EFFECTS OF SMALL POPULATION SIZE AND SELECTION ON THE EXPECTED TIME TO FIXATION OF A FAVORABLE ALLELE WITH MULTIPLE LOCI

by 632

LARRY SCOTT

B. S., Kansas State University, 1967

A MASTER'S REPORT

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Statistics
and Computer Science

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1970

Approved by:

R. F. Nassar

Major Professor

TABLE OF CONTENTS

Many laboratory populations as well as plant and animal breeding programs contain only a small number of individuals. It is important to note that under these conditions the success or failure of a gene is determined not only by its fitness but also by the random sampling of gametes due to the finite gene pool (random genetic drift). This concept was first investigated by R. A. Fisher in 1922. Later, Fisher (1930), Wright (1931), and Haldane (1932) worked out methods for finding the ultimate probability of survival of a mutant gene when the process of change in gene frequency is stochastic.

Wright (1945) was the first to use the Fokker-Plank equation (Kolmogorov forward equation) in his solution of gene frequency distributions under the joint effect of selection and random genetic drift.

Exact solutions for the entire process of change of gene frequencies were formulated by Kimura (1955, 1957, 1962, and 1964) using a continuous Markov chain in time and space. These solutions were extended by Robertson (1962) and Ewens (1963) to include overdominance and the expected time to fixation of a favorable allele.

Carr and Nassar (1969), using a finite Markov chain to describe the stochastic change in gene frequencies in a finite population, solved for the effects of selection and random sampling of gametes on the probability of and the expected time to fixation of a favorable allele in the one-locus case.

Of particular relevance to this work is the discovery that selection for overdominance accelerated instead of retarded the time to fixation when the equilibrium gene frequency was above 0.8. It is essential to know whether the same would be true in the case of multiple loci with and without linkage. Hence it is the purpose of this paper to investigate for n-loci the effects of small population size and selection on the expected time to fixation of a favorable allele in the case of overdominance.

MATHEMATICAL MODEL

Consider a population of finite size N, mating randomly with nonoverlapping generations. Let the model for one locus and two alleles be defined by

Genotype	Relative Fitness
AA	1 + s
Aa	1 + hs
aa	1

If we let x(t) be the number of A-genes in generation t, then Carr and Nassar (1969) have shown that, for this model, the stochastic process (x(0), x(1), x(2), . . ., x(t)) is a finite Markov chain with (2N - 1) transient states and (0, 2N) as its absorbing states. The finite Markov chain has a stochastic matrix, P, having as its ij^{th} element

$$P_{ij} = \binom{2N}{j} p^j (1-p)^{2N-j} \qquad \begin{matrix} i = 0,1,2, \ldots, 2N \\ j = 0,1,2, \ldots, 2N \end{matrix} \qquad (2.1)$$

where

$$p = \frac{2Ni(1 + hs) + i^2 s(1 - h)}{4N^2 + 4Nihs + i^2 s(1 - 2h)} \tag{2.2}$$

Probability of Ultimate Fixation

Feller (1957) has shown that the probability of ultimate fixation $f_i (i = 1, 2, \ldots, 2N)$ may be found by solving the system of equations

$$f_i - \sum_{j=1}^{M} P_{ij} f_j = f_i^{(1)} \tag{2.3}$$

where

$M = 2N - 1$

$f_i = \sum_{n=1}^{\infty} f_i(n)$, the probability of ultimate fixation, given that the initial state was i.

$f_i^{(n)} = Pr \cdot (\mathbf{X}(n) = 2N \,|\, \mathbf{X}(0) = i$ and $\mathbf{X}(r) < 2N$ $r = 1, 2, \ldots, n - 1)$, the probability that the population will be fixed in the n^{th} generation, given that the initial state was i.

Solving this system of equations and assuming no selection, we get

$$f_i = i/2N \qquad i = 1, 2, \ldots, M \tag{2.4}$$

which agrees with Kimura's solution for survival of a mutant gene.

If we rewrite (2.3) in matrix notation with $F = [f_i]$ an M by 1 vector, we get

$$F = QF = R \tag{2.5}$$

with

$$F = (I - Q)^{-1} R \tag{2.6}$$

where

$Q = [P_{ij}]$, the M by M stochastic matrix

$R = [P_{i,2N}]$, an M by 1 vector of transition probabilities from transient state $i (i = 1, 2, \ldots, M)$ to the absorbing state 2N.

Carr and Nassar (1969) have shown that $B = (I - Q)^{-1}$ does not exist but can be approximated. If we let

$$B_{(0)} = [\delta_{ij}/1 - P_{ij}]$$

and apply the recurrence relation

$$B_{(r)} = (2I - B_{(r-1)}\ (I - Q))B_{(r-1)} \qquad (2.7)$$

a sequence of approximations for B is obtained. They have also shown that this sequence converges rapidly to B.

Expected Time to Fixation

Let

t_i = the expected time to fixation of one or the other absorbing states, given initial state i.

E(T) = the expected time to fixation.

By making use of probability generating functions, Carr and Nassar (1969) demonstrated that the expected time to fixation can be expressed as

$$E(T) = (I - Q)^{-1}\delta, \qquad \delta' = [1, 1, \ldots, 1] \qquad (2.8)$$

where $(I - Q)^{-1}$ can be approximated by equation (2.7).

Multiple Loci

To generalize the one locus model to include multiple loci

we take, for example, the two-locus case. Let 0 and I represent the two alleles at each locus. If the probability of recombination between the two loci is c, then the ten possible genotypes are given in Table I.

If the population can be represented by $\mathbf{X}(t) = I(i_1, i_2, i_3)$ for any generation t (where i_1 = number of x_1, i_2 = number of x_2, and i_3 = number of x_3 in Table I), then we have a stochastic process $(\mathbf{X}(t);\ t = 0, 1, 2, \ldots)$ with a stochastic matrix P having

$$P_{IJ} = \frac{(2N)!}{j_1!\ j_2!\ j_3!\ j_4!}\ P_1^{j_1}\ P_2^{j_2}\ P_3^{j_3}\ P_4^{j_4} \qquad (2.9)$$

as its ij^{th} element.

P_{IJ} is defined as the probability of going from state $I(i_1, i_2, i_3)$ to state $J(j_1, j_2, j_3)$ in one generation, where $(j_1 + j_2 + j_3 + j_4 = 2N)$. From Table I

$$\begin{aligned} P_1 &= p^2(1 + s_{22}) + pq(1 + s_{21}) + pr(1 + s_{12}) \\ &\quad + pt(1 + s_{11})(1 - c) + qr(1 + s_{11})/\bar{w} \\ &= p(1 + s_1) - Dc(1 + s_{11})/\bar{w} \qquad (2.10) \\ &= \text{the probability of } x_1 \text{ gamete in generation } t + 1, \end{aligned}$$

where

D = linkage disequilibrium = $(pt - qr)$

$\bar{w}$ = average fitness of the population

$1 + s_i$ = the average fitness of the x_i gamete from Table I.

TABLE I

Genotype	Probability	Relative Fitness	Gametes			
			x_1 = 11	x_2 = 10	x_3 = 01	x_4 = 00
11/11	p^2	$1 + s_{22}$	1			
11/10	2pq	$1 + s_{21}$	1/2	1/2		
11/01	2pr	$1 + s_{12}$	1/2		1/2	
11/00 (Coup.)	2pt	$1 + s_{11}$	1/2(1-c)	1/2 c	1/2 c	1/2(1-c)
10/10	q^2	$1 + s_{20}$		1		
10/01 (Rep.)	2qr	$1 + s_{11}$	1/2 c	1/2(1-c)	1/2(1-c)	1/2 C
10/00	2qt	$1 + s_{10}$		1/2		1/2
01/01	r^2	$1 + s_{02}$			1	
01/00	2rt	$1 + s_{01}$			1/2	1/2
00/00	t^2	1				1

Note: $p + q + r + t = 1$.

Similarly,

$$P_2 = (q(1 + s_2) + Dc(1 + s_{11})/\bar{w} \qquad (2.11)$$

$$P_3 = (r(1 + s_3) + Dc(1 + s_{11})/\bar{w} \qquad (2.12)$$

$$P_4 = (t(1 + s_4) - Dc(1 + s_{11})/\bar{w} = 1 - P_1 - P_2 - P_3 \qquad (2.13)$$

It can be shown by methods similar to those used to obtain (2.5) and (2.8) that the probability of fixing a gamete $[a = x_1, x_2, x_3, x_4]$ is given by the equation

$$F_a = (I - Q)^{-1} R_a \qquad (2.14)$$

and the expected time to fixation is

$$E(T) = (I - Q)^{-1} \delta \text{ , (Carr and Nassar, 1969)} \qquad (2.15)$$

where $Q = [P_{IJ}]$ and M x M transition probability matrix,

$$M = \binom{2N + 3}{2N} - 4, \text{ and}$$

$R_a = [P_{I,I_a}]$ an M x 1 vector of transition probabilities from transient state $I = (i_1, i_2, i_3)$ to the absorbing state I_a, a = 11, 01, or 10.

If we wish to generalize further to n-loci, it can be shown that (2.14) and (2.15) are still appropriate. Some representative a values for n-loci are given in Table II.

SIMULATION

A program was written for the IBM 360/50 digital computer to simulate a diploid population of finite size N with non-overlapping generations. The model described earlier was used

TABLE II

Number of Loci	Gametes $a = x_1, x_2, \ldots, x_b;\ b = 2^n$
1	1, 0
2	11, 10, 01, 00
3	111, 110, 101, 100, 011, 010, 001, 000
5	11111, 11110, 11101, 11100, 11011, 11010
	11001, 11000, 10111, 10110, 10101, 10100
	10011, 10010, 10001, 10000, 01111, 01110
	01101, 01100, 01011, 01010, 01001, 01000
	00111, 00110, 00101, 00100, 00011, 00010
	00001, 00000

for 2, 3, 5, and 15 loci. An individual genotype was obtained as the sum of the appropriate values for the n-loci, $G = \sum_{i=1}^{n} G_i$ (Table III).

The phenotype of an individual was given by the expression

$$P_j = \sum_{i=1}^{n} G_i + e_j \qquad (3.1)$$

where

n = the number of loci

e_j = a random environmental effect simulated by adding twelve uniform random deviates (-1/2, 1/2) and multiplying by a constant C. Thus each e_j is distributed as an independent normal random variable with mean zero and variance C^2.

TABLE III

Genotype	Relative Fitness	Genotypic Value G_i	Frequency
$\frac{1}{1}$	1 + s	1	q^2
$\frac{1}{0}$	1 + hs	1.125	$2q(1 - q)$
$\frac{0}{0}$	1	0	$(1 - q)^2$

The phenotypic variance in the population was given by

$$\sigma_p^2 = \sigma_G^2 + \sigma_e^2 \qquad (3.2)$$

The genotypic value at a locus (Table III) was determined such that the equilibrium gene frequency ($\hat{q}$) in an infinitely large population would be 0.9 ($\hat{q} = \frac{h}{2h-1}$, $h = 1.125$). Since from previous work (Carr and Nassar, 1969), with one locus, strong selection with a case of overdominance ($\hat{q} = 0.9$, or $h = 1.125$) caused the greatest acceleration in time to fixation which meant that such acceleration if present in the case of multiple loci, could be demonstrated experimentally without difficulty. The relative fitness 1 + s and 1 + hs were determined as follows.

$$1 + s = \frac{P(\frac{1}{-})}{P(\frac{1}{1})} = \frac{\int_u^{\infty} e^{-1/2(u_1-u)^2/\sigma_p^2}}{\int_u^{\infty} e^{-1/2(u_3-u)^2/\sigma_p^2}} \qquad (3.3)$$

$$1 + hs = \frac{P(\frac{1}{-})}{P(\frac{0}{1})} = \frac{\int_u^{\infty} e^{-1/2(u_2-u)^2/\sigma_p^2}}{\int_u^{\infty} e^{-1/2(u_3-u)^2/\sigma_p^2}} \qquad (3.4)$$

where

u_1 = genotypic value of $\frac{1}{1}$

u_2 = genotypic value of $\frac{1}{0}$

u_3 = genotypic value of $\frac{0}{0}$

u = genotypic value of the population.

The integration is from u to ∞ since half of the population was selected.

The effective population number (N_e) was determined from the formula by Kimura and Crow (1963).

$$N_e = \frac{(N_{t-1} - 1)\bar{k}}{1 + V_k/\bar{k}}$$

where

$\bar{k}$ = the mean number of progeny per parent

V_k = the variance in progeny numbers.

Population of different sizes was generated and pair mating was at random with one sire mated to one dam. Selection

was directional ($s \neq 0$) or random ($s = 0$) and the size of the population was kept constant. To illustrate a 4-1-4 population (Table IV), start from four pair matings with four offsprings (2 male, 2 female) per mating. For the population size to remain constant, four males and four females were selected as parents of the next generation. A population under consideration was regarded as

1. being initially a random sample from a hypothetically large population in linkage equilibrium,
2. being a cross between two homozygous lines in coupling or repulsion phase.

Selection was practiced in each population until fixation. A population under selection was replicated and the time to fixation was the average time over replications. Runs were under independent assortment and linkage. For linkage all loci were equally spaced on one chromosome that varied in length according to whether linkage was tight or lose. For two, three, and five loci, one crossover was allowed. For fifteen loci, the loci were divided into three groups of five loci each. One crossover was allowed per group and only second crossover between groups. If a first crossover occurred in one group, the probability of a second crossover occurring in another group was determined as the product of the first crossover probabilities in the two groups. Crossovers were directional, proceeding from one end to the other on a chromosome.

RESULTS AND DISCUSSION

The results are presented in Tables IV, V, VI, VII, VIII, IX, and X. In the case of two loci in linkage equilibrium (Tables IV and V) and under independent assortment, the expected time to fixation (E(T)) for a 4-1-4 population (N_e = 10) and s = .2 and .8 was not significantly different from E(T) for no selection (s = 0). Thus an N_es of 8 did not cause any acceleration in time to fixation. Under tight linkage (crossover probability of .05) there was no change from results of independent assortment. However, for looser linkage (crossover probability of .25) an increase in selection intensity from Ns = 2 to Ns = 8 reduced the expected time to fixation from 34.32 to 19.82, but was still not significantly less than E(T) for independent assortment and no selection. Thus the expected time to fixation in the case of two loci and up to an Ns of 8 was not significantly less than that for no selection as was true with one locus. Linkage did not seem to change results from those of independent assortment.

For the case of 8-1-4 population (Table V) and independent assortment, an Ns of 14 caused an acceleration in time to fixation ($41.03 < 56.96$), also Ns = 16. Note that the time to fixation increased with N as was found from theoretical work by Carr and Nassar (1969) for the two-locus case. This increase is linear at lower values of N. Acceleration in time to fixation was still in effect with tight linkage and Ns = 16. This, however, disappeared with moderate linkage (.25). Thus it seems clear that as in the one-locus case the same phenomenon

of acceleration in time to fixation when overdominance is such that the equilibrium gene frequency is 0.9 can occur with two loci if selection intensity is strong enough (Ns = 14 or above). The same occurs with very tight linkage and strong selection. Strong selection with looser linkage, however, did not seem to cause acceleration. At an initial gene frequency of 0.6 strong selection (Ns = 16) caused acceleration in time to fixation with and without linkage. Acceleration in time to fixation for Ns = 16 occurred for three and five loci under independent assortment and linkage (Table VI). It is also obvious that for the same population size and selection intensity the time to fixation increased with an increase in number of loci. In the case of 15 loci (Table VII) acceleration in time to fixation occurred for an Ns of 8 (N = 10, s = .8) in variant with the case of 2, 3, and 5 loci. Another important thing to notice is the fact that linkage with relatively strong selection did retard time to fixation. It is known that linkage in small populations gives rise to linkage disequilibrium which can inflate the degree of overdominance as given in the study. From theoretical work with the one-locus case, Carr and Nassar (1969), it is known that stronger overdominance h = 1.125 can decrease the amount of acceleration and eventually cause retardation in time to fixation. Thus linkage disequilibrium can be the cause of retardation in these results. The magnitude of linkage disequilibrium is no doubt larger in the 15-locus case than in 2, 3, or 5-locus cases. This explains why retardation occurred for linkage and 15 loci only. It seems clear that acceleration in time to fixation can occur for multiple loci with and without

linkage when selection is strong and overdominance is relatively weak. However, if the number of loci is large, then linkage can cause retardation in time to fixation. It seems that linkage disequilibrium is the determining factor in preventing acceleration in time to fixation in multiple locus systems with a large number of loci. Under the conditions that were investigated it is certain that linkage disequilibrium was at a minimum due to the fact that populations were started as a random sample from a hypothetically large population in linkage equilibrium. As a result the effects of linkage disequilibrium was only manifested under linkage and large numbers of loci. It was of interest to determine whether retardation would take place in the case of smaller number of loci and without linkage if the population at the outset was in linkage disequilibrium. For that purpose two extreme initial situations of linkage disequilibrium, coupling and repulsion, were investigated. Results under these two conditions serve as an upper and lower limit to what might happen in reality.

Results of Table VIII clearly show that the expected time to fixation for the two-locus case was the same regardless of the linkage equilibrium phase in the population at the outset and regardless of linkage or no linkage. An exception was the case of coupling and tight linkage which reduced expected time to fixation. Runs with no selection were not attempted since for $Ns = 16$ it is known from our previous results that acceleration in time to fixation occurs.

With five loci (Table IX) acceleration in time to fixation again occurred for all linkage disequilibrium phases and

independent assortment. The time to fixation was less for the repulsion phase than for coupling and linkage equilibrium under independent assortment. However, with linkage the expected time to fixation was prolonged for the repulsion case and for tight linkage, it exceeded the time with no selection. This relative increase in time to fixation was greatest at initial gene frequency of 0.5 and had no effect when time was measured when the population reached a gene frequency of 0.8. This can be explained by the fact that with random mating and recombination the population is expected to approach linkage equilibrium at which time the effect on retardation is nonexistent.

For two loci in coupling phase with tight linkage expected time to fixation was accelerated, while with five loci no significant difference was shown. The reason for this seems to be that with tight linkage, gametes go to fixation in groups of two more readily than in groups of five. For 15 loci (Table X), as in the five-locus case, there seems to be no difference in time to fixation between coupling, repulsion, or linkage equilibrium. Linkage with coupling again did not produce any significant new results. The conclusion is that initial coupling or repulsion under independent assortment did not cause any deviations from results obtained for populations with linkage equilibrium. It seems that the conditions of random mating and no epistasis is most favorable for the population to rapidly reach linkage equilibrium. With tight linkage, however, repulsion phase can cause retardation in time to fixation, and this can occur for a smaller number of loci than in the case of initial linkage equilibrium.

TABLE IV

Expected time to fixation for a 4-1-4 (N_e = 10) population with two loci in the case of overdominance.

h	P_0	Linkage	Reps.	Selection s	$N_e s$	E(T)*
1.125	.8	I.A.	56	0.0	0	26.64 ± 2.99
			40	.2	2	27.85 ± 3.96
			40	.8	8	23.32 ± 2.93
1.125	.8	.05	40	0.0		25.88 ± 2.92
			70	.2	2	31.13 ± 2.94
			40	.8	8	20.72 ± 2.42
1.125	.8	.25	40	.2	2	34.32 ± 4.90
			40	.8	8	19.82 ± 3.29

*E(T) = expected time to fixation.

TABLE V

Expected time to fixation for an 8-1-4 (N_e = 20) population with two loci in the case of overdominance.

h	P_0	Linkage	Reps.	Selection s	$N_e s$	E(T)
1.125	.8	I.A.	56	0.0	0	56.96 ± 5.44
			47	.4	8	50.36 ± 4.71
			40	.7	14	41.03 ± 3.29
			40	.8	16	33.05 ± 3.89
		.05	50	0.0	0	53.27 ± 5.95
			48	.8	16	38.42 ± 4.20
		.25	40	.8	16	48.58 ± 2.84
1.125	.6	I.A.	16	0.0	0	86.31 ±12.57
			27	.8	16	42.96 ± 4.35
		.25	25	.2	4	69.96 ±10.45
			49	.8	16	51.14 ± 4.43

TABLE VI

Expected time to fixation for an 8-1-4 (N_e = 20) population with three and five loci in the case of overdominance. Initial gene frequency = .8.

No. Loci	h	Linkage	Reps.	Selection s	$N_e s$	E(T)
3	1.125	I.A.	16	0.0	0	79.56 ±13.86
			20	.8	16	45.90 ± 5.74
5	1.125	I.A.	24	0.0	0	90.54 ±13.60
			13	.8	16	59.23 ± 6.80
		.05	20	.8	16	53.75 ± 6.82
		.30	19	.2	4	83.84 ± 9.42
			24	.8	16	52.13 ± 2.69

TABLE VII

Expected time to fixation for a 4-1-4 and an 8-1-4 population for fifteen loci in the case of overdominance. Initial gene frequency = .8.

Population	h	Linkage	Reps.	Selection s	$N_e s$	E(T)
8-1-4	1.125	I.A.	10	0.0	0	64.70 ± 9.60
			10	.2	4	60.70 ± 5.87
4-1-4	1.125	I.A.	17	0.0	0	42.70 ± 2.85
			10	.2	2	46.60 ± 4.50
			10	.4	4	48.70 ± 3.89
			13	.8	8	31.08 ± 2.33
	1.125	.15	24	.4	4	68.08 ± 9.04
			20	.8	8	56.50 ± 6.54

TABLE VIII

Expected time to fixation of an 8-1-4 population with two loci and selection = .8 in the case of over-dominance. (h = 1.125) $N_e s = 16$.

Phase	Linkage	Reps.	$\hat{P}_0$	E(T)
L.E.	I.A.	38	.5	50.92 ± 4.64
			.6	48.18 ± 4.68
			.8	39.02 ± 4.33
Rep.	I.A.	35	.5	45.03 ± 4.46
			.6	42.48 ± 4.43
			.8	34.80 ± 4.50
	.05	35	.5	50.71 ± 3.83
			.6	42.57 ± 3.69
			.8	32.91 ± 3.26
	.25	20	.5	48.80 ± 5.21
			.6	44.70 ± 5.31
			.8	37.45 ± 5.11
Coup.	I.A.	29	.5	49.31 ± 4.60
			.6	45.52 ± 4.36
			.8	34.76 ± 4.07
	.05	20	.5	32.45 ± 3.96
			.6	30.85 ± 3.91
			.8	26.50 ± 3.55
	.25	20	.5	49.45 ± 5.98
			.6	46.85 ± 6.05
			.8	37.45 ± 5.51

Where Rep. means repulsion
Coup. means coupling
I.A. means independent assortment
Reps. is the number of repeats
$\hat{P}_0$ is initial gene frequency.

TABLE IX

Expected time to fixation of an 8-1-4 population with five loci and selection = .8 in the case of overdominance. (h = 1.125) $N_e s = 16$.

Phase	Linkage	Reps.	$\hat{P}_0$	E(T)
L.E.	I.A.	18	.5	66.61 ± 6.01
			.6	63.56 ± 6.10
			.8	53.94 ± 6.12
Rep.	I.A.	10	.5	48.70 ± 3.81
			.6	45.90 ± 3.78
			.8	39.30 ± 3.93
	.05	11	.5	137.72 ± 9.12
			.6	100.63 ± 6.79
			.8	54.00 ± 8.32
	.25	7	.5	85.57 ± 6.78
			.6	76.28 ± 6.70
			.8	60.14 ± 8.87
Coup.	I.A.	7	.5	67.71 ± 7.29
			.6	63.86 ± 7.23
			.8	55.29 ± 6.28
	.05	20	.5	59.75 ± 9.27
			.6	58.50 ± 9.24
			.8	53.55 ± 9.16
	.25	20	.5	54.70 ± 5.99
			.6	53.55 ± 5.99
			.8	49.00 ± 5.92

For an 8-1-4 population with selection = 0 (random mating) and initial gene frequency = 0.5, we have $E(T) = 100.85 \pm 7.52$.

TABLE X

Expected time to fixation of an 8-1-4 population with 15 loci and selection = .8 in the case of overdominance. (h = 1.125) $N_e s = 16$.

Phase	Linkage	Reps.	$\hat{P}_0$	E(T)
L.E.	I.A.	4	.5	84.00 ± 6.78
			.6	80.25 ± 7.52
			.8	68.25 ± 7.26
Rep.	I.A.	4	.5	81.25 ± 3.17
			.6	77.25 ± 1.66
			.8	66.75 ± 2.87
Coup.	I.A.	4	.5	92.00 ± 2.27
			.6	88.50 ± 2.33
			.8	80.75 ± 3.09
	.15	10	.5	85.90 ±14.28
			.6	84.60 ±14.29
			.8	78.80 ±14.58

SUMMARY

Computer simulation was used to study the expected time to fixation in a population of finite size N, for the case of overdominance with equilibrium gene frequency $\hat{q} = 0.9$. For populations initially in linkage equilibrium it was found that:

1. For a 4-1-4 ($N_e = 10$) population the effects of either linkage or selection did not significantly change the expected time to fixation from that of no selection unless there were more than five loci.
2. For an 8-1-4 ($N_e = 20$) population selection alone significantly decreased the expected time to fixation from that of no selection.
3. For both tight and loose linkage to retard the expected time to fixation we must have more than five loci. For gametes in either coupling or repulsion phase the expected time to fixation was found to depend on the number of loci and the phase of linkage. In the case of two loci tightly linked in coupling phase, the expected time to fixation is accelerated. For five loci in repulsion phase the expected time to fixation is highly sensitive to the type of linkage and is increased with tight linkage.

BIBLIOGRAPHY

Carr, R. N. and Nassar, R. F., 1969. Effects of selection and drift on the expected time to fixation or loss of a favorable allele. Submitted for publication.

Ewens, W. J., 1963. Numerical results and diffusion approximations in a genetic process. Biometrika 50:241-249.

Feller, W., 1957. An introduction to probability theory and its applications (2nd ed.). John Wiley, New York.

Fisher, R. A., 1922. On the dominance ratio. Proc. Roy. Soc. Edin. 42, 321-341.

Fisher, R. A., 1930. The distribution of gene ratios for rare mutations. Proc. Roy. Soc. Edin. 50, 205-220.

Haldane, J. B. S., 1932. The causes of evolution. Harper and Brothers, New York.

Kimura, M., 1955a. Random genetic drift in multi-allelic locus. Evolution 9, 419-435.

Kimura, M., 1955b. Stochastic processes and distribution of gene frequencies under natural selection. Cold Spring Harbor Symp. 20, 33-53.

Kimura, M., 1957. Some problems of stochastic process in genetics. Ann. Math. Statist. 20, 882-901.

Kimura, M., 1962. On the probability of fixation of mutant genes in a population. Genetics 47, 713-719.

Kimura, M., 1964. Diffusion models in population genetics. Methuen and Co., Ltd., London.

Kimura, M. and Crow, J. F., 1963. The measurement of effective population numbers. Evolution 17, 279-288.

Robertson, A., 1962. Selection for heterozygotes in small populations. Genetics 47, 1291-1300.

Wright, S., 1931. Evolution in Mendelian populations. Genetics 16, 97-159.

Wright, S., 1945. The differential equation of the distribution of gene frequencies. Proc. Nat. Acad. Sci. 31, 382-389.

EFFECTS OF SMALL POPULATION SIZE AND SELECTION ON THE EXPECTED TIME TO FIXATION OF A FAVORABLE ALLELE WITH MULTIPLE LOCI

by

LARRY SCOTT

B. S., Kansas State University, 1967

AN ABSTRACT OF A MASTER'S REPORT

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Statistics
and Computer Science

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1970

A genetic population of finite size N was simulated on the IBM 360/50 computer using a finite Markov chain. The effects of linkage and selection on the expected time to fixation were studied for multiple loci. It was found that:

1. For a 4-1-4 (N_e = 10) population the effects of either linkage or selection did not significantly change the expected time to fixation from that of no selection unless there were more than five loci.
2. For an 8-1-4 (N_e = 20) population selection alone significantly decreased the expected time to fixation from that of no selection.
3. For both tight and loose linkage to retard the expected time to fixation we must have more than five loci. For gametes in either coupling or repulsion phase the expected time to fixation was found to depend on the number of loci and the phase of linkage. In the case of two loci tightly linked in coupling phase, the expected time to fixation is accelerated. For five loci in repulsion phase the expected time to fixation is highly sensitive to the type of linkage and is increased with tight linkage.