An Evaluation of Multiple Peak Epistasis and Population Structure in Directional Selection Program

FRANK D. ENFIELD and FARHAD ANKLESARIA, USA

Dept. of Genetics and Cell Biology
University of Minnesota
St. Paul, Minnesota

SUMMARY

Computer simulation was used to compare the effectiveness of directional selection using a subdivide-merge type of breeding scheme with a single large undivided mass selected population in the presence of known multiple peak epistasis. The selected breeding structure very closely followed that investigated by Madelena and Hill (1972) in their evaluation of non-epistatic models. For the model used the subdivide-merge scheme was shown to be superior to the large mass selected population, both in the long and short term, within the range of allelic frequencies investigated in the model. The response to selection in both the subdivided and large populations was markedly influenced by initial linkage disequilibrium and close linkage. Negative genetic phase imbalance reduced (or reversed in some cases) the advantage gained by the subdivide-merge scheme. Positive disequilibrium slightly enhanced the subpopulation advantage. Simulation with mixed models (both dominant and multiple peak loci) showed an increasing advantage for the large, undivided population. In practical breeding terms, given the lack of direct experimemtal evidence for multiple peak epistasis as a major mode of gene action, and the possible negative effects of inbreeding in the small sublines, the single large, mass selected population still appears to be of greater overall utility.

INTRODUCTION

The primary objective of the work to be discussed was to evaluate by the use of computer simulation the conditions under which multiple peak epistasis is the driving force in determining limits in response to directional selection. Gene action of the multiple peak type and schemes for its effective utilization are central to Wright's shifting balance theory of evolution (Wright, 1977). Wright also has placed considerable emphasis on multiple peak epistasis as the primary kind of gene action in the inheritance of some quantitative traits of importance in plant and animal improvement programs . Wright (1963) states that with multiple peak epistasis the most effective selection methods would entail subdivision of a crossbreeding species into small populations sufficiently isolated to permit differentiation under the joint effects of random drift and intergroup selection. In practical breeding terms this requires a program which involves selection within small populations interrupted periodically by selection among and crossing of these populations to develop new set of lines with the process then repeated.

A number of studies have been done to determine the consequences of selection in a subdivided population versus a large undivided one. Many of these have been designed specifically to test Wright's theory. The approaches have been of three types: experimental (live organisms), analytical, and computer simulation studies.

In experimental studies done using organisms such as Tribolium and Drosophila the results have not been consistent . Rarely, however, have striking advantages been observed in terms of greater long term response to directional selection being obtained in subdivided populations as compared with large single populations. Bowman and Falconer (1960) examining litter size in mice as a function of inbreeding, used a scheme involving cyclic inbreeding and crossing. They found a slight gain in litter size at initial line crossing, attributing this to simple dominance at the loci and concluded that breeding scheme did not appear to be useful. Goodwill (1974), using *Tribolium castaneum* as the experimental organism, selected for pupa weight to look at the effects of cyclic between-within line selection schemes. He

found no significant differences between the large single mass selected population and the two schemes of within-between line selection with respect to rate of response. In a similar series of experiments with Tribolium, Katz and Enfield (1977) found no striking differences between different subdivide-merge selection schemes but did find them less effective than selection in a large single population. Rathie and Nicholas (1980),selecting for high abdominal bristle number in Drosophila, found no superiority of the subdivided population selection scheme over the large undivided population.

The simulation studies of Madalena and Hill (1972) set the stage for many of the experiments summarized above. In their work the only time that the between-within line selection procedures showed an advantage over a single large mass selected population was the case of loci where the favored allele was the recessive allele and started at low allelic frequencies in the foundation population. They considered only non-epistatic models in their analysis. More recently, Slatkin (1981) analytically confirmed Madalena and Hill's result: there is a better chance for selection with dominant gene action for the favorable allele in a single large population than in a subdivided one, the reverse being true in the case of an underlying recessive model. Since the latter seems to rarely be the case for traits of interest it leaves the primary question of interest to be centered on what conditions will make multiple peak epistatic gene interactions the driving force, as Wright postulates, in favoring the subdivide-merge scheme of within-between line selection.

METHODS

The Genetic Model

The model of population structure underlying the simulation used in this study closely follows the repeated-cycle of Madalena and Hill (1972) and is also somewhat similar to the experimental design of Katz and Enfield (1977) and Rathie and Nicholas (1980). All simulations to be reported involved comparisons of performance of a population under a subdivide-merge type of selection scheme and a population having identical genetic characteristics but maintained as a single large mass selected unit. The most important criteria of comparison were: (1) probability of fixation of the most favorable alleles, (2) average phenotypic value at the selection plateau, and (3) the average time taken to reach fixation at all loci. Each of the sets of simulations was replicated a large number of times (typically 90) and the statistics obtained were averaged over these replicates.

Each individual in a population had n loci that determined its genotypic value y according to the mode of gene action for those loci. Non-epistatic models of gene action were initially used to replicate some of the earlier experiments of Madalena and Hill. This provided an independent program replication of their results in addition to providing a check on our program. A two locus and a three locus multiple peak epistatic model (each with two alleles) were used in the later simulation experiments. The genotypic values for the two models are given in Table 1. Certain points about the model are worth noting : (1) the model is symetrical with all genes having equal effects, and (2) there is no dominance. The phenotypic value of an individual is the sum of its genotypic value and a normally distributed environmental effect e (independent between individuals, and with mean zero and variance σ_e^2).

Each replicate was initiated at generation 0 by sampling Q monoecious individuals from an infinite base population in Hardy-Weinberg equilibrium. In the subdivide-merge scheme the Q individuals were divided into m sublines with M (M = Q/m) individuals in each subline. The best N individuals within each subline were chosen by truncation selection to be parents the next generation. The within-subline selection was continued for T generations. At generation T the v lines with the best average performance were selected. The best N individuals in each of these v lines were saved and the resulting Nv individuals randomly mated to form generation T+1. This generation, a single population of size Q was randomly mated and the Q progeny divided into m sublines to start another among-within selection cycle.

Table 1. Genotypic values for epistatic models

Two-locus model

	A_2A_2	A_2a_2	a_2a_2
A_1A_1	2.0	1.0	0.0
A_1a_1	1.0	0.75	0.50
a_1a_1	0.0	0.50	1.0

Three-locus model

	A_2A_2	A_2a_2	a_2a_2
A_1A_1	A_3A_3 = 4.0	3.0	2.0
	A_3a_3 = 3.0	2.375	1.750
	a_3a_3 = 2.0	1.750	1.500
A_1a_1	3.0	2.375	1.750
	2.375	2.125	1.875
	1.750	1.875	2.0
a_1a_1	2.0	1.750	1.500
	1.750	1.875	2.0
	1.500	2.0	2.500

To summarize, each selection cycle was composed of T generations of within-line selection (with intensity i_w = N/M) followed by 1 generation of among-line selection (intensity i_b = v/m) and 1 generation of random mating to permit recombination. In the parallel undivided, mass selected population the individuals were maintained as a large population of size Q and truncation selection was practiced with Nm individuals selected to be parents each generation. The selection intensities were therefore the same (Nm/Q = N/M) as in the subdivided populations for those generations when within-line selection was practiced . The cycles of selection were continued until fixation had occurred at all loci or until a preset number of generations had elapsed.

Initial parameters of interest in the model investigated by the simulation were : (1) initial allelic frequencies, (2) number of loci, (3) types of gene action, (4) linkage effects, and (5) magnitude of environmental effects. The population state at any point in time could be summarized by mean phenotypic value, allelic frequencies, and level of linkage disequilibrium.

Since most simulation replicates were run until fixation had occurred at all loci the allele frequency for any given locus was either zero or one at termination. The running of a large number of replicates could therefore be thought of as tracking a binomial random variable where success could be equated to absorption at 1.0. The probability of absorption at 1.0 for a locus could be empirically determined by averaging the final allelic frequencies for the locus for all the replicates of a given simulation. Since the proportion of replicates fixing at 1.0 were available and the number of replicates known the standard error of the estimate could be determmined. Probabilities of fixation will be used in the Results and Discussion section as the probability of fixation at 1.0. Data on average time to fixation were also summarized.

The Parameter Sets

After initial testing of the simulation program the first group of tests were replications of a subset of Madalena and Hill's (1972) runs . These were followed by the simulations based on the two locus multiple peak epistatic models. Finally, three locus epistatic models and mixed epistatic and non-epistatic models were investigated. Preliminary runs of parameter sets that indicated potential benefits for the subdivide-merge scheme were investigated more closely. Certain parameters were kept invariant: the population size runs; the number of sublines (m) was 8; experiments were generally replicated 90 times; each replicate was maintained for 80 generations or fixation for all loci; the number of generations of within-line selection was 3, yielding a selection cycle of 5. The assumed values of some of these parameters were adopted from the values used by Madalena and Hill (1972). The total population size of 80 was selected to minimize random drift in the large population. The 8 sublines required a subpopulation size of 10 individuals, which was small enough that random drift could take place . Other relatively invariant parameters were selected based on what seemed reasonable from a practical breeding standpoint. Many of the other parameters took "high " and "low" values. Within-line selection intensities of .2 (i_w = 2/10) and .4 (i_w = 4/10) were used. Among-line selection intensities of .25 (i_b = 2/8) and .5 (i_b = 4/8) were explored.

The model assumes that many genes affect the trait and the effects of each is small relative to total phenotypic variance. The environmental variance was chosen to reflect this. First a maximum value of additive genetic variance was calculated subject to the constraint of equal allelic frequencies for all loci. Then assuming 25 to 50 such loci and heritabilities of about .3 a range was calculated for environmental variance. A value for σ_e^2 of 10 was used for most runs. In some cases an environmental variance of 1.0 was used to simulate major gene effects.

Linkage effects with initial high positive and negative linkage disequilibrium values as well as initial linkage equilibrium were investigated. Linkage values tested included free recombination and recombination values of .2 and .1 .

RESULTS AND DISCUSSION

Limitations on length of manuscript and time for presentation preclude an exhaustive presentation of all the results obtained. These will be published in more detail in a later manuscript so I will try to concentrate on the major highlights in this paper.

Non-epistatic models

I will only briefly summarize the work with the non-epistatic models since the results are highly consistent with the results of Madalena and Hill. Mention should be made of the fact that our simulation differed from theirs in the basic assumptions concerning the magnitude of gene effects. Their simulation used a low environmental variance giving rise to rather steady phenotypic plateau means over replicates. The loci simulated were, in other words, loci with large effects. With the larger environmental variance simulated in our work we have tended to concentrate on the probability of fixation of the favorable alleles as a more significant parameter. With a large environmental variance, picking up significant differences in mean phenotypic values requires an unreasonably large number of replicates to have adequate resolving power to compensate for the background noise. Nevertheless, this difference in modeling led to no obvious differences in results.

In the additive model the performance of the large population is markedly better if the initial allelic frequency for the favorable allele is low. (In our case .10). This is due to the chance fixation of the unfavorable allele in the small sub-populations which are more subject to the drift effects. As the initial allelic frequencies are increased any advantage for the large population disappears by the time starting frequencies from .5 to .7 are reached. If the gene action is dominance for the favorable allele the large, undivided population is again favored. Here both chance fixation and local inbreeding effects work to lower the subdivided population mean.

The model where the favorable allele is the recessive is the only non-epistatic kind of gene action where there was an advantage for the subdivide-merge scheme and here at only low (.10) initial allelic frequencies. In this case the local inbreeding effect works to the advantage of the sub-populations.

One other result described by Madalena and Hill was also also noted in our results : intense selection among lines in the subdivided population was detrimental to the final plateau but could produce a higher mean in the short term but rarely beyond the first cycle of selection.

Two-locus epistatic model

A brief summary of the characteristics of the two-locus model shown in Table 1 will be useful before examining the simulation results. The model has two adaptive peaks of unequal value. The model is symetrical in the sense that both genes have equal effects. There is no dominance for the alleles favoring the higher peak. An unstable equilibrium point or saddle occurs when the allelic frequency of the favorable allele for both genes is at .33... . In a very large population in the absence of drift selection would take the population to the higher peak if the initial allelic frequencies for the favorable allele at both loci were greater than .33 . If both are less than .33 the population would move to the lower peak with selection for increased phenotypic value. In the case of initial allelic frequencies on opposite sides of the equilibrium value of .33 the outcome is in doubt and is a function of the relative location of the two genes on the allelic frequency plane. The population can move toward a plateau only after both genes have frequencies on the same side of the equilibrium point.

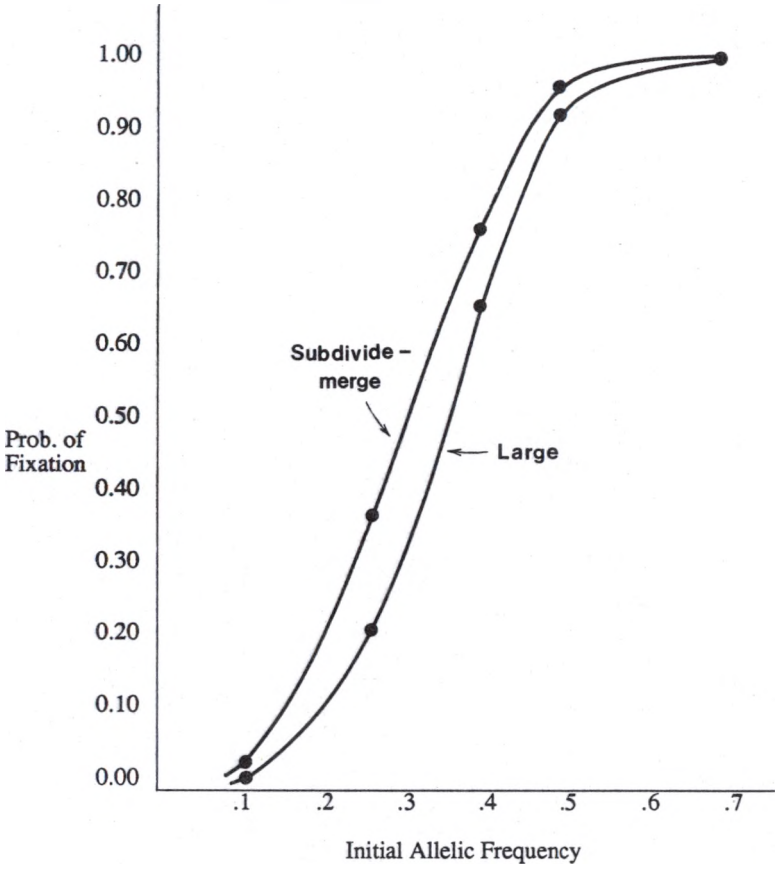
Table 2 presents the results of the first series of simulations for the two-locus model where the genes were unlinked and the population was initiated in linkage equilibrium. A subset of these data is presented graphically in Figure 1. In the first two comparisons of Table 2 when the genes start on opposite sides of the equlibrium point there is always a tendency

Table 2. Two-locus epistatic model

Population size	Q = 80	Recombination frequency	c = .5
No. of sublines	m = 8	Linkage disequilibrium	D = 0
Gens. of Selection	T = 3	Environmental variance	σ^2_e = 10.0

Selection intensity within	between	allele frequency q_1	q_2	Y ± 2 S.E.		fixation time	P(fix) ± 2 S. E.	
16/80		.20	.70	1.72	.11	16	.71	.10
2/10	4/8			1.63	.12	18	.70	.10
16/80		.20	.50	1.47	.12	17	.49	.10
2/10	4/8			1.47	.12	18	.56	.10
16/80		.10	.10	.96	.07	10	.02	.04
2/10	4/8			1.03	.08	13	.04	.04
2/10	2/8			1.02	.08	9	.03	.04
32/80		.10	.10	1.02	.07	14	.01	.02
4/10	4/8			1.06	.08	12	.02	.03
4/10	2/8			1.00	.07	9	.03	.04
16/80		.25	.25	1.28	.11	17	.22	.08
2/10	4/8			1.35	.12	16	.38	.10
2/10	2/8			1.25	.12	11	.31	.10
32/80		.25	.25	1.19	.12	26	.21	.08
4/10	4/8			1.30	.11	22	.24	.09
4/10	2/8			1.16	.12	16	.21	.08
16/80		.40	.40	1.74	.14	18	.67	.10
2/10	4/8			1.65	.12	18	.76	.08
2/10	2/8			1.68	.14	10	.69	.10
16/80		.50	.50	1.86	.09	15	.92	.06
2/10	4/8			1.85	.08	14	.95	.04

Figure 1. Probability of Fixation of Favorable Alleles for Different Initial Allelic Frequencies



Population size = 80 Sublines = 8
Lines selected = 4 No. selected / line = 2

No. of replications = 90

toward fixation for the favorable allele when compared with the mean starting allelic frequency . The probability of fixation of the favorable allele is very close to the initial frequency of the favorable allele for both population structures. For any given starting frequency between .1 and .5 when the starting frequency of the favorable allele is the same for both genes the subdivide-merge scheme is always favored (Figure 1) even though the difference is usually small. The difference is greatest when the starting frequencies are close to the unstable equilibrium point. In the subdivide-merge scheme there is a slight advantage to the system where the among-line selection is less intense if the more intense selection is practiced within lines. These same simulations were repeated with the only change in modeling being that the recombination frequency was changed to simulate a tight linkage model of .1 but the initial linkage equilibrium was maintained. The results were almost identical with the exception that the average time to fixation was shortened by slightly more than 2 generations on average over all replicates.

Table 3 provides the results for a series of runs where a recombination frequency of .2 was simulated and the populations were started in a negative phase of linkage disequilibrium. These results in general show that any advantage of the subdivide-merge scheme that had existed with free recombination or tight linkage but initial linkage equilibrium has been lost. The cases where the populations are initiated with allelic frequencies on the opposite side of the equilibrium frequencies now favor the large undivided population. In no case is there a distinct advantage for the sub-population structure.

Extended epistatic models

If the basic epistatic model is extended to three interacting loci (see Table 1) the range of initial allelic frequencies at which the subdivide-merge scheme shows any advantage is reduced. These results are summarized in Table 4. When the initial allelic frequencies are at .25 there appears to still be a slight advantage for the subdivided populations. This would suggest that as the number of interacting genes increase and the corresponding number of peaks and valleys increase i.e. the interactions become more complex, the advantage of the subpopulation appears to decline. This seems somewhat counter-intuitive to expectations. This model needs further investigation, probably with a reduction in the environmental variance, to ascertain if the results still hold if the genes were major effect genes.

Finally, the results of simulations using a mixed model (2 loci of the multiple peak type and 2 non-epistatic and completely dominant for the favorable allele) are presented in Table 5 . The intent here was to ascertain whether the subdivide-merge scheme still retained its advantage when other loci which act in ways that are known to be best exploited by a single, large mass selected population are included. The second set of results included in this table are for a situation where the genes are assumed to have major effects. In this case the environmental variance was reduced to 1.0 from the 10.0 used in the other simulations. In the cases where the environmental variance was large the subdivided routine maintained an advantage when measured in terms of the plateau mean or probability of fixation as long as the between-line selection was not too intense. With the small component of environmental variance the intense selection among lines was not as detrimental. The most striking advantage for the subdivide-merge scheme in this analysis was the case of the epistatic genes being initiated at .2 and .5 with the low environmental variance. The probability of fixation for the most favorable alleles at the epistatic genes differs by .4 in the two systems while the dominant favorable alleles always get fixed positively in the two systems. These results tend to emphasize the importance of the assumptions concerning the magnitude of gene effects when comparing the two selection systems.

The results taken in total confirm Wright's hypothesis that a subdivide-merge breeding scheme can be the most effective breeding scheme when multiple peak epistasis is important. Its relative advantage will, however, depend on many other parameters some of which may be sufficiently important that they negate the advantage of the system or even make it the least effective. Initial allelic frequencies have a major effect on outcome as well as the initial state of linkage disequilibrium. From a practical breeding point of view this means that the choice of

Table 3. Two-locus epistatic model
Linkage : Negative initial linkage disequilibrium

Population size	Q = 80	Recombination frequency	c = .2
No. of sublines	m = 8	Linkage disequilibrium	D < 0
Gens. of selection	T = 3	Environmemtal variance	$\sigma^2_e = 10.0$

Selection intensity within	between	allele frequency q_1	q_2	Y ± 2 S.E.		fixation time	P(fix) ± 2 S.E.	
16/80		.20	..70	1.58	.12	18	.67	.10
2/10	4/8			1.41	.14	20	.53	.10
16/20		.20	.50	1.34	.13	18	.35	.10
2/10	4/8			1.24	.13	17	.31	.09
16/80		.10	.10	.92	.06	9	.00	.00
2/10	4/8			1.04	.07	11	.03	.03
2/10	2/8			.97	.07	8	.01	.02
32/80		.10	.10	1.09	.09	14	.04	.04
4/10	4/8			.99	.06	11	.03	.03
4/10	2/8			.92	.07	8	.00	.00
16/80		.25	.25	1.16	.11	16	.19	.08
2/10	4/8			1.16	.11	13	.19	.08
2/10	2/8			1.07	.11	9	.14	.07
32/80		.25	.25	1.17	.10	24	.19	.08
4/10	4/8			1.15	.11	20	.20	.08
4/10	2/8			1.06	.12	13	.17	.08
16/80		.40	.40	1.59	.12	18	.56	.10
2/10	4/8			1.41	.12	17	.55	.10
2/10	2/8			1.25	.14	11	.47	.10
16/80		.50	.50	1.90	.10	17	.87	.07
2/10	4/8			1.76	.15	18	.82	.07

Table 4. Three-locus epistatic model

Population size	Q = 80	Recombination frequency	c = .5
No. of sublines	m = 8	Linkage disequilibrium	D = 0
Gens. of selection	T = 3	Environmental variance	$\sigma^2_e = 10.0$

Selection intensity within	between	allelic frequency q_1	q_2	q_3	Y ± 2 S.E.		fixation time	P(fix) ± 2 S.E.	
16/80		.20		.70	3.63	.16	20	.82	.08
2/10	4/8		.50		3.63	.16	20	.83	.08
16/80		.10		.10	2.43	.07	10	.00	.00
2/10	4/8		.10		2.46	.07	15	.00	.00
2/10	2/8				2.50	.07	10	.02	.02
32/80		.10		.10	2.49	.08	18	.00	.00
4/10	4/8		.10		2.50	.07	15	.00	.00
4/10	2/8				2.54	.06	10	.00	.00
16/80		.25		.25	2.74	.13	21	.17	.08
2/10	4/8		.25		2.82	.14	18	.23	.08
2/10	2/8				2.62	.14	12	.22	.08
32/80		.25		.25	2.71	.12	29	.16	.08
4/10	4/8		.25		2.74	.12	31	.20	.06
4/10	2/8				2.72	.14	20	.20	.08
16/80		.40		.40	3.57	.16	22	.74	.08
2/10	4/8		.40		3.55	.16	22	.69	.08
2/10	2/8				3.34	.16	14	.67	.10
16/80		.50		.50	3.86	.12	18	.93	.06
2/10	4/8		.50		3.83	.10	17	.91	.06

Table 5. Mixed model

Two-locus multiple peak epistatic genes (loci 1 and 2)

Two dominant favorable genes (loci 3 and 4)

Population size	Q = 80	Recombination frequency	c = .5
No. of sublines	m = 8	Linkage disequilibrium	D = 0
Gens. of selection	T = 3	Environmental variance	$\sigma^2_e = 10.0$

Selection intensity within	between	allelic frequency q_1, q_3	q_2, q_4	Y ± 2 S.E.		fixation time	P(fixation) 1,2	3,4
16/80		.20	.70	3.62	.18	35	.60	.99
2/10	4/8			3.67	.17	29	.80	.96
16/80		.10	.10	2.67	.17	37	.00	.82
2/10	4/8			2.80	.16	42	.04	.87
2/10	2/8			2.53	.21	29	.04	.62
16/80		.25	.25	3.16	.14	35	.22	.99
2/10	4/8			3.28	.18	31	.30	.97
2/10	2/8			3.04	.18	19	.24	.89
16/80		.50	.50	3.87	.12	32	.84	.99
2/10	4/8			3.86	.12	26	.88	.99
				Environmental variance $\sigma^2_e = 1.0$				
16/80		.20	.70	3.93	.08	21	.92	1.00
2/10	4/8			3.93	.06	16	.95	1.00
16/80		.20	.50	3.50	.14	20	.50	1.00
2/10	4/8			3.91	.08	16	.91	1.00
16/80		.10	.10	3.03	.04	20	.02	1.00
2/10	4/8			3.03	.04	25	.02	1.00
2/10	2/8			3.02	.06	19	.03	1.00
16/80		.25	.25	3.58	.12	21	.62	1.00
2/10	4/8			3.61	.12	16	.64	1.00
2/10	2/8			3.66	.14	12	.65	1.00
16/80		.40	.40	3.92	.07	20	.94	1.00
2/10	4/8			3.97	.04	17	.98	1.00
2/10	2/8			3.98	.02	12	1.00	1.00

foundation material may be as important as the selection system employed. Probably more importantly, it is apparent from these results combined with those of Madalena and Hill that ,on average, multiple peak gene action must be the overriding kind of gene action for the subpopulation program to be the most effective. The experimental results with laboratory organisms would raise serious questions whether this is the case for the traits evaluated thus far.

REFERENCES

BOWMAN, J.C. and FALCONER, D.S. 1960. Inbreeding depression and litter size in mice. Genet. Res. 1 ,262-274.

GOODWILL, R. 1974. Comparison of three selection programs using *Tribolium castaneum*. J.Hered. 65 , 8-14.

KATZ, ALAN J. and ENFIELD, FRANKLIN D. 1977. Response to selection for increased pupa weight in *Tribolium castaneum* as related to population structure. Genet. Res. 30 ,237-246.

MADALENA, F. E. and HILL, W. G. 1972. Population structure in artificial selection programmes, simulation studies. Genet. Res. 20 , 75-99.

RATHIE, K.A. and NICHOLAS, F.W. 1980. Artificial selection with differing population structures Genet. Res. 36 , 117-131.

SLATKIN, M. 1981. Fixation probabilities and fixation times in a subdivided population. Evolution 35. 477-488.

WRIGHT S. 1963. Plant and animal improvement in the presence of multiple selective peaks. Statistical Genetics and Plant Breeding pp. 116-122. Pub. No. 982 ,Natl. Acad. Sci. , N.R.C. Washington D.C.

WRIGHT , SEWALL 1977. Shifting balance theories of evolution. Evolution and the Genetics of Populations. Vol. 3 pp. 443-473. University of Chicago Press, Chicago.