

SELECTION IN FINITE POPULATIONS. II. THE SELECTION LIMIT
AND RATE OF RESPONSE FOR A MONTE CARLO
SIMULATION MODEL¹

PHILIP W. HEDRICK²

Department of Biology, University of Chicago

Received September 10, 1969

KIMURA (1957) developed a general theory to give the probability of fixation of a favorable allele in a finite population based on a single locus, two allele model. By assuming no linkage or epistatic interactions between loci, ROBERTSON (1960) extended the single-locus theory to predict the limits to directional selection. At the present time there is no general solution for multilocus systems in which there is linkage or epistasis. In order to investigate the effects of linkage and epistasis on the selection limit, a number of workers have employed computer simulation (GILL 1965; YOUNG 1967; QURESHI and KEMPTHORNE 1968).

In a series of papers, LATTER examined a two locus system where the genes had an additive effect on the selected trait. The two loci were in initial linkage equilibrium and several intensities of linkage were studied. LATTER (1965b) found that with no linkage ($c=.5$) between the two loci that the selection limit was generally close to the probability of fixation as given by $u(q)$ (KIMURA 1957).

$$u(q) = \frac{\int_0^{q_0} e^{-2Nsx(2h-1)(1-x)-2Nsx} dx}{\int_0^1 e^{-2Nsx(2h-1)(1-x)-2Nsx} dx} \quad (1)$$

where q_0 is the initial gene frequency of the favorable allele A , N is the effective population size, and the relative fitnesses of the genotypes AA , Aa , and aa are $1+s$, $1+hs$, and 1 , respectively. But when the population size was small and s was large, LATTER found that his simulated selection limit was lower than the prediction given by $u(q)$.

With tight linkage ($c \leq .05$) between two loci of the same initial gene frequency and of equal gene effect, LATTER observed that linkage reduced the selection limit most when

$$\frac{u(q_0) - q_0}{1 - q_0} = .6 \text{ to } .7 \quad (2)$$

HILL and ROBERTSON (1966) also examined a two locus system where the genes had additive action. The loci were assumed to be in initial linkage equilibrium. They observed the greatest reduction in gene frequency response with tight linkage when the loci had a low initial gene frequency and large, equal effects.

¹ Adapted from a thesis presented in partial fulfillment of the Ph.D. degree at the University of Minnesota. The investigation was supported by P.H.S. Grants GM-3090-12 and GM-01156-06.

² Present address: Department of Systematics and Ecology, University of Kansas, Lawrence, Kansas.

Both LATTER and HILL and ROBERTSON observed that linkage reduced the time to half of the gene frequency response (half-life). With tight linkage and low initial gene frequency, LATTER found that the time to 95% of the gene frequency response was dramatically increased. HILL and ROBERTSON also stated that response continued for two intermediate recombination levels ($c = .125$ and $.03125$) after response had ceased for the complete linkage and no-linkage cases.

The first objective of this investigation was to examine the accuracy with which $u(q)$ predicts the selection limit for genes of additive, dominant and recessive gene action in a genetic system of thirty chromosomes, each with two loci. Selection limit results obtained using a transition matrix approach and the estimated effective population size were used to interpret these findings. The second aspect of the investigation compared the simulated rate of gene frequency response in terms of half-life time to the expected rate computed from matrix iteration. The effect of complete linkage between the pairs of loci on the selection limit and the rate of gene frequency response was also studied.

MODEL AND METHODS

The genetic system simulated in this study is that of directional selection for a quantitative trait. Sixty loci contributed to the quantitative character such that the phenotypic value for any individual was

$$P = \sum_{i=1}^{60} \gamma_i + e$$

with the values of γ_i being $2u$, $u+au$, or zero depending on whether the genotype at the i -th locus was AA , Aa , or aa . The environmental effect, e , was random relative to the genotypic value and was normally distributed. The genotypic variance changed as the gene frequency changed by selection but the phenotypic variance was kept constant throughout an experiment by making the required adjustments in the environmental variance. There was no interaction between the different loci, i.e., no epistasis.

The loci were arranged into thirty chromosomes, each with two loci and there were two alleles per locus. All the loci had the same genotypic values and expected initial gene frequency in a given experiment. The linkage between the two loci on a chromosome was either complete ($c=0$) or absent ($c=.5$). The experiments included all combinations of six initial gene frequencies, three degrees of dominance, three levels of gene effects and the two linkage intensities. Each experiment was replicated ten times when all sixty loci contributed to the genotypic value and fifteen times in a few cases where only forty loci were used.

In all the experiments sixty individuals were randomly drawn from a conceptually infinite base population to form the zero generation. This generation was assumed to be in linkage equilibrium. Then the twenty individuals with the highest phenotypic values were selected to be the parents of the first generation. The parents were randomly ordered and adjacent individuals in the array were pair mated. Sixty progeny were generated so that every parental pair had an equal probability of being the parents for any individual. The twenty progeny with the highest phenotypic value were selected as parents for the next generation, and so on.

Since it was impractical to continue selection until all sixty loci were fixed, the experiments were terminated ten generations after the expected time needed for three-quarters of the gene frequency response. At this point in all the cases with no linkage and the runs with complete linkage and recessive gene action, the loci still segregating were assigned to one of the two fixed-gene classes. The proportions of the loci added to the favorable and unfavorable classes were calculated from $u(q)$ using the gene frequency in the terminal generation. After the segregating loci were assigned to terminal classes, the proportion of loci in the favorable fixed class was defined as the selection limit, q_f .

A different approach was used to estimate the selection limit in the runs with complete linkage when the genes were either dominant or additive. The logic of this approach was based on two observations. First of all, the number of loci segregating at the termination of the experiments was less when there was complete linkage than when there was no linkage. Secondly, the gene frequencies of the segregating loci for the complete linkage runs were consistently less than for the no-linkage runs. Based on these observations and realizing that $u(q)$ increases as q increases, the expected difference in the selection limit between the complete and no-linkage cases would be greater than the difference in gene frequency observed in the final generation of the experiment. Therefore, for dominance and additivity, the difference in the gene frequency for the two linkage intensities in the last generation of the experiment was used as the measure of the linkage effect. For recessive genes, the effect of linkage was measured by the difference in the selection limits for the complete and no-linkage runs.

Using $q_f - q_0$ as the gene frequency response for all the no-linkage cases and complete linkage with recessive gene action runs, the times to one-half and three-quarters of the average gene frequency response were found. For linkage with dominance or additivity, in order to compute the half and three-quarters lives, the selection limit was defined as the selection limit with no linkage minus the difference in gene frequency at the end of the experiment between the two intensities of linkage. Since this is a maximum estimate of the selection limit for the no-linkage runs the gene frequency at half or three-quarters life and the time to half or three quarters life are also maximum values.

The simulation parameters were set up so that the simulated results could be compared to theoretical values of the selection limit and the rate of response. To do this the expected selective value was calculated using the expression, $s = 2iu/\sigma_p$ (HALDANE 1931) where i is the intensity of selection (standardized selection differential), u is half the difference in genotypic value between the two homozygotes, and σ_p is the phenotypic standard deviation. The expected population size was computed as the number of parents plus one-half (WRIGHT 1931). WRIGHT'S approximation excludes the possibility of self-fertilization and assumes there is no selection. Equivalent levels of dominance for the simulated model and the theoretical model are $a = 0$ and $h = 1/2$ for additivity, $a = 1$ and $h = 1$ for dominance, and $a = -1$ and $h = 0$ for recessivity.

RESULTS

(i) *Selection limit:* What is the validity of using $u(q)$ to predict the average selection limit of unlinked loci in a multilocus system? As an attempt to answer this question a comparison between the simulated selection limit and $u(q)$ was made in Table 1 for the percent difference in response

$$\frac{u(q) - q_f}{u(q) - q_0} \times 100$$

and the absolute difference $u(q) - q_f$.

Generally $u(q)$ was within one standard error of the simulated selection limit with $Ns=2.05$ (the SE of q_f between replications ranged from .012 to .026 and the binomial SE from .007 to .020). For dominance there were several instances when the probability of fixation was more than one standard error greater than the selection limit. With the larger Ns values, the agreement of q_f and $u(q)$ was not as good. In general for these runs the percent difference (reduction) in response was greatest for low gene frequencies and for recessive gene action. In nine cases this reduction in response exceeded twenty percent.

In an effort to find the reason for this reduction in response, the selection limit obtained from transition matrix iteration and an estimate of the effective population size were used. The transition matrix selection limits were taken

TABLE 1

The percent and actual (in parentheses) difference between the probability of fixation, $u(q)$, and the selection limit, q_f , for $c = .5$

q_0	Additive	Dominant	Recessive
		$Ns = 2.05$	
.0625	-3.5 (.006)	.4 (.001)	11.2 (.009)
.125	.0 (.000)	8.4 (.035)	-3.2 (.005)
.25	1.5 (.006)	8.4 (.043)	3.6 (.010)
.375	-5.4 (.023)	4.9 (.024)	8.1 (.028)
.5	-2.1 (.008)	.5 (.002)	9.4 (.033)
.75	-4.1 (.009)	.0 (.000)	-3.2 (.007)
		$Ns = 4.1$	
.0625	19.8 (.067)	17.2 (.092)	13.7 (.019)
.125	12.2 (.063)	15.2 (.107)	24.8 (.065) ⁵
.25	8.4 (.052)	7.0 (.050) ³	22.1 (.097)
.375	3.6 (.021)	3.1 (.019)	16.9 (.084)
.5	4.6 (.022)	.4 (.002)	12.7 (.058)
.75	.8 (.002)	.4 (.001)	.0 (.000)
		$Ns = 8.2$	
.0625	27.8 (.161) ¹	21.6 (.171) ⁴	20.7 (.045) ⁶
.125	23.7 (.177)	8.6 (.073)	41.4 (.166)
.25	7.5 (.055) ²	2.1 (.016) [†]	44.6 (.267) ⁷
.375	5.8 (.036)	.0 (.000) [†]	21.2 (.126) [†]
.5	1.0 (.005)	.0 (.000) [†]	17.3 (.086) [†]
.75	.0 (.000)	.0 (.000)	.0 (.000) [†]

† 40 loci affecting character

Effective population size calculated for the seven experiments indicated by superscripts 1...7 (Table 2)

from a model in which the expected gene frequency after one generation was

$$q' = q + sq(1 - q)[h + (1 - 2h)q] / [1 + sq^2 + 2hsq(1 - q)] \quad (3)$$

as in method B of HEDRICK (1970). The initial vector of gene frequency distribution, Y_0 , though was given by

$$P_i = \binom{2N}{i} (q_0)^i (1 - q_0)^{2N-i} \quad (4)$$

where P_i is the probability of i favorable genes when the initial gene frequency is q_0 . The limit given by this approach should be closer to the simulated selection limit than $u(q)$ since it uses a better approximation of gene frequency change and takes into account sampling variation in the initial number of favorable alleles at a locus.

The probability of fixation is directly dependent on population size, so that a reduction in the population size would decrease $u(q)$. Assuming that the effective population size was twenty and one-half would cause an overestimate in $u(q)$ if the population size was actually less. To find out what the effective population size actually was for several runs, it was calculated from the average inbreeding coefficient. The relationship used in this computation was

$$1 - F_x = (1 - (1/2N))^x (1 - F_0) \quad (5)$$

TABLE 2

The effective population size, N , and the expected reduction from $u(q)$ where the population size was assumed to be 20.5 due to N and from matrix iteration. The ratio of observed response to the expected response is given in the right-hand column

Model (see Table 1)	N	Reduction from $u(q)$			Ratio
		N	Matrix	Expected	
1	17.7	.057	.093	.150	.97
2	16.9	.017	.025	.042	.99
3	18.1	.022	.039	.061	1.01
4	18.3	.037	.144	.181	1.02
5	17.5	.030	.013	.043	.88
6	14.0	.049	.012	.061	1.10
7	8.2	.214	.036	.250	.89

where F_x is the inbreeding coefficient in the x -th generation, F_0 is the inbreeding coefficient in the zero generation (assumed to be zero), and N is the effective population size (FALCONER 1960). The inbreeding coefficient for each generation was computed from the pedigree of each individual in that generation. The expression $\ln(1 - F_x) = bx$ was fitted from t , the second generation, to T which was approximately the number of generations to half the gene frequency response. Then by setting $b = \ln(1 - (1/2N))$ and solving for N , an estimate of the effective population size was found.

The effective population size was calculated for the seven experiments indicated by superscripts in Table 1. The effective size and the reduction in the probability of fixation expected because of the reduction in the population size are given in Table 2. The largest reduction in population size was seen for recessive genes with large effect (Experiments 6 and 7). For the other experiments, the reduction in population size was no greater than four. On the other hand, the differences accounted for by the matrix iteration were generally greatest in the dominant and additive cases (Experiments 1-4). An expected amount of reduction from $u(q)$ was computed by summing the effects of the reduced population size and those accounted for by the matrix iteration. The ratio of the observed response over the expected response,

$$\frac{O(q_t) - q_0}{E(q_t) - q_0},$$

is given in the last column of Table 2. For all the models the ratio is close to unity and particularly for the additive and dominant cases, models 1-4.

All fifty-four combinations of dominance, initial gene frequency and gene effect were also run with complete linkage between the two loci on each chromosome. For all these runs the same initial arrays of gametes as in the corresponding no-linkage case were used. As a measure of the effect of linkage on the selection limit for additive or dominant genes, the absolute difference in gene frequency response in the last generation for the two intensities of linkage is given in Table 3. For recessive genes the difference in the selection limit and the percent

TABLE 3

The percent and actual (in parentheses) difference between the selection limits with $c = .5$ and the selection limit with $c = 0$

q_0	Additive	Dominant	Recessive
		$N_s = 2.05$	
.0625	29.9 (.029)	18.9 (.039)	3.4 (.006)
.125	19.9 (.045)	11.8 (.037)	5.9 (.016)
.25	19.4 (.058)	11.7 (.044)	17.2 (.046)
.375	9.4 (.036)	6.8 (.025)	-1.8 (.006)
.5	10.7 (.033)	14.5 (.050)	-5.3 (.018)
.75	7.4 (.014)	1.6 (.003)	1.3 (.003)
		$N_s = 4.1$	
.0625	7.2 (.018)	19.2 (.068)	.8 (.001)
.125	7.0 (.027)	8.9 (.040)	-3.5 (.007)
.25	18.0 (.089)	21.1 (.115)	18.1 (.062)
.375	6.6 (.027)	8.5 (.043)	7.5 (.031)
.5	.9 (.003)	9.9 (.041)	9.2 (.046)
.75	3.1 (.007)	2.8 (.006)	.0 (.000)
		$N_s = 8.2$	
.0625	13.7 (.055)	22.3 (.125)	13.3 (.023)
.125	17.3 (.097)	22.3 (.168)	-3.6 (.008)
.25	18.8 (.123)	18.6 (.119)	7.8 (.026)
.375	6.1 (.035)	7.5 (.040)	11.6 (.054)
.5	5.5 (.027)	2.3 (.010)	-3.2 (.013)
.75	.0 (.000)	0.0 (.000)	.0 (.000)

difference in total gene frequency response for the two linkage intensities are given.

With additivity or dominance, the percent reduction was generally greatest at low initial gene frequencies. In several instances the reduction was twenty percent or greater. For recessive genes the reduction due to linkage was generally less than for additive or dominant genes. The largest reduction for recessive genes was generally also at a low initial gene frequency.

(ii) *Rate of gene frequency response:* In Figures 1, 2, and 3 the times for half the gene frequency response observed in the simulation runs are compared to the half-life times computed from matrix iteration (as above). The solid lines represent the time to half of the gene frequency response, $1/2(q_f - q_0)$, for matrix iteration corresponding to a population size of twenty. The circles indicate the time to half of the simulated gene frequency response when there is no linkage. With linkage and dominance or additivity, the maximum (as defined in METHODS) half-life is represented as triangles. For recessive genes and linkage, the triangles indicate the time to half the observed gene frequency response.

The time to half-life with independent assortment in the simulated system was longer in every case than that calculated using matrix iteration. The average increase in time to half-life for all three levels of dominance was approximately three generations. With complete linkage and dominance or additivity, the maxi-

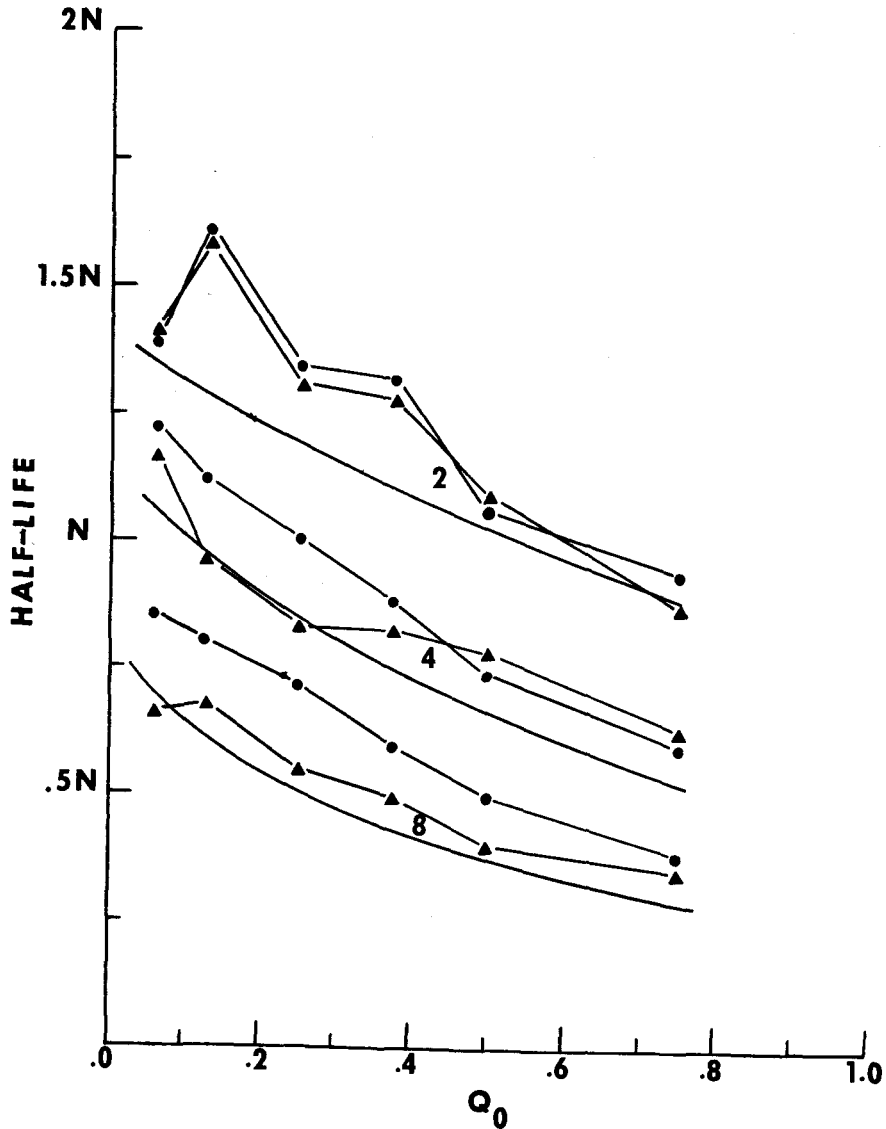


FIGURE 1.—The time in generations to the half-life of gene frequency response for additive gene action for N_s s equal to 2, 4, and 8; solid lines, transition matrix; circles, simulation with $c = .5$; triangles, simulation with $c = 0$.

imum half-life was almost always less than for independent assortment. This difference increased as N_s s became larger, particularly at low initial gene frequencies. For recessive genes and linkage, the decrease in half-life time from independent assortment was also largest at low initial gene frequencies and high N_s s values. The times for three-quarters of the response for both linkage and no linkage showed much the same trends.

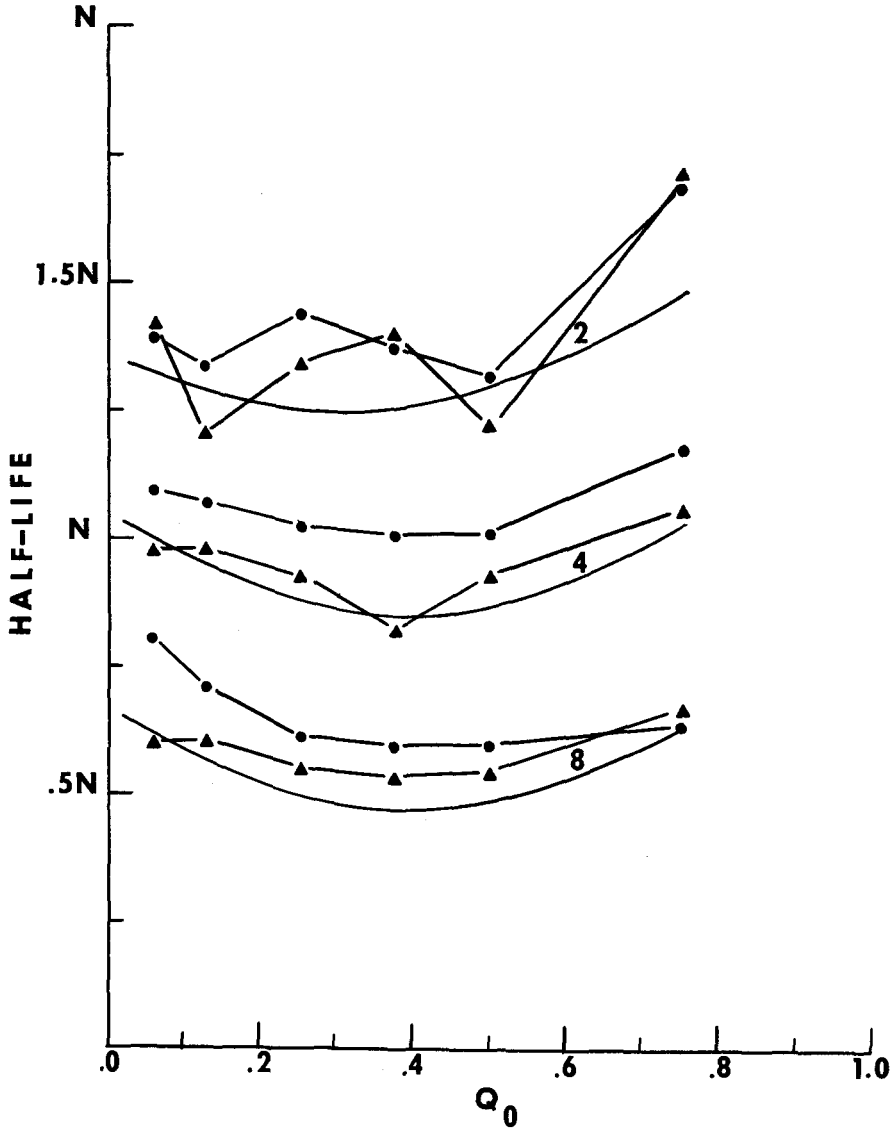


FIGURE 2.—As Figure 1 but for dominant gene action.

DISCUSSION

There was apparent agreement between the selection limit for no linkage and the probability of fixation when Ns was small. The values at the actual limit may not actually be as close to $u(q)$ as those indicated since q_f was partially determined using $u(q)$. Since the simulation model differed from the model used for $u(q)$ in several ways that lower the selection limit, using $u(q)$ to find q_f would increase the selection limit above its actual value. On the other hand, it should be noted that the formula

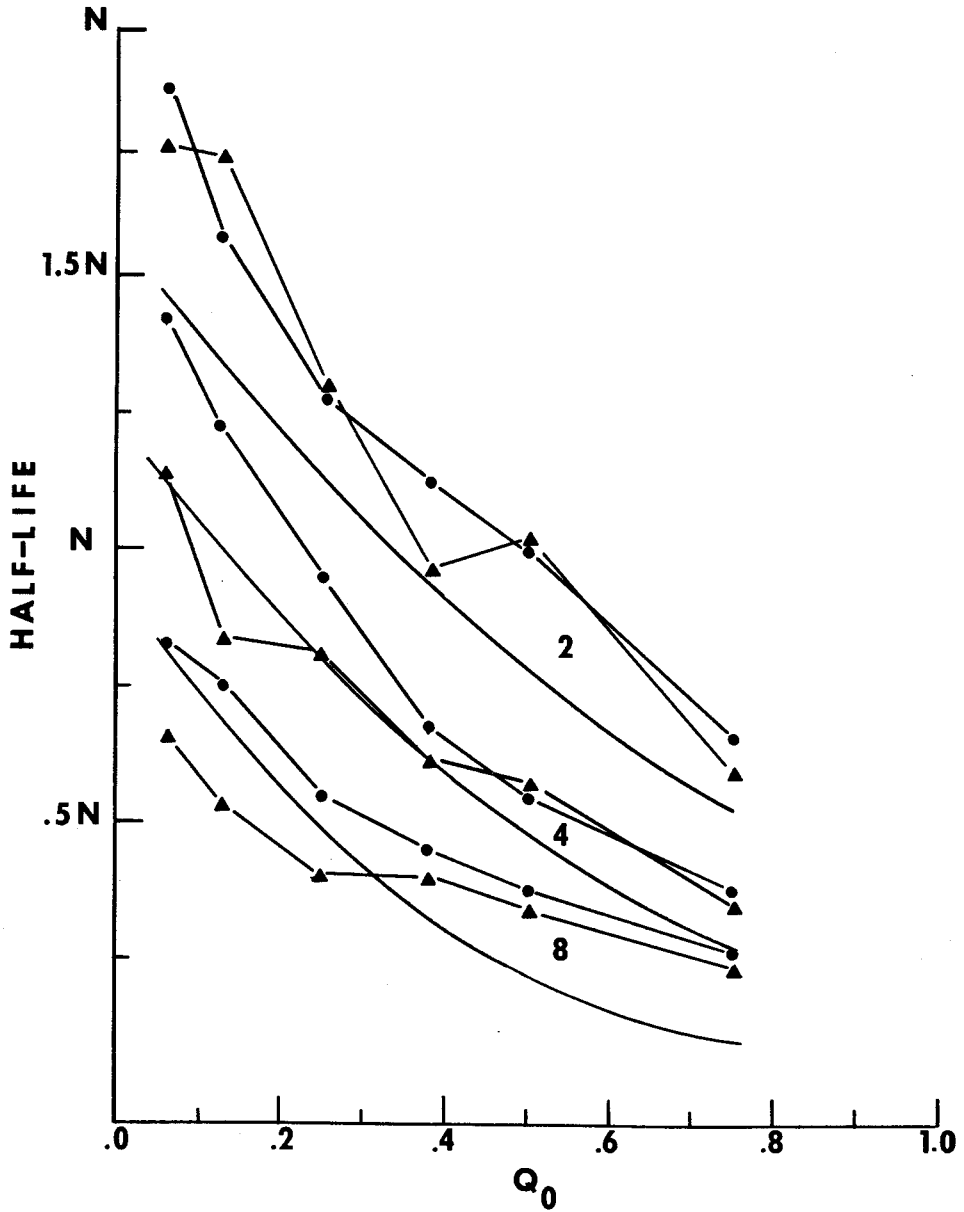


FIGURE 3.—As Figure 1 but for recessive gene action.

$$s = 2iu/\sigma_p$$

is not exact and that the error resulting from it increases as the gene effect,

$$2u/\sigma_p,$$

increases (LATTER 1965a). For the largest gene effect (.30) used in this study, though, the error is between two and three percent.

To see how the population size, and as a result, the selection limit, was reduced when N_s was large, L. P. JONES used the pedigrees of the runs from which the effective population size was calculated and computed the variance of proportion of genes (JAMES 1962). By looking back for five generations, it was possible to evaluate how parental pairs in different runs had contributed to future generations. For runs with recessive genes he found that the variance of the proportion of genes was quite high, and in some cases all individuals had descended from the same parental pair five generations previous. In contrast to this, the variance for dominant or additive runs was close to the expected value and most of the parental pairs had descendants in the present generation.

Complete linkage generally reduced the gene frequency in the final generation for additive and dominant genes. For recessive genes the selection limit was often quite similar for the two levels of linkage. HILL and ROBERTSON (1966) observed a reduction in response of almost thirty percent for a run with parameters similar to the additive run in which the initial gene frequency was .125 and N_s was equal to 8.2. The observed reduction in this experiment was 17.3% in the last generation. If the run was continued till complete fixation, the reduction would have been greater since a number of chromosomes in the complete linkage run were segregating only for repulsion gametes.

An effort was made to find a quantitative guide to describe when linkage had the greatest effect for dominant and recessive genes as LATTER has done for additive genes. In fact, it was not possible to critically evaluate the applicability of his estimate for additive genes in the sixty-locus system much less extend it to other levels of dominance. The only generalization really possible from these data is that linkage had the greatest effect in reducing response at gene frequencies of .25 or below.

When the comparison of half-life times for matrix iteration and simulation with no linkage was made, the half-life for the simulation was longer in every instance. Part of this elongation was due to the matrix method used for the comparison. A method based on genotypic specification which would more closely mimic the simulation as used by HEDRICK (1970) would be better if equivalent population size was possible. Comparison of the two matrix methods showed that a maximum of five percent of the increased half-life time could be accounted for by the matrix using genotypic specification.

The reduction in half-life time due to linkage was greatest for large N_s and low q_0 . This effect was caused by the combination of a large N_s , which resulted in a high selection limit, and a low q_0 , reducing the probability of the favorable coupling gamete (++) being present in the initial population. For a number of runs the array of gametes in the final generation was examined. From this it was found that for dominant genes more than ninety percent and for additive genes about forty percent of the segregating chromosomes were complementary repulsion types, i.e., only (+ -) and (- +) gametes were present. Since in the chromosomes where only repulsion types are segregating the maximum gene frequency for the chromosome has already been reached, it follows that the

reduction in half-life due to linkage was actually greater than given in Figure 2 for dominant genes and somewhat larger than given in Figure 1 for additive genes.

LATTER (1966) found that tight linkage between two additive loci of equal effects with a low q_0 had an extended period of response. This effect was attributed to a period of segregation of the two repulsion types, $(- +)$ and $(+ -)$, until an effective crossover produced a $(+ +)$ gamete and a period of late response. The high proportion of repulsion segregants observed in both the additive and dominant runs here indicates a potential for prolonged response if the recombination fraction was very small. HILL and ROBERTSON (1966) have suggested that LATTER's observations may be applicable only to the special case where the gene effects at the two loci are quite similar. When there is dominance and the segregation of only the two repulsion types, a situation of pseudo-overdominance exists. The two loci could be quite different in effect and their segregation still maintained by the force of selection. With a large population size, segregation for an indefinite period of time could occur while waiting for the effective crossover event.

I wish to thank L. P. JONES and R. E. COMSTOCK for their assistance and encouragement during the investigation and Dr. COMSTOCK for his thorough reading of the manuscript. The comments of W. G. HILL and B. D. H. LATTER on the manuscript were very helpful.

SUMMARY

The probability of fixation of a favorable allele as given by $u(q)$ (KIMURA 1957) was close to results simulated for a system of thirty chromosomes, each with two loci when there was a selective value of .1 for all loci and no linkage. With a selective value of .2 or .4, the simulated selection limit was lower than the $u(q)$ value. This difference was accounted for by computing an estimate of the effective population size and by the reduction expected from a transition matrix approach (HEDRICK 1970). The effect of complete linkage on the response was most pronounced at low initial gene frequencies. Linkage did not reduce the response as much for recessive genes as for dominant or additive genes.—The time to half of the gene frequency response was somewhat higher for the simulated results than for the matrix iteration. Part of this difference can be attributed to the matrix method used. The time to half-life was reduced greatly with complete linkage when there was a low initial gene frequency and a high Ns value. It is suggested that a low recombination level may lengthen the response time, particularly for dominant genes.

LITERATURE CITED

- FALCONER, D. S., 1960 *Introduction to Quantitative Genetics*. Ronald Press, New York.
GILL, J. L., 1965 Selection and linkage in simulated genetic populations. *Australian J. Biol. Sci.* **18**: 1171-1187.
HALDANE, J. B. S., 1931 A mathematical theory of natural and artificial selection. VII. Selection intensity as a function of mortality rate. *Proc. Cambridge Phil. Soc.* **27**: 131-136.

- HEDRICK, P. W., 1970 Selection in finite populations. I. The probability of fixation and rate of response using transition matrix iteration. *Genetics* **65**: 157-173.
- HILL, W. G. and A. ROBERTSON, 1966 The effect of linkage on limits to artificial selection. *Genet. Res.* **8**: 269-294.
- KIMURA, M., 1957 Some problems of stochastic processes in genetics. *Ann. Math. Statist.* **28**: 882-901.
- JAMES, J. W., 1962 The spread of genes in random mating control populations. *Genet. Res.* **3**: 1-10.
- LATTER, B. D. H., 1965a The response to artificial selection due to autosomal genes of large effect. I. Changes in gene frequency at an additive locus. *Australian J. Biol. Sci.* **18**: 585-598.
- , 1965b The response to artificial selection due to autosomal genes of large effect. II. The effects of linkage on limits to selection in finite populations. *Australian J. Biol. Sci.* **18**: 1009-1023.
- , 1966 The response to artificial selection due to autosomal genes of large effect. III. The effects of linkage on the rate of advance and approach to fixation in finite populations. *Australian J. Biol. Sci.* **19**: 131-146.
- QURESHI, A. W. and O. KEMPTHORNE, 1968 On the fixation of genes of large effects due to continued transaction selection in small populations of polygenic systems with linkage. *Theoret. Appl. Genetics* **33**: 249-255.
- ROBERTSON, A., 1960 A theory of limits in artificial selection. *Proc. Royal Soc. London B* **153**: 234-249.
- WRIGHT, S., 1931 Evolution in Mendelian populations. *Genetics* **16**: 97-159.
- YOUNG, S. S. Y., 1967 Computer simulation of directional selection in large populations. II. The additive \times additive and mixed models. *Genetics* **56**: 73-87.