frequency equilibrium. Both within-line and between-line selection produce appreciable linkage disequilibrium only when selection is disruptive. The results also suggest that much of the two-locus equilibrium structure can be predicted from a knowledge of single-locus equilibria.

Key words: Locus models – Between-line selection – Within-line selection

Introduction

The consequences of full-sib mating for two diallelic autosomal loci in the absence of mutation and selection were considered briefly by Jennings (1917) and in greater detail by Haldane and Waddington (1931). Reeve and Gower (1958) derived the 19 × 19 generation matrix for the case of an unselected pair of alleles at one locus linked to a pair at another locus with equal selection against both homozygotes. They considered within-line selection and the special case where selection acts equally within and between lines. In this paper we consider general models for continued full-sib mating with two diallelic autosomal loci, taking account of linkage, mutation and selection within and among lines.

Two approaches to this problem are developed. The first approach follows that of earlier workers by deriving the full probability transition matrix for full-sib mating with selection within lines and mutation. While it is computationally possible to obtain exact solutions to the equilibrium system for this formulation, the computational effort is nearly prohibitive. The introduction of between-line selection, as done in the one-locus case by Clegg and Kidwell (1974), further complicates the computational problem. Consequently, a second approach to solving the equilibrium system is presented. This second approach rests on replacing the transition matrix by a series of linear operators. The properties of the individual operators permit certain approximations which allow a rapid iterative solution of the equilibrium system.

General Models

Within-Line Selection

The model is developed with complete generality, but only a few specific cases are considered in detail. The assumptions include two diallelic loci, constant selective values for each genotype which may differ between the sexes, constant recombination rates which may differ between sexes, and constant mutation rates for each locus. Mutation at each locus is assumed to occur independently after gamete formation, i.e. after crossing over occurs. Parents within each line (= mating type) are chosen at random after selection. Selection is assumed to occur only by viability differences.

Let μ_a be the probability that a gamete expected to

0040-5752/79/0054/0133/$ 1.00

Table 1. Genotypes with their gamete pair (AB = 1, Ab = 2, aB = 3, ab = 4), identification and within-line fitness values

Genotype	Gamete pairs	Identification	Relative Fitness	
			Males	Females
AB/AB	(1,1)	0	w_{m0}	w_{f0}
AB/Ab	(1,2) or (2,1)	1	w_{m1}	w_{f1}
AB/aB	(1,3) or (3,1)	2	w_{m2}	w_{f2}
AB/ab	(1,4) or (4,1)	3	w_{m3}	w_{f3}
Ab/Ab	(2,2)	4	w_{m4}	w_{f4}
Ab/aB	(2,3) or (3,2)	5	w_{m5}	w_{f5}
Ab/ab	(2,4) or (4,2)	6	w_{m6}	w_{f6}
aB/aB	(3,3)	7	w_{m7}	w_{f7}
aB/ab	(3,4) or (4,3)	8	w_{m8}	w_{f8}
ab/ab	(4,4)	9	w_{m9}	w_{f9}

carry the allele A carries the allele a, i.e. that A has mutated to a, and ν_a be the probability of reverse mutation. The symbols μ_b and ν_b are defined similarly for the B locos. We designate r_m and r_f as the probability of recombination in males and in females respectively, where ($0 \leqslant r_m, r_f \leqslant .5$). The relative fitness of each sex and genotype is defined in Table 1. Individuals of each of the 10 male genotypes can mate with individuals of each of the 10 female genotypes to produce 100 distinct mating types or lines. In this paper the mating types are identified by a two digit number; the first digit designates the genotype number (Table 1) of the male and the second digit the corresponding number of the female. For example, mating type 34 is $3 = AB/ab^{\delta\delta} \times 4 = Ab/Ab^{\varphi\varphi}$. We note here for later use in computing that the number of mating types may be reduced to 55 by combining the off-diagonal elements which are the reciprocal mating types.

The probabilities that the parents of any generation, t, are of each mating type can be expressed as the product of the t^{th} power of a 100 x 100 generation matrix and 100 x 1 column vector of their probabilities in some initial generation, 0, as follows:

$$[p_{i,t}] = [a_{i,j}]^t \times [p_{j,0}] \quad \ldots\ldots (1)$$ where

$[p_{i,t}]$ is the 100 x 1 column vector of probabilities that a randomly chosen full-sib pair in generation t will be of mating type i. $[a_{i,j}]$ is a 100 x 100 generation matrix. The elements $a_{i,j}$ are the conditional probabilities that mating type j in one generation will give rise to mating type i in the subsequent generation. $[p_{j,0}]$ is a 100 x 1 column vector of the relative frequencies of the mating types among adult parents in some initial generation, 0.

Calculation of the $a_{i,j}$ is straight-forward, but tedious. The procedure is outlined briefly below. The males and females of each mating type produce 4 kinds of gametes with a frequency distribution determined by the mutation and recombination rates. The distribution of gametes produced by males and females of 3 representative genotypes is given in Table 2. The distributions for the seven remaining genotypes are calculated similarly. Within mating types, each of the 4 kinds of male gametes can unite with each of the 4 kinds of female gametes to produce 16 zygotic combinations of each sex. Their frequency distribution is determined as the product of the gametic frequency distributions of the two sexes. There are, of course, only 10 distinct kinds of zygotes and their frequencies may be obtained by combining the off-diagonal reciprocals. Following zygote production each of the 10 distinct kinds of male zygotes can mate with each of the 10 distinct kinds of female zygotes to again from 100 mating types.

We describe in detail the calculation of $a_{34,45}$. Any other $a_{i,j}$ can be calculated similarly. Consider the pair of mating type 45. We are interested in the progeny pairs

Table 2. Gametic array produced by three representative genotypes of each sex. The model assumes that the mutations occur independently and after crossing over. g_{ijk} = the frequency of the k^{th} gamete produced by the j^{th} sex of the i^{th} genotype. 1 = 0,1,...9; j = 1,2 (1 = male, 2 = female); k = 1,2,3,4

Genotype	Gametes	Frequency	Designation	
			Male	Female
AB/AB	*AB*	$(1-\mu_a)(1-\mu_b)$	$g_{0,1,1}$	$g_{0,2,1}$
	Ab	$(1-\mu_a)\mu_b$	$g_{0,1,2}$	$g_{0,2,2}$
	aB	$\mu_a(1-\mu_b)$	$g_{0,1,3}$	$g_{0,2,3}$
	ab	$\mu_a\mu_b$	$g_{0,1,4}$	$g_{0,2,4}$
AB/Ab	*AB*	$(1-\mu_a)(1-\mu_b+\nu_b)/2$	$g_{1,1,1}$	$g_{1,2,1}$
	Ab	$(1-\mu_a)(1+\mu_b-\nu_b)/2$	$g_{1,1,2}$	$g_{1,2,2}$
	aB	$\mu_a(1-\mu_b+\nu_b)/2$	$g_{1,1,3}$	$g_{1,2,3}$
	ab	$\mu_a(1+\mu_b-\nu_b)/2$	$g_{1,1,4}$	$g_{1,2,4}$
AB/ab	*AB*	$[(1-r_m)/2][(1-\mu_a)(1-\mu_b)+\nu_a\nu_b] + (r_m/2)[\nu_a(1-\mu_b)+\nu_b(1-\mu_a)]$	$g_{3,1,1}$	$g_{3,2,1}$[a]
	Ab	$[(1-r_m)/2][(1-\mu_a)\mu_b+\nu_a(1-\nu_b)] + (r_m/2)[(1-\mu_a)(1-\nu_b)+\nu_a\mu_b]$	$g_{3,1,2}$	$g_{3,2,2}$[a]
	aB	$[(1-r_m)/2][\mu_a(1-\mu_b)+\nu_b(1-\nu_a)] + (r_m/2)[\mu_a\nu_b+(1-\nu_a)(1-\mu_b)]$	$g_{3,1,3}$	$g_{3,2,3}$[a]
	ab	$[(1-r_m)/2][\mu_a\mu_b+(1-\nu_a)(1-\nu_b)] + (r_m/2)[\mu_a(1-\nu_b)+\mu_b(1-\nu_a)]$	$g_{3,1,4}$	$g_{3,2,4}$[a]

[a] The formulae for the females differ from those for the males in that r_f replaces r_m

that are of type 34, i.e. the male of genotype 3 = *AB/ab* = (1,4) or (4,1) and the female of genotype 4 = *Ab/Ab* = (2,2). The frequency of the *AB/ab* genotype in the male progeny is $g_{4,1,1}\ g_{5,2,4} + g_{4,1,4}\ g_{5,2,1}$ where the $g_{i,j,k}$ are calculated as in Table 2. Similarly, the frequency of females of genotype 4 is $g_{4,1,2}\ g_{5,2,2}$. Each of these is subject to selection so the number of adult pairs of type 34 measured relative to the total progeny of pairs of type 45 will be $S_{34,45} = W_{m3}\,(g_{4,1,1}\ g_{5,2,4} + g_{4,1,4}\ g_{5,2,1})\,W_{f4}\,(g_{4,1,2}\ g_{5,2,2})$.

To obtain frequencies in the new generation we must normalize by

$$S_{\cdot\cdot,45} = \sum_{i=0}^{99} S_{i,45}. \quad \text{Hence, } a_{34,45} = S_{34,45}/S_{\cdot\cdot,45}.$$

Of course, at equilibrium $[p_{i,t}] = [p_{i,t-1}]$ for all i. General equilibrium solutions can be obtained by the method suggested by J.F. Crow and described in detail by Clegg and Kidwell (1974) for the single locus case.

Between-Line Selection

The model of between-line selection is similar to the single locus model of Clegg and Kidwell (1974). We assume an infinite number of full-sib pairs (populations of size 2), whose frequency distribution is given by equation 1. The mating types are assumed to reproduce at constant rates, b_i, per generation. Between-line selection is assumed to follow within-line selection and formation of mating types. Thus the number of pairs of mating type i, after between-line selection will be

$$n_i = b_i \sum_{j=0}^{99} a_{i,j}\, p_{j,t-1}.$$

The relative frequency will be $p_{i,t}$ where

$$[p_{i,t}] = [b_{i,i}] \times [a_{i,j}] \times [p_{j,t-1}]$$

and $[b_{i,i}]$ is the 100 × 100 diagonal matrix whose i,i^{th} element is the renormalized between-line selective value

$$b_{i,i} = b_i / \sum_{j=0}^{99} n_j.$$

The matrix $[a_{i,j}]$ is the previously defined 100 × 100 generation matrix. The vector $[p_{i,t}]$ is the previously defined 100 × 1 column vector of mating type frequencies of adult parents.

It is important to note that the matrix

$$[d_{i,j}] = [b_{i,i}] \times [a_{i,j}]$$

is not a probability transition matrix because its column sums are not 1 and its value depends on $[p_{j,t-1}]$ so that the computational methods described earlier cannot be employed.

Computational Methods

The equilibrium equations are much too complicated to permit analytical deduction and generalization. An alternative approach is to evaluate them numerically for a number of parameter values reflecting different specific models of within- and between-line selection. The amount of computer time required to evaluate the exact solutions proved to be excessive, because, after some simplification, fifty-five 54 × 54 determinants must be evaluated each time. Instead, we use a rapid and accurate approximation. It is described in the following paragraphs and the results of the numerical analysis are discussed in the concluding section of the paper.

The programming is greatly simplified by exploiting all the symmetries of the situation by distinguishing heterozygotes on the basis of which allele came from which parent. When this is done, there are $256 = 2^8$ mating types that fall into 55 categories, or orbits, with lines in the same orbit having equal frequencies after at most two generations. A description of the methods used can be obtained by writing to the Division of Biology and Medicine, Brown University, Providence, Rhode Island 02912, Attn: Section of Population Biology and Genetics.

Each of the 4 pure lines is an orbit in itself. There are 56 lines that are pure at one locus and heterozygous at the other. (A line is classed as heterozygous if both alleles are present even though a heterozygous line, such as *AB/AB* × *Ab/Ab*, may be made up of homozygous individuals.) These constitute 16 orbits and there are 35 orbits made up of the 196 lines that are heterozygous at both loci.

If we let p_i and p_i', i = 1, 2,..., 55 be the frequency of the i-th line in two successive generations then

$$p_i' = \sum_{i=1}^{55} c_{i,j}\, p_j / \bar{w}$$

where $[c_{i,j}]$ is a constant matrix, but not a probability matrix, and $\bar{w}$ is a number which measures the average between-line fitness. As noted earlier, $\bar{w}$ is a linear combination of the p's, however, at equilibrium, when $p_i' = p_i$, it will have some numerical value, say λ. Thus the equilibrium values of $p_1, p_2, \ldots, p_{55}$ will be an eigen-vector of the eigen-value problem

$$\lambda \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_{55} \end{bmatrix} = \begin{bmatrix} c_{1,1} & c_{1,2} & \cdots & c_{1,55} \\ c_{2,1} & c_{2,2} & \cdots & c_{2,55} \\ \vdots & \vdots & & \vdots \\ c_{55,1} & c_{55,2} & \cdots & c_{55,55} \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_{55} \end{bmatrix}$$

We order the orbits so $p_1 \ldots, p_4$ are the frequencies of the pure lines, $p_5 \ldots p_{20}$ are those of the lines pure at one locus and $p_{21}, \ldots, p_{55}$ are those of the lines with both

alleles at both loci. Then the $c_{i,j}$ with i = 5,6, ..., 20 and j = 1,2,3,4 are of the same order of magnitude as the mutation rates. So, too, are those with i = 21,22,...,55 and j = 5,6,...,20, while those with i = 21,22,...,55 and j = 1,2,3,4 are of the order of the squares of the mutation rates.

Accordingly, we partition the matrix $[c_{i,j}]$ as

$$\begin{bmatrix} I+\theta A & B & C \\ \theta D & E & F \\ \theta^2 G & \theta H & K \end{bmatrix}$$

where θ is a number of the same order of magnitude as the mutation rates. It is important to realize that even though there is no general parameterization that will give $[c_{i,j}]$ this form with constant submatrices A,B,...,K, nevertheless, for any particular selection, mutation, and recombination values, we can choose θ arbitrarily and will then have definite values for A, B,..., K.

It proves to be in convenient to seek the eigen-vector normalized by the condition $\Sigma_{i=1}^{55} p_i = 1$. Accordingly, we seek a positive eigen-vector

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix}$$

where

$$X = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}, \quad Y = \begin{bmatrix} x_5 \\ x_6 \\ \vdots \\ x_{20} \end{bmatrix}, \text{ and } Z = \begin{bmatrix} x_{21} \\ x_{22} \\ \vdots \\ x_{55} \end{bmatrix},$$

without committing ourselves, yet, to any normalization. When this is found, we can set $p_i = x_i / \Sigma_{i=1}^{55} x_i$.

We seek to solve

$$\lambda X = X + \theta AX + BY + CZ \quad (2)$$
$$\lambda Y = \theta DX + EY + FZ \quad (3)$$
$$\lambda Z = \theta^2 GX + \theta HY + KZ \quad (4)$$

with power series of the form

$$\lambda = \Sigma_{n=0}^{\infty} \lambda_n \theta^n, X = \Sigma_{n=0}^{\infty} X_n \theta^n,$$

$$Y = \theta \Sigma_{n=0}^{\infty} Y_n \theta^n, \text{ and } Z = \theta^2 \Sigma_{n=0}^{\infty} Z_n \theta^n.$$

For the terms in θ^0 we get just $\lambda_0 X_0 = X_0$, so we must take $\lambda_0 = 1$. Then the terms in θ^1 give

$$\lambda_0 X_1 + \lambda_1 X_0 = X_1 + AX_0 + BY_0$$

$$\lambda_0 Y_0 = DX_0 + EY_0.$$

Since $\lambda_0 = 1$,

$$\lambda_1 X_0 = \{A + B(I-E)^{-1} D\} X_0.$$

We must choose λ_1 to be the dominant eigen-value of the 4 x 4 matrix $L = A + B(I-E)^{-1} D$ and let X_0 be a corresponding eigen-vector. (Since E is a non-negative matrix of norm less than 1, $(I-E)^{-1}$, and hence L, are also non-negative). It can be chosen so that its elements are all positive. For most purposes, $X = X_0$, $Y = \theta Y_0 = \theta(I-E)^{-1} DX_0$, $Z = 0$ is a very good approximation to the solution.

Equations (2)-(4) pose an eigen-vector problem, so solutions are only determined up to an arbitrary scalar factor. For higher order approximations the presence of an arbitrary constant is very inconvenient. Any of our approximate solutions $X = \Sigma_{n=0}^{N} X_n \theta^n$ can be resolved into its components in terms of the eigen-vectors of L. Since the constant term, X_0, is an eigen-vector, we can normalize by insisting that the component of X in that direction is precisely the X_0 originally chosen. In other words, we can require that $X_1, X_2, \ldots$ all lie in the subspace spanned by the other eigen-vectors. We shall make this more explicit below, but the idea is that, since λ_1 is the dominant eigen-value of L, when we restrict λ_1-L to this subspace it is 1-to-1.

We have already chosen $\lambda_0 = 1, \lambda_1, X_0$, and Y_0 so that the equations (2)-(4) are satisfied to terms of order θ^k for k = 0 and 1. For $k \geqslant 2$, the terms in θ^k are

$$\Sigma_{i=0}^{k} \lambda_i X_{k-i} = X_k + AX_{k-1} + BY_{k-1} + CZ_{k-2} \quad (5)$$

$$\Sigma_{i=0}^{k-1} \lambda_i Y_{k-i-1} = DX_{k-1} + EY_{k-1} + FZ_{k-2} \quad (6)$$

$$\Sigma_{i=0}^{k-2} \lambda_i Z_{k-i-2} = GX_{k-2} + HY_{k-2} + KZ_{k-2} \quad (7)$$

Suppose $\lambda_0 = 1, \lambda_1, \ldots, \lambda_{k-1}, X_0, \ldots, X_{k-2}, Y_0, \ldots, Y_{k-2}$, and $Z_0, \ldots, Z_{k-3}$ have all been determined. To satisfy (7) we need only set

$$Z_{k-2} = (I-K)^{-1} \{GX_{k-2} + HY_{k-2} - \Sigma_{i=1}^{k-2} \lambda_i Z_{k-i-2}\}.$$

From (6) we must have

$$Y_{k-1} = (I-E)^{-1} \{DX_{k-1} + FZ_{k-2} - \Sigma_{i=1}^{k-1} \lambda_i Y_{k-i-1}\}$$

which, substituted in (5) yields (8)

$$\lambda_k X_0 + (\lambda_1 - L) X_{k-1} = B(I-E)^{-1}$$

$$\{FZ_{k-2} - \Sigma_{i=1}^{k-1} \lambda_i Y_{k-i-1}\} + CZ_{k-2} - \Sigma_{i=2}^{k-1} \lambda_i X_{k-i} \quad (9)$$

Equation (9) is solved by resolving its right hand side into its components in the X_0 direction and in the subspace spanned by the other three eigen-vectors of L. Since λ_1-L is a 1-1 map of the subspace onto itself, we can choose λ_k and X_{k-1} so that the left hand side of (9) is equal to the right hand side.

To be quite explicit, let Q be the matrix whose columns are the eigen-vectors of L, with X_0 first. Now $Q^{-1}LQ$ is the diagonal matrix of the eigen-values of L and

$$Q^{-1}X_0 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Then let

$$M = \lambda_1 I - Q^{-1}LQ = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \alpha & 0 & 0 \\ 0 & 0 & \beta & 0 \\ 0 & 0 & 0 & \gamma \end{bmatrix}$$

where α, β, and γ are strictly positive. If we multiply equation (9) by Q^{-1} we get an equivalent equation

$$\lambda_k Q^{-1}X_0 + MQ^{-1}X_{k-1} = U \qquad (10)$$

where

$$U = \begin{bmatrix} u_0 \\ u_1 \\ u_2 \\ u_3 \end{bmatrix}$$

is Q^{-1} times the right hand side of (9). Clearly, we must have $\lambda_k = u_0$, but since M is singular, equation (10) does not determine X_{k-1} uniquely. However, it is natural enough to choose the solution

$$X_{k-1} = Q \begin{bmatrix} 0 \\ u_1/\alpha \\ u_2/\beta \\ u_3/\gamma \end{bmatrix}$$

It is not self-evident that the power series we obtain in this way must converge. However, the implicit function theorem guarantees the existence of a solution which is analytic in θ and we have seen that its power series can only be the ones we have determined.

Numerical Results

A virtually unlimited number of specific models can be generated from the combination of within- and between-line fitness schemes. Moreover, the potential number of parameters is unmanageably large. For example, there are 55 different mating types if we do not distinguish orderings by sex; this results in 55 separate between-line fitness parameters. To reduce the number of between-line parameters we consider only multiplicative selection, i.e. selection models where the line fitness (b_i for the i^{th} line) is the product of fitness values assigned to genotypes. Assuming multiplicative between-line fitness and that genotypic fitness is the same over sexes reduces the parameterization from 55 to 10. To further reduce the number of between-line fitness parameters only special models of selection have been investigated. The most elementary mod-

Table 3. Models of Within- and Between-Line Selection

	Fitness Model					
	Within-Line Selection			Between-Line Selection		
Genotype	I	II	III	I	II	III
AB/AB	1	1	1+2s	1	$(1-2t)^2$	1+2t
AB/Ab	1+s	1−s	1	1+t	(1−t) (1−2t)	1
AB/aB	1+s	1−s	1	1+t	(1−t) (1−2t)	1
AB/ab	$(1+s)^2$	$(1-s)^2$	1+2s	$(1+t)^2$	$(1-t)^2$	1+2t
Ab/Ab	1	1−2s	1−2s	1	1−2t	1−2t
Ab/aB	$(1+s)^2$	$(1-s)^2$	1−2s	$(1+t)^2$	$(1-t)^2$	1−2t
Ab/ab	1+s	(1−s) (1−2s)	1	1+t	1−t	1
aB/aB	1	1−2s	1−2s	1	1−2t	1−2t
aB/ab	1+s	(1−s) (1−2s)	1	1+t	1−t	1
ab/ab	1	$(1-2s)^2$	1+2s	1	1	1+2t

els investigated are presented in Table 3. These models assume a single parameter for genotypic fitness within lines (s) and a single between-line fitness parameter (t). Model I assumes multiplicative overdominance. Model II represents a case of directional selection where the *AB/AB* genotype is favored within lines, while between-line selection favors the *ab/ab* genotype. Model III presents an extreme case of disruptive selection where genotypes *AB/AB* and *ab/ab* are favored over all other genotypes.

Table 4 reports a sample of numerical results contrasting the effect of mutation and within-line selection on gene frequencies, heterozygosity and the correlation in allelic state over loci. All numerical results reported in the Tables were obtained using the approximations outlined in the section on Computational Methods. Exact solutions to equation (1), for a few parameter values, were also obtained to check the accuracy of the approximate solutions. Very close agreement between the exact and approximate solutions was observed in all cases. For each choice of mutation and selection parameters, runs were made over a range of recombination varying from 0.01 to 0.5. The final equilibrium values were always essentially independent of the degree of linkage assumed between the loci. The explanation can be readily seen from the frequency of double heterozygotes, which are usually two or more orders of magnitude below the mutation rates due to the intense inbreeding of the full-sib mating scheme. New gametic types arise from mutation much more frequently than they do from recombination. Consequently, recombination has no influence on the final equilibrium structure. For the same reason, the correlation in allelic state among loci (standardized linkage disequilibrium) is always close to zero for models I and II at equilibrium. Mutation, which acts independently among loci, results in independence at the level of the gametic frequency distribution. Extensive numerical investigation of models of directional selection and models of selection favoring heterozygotes has failed to reveal any cases where the level of linkage disequilibrium was appreciable. Reference to the equilibrium gene frequencies and mutation rates also shows that mutation plays an important role in determining the final equilibrium structure as is the case for single-loci (Clegg and Kidwell 1974). Within-line selection is relatively inefficient compared to mutation because the within-line variance at equilibrium is small.

The only class of models which lead to linkage disequilibrium at equilibrium are models of disruptive selection (e.g. Model III). The presence of mutation stabilizes the internal equilibrium under these selection schemes. Hence, the extreme homozygous genotypes can be simultaneously maintained.

Table 5 contrasts the equilibrium effects of between- and within-line selection. Unlike within-line selection, between-line selection has the predominant influence on gene frequency equilibrium. This, of course, follows from the fact that virtually all the genetic variance is expressed between lines rather than between genotypes within lines. Like within-line selection, between-line selection produces appreciable linkage disequilibrium only when the selection is disruptive (Model III). All models of symmetrical and directional selection investigated lead to zero or very small values of linkage disequilibrium regardless of the degree of epistasis assumed.

In seeking to apply these results to experimental inbreeding programs, an important caveat should be observed. The equilibrium distributions are a function of

Table 4. The effect of within-line selection and mutation on equilibrium structure. The selection models are defined in Table 3. Mutation rates are assumed to be equal at the two loci, i.e. $\mu_a = \mu_b$; $\nu_a = \nu_b$. The recombination fraction is .01. p(A) and p(B) are the frequencies of alleles *A* and *B*. h(A) and h(B) are the frequencies of heterozygotes at the *A* and *B* loci and h(D) is the frequency of double heterozygotes. r is the correlation in allelic state over loci Since the models are symmetric p(A) = p(B) and h(A) = h(B)

Model	μ	ν	s	p(A) = p(B)	h(A) = h(B)	h(D)	r
I	10^{-5}	10^{-5}	.1	.5000	1.50×10^{-4}	3.87×10^{-7}	0.0
	10^{-5}	10^{-5}	.5	.5000	3.24×10^{-4}	2.83×10^{-7}	0.0
	10^{-5}	10^{-6}	.1	.0909	2.73×10^{-5}	1.28×10^{-9}	0.0
	10^{-5}	10^{-6}	.5	.0909	5.90×10^{-5}	9.37×10^{-9}	0.0
II	10^{-5}	10^{-5}	.1	.7093	1.13×10^{-4}	1.93×10^{-8}	0.0
	10^{-5}	10^{-5}	.4	.9981	4.20×10^{-5}	2.60×10^{-9}	0.0
	10^{-5}	10^{-6}	.1	.1961	3.12×10^{-5}	1.48×10^{-9}	0.0
	10^{-5}	10^{-6}	.4	.9816	4.13×10^{-5}	2.51×10^{-9}	0.0
III	10^{-5}	10^{-5}	.1	.5000	9.96×10^{-5}	1.74×10^{-8}	+.6692
	10^{-5}	10^{-5}	.4	.5000	3.45×10^{-5}	6.59×10^{-9}	+.9995
	10^{-5}	10^{-6}	.1	.0284	1.25×10^{-5}	4.20×10^{-10}	+.3160
	10^{-5}	10^{-6}	.4	.0099	3.76×10^{-6}	1.31×10^{-10}	+.9976

Table 5. Joint effects of mutation coupled with selection within and between lines on gene frequency [p(A) and p(B)], heterozygosity [h(A), h(B) and h(D)] and the correlation in allelic state between loci (r)

Fitness model		Parameter values				Equilibrium statistics			
Within	Between	μ	ν	s	t	p(A) = p(B)	h(A) = h(B)	h(D)	r
I	I	10^{-5}	10^{-5}	0	.001	.5000	1.21×10^{-4}	2.26×10^{-8}	.0000
		10^{-5}	10^{-5}	.4	.001	.5000	2.75×10^{-4}	1.86×10^{-7}	.0000
I	II	10^{-5}	10^{-5}	0	.001	.0025	1.19×10^{-4}	2.14×10^{-8}	.0000
		10^{-5}	10^{-5}	.4	.001	.0047	2.68×10^{-4}	1.64×10^{-7}	.0000
		10^{-5}	10^{-6}	0	.001	.0003	1.19×10^{-5}	2.15×10^{-10}	.0000
		10^{-5}	10^{-6}	.4	.001	.0005	2.69×10^{-5}	1.66×10^{-9}	.0000
II	II	10^{-5}	10^{-5}	0	.001	.0025	1.19×10^{-4}	2.14×10^{-8}	.0000
		10^{-5}	10^{-5}	.4	.001	.0261	4.01×10^{-4}	3.62×10^{-7}	−.0002
		10^{-5}	10^{-6}	0	.001	.0003	1.19×10^{-5}	2.15×10^{-10}	.0000
		10^{-5}	10^{-6}	.4	.001	.0026	4.11×10^{-5}	3.87×10^{-9}	−.0000
III	III	10^{-5}	10^{-5}	.1	0	.5000	9.96×10^{-5}	1.84×10^{-8}	+.6692
		10^{-5}	10^{-5}	.2	0	.5000	6.80×10^{-5}	1.25×10^{-8}	+.9330
		10^{-5}	10^{-6}	.1	0	.0284	1.25×10^{-5}	4.20×10^{-10}	+.3160
		10^{-5}	10^{-6}	.2	0	.0132	7.61×10^{-6}	2.55×10^{-10}	+.7393
		10^{-5}	10^{-5}	0	.001	.5000	1.18×10^{-4}	2.15×10^{-8}	+.9949
		10^{-5}	10^{-5}	.1	.001	.5000	8.48×10^{-5}	1.51×10^{-8}	+.9978
		10^{-5}	10^{-6}	0	.001	.0001	1.19×10^{-5}	2.16×10^{-10}	+.0947
		10^{-5}	10^{-6}	.1	.001	.0001	8.49×10^{-6}	1.52×10^{-10}	+.3307

mutation rates which are ordinarily small. Consequently, the time to achieve equilibrium will be large especially if between line selection is insignificant and the initial distribution features complete homozygosity.

A natural question to consider is whether the equilibrium results for two-loci can be predicted from a knowledge of one-locus statics. In most instances the answer to this question appears to be yes. Our numerical calculations have failed to reveal any dependence of equilibria on linkage. Mutation, which acts independently over loci, plays a major role in determining equilibrium distributions, and linkage disequilibrium is usually close to zero. For these reasons, we conclude that two-locus equilibria will be closely approximated by the appropriate one-locus distributions for most models of selection.

Literature

Clegg, M.T.; Kidwell, J.F.: Selection and mutation within and among full-sib lines. J. Hered. **65**, 48-55 (1974)

Haldane, J.B.S.; Waddington, C.H.: Inbreeding and linkage. Genetics **16**, 357-374 (1931)

Jennings, H.S.: The numerical results of diverse systems of breeding with respect to two pairs of characters, linked or independent, with special relation to the effects of linkage. Genetics **2**, 97-154 (1917)

Reeve, E.C.R.; Gower, J.C.: Inbreeding with selection and linkage. II. Sib-Mating. Ann. Hum. Genet. **23**, 36-49 (1958)

Received February 1, 1978
Communicated by A. Robertson

Dr. F.M. Stewart
Department of Mathematics
Brown University, Providence
Rhode Island 02912 (USA)

Dr. M.T. Clegg
Department of Botany
University of Georgia, Athens
Georgia 30602 (USA)

Dr. J.F. Kidwell
Division of Biology and Medicine
Brown University, Providence
Rhode Island, 02912 (USA)