

Effect of selection, mutation, and linkage on the equilibrium structure of selfing systems

ABSTRACT: One-, two- and three-locus models of selection and mutation in completely self-fertilizing populations are examined. Equilibrium frequencies can be determined for these systems. Numerical analyses indicate that random genetic drift attributable to selfing plays a major role in determining equilibrium frequencies, even when strong directional selection is operating. Linkage has no effect on marginal gene frequencies or single-locus heterozygosity at equilibrium. The only model of those examined which leads to linkage disequilibrium is disruptive selection. In that model, selection reinforced the effect of selfing in favoring the homozygous genotypes.

Isaac Kohane

James F. Kidwell

SELF-FERTILIZATION represents one extreme from the wide spectrum of mating systems found in nature. Moreover, it is a common mode of reproduction among plant species, particularly among colonizing species. Interest in self-fertilizing systems stems, not only from its prevalence, but also because it represents a type of sexual reproduction where the presumed advantages of sexuality have been foregone. Aside from questions of evolutionary significance, self-fertilization is often employed in plant breeding strategies as an aid in manipulating genetic materials. It is not surprising that this system of mating has long been an object of theoretical investigation. In fact, Mendel⁷ calculated the rate of loss of heterozygosity at a single diallelic locus under self-fertilization in the absence of selection and mutation.

Studies of the dynamics of selfing systems usually focus on the rate of loss of heterozygosity under various assumptions about patterns of selection. Haldane² considered the short term effects of recurrent forward mutation at a single diallelic locus with no selection in a self-fertilizing line and derived the expected frequency of heterozygotes due to mutation of fully viable autosomal genes. Kidwell⁵ derived the equilibrium frequencies for a single autosomal diallelic locus subject

to forward and reverse mutation with no selection in an infinite population of self-fertilizing organisms. In this paper we extend those results to include the effect of selection. We also consider equilibrium gene and genotype frequencies for two and three diallelic autosomal loci under selection and subject to recurrent fore and back mutation.

One-locus case

Although it is both cumbersome and tedious, detailed derivations of the equilibrium conditions for the one-locus case are presented so that the results for two- and three-loci can be given more concisely. In addition to complete selfing, the assumptions include a single diallelic autosomal locus with constant selection values, constant mutation rates, and normal Mendelian segregation. Selection is assumed to occur through viability differences only. Let μ and ν represent the mutation rates from A_1 to A_2 and from A_2 to A_1 , respectively. The relative fitness of each of the three genotypes is designated as w_1 for A_1A_1 , w_2 for A_1A_2 and w_3 for A_2A_2 . Let $P_{(1,t-1)}$, $P_{(2,t-1)}$, and $P_{(3,t-1)}$ be the frequencies of the genotypes A_1A_1 , A_1A_2 and A_2A_2 in any generation $t-1$. The P 's obviously must sum to 1. The relation between the genotype frequencies

The authors are student and professor in the Division of Biology and Medicine, Brown University, Providence, Rhode Island 02912. They thank Michael Clegg, Peter Karciva, Peter Smouse, and Frank Stewart for invaluable advice and suggestions concerning the analysis and preparation of the manuscript. Please address reprint requests to Dr. Kidwell.

© 1983, American Genetic Association.

in generation $t-1$ and relative numbers in generation t can be presented as follows:

Genotype and frequency in any generation $t-1$	Genotype and frequency in the next generation, t		
	A_1A_1	A_1A_2	A_2A_2
$A_1A_1: P_{(1,t-1)}$	$w_1(1-\mu)^2 P_{(1,t-1)}$	$2w_2\mu(1-\mu)P_{(1,t-1)}$	$w_3\mu^2 P_{(1,t-1)}$
$A_1A_2: P_{(2,t-1)}$	$\frac{w_1(1-\mu+\nu)^2}{4} P_{(2,t-1)}$	$\frac{w_2(1-\mu+\nu)(1+\mu-\nu)}{2} P_{(2,t-1)}$	$\frac{w_3(1+\mu-\nu)^2}{4} P_{(2,t-1)}$
$A_2A_2: P_{(3,t-1)}$	$w_1\nu^2 P_{(3,t-1)}$	$2w_2\nu(1-\nu) P_{(3,t-1)}$	$w_3(1-\nu)^2 P_{(3,t-1)}$

and

$$\left. \begin{aligned} P_{(1,t)} &= w_1 \left\{ (1-\mu)^2 P_{(1,t-1)} + \frac{(1-\mu+\nu)^2}{4} P_{(2,t-1)} + \nu^2 P_{(3,t-1)} \right\} \\ P_{(2,t)} &= w_2 \left\{ 2\mu(1-\mu)P_{(1,t-1)} + \frac{(1-\mu+\nu)(1+\mu-\nu)}{2} P_{(2,t-1)} + 2\nu(1-\nu)P_{(3,t-1)} \right\} \dots 1.1 \\ P_{(3,t)} &= w_3 \left\{ \mu^2 P_{(1,t-1)} + \frac{(1+\mu-\nu)^2}{4} P_{(2,t-1)} + (1-\nu)^2 P_{(3,t-1)} \right\} \end{aligned} \right\}$$

Since the derivation of the factors $(1-\mu)^2$, $(1-\mu+\nu)^2/4$, etc. may not be self-evident, Kidwell⁵ illustrated their calculation for the heterozygote A_1A_2 . It is reproduced here for the convenience of the reader. There are four possible outcomes at gametogenesis, each with a known probability of occurrence:

- 1) A_1 does not mutate and A_2 does not mutate, $P(1) = (1-\mu)(1-\nu)$;
- 2) A_1 does not mutate and A_2 mutates to A_1 , $P(2) = \nu(1-\mu)$;
- 3) A_1 mutates to A_2 and A_2 does not mutate to A_1 , $P(3) = \mu(1-\nu)$;
- 4) A_1 mutates to A_2 and A_2 mutates to A_1 , $P(4) = \mu\nu$.

Hence, the frequency of the A_1 gamete is $(1-\mu+\nu)/2$ and the frequency of the A_2 gamete is $(1+\mu-\nu)/2$. On selfing the frequency of the A_1A_1 genotype is obviously $(1-\mu+\nu)^2/4$, etc.

Letting $w_1(1-\mu)^2 = a_{11}$; $2w_2\mu(1-\mu) = a_{12}$, etc., we can write the recursion equations 1.1 in matrix form, i.e.,

$$\begin{bmatrix} P_{(1,t)} \\ P_{(2,t)} \\ P_{(3,t)} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \times \begin{bmatrix} P_{(1,t-1)} \\ P_{(2,t-1)} \\ P_{(3,t-1)} \end{bmatrix} \dots 1.2$$

or $[P_{i,t}] = [A_{ij}] \times [P_{i,t-1}]$.

The elements of the matrix $[A_{ij}]$ are the expected rates that genotype j in one generation will give rise to genotype i in the next generation. It should be noted that $[A]$ is

neither a probability transition matrix nor a matrix of probabilities. The column vector $[P_{i,t}]$ does not in general sum to unity and must be normalized by dividing each element of $[P_{i,t}]$ by the sum of the three elements prior to substituting it for the vector $[P_{i,t-1}]$ in each generation. In this model the selective value of each progeny genotype is independent of the parental genotype from which it arose.

At equilibrium

$$[P_{(i,t)}] = [P_{(i,t-1)}] \text{ for all } i.$$

Although analytical methods are available and feasible for the single-locus case, they were not used because the equilibrium expressions are much too detailed to permit biological generalizations. Instead, equilibrium frequencies were obtained by iterating the recursion equation 1.1 until $[P_{(i,t)}] - [P_{(i,t-1)}] = \text{some arbitrary tolerance, } K, (10^{-11} \text{ in this case})$. An alternative procedure determines the dominant root of the characteristic equation of $[A_{ij}]$. The former method has the advantage of showing how the equilibrium values are approached, yet requires no more computation than do numerical routines for calculating dominant eigenvalues. The dominant root method was, however, used to confirm some of our conclusions in the two-locus case.

Equations 1.1 were evaluated for a large number of parameter values reflecting several different models of selection. A small but representative sample is given in Table 1. The most obvious conclusion is that selfing is the

major force in changing genotype frequency. This is dramatically illustrated in Table 1 and Figure 1, for the case of equal fore and back mutation rates and selective neutrality. Even under complete dominance and with mutation favoring the recessive allele, the frequency of the heterozygote is always less than that of the recessive homozygote. When selection is reinforced by selfing, the favored homozygote increases far more rapidly than with the neutral model, as shown in Figure 2. When selection directly opposes the effect of selfing,

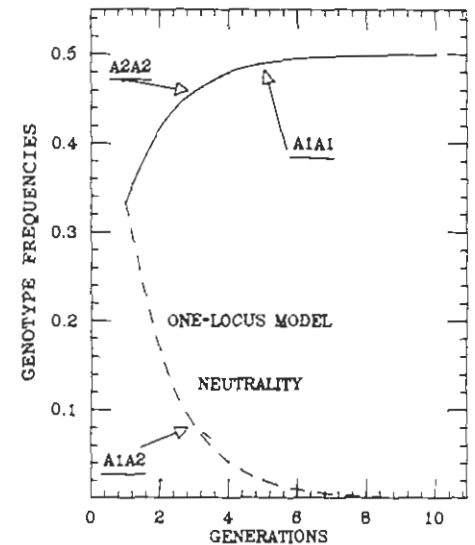


FIGURE 1 The approach to genotype frequency equilibrium of a single locus under neutrality ($\mu = \nu = 10^{-5}$; $w_1 = w_2 = w_3 = 1$).

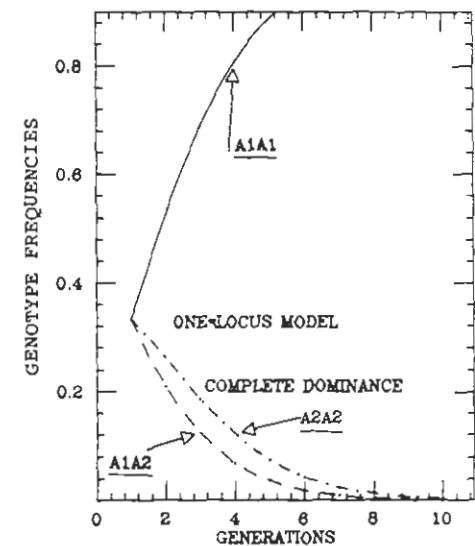


FIGURE 2 The approach to genotype frequency equilibrium of a single locus with complete dominance ($\mu = 10^{-5}$, $\nu = 10^{-6}$; $w_1 = w_2 = 1$, $w_3 = 0.5$).

i.e., the heterozygote is favored, except for extreme selection values the effect of selfing is moderated only slightly and still predominates, as can be seen by examination of Table I and Figure 3. Finally, it is obvious from Table I and the figures that mutation is a trivial force unless the selection coefficients are extremely small.

Two-locus case

The consequences of self-fertilization for two autosomal loci in the absence of mutation and selection have been considered by Diamantis¹, Haldane and Waddington³, Karlin⁴, Kimura⁶, Robbins⁸, and Wright⁹. Here we consider the joint effects of self-fertilization, linkage, selection, and mutation. The assumptions include: 1) two diallelic autosomal loci, 2) constant selective values for each genotype, 3) selection occurs only through viability differences, 4) constant fore and back

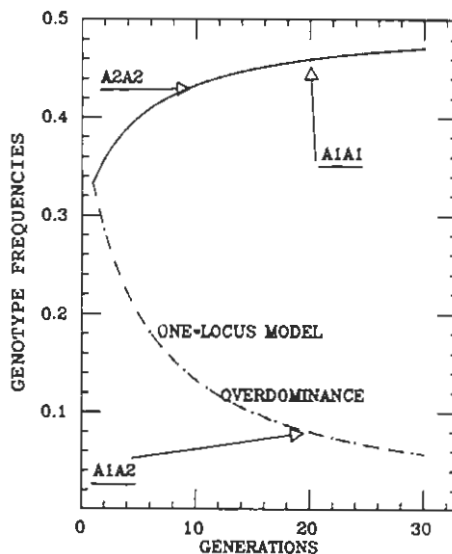


FIGURE 3 The approach to genotype frequency equilibrium of a single locus with overdominance ($\mu = \nu = 10^{-5}$; $w_1 = w_3 = 0.5$, $w_2 = 1.0$).

Table I. Single locus equilibrium genotype and allele frequencies for four selection models and some representative parameter values

Parameter			Frequency of					
μ	ν	w_1	w_2	w_3	A_1A_1	A_1A_2	A_2A_2	A_1
Neutrality								
10^{-5}	10^{-5}	1.0	1.0	1.0	4.99980×10^{-1}	3.99980×10^{-5}	4.99980×10^{-1}	5.00000×10^{-1}
10^{-5}	10^{-6}	1.0	1.0	1.0	4.98193×10^{-1}	2.19350×10^{-5}	5.01785×10^{-1}	4.98204×10^{-1}
Complete dominance								
10^{-5}	10^{-5}	1.0	1.0	0.5	9.99950×10^{-1}	3.99988×10^{-5}	9.99980×10^{-6}	4.98203×10^{-1}
10^{-10}	10^{-10}	1.0	1.0	0.5	9.99999×10^{-1}	3.99999×10^{-10}	9.99999×10^{-11}	4.98204×10^{-1}
10^{-5}	10^{-6}	1.0	1.0	0.5	9.99950×10^{-1}	3.99984×10^{-5}	1.00000×10^{-5}	9.99970×10^{-1}
Incomplete dominance								
10^{-5}	10^{-5}	1.0	0.9	0.5	9.99959×10^{-1}	3.27266×10^{-5}	8.18177×10^{-6}	9.99975×10^{-1}
10^{-5}	10^{-6}	1.0	0.9	0.5	9.99959×10^{-1}	3.27263×10^{-5}	8.18201×10^{-6}	9.99975×10^{-1}
10^{-5}	10^{-5}	1.0	0.5	0.1	9.99986×10^{-1}	1.33333×10^{-5}	3.70387×10^{-7}	9.99993×10^{-1}
10^{-5}	10^{-6}	1.0	0.5	0.1	9.99986×10^{-1}	1.33333×10^{-5}	3.70394×10^{-7}	9.99993×10^{-1}
Overdominance								
10^{-5}	10^{-6}	0.5	1.0	0.5	4.95425×10^{-1}	7.56078×10^{-3}	4.97014×10^{-1}	4.99205×10^{-1}
10^{-5}	10^{-5}	0.5	1.0	0.5	4.95314×10^{-1}	9.37116×10^{-3}	4.95314×10^{-1}	5.00000×10^{-1}
10^{-5}	10^{-6}	0.01	1.0	0.01	5.05041×10^{-3}	9.89899×10^{-1}	5.05060×10^{-3}	5.00000×10^{-1}

Table II. Four two-locus selection models

Genotype	Fitness	Model			
		1	2	3	4
A_1B_1/A_1B_1	$w_1 =$	1	1	1	$1 + 2s$
A_1B_1/A_1B_2	$w_2 =$	$1 - s$	$1 + s$	$1 - s$	$1 + s$
A_1B_1/A_2B_1	$w_3 =$	$1 - s$	$1 + s$	$1 - s$	$1 + s$
A_1B_1/A_2B_2	$w_4 =$	$1 - 2s$	$(1 + s)^2$	$(1 - s)^2$	1
A_1B_2/A_1B_2	$w_5 =$	$1 - 2s$	1	$1 - 2s$	1
A_1B_2/A_2B_1	$w_6 =$	$1 - 2s$	$(1 + s)^2$	$(1 - s)^2$	1
A_1B_2/A_2B_2	$w_7 =$	$1 - 3s$	$1 + s$	$(1 - s)(1 - 2s)$	$1 + s$
A_2B_1/A_2B_1	$w_8 =$	$1 - 2s$	1	$1 - 2s$	1
A_2B_1/A_2B_2	$w_9 =$	$1 - 3s$	$1 + s$	$(1 - s)(1 - 2s)$	$1 + s$
A_2B_2/A_2B_2	$w_{10} =$	$1 - 4s$	1	$(1 - 2s)^2$	$1 + 2s$

mutation rates that may differ between loci, 5) mutation at each locus occurs after gamete formation, i.e., after crossing over has occurred.

Let μ_a , ν_a , μ_b , and ν_b be the fore and back mutation rates for the A and B loci, respectively. The probability of recombination is denoted by r , where $(0 \leq r \leq 0.5)$. (We originally developed the model permitting different recombination fractions for male and female gametophytes. This feature is omitted here in the interest of simplicity.) The fitness of each genotype is designated as follows:

$$\text{Genotype} \quad \begin{array}{ccccc} A_1B_1 & A_1B_1 & A_1B_1 & A_1B_1 & A_1B_2 \\ A_1B_1 & A_1B_2 & A_2B_1 & A_2B_2 & A_1B_2 \end{array}$$

$$\text{Fitness} \quad w_1 \quad w_2 \quad w_3 \quad w_4 \quad w_5$$

$$\text{Genotype} \quad \begin{array}{ccccc} A_1B_2 & A_1B_2 & A_2B_1 & A_2B_1 & A_2B_2 \\ A_2B_1 & A_2B_2 & A_2B_1 & A_2B_2 & A_2B_2 \end{array}$$

$$\text{Fitness} \quad w_6 \quad w_7 \quad w_8 \quad w_9 \quad w_{10}$$

As in the one-locus case, the relation between the genotype frequencies in generation $t-1$ and relative numbers in generation t is:

$$[P_{(t,t)}] = [A_{ij}] \times [P_{(t,t-1)}] \quad \dots 2.1$$

where $[P_{(t,t)}]$ and $[P_{(t,t-1)}]$ are 10 by 1 column vectors and $[A_{ij}]$ is a 10 by 10 matrix of rates analogous to those in the one-locus models.

Calculation of the a_{ij} is straightforward but tedious. Space limitations preclude writing out the entire set of 100 rates; hence their calculation is illustrated for $a_{6,4}$ —the rate that an individual of genotype 4, A_1B_1/A_2B_2 , in generation $t-1$ will give rise to an individual of genotype 6, A_1B_2/A_2B_1 , in generation t . The parent will produce four types of gametes in the following proportions:

Gamete	Frequency
A_1B_1	$[(1-r)/2] \{ (1-\mu_a)(1-\mu_b) + \nu_a\nu_b \}$ $+ [r/2] \{ \nu_a(1-\mu_b) + \nu_b(1-\mu_a) \} = g_{4,1}$
A_1B_2	$[(1-r)/2] \{ \mu_b(1-\mu_a) + \nu_a(1-\nu_b) \}$ $+ [r/2] \{ (1-\mu_a)(1-\nu_b) + \nu_a\mu_b \} = g_{4,2}$
A_2B_1	$[(1-r)/2] \{ \mu_a(1-\mu_b) + \nu_b(1-\nu_a) \}$ $+ [r/2] \{ \mu_a\nu_b + (1-\nu_a)(1-\mu_b) \} = g_{4,3}$
A_2B_2	$[(1-r)/2] \{ \mu_a\mu_b + (1-\nu_a)(1-\nu_b) \}$ $+ [r/2] \{ \mu_a(1-\nu_b) + \mu_b(1-\nu_a) \} = g_{4,4}$

where the subscripts on the g 's refer to the parental genotype, and type of gamete produced. The frequencies are designated:

$$g_{4,1}, g_{4,2}, g_{4,3}, g_{4,4}$$

thus

$$a_{6,4} = 2g_{4,2} \times g_{4,3} \times w_6$$

where w_6 is the fitness of genotype 6. At equilibrium

$$[P_{(t,t)}] = [P_{(t,t-1)}] \quad \text{for all } i.$$

In order to study this two-locus system, the four selection models described in Table II were analyzed. Model 1 is additive directional selection against the recessive allele; model 2 represents multiplicative overdominance; model 3 is multiplicative directional selection favoring A_1B_1/A_1B_1 , and model 4 is disruptive selection. The recurrence relation 2.1 was used to obtain numerically the equilibrium frequencies for these four selection models using a number of parameter values. The results are given in Table III. These are but a small sample of a virtually infinite number of selection models and parameter values that can be constructed. Caution should be exercised in attempting to draw general conclusions from this small set of models that are in many ways artificial and employ a very restricted set of parameter values.

Interest at the two-locus level centers on the

influence of linkage on equilibrium gene and genotype distributions, selection and mutation having effects similar to those observed in the one-locus models. As expected, the degree of linkage has no effect upon the equilibrium frequencies but tight linkage slows the approach to equilibrium. This transient effect is illustrated in Figures 4 and 5 in which three representative genotypes of the total of 10 were graphed. In the case of tight linkage, the slower approach to equilibrium allows the double heterozygote to decrease less rapidly than the single heterozygote due to overdominant selection. However, in the long run, the effect of selfing predominates and the frequency of the double heterozygote is less than that of the single heterozygote. In the absence of linkage, the selfing effect predominates from the outset. This is shown in Figures 4 and 5.

Disruptive selection generates the maximum amount of linkage disequilibrium which is defined as

$$d = ru - st,$$

where r , s , t , and u denote the relative frequencies of the gametic types A_1B_1 , A_1B_2 , A_2B_1 , and A_2B_2 , respectively. This is not due to the tightness of linkage but to the fact that selection favors the genotypes A_1B_1/A_1B_1 and A_2B_2/A_2B_2 equally, and that selection is reinforced by selfing. The other models generate little, if any, linkage disequilibrium.

Three-locus case

Extension to the three-locus case is straightforward, but the level of complexity increases substantially. All of the assumptions

Table III. Joint effects of linkage, mutation, selection and selfing on gene frequency [$P(A_1)$ and $P(B_1)$], heterozygosity [$h(A)$, $h(B)$ and $h(D)$], and linkage disequilibrium. The selection models are defined in Table II. Mutation rates are assumed to be equal at the two loci, $\mu_A = \mu_B$; $\nu_A = \nu_B$. $P(A_1)$ and $P(B_1)$ are the frequencies of the alleles A_1 and B_1 . $h(A)$ and $h(B)$ are the frequencies of the single heterozygotes; $h(D)$ is the frequency of double heterozygotes and d is the degree of linkage disequilibrium. Since the models are symmetric, $P(A_1) = P(B_1)$ and $h(A) = h(B)$.

Fitness model	Parameter values				Equilibrium statistics			
	μ	ν	s	r	$P(A_1) = P(B_1)$	$h(A) = h(B)$	$h(D)$	d
1	10^{-5}	10^{-5}	0.01	0.01	9.991892×10^{-1}	3.920727×10^{-5}	2.232828×10^{-9}	-2.351461×10^{-8}
	10^{-5}	10^{-6}	0.01	0.01	9.991878×10^{-1}	3.917885×10^{-5}	2.229624×10^{-9}	-2.366824×10^{-8}
	10^{-5}	10^{-5}	0.15	0.01	9.999680×10^{-1}	2.956495×10^{-5}	1.056244×10^{-9}	6.711218×10^{-11}
	10^{-5}	10^{-6}	0.15	0.01	9.999680×10^{-1}	2.956449×10^{-5}	1.056220×10^{-9}	6.711295×10^{-11}
	10^{-5}	10^{-5}	0.01	0.50	9.991892×10^{-1}	3.920727×10^{-5}	1.537003×10^{-9}	-2.352587×10^{-8}
	10^{-5}	10^{-6}	0.01	0.50	9.991877×10^{-1}	3.917885×10^{-5}	1.534776×10^{-9}	-2.367959×10^{-8}
	10^{-5}	10^{-5}	0.15	0.50	9.999680×10^{-1}	2.956503×10^{-5}	8.410795×10^{-10}	1.110611×10^{-10}
	10^{-5}	10^{-6}	0.15	0.50	9.999680×10^{-1}	2.956457×10^{-5}	8.410509×10^{-10}	1.110629×10^{-10}
2	10^{-5}	10^{-5}	0.1	0.01	5.000000×10^{-1}	4.888587×10^{-5}	4.095392×10^{-9}	1.405126×10^{-15}
	10^{-5}	10^{-6}	0.1	0.01	4.982053×10^{-1}	2.680935×10^{-5}	1.231764×10^{-9}	1.789177×10^{-12}
	10^{-5}	10^{-5}	0.4	0.01	5.000000×10^{-1}	9.341974×10^{-5}	1.177457×10^{-7}	1.874369×10^{-15}
	10^{-5}	10^{-6}	0.4	0.01	4.982811×10^{-1}	5.122003×10^{-5}	3.925641×10^{-8}	1.055523×10^{-8}
	10^{-5}	10^{-5}	0.1	0.50	5.000000×10^{-1}	4.888553×10^{-5}	2.389795×10^{-9}	1.419004×10^{-15}
	10^{-5}	10^{-6}	0.1	0.50	4.982052×10^{-1}	2.680925×10^{-5}	7.187356×10^{-10}	4.312089×10^{-14}
	10^{-5}	10^{-5}	0.4	0.50	5.000000×10^{-1}	9.331540×10^{-5}	8.707763×10^{-9}	2.519686×10^{-15}
	10^{-5}	10^{-6}	0.4	0.50	4.982125×10^{-1}	5.117884×10^{-5}	2.619273×10^{-9}	1.491494×10^{-12}
3	10^{-5}	10^{-5}	0.1	0.01	9.999509×10^{-1}	3.272657×10^{-5}	1.416416×10^{-9}	6.292400×10^{-12}
	10^{-5}	10^{-6}	0.1	0.01	9.999509×10^{-1}	3.272561×10^{-5}	1.416353×10^{-9}	6.292694×10^{-12}
	10^{-5}	10^{-5}	0.4	0.01	9.999904×10^{-1}	1.714277×10^{-5}	3.247162×10^{-10}	4.025319×10^{-12}
	10^{-5}	10^{-6}	0.4	0.01	9.999904×10^{-1}	1.714276×10^{-5}	3.247168×10^{-10}	4.025485×10^{-12}
	10^{-5}	10^{-5}	0.1	0.50	9.999509×10^{-1}	3.272658×10^{-5}	1.071029×10^{-9}	2.843433×10^{-24}
	10^{-5}	10^{-6}	0.1	0.50	9.999509×10^{-1}	3.272561×10^{-5}	1.070966×10^{-9}	3.360421×10^{-24}
	10^{-5}	10^{-5}	0.4	0.50	9.999904×10^{-1}	1.714278×10^{-5}	2.938748×10^{-10}	6.139231×10^{-26}
	10^{-5}	10^{-6}	0.4	0.50	9.999904×10^{-1}	1.714276×10^{-5}	2.938742×10^{-10}	7.431701×10^{-26}
4	10^{-5}	10^{-5}	0.1	0.01	5.000000×10^{-1}	3.384583×10^{-5}	1.516914×10^{-9}	2.499408×10^{-1}
	10^{-5}	10^{-6}	0.1	0.01	4.978986×10^{-1}	1.855144×10^{-5}	7.628745×10^{-10}	2.499631×10^{-1}
	10^{-5}	10^{-5}	0.4	0.01	5.000000×10^{-1}	2.545432×10^{-5}	6.939918×10^{-10}	2.499793×10^{-1}
	10^{-5}	10^{-6}	0.4	0.01	4.975382×10^{-1}	1.394362×10^{-5}	3.487772×10^{-10}	2.499826×10^{-1}
	10^{-5}	10^{-5}	0.1	0.50	5.000000×10^{-1}	3.384590×10^{-5}	1.133563×10^{-9}	2.499408×10^{-1}
	10^{-5}	10^{-6}	0.1	0.50	4.978981×10^{-1}	1.855146×10^{-5}	5.700730×10^{-10}	2.499631×10^{-1}
	10^{-5}	10^{-5}	0.4	0.50	5.000000×10^{-1}	2.545440×10^{-5}	5.864977×10^{-10}	2.499793×10^{-1}
	10^{-5}	10^{-6}	0.4	0.50	4.975378×10^{-1}	1.394365×10^{-5}	2.947515×10^{-10}	2.499826×10^{-1}

of the two-locus model hold. Let $\mu_a, \mu_b,$ and μ_c be the forward mutation rates and $\nu_a, \nu_b,$ and ν_c be the back mutation rates for the three-locus model. Under the stated assumptions the allele at each locus in a gamete pro-

duced by meiosis can mutate to the other. It is convenient to let

$$y_a = 1 - \mu_a; y_b = 1 - \mu_b; y_c = 1 - \mu_c;$$

$$z_a = 1 - \nu_a; z_b = 1 - \nu_b \text{ and } z_c = 1 - \nu_c.$$

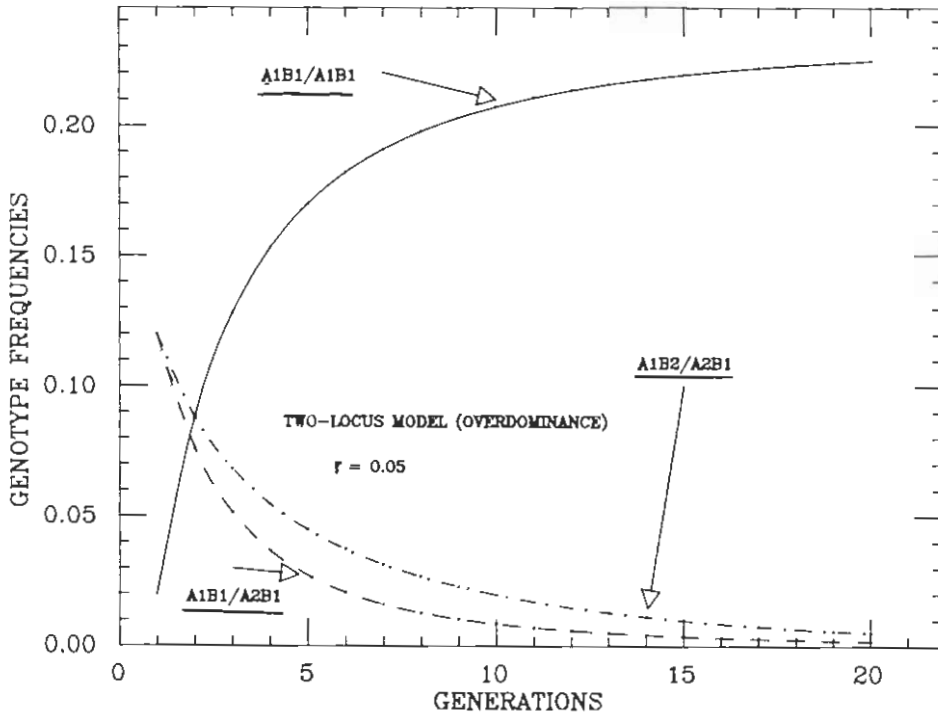


FIGURE 4 The approach to genotype frequency equilibrium for three of the 10 possible genotypes for two overdominant loci ($\mu = \nu = 10^{-5}$; $s = 0.4$ see Table I; $r = 0.05$).

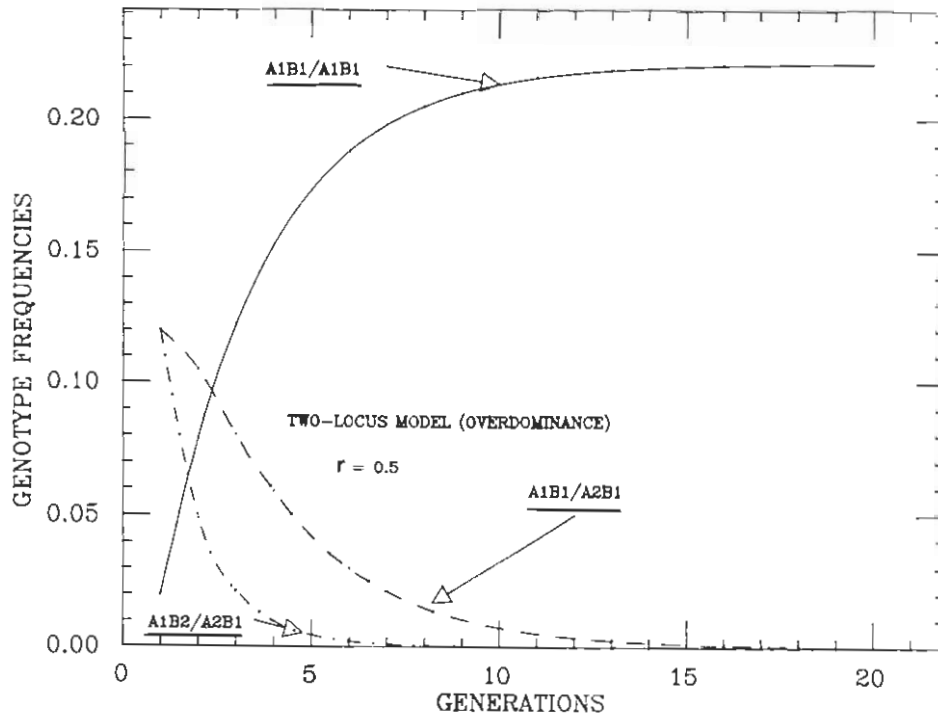


FIGURE 5 The approach to genotype frequency equilibrium for three of the 10 possible genotypes for two overdominant loci ($\mu = \nu = 10^{-5}$; $s = 0.4$ see Table I; $r = 0.5$).

The recurrence relations and equilibrium equations are of the same form as for the one- and two-locus cases, i.e.,

$$[P_{(i,j)}] = [A_{ij}] \times [P_{(i,j-1)}] \dots 3.1$$

where $[P_{(i,j)}]$ and $[P_{(i,j-1)}]$ are 36 by 1 column vectors and $[A_{ij}]$ is a 36 by 36 matrix.

Calculation of the a_{ij} is again straightforward but much more tedious than in the two-locus case. The linear order of the loci is arbitrarily taken as $A - B - C$. Consider the genotype $A_1B_1C_1/A_2B_2C_2$ where the subscripts designate parental origin of the chromosome and may or may not also designate different alleles. At meiosis, four mutually exclusive events are possible, one of which must occur. Each event leads to the production of two unique gametes.

Event	Description	Gametes
1	A crossover between A and B , but not between B and C	$A_1B_2C_2$ $A_2B_1C_1$
2	A crossover between B and C , but not between A and B	$A_1B_1C_2$ $A_2B_2C_1$
3	A crossover between A and B and between B and C , i.e., a double crossover	$A_1B_2C_1$ $A_2B_1C_2$
4	No crossovers	$A_1B_1C_1$ $A_2B_2C_2$

We denote the probabilities of these events as $C_1, C_2, C_3,$ and C_4 , respectively, and observe that $C_1 + C_2 + C_3 + C_4 = 1$. The C 's are related to the conventional recombination fractions as follows:

$$r_1 = C_1 + C_3$$

$$r_2 = C_2 + C_3$$

$$r_3 = C_3$$

where r_1 is regarded as the probability of a crossover between A and B ; r_2 is the probability of a crossover between B and C ; and $r_3 = C_3$ is the probability of a double crossover. It is customary to think of $r_3 = r_1r_2$, but this restriction is not essential.

The C 's can be expressed in terms of the r 's as follows:

$$C_1 = r_1 - r_3$$

$$C_2 = r_2 - r_3$$

$$C_3 = r_3$$

$$C_4 = 1 - (C_1 + C_2 + C_3) = 1 - r_1 - r_2 + r_3$$

Calculation of the a_{ij} proceeds as follows: 1) Determine the gametic output and frequency of each of the j ($= 36$) parental gen-

Table IV. Gamete frequency after meiosis followed by mutation. The gametes in the top row mutate to those in the left hand column with the frequencies given in the body of the table, e.g., $A_2B_1C_2$ mutates to $A_1B_2C_2$ with a frequency of $\nu_3\mu_bz_c$

Gametes	Gametes							
	$A_1B_1C_1$	$A_1B_1C_2$	$A_1B_2C_1$	$A_1B_2C_2$	$A_2B_1C_1$	$A_2B_1C_2$	$A_2B_2C_1$	$A_2B_2C_2$
$A_1B_1C_1$	$y_a\nu_b\nu_c$	$y_a\nu_b\nu_c$	$y_a\nu_b\nu_c$	$y_a\nu_b\nu_c$	$\nu_a y_b\nu_c$	$\nu_a y_b\nu_c$	$\nu_a\nu_b\nu_c$	$\nu_a\nu_b\nu_c$
$A_1B_1C_2$	$y_a\nu_b\mu_c$	$y_a\nu_bz_c$	$y_a\nu_b\mu_c$	$y_a\nu_bz_c$	$\nu_a y_b\mu_c$	$\nu_a y_bz_c$	$\nu_a\nu_b\mu_c$	$\nu_a\nu_bz_c$
$A_1B_2C_1$	$y_a\mu_b\nu_c$	$y_a\mu_b\nu_c$	$y_a z_b\nu_c$	$y_a z_b\nu_c$	$\nu_a\mu_b\nu_c$	$\nu_a\mu_b\nu_c$	$\nu_a z_b\nu_c$	$\nu_a z_b\nu_c$
$A_1B_2C_2$	$y_a\mu_b\mu_c$	$y_a\mu_bz_c$	$y_a z_b\mu_c$	$y_a z_bz_c$	$\nu_a\mu_b\mu_c$	$\nu_a\mu_bz_c$	$\nu_a z_b\mu_c$	$\nu_a z_bz_c$
$A_2B_1C_1$	$\mu_a y_b\nu_c$	$\mu_a y_b\nu_c$	$\mu_a\nu_b\nu_c$	$\mu_a\nu_b\nu_c$	$z_a y_b\nu_c$	$z_a y_b\nu_c$	$z_a\nu_b\nu_c$	$z_a\nu_b\nu_c$
$A_2B_1C_2$	$\mu_a y_b\mu_c$	$\mu_a y_bz_c$	$\mu_a\nu_b\mu_c$	$\mu_a\nu_bz_c$	$z_a y_b\mu_c$	$z_a y_bz_c$	$z_a\nu_b\mu_c$	$z_a\nu_bz_c$
$A_2B_2C_1$	$\mu_a\mu_b\nu_c$	$\mu_a\mu_b\nu_c$	$\mu_a z_b\nu_c$	$\mu_a z_b\nu_c$	$z_a\mu_b\nu_c$	$z_a\mu_b\nu_c$	$z_a z_b\nu_c$	$z_a z_b\nu_c$
$A_2B_2C_2$	$\mu_a\mu_b\mu_c$	$\mu_a\mu_bz_c$	$\mu_a z_b\mu_c$	$\mu_a z_bz_c$	$z_a\mu_b\mu_c$	$z_a\mu_bz_c$	$z_a z_b\mu_c$	$z_a z_bz_c$

Table V. Gametic output of the genotype $A_1B_1C_1/A_1B_2C_2$ due to meiosis and mutation

Gamete and its freq. after meiosis and before mutation	Gamete frequency after meiosis and mutation							
	$A_1B_1C_1$	$A_1B_1C_2$	$A_1B_2C_1$	$A_1B_2C_2$	$A_2B_1C_1$	$A_2B_1C_2$	$A_2B_2C_1$	$A_2B_2C_2$
$A_1B_1C_1$ $1/2(C_1 + C_4)$	$y_a\nu_b\nu_c$	$y_a\nu_b\mu_c$	$y_a\mu_b\nu_c$	$y_a\mu_b\mu_c$	$\mu_a y_b\nu_c$	$\mu_a y_b\mu_c$	$\mu_a\mu_b\nu_c$	$\mu_a\mu_b\mu_c$
$A_1B_1C_2$ $1/2(C_2 + C_3)$	$y_a\nu_b\nu_c$	$y_a\nu_bz_c$	$y_a\mu_b\nu_c$	$y_a\mu_bz_c$	$\mu_a y_b\nu_c$	$\mu_a y_bz_c$	$\mu_a\mu_b\nu_c$	$\mu_a\mu_bz_c$
$A_1B_2C_1$ $1/2(C_2 + C_3)$	$y_a\nu_b\nu_c$	$y_a\nu_b\mu_c$	$y_a z_b\nu_c$	$y_a z_b\mu_c$	$\mu_a\nu_b\nu_c$	$\mu_a\nu_b\mu_c$	$\mu_a z_b\nu_c$	$\mu_a z_b\mu_c$
$A_1B_2C_2$ $1/2(C_1 + C_4)$	$y_a\nu_b\nu_c$	$y_a\nu_bz_c$	$y_a z_b\nu_c$	$y_a z_bz_c$	$\mu_a\nu_b\nu_c$	$\mu_a\nu_bz_c$	$\mu_a z_b\nu_c$	$\mu_a z_bz_c$
Sums	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8

otypes as a result of meiosis. For example, for $a_{1,4}$ this would be (gametes, above; frequencies, below):

$A_1B_1C_1$	$A_1B_1C_2$	$A_1B_2C_1$	$A_1B_2C_2$
$1/2(C_1 + C_4)$	$1/2(C_2 + C_3)$	$1/2(C_2 + C_3)$	$1/2(C_1 + C_4)$

where the 4th parental genotype is $A_1B_1C_1/A_1B_2C_2$.

2) Each of the gametes produced by meiosis can give rise to one of eight distinct types as a result of mutation or its absence at each of the three loci. For the reader's convenience, the first row of Table IV gives the eight possible kinds of gametes produced by meiosis. (Most parental genotypes will not produce all eight gametic types.) The left hand column gives the

eight possible gametic types that can arise as a result of mutation or its failure. The body of the table gives the associated probabilities. We will use the term "mutation coefficients" to identify them. Next we construct a two-way table similar to Table V. The top row identifies each of the eight possible gametic types. The left hand column lists the gametic types produced by meiosis and their frequency before mutation. Each element in the body of the table is the product of the frequency before mutation and the mutation coefficient. The frequency of each gametic type is given by the sum of the frequencies in each column. In this example, the frequency of $A_1B_1C_1$ is:

$$S_1 = 1/2(C_1 + C_4)\{y_a\nu_b\nu_c + y_a\nu_b\mu_c\} + 1/2(C_2 + C_3)\{y_a\nu_b\nu_c + y_a\nu_bz_c\}$$

3) The zygote production for each of the 36 selfing genotypes is obtained by constructing an 8 by 8 matrix of genotype frequencies by premultiplying a row vector of the frequencies of the eight gametic types by its transpose. The eight diagonal elements correspond to the homozygotes. Reciprocal off diagonals are combined to give the 28 heterozygotes. These genotype frequencies are multiplied by the appropriate fitness values and when properly identified used to form one of the 36 row or column vectors of $[A_{ij}]$. Equilibrium genotypic frequencies are calculated in the same manner as for one- or two-locus models.

The four selection models examined in the two-locus case were extended to three loci and numerical solutions for a number of parameter values obtained. The numerical results for three loci are similar to those obtained for two loci.

In conclusion, it should be noted that the theoretical foundations for the formulation of the recurrence equations have existed since the early decades of this century, but only recently has the availability of high speed computers made the numerical analysis possible.

References

- DIAMANTIS, B. An application of the generating function to the solution of a genetic problem. *J. Appl. Prob.* 10:437-440. 1973.
- HALDANE, J. B. S. The amount of heterozygosity to be expected in an approximately pure line. *J. Genet.* 32:375-391. 1936.
- and C. H. WADDINGTON. Inbreeding and linkage. *Genetics* 16:357-374. 1931.
- KARLIN, S. Equilibrium behavior of population genetic models with non-random mating. *J. Appl. Prob.* 5:231-313. 1968.
- KIDWELL, J. F. Heterozygosity due to mutation in a self-fertilizing population. *J. Hered.* 60:41-42. 1969.
- KIMURA, M. A probability method for treating inbreeding systems, especially with linked genes. *Biometrics* 19:1-17. 1963.
- MENDEL, G. Experiments on plant hybrids. The original paper was published in *Verhandlungen des naturforschenden Vereines in Brünn* 4 (1865), Abhandlung, p 3-47. The reference here is to the translation in *The Origin of Genetics: A Mendel Source Book*. C. Stern and E. Sherwood, Eds. W. H. Freeman Co., San Francisco. 1966.
- ROBBINS, R. B. Some applications of mathematics to breeding problems. III. *Genetics* 3:375-389. 1918.
- WRIGHT, S. Inbreeding and homozygosity. *Proc. Natl. Acad. Sci.* 15:411-420. 1933.