

This dissertation has been 63-5181
microfilmed exactly as received

GILL, John Leslie, 1935-
EFFECT OF POPULATION SIZE, SELECTION
INTENSITY, LINKAGE AND NON-ADDITIVE
VARIABILITY UPON GENETIC CHANGE IN
SIMULATED POPULATIONS.

Iowa State University of Science and Technology
Ph.D., 1963

Biology-Genetics
University Microfilms, Inc., Ann Arbor, Michigan

EFFECT OF POPULATION SIZE, SELECTION INTENSITY, LINKAGE
AND NON-ADDITIVE VARIABILITY UPON GENETIC CHANGE
IN SIMULATED POPULATIONS

by

John Leslie Gill

A Dissertation Submitted to the
Graduate Faculty in Partial Fulfillment of
The Requirements for the Degree of
DOCTOR OF PHILOSOPHY

Major Subject: Animal Breeding

Approved:

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

Head of Major Department

Signature was redacted for privacy.

Dean of Graduate College

Iowa State University
Of Science and Technology
Ames, Iowa

1963

TABLE OF CONTENTS

	Page
INTRODUCTION	1
REVIEW OF LITERATURE	4
GOALS	24
MECHANICS OF SIMULATION	31
DISCUSSION OF RESULTS	55
Analysis of Genotypic Means	55
Population size	62
Selection intensity	110
Environmental variation	162
Linkage	177
Achieved and Predicted Selection Differentials	186
Estimation of Realized Heritability from Selection Experiments	195
SUMMARY AND CONCLUSIONS	203
LITERATURE CITED	221
ACKNOWLEDGMENTS	229
APPENDIX A: FIGURES	230
APPENDIX B: TABLES	297

INTRODUCTION

In attempting to understand the basis for genetic improvement of plants and animals our research objectives frequently are stated in terms of testing predictions based on mathematical formulations and prior estimates of genetic parameters. Frequently the scope of the experiment required for reliable results is beyond the available resources of time and facilities. Sometimes the mathematical treatment of the problem may be inadequate or the underlying assumptions are questionable. In such cases seeking answers by doing the research exclusively with each species of plant or animal is inefficient if partial solutions to important questions can be found by other means. Until recently the major alternative has been to conduct the research with laboratory organisms. However, in the last few years numerical methods have been introduced into quantitative genetics research. The chief technique has been to create populations with high-speed computers and to require their members to perform according to the laws of probability associated with specified genetic systems. This technique of simulating stochastic processes through the use of repetitive sequences involving random numbers has been termed the Monte Carlo technique. With the Monte Carlo technique, genetic forces and mechanisms can be simulated and observed where true population parameters, modes of gene action and

linkage relationships are known exactly. This is something that can never be achieved in even the most accurately controlled laboratory or field experiment.

The more basic the question we ask, the more likely it is that we can find an answer by Monte Carlo which would apply reasonably well to plant and animal breeding. Conversely, the more the question concerns only one species or only one of its products, the less likely is Monte Carlo to provide an answer which is dependable or complete.

The two major areas in quantitative genetics where Monte Carlo techniques appear to be most useful at this time are:

- (1) Where inferences, based on results of experiments, or mathematical formulations, based on a body of underlying assumptions, need confirmation, and
- (2) Where mathematical formulations, concerning a problem relating to economic species, are lacking because of the complexity of the situation, and the answers are not readily obtainable from experiments.

For example, the development of the mathematical theory of genetic selection for polygenic characters has necessarily been restricted almost entirely to the case of genetic effects originating from the genes at one locus or a number of

independently segregating loci without epistasis. A mathematical description of the basic facts of polygenic inheritance in complex genetic situations results in equations too cumbersome for solution. Simplifying assumptions lead to large departures from reality, somewhat in proportion to the simplification achieved. For the polygenic case we long have had formulations for selection without epistasis and for the effects of linkage on the approach to equilibrium under random mating without selection, but no one has been able, mathematically, to derive valid equations for selection in finite populations in the presence of linkage and epistasis.

The present study was undertaken as a feasible approach to increased understanding of genetic selection. Perhaps its major usefulness will not be in obtaining direct answers to problems but in clarifying thoughts and definitions of the problems to be investigated on the economic species themselves.

REVIEW OF LITERATURE

Because of the broad scope of the problem of understanding genetic selection and its multitudinous ramifications, the literature on various aspects of the subject is understandably quite diverse in origin and nature and comprises an imposing record of research. The impracticality of undertaking a comprehensive review of the entire field of research in quantitative genetics concerning selection suggests that an intensive review of those studies concerned with the Monte Carlo approach, using simulated genetic populations, would be preferable to a cursory and, perhaps, superficial treatment of the corporate testimony. Consequently, this premise has been followed to the extent that the more classical approaches to the study of genetic selection have been relegated to the section on discussion of results, where points pertinent to the topic at hand will be introduced discriminately.

Fraser (1957a) introduced the simulation of genetic systems by digital computers with a program which allowed for the following types of problems.

- (1) Effects of linkage on the efficiency of selection
- (2) Competitive efficiencies of alleles as influenced by the parameters of population size, selection intensity, etc.
- (3) Comparison of efficiencies of different breeding

plans for varying degrees of inter-locus interactions.

The binary symbols 1 and 0 were substituted for normal and mutant alleles and a , ah and 0 symbolized the phenotypic contributions of genotypes AA, Aa and aa, respectively. Therefore, $h = 1$ for complete dominance and $h = \frac{1}{2}$ for no dominance. Using diagonal matrices to represent many loci he showed that the phenotypic value could be determined by

$$P_i = (A\&B)a_i + (A\equiv B)a_i h_i,$$

where A is the maternal gamete, B is the paternal gamete, (A&B) is the logical product of A and B and (A≡B) is the logical equivalent of A and B. The phenotypic value P_{AB} then is the sum of the diagonal elements of matrix F_i . The logical product of two arrays of alleles which are ones or zeros simply produces ones where both alleles are one and gives zeros elsewhere, i.e., it identifies loci which are homozygous for the favorable allele. The logical equivalent of the same two arrays of alleles produces zeros when two alleles are equivalent in state and gives ones when they are not, thus identifying the heterozygous loci.

Segregation was simulated by testing a random number, r_i , in the range $0 < r_i < 1$, against 0.5 for each digital position of (A≡B) which contained 1, i.e. for heterozygous loci. Recombination was achieved by transforming a vector of frequencies of recombinants, (f_i) , by sequential summa-

tion to give (F_i) . For example, if a genotype AB/ab exists the frequencies of the four types of gametes which may be produced are $\frac{1}{2}(1-r)[AB]$, $\frac{1}{2}r[Ab]$, $\frac{1}{2}r[aB]$ and $\frac{1}{2}(1-r)[ab]$, where r is the recombination frequency. Then if r should be .2 say, the frequencies would be .4, .1, .1 and .4, respectively. Sequential summation of the array of frequencies $[f_i]$ provides the array $[F_i] = [.4, .5, .6, 1.0]$. Then a random number, r_i , in the range $0 \leq r_i \leq 1$ was generated and tested across (F_i) until $F_i < r_i < F_{i+1}$, indicating the i th term in a vector of types of gametes as the gamete produced.

Environmental effects were included by specifying a function $x = f(r)$ such that "if r is a random number in the range 0 to 1, then x is a random normal deviate."

Selection was accomplished by ordering phenotypes of the progeny and selecting the corresponding genotypes of top, bottom or middle phenotypes according to the type of selection. No account was taken of variation in number of progeny per parent.

Inter-locus interactions were simulated by specifying three matrices of order n (three genotypic states, n loci) in which the rows specified the locus modified and the columns specified the modifier locus, with a modulus of four. Zero indicates no interaction; one, two and three specify types of non-linear interaction. The chief diffi-

culty with this system is the large amount of time and memory space required for evaluation.

In a concurrent paper Fraser (1957b) characterized the effects of linkage on rates of advance under selection. He concluded that detection of the occurrence of linkage in a genetic system from heterogeneity of the rate of advance under selection is unlikely to be possible in large populations unless linkage is fairly tight, i.e. $r \approx 0.005$. Here the simulated study is somewhat in conflict with the polygene concept of Mather (1943) which he based initially on evidence from the results of selection experiments with Drosophila melanogaster. However, the simulation study suffered from two major defects: (1) limitation of loci to seven or less and (2) lack of provision for dominance relations other than complete dominance.

The 20 runs of the simulation program, each replicated, were divided into two sets, P and Q. In the P set the genetic structure of the original population was such that two cross-overs must occur before homozygous individuals could occur. The Q set required four cross-overs. Two sizes of population were considered; a medium-sized population with low selection intensity (50/100) and a small population with high selection intensity (4/40). The six loci were divided into two linkage groups and runs were made at five different values of intra-group linkage rela-

tions for adjacent loci (.5, .25, .05, .025, .005).

In the P set of runs the effect of linkage was marked only for $r = .005$ and the small populations advanced at a distinctly slower rate despite the stronger intensity of selection. In the Q set of runs the larger populations showed a clearer relation of the rate of advance under selection to the amount of recombination, and the higher selection intensity in the small populations was effective.

Considering all runs, distinct discontinuities in the rate of advance were observed only for recombination values of $r = .005$.

Barker (1958a and b) continued the series on simulation of genetic systems with concurrent papers on selection between alleles at an autosomal locus and at a sex-linked locus. In the first of these he simulated one autosomal locus with two alleles segregating in a bisexual population with no overlap of generations. Large and small population sizes and coefficients for four selection processes (survival from fertilization to sexual maturity, reproductive selection, selection between gametes at meiosis in heterozygotes and selection of successful gametes) were specified by using adaptive values estimated for certain genotypes of Drosophila melanogaster in selection experiments with large and small populations. The large experimental and simulated populations agreed fairly closely as to proportions of

genotypes over a period of 15 generations except for less variation in the simulated population. This may have been due to constant adaptive values and lack of generation overlap in the simulated population or to chromosomal sampling variation in the experimental population. The author concluded that generation overlap has little effect except on variation.

In the small populations, the simulated population changed genotypic proportions more slowly from the third to the twelfth generations than did the experimental population. This was explained by the constant reproductive coefficients of the simulated population versus possible changing coefficients for the experimental population.

The chief asset of the experiment was in showing that it is possible to simulate the operations of natural selection between two alleles at an autosomal locus.

In the paper concerning selection between alleles at a sex-linked locus, Barker described striking differences between experimental and simulated populations. A genotype which was eliminated from a large experimental population of Drosophila pseudoobscura by generation nine was tending toward equilibrium in the simulated population. The simulated populations agreed closely with one another and fluctuated at random around the expected curve, i.e. the simulation provided, for given selective values, an accurate descrip-

tion of the expected changes in frequency in a large population.

The differences between the results from experimental and simulated populations emphasizes the inadequacy of available methods for estimating the fitness of competing genotypes.

Fraser (1960a) presented the fifth paper of the Sydney series on simulation of genetic systems at an international symposium sponsored by the Biometrics Society and the International Union of Biological Sciences, held at Ottawa in 1958. He reiterated the simulation procedure described in the first paper of the series and again emphasized the point made in the second paper; that large populations (100) with low selection intensities ($\frac{1}{2}$) showed no periods of slow response followed by periods of sudden response, as proposed by Mather (1943), for any linkage values. According to Mather's hypothesis, individual genes of polygenic units occur in balanced combinations. Fraser pointed out that it seems unlikely that such combinations would be favored by selection to any marked extent unless some epistasis occurs. He presented data from a simulation study designed to evaluate this premise. In this study the parameters were allowed to vary in each simulation trial except that the degrees of dominance and epistasis and the tightness of linkage were under genetic control within fixed ranges. Consequently,

when selection against phenotypic extremes was simulated, response to selection was measured by the reduction of phenotypic variability that occurred. This reduction was analyzed into components due to homozygosity, dominance, epistasis and linkage. He concluded that "it appears from the available data that selection for decreased variability will strongly favor epistasis if this can be varied to a form that relates genotype/phenotype by a sigmoid function. The existence of such functions is the basis of Waddington's concept of developmental canalization."

If intermediates are favored then deviation from the optimum value is disadvantageous. Therefore, selection will operate against the causes of deviation, and will tend to produce more stability so that development is "canalized" along the path that leads to the optimal phenotypic expression. The role ascribed to selection is its discrimination against alleles that increase variability.

Unfortunately the epistasis which Fraser simulated was not specific in nature. It involved a non-linear transformation of the "additive + dominance phenotype". The 20 loci were divided into A, D, E and F sub-genotypes of five loci each. The A sub-genotype was purely additive in the absence of modifying sub-genotypes; dominance of these loci was determined by the D sub-genotype; epistasis of the A loci was determined by E and F sub-genotypes as being first

and second order deviations from linearity. The relation of D, E and F sub-genotypes to the d, e and f coefficients of dominance and epistasis was determined on the basis of an additive action of the constituent loci, i.e., there were eleven values for d, e or f. Phenotypic values were computed in three steps:

- (1) From A, count x loci homozygous for 1-alleles and y heterozygous loci.
- (2) From D, count 1-alleles for d. Then Additive + Dominance components = $X = 2x + dy$.
- (3) From E and F, count 1-alleles for e and f. Then $r = X + eX^2 + fX^3$.

This formulation, while simulating inter-locus interactions of a sort, is far removed from the classical concepts of epistatic interactions, and may or may not be a realistic phenomenon when associated with a particular selection scheme. Fraser admitted that it is more generally valid to base any analysis of epistasis on the interactions of individual genes. However, using this system, he found that selection against extremes caused a shift of the D sub-genotype towards the d coefficient having a value of 1.0 in a scale from 2.0 to 0, i.e., almost no dominance, and a shift of the E and F sub-genotypes towards the e coefficient having a large positive value and the f coefficient having a small negative value. The result at the end was

the sigmoid relationship of genotype to phenotype.

Fraser (1960b) continued the discussion of epistasis in the sixth paper of the Sydney series on simulation of genetic systems, concentrating on the problem of existence of genetic variability in the absence of phenotypic variability and the possibility that selection could modify epistatic interactions and reduce the trend towards homozygosity. He explained a modified simulation program in which forty loci instead of twenty were involved, number of offspring was increased, and a new, faster method of simulating segregation and recombination, through identification of genotypic state by the logical product of the maternal and paternal gametes, was used rather than the slower method of the "random walk". Using this program he found that selection against extreme phenotypes produced a much greater reduction of variability for a specified deviation from the initial gene frequency in the epistatic runs than in the additive runs. The effect of selection on fixation in the epistatic runs was small, as was the effect on deviations from original gene frequencies in small populations. But the effect of selection on deviations from original gene frequencies in large populations was significant. The magnitude of the effect may be related to the marked effect of population size on the percentage of fixed loci and deviations from original gene frequency.

In contrast to the simple additive system where selection against phenotypic extremes led to fixation at a slow rate, such selection operated, in complex epistatic systems, to modify the relation of the genotype to the phenotype into a sigmoid function. If this function were genetically fixed, it would have unexpected effects on the progress of selection towards extreme phenotypes. Initially, selection would cause the usual distribution shift of gene frequencies at the additive loci but then phenotypic variance would decrease sharply, due to an increase in the proportion of genotypes having the same phenotype. Further selection for the extreme phenotype would be against genetic extremes resulting in fixation of the additive loci.

Fraser (1960c) presented the seventh paper of the Sydney series on simulation of genetic systems as a direct extension of the previous one on epistasis. He was able to show that the trend towards genetic fixation was predominately controlled by the population parameters of reproductive rate and number of parents, verifying the conclusion of the previous paper that selection against extremes is a minor contributor to the incidence of genetic fixation of a complex genetic system.

Various hypotheses have been provided for the apparent lack of intense selection against extremes of a "canalized" character, one on the direct effects of the character on

some unmeasured aspect of fitness, another on a pleiotropic relationship of the character to other more important characters which are under intense selection. The Monte Carlo results indicated that neither explanation is necessary. In Fraser's terminology, "Selection operative at limits far removed from a canalization zone can cause a sufficient reduction of phenotypic variance." The reason is that selection operates against extreme values of the epistatic coefficients as determined by the quadratic and cubic sub-genotypes, causing a greater degree of "canalization" than is expected from measurements of reproductive fitness. Genetic substitution in the sigmoid functional relationship of genotype-phenotype would produce an opposite effect to the same substitution in a linear additive function because the model is based on a cubic equation. Fraser suggests that a model based on a higher-order equation would not show this unexpected divergence between additive and epistatic systems.

The epistatic rationale presented by Fraser in the last three papers of the series seems clearly dependent upon the premise that some sort of non-linear interaction, based on a cubic equation, exists as a real genetic phenomenon. Perhaps such interactions are related to so-called "canalized" characters but it is worth noting that the consequences of selecting against phenotypic extremes may be quite different

when the epistatic relationship is of a different form.

The chief virtue of the Sydney series of papers on simulation of genetic systems would seem to be that it introduced the simple basic tool, known as Monte Carlo, to the field of quantitative genetics. While most of the work thus far has been restricted to natural selection problems, the ground work has been laid for application to artificial selection and, indeed, for problems not involving selection. The basic simulation techniques presented should prove invaluable to other experimenters and the example has been set for the experimenter to parallel his experiments with theoretical analyses aimed at elaboration of concepts derived from data, or examination of the parameters to define the values which will give an optimum chance of success.

Martin and Cockerham (1960) presented the results of the first Monte Carlo application to quantitative genetics made in the United States at the 1958 Ottawa symposium on biometrical genetics. The chief point of interest was that tight linkage can slow progress from selection when the populations are initially in linkage equilibrium.

Selection was by upper truncation of phenotypes and selected parents were mated at random with replacement of sampled individuals. The program allowed for any degree of dominance but no epistasis, and effects of alleles were equal at all loci. The recombination frequency was the

same between all adjacent loci (not more than 25 in number). Environmental variance, when not zero, was made equal to the expected additive genetic variance in the initial generation. All runs started with expected gene frequencies of 0.5 and in linkage equilibrium except when distinguished as loaded repulsion.

With the initial population in linkage equilibrium with five loci, tight linkage (0.01) slowed progress under the additive model compared with free recombination only when no environmental variation was included. Under the dominance model no apparent effects were noted for selection intensity, environmental variation or recombination frequency. When the initial population had five loci in repulsion, the only significant change from the equilibrium situation was the large effect of level of recombination under both the additive and dominance models.

With the initial population in linkage equilibrium with twenty loci, tight linkage (.01) slowed progress under both models and resulted in fixation of some unfavorable alleles when selection was intense under the additive model. Intensity of selection seemed to affect progress only when free recombination was simulated.

Bohidar (1960) introduced sex-linkage and varying degrees of dominance into a basic simulation program for genetic selection without epistasis or environmental

variation. The most noteworthy fact shown in this study was that linkage resulted in the formation of plateaus of genetic merit when selection was by upper truncation, and tight linkage sometimes resulted in the formation of successive plateaus. In populations initiated in the coupling phase of linkage, selection and linkage complemented each other, i.e. the tighter the linkage, the faster the progress. However, in populations initiated in maximum repulsion with no dominance, progress depended upon the relative strength of the two opposing forces, linkage and selection. Selection points were achieved with linkage values of .5, .3 and .1 but plateaus were formed with linkage values of .03 and .003. With linkage values of .015 and .007, the two opposing forces established a partially balanced relationship where sometimes one, sometimes the other determined progress. For a few generations complete balance was sometimes achieved, resulting in a temporary plateau. At this point random drift became the effective force and the population moved to genetic fixation.

In discussing the effects of degree of dominance, the author concluded that the variability of the population was maintained longer under a complete dominance model than for no dominance and thus selection was sensitive over a longer period of time. In addition, populations under complete dominance drifted toward fixation at a slower rate. With

complete dominance and intermediate truncation selection the selection point was readily attained, especially with tight linkage, in contrast to the no dominance case, where a fluctuating plateau was invariably favored. With the overdominance model and the population in initial maximum repulsion, progress depended largely on random drift. Variability was maintained but the opposing forces of selection and linkage became weak, with random drift leading to a fluctuating plateau when selection was for upper extremes. The progeny and parent populations formed two distinct plateaus. As linkage was increased the two plateaus pulled apart to become non-overlapping fluctuating plateaus. For intermediate truncation selection under overdominance, random drift formed a fluctuating plateau when little or no linkage was simulated, but as linkage tightened, the selection point was achieved, i.e., selection and random drift became complementary to each other in this situation.

Baker and Comstock (1961) presented results of a recent Monte Carlo study involving quantitative genetics at a symposium on statistical genetics and plant breeding at Raleigh, North Carolina, in 1961. Proceedings of this symposium are presently being printed by the National Research Council of the National Academy of Sciences in Washington, D. C. The Monte Carlo study was directed toward evaluating the premise that the average genotype of the heterozygote at a specific

locus can be enhanced by effects of linked loci under specified conditions. This premise was supported by the results of mathematical analysis by Robinson and Comstock (1955), which related enhancement of the heterozygote to the amount of linkage disequilibrium. The Monte Carlo results showed that the mean linkage disequilibrium increased more for adjacent loci than for those separated by five loci in selected populations. In all cases disequilibrium was much larger in selected populations than in random ones.

The study was made with complete dominance and equal genetic effects for all loci with two initial parents from a conceptually infinite population assumed to be in linkage equilibrium. The 35 loci simulated were assumed to be on one chromosome with uniform recombination frequency for adjacent loci and no interference. The initial gene frequencies were 0.5 and there were no inter-locus interactions, i.e. no epistasis. The number of parents subsequent to the initial generation varied from two to ten and the offspring population size from four to one hundred. The variance due to the combined effect of the environment and those genes affecting the quantitative character which were not traced, i.e. simulation of approximately ten other chromosomes, varied from zero to 324. Each run was continued for e generations, the number required for an expected decrease in heterozygosity of 50% to occur. The primary issue concerning

how selection and linkage affect the rate of fixation in finite populations or, as they put the question, "Does linkage aid in the persistency of polymorphism?", was discussed on the basis of comparisons among control results, selection-free-recombination results and selection-linkage results.

The results indicated that selection-linkage sets consistently produced a lower proportion of fixed loci than other sets but the differences were not statistically significant in many cases. In addition, selection-linkage sets resulted in genetic means which were just as high as those for selection-free-recombination sets at the same level of environmental variance and the mean level of genetic variance over all generations was much larger in selection-linkage results than in random results.

The fact that linkage of .01 did not hinder the progress of the genotypic mean is in direct contradiction to the results of Martin and Cockerham (1960) and those of Bohidar (1960), where tight linkage resulted in the formation of plateaus of genetic merit. However, the fact that Bohidar's simulation program included no variation due to environment, or to the remainder of a large genome which probably affects most quantitative characters, probably resulted in a considerably larger proportion of fixed loci at a given generation in the selection scheme, giving at least a possible

explanation for the differences observed. Baker and Comstock's (1961) method of simulating extraneous variation may be partially responsible for the negative coefficients of regression of genetic mean on generation which they encountered in most parameter sets, although they dismissed them as due to inbreeding depression because of finite population size.

Larger genetic variance in selection-linkage results was probably due to the linkage disequilibrium contribution to dominance variance described briefly by Robinson and Comstock (1955) and in more detail by Robinson et al. (1960).

Discussion in the three studies which followed the introduction of Monte Carlo techniques to quantitative genetics by the Sydney series of papers was concentrated largely upon the revelation of the effects of linkage on the progress of the genetic mean and on the rate of fixation of genes in finite populations under selection. Even in this area of primacy, agreement was not reached as to the significance of low recombination values as forces upon genetic progress. Innate differences in the simulation procedures, as well as different goals and choices of parameters, undoubtedly contributed to the confusion. Despite obvious and, possibly, unknown shortcomings of the simulations used, these studies ratify the conclusion that the basic simplicity and unusual applicability of the Monte Carlo technique to quantitative

genetics provides a valuable addition to the arsenal of tools with which to assault the enigmas of inheritance. These enlarged opportunities should serve to focus the course of investigation and prod the imagination of other inquisitors.

GOALS

In any discussion of goals it is appropriate to consider the relation of one's own specific goals of the research at hand to the more general goals which embody the thought of at least a majority of the competent workers in the field.

The deductive treatment of selection theory has been largely developed by Fisher, Wright and Haldane. Their conclusions concerning the limited combinations of parameters and restricted situations which they treated, surely, have been key pointers to the direction of pursuits of their contemporaries and the ensuing generation of investigators.

The resolve to quantify the joint effects of selection, epistasis and linkage in finite populations has been primary in the minds of mathematical geneticists for years but, in view of the assumptions they have made to simplify the problem, the multitude of formulations that have been derived can scarcely be said to have contributed more than some good educated guesses as to the true state of affairs. A major deficiency of quantitative genetics is the void between the mathematician working with extremely simple genetic models and the experimenter working with organisms of extreme genetic complexity. The Monte Carlo method circumvents the necessity for advanced mathematical techniques required for solutions to even the simplest genetic models and, to a

large degree, avoids the inadequacy of classical methods for solution of the problems involved in actual genetic systems.

Fraser (1960b) pointed out that genetic models can be devised, programmed, and tested within weeks or months, certainly with sufficient speed for an experimenter to examine many of the theoretical consequences of his ideas before he devises experiments with live organisms. He regarded this as the primary goal in using the Monte Carlo method, though he recognized its use in the methodical examination of the importance of variables determining the effectiveness of selection.

Martin and Cockerham (1960) state that they have come to view the procedure of empirical machine studies as one of detection and suggestion. It is an excellent method of examining the consequences of the assumptions necessary in simplifying genetic systems for mathematical analysis, and of detecting effects of variations in parameters or suggesting how parameters operate. However, since a thorough understanding requires a formulated theory, they believe that the empirical procedure is properly a tool to this end and is not to be used to generate the end product or total response surface.

Robertson (1960) also suggests that the Monte Carlo method is probably the most satisfactory way of handling the

problem of linkage and selection in a population of limited size.

It appears that one may define two major goals or areas of investigation in quantitative genetics which can be considered proper objects of research using the tool of Monte Carlo:

- (1) Examination of theoretical consequences of ideas in order to design better experiments with organisms and comparison of the Monte Carlo results with those found experimentally.
- (2) Examination of the importance and modes of operation of parameters involved in complex genetic systems and relation of the insights obtained to the appropriate existing theory when possible, or suggestion of new theoretical considerations to be examined mathematically.

Now if one is intent upon making investigations in this latter area, more specific questions arise. What genetic systems should be simulated? Which factors are of most importance and how should they be varied? In answer to the first of these questions, it seems clear from the discussion above and from observation of the goals of previous Monte Carlo studies that selection theory for complex finite populations is a principal target. Thus the broad objective of this study is defined as the determination of the effects

of population size, degree of truncation selection, environmental variation, linkage and mode of gene action upon the progress of genetic populations under selection. Obviously, any comprehensive study of these effects should include an attack on the interactions among them and comparisons of various approaches to simulation of realistic situations as conceived in the notions and theories extant at the time. This involves a series of coordinated investigations of which the present effort is an initial study of the general effects of a wide variety of variables and parameters. This pilot study will then be used as a guide for the concentration of subsequent efforts in areas deemed potentially most fruitful.

The design chosen for this initial study was a 4^4 factorial plan for each mode of gene action to be studied. It would be statistically advantageous to be able to study various fractional levels of components of genotypic variance, associated with mode of gene action, within the framework of a factorial plan, i.e., to consider the effects of several levels of the factors of additive, dominance and epistatic variance. However, the meaningful comparisons involve percentages of a whole, precluding the achievement of orthogonal arrays of such levels. Therefore, one must resort to identifying desired levels of these components with particular modes of gene action to be studied separately.

The employment of nine different types of gene action should be sufficient for this purpose. These may be classified into three basic groups.

Simple classical models include:

- (1) Additive
- (2) Complete dominance
- (3) Overdominance.

Classical epistatic models include:

- (4) Optimum gene number
- (5) Duplicate factors
- (6) Complementary factors.

Some conditional epistatic models are as follows:

- (7) Only additive by additive genetic variation initially
- (8) Only additive by dominance genetic variation initially
- (9) Only dominance by dominance genetic variation initially.

One full replicate of the possible combinations in a 4^4 factorial plan with nine models amounts to 2,304 trials. The desire to temper costs and the exploratory nature of a pilot study dictated the use of a 1/16 fractional replicate of the basic plan. This plan is adapted only to the most preliminary type of study. Because of the high fractionation, only the main effects of the factors are suitable for

investigation, i.e., nearly all interactions are confounded. The pooled sum of squares for interactions, with three degrees of freedom, may be used as an estimate of error if the interactions are assumed to have negligible effects. However, the plan has the advantage of balance, because of the orthogonality of the arrays of levels, which some other designs do not have.

The factors deemed of most importance and the levels chosen to represent the effects of each, in as broad a manner as possible with the given facilities, are:

- (1) Parent population size (8, 12, 16, 32)
- (2) Selection intensity (1/2, 1/4, 1/6, 1/8)
- (3) Environmental variation (0, $\sigma_G^2/3$, σ_G^2 , $3\sigma_G^2$),

where σ_G^2 is the expected genotypic variance in the initial generation of progeny,

- (4) Linkage (.005, .05, .2, .5),

in terms of equal recombination frequencies between adjacent loci on each chromosome. The details of adapting these variables to a coherent simulation scheme and the assumptions which must be made will be given in a section on mechanics.

The specific results of primary interest are the genotypic and phenotypic means and variances of the selected and unselected populations in each generation, and those of secondary interest are the amount of allelic fixation and

frequencies of unfixed genes, changes in components of genotypic variance, achieved selection differentials and the heritability realized from selection. From these data the mode of operation of each parameter and the general force exerted on the population under selection can be investigated.

MECHANICS OF SIMULATION

It is not the purpose here to describe in detail the method of programming a genetic system and its modification for various computers since each machine has operations peculiar to itself. However, the logic of simulating, realistically, the processes involved, and the inevitable body of accompanying assumptions are of sufficient importance to require inclusion of some details.

The population simulated is of the unisexual diploid type in which quantitative characteristics are expressed in both sexes. As indicated previously, the principal pressures to which the population is subjected are those of selection, linkage, dominance, epistasis, environmental variation and random fluctuation due to finite size. Selected parents are mated at random by sampling with replacement, and each mating results in one offspring. Thus both full-sib and half-sib families may be represented among the offspring. To avoid confusion between the effects of natural and artificial selection, the population is assumed to be free from mutation and differential viability.

Simulation of overlapping generations is extremely difficult because of the many ways in which it could be specified, and because of the many parameters which would be required. Haldane (1926) found expressions for the progress of slow selection in a Mendelian population where generations

overlap and noted that the changes were very similar to those which occur when generations are separate. Barker (1958a) suggests that overlapping generations have little effect on the comparison of results in simulated and experimental populations except for the relative lack of generation to generation variation in the simulated populations.

It is clear that one cannot easily simulate the number of chromosomes and loci thought to be involved in the development and expression of most quantitative traits among our economic organisms. Therefore, adoption of a plan which allows for as large a number of loci as can be expediently handled by the available computer is the only practicable solution. The repercussions which this may have on the results in general and on the dynamics of changing gene frequencies, in particular, must be steadfastly borne in mind. In the present study the largest adaptable number proved to be 40 loci because of the 40 binary "bit" structure of one computer "word" of memory. There is some argument for the simulation of a single autosomal chromosome with extrapolation of results to the plural case but the more flexible situation of eight autosomal chromosomes with five loci each has been used in this study. The alleles located on the first member of each pair of chromosomes are allocated to a single word of memory space. The corresponding alleles, located on the second member of each pair of chromosomes, are

allocated to another word of memory. Other concessions to expediency and machine eccentricity include the limitation of alleles to two per locus, uniform recombination frequency for adjacent loci on the same chromosome with no interference, equal genetic effects for all loci and the restriction of epistasis to sequential pair-wise interactions of loci.

The assumption of equal genetic effects at all loci is disturbing to some investigators. Any deviation from this, in a real situation, contributes somewhat to larger genetic variation in a population with intermediate gene frequencies. Perhaps later refinements of simulation procedure will make it possible to avoid this criticism. However, since the present simulation is already limited to forty loci it would seem that the main difference which would result if unequal effects were simulated would be the acceleration of fixation of genes with large effects, leaving an even smaller number of loci segregating and affecting progress. Therefore the base for inferring the significance of differences in later generations would be even narrower and more subject to sampling variation than the one proposed, which is already limited.

The restriction of epistasis to pair-wise interactions may or may not be serious depending upon one's point of view. Fisher (1918) pointed out that while more complex

connections could doubtless exist, the number of unknowns introduced by dual epistasis alone is more than can be determined by existing data. He expressed the opinion that it is very improbable that any statistical effect of a different nature is actually produced by more complex somatic connections. On the other hand Lush (1948) states: "Even if it be true that most genes interact nearly additively and thus have trifling or zero epistatic interactions, yet the number of possible interactions is so large that, even if only a tiny fraction of these are real, they could furnish a large total amount of epistatic variance." This argument leads to a reasonable hypothesis that those who study quantitative characters affected by many genes should expect to find epistasis more important than those who study only two pairs of genes at a time. The additional restriction of epistasis to interactions of sequential pairs of adjacent loci may be serious because such a procedure accounts for only a small fraction of all possible interacting pairs. If n loci exist, $n(n-1)/2$ interactions between two loci are possible. Simulation of $n/2$ sequential interactions accounts for only $1/(n-1)$ of the possible total of dual interactions. Regardless of the view taken, as to the severity of these restrictions, simulation of the more complex epistatic interactions awaits further refinement of programming technique.

The modes of gene action studied are listed in Table 26 as genotypic values for a pair of loci, along with the expected genotypic means and variances of the offspring created from an initial parent population which is completely heterozygous but with random association of coupling and repulsion linkage phases.

No particular reason exists for listing the genotypic values in a two-locus table for the non-epistatic models except that it is convenient for purposes of programming the computer. The particular coding of the genotypic values is purely arbitrary but pragmatical because whole integers are associated with an interacting pair of loci in all cases, but there is some semblance of comparison among the models because of equal or similar parameters in the initial generation. Kempthorne (1954) and Cockerham (1954) both provided partitions of total genotypic variance, under random mating, for a generalized gene model without linkage. Although Cockerham's procedure is limited to consideration of two alleles per locus, the mechanics of partitioning the variance are somewhat easier than they are with Kempthorne's method, which is general for any number of alleles. In this study, all models were restricted to two alleles per locus and the partitions of variance were accomplished by Cockerham's method. The relation of these partitions and the mean to gene frequency, for each of the nine models

used in this study are given by Figures 57-66 in Appendix A.

Horner et al. (1955) have discussed non-allelic gene interactions for the classical epistatic models and Horner (1956) listed the components of variance at several gene frequencies.

The epistatic models in this study involve pair-wise epistasis, i.e., interacting pairs of loci in sequence along eight chromosomes of five loci each, which results in 16 intra-chromosomal interactions and four inter-chromosomal interactions. These types of interaction are confounded with linkage when it exists but the procedure does offer a sample of the interactions possible with dual epistacy.

The optimum number model used is based on the equation

$$\text{yield} = 11 - (\text{number of plus alleles} - K)^2$$

where $K = 2$ for intermediate preferred. Wright (1935) first proposed this type of model. Lush (1948) hypothesized that an intermediate is especially likely to be optimum for populations which occupy an ecological niche where they are hemmed in on all sides by competitors better adapted to the adjoining conditions. Mather's emphasis (1943) on balance and on the role played by linkage rests on the assumption that the intermediate is favored. Under the optimum number model, few of the preferred individuals will be heterozygous at many loci, after several generations, because of selection and random drift, i.e., they will be AAbb or aaBB at

most interacting loci. Fisher (1930) emphasized that intermediate gene frequencies are likely to be unstable in this case. Therefore, not much epistatic variance from the intermediate being optimum can be expected in populations which have been under selection for a long time. When all gene frequencies are equal to .5 under the Hardy-Weinberg distribution, the epistatic variance component, which happens to be entirely additive-by-additive (σ_{AA}^2), makes up 2/3 of the total genotypic variance and dominance comprises the remainder (Figure 60). Thus, the average effect of a gene substitution is zero. If half of the genes are at frequencies of .9 and half are at frequencies of .1, as they are likely to be after many generations of selection, the epistatic component of variance is reduced to less than 1/3 of the total genotypic variance and additive variance measures more than half the total. For a given change in gene frequency, changes in the mean increase as gene frequency tends toward zero or one.

Duplicate genes were found early in genetics. In this model, substitution of either gene can produce the outward phenotypic change but does not if the other has already produced the effect. This is a simplified case of thresholds. At equilibrium gene frequencies of .5, epistatic variance (σ_I^2) contributes about 60 percent of the total genotypic variance (Figure 61). Additive (σ_A^2) and

dominance (σ_D^2) variance make up the remainder in the proportion 2:1. The epistatic variance consists of equal portions of additive by additive (σ_{AA}^2) and additive by dominance (σ_{AD}^2) variance of the same magnitude as σ_A^2 , and a small amount of dominance by dominance variance (σ_{DD}^2). If the mean frequency of favorable alleles moves only as far away as .7, total genotypic variance is drastically reduced to approximately 13 percent of the original, and most of this, about 85 percent, consists of epistatic variance. Consequently, one might postulate that random drift is likely to be a more powerful force than selection after a few generations. For a given change in gene frequency, changes in the mean become smaller as gene frequency tends toward unity.

Complementary genes were discovered when Bateson obtained purple-flowered sweet peas by crossing two races which each has white flowers. In general, genes are complementary when neither can produce an effect unless the other is present. Lush (1948) pointed out that epistasis is important in this case only when the dominant genes are rare. As gene frequencies change from .5 to .8 under selection, the proportions of σ_A^2 , σ_D^2 and σ_I^2 change from 4: 2: 1 to approximately 1:2:0 and the total variance is about 30 percent of the original (Figure 62). The epistatic variance is not zero as long as variation remains but it is negligible

for gene frequencies exceeding .75. For a given change in gene frequency, changes in the mean become slightly smaller as gene frequency tends toward unity. In this respect the model resembles the complete dominance case (Figure 58).

The additive-by-additive conditional epistatic model is characterized by having only σ_{AA}^2 at gene frequency of $p = .5$, but there is positive regression of offspring on parent under random mating. The average effect of a gene substitution in the initial population is zero. However, the proportion of additive variance increases rapidly (14 percent at $p = .6$, over 50 percent at $p = .75$, 90 percent at $p = .9$) as gene frequency increases, and σ_{AA}^2 decreases correspondingly (Figure 63). These figures are based on proper proportions for either p or $(1-p)$ since the model has two "peaks" of genetic merit. For a given change in gene frequency, changes in the mean become larger as gene frequency tends toward zero or one.

The additive-by-dominance conditional epistatic model contains only σ_{AD}^2 at gene frequency of $p = .5$. The average effect in the initial population is zero. The partitions of hereditary variance are not easily followed as gene frequency changes in this case because it is a multiple-peaked model and two of the three peaks are single-heterozygote combinations of the two-locus genotypic combinations (Table 26). Following the progress of gene frequency toward the highest,

or homozygous peak, one observes that the total genotypic variance increases rapidly up to $p = .85$ and the content rapidly changes from σ_{AD}^2 , first to increasing amounts of σ_{AA}^2 and σ_D^2 up to $p = .8$, and more and more to σ_A^2 beginning at $p = .6$ and accelerating rapidly thereafter. For a given change in gene frequency, changes in the mean increase as gene frequency tends toward unity (Figure 64). However, when the progress of gene frequency is toward either of the secondary peaks, total genotypic variance decreases somewhat with σ_{AD}^2 changing, primarily, to σ_D^2 , with only small amounts of σ_A^2 and σ_{AA}^2 coming in. In this situation, the frequency of one gene of an interacting pair must decrease twice as rapidly as the frequency of the second gene increases, in order to obtain even a slight increase in the mean (Figure 65). The net effect of selection should be a compromise of these two situations. Peculiar as it may seem after looking at the genotypic values of the model, no "valley" of genetic merit exists between the mean genotype at $p = .5$ and the homozygous peak. In fact, a contour map of means shows that the introduction of additive variance into the population requires less random drift in the direction of the homozygous peak than toward the secondary peaks.

The dominance-by-dominance conditional epistatic model contains only σ_{DD}^2 at gene frequency of $p = .5$ and never achieves any useful genetic variance, i.e., σ_A^2 or σ_{AA}^2 ,

regardless of the change in p (Figure 66). Following progress toward one of the four single-heterozygote peaks, one observes that the total variance remains constant, but its composition changes from σ_{DD}^2 , partly to σ_{AD}^2 , and later, partly to σ_D^2 . This is a model in which genetic drift should be the predominating force. Under this model, the mean of a completely heterozygous population cannot be improved.

Other models which may be studied in the future include multiplicative or geometric interactions, inhibitor genes, threshold effects, reversal genes or various other types of modifier genes.

In any genetic population the effective size of the population is related to the number of parents in some manner. For this reason parent population size, rather than number of offspring, should be the primary specification in a simulation study. With equal numbers of each sex and using a system of mating by random sampling of selected males and females with replacement, the levels of N used in this study 8, 12, 16 or 32 parents, should produce, in the absence of other forces, reductions in the panmictic index of 6.25%, 4.67%, 3.12% and 1.56%, respectively, per generation, due to the finite size of the population. Therefore these levels are equally spaced for the approximate inbreeding coefficient $F \approx 1/2 N$.

Selection intensity for a given run, or parameter set,

is constant from generation to generation for 30 generations, or until genetic fixation of alleles at all 40 loci occurs. In the present study, all selection is by upper truncation of phenotypes, although the simulation program is designed to handle lower truncation or intermediate selection also. The selection intensities of $1/2$, $1/4$, $1/6$ and $1/8$, when combined with the various parent population sizes, specify progeny populations ranging from 16 to 256 in number, and correspond to selected population means which are expected to be .8, 1.27, 1.5 and 1.65 standard deviations, respectively, above the mean of an unselected population which is conceptually infinite and normally distributed. The values are somewhat smaller but of similar proportions for small populations where lack of extremes produces some non-normality. Therefore one should expect differences between consecutive levels in the approximate ratio 6 : 3 : 2. The selection intensities are equal in both sexes. Animal breeders may wish to convert these intensities to unequal ones by using tables of z/b provided by Lush (1945) and others. For example, $1/6$ selection in each sex is approximately equivalent to saving 2 percent of the males and 60 percent of the females from large populations.

Considerable controversy exists about the subject of simulating environmental or extraneous variation. Succinctly, the two points of view are: (1) Allow for levels

of environmental variation relative to the expected genetic variance in the initial population and consider the differences in terms of relative magnitude of environmental effect or as broad heritabilities associated with the number of loci simulated, and (2) allow for variance due to the combined effect of the environment and those genes, not simulated, which may affect the quantitative character. Therefore, the differences are associated with the entire genome of one or more species of organisms. This procedure should simulate, more accurately, the real magnitude of average gene effects in relation to the total amount of phenotypic variability. A general criticism of the first point of view is that the number of loci simulated is likely far less than the number associated with most quantitative characters of economic importance. Consequently the average genetic effects of single genes in the simulated populations are large in proportion to the total variance, affecting the dynamics of changing gene frequency and resulting in more genetic fixation at any given generation in the selection scheme. Furthermore, one should not try to extrapolate conclusions concerning heritability, based on components of variance, to real systems that are surely based on considerably more loci. The second method of simulation, while approximating locus-by-locus dynamics of real organisms with higher probability, fails to allow for direct comparisons

between environmental levels as constitutional effects in the simulated population and therefore does not permit direct inferences concerning the magnitude of these effects in relation to the other factors in the study.

The key to choosing one of these methods appears to be associated with the primary objectives of the research. If one is chiefly interested in the intrinsic locus-by-locus effects upon changing populations, and is not much concerned with the effects of environment upon general population parameters such as the genotypic mean, the obvious choice is the second procedure. But if the principal goals include observation of general population parameters such as means and variances, which are based on the total effect of all loci, and the manner in which specific levels of environmental variance affect these parameters, including the relative magnitude of the effects among the levels of this factor and other factors as well, then it seems logical that one should choose the first method.

Because of the nature of the specific primary objectives of this study, and because it was necessary to incorporate specific levels of environment into a balanced factorial plan so that orthogonal comparisons could be made, method number one was used. However, in subsequent studies either method may be used, depending upon the objectives involved and the inferences made from this study concerning

the importance of simulating different levels of environment. The four levels of environment actually simulated in this study are 0, $\sigma_G^2/3$, σ_G^2 and $3\sigma_G^2$, where σ_G^2 is the expected genotypic variance in the initial generation of progeny. If one wishes to talk about heritability in terms of 40 loci, then these levels correspond to levels of heritability, in its broad sense, of 1.0, .75, .50 and .25, respectively.

To simulate linkage, recombination values of .005, .05, .2 and .5 were applied to the adjacent loci of the eight chromosomes of five loci each, with the frequency of crossover being uniform for all loci on the same chromosome for a given run. No interference was simulated. If one wishes to consider the end loci of two consecutive linearly arranged chromosomes as being also "adjacent", the mean recombination values of all "adjacent" loci would be approximately .06, .1, .25 and .5 instead of the levels listed above. From either point of view, the ratio of differences which one might expect if the effects of linkage are linear, is 1:3:6 for consecutive levels. However, it is doubtful if the real effects of linkage are ever linear over so broad a range. Some support for this view is indicated by reports of varying coefficients of coincidence in Drosophila melanogaster and other organisms (Bailey, 1961).

In consideration of the actual computer program, only those points of methodology and logic which seem potentially

useful to a wide variety of machines and simulation procedures will be discussed.

The machine which was used is a binary computer with 16,384 direct-access memory locations. Since many genetic mechanisms are also binary, some simplifications in logic were possible which might not be practicable in computers which, functionally, are non-binary. However, this advantage is partially offset by the minor complexity of working with a sexa-decimal number system.

The Monte Carlo method is so called because it is based on the simulation of stochastic processes. Although there are many different types of Monte Carlo analyses, they all include as a common feature the use of sets of random or pseudo-random numbers. Although the definition is subject to controversy, random numbers are probably truly random only conceptually, while pseudo-random numbers are those which are used in random process simulation, and which have fulfilled as many criteria of randomness as possible.

The multiplicative congruential method for generating uniformly distributed pseudo-random numbers on a digital computer has been discussed, tested and used by numerous persons since it was proposed by Lehmer in 1951. On a binary computer with a word size of P bits, the method most frequently takes the form $X_{i+1} = \lambda X_i \pmod{2^P}$, where λ is a fixed odd integer, and the starting number is taken from

the uniform distribution on the unit interval. The procedure was modified by Rotenberg (1960) and Greenberger (1961) to the form $X_{i+1} = (2^a+1)X_i+C$, with $a \geq 2$ and C odd. This yields the maximum period of 2^P , which is $2^{40} \approx 10^{12}$ in the computer which was used, and has the advantage that multiplication by λ can be accomplished by one shift order and one add order.

In the present study a was set equal to nine and $C = 1$, making $\lambda = 2^9 + 1 = 513$ and $X_{i+1} = 513X_i + 1$. Using an approximate formula, the serial correlation of the random numbers produced by this method is $\rho \approx .002$. The method has passed many tests for randomness, including distribution in equal intervals, mean and variance checks, runs up and down and runs above and below the mean. Results of these tests and others have been tabulated by Meyer (1956) and Rotenberg (1960).

To insure that phenotypes created from genotypes, conceptually distributed normally, by adding or subtracting values for good and poor environments, are also distributed normally, it is necessary to generate random normal deviates, or store tabulated values in the computer. Because the generation process is relatively simple, and because of the heavy demands on memory space in many computers, it seems more economical to generate the deviates instead of storing them. The general procedure used was to generate

seven uniformly distributed random numbers in the range $-1 < r_i < +1$, add them together and code the variance so that the sum is normally distributed with mean zero and variance equal to one. These numbers then are in the range $-7 < e_j < +7$. The variance of a uniformly distributed variable equals the square of the range divided by 12. Therefore $V(r_i) = (2)^2/12 = 1/3$. To find the constant needed to make e_j a $N(0,1)$ variable, one notes that $e_j = C \sum_{i=1}^7 r_i$ gives $V(e_j) = 7C^2/3$. Then if $V(e_j)$ is to equal 1, put $1 = 7C^2/3$ or $C^2 = 3/7$ and $C = \sqrt{3/7}$.

Perhaps an easier method is to make $e_j = \sum_{i=1}^{12} r_i$ giving $V(e_j) = 4$. Then $C = \frac{1}{2}$ and the coding can be accomplished by shifting each e_j one place to the right instead of by multiplication, which is slower by a factor of 20 in the computer which was used. Taking the sum of three uniformly distributed random numbers would achieve the unit variance without coding. Despite the fact that the range is only ± 3 , this distribution, which is based on three interlocking parabolas, is a surprisingly close approximation to the normal distribution.

Because of the length and complexity of the simulation procedure, the program was divided into several logical blocks or subroutines. The initial block consists of instructions which put in the numerical constants and particular values of the variables to be used in the program

and create a specified number of genotypes of each sex to be used as initial parents. The essential input associated with each model of gene action consists of the nine two-locus genotypic values, and expected genotypic standard deviation of the offspring of the initial parents (σ_G) and, if selection is to be generalized for intermediate selection as well as upper and lower truncation, the selection point or goal (C) from which deviations of individual phenotypes may be calculated. The generalized selection goal for upper truncation selection is $C = mG' + \sigma_G\sigma_{E'}$, where m is the number of pairs of loci, G' is the maximum two-locus genotypic value for a given model and $\sigma_{E'}$ is the maximum random normal deviate generated, in standard deviation units. The required input associated with each run, or parameter set, is the particular level of each factor being studied, in this case, number of parents, selection intensity, environmental variance and recombination value.

In the process of generating the initial parents, random numbers are used to choose 1 or 0 alleles of the first chromosome of each pair with equal probability. By taking the complement of this array of 40 alleles, the corresponding alleles of the second chromosome of each pair are generated so that they are alternate to those of the first chromosome at every locus, producing the desired, completely hetero-

zygous individual. In addition to having the initial gene frequency equal to .5 at every locus, the linkage association, between adjacent loci in coupling or repulsion phases, is random, not only with respect to location, but also with respect to epistatic interactions.

The second subroutine in the program is concerned with reproduction. After the initial generation, parents to be "mated" are chosen randomly from the selected group of each sex and the processes of segregation, recombination and fertilization are simulated by a random mask method which produces a gamete from each parent. This method was first developed by Schweppe and Bohidar at Iowa State University, and was used by Bohidar (1960). This method is faster than the random transform of a vector of genetic recombination frequencies, used by Fraser (1957a) for a small number of loci, and very much faster than the random walk method which he proposed for linkage of large numbers of loci (Fraser, 1960a).

A mask is merely a collection of 1's and 0's constructed from specified probabilities of the linkage relations between loci, i.e., a 1 or 0 indicates whether the first or second chromosome of a pair shall provide the allele from the locus corresponding to that binary bit. Consequently, when tight linkage is desired, i.e., high probability of no crossover between adjacent loci on the same chromosome,

the mask will consist chiefly of sequences of 1's or 0's, with an occasional alternation indicating either a crossover or a change from one chromosome to the next. It is necessary to decide at random from which chromosome of a pair the allele from the first locus of each chromosome shall come. Otherwise, gametes with a preponderance of alleles from one of the chromosomes of each pair would always result, violating Mendel's first law concerning segregation. Whether the remaining alleles come from that chromosome or its partner depends upon the freedom for crossovers to occur, i.e., the specified recombination frequencies.

Logical algebra is used to procure the alleles from the parent as indicated by the mask. As an example, consider the simple genotype 1101/1010. Assume that a random mask has been produced from specified linkage relations and turns out to be 1001. The logical product of the mask and the first, or paternal half of the genotype is $(1101)(1001) = (1001)$, and the logical product of the complement of the mask and the second, or maternal half of the genotype is $(1010)(0110) = (0010)$. Adding the two logical products gives $(1001)+(0010) = (1011)$, which is the same gamete one would obtain by picking out the alleles indicated by the mask on sight, i.e., first and fourth alleles from the paternal half and second and third alleles from the maternal half of the

genotype. Then gametes are produced alternately from male and female parents and successive pairs of gametes are stored together as genotypes of offspring, completing the fertilization simulation.

The third subroutine is entitled genotypic evaluation. The purpose of this block of the program is to calculate numerical values of the genotypes of the offspring created in the reproduction subroutine by identifying the genotype at each locus and applying the values given in Table 26. Because of the dual epistasis involved, the genotypes must be evaluated by pairs of loci, each individual's genotypic value consisting of the sum of 20 two-locus values from the array of nine possible genotypes for each pair. Then the genotypic mean and variance of the population may be computed. Again, logical algebra is helpful in identifying the genotypes. Using the simple genotype 011/001 for illustrative purposes, the identification of the homozygous 1, heterozygous and homozygous 0 allelic states can be provided by the logical product, logical equivalent, and logical not-sum, respectively, of the paternal and maternal halves of the genotype. As shown by Fraser (1957a), the three results are 001, 010 and 100, respectively. For computers which have only the logical product instruction, the logical not-sum can be obtained by taking the logical product of the complements of the genotypic halves and the

logical equivalent can be implied by exclusion. In addition, the major features of this subroutine can be used to compute genotypic or gene frequencies among the selected or unselected populations.

The fourth subroutine is concerned with phenotypic evaluation. In this part of the program, random normal deviates, which simulate environmental values, are added to the genotypic values to create phenotypes. The standard equation $P = G + E$ is actually computed as $P = G + \sigma_G R d_j$, where σ_G is the expected genotypic standard deviation among the initial generation of offspring, d_j is a $N(0,1)$ random deviate and R is the square root of the desired ratio of environmental variation to genotypic variation, σ_E^2/σ_G^2 . The phenotypic mean and variance are then computed from the P values.

The fifth subroutine concerns selection. The process in this study involves upper truncation of a given fraction, $b = n/N$, of the population of phenotypes. For generality, unsigned deviations from the selection goal are ordered by ascending magnitude. However, the procedure is complicated because the values being ordered cannot be used directly in the reproduction simulation of the next generation. Therefore, the genetic identity of each phenotype must be retained in order to simulate production of gametes from the corresponding genotype. If n of N individuals are to be selected,

only the first n random values are ordered, then succeeding ones are placed in their proper ranking among these n and the individual which ranks $(n + 1)$ is dropped each time.

The last subroutine of the genetic simulation concerns evaluation of the individuals selected to be parents of the next generation. Its function includes computation of genotypic and phenotypic mean and variance and relocation of the genotypes of individuals which are to serve as parents of the next generation.

The remainder of the program contains computer laboratory library subroutines such as floating point arithmetic and floating point output. The basic output of this study consists of the genotypic and phenotypic means and variances of both selected and unselected populations, genotypic and gene frequencies of the selected population and the genotypes of the initial parents. In addition, the last random uniform number and last random normal deviate used in each run, or parameter set, is recorded in case a re-run of the following set from the same random start is desired because of loss of data or for other reasons. Experience has indicated that it would be advantageous to decide exactly which summary statistics of the basic output are needed and to incorporate that procedure directly into the simulation program as much as possible.

DISCUSSION OF RESULTS

Analysis of Genotypic Means

The progress of the genotypic mean of the unselected population was recorded for each of 16 runs, or parameter sets, associated with each of the nine models of gene action. The content of each of the parameter sets was derived from the orthogonal arrays of a 1/16 fractional replication of a 4^4 factorial plan. The particular fraction used was selected at random and is associated with the confounded defining contrasts $AB^2C = BCD^2 = AB^3D^2 = AC^3D^3 = ABC^2D$, where A, B, C and D are the four factors. Kempthorne (1952) illustrated the way in which each main effect can be partitioned into three orthogonal comparisons, each with one degree of freedom. These are $A' = a_3 + a_2 - a_1 - a_0$, $A'' = a_3 - a_2 - a_1 + a_0$ and $A''' = a_3 - a_2 + a_1 - a_0$, where a_i represents the effect of the i th level of factor A. These comparisons may be related to linear, quadratic and cubic effects of A by the relations $A_L = 3a_3 + a_2 - a_1 - 3a_0 = 2A' + A'''$, $A_Q = a_3 - a_2 - a_1 + a_0 = A''$ and $A_C = a_3 - 3a_2 + 3a_1 - a_0 = A' + 2A'''$.

Each factorial effect is confounded with five aliases, its generalized interactions with the defining contrasts. The main effects are estimable if their aliases, which are all interactions, are assumed to be negligible. Comparisons among the main effects are still orthogonal because each

level of a factor occurs the same number of times, once, in this plan, with every level of the other factors, in some parameter set of the design. The two-factor interactions are too intertwined to be estimable.

Because the fraction of the full replication was selected at random and the designations A, B, C and D were applied to the factors studied without foreknowledge of the particular aliases involved, one need not hedge in discussing the possibilities of being misled by the assumption of negligible interactions.

In the plan which was used, most of the aliases of main effects are three-factor and four-factor interactions. Evidence for the existence of these interactions, involving the factors as assigned, A = environmental variance, B = population size, C = selection intensity and D = linkage, is either lacking or unconfirmed.

Main effects which have two-factor interactions for aliases are $A = BC^2 = BD = CD^2$, $B = AC^2 = CD^2$ and $C = AB^2 = BD^2$. The main effect of D, linkage, is not confounded with any two-factor interactions. The effect of environmental variation is confounded with the interactions of population size with selection and with linkage, and also with the interaction of selection and linkage. The effect of population size is confounded with the interactions of selection intensity with environmental variation and with

linkage. The effect of selection intensity is confounded with the interactions of population size with environmental variation and with linkage.

Direct evidence is lacking for the existence of the interactions of linkage with population size or selection intensity. The same is true of the interaction between population size and environment. However, the existence of the interaction between selection and environment, and the one between selection and population size, both have been inferred from various selection experiments. These interactions are important in relation to inferences about the main effects of population size and environment, respectively.

The different gene models which were studied were not incorporated into the factorial plan. Therefore, inferences concerning differences in main effects among models are founded on a firmer basis than those concerning differences within a particular model.

In an investigation such as this one, which is primarily of the screening type designed to pick out the most important factors, the conclusion may be drawn that some factors are important but others are relatively unimportant. If further experimentation is confined to the factors considered to be important and interactions are estimated, little or no harm has been done by the assumption of negligible interactions in the first study. However, if the investiga-

tion is part of a program of fundamental research and the investigator wrongly concludes either that factor has a beneficial effect or that it has a deleterious effect, in the absence of further experiments, energy may be dissipated in creating incorrect theories to explain the inferred effect or in disputes with colleagues.

In view of these reservations, conclusions drawn in the present discussion are subject to some misinterpretation especially with regard to the main effects of population size and environmental variation, which are confounded with interactions which may not be negligible. Therefore, the conclusions necessarily must be treated with some caution.

The fractional replication of the design which was used involves comparisons of the parameter sets listed in Table 25, Appendix B. The order of the sets has been changed from the original plan for A, B, C and D in order to clarify presentation.

The progress of the genotypic mean of the unselected population was recorded in each generation of a given run or set. Each run was continued for 30 generations or until complete fixation of all alleles in the population was achieved, if that occurred first. The formal analysis of the results in each generation of each set for every model would be costly and time consuming, and would produce an unwieldy and unneeded amount of data. The solution which

was chosen was that of averaging the results over each five consecutive generations and testing these differences. This amounts to testing the differences in linear regression of genotypic mean on generation number over five-generation intervals. The injustice done to the data by assuming linearity of the results seems negligible when offset by the advantage of elimination of some degree of random sampling from generation to generation, by the shortness of the interval and because of the inconclusiveness of results taken from any single generation.

The differences between levels could be tested by using the mean square for pooled interactions, with three degrees of freedom, for an error term. However another method gives more precision to the tests and, at the same time, allows a test to be made of the significance of the mean square for pooled interactions. This involves repeating the basic 16 parameter sets using different random starts in the computer. One may question the procedure of using a factorial plan fractionated so highly as 1/16 and then repeating it instead of using a larger fraction. However, the repeat runs were much faster and less expensive than the original ones because the program was reduced to the computation and output of only the genotypic mean of the unselected population. Consequently, the total resources used could not have produced a larger fraction of the same plan.

Let F_{ij} = the genotypic mean over five generations in the i th replicate of the four populations with the j th level of factor F , where $i = 1, 2$ and $j = 1, 2, 3, 4$. Also G_{ij} = the genotypic mean over five generations in the i th replicate of the population with the j th parameter set, where $i = 1, 2$ and $j = 1, 2, \dots, 16$. Then the sum of squares for levels of factor F equals $\frac{1}{2} \sum_{j=1}^4 F_{.j}^2 - (F_{..}^2/8)$, where the replacement of subscripts by periods indicates that summation over the range of the subscripts has been made. If the sums of squares for the four factors involved are called A, B, C and D , then the sum of squares for interactions is $I = \frac{1}{2} \sum_{j=1}^{16} G_{.j}^2 -$

$$(G_{..}^2/32) - (A+B+C+D). \text{ The sum of squares for error is } E = \sum_{i=1}^2 \sum_{j=1}^{16} G_{ij}^2 - \frac{1}{2} \sum_{j=1}^{16} G_{.j}^2 \text{ and the total sum of squares is } T = \sum_{i=1}^2 \sum_{j=1}^{16} G_{ij}^2 - (G_{..}^2/32).$$

The analysis of variance is straightforward and allows for equal precision in the tests of all factors. Three degrees of freedom are associated with the sum of squares for each factor and for pooled interactions, and 16 degrees of freedom are associated with the sum of squares for error derived from replication.

In order to determine which differences between levels of a particular factor were responsible for a statistically significant mean square, the multiple range test described

by Duncan (1955) was employed, using the critical values as modified by Harter (1960). This procedure employs special protection levels based on degrees of freedom which assure high probability of accepting the significance of mean differences when, in fact, the differences are real. This probability is slightly less than that associated with some other sequential testing procedures but the corresponding protection against making type I errors is greater. The necessity of limiting the use of the multiple range test to cases where the F-test among all means is significant has not been definitely established but it is accepted as standard procedure by most investigators. If the tabulated critical value is denoted by $Z_{n,p}$, where n equals the number of degrees of freedom associated with the error mean square, $\hat{\sigma}_E^2$, and p equals the number of means included in the range of magnitude of the two means being compared, F_{ij} and F_{ij}' , a difference, $F_{ij} - F_{ij}'$, is considered to be significant at the five percent level of probability when it exceeds $\hat{\sigma}_E Z_{n,p}$.

Figures 1-4, in Appendix A, illustrate the statistical significance of differences among means of simulated populations which are due to the different levels of population size (N), selection intensity (S), environmental variation (E) and linkage (L). In general, the test of the mean square for pooled interactions was statistically significant at the

probability level $\alpha = .01$. Therefore, the results for interactions are not shown in Figures 1-4. No illustration is given for the additive-by-dominance conditional epistatic model because no differences were statistically significant even at the probability level $\alpha = .25$.

Figures 5-40 show the mean genetic progress of the progeny populations for the four replicated parameter sets which have the same level of any factor in common for each of the nine models of gene action.

Each of the four factors, as it relates to the progress of the genotypic mean, will be considered in turn.

Population size

The most significant relationships concerning population size apply only to the breeding population and not to the total number of individuals in a group. Wright (1931) and others have discussed circumstances which make the effective number even smaller than the number of parents. If the population size fluctuates greatly from one generation to another, the effective number is much closer to the minimum number than to the maximum. If a difference in numbers of breeding males and females exists, the effective number is closer to that of the less numerous sex. The case when population numbers remain constant and the number of parents of each sex is the same has been simulated in this study. In such a population two presumably random

samples of gametes are produced and these are united at random to form zygotes. In this respect, the simulated populations may differ from real populations in which the conditions of random sampling of gametes are probably never met exactly. Also, in real populations, the number of surviving offspring left by different parents may vary considerably because of natural selection. Lush (1946) expressed the opinion that this is the major cause for chance deviations in gene frequency being so much larger than indicated by finite size of the population.

Theoretical effect of inbreeding Restriction of the size of a breeding population has an effect which Crow (1954) has described in three ways:

- (1) Inbreeding effect; a decrease in the average proportion of heterozygous loci.
- (2) Random extinction; Fisher's appellation for random fixation and loss of alleles.
- (3) Increase in the average variance of the distribution of gene frequencies.

Inbreeding is the mating of individuals which are related to each other more closely than the average relationship within the population. The primary effect of inbreeding is to make more of the loci homozygous. The rate of decrease of heterozygosis in systems of mating more complicated than self-fertilization was first worked out from the

recurrence relation between successive generations by Jennings (1914) and other workers. Wright (1921a) generalized formulas that would express the average consequences of inbreeding for any defined mating system in terms of gene frequencies of .5 and later (Wright, 1922) proved that these formulas applied to any gene frequencies. In this latter paper he proposed the quantity F as an inbreeding coefficient giving "the departure from the amount of homozygosis under random mating toward complete homozygosis." The inbreeding coefficient does not include any effects due to selection and, according to Lush (1948), F expresses only the most probable result because of Mendelian sampling variations. Later Wright (1931) extended the concept to include inbreeding due to the finite size of a population and showed the relation of this to the random drift in gene frequency.

Bartlett and Haldane (1934) extended the scope of Wright's method by using matrix algebra. This work has been further developed by Fisher (1949) and was presented by Kempthorne (1957) as the generation matrix theory of inbreeding. The method is more descriptive of the population than Wright's F but requires matrices with large numbers of elements for some systems of mating, whereas the inbreeding coefficient alone may be obtained more readily by the method of path coefficients. Malécot (1948) has shown how the

general formula for F_t , given by the method of path coefficients, can also be demonstrated directly from the theory of probability.

Coefficients of inbreeding other than F have been suggested. Bernstein (1930) suggested a coefficient, α , to describe the departure from panmixis. It is currently used by some investigators in human genetics and is identical with F . Fisher (1949) proposed an absolute measure, λ , of the amount of progress made in inbreeding, which is the same as $(1 - F)$ for a steady rate of decline in heterozygosity.

Wright (1931) has shown that heterozygosity decreases by a proportion which is approximately $1/2N$ each generation, where N is the total size of the breeding population. The amount of heterozygosity is, therefore, proportional to $P = (1 - F)$ for inbred populations derived from random mating populations. Wright (1951) has called F the fixation index and P the panmictic index. For the case of random mating with effective population size equal to N , it may be shown that $F_t = 1/2N + (1 - 1/2N)F_{t-1}$. In terms of the panmictic index $P_t = (1 - 1/2N)P_{t-1}$. After t generations the panmictic index would become approximately $P_t = P_0 e^{-t/2N}$ as given by Malécot (1948). He pointed out that, in order for P to decrease to a fraction, X , of its value in the original generation, the number of generations required is the solution of $e^{-t/2N} = X$, or $-t = 2N \ln X$.

In the simulated population used in this study, $P_1 = 1$, i.e., the first generation of offspring have the Hardy-Weinberg distribution at each of the 40 loci. Suppose one considers the time required for random drift to decrease heterozygosity until only two heterozygous loci are segregating in the average individual. This represents $1/20$ of P_1 . Therefore $-t = 2N \ln(.05)$ or $t = 6N$. Consequently, in the simulated population of size eight, random drift alone would lead to the case specified in approximately 48 generations and to complete fixation shortly thereafter. Even in 20 generations, only 27 percent of the original heterozygosity would remain, corresponding to an average of almost 11 heterozygous loci still segregating. However, for the largest population simulated, 32, the corresponding figures are 72 percent and 29 loci. Obviously, in a population under selection for a homozygous maximum the changes would be even more rapid.

The same effect may be described in terms of random extinction, i.e., fixation and loss of alleles or "decay of variance." According to Fisher (1930) and Wright (1931), in a population without mutation or selection, the gene frequencies scatter until all frequencies become equally likely in a few generations. Then, as $1/2N$ of the genes drift into fixation or loss each generation, the proportion of unfixed loci falls off by the same amount so that after

t generations the number remaining unfixed is given by $L_t = L_0 e^{-t/2N}$, which is the same as the formula for decrease in heterozygosity. Crow (1954) pointed out that this formula also describes the rate of decay in variance since the intra-population variance in gene frequency is also proportional to the amount of heterozygosity.

Relation of gene frequency to the mean The random drift of gene frequency or inbreeding effect due to the finite size of the population is related to changes in the mean of the population. In the notation of Lush (1948), the average phenotypic values for genotypes AA, Aa and aa are $y + 2x$, $y + hx$ and y , respectively. The corresponding frequencies are $q - p/2$, p and $1 - q - p/2$, where q is the gene frequency and p is the fraction of heterozygotes. With one pair of genes, the mean of this system is $2q = p(h-1)$. Under inbreeding p takes the form $2q(1-q)(1-F)$ and the mean $M = 2q + 2q(1-q)(1-F)(h-1)$, where $h=1$ with no dominance, $h=2$ for complete dominance and $h > 2$ for overdominance. The change in the mean due to inbreeding can be described by $dM/dF = -(h-1)2q(1-q)$. From this one may note that the phenomenon known as inbreeding depression is dependent on the degree of dominance, being zero for the case of no dominance and possibly quite large for extreme overdominance.

Lush (1948) pointed out that inbreeding will also decrease the mean when interacting genes are dominant or

intermediate in their epistatic effects. The former case includes the complementary and duplicate types of gene action used in this study and the latter includes the optimum number and dominance-by-dominance models. The additive-by-dominance model is multiple "peaked" and includes both cases. However, the additive-by-additive model includes a dominant peak and a recessive peak, but no dominance deviations of any kind are involved, so the inbreeding effect would be expected to be zero.

Wright (1951) expressed the relation of the mean under inbreeding (M_F) to the mean under random mating (M_0) and the mean of lines completely fixed without selection (M_1) by the equation $M_F = M_0 + F(M_1 - M_0)$. This formula is appropriate only for populations without epistasis.

Kempthorne (1957) expressed the mean of an inbred population, μ_I , in terms of the mean of the original random mating population, μ_R , and the average dominance deviations in that population, by the equation $\mu_I = \mu_R + FD_1 + F^2D_2 + F^3D_3 + \dots$, where the subscripts refer to the number of loci involved in a particular type of dominance deviation. If there are no epistatic deviations which involve only dominance, i.e., no D_2 , D_3 , etc., the mean of the inbred population is linearly related to F in spite of the existence of other epistatic deviations. This linear relation should hold for all but three of the models used in this

study. Those are the duplicate factor, complementary factor and dominance-by-dominance epistatic models. In the first two of these models, only trivial amounts of σ_{DD}^2 exist at gene frequencies higher than .7 (Figures 61 and 62). In the third model, $D \times D$, σ_{DD}^2 decreases from 100 percent of the total variance at gene frequency of .5 to less than 50 percent at .8 and to 13 percent at .9, when one gene of an interacting pair increases to those frequencies and the other frequency remains constant (Figure 66). If both genes increase or decrease in frequency, the loss of σ_{DD}^2 is even more rapid.

Figure 13 shows that the mean of populations under the $D \times D$ model, which are as large as 32, is depressed little over 30 generations in comparison to the mean of smaller ones, where random drift can result in more drastic changes in gene frequency. These changes can result in rapid fixation of one gene of an interacting pair so that the remaining variance is due to segregating alleles at the other locus, which are then acting in an overdominant manner. Thus, in this situation, inbreeding depression is large but it is related to D_1 deviations and not to the epistatic ones (D_2). However, one may conjecture that situations exist in which the contributions of D_2 , D_3 , etc., to the change in the mean upon inbreeding may be negative or positive, and the magnitude of the change possibly could be rather large,

in some cases, in proportion to the fraction of the total genotypic variance accounted for by those deviations.

The variance, of course, is also affected by inbreeding. The relationship $\sigma_{AF}^2 = (1+F)\sigma_{AR}^2$ has been used to compare the additive variance of an inbred population to that of a population mated at random. Kempthorne (1957) has pointed out that this formula is accurate only when no dominance exists or when gene frequency equals .5, both very special cases. Consequently, the use of the equation to adjust the additive variance of an inbred population is questionable. Robertson (1952) has shown that the variance within lines will increase in the early stages of inbreeding if dominant genes are at relatively high frequency.

Unfortunately, the relation of gene frequency to the mean is not simple when selection is operating on a population of finite size. Chance variations are random in the direction in which they change gene frequency, but constant selection causes gene frequency to change steadily in one direction or toward an equilibrium. Robertson (1961) has shown that one may expect the inbreeding under individual selection to be greater than that calculated from the actual number of parents when both the intensity of selection and the heritability of the character are high. He gave the ratio of actual numbers to effective numbers as $1+(4i^2\sigma_b^2)/(\sigma_b^2 + \sigma_w^2)$, where i^2 is the squared selection differential,

σ_b^2 is the variance between families and σ_w^2 is the variance within families.

The strength of random drift largely depends on the size of the population, whereas the strength of selection does not. However, certain selection dynamics are affected by the number of loci involved. Because random drift and selection frequently are opposing forces, the observed changes of mean in the small simulated populations, with only 40 loci, may more closely resemble actual changes in slightly larger populations in which more loci are available to the same selection pressure.

The rate of change of gene frequencies under selection has been dealt with primarily by Haldane (1923, 1926), Fisher (1930) and Wright (1931). Wright has been especially concerned with the steady state distribution of gene frequencies. In the 1931 paper he concluded that, in small natural populations, there is nearly complete fixation, little variation, little effect of selection and thus a static condition, whereas in slightly larger populations there is continual random shifting of gene frequencies but relatively rapid, continuing changes due to selection, even under static conditions. Fisher (1930) developed what he called "the fundamental theorem of natural selection," that the rate of increase in fitness of any organism at any time is equal to its genetic variance in fitness at that time.

The basic tenets of this theorem have been adapted to artificial selection theory by some investigators. Later, Fisher (1941) defined a quantity, $\lambda = Q^2/Pr$, which measures the extent of departure from random mating genotypic frequencies of $P(AA)$, $2Q(Aa)$ and $R(aa)$. He showed that λ must remain constant for the rate of change in fitness to be equal to the genetic variance, i.e., for changes in genotypic frequencies to be ascribed to the changes in gene frequency.

Crow and Kimura (1955) pointed out that a difficulty in the utility of λ is that, except for random mating ($\lambda=1$), λ is not invariant with changes in gene frequency. In this respect Wright's F is more useful for consanguineous mating systems because it is a function of the mating system and independent of gene frequencies. However, λ is more general because it may be used for any mating system.

Extension of Fisher's concept to more than two alleles has been made by Kempthorne (1957) for general quantitative selection with no variation in fitness. He also showed that the change in the mean could be expressed for changing values of λ as

$$\Delta\mu = 2\sum_k \Delta p_k \alpha_k + \frac{1}{2} \sum_{\substack{ij \\ i \neq j}} p_{ij} \frac{\Delta\lambda_{ij}}{\lambda_{ij}} d_{ij}, \text{ where the } p_k \text{ are}$$

gene frequencies, the α_k are average gene effects described by Fisher (1941), the p_{ij} are genotypic frequencies and

\bar{d}_{ij} is $y_{ij} - \alpha_i - \alpha_j$, the dominance deviation attached to the genotype A_iA_j which has the value y_{ij} . For the two-allele case the formula is simply $\Delta\mu = 2(\Delta p \alpha_1 + \Delta q \alpha_2) + Q\Delta\lambda d_{12}/\lambda$. In general the change in the population mean consists of the first part, attributable to changes in gene frequency, and the second part attributable to the mating system. Obviously, if there is no dominance the mating system does not matter. Also if the changed population is obtained by selection in a random mating population followed by random mating of selected individuals, then $\Delta\mu = 2(\Delta p \alpha_1 + \Delta q \alpha_2)$ because λ will not change. Unfortunately, this is true only for conceptually infinite populations. In random mating finite populations with selection, such as those simulated in this study, the change in the mean due to n loci segregating, assuming equal gene effects for all loci and no interlocus interactions, becomes $\Delta\mu = 2 \sum_{i=1}^n (\Delta p_i \alpha_i + \Delta q_i \alpha_i) + \sum_{i=1}^n Q_i \Delta \lambda_i d_i^2 / \lambda_i$. Translating this into terms of the genotypic values used in this study, the formula, when there is complete dominance, becomes

$$\Delta\mu_{n,n} = 4 \sum_{i=1}^n \Delta p_i q_i + 2 \sum_{i=1}^n p_i q_i P_i R_i \left[\frac{Q_i^2}{P_i R_i} - \frac{Q_i^2}{P_i R_i} \right] / Q_i$$

where p_i and q_i are allelic frequencies, P_i , $2Q_i$ and R_i are

genotypic frequencies and subscripts with primes refer to generation $n+1$.

The formula for the overdominance case is quite similar.

$$\Delta\mu_{n,n'} = 4 \sum_{i=1}^n [\Delta p_i (p_i - q_i) + P_i R_i [Q_i (1 - Q_i) + P_i R_i] \left[\frac{Q_i^2}{P_i R_i} - \frac{Q_i^2}{P_i R_i} / Q_i \right]]$$

Kempthorne (1957) has translated the formula into terms of fitness as described by Fisher (1930). In this case, $\Delta v = \sigma_A^2(f) + \frac{1}{2} \sum_{i \neq j} \sum_j P_{ij} \Delta \lambda_{ij} d_{ij}(f) / \lambda_{ij}$. Kimura (1958) has developed a simpler proof based on the Malthusian parameter, a , and coefficients of departure from random mating. His coefficient of departure involving two loci is $\theta_{ijkl} = P_{ijkl} / P_{ij} P_{kl}$, where P_{ijkl} is the frequency of the genotype $A_i A_j B_k B_l$, P_{ij} is the frequency of the genotype $A_i A_j$ without regard to locus B (i.e., $P_{ij} = \sum_k \sum_l P_{ijkl}$), and P_{kl} is similarly defined for $B_k B_l$. The coefficients for the individual loci are $\theta_{ij} = P_{ij} / X_i X_j$ and $\theta_{kl} = P_{kl} / Y_k Y_l$, where the X's and Y's are allelic frequencies. The epistatic effect ϵ_{ijkl} can be further subdivided into interaction components, each with its own coefficient of departure from random combination. For example the interaction of the dominance effect at locus A with the additive effect at locus A with the additive effect at locus B, $(d\beta)_{ijk}$, has

the coefficient $\theta_{ijk} = P_{ijk}/P_{ij}Y_k$. Kimura then gives the rate of increase in population fitness for an arbitrary number of loci as

$$\frac{d\bar{a}}{dt} = \sigma_A^2 + \overline{\frac{da}{dt}} + \overline{\sum \epsilon \frac{d}{dt} \log \theta},$$

where ϵ stands for a dominance or epistatic interaction effect, θ is the corresponding coefficient of departure from random combination, \underline{a} stands for the Malthusian parameter, which includes the effect of overpopulation, genetic variance and deterioration of environment, and the overbars indicate mean values. The summation extends over all possible intra- and inter-locus interactions.

Kimura (1958) then considers the applications of his formula for increase in population fitness to artificial selection. In notation adapted to dual epistasis, and ignoring terms for improvement of environment, the approximate result is

$$\Delta\bar{Y} = h^2 I + \left(\frac{\Delta F}{1-F} \sum_{m=1}^n p_m d_{mm} - F \sum_{m=1}^n \Delta p_m d_{mm} \right) +$$

$$\sum_{m'} \left(\sum_{ijkl} \epsilon_{ijklm'} \Delta p_{ijklm'} \right),$$

where $h^2 = \sigma_A^2 / \sigma_P^2$, or heritability in the narrow sense,

I = the selection differential,

F = Wright's coefficient of inbreeding,

p_m = gene frequency at locus m ,

d_{mm} = the dominance deviation at locus m ,

$\epsilon_{ijklm'}$ = the epistatic effect of a pair of loci denoted by m'

and

$P_{ijklm'}$ = the frequency of the genotype $A_iA_jB_kB_l$ at the pair of loci denoted by m' .

Thus, the rate of increase in the mean is expressed in a sum of three terms:

- (1) A term due to additive genetic variance, which is determined by summation of the product of changes in gene frequency and the additive effects at those loci (this term is usually referred to as predicted response to selection in a random mating, conceptually infinite population).
- (2) A term due to the effects of dominance in a non-random mating population, e.g., inbreeding depression due to finite size of population.
- (3) A term due to the effects of interlocus interactions involving only two loci each, i.e., dual epistasis, which is general for linkage disequilibrium.

This formula by Kimura is possibly the most comprehensive yet devised to describe the relation between gene frequency and the mean of a population of finite size which is being selected in the presence of linkage, dominance and epistasis. In order to assess the merits of this formulation one must know the changes in gene and genotypic

frequencies over time, the additive gene effects at all loci, the dominance and epistatic deviations for all types of intra- and inter-locus interaction, the fraction selected, the total variance in the population, the phenotypic reach or selection differential and the inbreeding coefficient, including its change with time. A point of concern is the fact that the change in the frequency of the ordered genotype $A_iA_jB_kB_l$, i.e., ΔP_{ijkl} , is expressible in terms of $\Delta \log \theta_{ijkl}$, where $\theta_{ijkl} = P_{ijkl}/P_{ij}P_{kl}$. It is easily seen that the θ for many loci will be undefined in a population of finite size under selection because of incomplete distribution of genotypes. Consequently, the effect of θ over time cannot be assessed for many segregating loci even though its value may be known at one end of the interval. Thus Kimura's coefficients of departure from random mating suffer from the same inadequacies in finite populations as does Fisher's λ , which Kempthorne (1957) incorporated into his fundamental equation of selection theory. The chief difference is that the limitation applies only to the epistatic term of Kimura's equation while not affecting the term due to dominance. However, Kimura's dominance term is not as general as Kempthorne's because it is useful only for random or consanguineous mating systems.

Monte Carlo differences in means In considering differences in the observed means of the simulated populations

in this study which are due to population size, one may readily see from Figures 1-4, Appendix A, that random drift is more important, in general, in small populations which are affected by dominance than in those which are not affected, i.e., the differences among population sizes of 8, 12, 16 and 32 are especially noticeable with the overdominance and dominance-by-dominance (DxD) models, and to a lesser degree with complete dominance and complementary factor models. The only other case in which the effects differ with any probability of consequence is with the additive model.

Kojima (1961) has shown that size of population does not cause a serious difference between the gain from selection expected in an infinite population using the usual prediction equation, selection differential times heritability, and the expected gain from his formula for a small population, except when dominance exists. He pointed out that the joint effects of the finite size of population and dominance gene action could amount to a considerable bias in the usual prediction equation, and that such a bias can be largely accounted for by inbreeding depression.

Consequently, one may ascribe a sizeable portion of the observed significant differences to inbreeding depression in all of the simulation models except the additive one. The effect is stronger, as expected, when overdominance is

present. In that case, the observed differences were significant at $\alpha < .005$ over the entire interval of 30 generations.

The differences in means of populations under complementary and DxD gene action (Figures 2 and 4) did not reach statistical significance for about five generations, but they attained high levels of significance thereafter. The probable change in percentage components of genetic variance (Figures 62 and 66) accompanying the expected direction of change in gene frequencies due to selection, shows that the proportion of total variance which is accounted for by dominance increases slowly at first and then more rapidly as gene frequency changes in these models. The same phenomenon is true for the complete dominance model except that the percentage of dominance variance increases more slowly as gene frequency changes (Figure 58). This is reflected in the probabilities shown in Figure 1, where the significance of differences due to population size did not reach .05 until 15 generations of selection were completed.

The statistically significant differences among means observed in different sizes of population under the additive model after generation 15 (Figure 1) cannot, of course, be attributed to inbreeding depression resulting from dominance. These differences are reflections of the decrease in additive variance due to fixation of some loci. Therefore,

the differences must be expressed in terms of the rate of approach to homozygosity, or random drift, as a force opposing selection.

Figures 5 - 13, Appendix A, illustrate the mean genetic progress of populations under various models of gene action for the four replicated parameter sets which have the same population size in common. None of the differences observed in populations with optimum number, duplicate factor, additive-by-additive (AxA) or additive-by-dominance (AxD) types of gene action (Figures 8, 9, 11 and 12, respectively), even remotely approached statistical significance, i.e., none were significant at $\alpha < .25$. The additive-by-additive model (AxA) does not contain dominance variance of any kind at any level of gene frequency. Consequently, inbreeding depression is not a factor. In the additive-by-additive and optimum models selection is effective at the higher gene frequencies, i.e., both have high proportions of additive-by-additive variance, σ_{AA}^2 , at gene frequencies near .5, but have less of this and larger proportions of additive variance for other gene frequencies (Figures 60 and 63). The change in components of variance is noticeable for gene frequencies only slightly removed (.05) from .5 for the optimum number case, allowing for effective selection even though the favored genotypes are combinations of alleles which average .5 in frequency. In

the AxA case, the increase in σ_A^2 is achieved more slowly in relation to changing gene frequency, but the existence of two homozygous peaks of genetic merit insures that any gene removed from a frequency of .5 will be pushed farther toward one of the limits, 0 or 1, as selection proceeds over time. Therefore, the force of selection predominates over random drift to the extent that populations of size eight reach homozygosity little faster than those of size 32. Genetic progress was even more rapid, in populations of all sizes, under the AxA conditional epistatic model, than with additive action, especially for the smaller populations.

The duplicate factor model is of such nature that one might conjecture that selection would be rendered almost powerless as a force opposing random drift. This is because all two-locus genotypic combinations except the double-homozygous recessive have the same value (Table 26). Therefore, one would expect differences in population size to have a noticeable effect upon the mean, especially so because dominance variance is involved. However, the observed results do not behave in this way. It appears that almost any significant amount of artificial selection, i.e., $< \frac{1}{2}$ selected, will exclude the undesirable double homozygote sufficiently to raise the mean consistently, regardless of population size. It should be noted that, in this model, random fixation can take one of four forms, fixation in AABB,

AAbb, aaBB or aabb states for interacting pairs of loci. Fixation in any of the first three states contributes maximum genotypic value. Thus, the force of random drift is roughly $1/3$ as powerful, from this point of view, as in a model such as the additive one. Rather than being an antithetical force to selection, it seems possible that random drift may work in conjunction with selection. This conclusion is confirmed by the results from parameter sets including the smallest populations, 8 and 12, with the weakest selection intensity, $\frac{1}{2}$. In both of these cases, fixation at maximum genotypic value for the population was reached in only 12 generations. Thus, it appears that the forces of random drift and selection may be somewhat complementary in this situation, rather than being opposite in effect.

The additive-by-dominance model (AxD) is a complex case and the reason which can explain the lack of observed differences in the mean due to population size (Figure 12) is not immediately obvious. However, if one studies the genotypic values for a pair of interacting loci, shown in Table 26, it can be seen that the intra-locus effects at locus A are overdominant, in the usual sense of the word, when combined with bb, additive when combined with Bb and also overdominant when combined with BB. The term overdominance here refers to the general situation where the

value of the heterozygote is not within the range of values of the homozygotes (Kempthorne, 1957, p. 317). In the latter case the value of Aa is less than the values of AA or aa if it is combined with BB, or, in the notation of Fisher (1918), $\underline{a} = +1$, $d = -2$ and $-\underline{a} = -1$. Consequently, if selection is effective in increasing the average gene frequency at most loci, i.e., if the population progresses toward the highest, homozygous peak of genetic merit rather than in the direction of either of the subordinate peaks, the overdominant situation where $d = -2$ will predominate. Then the effect of inbreeding will be almost twice as large as in the usual case of inbreeding depression with complete dominance (Lush, 1948, p. 260 or Falconer, 1960, p. 251) but in the opposite direction. A different term is needed for this phenomenon. Perhaps one could term it inbreeding "uplift" or "elevation" or even "sublimation". However, if selection proceeds in the direction of either or both of the peaks of lesser merit, inbreeding depression will develop.

Even though the differences in means due to population size (Figure 12) were not statistically significant, the trends may portend situations which are real. For example, in the first 10-15 generations, the means of small populations increased considerably faster than those of slightly larger populations but the trend was reversed in the next 15-20 generations. As mentioned previously, a contour map

of the expected means and variances for this model shows (1) that no "valley" of genetic merit exists between a completely heterozygous population and the primary, homozygous peak (Figure 64), (2) the magnitude of random change in gene frequency from .5, required to achieve more than a negligible amount of additive variance for selection to act upon, is considerably less in the direction of the primary peak than toward either of the subordinate peaks and (3) as gene frequency moves farther toward the primary peak, the amount of additive variance (σ_A^2) increases very rapidly, while in the corresponding situation toward the subordinate peaks, the proportion of σ_A^2 increases very slowly at first but a bit more rapidly over time (Figures 64 and 65).

Utilizing the foregoing observations, the argument may proceed that the probable course of events should be (1) early effectiveness of selection in changing gene frequency rather rapidly in the direction of homozygosity for a majority of interacting pairs of loci, with assistance from inbreeding "uplift" after gene frequency has changed enough for the overdominant situation where $d = -2$ to predominate, and (2) exhaustion of additive variance among the loci proceeding toward homozygosity, with the result that selection is applied more and more to the remaining loci whose gene frequencies have drifted toward the subordinate peaks, some as far as fixation of the undesirable allele in the smaller

populations, because of inbreeding depression, and fewer of these in the larger populations.

The sharp, rapid decline of the observed amount of genotypic variance, beginning about the sixth generation in the smallest simulated populations, and later, with correspondingly less severe decreases, in the larger populations, is consonant with the hypothetical result proposed.

Multiple comparisons, i.e., the testing of all possible pairs, of the levels of population size for the models in which general statistical significance of effects was observed (Figures 1-4) revealed that the mean of populations of size 32 was always significantly larger than the mean of populations of size 8, if any pair differences were significant at any point in time over the period of 30 generations. In populations with additive gene action (Figure 5), 32 versus 8 was the only pair difference which was statistically significant. This condition was not achieved until after the 20th generation. Tests of differences among means in the interval of generation 16-20, revealed none which were statistically significant, illustrating the point that the mean square for the general effects of a factor may be statistically significant at some probability level, α , yet tests of all differences among levels of the factor may yield no significant results at the same probability level. This phenomenon can be accounted for by the accumu-

lation of small differences among several levels of a factor, none of which are deemed worthy to be called significant.

The custom of using the five percent level of probability to demarcate significance when testing hypotheses has led many experimenters to place complete confidence in this level of risk protection for controlling type I errors. Whether or not this is the most appropriate level depends upon the nature of the experiment, the possible economic relationship to the research, the number of replications used and the ratio of the damage done by a type I error and a type II error. Avoidance of the rigidity of custom, plus the relevance of sequential tests over time involved in this study, led to the method used in Figures 1-4 for presenting the probabilities associated with observed differences.

Returning to the discussion of results under additive gene action, one may ask why other pair differences were not statistically significant and why general significance was not realized in the first 15 generations. The answer to the first question involves the fact that inbreeding depression does not occur when variance due to dominance, or to interactions which involve only dominance, is lacking. Consequently, differences in means which do occur are relatively small and are due to changes in gene frequency caused by selection and the effect of population size on the strength

of random drift, i.e., they must be explained in terms of the panmictic index, $P_t = P_0 e^{-t/2N}$, and selection. The values of P for $t = 20$ generations are approximately .72 and .27 for populations of size 32 and 8, respectively. Thus, a difference of about 45 percentage points of original heterozygosity was reflected in the mean genotypic values of the populations as approximately $237 - 224 = 13$ on the arbitrary scale where original means were at 200, i.e., the loss of almost three times as much heterozygosity due to random drift in the small populations as in the large ones was reflected in the ratio of mean genetic progress of approximately 2:3. Or, putting it another way, the loss of only 1/3 as much heterozygosity in the large populations as in the small ones resulted in an additional 1/3 gain for the mean. This numerical correlation is a spurious one derived from the way in which the gain is stated, but the existence of a true relation of considerable magnitude between loss of heterozygosity and change in the mean is undoubted.

Robertson (1961) indicated that inbreeding will be greater than that calculated from the actual number of parents when both the intensity of selection and the heritability of the character are high. A comparison of results from some of the parameter sets (N = population size, S = selection intensity, E = environmental variance and L = linkage) appears to corroborate this point. The population

associated with the parameter set (N,S,E,L) should take $t = 6N$ generations to obtain .05 heterozygosity by genetic drift. The comparison with the observed generation of fixation (t') is as follows:

$$\begin{array}{lll} (8, 1/4, 3\sigma_G^2, .005), & t = 48, & t' = 16, \\ (16, 1/4, \sigma_G^2, .2), & t = 96, & t' = 29, \\ (8, 1/6, 0, .2), & t = 48, & t' = 9, \\ \text{and } (16, 1/6, \sigma_G^2/3, .005), & t = 96, & t' = 21. \end{array}$$

Obviously, selection increases the speed of fixation but the relevant point here is that fixation occurring in a population with the parameters in set 2 is expected to take twice as long as fixation with set 1 and likewise with sets 4 and 3, whereas the observed ratios are 1.775 and 2.333, respectively, i.e., the difference in sets with high selection intensity involved more inbreeding due to size of population than the difference in sets with lower selection intensity. However, it must be noted that the differences in environmental variation involved in the comparisons could have contributed to this phenomenon. It seems unlikely that this contribution is large but if the effect were only to increase the difference in number of generations required for fixation by two for the first two sets, and to make a corresponding decrease between sets 3 and 4, the observed ratios would both be exactly 2.0, which is the expected value. Linkage would not contribute to differences in the

comparison, because these effects cancel. It is unfortunate that comparisons cannot be made where parameter sets are identical except for the factor in question, but this is one of the burdens which must be borne when fractionally replicated designs are used. An additional handicap is the fact that useful comparisons involving wider differences in population size and in selection intensity were not possible with the parameter sets available because of the particular nature of the orthogonal combinations, i.e., high selection intensity occurring with low levels of environmental variance and vice versa, which appeared in the possible comparisons involving populations of size 32 and selection intensities of $1/2$ and $1/8$. Comparisons involving these wider differences in factor levels might support Robertson's hypothesis more conclusively.

Apparently the observed differences in mean genetic progress of any two populations with less difference in size than 32 and 8 are not reflections of true differences because of sampling error, or else the design lacked the precision necessary to distinguish them, a phenomenon involving the small magnitude of the differences in relation to the magnitude of sampling errors. However, the observed trends over the period of 30 generations consistently favor the direction expected. Also, the proportional differences involving populations of size 16, 12 and 8 are about as ex-

pected from calculation of the panmictic index, but the differences involving populations of 32 with the others appear to be larger than expected in relation to the magnitude of all other differences. However, the differences between populations of 32 and 16 are expected to be somewhat larger than differences between the other consecutive levels after several generations, i.e., the change in the panmictic index, P , is such that levels of N which were equally spaced for $(1/2N) = (1 - P_1)$ are not so spaced for $(1 - P_t)$ when the number of generations, t , is large. This deviation is characterized numerically by the fact that the differences in means between populations of size 32 and those of size 16 and 12 approached statistical significance rather closely in later generations, while other differences did not.

General statistical significance of differences in means resulting from the size of population was not realized in the first 15 generations, probably because the force of selection was much stronger than random drift during that period. Consequently, populations of all sizes advanced rather rapidly at first, with consistent but small differences in mean due to random drift not becoming sufficiently large in magnitude relative to sampling errors until the advance under selection had exhausted a considerable amount of the original genetic variance, weakening the effect of selection.

In populations with complete dominance (Figure 6) paired

comparisons gave results much like those observed with additive gene action, i.e., the only statistically significant difference was that between means of populations of size 32 and size 8 after generation 20. In the last five generations, 26-30, the difference was significant at $\alpha = .01$ with complete dominance. In the interval of generations 16-20, no pair differences statistically supported the general significance which was indicated, but the comparison of 32 versus 8 yielded a probability close to $\alpha = .05$. In the later generations all pair differences except 16 versus 12 approached statistical significance.

Examination of Figures 5 and 6 will show the difference in trends observed with additive gene action and with complete dominance. With additivity, selection is much stronger than random drift in the first 10 generations, but with dominance, inbreeding depression appears to be developing quickly, possibly having considerable effect from the start. It is not until later, when gene frequencies have changed to values that enable dominance to obscure almost all unfavorable alleles from discrimination by selection, that the effect of population size is enough stronger than selection so that differences in the means are of sufficient magnitude, relative to sampling errors, to achieve statistical significance.

The smallness of the difference between the means for

populations of size 16 and 12 is more understandable after checking the results of individual parameter sets. In the four populations of size 12, levels of selection intensity and environmental variation were both high or low, while in populations of size 16, high levels of selection intensity were associated with low levels of environmental variation and vice versa, as a result of the choice of the particular fraction of the experimental design used. Therefore, all populations of size 12 produced similar changes in the mean under selection but the four populations of size 16 produced four strikingly different changes in the mean according to the desirability of the various associations of selection intensity and level of environmental variation. The performance of the population of size 16 with $\frac{1}{2}$ selected, and with the highest level of environmental variation, was noticeably poorer than that of the other populations of the same size. In fact, the regression of population mean on generation number was negative. This particular combination of parameters alone appears to be responsible for the failure of the average performance of populations of size 16 to exceed the performance of populations of size 12 to the degree expected. Consequently, the existence of a real interaction between selection and environment is implied.

Comparisons of the means of populations for all pairs of levels of population size, with overdominance present

(Figure 7), revealed statistical significance for almost every comparison over the period of 30 generations. A notable exception was the failure of the mean for populations of size 16 to exceed significantly that for populations of size 12 except for the last five generations (26-30). This exception undoubtedly may be countered by the same argument given above in the discussion of trends under dominance, i.e., an interaction exists between selection and environment. Consultation of charts showing the progress of the mean associated with individual parameter sets confirms this suggestion, again showing the low mean for the population of size 16 when selection was weak and environmental variation high. However, the 16-12 difference gradually approached statistical significance throughout the 30 generations of selection, finally attaining it after generation 25.

The general level of significance of differences in means among levels of population size when overdominance was present was the highest for any factor or model investigated. Differences were significant at $\alpha < .001$ over the entire period of 30 generations. The illustration in Figure 2 does not give proper credit at this point. However, such high levels of significance were rather rare for other factors and models.

The apparent reason for such strong effects of popula-

tion size is the phenomenon of overdominance occurring in a mass selection scheme. If one could be certain that overdominance was the predominate type of gene action, inbred lines would be formed, selected and crossed to take advantage of it, knowing that mass selection cannot long maintain any gains accomplished because the favored genotypes are heterozygous. But if one does not know that overdominance predominates when, in fact, it does, the results should be similar to those observed in the simulated populations, i.e., very strong inbreeding depression resulting in negative regression of mean on generation number. The magnitude of this regression is, of course, smaller for the larger populations and somewhat smaller for intense selection than for weak selection. Thus random drift and segregation of heterozygous genotypes almost completely overpower mass selection for an overdominant character.

The complementary factor model is strongly suggestive of the complete dominance model in that its epistatic effects exhibit dominance. Consequently, it is not surprising that the results of mean genetic progress by level of population size (Figure 10) resemble those of dominance (Figure 6). In the complementary case, the only difference in means which was statistically significant was that of populations of size 32 versus those of size 8. This difference began to approach statistical significance after five generations, not

achieving it until after ten.

Again, as in dominance, the trends illustrate that the mean of populations of size 16 does not exceed that of populations of size 12 by as much as expected in relation to the magnitude of other differences. However, the restriction is not as severe as in the dominance case. Perhaps the fact that the double recessive homozygote for a pair of interacting loci is not penalized in genetic merit, as it is with dominance, reduces the severity of the interaction of selection with environment hypothesized above. If, as it appears, the low level of selection primarily is responsible for the interaction with environmental levels in the presence of dominance this postulation should have some merit.

The last model for which significant differences in the mean of populations of different size should be discussed is the conditional epistatic one in which dominance-by-dominance variation comprises the total genotypic variance if gene frequencies are equal to .5. When one perceives that the model essentially consists of two-locus interactions of the two kinds of overdominance, heterozygote superior or inferior to both homozygotes (Table 26), and that the model is relatively free from the effects of selection because σ_A^2 and σ_{AA}^2 do not exist in such populations at any gene frequencies (Figure 66), the results illustrated in Figure 13 are not

at all surprising. Except for smaller magnitude, the negative regressions of mean on generation number and the differences between the means of populations of different sizes are remarkably like those illustrated for conventional overdominance (Figure 7).

The statistically significant individual comparisons of means include gradually increasing importance of the difference between means of populations of size 32 and size 8 over the entire period, the difference involving populations of 16 and 8 after 15 generations and all differences among levels of population size, except for 32 versus 16, after 20 generations.

Relation of random drift to fixation observed The relation of population size to random drift, as it affects fixation of alleles and the mean frequency of genes still segregating, is illustrated in Tables 1-8, Appendix B, for all of the genetic models which were simulated except for the additive case. In most cases the mean frequency of the unfixed genes appears to have little relation to population size after several generations of selection, because many alleles are fixed and the number of genes included in the computation of the mean frequency decreases rather rapidly.

With four of the models, complete dominance, duplicate factors, complementary factors and AxD (Tables 1, 4, 5 and 7, respectively), selection favors one allele over the

other. In these cases, increasing the population size results in considerably less fixation of both alleles, as expected, although the effect is more noticeable for complete dominance and complementary factors than for the other two gene models, especially for populations of size 32, where no fixation of the deleterious allele occurred at any level of selection over the entire 30 generation period. With the duplicate factor and AxD gene models, some fixation of the deleterious allele could not be avoided even in populations of size 32, where some of these alleles were fixed as early as generation 10. In the duplicate factor model, however, the definition of a deleterious allele is conditional upon the genotype existing at the other locus of an interacting pair. Thus, selection sometimes aids in the fixation of either allele. In the AxD model, the two single-heterozygote peaks of genetic merit are the source of similar results.

If one relates the total amount of fixation to that expected from random drift alone, the relative effect of selection on fixation in different sizes of population may be assessed. From the panmictic index for generation t , $P_t = P_0 e^{-t/2N}$, as given by Malécot (1948), the proportion of the initial total heterozygosity which should be lost by random drift in 20 generations is 73, 56, 48 and 28 percent for populations of size 8, 12, 16 and 32, respectively.

Fixation corresponding to these percentages has taken place in about 15 generations in simulated populations with complete dominance. However, the implication that selection has increased the speed of fixation by about 5 generations is not exact because loss of heterozygosity and fixation are not quite identical due to the existence of some genes in a homozygous state in a majority of individuals in the population but not in all members. Consequently selection should have an even stronger effect on fixation than implied.

Knox (1962) has studied the stochastic behavior of time to fixation in random-mating populations of fixed size with no selection or mutation, using an absorbing Markov chain with Bernoulli transition probabilities and approximate moments from diffusion theory. The theory is based on populations involving a single locus, but it can be adapted for multiple loci if free recombination exists. Even this restriction appears to be of little importance in models of gene action where one allele is favored over the other. At least the extrapolation to the case of multiple loci seems reasonable after noting that no important differences in fixation percentage due to linkage were observed in the simulated populations with such gene action.

Knox (1962) gives the mean time to complete fixation for completely heterozygous populations of 10, 20, 30 and

50 alleles as 13, 26, 40 and 68 generations, respectively. These correspond to populations of 5, 10, 15 and 25 individuals. The results are almost linear with respect to population size. Therefore, free interpolation and extrapolation to populations of size 8, 12, 16 and 32 yields approximate fixation times of 21, 32, 43 and 86 generations, respectively. Results from the simulated populations of size 8 and 12 indicate that mean fixation could occur that fast due to random drift alone. Knox's estimates are subject to large standard errors, approximately 75 percent as large as the estimates themselves, but the simulation results suggest that his estimates are probably reasonable.

It seems likely that approximate estimates of fixation time due to random drift, which are derived from the panmictic index, are too large as Baker and Comstock (1961) suggested they might be. For example, calculation of the loss of 95 percent heterozygosity by the relation $-t = 2N \ln (.05)$, or $t = 6N$, gives 48, 72, 96 and 192 generations for populations of 8, 12, 16 and 32 respectively. As Lush (1946) noted, overestimation of fixation time is probably due, primarily, to the fact that different parents leave varying numbers of surviving offspring.

Models such as overdominance, optimum number and DxD (Tables 2, 3 and 8), which do not favor one allele over the other under selection, also produce a negative correlation

between population size and total amount of fixation, but the fixation is equally divided between alleles as it would be without selection. Selection and random drift are antagonistic forces with overdominance and DxD gene models. Thus, there is less fixation in these two models for all sizes of population than that which occurs under a model such as the optimum number case, where the forces of selection and random drift complement each other for any loci linked in repulsion or not linked.

With the AxA model (Table 6), selection does not favor either allele, and population size appears to have little effect on the rate of fixation because selection is much stronger than random drift even though they are complementary in effect.

Comparison of Monte Carlo results with the theoretical relation of gene frequency to the mean The effects of population size on rate of fixation are quite similar to the effects on progress of the mean, indicating substantial correlation between changes in gene frequency and changes in the mean. The exact specification of such a relation involving small populations awaits further developments, despite the efforts of Fisher (1930), Kempthorne (1957) and Kimura (1958) described earlier in this section. As a test of the current thought on the nature of the relationship, some of the simulated populations were compared with

Kempthorne's fundamental equation of selection theory.

For the two-allele case the change in mean takes the form

$$\Delta\mu = 2(\Delta p\alpha_1 + \Delta q\alpha_2) + Q\Delta\lambda d_{12}/\lambda,$$

where the α 's are average gene effects and λ measures the departure from random mating genotypic frequencies in the manner defined by Fisher (1941). Changes in gene frequencies are symbolized Δp and Δq for the two alleles, Q represents one-half of the fraction of heterozygosis and d_{12} is the dominance deviation attached to the genotype A_iA_j .

The first term of the equation is attributable to changes in gene frequency and it is the only contribution to change in the mean if no dominance exists or if mating is random. Calculation of this term, for the populations associated with all parameter sets under complete dominance and overdominance, was accomplished by specifying the simulated average effects in terms of gene frequencies, and then summing over all loci. Such computation is not appropriate for epistatic models, where average effects are conditional upon the genotypic state at another locus. The equations are $\Delta\mu = 4\sum\Delta p_i q_i$ for complete dominance and $\Delta\mu = 4\sum\Delta p_i (p_i - q_i)$ for overdominance.

Comparison of the computed means with the observed ones indicated that the equations consistently overesti-

mated the change in the mean over five generation intervals although the differences were much smaller after generation 15. Obviously, some portion of the discrepancy is due to inbreeding depression caused by dominance, which was not accounted for in the simple equations involving average gene effects. The fact that the discrepancies were larger with overdominance supports this contention. Some discrepancy is due to the fact that the predicted changes in the mean are based on small changes in gene frequency, whereas, in this study, changes in gene frequency were large. Another portion of the discrepancy is due to the fact that the average effects of genes do not remain constant over an interval even as short as 5 generations, especially in populations which simulate only 40 loci. The remaining portion of the discrepancy is due to the mating system. Theoretically, the simulated populations were mated at random, but small populations develop some inbreeding due to finite size alone. Therefore, the discrepancies observed were larger for the smaller populations because the simple equations do not account for this effect.

Inclusion of the second term of Kempthorne's (1957) fundamental equation would be expected to rectify discrepancies, between the observed change in the mean and that calculated from gene frequencies, which are due to dominance and to non-random mating.

Converting the λ 's and dominance deviations to terms involving only gene frequencies and genotypic frequencies by using specific simulation values, the contribution of the second term to change in the mean becomes

$$2\sum p_i q_i P_i R_i \left[\frac{Q_i^2}{P_i' R_i'} - \frac{Q_i^2}{P_i R_i} \right]$$

for complete dominance and

$$4\sum P_i R_i [Q_i (1-Q_i) + P_i R_i] \left[\frac{Q_i^2}{P_i' R_i'} - \frac{Q_i^2}{P_i R_i} \right]$$

for overdominance, where p_i and q_i are allelic frequencies, P_i , $2Q_i$ and R_i are genotypic frequencies and subscripts with primes refer to values at the upper end of the time interval over which change is being measured.

Surprisingly, the inclusion of this term in the equation does not improve the relation to the observed change in the mean consistently. In fact, Figures 41-44, which illustrate the relation with and without this term for populations of each size under complete dominance and overdominance, show that the relation may be even poorer at times than when only the simple equation involving gene frequencies and additive effects was used. This was particularly true for large populations under strong mass selection for an overdominant character.

The chief reason for the poor relationship appears to be the fact that the λ 's for many loci are undefined in a population of finite size under selection because of the

incomplete distribution of genotypes. Consequently, the effect of λ over time cannot be assessed for many segregating loci even though its value may be known at one end of the interval. If either homozygote class, for any locus i , is not represented in the population, i.e., if P_i or R_i is zero, then $\lambda_i = Q_i^2/P_iR_i$ is undefined. This situation occurs frequently in small populations with complete dominance because selection acts to reduce R_i rather quickly for many loci.

When populations involving an overdominant character are selected, this situation does not occur so often, but another phenomenon contributes to the problem. Segregation of predominately heterozygous genotypes and recombination of predominately homozygous genotypes seems to run in cycles so that the value of Q fluctuates strongly. If Q decreases over time, $\Delta\lambda$ is negative but the magnitude of contribution to $\Delta\mu$ is tempered considerably because the effect is divided by λ_0 , the value involving a large Q at the beginning of the time interval. However, if Q increases over time, as would occur when a majority of individuals, which are homozygous at many loci, are mated to the opposite homozygous types, $\Delta\lambda$ is positive, indicating a positive contribution of non-random mating to $\Delta\mu$, and the magnitude of the effect is enhanced because the term is divided by λ_0 , which involves a small Q at the beginning of the time

interval. Thus, a few loci contributing positive $\Delta\lambda$'s can offset many loci contributing negative $\Delta\lambda$'s to the change in the mean. This effect is particularly noticeable in the larger populations with overdominance and strong selection (Figure 44) because many alleles drift toward complete homozygosity in the total population but they recombine instead of becoming fixed (Table 2).

The second term of the general equation does not significantly improve the relation of change in gene frequency to change in the mean with any consistency for another reason. The dominance deviations do not remain constant. They change rather rapidly because of large changes in gene frequency. Linkage disequilibrium may contribute to dominance variation (Robinson and Comstock, 1955) and, thus, to the magnitude of change in dominance deviations.

Sampling of gametes does not appear to be an important factor, for the calculated relation of change in gene frequency to change in the mean was equally poor when applied to the observed changes in selected or unselected populations.

Undoubtedly, closer relationship of formula to fact could be attained by shortening the time interval of comparison to one or two generations to reduce fluctuations in additive effects and dominance deviations, but the generality of the term for non-random mating would still suffer

from numerous undefined values of λ , for which small populations are responsible.

Crow and Kimura (1955) pointed out that, even though λ is general for any mating system, it is not independent of changes in gene frequency, which may be rather drastic in small populations, but Wright's F is a function only of the mating system and, therefore, should be more useful for defining the effects of dominance in finite populations under random or consanguineous systems of mating. This contention leads one back to the Fisherian proposition that gene frequency can properly be related to the mean only in terms of the additive effects under random mating.

Kimura (1958), however, developed another equation for relating the change in gene frequency to changes in the mean. The formula was presented earlier in this section. His term for the effect of dominance is weighted by F rather than λ , i.e., $\frac{\Delta F}{1-F} \sum p_m d_{mm} - F \sum \Delta p_m d_{mm}$. He also included a term for the effects of interlocus interactions, involving two loci each, which allows epistatic models to be evaluated too. However this term involves coefficients of departure from random mating which suffer the same inadequacies as Fisher's λ for small populations. Therefore, it seems unlikely that the tedium associated with computation of the expected change in the mean and relation of it to the change observed in small simulated populations with

epistasis, would yield fruitful results. Perhaps a Monte Carlo investigation could be designed specifically to assess the veracity of such a suggestion.

Conclusions with respect to population size Because of the conglomerate nature of the relation of population size to the genetic mean, some sort of summarization seems appropriate at this point before proceeding to the discussion of selection intensity as it affects the mean.

In general, the effects of population size on the mean are small or even negligible in relation to the force of selection for populations without dominance variation of some type, but the smaller populations show more inbreeding depression when complete dominance, overdominance, complementary or dominance-by-dominance (DxD) gene models exist.

The duplicate factor model also contains dominance but random genetic drift and selection proved to be complementary forces with selection the stronger of the two so that population size had little effect on genetic progress.

The additive-by-dominance (AxD) conditional epistatic model also contains dominance at gene frequencies other than .5. Statistically significant differences in the means due to population size were not found, but general trends indicated that the forces of random drift and selection are complementary for about 10-15 generations, while σ_{AD}^2 predominates, but become antagonistic later as σ_D^2 predominates.

Thus, inbreeding "uplift" occurs first and the usual inbreeding depression evolves later.

Results with the complementary factor type of gene action differ little from those with complete dominance because the epistasis involved is much like dominance in nature and the epistatic variance is negligible for gene frequencies above .7.

Indications of the existence of an interaction between selection and environment were observed in small populations with complete dominance, and its existence is suspected when overdominance or complementary gene action is involved.

A negative regression of genetic mean on generation number was observed for populations of all sizes under mass selection with overdominance or DxD gene models. The DxD model is merely the epistatic equivalent of overdominance, a relation somewhat like that of the complementary model to the dominance one. The DxD inter-locus interactions involve the two forms of overdominance, heterozygote superior or inferior to both homozygotes.

The optimum number and additive-by-additive (AxA) models contain no dominance variation. Random drift and selection are complementary, and selection is the stronger of the two forces. In fact, genetic progress was more rapid in populations of all sizes with the AxA gene model than with the

simple additive one. Differences observed in the means of populations of different size were of negligible practical significance.

The mean frequency of unfixed alleles, in the populations simulated, is not strongly related to population size after several generations of selection because the number of genes still unfixed is markedly reduced.

The effects of differences between levels of population size on the rate of fixation of alleles are larger when the forces of random drift and selection are strongly antithetical (e.g., overdominance and $D \times D$ models) than when they are complementary (e.g., optimum number, duplicate factor and $A \times A$ models), especially when the force of selection is much stronger than that of drift (e.g., in the $A \times A$ model).

Robertson's (1961) theory that the inbreeding effect is larger than the amount calculated from population size when both selection and heritability are high, was given tentative confirmation by limited comparisons involving populations of almost the same size under selection at two adjacent levels of intensity. However, such confirmation is conditional upon the supposition of small environmental influence. Comparisons involving a wider range of levels of selection and population size might corroborate the contention more conclusively.

Ample indication was given that the mean time to allelic fixation which is implied by the panmictic index, a function of population size, overestimates the actual time, possibly because the number of surviving offspring varies among parents. The values derived by Knox (1962), using an absorbing Markov chain with Bernoulli transition probabilities, are probably more nearly correct.

Current formulae for relating the change in gene frequency to change in the mean of populations under selection generally overestimate the change because terms for the effects of dominance or epistasis, under non-random mating, are not independent of drastic changes which sometimes occur in finite populations. Under selection for a homozygous maximum, this deficiency is related to the fact that not all genotypic states are represented for many loci in a small population. In populations under mass selection for an overdominant character, or other heterozygous maximum, the problem is manifested in cycles of types of mating, where segregation of predominately heterozygous genotypes exceeds the amount of recombination of predominately homozygous genotypes of opposite state, and vice versa.

Selection intensity

Theoretical statistical studies concerning mass selection may be classified into two general groups according to the mode of attack. The first category includes investiga-

tions concerning the rate of change of gene frequencies under selection. The second group includes prediction equations.

Effect of selection on gene frequency This problem has been dealt with primarily by Haldane (1923, 1926), Fisher (1930) and Wright (1931, 1945, 1951). Much of the elaboration of these and many later workers involves descriptive formulas which relate changes in gene frequencies and genotypic frequencies to changes in the mean of populations under natural selection, given the additive and dominance effects and, in some cases, the epistatic effects at each locus. Thus, equations such as the one given by Kempthorne (1957), which was discussed previously in the section concerning population size, have been derived from the basic concepts given by Fisher (1930) and Wright (1931).

Wright (1945) introduced the forward Kolmogorov (Fokker-Planck) diffusion equation into population genetics in order to obtain the probability density of the stochastic process which describes the change in gene frequency in the population as a function of time. Fisher (1922) had previously given a particular case, though his work contained an error which was corrected later. The system of differential equations which describes the diffusion process involves functions for the infinitesimal mean and variance

of the change in gene frequency. Cases involving random drift, mutation and migration in populations under natural selection have been worked out by Malécot (1948), Crow and Kimura (1955), Kimura (1958) and by others. This type of approach assumes that gene frequency is a continuous variable, whereas the change in gene frequencies is best described by a stochastic process made up of discrete values $0, \frac{1}{2N}, \frac{2}{2N}, \dots, 1 - \frac{1}{2N}, 1$. Unless N is rather large, as it is in many natural populations, the assumption of a continuous probability density may result in serious error. However, Watterson (1962) has given conditions under which the continuous diffusion approximation should be adequate for all but very small populations.

Prediction equations The second category of theoretical selection studies includes investigations which involve prediction of change in the genetic mean based on estimates of general population parameters such as genetic means and variances from populations which are conceptually distributed normally. This type of study is more appealing to investigators in the applied fields of plant and animal breeding than is the type of study mentioned previously, because prediction equations tend to imbue minds with visions of economic success, whereas descriptive ones, such as those involving changes in gene frequency, do not.

The basic idea of prediction equations is to relate the

change in the population mean, or response to selection ($\Delta\mu$), to the selection differential, or average phenotypic superiority of the selected parents (ΔP). One of the simplest relations for mass selection involves the regression of offspring on mid-parent, i.e., $\Delta\hat{\mu} = b_{OP} \Delta P$.

The simple linear coefficient of regression was described by Gauss, more than a hundred years ago, as an expression of the linear change in units of Y for each unit of X. Its value also is chosen to minimize the sum of squared deviations of Y from the regression line whose slope is $b_{YX} = r_{YX} \sigma_Y / \sigma_X$, where r_{YX} is the correlation between y and x. The mathematical foundation of most of the work that has been done specifically for selection was provided by Karl Pearson at the end of the last century. His primary interest was in the effects of natural selection on correlation and variability.

Eisenhart (1939) showed that the regression coefficient is not biased when the distribution of the independent variable has been narrowed by truncation selection, provided the association is linear, although its sampling error is increased. The correlation coefficient is reduced by such selection, compared with its size in the unselected population. Selection of the dependent variable (Y) leads to estimates of regression which are also biased. Eisenhart (1939), Winsor (1946) and Cochran (1951) have summarized the

limitations which random environmental errors of measurement and non-normality of X and Y impose upon prediction equations.

The regression of offspring on the mean of the parents is the most nearly unbiased estimate of effective heritability, excluding selection experiments, that one may obtain (Dickerson, 1959). This is true if one can assume that environmental correlations between parent and offspring are negligible, as they are in this study. Heritability (h^2) is defined, in the "narrow" sense of Lush (1945), as the proportion of the total variance in the population that is attributable to the average effects of genes. In the broad sense of the word, heritability is the proportion of the total variance in the population that is attributable to genetic effects of all kinds. The customary symbol, h^2 , is derived from Wright (1921b), who referred to it as the degree of determination of variation by heredity. The symbol commonly is used to denote heritability in any sense of the word which the writer defines. Effective heritability includes some variation due to inter-locus interactions of additive effects. Therefore, $b_{OP} = (\sigma_A^2 + .5\sigma_{AA}^2 + .25\sigma_{AAA}^2 + \dots) / \sigma_P^2$ probably more closely estimates effective heritability than it does either of the values formally defined.

Prediction of response to selection is the chief

purpose of estimating heritability. Therefore h^2 is usually incorporated into the linear equation, giving $\Delta\hat{\mu} = h^2\Delta P$. Of course, the estimate of h^2 which is used in the equation must be derived from a previous generation of results of $b_{o\bar{p}}$, or from some other source. If one knew $b_{o\bar{p}}$ in the present situation he would also know $\Delta\mu$ and there would be no need to predict. Falconer (1960) has discussed the limitations involved in extrapolating the value for h^2 .

The prediction of response to selection is valid, in principle, for only one generation because the basic effect of selection is to change gene frequencies, and this results in changes in the heritability associated with the progeny population. However, selection experiments have shown that heritability estimates are reasonably accurate for predicting response for several generations.

In the case when the gene effects are not additive, one will not know the change in gene frequency, so that no adequate means exist for predicting the population which will result by random mating, or any other system of mating, of the selected individuals. Consequently, basic regression formulas such as those discussed by Cochran (1951), who extended and elaborated the work of Perotti (1943), are not particularly good predictors except for the linear situation of additivity.

Recently, efforts have been made to incorporate the

best qualities of both types of work, i.e., relating changes in gene frequency to changes in the mean, and formulating predictions using components of genetic variance and selection differentials, into a single comprehensive formula for describing or predicting the expected gain. One of the most elaborate of these formulas, given by Kimura (1958), was presented previously in the discussion concerning population size. He has retained the basic regression expression, $h^2\Delta P$, to account for additive effects associated with changes in gene frequency, and utilized one of the descriptive equations which tie dominance deviations to individual changes in the frequencies of genes, weighting this term by the independent and predictable inbreeding coefficient F . In trying to account for epistatic deviations, he reverted to a basic technique used by his predecessors (e.g., Fisher, 1949 and Kempthorne, 1957); a method which is more general than F for expressing the effect of the departure from random mating but one which is not independent of drastic changes in gene frequency, and therefore, appears to be inadequate for finite populations. Despite the limitations, Kimura's formula appears to be the most comprehensive formula available for descriptive purposes.

Kojima (1961) has developed a formula to express the effects of dominance and size of population on response to mass selection for the single-locus case. Generalization

to the multiple-locus case is possible only when epistasis and linkage disequilibrium are not present. The generalized formula is:

$$E(\Delta\bar{y}) = k\sigma_A^2 \pm \frac{1}{2} k^2 \sum_i (\sigma_\alpha^2 \sigma_\beta)_i \pm \sum_i V(\Delta p_i) (\sigma_\beta)_i / p_i q_i .$$

In the first term of the equation, σ_A^2 is the additive genetic variance as a fraction of the total variance, i.e., $\sigma_A^2 \approx h^2$. Thus this term is equivalent to the conventional prediction equation, $h^2\Delta P$, except that k is a generalized selection differential, since it is not restricted by the size of sample or the form of the phenotypic distribution. Kojima showed, however, that his k is not appreciably different from the usual selection differential. This fact bolsters one's faith in the usual method of computing theoretical selection differentials for finite populations from tables of ranked normal deviates (Fisher and Yates, 1943), upon the assumption of normality of the phenotypic distribution.

The second term represents a positive or negative contribution due to dominance, where σ_α^2 and σ_β^2 represent additive and dominance variance due to a particular locus. The magnitude of this contribution is proportional to the gain due to the linear effects of genes ($k\sigma_\alpha^2$) and to the degree of dominance measured by σ_β .

The third term does not depend on the linear effect of the genes but is proportional to the variance in the change of gene frequency. This variance is made up of variation

due to random genetic drift and to a factor which modifies the sampling variance as a function of the genetic effects and the selection differential. The sign of the third term also depends on the type of dominance.

The last two terms are negative when the heterozygote is better than the mean of the two homozygotes, and are positive otherwise. These are the terms which serve to point out possible biases in the usual prediction equation, $h^2\Delta P$. The first part of the deviation of $h^2\Delta P$ from $E(\Delta\bar{Y})$ is due to the non-linearity of gene action with respect to the allelic substitution at a particular locus. The adjustment is not restricted by the sizes of the genetic sample used. The second part of the deviation represents a general effect of inbreeding when genes exhibit dominance. When the term is negative, the effect is the usual inbreeding depression due to finite populations.

Since Kojima concluded that size of population does not cause a serious difference between the gain expected from $h^2\Delta P$, and that from $E(\Delta\bar{Y})$, except as it contributes to inbreeding depression when dominance gene action exists, it would seem that his formulation has little utility over the standard method of describing changes due to additive effects by regression, and those due to dominance by Wright's coefficient of inbreeding, F , as Kimura (1958) did in the first two terms of his equation.

Griffing (1960a) acknowledged Kimura's 1958 paper as the only generalized treatment of the theoretical response of a population to continuous artificial selection. He also credited Lush (1948) and Kempthorne (1957) with the intuitive suggestions, concerning partial transmission of epistatic variation from parent to offspring, which led him to develop an argument for the general problem of description and prediction associated with truncation selection based on the individual phenotype, with arbitrary dominance, epistasis, linkage and number of loci.

Lerner's (1954) work stimulated Griffing to show that the contributions of additive-by-additive epistatic components of variance to the responses of selection and to relaxation following selection, mimic those due to natural selection. It generally had been assumed that natural selection acting antagonistically to artificial selection was responsible for the phenomenon termed genetic homeostasis by Lerner.

For selection based on genotypes generated by alleles at two loci which may be linked, Griffing followed the generalized gene model of Kempthorne (1957), in which the genotypic value is characterized as follows:

$$\begin{aligned}
 G_{ijkl} = & \alpha_i^1 + \alpha_j^1 + \alpha_k^2 + \alpha_l^2 + \delta_{ij}^1 + \delta_{kl}^2 + (\alpha\alpha)_{ik} + (\alpha\alpha)_{il} \\
 & + (\alpha\alpha)_{jk} + (\alpha\alpha)_{jl} + (\alpha\delta)_{ikl} + (\alpha\delta)_{jkl} + (\delta\alpha)_{ijk} \\
 & + (\delta\alpha)_{ijl} + (\delta\delta)_{ijkl},
 \end{aligned}$$

where α_u^a = additive genetic effect of the A_u^a allele,

δ_{uv}^a = dominance effect for the $A_u^a A_v^a$ genotype,

$(\alpha\alpha)_{ik}$ = additive-by-additive epistatic effect associated with genes A_i^1 and A_k^2 ,

$(\alpha\delta)_{ikl}$ = additive-by-dominance epistatic effect associated with the gene A_i^1 and the genotype $A_k^2 A_l^2$,

and $(\delta\delta)_{ijkl}$ = dominance-by-dominance epistatic effect associated with the genotypes $A_i^1 A_j^1$ and $A_k^2 A_l^2$.

The total genotypic variance in a random mating population may be partitioned as $\sigma_G^2 = \sigma_A^2 + \sigma_D^2 + \sigma_{AA}^2 + \sigma_{AD}^2 + \sigma_{DD}^2$. A two-allele partition was given by Cockerham (1954) and the general partition was given by Kempthorne (1954). Both partitions are accurate for unlinked genes, and the bias in the epistatic components probably is rather small if linkage equilibrium exists. Linkage disequilibrium may bias all the components. Robinson and Comstock (1955) have shown that σ_D^2 is biased upward, and that σ_A^2 is biased upward if coupling predominates or downward if repulsion predominates.

The frequency of the genotype $(A_i^1 A_k^2) (A_j^1 A_l^2)$ following selection is $f_{ik}^0 f_{jl}^0 [1 + (\Delta P / \sigma_P^2) G_{ijkl}^0]$, and the genotypic mean of the selected parents is $\mu_S = \sum_{ijkl} f_{ik}^0 f_{jl}^0 [1 + (\Delta P / \sigma_P^2) G_{ijkl}^0] G_{ijkl}^0$. The frequency of the progeny genotype

turns out to be: $f_{ik}^1 f_{jl}^1 \approx f_{ik}^0 f_{jl}^0 + (\Delta P/\sigma_p^2) f_{ik}^0 f_{jl}^0 [\alpha_1^1 + \alpha_j^1 + \alpha_k^2 + \alpha_l^2 + (\alpha\alpha)_{ik} + (\alpha\alpha)_{jl}]$.

The increase in genotypic mean of the progeny population, then, is $\Delta\mu_1 = \sum_{ijkl} f_{ik}^1 f_{jl}^1 G_{ijkl}^0$ or $\Delta\mu_1 \approx (\Delta P/\sigma_p^2) [\sigma_A^2 + \frac{1}{2} \sigma_{AA}^2]$ which, obviously is not the same as the usual linear prediction, $\hat{\Delta\mu} = h^2 \Delta P$, unless h^2 is the regression of offspring on mid-parent.

Griffing derived the predicted change in mean after n generations of selection as

$$\Delta\mu = (\Delta P/\sigma_p^2) [n\sigma_A^2 + \sum_{i=1}^n (1-r)^{i-1} \sigma_{AA}^2/2],$$

where r is the recombination frequency between the two loci.

It is clear that the extent of the influence of σ_{AA}^2 is largely determined by the magnitude of the recombination value, which has the range $0 \leq r \leq \frac{1}{2}$. Thus, if $r = 0$ (i.e., there is no recombination), $\Delta\mu = (\Delta P/\sigma_p^2) [n\sigma_A^2 + n\sigma_{AA}^2/2]$, or if $r = \frac{1}{2}$ (i.e., loci are independent), $\Delta\mu = (\Delta P/\sigma_p^2) [n\sigma_A^2 + (2 - [\frac{1}{2}]^{n-1}) \sigma_{AA}^2/2]$.

Therefore, if the loci exhibit a low recombination value considerable effect can be generated by σ_{AA}^2 , and if the loci are independent the maximum contribution after n generations of selection is approximately σ_{AA}^2 , for large n .

Furthermore, if $r \neq 0$, the increment changes between consecutive generations of selection are not equal. Thus, $\Delta\mu_{n,n-1} = (\Delta P/\sigma_p^2) [\sigma_A^2 + (1-r)^{n-1} \sigma_{AA}^2/2]$. Hence, the influence

of σ_{AA}^2 diminishes as the number of generations in the selection program increases. This causes a departure from linearity of the response of selection with time. In actual populations, the long-time response to selection should result in an asymptotic approach to the goal of selection, whether it is that of homozygosity or a stable equilibrium.

Comparison of Monte Carlo results with the theoretical prediction of change in the mean A comparison of the observed results, from simulated populations involving dual epistasis, and the results predicted by Griffing's equation, should show the nature of the limitations of the derived equation. Griffing's first assumption is that the populations are infinite in size. In finite populations, genetic drift will undoubtedly affect the reliability of the prediction. Secondly, discrepancies may result because it was assumed that the effects of individual genes are small so that the square and products of the quantities (gene effect/total phenotypic standard deviation) can be neglected. This assumption is particularly important for the comparison with simulation results because of the mechanics of simulation. With only 40 loci and no extraneous variation due to the remainder of the postulated actual genome which was not simulated, the effects of simulated genes are bound to be relatively large in relation to the total variance. Here is a point of contention for those who would

simulate extraneous variance for loci not taken into account (Baker and Comstock, 1961). Therefore, discrepancies which are due to the relative size of gene effects cannot necessarily be ascribed totally either to bias in the prediction or in the simulation. Perhaps the situation can be overcome simply by simulating more environmental variance.

The errors introduced by approximations tend to accumulate so that the basis of prediction becomes more subject to error as the mean of the selected population becomes farther removed from its original position.

Finally, the assumption that the parameters are constant, and do not change after several generations of selection, is likely to be as serious a flaw as the failure to account for random drift. However, discrepancies in the comparison with simulation results which are due to changing parameters are likely to be magnified for the same reasons given above in the discussion on the relative size of gene effects.

Predictions for comparative purposes were made for all parameter sets of four models, additive, complete dominance, optimum number and additive-by-additive (AxA). The first two of these do not involve σ_{AA}^2 . Therefore, the predictions made for them are the usual linear ones, but the results should give indications of the type and severity of errors due to some of the assumptions. Predictions were made on

two bases for each population: (1) The parameters which were used to simulate the original population, and (2) parameters measured in an early generation of the Monte Carlo results. The second method was added because the first one involves some situations which are unique and peculiar to the original population (e.g., gene frequency of .5 at every locus). This method also provides a measure of bias due to the change in parameters from their original values. The two results are given in Figures 45-56 as \hat{G} and \hat{G}' , respectively, along with the observed progress of the mean (G).

For the additive model, typical cases for populations of size 8 and 32 are presented in Figures 45 and 46. It is rather clear that random drift is important over several generations of selection. Thus the effects of finite population size are likely to contribute significantly to bias in Griffing's formula. Although the predictions using parameters observed in the first generation of Monte Carlo results were slightly superior to the predictions using the parameters of simulation for the two cases illustrated, the reverse was true in about half of the total cases examined, as one may see from the relations presented in Table 17, which have been extrapolated linearly to generation 30. The difference between the two predictions was small, in most cases, in relation to the general difference between

predicted and observed results, indicating that the genetic parameters did not change much, or in a consistent manner in one generation, although they probably changed considerably after several generations. This change becomes increasingly noticeable after two or three generations of selection in the smaller populations, and becomes larger after 4-8 generations in the larger ones. However, the magnitude of the differences and the rate of change may be influenced by the fact that each gene in the simulated population has a relatively large effect because the number of loci is limited.

Results presented in Figures 47 and 48 and in Table 18, for the complete dominance case, are similar to the additive ones except that \hat{G}' was predicted from parameters observed in generation six of the Monte Carlo results, in order to assess the magnitude of change in parameter values, and the influence of random drift over a different period of time. Although both of the predictors were, in general, no better than those for the additive case, it should be pointed out that the difference between the two predictions was larger and exhibited a trend which may be seen in Table 18. Prediction of the mean in generation 30 obviously should be easier from generation six than from generation zero, but this proved to be true consistently only for the smallest populations, whereas, for the larger populations, prediction based on original parameters was consistently

better. This would indicate that most of the damage in genetic merit, resulting from fixation caused by random drift, had already occurred after only six generations in populations of size 8, and some of it had occurred by then in populations of size 12. However, little or no damage had been done at that time in the larger populations, leading one to the conclusion that discrepancies between predicted and observed means in the larger populations, at that point, are largely due to changes which selection caused in genetic parameters. This contention was corroborated by comparing estimates of σ_A^2 in generation six against the level simulated. The estimate of σ_A^2 includes error due to linkage disequilibrium and inbreeding caused by finite size of population. Only the population of size 32 which was under the most intense selection, 1/8, maintained an amount of σ_A^2 comparable to the original amount. The other populations showed decreases in σ_A^2 amounting to as much as 1/3 of the original amount.

In both additive and dominance cases it was not clear whether intensity of selection, level of environmental variation or combinations of the two, influenced the magnitude of discrepancies between the two predictions, although there was some indication that strong levels of selection coupled with minimal amounts of environmental variation contributed to such differences as existed.

To consider some cases which will test Griffing's theory, one must turn to models which involve σ_{AA}^2 , preferably in amounts which can be considered to be of more than trivial importance. The only simulated models which meet this criterion, over more than a narrow range of gene frequency or other special conditions, are the optimum number one and the additive-by-additive (AxA) model. The optimum number model involves σ_{AA}^2 as 2/3 of the total genotypic variance when gene frequency is .5, and still 1/3 of the total when the frequencies are equally divided between .9 and .1 (Figure 60). The AxA model is defined as involving only σ_{AA}^2 as genotypic variance when gene frequency is .5, but the proportion of σ_{AA}^2 rapidly decreases as gene frequencies are changed from .5 toward 1.0, σ_A^2 then becoming more important (Figure 63). At gene frequency of .9, σ_{AA}^2 makes up about 1/8 of the total genotypic variance.

The most obvious conclusions, which may be derived from examining the correspondence between predicted means and the observed means for populations involving all 16 parameter sets under the optimum number model, are: (1) a strong positive correlation exists between the "goodness of fit" of the predicted curves of the mean and the amount of recombination involved, i.e., when linkage was quite tight (.005) both the mean predicted from parameters of simulation (\hat{G}) and the one predicted from parameters observed in

the second generation of Monte Carlo results (\hat{G}'), far over-estimated the observed progress in the mean, (2) the size of population affects the accuracy of prediction considerably, i.e., random drift in the smaller populations contributes heavily to discrepancy and (3) \hat{G} consistently is a better predictor than \hat{G}' when linkage is reasonably close (.05 and .005), but it is not particularly better and, in some cases, it is worse than \hat{G}' for low levels of linkage or free recombination (.2 and .5).

Most of the discrepancy between the predicted and observed means, when recombination was extremely limited, was due to the fact that the amount of σ_{AA}^2 declined considerably from the value used in the prediction equations, with differences among populations of different size being due to random drift (Figures 49 and 50). The fit of predicted curves was, in general, much better when free recombination existed because the maximum contribution of epistatic variance to the prediction was σ_{AA}^2 , this maximum being achieved and remaining constant after 15 generations of selection. Thus, the decline in σ_{AA}^2 from the original value in the populations did not contribute as much to accumulated errors in the prediction equation when there was no linkage. However, the effect of random drift was still evident in populations of different size (Figures 51 and 52).

The third point, concerning the relative reliability of

the two predictors, derives from the curious phenomenon that the optimum number model does not involve any additive variance (σ_A^2) at gene frequency of .5, but considerable amounts are generated by even small changes in gene frequency. Therefore the predictor, \hat{G} , which utilizes the parameters of simulation of the original heterozygous population, involves no σ_A^2 , whereas the predictor, \hat{G}' , utilizes the amount of σ_A^2 generated after two generations of mating and one generation of selection have occurred. This amount varied from 10 to 60 percent of the total genotypic variance. Much larger amounts were generated quickly in small populations than in larger ones but the amount does not seem to be related to the intensity of selection. Therefore, one must conclude that random drift is responsible for most of the difference. Because of this situation, \hat{G} was a better predictor than \hat{G}' when linkage was tight, even though both grossly overestimated the observed genetic progress because of large predicted contributions of σ_{AA}^2 . However, for low levels of linkage or for free recombination, \hat{G} involved no predicted contribution from σ_A^2 and very little from σ_{AA}^2 . Therefore, progress was underestimated by \hat{G} after a few generations of selection, while \hat{G}' still overestimated progress badly, except for some of the larger populations, because the contribution of σ_A^2 was extrapolated linearly from the earli-

er estimate. In the first few generations, \hat{G} was still somewhat better as a predictor than \hat{G}' , not underestimating progress for at least 5 generations, and sometimes for longer periods in the cases involving the larger populations. The larger populations were slower in generating σ_A^2 , and thus, slower in changing the mean than the smaller populations, where random drift, rather than selection, was the early effective force.

The means for populations with the AxA gene model were predicted with parameters of simulation (\hat{G}) as before, but \hat{G}' was derived from parameters observed in generation six of the Monte Carlo results in order to ascertain whether the effects of random drift would still be as strong as if results from earlier generations were used.

As in the optimum number situation, the estimation from simulated parameters, \hat{G} , was based solely on the contributions of σ_{AA}^2 . Therefore, \hat{G} was a much better predictor of the progress of the mean when tight linkage existed, especially for the larger populations, although there was a tendency to underestimate in the early generations, and to overestimate later (Figures 53-56). When low levels of linkage or free recombination existed, \hat{G} badly underestimated the means of all populations after 2 or 3 generations of selection. Because it is known that σ_A^2 is generated rather quickly and in large amounts when gene frequencies

change from .5 (Figure 63), the fact that \hat{G} was a reasonably good estimator for any case appears to be purely accidental, since it is based solely on the contribution of σ_{AA}^2 , weighted by the tightness of linkage.

Predictions of the progress of the mean using parameters observed in generation six of the Monte Carlo results (\hat{G}') were erratically above some observed means and below other means of small populations with tight linkage, but they were consistently above those of the larger populations. With low levels of linkage or free recombination, these differences tended to be diminished so that prediction was relatively good. The relation of population size to the amount of σ_A^2 available at generation six was almost completely reversed from the relation observed at generation two under the optimum number model, i.e. for the AxA case, the lower levels of σ_A^2 available were associated with small populations. However, σ_{AA}^2 had been reduced drastically in the smaller populations by generation six while this was not generally true for the larger ones, so that the proportion of σ_A^2 in relation to the total genetic variance was still higher in small populations. Selection intensity was involved in the determination of the amounts of total genotypic variance because of the effect on the rate of changing gene frequency. Apparently random drift had changed gene frequency so that

considerable amounts of σ_A^2 were generated in the smaller populations in the first few generations, and selection acted upon it rather immediately, reducing total variance, whereas the generation of σ_A^2 in the larger populations took place more slowly because selection was not aided much by random drift in making changes in gene frequency.

A general impression of Griffing's equation is that the contribution attributed to σ_{AA}^2 under tight linkage, while possibly accurate for a generation or two, is generally far too large over several generations of selection. Random drift, as well as selection intensity, appears to have considerable influence in changing the genetic parameters drastically in a few generations in small populations, so that extrapolation over several generations is erratic. The magnitude of the discrepancies noted between predicted means and those observed in Monte Carlo results is probably larger than it would be in an applied situation because of restrictions in the mechanics of simulation.

Monte Carlo differences in means To help ascertain the nature of the way in which varying intensities of selection affect the change in the population mean, and the manner in which the effects are related to those of other factors, one may peruse the usual linear prediction equation, $\hat{\Delta\mu} = h^2 \Delta P$. The particular h^2 involved may be an estimate of σ_A^2 / σ_P^2 , so that only additive effects are involved, or it may be an

estimate such as the regression of offspring on mid-parent, so that at least part of the epistatic contributions are considered. The selection differential, ΔP , is a function of selection intensity almost entirely, except that the theoretical expectation is slightly different for small populations. Selection and random drift may work antagonistically or in concord to change σ_A^2 as they change gene frequency. All the effects due to dominance, epistasis and environment, as well as the average gene effects contribute to the magnitude of σ_p^2 . Selection may work to change σ_p^2 , and linkage may affect the rate and magnitude of the change through its influence on the non-environmental effects. Therefore, the general nature of selection is that it influences all three of the basic components of the simple prediction equation. Selection has influence as a rather constant intensity or differential, ΔP , but also it is a force for changing σ_A^2 and σ_p^2 through its effect on gene frequency.

Given the general nature of selection intensity, one would like to know something of the magnitude of the effects of different degrees of it upon the mean. Figures 1-4 give the probabilities of significant differences among the means of simulated populations being due to differences among levels of selection intensity.

The force of selection upon the mean in the simple

additive situation (Figure 1) appears to be about as one would expect, i.e., large differences appear in the early stages of selection and then gradually decline as fewer and fewer segregating loci are available for change. The sudden drop in the probability value for the period including generations 11-15 appears to be a sampling phenomenon. Careful checking showed the error mean square to be actually more than twice as large as the corresponding values in adjacent periods. It seems unlikely that such a peculiar phenomenon could be ascribed to the actual forces at work in the simple linear situation. Perhaps one should keep in mind the remonstrance of Fisher (1960) that "The one chance in a million will undoubtedly occur, with no less and no more than its appropriate frequency, however surprised we may be that it should occur to us."

The differences among means due to level of selection intensity, when complete dominance was present, did not attain general statistical significance in the first ten generations of selection, although the probability of significance increased steadily over the entire 30 generations, in contrast to the additive situation (Figure 1). This observation is in general agreement with the statement by Bohidar (1960) that, "Dominance makes selection sensitive longer." The main reason for this is probably the decline in rate of fixation and change in gene frequency because

most genotypes are indistinguishable at the frequencies attained after several generations of selection. Negative regressions of genetic merit on generation number were observed only for populations with the two parameter sets which involved the smallest population size (8) in combination with low levels of selection intensity ($\frac{1}{2}, \frac{1}{4}$) and high levels of environmental variation ($\sigma_G^2, 3\sigma_G^2$). Baker and Comstock (1961) found similar results for parent populations of size 10 but the regressions for populations of size 6 tended to be negative regardless of the levels of selection and environment simulated.

The general pattern of significance of differences among means due to selection for an overdominant character (Figure 2) was quite similar to the trend observed for the complete dominance case. However, in the overdominance situation the probabilities, $1-\alpha$, were slightly higher over the entire period of 30 generations. Since the regression of genetic merit on generation number was negative for all parameter sets, one may surmise that random drift is more powerful than the force of selection, but that the stronger intensities of selection come considerably closer to bringing about a balance of the two antithetical pressures than do the less severe ones. Bohidar (1960) observed this same phenomenon.

The variability of the force of selection upon the

genetic mean over the range of intensity simulated was rather small for a few generations under the optimum number model (Figure 2). However, it became increasingly significant until approximately generation 20 and then declined sharply. Under the optimum number model no additive variance exists in completely heterozygous populations such as the original simulated ones. Therefore, any level of selection is rather ineffective in changing the mean until random drift has caused the gene frequencies to change enough to generate tangible amounts of additive variance. After this has occurred, inherent differences in the intensity of selection become apparent. Later, when selection at high intensities for 18-20 generations has exhausted most of the additive variation, these populations tend to reach plateaus of genetic merit, whereas the populations under less intense selection continue to progress genetically, narrowing the magnitude of observed differences.

As noted previously, in the discussion concerning population size, the duplicate factor model is of such nature that almost any significant amount of artificial selection pressure, i.e., $< \frac{1}{2}$ selected, results in the exclusion of the one undesirable genotype, the double homozygote recessive for each pair of interacting loci, sufficiently to raise the mean consistently regardless of population size. Therefore, random drift toward any of the other fixation

states is complementary to selection. The only probable consequences of different intensities of selection might be the small differences in time needed to reach a particular level of genetic merit (Figure 8). In general, though, there were no differences of statistical consequence due to selection intensity (Figure 3).

As might be expected, the effects of selection under the complementary factor model (Figure 3) were very much like those observed for the complete dominance and overdominance cases. The generally low levels of epistatic components of genetic variance, and the inherent similarity of the complementary factor model to the dominance case, account for the parallel results, i.e., they explain why intensity of selection causes differences in means to become statistically significant after a few generations of selection and remain so over a considerable period of time.

Selection in populations with the additive-by-additive conditional epistatic model (Figure 4) produces results similar to those for the optimum number model, with regard to differences in the mean. The models are similar in that neither involves additive variance at gene frequency of .5, nor dominance variance at any frequency, but both involve considerable amounts of additive variance as soon as random drift moves gene frequency from .5 (Figures 60 and 63), and neither favors one allele over the other at a single locus,

although the epistatic interactions cause loci linked in coupling to be favored genotypes in the AxA model and those linked in repulsion to be favored genotypes with the optimum number model (Table 26). Consequently, after additive variance is generated in the AxA model, different levels of selection have varying effects on the progress of the mean, quite irrespective of the effects of random drift, until most of the additive variation is exhausted.

The discussion of the nature of the additive-by-dominance (AxD) model in the section on population size is relevant now to the fact that no differences of statistical consequence were observed among means of populations under different intensities of selection. It has been pointed out that, if selection is at all successful in moving the frequencies of a majority of genes toward 1.0, then an interaction of the additive effects at one locus with an overdominant situation at another, where the heterozygote is inferior to both homozygotes, would be the prevailing situation so that inbreeding would complement selection, possibly obscuring differences between selection intensities if they exist. Although no differences between means due to selection intensity were statistically significant, the trend illustrated in Figure 21 indicates that $\frac{1}{2}$ selection possibly is not as effective as stronger levels in moving a majority of gene frequencies quickly toward 1.0.

The dominance-by-dominance (DxD) conditional epistatic model essentially is an interaction of the two types of overdominance, heterozygote superior or inferior to both homozygotes, one type occurring at one locus, the second at another locus. At gene frequencies near .5, a balance of the two types allows random drift to be much more powerful than selection of any given intensity, but at gene frequencies removed from .5 the conventional overdominance predominates so that stronger levels of selection are more effective than the weaker intensities against random drift. This should result in increased differences over many generations among means of populations under different degrees of selection. Observations of the simulated populations corroborate this contention (Figures 4 and 22).

Multiple comparisons of the levels of selection intensity for the models in which general statistical significance of effect on the mean was observed (Figures 1-4), revealed that every statistically significant comparison of a pair of levels involved the inferior performance of populations from which $\frac{1}{2}$ were selected as parents compared to the performance of populations selected more intensely especially those selected as stringently as $\frac{1}{8}$.

It has been shown previously that selection intensities of $\frac{1}{8}$, $\frac{1}{6}$, $\frac{1}{4}$ and $\frac{1}{2}$, when combined with the various parent population sizes, specify progeny populations

ranging from 16 to 256 in number, and correspond to selected population means which are expected to be 1.65, 1.5, 1.27 and .8 standard deviations, respectively, above the mean of the unselected population. Therefore, one should expect differences in means, associated with two consecutive levels of selection, in the approximate ratio 2:3:6. Hence, it is not surprising that populations from which only the best half of the individuals are selected are inferior to those in which selection is more discriminatory.

For the simple additive case, the means of populations under the weakest selection intensity ($\frac{1}{2}$) were statistically inferior, over the first ten generations to those of populations selected quite intensely ($\frac{1}{8}$ or $\frac{1}{6}$), but they were not inferior to the means of populations with $\frac{1}{4}$ selection. However, the general trends shown in Figure 14 indicate that an intensity of $\frac{1}{4}$ is probably also better than $\frac{1}{2}$ and that major proportions of most differences are maintained far beyond 10 generations, despite the lack of statistical significance. If the trends are at all accurate, $\frac{1}{8}$ selection gives no better results than $\frac{1}{6}$ at any period of selection.

When complete dominance existed, the general significance of differences in means due to selection from generation 10 through 20, was not supported by any statistically significant pair differences, but, from generation 20

through 30, the means of populations with $1/8$ selection were statistically superior to the means of those with $1/2$ selection. Figure 15 indicates that $1/2$ selection was possibly inferior to the other levels throughout the period, and shows that differences among the three stronger intensities were, perhaps, trivial in nature.

The strong evidence, after a few generations, for significant differences among the means of populations under varying degrees of selection for an overdominant character, was supported by statistically important individual differences which showed that $1/2$ selection was an inferior force for combating random drift compared to all the higher levels of selection until generation 20, and that it was inferior compared to intensities of $1/8$ and $1/6$ thereafter (Figure 16). The general strength of selection in relation to the force of random drift may be assessed by comparing Figures 7 and 16.

No paired comparisons of differences in means due to selection were statistically significant for populations with optimum number gene action, but the trends revealed in Figure 17 are about as expected, i.e., the observed means were consistently higher for stronger selection. Inspection of the results for individual parameter sets revealed that the magnitude of genetic progress made at any intensity of selection was somewhat influenced by the

associated amount of environmental variation, especially at the high level, $3\sigma_G^2$.

As expected, the results of tests of paired differences for the complementary factor model were much like those for complete dominance. Again, 1/2 selection produced statistically inferior population means, this time after 10 generations, compared to all the higher intensities. Figure 19 illustrates that this trend began to develop from the early generations, growing in magnitude over the entire period. A possible reason for the later development of statistical significance, compared to the dominance case, is that the recessive double homozygote under complementary action is not discriminated against any more than a two-locus genotype with only one locus homozygous recessive. Thus, the power of selection to avoid the genotypic state of least genetic merit is not nearly as discriminatory as it is for the non-epistatic dominance situation.

As indicated previously, selection is a stronger force than random drift for changing the mean of a population with the additive-by-additive gene model. However, it is significantly stronger only during the period when σ_A^2 is available in non-trivial amounts, i.e., from about the 5th to the 15th generation of selection. In this interval, at least, the progress of the means of populations with 1/2

selection was less than the progress of those selected more strongly. Figure 20 shows that, subsequent to generation 15, populations with 1/2 selection tend to continue to advance in genetic merit while those which have been selected more highly tend to reach plateaus of genetic merit, so that differences due to selection are no longer statistically apparent. Whether the populations with 1/2 selected each generation would have continued to advance until they surpassed the other populations, as Dempster (1955) and Robertson (1960) have indicated they should, is now a moot point since the simulation procedure was never continued past 30 generations. Only the results from additive-by-additive cases show this interesting possibility in as brief a period as 30 generations. As noted previously, the rate of advance under this model is higher than for any other model simulated, including the simple additive one. This phenomenon probably can be accounted for by the fact that the AxA model is not identical with AxA gene action or σ_{AA}^2 except at gene frequencies of .5. At higher frequencies, most of the variance is σ_A^2 rather than σ_{AA}^2 (Figure 63), and the existence of two peaks of genetic merit (Table 26), one for double homozygotes of one allele and a second for double homozygotes of the other type of allele, enables selection to achieve its goal quickly, no matter which way the original gene frequencies

drift. Some of the populations being selected with other models of gene action might show the plateauing of genetic merit in intensely selected populations, and the subsequent recovery of populations selected less highly, more distinctly over a longer period of time. Of course, some populations may reach a selection limit or equilibrium while still retaining genetic variation, because continued selection for heterozygotes (e.g., overdominance).

Robertson (1960) showed that the optimum intensity of selection for reaching the theoretical limit in artificial selection may encompass a rather broad range of intensity for populations as large as 50. However for populations as small as 10, the theoretical optimum appears to be restricted to selection of very nearly $1/2$ of the population each generation. Robertson assumed that no linkage existed, which he suggested might lower the optimum selection intensity when it does exist, and that the quantity a/σ would remain constant throughout the selection, where a represents the average effect of a gene substitution and σ^2 is the total variance in the population. However, he pointed out that, "The smaller the number of genes contributing to any given additive genetic variance (and in consequence the greater their individual average effect) the lower will be the possible advance by selection and the quicker will it be reached." Thus, limitation of the

simulated populations to 40 loci may contribute to changing values of a/σ to the extent that the comparison of the results to the theory of limits is partially invalidated. Complete fixation at maximum genetic value did occur in some simulated populations under intense selection, but most of these were populations associated with parameter sets in which no environmental variation was involved. Other populations, involving environmental variations, appeared to be approaching non-maximum plateaus of genetic merit after 30 generations. Robertson (1955b) has discussed reasons why one may not obtain the expected response to selection. These are keynoted by the possibility of existence of genes with large or varying effects.

Effect of selection on fixation observed The discussion of size of gene effects and number of loci is related to the problem of evaluation of the effect of selection on mean fixation of alleles and mean frequency of unfixed genes (Tables 9-16). Surprisingly, in most cases, selection intensity appears to have little relation to the mean frequency of unfixed genes after several generations of selection. Apparently, fixation of alleles occurs at a large proportion of loci, because of limited population size, resulting in erratic mean frequencies based on a few loci still segregating.

In the complete dominance case (Table 9), intensity of

selection seems to have its main effect on the ratio of the numbers of favorable and undesirable alleles fixed, rather than on total amount of fixation, which was as high as 80 percent after 25 generations in some cases. The ratio of favorable to unfavorable alleles fixed was about 3:1 for a selection intensity of 1/2, 5:1 for 1/4, 10:1 for 1/6 and 15:1 for 1/8 after 25 generations of selection.

Falconer (1960) gives the difference in gene frequency between an unselected population and the selected group as $\Delta q = -ipq\alpha/\sigma_p$, where i is the standardized selection differential, $\Delta P/\sigma_p$, p and q are frequencies of favorable and unfavorable alleles, and α is the average effect of a gene substitution. This formula enables one to translate the intensity of selection into the approximate coefficient of selection against the unfavorable allele for the dominance case, $s = (Y_{AA} - Y_{aa})\Delta P/\sigma_p^2$. Thus, the coefficient of selection operating on any locus is directly proportional to the intensity of selection, and to the difference of value between the two homozygotes, expressed in terms of the phenotypic standard deviation. Unfortunately, this simple relationship does not hold for epistatic conditions. The theoretical equivalent for s , using the genotypic values simulated, is $s = 2z/b\sigma_p$, where z is the ordinate of a $N(0,1)$ curve at the truncation point indicated when a proportion b is selected. Using tabulated values of z/b (Lush,

1945) and an average σ_p over the four simulated levels of variation, the approximate \underline{s} at each locus when selection begins, in the complete dominance case, is .2, .33, .4 or .43 for selection intensities of 1/2, 1/4, 1/6 and 1/8, respectively.

Using these values for s and the table given by Lush (1945, p. 126) for the approximate time required for selection to increase the frequency of a favored gene by various amounts under random mating, one may determine that a selection intensity of 1/2 for a complete dominant should change the frequency from .5 to .8 in about 20 generations. Calculation of the mean frequency of fixed and unfixed alleles from Table 9 of the results from simulated populations, gives .71 at generation 20. Selection intensities of 1/4, 1/6 and 1/8 should require about 30, 25 and 23 generations, respectively, to change gene frequency from .5 to .9. Again, the corresponding mean frequencies of all alleles, calculated from generation 25 in Table 9, are .81, .84 and .88. These results are strikingly close to the theoretical ones considering that random drift is involved. However only 40 loci were simulated, and some of the loci were already fixed at generation 25. Therefore, selection had already achieved those frequencies at some earlier generation. Lush (1948) has shown that s should be larger for a small number of genes than for a large number.

Therefore, random drift and early fixation due to selection account for deviations from theoretical approximations in the direction of the observed results, while the effect of restriction of the number of loci on the size of s should cause deviations in the opposite direction. Consequently, the difference in frequency can be attributed to the difference in force between the conditions. The relative magnitude of each is uncertain but the discrepancies are in the direction which random drift would take them so that one may postulate that restricting simulation to 40 loci is, perhaps, not as severe as one might at first presume.

In the overdominance case, selection appears to work against the force of random drift which leads to fixation (Table 10). At the weaker selection intensities of $1/2$ and $1/4$, the amount of fixation of genes responsible for an overdominant character was nearly 50 percent in 25 generations. But when selection was as strong as $1/8$, only 38 percent of the genes were fixed. In all simulated cases, selection for a heterozygous genotype (i.e., the overdominant cases) resulted in less fixation than in corresponding populations under selection for homozygous maximums.

Selection for an intermediate genotype, which is not necessarily heterozygous, takes place under the optimum number model. The results in Table 11 indicate that, in

this case, selection is a minor contributor to fixation. This is not surprising in view of the fact that paired comparisons revealed no statistically significant differences in means of populations under different intensities of selection.

Table 12 illustrates that selection intensity of $1/2$ is essentially as effective as stronger levels in preventing fixation of most of the recessive alleles for characteristics determined by duplicate factors. Most of the fixation of these alleles probably resulted from the fact that combination of the homozygous recessive alleles at one locus with homozygous dominant ones at an interacting locus produces a genotype which is not inferior to the one which is homozygous dominant at both loci, i.e., AABB, AAbb and aaBB all have the same genotypic values. If fixation occurred at any of these three states with equal probability, the ratio of fixation of dominant alleles to recessive ones would be 2:1. The ratios which were observed, for the four selection intensities which were studied, varied from 1.75:1 to 2.25:1 at generation 10, although these increased slightly in succeeding generations. Therefore, it would seem likely that little or no fixation of alleles in the inferior aabb state took place. Examination of the arrays of genotypic frequencies at generation 25 for each parameter set revealed that not a single

pair of interacting loci was fixed in the aabb state.

An additional similarity of the complementary factor model to the complete dominance case may be observed by comparing Table 13 to Table 9. In both cases, intensity of selection appears to have its main effect on the proportion of undesirable alleles fixed rather than on total amount of fixation, which was also similar for the two models. Small differences between the models possibly are due to the fact that the double homozygous recessive genotype, aabb, of two pairs of genes acting in a complementary manner is not inferior in genetic merit to any genotype involving the homozygous recessive at only one of the loci.

The strong power attributed to selection, as a force for changing the mean under the additive-by-additive (AxA) gene model, and the plateaus of genetic merit reached under intense selection, are emphasized and explained to some degree by the results given in Table 14. Even weak selection ($1/2$) resulted in 90 percent fixation by generation 25. Intensities of $1/4$ and $1/8$ each resulted in 100 percent fixation of the genes simulated, in all populations, by the 25th generation, and 95 percent of the loci involved fixed alleles when the intensity was $1/6$. The latter figure roughly corresponds to an average of only two unfixed loci in each population considered, and the mean frequency of the

favorable allele among these was .736. Consequently, little or no genetic variation remained in the populations selected intensely for 25 generations, so that the plateaus which were reached were actually limits for selection. Because of the fact that the aabb genotype, as well as the AABB genotype, is a maximum, the proportions of A alleles and a alleles (or B and b) which are fixed do not indicate if any fixation has occurred in the AAbb or aaBB states, which are the least desirable ones. Figure 20 indicates the relative amounts of this type of fixation at generation 30 for selection intensities of 1/4, 1/6 and 1/8.

Robertson's (1960) postulation that the ultimate limit for 1/2 selection should be higher than the limit for more intense selection, receives no conclusive concurrence or denial from the results given in Table 14 and Figure 20, but it seems unlikely that the mean level of genetic merit in the simulated populations with 1/2 selection would surpass the level of those with 1/8 selection after any number of generations of selection continued beyond 30. The reason is that 90 percent of the loci in populations with 1/2 selection had already been fixed after 25 generations, whereas the mean of these populations had advanced only 84 percent as far as the mean of populations with 1/8 selection, in which all genes were fixed.

There appears to be no relation of consequence between

intensity of selection and the mean amount of fixation under the additive-by-dominance conditional epistatic model (Table 15). Therefore, inferior means associated with populations with $1/2$ selection (Figure 21) cannot be explained on the simple basis of proportion of favorable and unfavorable alleles fixed. There is a distinct possibility, though, that the force of selection at this level is not as effective as the force at higher levels in overcoming the force of random drift, which would tend to fix some alleles in the double homozygous recessive state, $aabb$. Apparently, interacting genes whose frequencies change in the direction of either of the subordinate peaks of genetic merit, which are relatively close to the values for $Aabb$ and $aaBb$ genotypes (Table 26), tend to become fixed in the $AAbb$ or $aaBB$ states with considerably higher probability than in the inferior $aabb$ state when selection is as intense as $1/4$, and with somewhat less margin of probability with $1/2$ selection.

Intensity of selection appears to be relatively unrelated to fixation of alleles in populations under the dominance-by-dominance ($D \times D$) conditional epistatic model, as it should be, because no variance other than σ_D^2 , σ_{AD}^2 or σ_{DD}^2 exists in such populations at any gene frequencies. Consequently, the approach to fixation is almost entirely due to random drift, with possible modifications due to

linkage because two crossovers are required for any interacting pair of loci to reach one of the four fixation states, AABB, AAbb, aaBB or aabb, from the original state AaBb. Knox (1962) has shown that the mean absorption time, i.e. time to total fixation of an allele at one locus in a random mating population of 20 individuals, or about 3 more than the average number simulated in this study, should be about 53.7 generations. The values for 5, 10 and 15 individuals are 12.6, 26.2 and 40 generations, respectively. Therefore, it seems likely that complete fixation of a population of alleles involving 17 individuals would occur in about 46 generations, and that fixation in 25 generations would amount to approximately 50 percent. Table 16 shows that the average proportion of total fixation which had occurred at generation 25 under selection intensities of $1/2$, $1/4$, $1/6$ and $1/8$ was .532, .481, .487 and .500, respectively. The average of these results is exactly 50 percent fixation. The mean amount of recombination or linkage in the simulated populations was almost .2, but division of the populations into groups, with distinctly different degrees of linkage between groups and a common degree of linkage within the group, failed to show any important difference in amount of fixation due to linkage at generation 25. In fact, populations which involved .005 linkage and .2 linkage resulted in exactly the same

proportion of mean fixation in 25 generations (.4875). Populations with linkages of .05 and .5 also resulted in the same amount of fixation (.5125). Consequently, the results fit Knox's theory rather well, and linkage appears to be a minor factor in the approach to fixation over a period as long as 25 generations. However, it should be pointed out that populations involving tight linkage did show less fixation at generation 5 than those with more recombination. Approximately 13 percent fixation resulted from .005 and .05 linkage, and almost 20 percent from .2 and .5. However, these differences were largely dissipated by generation 10.

Conclusions with respect to selection A general synopsis of the more germane points of discussion, concerning the relation of the force of selection to the mean genetic progress of small populations, will be presented before proceeding to the discussion of the effects of environmental variation.

Comparison of the observed results from simulated populations involving dual epistasis to results predicted from the equation developed by Griffing (1960a), as a description of the progress of the mean under truncation selection on the individual phenotype with arbitrary dominance, epistasis, linkage and number of loci, revealed the nature of the limitations imposed upon the equation by assumptions which were necessary for the mathematical derivation. The

basic assumptions were that the populations involved are infinite in size, that the genetic population parameters are constant and do not change from generation to generation. A general impression received from examination of the results from simulated populations is that the predicted contribution to change in the mean attributed to σ_{AA}^2 under tight linkage, while possibly accurate for a generation or two, is generally far too large over several generations of selection. Random drift due to finite size of population appears to have considerable influence in changing the genetic parameters drastically in a few generations. Selection also changes these parameters so that extrapolation over several generations is quite erratic. However, the magnitude of the discrepancies noted between predicted means and those observed in Monte Carlo results probably is larger than it would be in an applied breeding situation because of restrictions in the mechanics of simulation. The primary restriction is limitation of the number of loci and consequent enforcement of a situation involving individual gene effects which are large relative to the amount of total phenotypic variation. This situation might be remedied by simulating more environmental variation, so that the dynamics of changing parameters would more closely resemble the changes which probably occur with most quantitative characters.

In general, selection was rather effective in advancing the genetic mean of small populations under all models of gene action in which the genotype of highest merit is homozygous. However, selection was ineffective, or at least weaker than random drift, in small populations under mass selection for a character which involves only a heterozygous genotype as optimum.

Major differences in genetic means were produced by different intensities of selection when a single peak of genetic merit existed, i.e., when the selection goal involved only one particular genotype (e.g., Additive and Overdominance models) or a small contiguous group of genotypes including only one maximum fixation state (e.g., Complete dominance and Complementary factor models). Minor differences may be observed over relatively short periods of a few generations when two homozygous peaks of genetic merit exist (e.g., Optimum number and $A \times A$ models), but no important differences resulted from different levels of selection when multiple goals existed (e.g., Duplicate factor, $A \times D$ and $D \times D$ models).

The presence of large amounts of dominance variance intensified the differences in mean genetic progress of populations under different levels of selection.

The regression of genetic merit on generation number was negative for all parameter sets involving populations

simulated with overdominance or DxD gene models. Negative regressions were also observed for populations with the two parameter sets which involved the smallest population size, 8, in combination with low levels of selection intensity (1/2, 1/4) and high levels of environmental variation (σ_G^2 , $3\sigma_G^2$) when complete dominance existed, and for one of these combinations in each of two other models, the complementary factor and optimum gene number models. These results indicate the inability of selection to overcome the force of random drift under the associated conditions.

Multiple comparisons of the levels of selection intensity, for the models in which general statistical significance of effect on the mean was observed, revealed that every statistically significant comparison of a pair of levels involved the inferior performance of populations from which 1/2 were selected as parents compared to the performance of populations selected more intensely, especially those selected as stringently as 1/8. This observation is in accordance with the approximate ratio, 2:3:6, of expected differences in means of two populations under two consecutive selection intensities in the array (1/8, 1/6, 1/4, 1/2), i.e., the difference in mean due to selecting 1/4 instead of 1/2 is expected to be even larger than the difference due to selecting 1/8 instead of 1/4. This phenomenon was especially emphasized under complete

dominance and complementary gene action after several generations of selection, probably because of the preponderance of dominance variance as a proportion of total genotypic variance at gene frequencies above .7.

Under the additive-by-additive (AxA) conditional epistatic model genetic progress was even more rapid than that achieved in the simple additive situation. This phenomenon was explained by the fact that the AxA model contains chiefly additive variance instead of σ_{AA}^2 except at gene frequencies of .5, and by the fact that two peaks of genetic merit exist. Selection appeared to be stronger than random drift, as a force for changing the mean, only during the period when additive variance became available in more than trivial amounts, i.e., from generation 5 to 15 when selection was intense. Subsequent to generation 15, populations with 1/2 selection tended to continue to advance in genetic merit while those selected more highly tended to reach plateaus of genetic merit, so that differences due to selection were statistically apparent. The postulations of Dempster (1955) and Robertson (1960) that the ultimate limit for 1/2 selection should be higher than the limit for more intense selection did not receive conclusive concurrence or denial from the results of simulated populations, but it seems unlikely that the mean level of genetic merit in the simulated populations under the AxA model and 1/2

selection would surpass the level of those with $1/8$ selection after any number of generations of selection continued beyond 30. The reason is that all of the loci in populations with $1/8$ selection had already been fixed after 25 generations resulting in a permanent genetic plateau of known level, whereas the mean of populations with $1/2$ selection had advanced only 84 percent as far, but 90 percent of the loci already were fixed. However, limitation of the simulated populations to 40 loci may contribute to changing values of the average gene effects as proportions of the total variance, a/σ is assumed to remain constant, is partially invalidated.

In most cases selection intensity appeared to have little relation to the mean frequency of unfixed genes after several generations of selection, regardless of the gene model involved. Apparently, fixation of alleles at a large proportion of loci results in erratic mean frequencies based on the few loci still segregating.

Using formulas given by Falconer (1960) for translating the intensity of selection into the approximate coefficient of selection against the unfavorable allele for the complete dominance case, and by Lush (1945) for

calculating the approximate time required for selection to increase the frequency of a favored allele by a given amount, it was observed that the mean gene frequencies of small simulated populations, which were selected for several generations, were strikingly close to the theoretical values, considering that random drift was involved. Partial cancellation of the deviations due to random drift could occur because of the effect of the small number of loci on the size of the coefficients of selection.

In the complementary factor situation, as well as for the complete dominance model, intensity of selection appears, primarily, to affect the proportion of undesirable alleles fixed rather than the total amount of fixation.

In all overdominant cases, selection for the maximum heterozygous genotype resulted in less fixation than in corresponding populations under selection for homozygous maximums.

Selection in the optimum number model, where maximum genotypes are not necessarily heterozygous, is only a minor contributor to fixation in relation to random drift.

Examination of genotypic frequencies after 25 generations of selection under the duplicate factor model revealed that no fixation had occurred, which involved any pair of interacting loci in the inferior aabb state, in populations associated with any parameter set. Thus,

because there are no other genetically inferior genotypic states associated with this model, $1/2$ selection is as effective as stronger intensities.

Under the additive-by-dominance (AxD) conditional epistatic model interacting genes whose frequencies change in the direction of either of the subordinate peaks of genetic merit, which are relatively close to the values for the Aabb and aaBb genotypes, tend to become fixed in the AAbb or aaBB states with considerably higher probability than in the inferior aabb state, when selection is as intense as $1/4$, and with somewhat less margin of probability with $1/2$ selection. This explains the somewhat lower means observed in populations under the latter intensity of selection. This difference in means cannot be explained in terms of selection achieving more fixation in the AABB state, i.e., in the primary peak of genetic merit. Thus, dominance variance is again implicated as an intensifier of differences due to selection intensity.

The approach to fixation in populations under the dominance-by-dominance (DxD) conditional epistatic model is relatively free from the effects of selection because linear and non-linear additive components of genotypic variance do not exist in such populations at any gene frequencies. Consequently, the approach to fixation is almost entirely due to random drift, with possible modifications

due to linkage. The average proportion of total fixation which occurred in simulated populations after 10 or more generations of selection corresponded almost exactly to values derived from tables of mean absorption time given by Knox (1962), regardless of the amount of linkage involved. However, after only five generations of selection, the amount of fixation in populations with tight linkage (.005, .05) was only 65 percent of that observed in populations with more recombination (.2, .5).

Environmental variation

Theoretical considerations Wright (1918, 1920) and Fisher (1918), independently of each other, developed the first comprehensive techniques of dealing with the statistical separation of environmental and genetic variation in general populations. Geneticists, however, have tended to gather all unexplained sources of variation into one convenient term which they think of as an error variance rather than as environmental variance. Perhaps this occurs because the environmental variation is difficult to measure or control, and the factors responsible are difficult to identify.

Some of the problems of fundamental importance, concerning environmental variation, cannot be studied easily with the Monte Carlo technique, but an appreciation of the consequences of some of the major phenomena may be important in relating the results from simulated populations to the

problem of prediction in real ones.

Dempster and Lerner (1950) pointed out that if the environmental variance is not independent of the average genotypic value, the regression of genotype on phenotype may be non-linear. Therefore, the proportion of the phenotypic selection differential which will be realized will depend on selection intensity and other factors. This will generally lead to bias of the estimate of genetic gain.

Comstock (1960) showed that information about the relative magnitude of different components of genetic variance, obtained from an experiment confined to one stratum of the pertinent environments, will be biased by variance due to genotype-environment interaction.

Whether genotype-environment interactions in actual populations are important factors in reducing heritability (e.g., Mason, et al., 1958) or not (e.g., King et al., 1959) depends upon the species and characteristics being investigated, and upon the specific environments in which the investigations take place. Such sources of bias in prediction equations have not been simulated in this study, although the Monte Carlo technique may prove to be valuable in assessing the nature of such bias in future investigations.

Robertson and Reeve (1952a) found that highly inbred lines of *Drosophila* tend to have much higher variance in body size than do crosses between them, so that the

environmental variance is not constant for all genotypes and tends to be smaller in heterozygotes than in homozygotes, regardless of whether the crosses were made from selected or unselected lines. They concluded that it seems probable that many quantitative characters would show the same tendency for environmental variability to decrease as heterozygosity increases if the individual genes have small effects. They also suggested that heterosis and reduced susceptibility to environmental variations are both manifestations of the same phenomenon of heterozygosity. However, other investigators have been unable to find dependence of environmental variance on heterozygosity. Falconer (1960) summarized these views and, pointed out that we must recognize the possibility that the environmental variance may be changed if the frequencies of genotypes are changed by selection or some other means. Wright (1920) observed essentially the same phenomenon in the amount of spotting in guinea pigs many years before. In a later paper, Reeve and Robertson (1953) concluded that environmental variables affecting only one character appear to differ less in their effects on homozygotes and heterozygotes than do those with effects on two or more characters. Therefore, the results from populations simulated in this study probably correspond more closely to those which one would expect for quantitative characters which are

particularly influenced by specific environmental factors, rather than to results for characteristics such as vigor, health or longevity which are influenced primarily by general environmental factors.

The variation caused by genetic segregation is basically discontinuous. However, the phenotypic distributions for most quantitative characters are continuous because of the simultaneous segregation of many genes, perhaps hundreds, which affect the character, and because of variation arising from non-genetic causes. The populations which have been simulated in this study involve only 40 loci which are presumed to affect a single character. Also the environmental variation which has been simulated is limited to an amount which is a function of the total genotypic variance due to the 40 loci which are segregating in the initial population. Therefore, the magnitude of the average gene effects, relative to other sources of variation, probably is somewhat larger in the simulated populations than it is for most actual metric characters. In fact, some of the simulated populations were selected in the absence of environmental variation in order to study the nature of the difference in genetic progress with and without environmental variation. This situation, in combination with the fixation of a high proportion of loci after several generations of selection, may lead to recognizable discontinuity

in the phenotypic distribution.

Monte Carlo differences in means Figures 1-4 show that major differences in population means over a considerable period of time were produced by selection in the presence of different levels of environmental variation when a single homozygous peak of genetic merit existed (e.g., Additive, Complete Dominance and Complementary Factor models). Minor differences were observed over relatively short periods of a few generations for other models. In general, differences in the amounts of simulated environmental variation between populations were important in affecting the relative progress of the mean only when these differences were expressed in populations in which the mean was changing rapidly because of other forces. In most cases, these periods coincided with the periods in which selection was the most effective. In the discussion concerning selection, it was shown that differences in means due to selection intensity were important, primarily, when the highest relative amounts of additive variance were available. This implies that the best progress was made under the highest effective heritabilities, because the environmental variance was constant, i.e., the environmental deviations were drawn from the same distribution in each generation.

In the additive case, consideration of differences in

means between populations associated with all possible pairs of levels of environmental variation, showed that the most important difference involved populations in which there was no environmental variation in contrast with populations which received an amount which was three times as large as the original genotypic variance ($3\sigma_G^2$). This difference was relatively significant, statistically, for about 20 generations, i.e., until most of the populations were approaching plateaus of merit due to fixation and exhaustion of additive variation (Figure 23). However, in the early stages of selection, the difference between means of populations with $\sigma_E^2 = \sigma_G^2/3$ and those with $\sigma_E^2 = 3\sigma_G^2$ for environmental variance, was also statistically significant. If the trends are at all accurate, it appears that environmental variance equivalent to $3\sigma_G^2$ is quite detrimental to progress relative to lesser amounts over many generations of selection, as one would expect.

Under complete dominance, the means of populations with environmental variance equivalent to σ_G^2 or $3\sigma_G^2$ were statistically inferior to those of populations with only genetic variance after five generations of selection (Figure 24). These differences persisted for the entire 30 generations of selection, much in the same manner that selection differences persisted. Bohidar's (1960) statement concerning the effect of dominance on the sensitivity of selection

might, therefore, be modified to read: "Dominance makes selection sensitive longer"--especially when the environmental variation is high. Unlike the additive case, the trends indicate that $\sigma_E^2 = \sigma_G^2$, as well as $\sigma_E^2 = 3\sigma_G^2$, is especially detrimental to genetic progress. Because of the fact that additive genetic differences are partially obscured by dominance, one may suspect that less environmental variance is required to complete the obscurity, and thus slow genetic progress, than is required in the additive case to make the same effect on progress. Baker and Comstock (1961) observed similar results consistently only when selection intensity was high (1/10). This dependence on high selection intensity was probably due to the fact that they simulated variance due to the combined effect of the environment and those genes which were not simulated directly, but which were postulated to affect the quantitative character in question. Martin and Cockerham (1960) simulated environmental variance equal to the expected additive genetic variance in the initial population and compared the results to those from populations with only genetic variance. In populations with 20 loci, their results were quite similar to those from this study, but, in populations with only 6 loci, they observed slowed progress due to σ_E^2 in the additive case but not with dominance.

In the overdominant situation, differences among means

due to level of environmental variation were much slower in attaining statistical significance than they were under complete dominance (Figures 1 and 2). If one compares the theoretical slope of the regression of the mean on gene frequency (Figure 59) to the observed slopes of the regressions of the means on generation number under different levels of environmental variation (Figure 25), it may be observed that gene frequency increased very rapidly at first because of random drift and in spite of the force of selection. However, as gene frequency reaches .7 and higher, the proportion of additive variance in the population increases considerably (Figure 59) so that selection begins to counteract the force of random drift more effectively, and the decline in the mean is slowed (Figure 25). Therefore, the various amounts of environmental variability which are present in the populations finally begin to act differentially upon the change in the mean under selection. The result was that populations without environmental variation were able to achieve an equilibrium in 25-30 generations, while the others could not. Bohidar (1960) observed this same equilibrium, or "valley" of merit, in populations under selection for an overdominant character without environmental variation.

The optimum number model was the only epistatic model which involved statistically significant differences in

means of populations which could be accounted for by the difference in a pair of levels of environmental variation, although general differences due to non-genetic variance were statistically significant at times in populations under the complementary and DxD models (Figures 2-4). The differences in means of populations with only genetic variance and those of populations with $\sigma_E^2 = 3\sigma_G^2$ were statistically important between generations 5 and 20 but not afterward (Figure 26). After generation 20 the populations with non-genetic variation continued to progress in mean value, while those without it tended to reach a plateau of genetic merit. A check of the populations associated with individual parameter sets showed that the plateaus were permanent, due to complete fixation, when selection intensity was high (1/6, 1/8), and that the genotypic variation was nearly exhausted by generation 20 in the population with 1/4 selection. However, in the population with 1/2 selection and tight linkage (.005), a slightly fluctuating plateau at a mean value of approximately 210 occurred between generations 20 and 30. A considerable amount of genotypic variance was still available during this period, varying from about 6 to 16, i.e., 20-50 percent of the original variation. If Figure 60 is at all accurate in representing the true partition of variance, additive variation should account for 30-65 percent of the genotypic

variance available at those levels, and most of the rest of it should be σ_{AA}^2 . This indicates that the plateau probably is not caused by lack of additive variation, i.e., it is not likely that the plateau could be due to epistasis alone. A more probable conclusion is that a combination of factors is involved, i.e., epistasis, the force of random drift opposing weak selection, and the presence of tight linkage. A check of the genotypic frequencies at generation 20 and 25 showed that only eight loci were still segregating. Three of these were adjacent on the first chromosome, and thus were tightly linked. The other five were linked adjacently on the third chromosome. The frequencies of the alleles at all of these loci were between .3 and .7 at both times, indicating that much of the variance was probably epistatic (Figure 60). Perhaps selection is ineffective in utilizing the additive variance which does exist, because the lack of environmental variation allows a situation to exist which is almost completely balanced in its effect on the mean, both genetically and phenotypically, until a crossover occurs. If this is true, some of the plateaus which Bohidar (1960) observed in populations under tight linkage may have been due to a similar balance without environmental variation under non-epistatic conditions.

Differences among means of populations which involved duplicate factors and different levels of environmental

variation never were statistically significant (Figure 3) although Figure 27 indicates that many individuals in populations with $\sigma_E^2 = 3\sigma_G^2$, probably carried the aabb inferior genotypes at several loci, at least during the first few generations of selection. In fact, more than 11 percent of all loci in such populations were fixed for the inferior allele by the fifth generation, compared with less than four percent fixed in populations with genetic variance only. However, after 20 generations of selection, the corresponding amounts were 20 and 18 percent. Thus, it appears that large amounts of environmental variation are detrimental to the progress of the mean by selection following segregation of the original heterozygous genotypes, but that the damage to the potential progress of the population over a long period of time is of little consequence.

With complementary gene action, differences among means due to environmental variation were statistically significant over most of the period of selection (Figure 3). However, tests between all possible pairs of environmental levels revealed no differences of statistical importance. This is another instance where many small differences, which were not deemed worthy of being called statistically significant, can add up to general significance collectively (Figure 28). It should be noted that these differences are

not as large as the corresponding ones under complete dominance (Figure 24). Again, the differences between models probably is due to the difference in merit associated with the aabb genotype (Table 26).

Under the AxA model, differences due to environment approached statistical significance only for the relatively short period when selection was acting on the maximum variance which had been generated by changes in gene frequency over the first ten generations of selection (Figure 4). Most of the difference at that time (generations 10-15) appears to be due to the efficiency of selection in advancing the means of populations which involved no environmental variation (Figure 29).

No differences of statistical consequence were caused by environmental variation in populations with the additive-by-dominance model, but Figure 30 indicates that there is a strong possibility that progress in populations with $\sigma_E^2 = 3\sigma_G^2$ is consistently inferior to that of populations with lesser amounts of non-genetic variation.

Figure 4 indicates statistical significance of differences among means due to environmental variation only after 25 generations, for populations involved with the DxD model. The similarity of the effect of this interaction of types of overdominance, heterozygote superior or inferior to both homozygotes, to the effect of conventional over-

dominance without epistasis is apparent if one compares Figures 25 and 31. The same approach to an equilibrium or "valley" of genetic merit for populations without environmental variation may be observed in both cases.

Differences in the mean proportions of loci fixed and in the mean gene frequency of unfixed loci, accounted for by the level of environmental variation, were negligible except in those cases which have been mentioned previously.

Conclusions with respect to environmental variation

Three major complications are involved in the partitioning of the variance of actual populations into genotypic and environmental components. They are (1) correlation between genotypic value and environmental deviation, (2) genotype-environment interaction and (3) dependence of environmental variance on genotype. These are the major sources of error in the partition of variance and may well be the major risks involved in relating the results due to environment in simulated populations to the problem of prediction in real ones, because the only reasonable method of simulating environmental variation is to make it uniform from generation to generation and completely independent of genotype. It is uniform in the sense that the environmental deviations differ from individual to individual in a random manner and the amount of environmental variation within the total population is constant in each generation.

If one limits the amount of simulated environmental variation to some function of the genotypic variance associated with a small number of simulated loci, the magnitude of the average gene effects, relative to other sources of variation, is somewhat larger in simulated populations than it probably is in most actual metric characters. Therefore, the dynamics of change may be somewhat different too.

If environmental variation is not simulated, and one limits the number of loci simulated rather severely, recognizable discontinuity in the phenotypic distribution may result after several generations of selection because of fixation of a high proportion of loci.

Differences in population means over a considerable period of time were produced by selection in the presence of different levels of environmental variation for additive, complete dominance and complementary factor models. Minor differences were observed over relatively short periods of time for other models. In general, differences in the amounts of simulated environmental variation between populations were important in affecting the relative progress of the mean only when these differences were expressed in populations in which the mean was changing rapidly because of selection.

With complete dominance, less environmental variation

was required to slow genetic progress than was required in the additive case to make the same effect.

In the overdominance situation, various amounts of environmental variation did not begin to act differentially upon the change in the mean under selection to an important extent for several generations, i.e., not until random drift had caused gene frequencies to deviate strongly from .5 so that the proportion of additive variance was considerably increased. Populations without environmental variation were able to achieve an equilibrium in 25-30 generations, whereas the others could not.

The optimum number model was the only epistatic model in which the difference in means, due to a pair of levels of environmental variation, was statistically significant. For other epistatic models, it appears that large amounts of environmental variation are detrimental to the progress of the mean by selection following segregation of the original heterozygous genotypes, but that the damage to the potential progress of the population over a long period of time is of little consequence.

With the optimum number model, populations with non-genetic variation continued to advance in merit after 20 generations of selection, whereas those without it tended to reach a plateau of genetic merit. These plateaus were permanent, because of fixation, when selection intensity

was strong, but a population with $1/2$ selection and tight linkage (.005) developed a slightly fluctuating plateau even though 20-50 percent of the original genotypic variance was still available. The plateau was postulated to be caused by the combined effects of epistasis, random drift opposing weak selection, and tight linkage.

Differences in the mean proportions of loci fixed and in the mean frequency of unfixed loci, accounted for by the level of environmental variation, were negligible in most cases.

Linkage

Theoretical considerations It has been known for many years that, over a long period of time, all loci in a population tend to approach random combination. However, linkage is important as a short-term phenomenon because it slows the rate of approach to that random condition. If the double heterozygotes which are in coupling and repulsion phases are not equally numerous, the number of cross-overs from each phase to the other will also be unequal. As described by Lush (1948), for example, in the case of random mating without selection, the gametic array goes c of the way each generation from where it is toward the equilibrium condition in which the genes are combined at random, c being the recombination fraction. The importance of this decay in disequilibrium frequently is overlooked. For ex-

ample, a correlation between two characters is sometimes assumed to be evidence of linkage between the genes concerned, whereas association between characters is more likely to be evidence of pleiotropy, because any association due to linkage is not likely to persist for many generations even if the linkage is close. However, Robertson and Reeve (1952b) pointed out that the distinction between a pleiotropic effect and a linkage effect may often be unimportant in practice, since it may be easier to modify the effect by selection of the genetic background than to eliminate it by breaking a tight linkage if that exists.

Lush (1954) has summarized the main reasons why linkage disequilibrium may exist. They are (1) recent emergence of the population from a cross of divergent strains, (2) increase in the proportion of repulsion combinations due to selection (especially, selection for a genetic intermediate) and (3) excess of coupling combinations because of positive assortive mating. The second of these factors is the one which is of special importance to the theory of mass selection. Baker and Comstock (1961) found that, when dominance is present, selection contributes to linkage disequilibrium in some simulated populations but does not in others.

Mather's (1943) theory of "polygenic balance" is based

on the idea of selection favoring intermediate values of metric characters. Intermediate individuals tend to carry more loci in the repulsion phase than in the coupling phase. Therefore, selection will favor repulsion chromosomes and tend to build up balanced combinations of genes which contribute the minimal amount of variance. According to Mather, this allows "potential" genetic variability to be stored in latent form. However, selection must be strong enough to maintain the balance against the continuous recombination which occurs, if genetic variability is to be affected significantly.

The partitions of genotypic variance given by Kempthorne (1954) and by Cockerham (1954) are accurate only under the conditions of random mating and no linkage. Cockerham (1956) stated that linkage causes a bias in the estimation of epistatic components from the covariance of relatives, but not in the estimation of the additive (σ_A^2) or dominance (σ_D^2) components unless disequilibrium exists. However, some recent unpublished results by Schnell (1961) show that, if epistasis is present, the estimates of additive and dominance components are biased by linkage, even in equilibrium populations, except for the particular case of estimation of additive variance from the covariance of parent and offspring.

Robinson and Comstock (1955) have derived the extra

contribution which linkage disequilibrium makes to the additive and dominance components of genotypic variance in the absence of epistasis. The contribution to the additive component is

$$\sum_i \sum_j (pt-rs) [1+(1-2q_i) a_i] [1+(1-2q_j) a_j] u_i u_j,$$

where p , r , s and t are the frequencies of gametes $B_i B_j$, $B_i b_j$, $b_i B_j$ and $b_i b_j$, respectively, $a_i u_i$ is the deviation of the heterozygous effect from the mean of the homozygous effects at locus i , u_i is half the difference between the homozygotes, and q_i is the frequency of the favored allele at locus i . The quantity $(pt-rs)$ expresses the linkage disequilibrium. The contribution of disequilibrium to additive variance is positive for an excess of coupling heterozygotes, and is negative for an excess of repulsion heterozygotes. The contribution to the dominance component is $2 \sum_i \sum_j (pt-rs)^2 a_i u_i a_j u_j$. This quantity is always positive.

In populations under selection, any linkage disequilibrium which exists is almost certain to make a negative contribution to σ_A^2 , relative to the amount which would occur without linkage, because selection tends to increase the proportion of repulsion heterozygotes. Therefore, the sign of the effect on total genotypic variance (σ_G^2) will depend on whether disequilibrium disturbs σ_A^2 more than σ_D^2 or vice versa. If the observed σ_G^2 in a simulated random mating

population, with selection and linkage but no epistasis, is less than the amount computed from the gene frequencies of that population ($\hat{\sigma}_G^2$), then estimates of σ_A^2 , made by the formulas of Kempthorne (1954) or Cockerham (1954), probably will be biased upward more than the estimates of σ_D^2 are biased downward by linkage disequilibrium. If σ_G^2 is larger than $\hat{\sigma}_G^2$, estimates of σ_A^2 probably will be biased upward less than those of σ_D^2 are biased downward. However, in this latter situation, inbreeding due to the finite size of the population also could make σ_G^2 larger than it would be in a truly random mating population. Under inbreeding, individuals which are heterozygous at a particular locus will also be heterozygous for a segment of chromosome in which the locus lies, because of linkage. The length of such segments at any given time depends upon the degree of inbreeding and the recombination value. Fisher (1949) has treated this subject in some detail. Linkage, in this manner, modifies the effect of inbreeding on total genotypic variance much in the same way it modifies the approach to equilibrium.

Comparisons of the observed and estimated genotypic variance were made for populations of the largest size available (32), one with tight linkage (.005) and one with free recombination (.5), for the complete dominance and overdominance cases. Results for generations 5 and 15 are:

	Generation 5		Generation 15	
	σ_G^2	$\hat{\sigma}_G^2$	σ_G^2	$\hat{\sigma}_G^2$
Dominance (.005)	23.9	16.0	7.3	6.0
(.5)	30.8	21.5	12.6	10.0
Overdominance (.005)	72.9	39.8	77.9	39.6
(.5)	57.6	39.9	56.5	39.8.

The results for dominance are essentially similar with and without linkage, i.e., the ratio $\hat{\sigma}_G^2/\sigma_G^2$ is nearly the same in both cases. The magnitude of variance observed at a given generation in the two populations varies because of the effect of different selection intensities on exhaustion of additive variation. It appears that inbreeding, not linkage disequilibrium, is responsible for the error in estimating σ_G^2 when one uses the standard formulas based on gene frequencies of conceptually random mating populations. However, in the overdominance situation, the ratio $\hat{\sigma}_G^2/\sigma_G^2$ is much smaller for populations with tight linkage than for those with free recombination, indicating that linkage disequilibrium due to selection is the source of considerable bias. The direction of the bias indicates that the effect of linkage disequilibrium on σ_D^2 probably is much larger than the effect of σ_A^2 . It appears that the magnitude of the effect of inbreeding and that of linkage disequilibrium are about equal, although the difference in the results at generation 5 and those at generation 15 indicates that the disequilibrium may increase over longer periods of selection

for the genetic intermediate.

Except for the paper by Kimura (1958), there has been little attempt to extend the mathematical theory of selection to include both linkage and epistasis. Griffing (1960b), however, has developed a method for estimating the average recombination value in populations under selection in order to adapt his generalized two-locus theory (Griffing, 1960a) to more complex genetic situations. His derivation involves the assumption that the effects of epistatic interactions involving three or more loci are negligible. The index of average recombination is $\bar{y} = (p+m-1)/2(p+m)$, where there are m chromosome pairs and p is the average number of chiasmata per nucleus. He assumed (1) that there are a very large number of active loci scattered at random over the chromosome set, (2) the chromosomes are not drastically different in size, and (3) a chiasma invariably is associated with a genetic cross-over. Griffing gave a simpler formula, $\bar{y} = (m'-1)/2m'$, for obtaining an estimate of the average recombination value, which is a rough approximation. The value m' is twice the haploid chromosome number. The average recombination value then can be incorporated into the prediction equation given previously (Griffing, 1960a):

$$\Delta\mu = (\Delta P/\sigma_P^2) [n\sigma_A^2 + \sum_{i=1}^n (1-\bar{y})^{i-1} \sigma_{AA}^2 / 2].$$

Monte Carlo differences in means

The effect of dif-

ferent levels of linkage on the differences of means observed among simulated populations was statistically significant in only one case; during the first five generations under the dominance-by-dominance model (Figure 4). The effect reached the probability level, $\alpha = .25$, for only two other models, the overdominance case (Figure 2) and the complementary model (Figure 3). If the effect of linkage is important at all in these two cases, it is more important after 15 generations of selection. In both the overdominance and DxD situations, it appears that tight linkage (.005) is somewhat more effective than levels which involve more recombination, in hindering the depression of the mean after a long period of selection (Figures 34 and 40), although it appears to be effective somewhat sooner under overdominance than in the DxD case. It is probable that this delaying action is caused by a buildup of linkage disequilibrium due to selection for a genetic intermediate. Conversely, in the complementary factor case (Figure 37), it appears that tight linkage (.005) is a hindrance to genetic progress. This result is quite reasonable because some cross-overs are required in obtaining the two-locus homozygous genotypes, which are desired, from the completely heterozygous original population. Practically all of the damage to the eventual progress of the mean was done in the

first 10-12 generations.

There were no significant differences in means due to linkage in populations under the additive or complete dominance models (Figures 32 and 33). Baker and Comstock (1961) also observed that linkage did not impair progress of the genotypic mean in simulated populations with dominance.

There were no statistically significant differences in means due to linkage in populations under the optimum number model, but the results shown in Figure 35 indicate that tight linkage (.005) may be responsible for slow progress in genetic merit over the first 10-15 generations. Since the superior genotypes involve a preponderance of repulsion phase combinations, tight linkage probably hinders the rate of obtaining cross-overs from the coupling phases which existed in the original population. Figure 36 illustrates the mean genetic progress by level of linkage with duplicate gene action. As in the complementary factor situation, tight linkage appears to slow genetic progress at first, but, unlike the result in the complementary case, the effect on potential progress seems to be negligible. The basic difference in the two models, i.e., the relative merit of the AAbb and aaBB fixation states (Table 26), probably is responsible for the observed differences in means. Differences in means due to linkage in populations under the AxA

and AxD models were not statistically significant, and no particularly enlightening trends are evident in Figures 38 and 39.

The mean proportion of loci fixed and the mean gene frequency of unfixed loci were essentially unaffected by different levels of linkage. Baker and Comstock (1961) observed less fixation in simulated populations with tight linkage than in those with free recombination, using the complete dominance model, but the difference was not statistically significant in many cases. In the complete dominance cases in this study, the proportions of loci fixed after 30 generations of selection were nearly identical in populations with recombination values of .005, .05, .2 or .5.

Achieved and Predicted Selection Differentials

The selection differential is a measure of the selection pressure applied to a population. It is the mean phenotypic value of the individuals selected as parents, expressed as a deviation from the population mean. The magnitude of the selection differential (ΔP) depends on two major factors. These are the proportion of the population included among the selected group (b) and the phenotypic standard deviation of the character (σ_p). Different methods of selection can be compared by using a generalized measure

of the selection differential, $\Delta P/\sigma_P$, which usually is referred to as the intensity of selection, i . Thus, the usual linear prediction equation, $\Delta\mu = \Delta Ph^2$, may be written $\Delta\mu = i\sigma_P h^2$.

For a constant percentage, b , of individuals selected, and constant σ_P , the intensity of selection for any one gene becomes weaker the more genes there are which affect net merit. In addition, dominance, epistasis and environmental effects weaken the intensity of selection for each gene. In this study, the intensity of selection for individual genes probably is much stronger than it is on the average in actual populations, because only 40 loci were simulated. However, the intensity may be similar to that associated with genes which have large effects in real populations.

If the distribution of phenotypic values is normal, the theoretical selection intensity can be determined from tables of the normal curve of error. For populations larger than 50, the mean of the truncated portion of a normal distribution is a good approximation to the expected mean of the selected population. Thus, for a population distributed as $N(\mu, \sigma_P)$, the expected selection differential is $(z/b)\sigma_P$. The ordinate $z = (1/\sqrt{2\pi})e^{-x^2/2}$, at the point $x = t$, where t is the deviation in standard units of the lowest selected phenotype from the mean of a

normal distribution having zero mean and unit variance. The proportion of individuals selected is given by $b = \int_t^{\infty} z dx$. Values of z/b have been tabulated by Lush (1945) and by others. Falconer (1960) has illustrated how z/b varies with b . The relation is nearly linear for $.2 < b < .8$. However, for other values of b , the curvilinearity is important in comparing selection plans. For example, if five percent of the individuals in a population are selected, $z/b = 2.06$. If ten percent are selected, $z/b = 1.75$. Thus, selection which apparently is twice as strong, results in a selection intensity which is only about 20 percent larger.

Lush (1948) has pointed out that slight departures from normality are common in biology, so that the second decimal of z/b values frequently is inaccurate in an actual case, and even the first decimal may be a little wrong. If the actual distribution is flat-topped, with fewer extremes than under a normal curve, then extremely heavy culling will have less effect than indicated by the normal values, while moderate culling is a bit more effective. Such discrepancies may be more important if selection is unequal in the two sexes, as it usually is in practical animal breeding situations. The proportions selected equally in each sex in the simulated populations in this study ($1/8$, $1/6$, $1/4$, $1/2$) all are rather moderate selection intensities compared to the proportions of males

sometimes selected in actual breeding programs. For example, 1/6 selection rather closely corresponds to selecting two percent of the males and 60 percent of the females, and 1/8 selection corresponds to one percent males and 60 percent females saved, respectively. However, in populations with unequal numbers of parents, the inbreeding effects are not related to the equal number case in the same manner as the selection differentials are related.

When selection is made from a small number of individuals, the mean deviation of the selected group is a little smaller than indicated by z/b . For populations of size 50 or less, the intensity of selection may be found from tables of deviations of ranked data (Fisher and Yates, 1943, Table XX). This table gives the average deviate of the r th largest of samples of n observations drawn from a normal distribution having unit variance. The deviation is

$$\mu_r = \int_{-\infty}^{\infty} \frac{n!}{(r-1)!(n-r)!} p^{n-r} q^{r-1} x z dx, \text{ where } z \text{ is the ordinate}$$

of the normal curve, and p and q are the probabilities of falling short of and exceeding x . For upper truncation selection, the expected mean of the selected population, when $b = r/n$, is the mean of the deviations for the r th largest observation and the $r-1$ observations which exceed it. This value roughly corresponds to the z/b value for large populations but it is always smaller.

Nordskog and Wyatt (1952) drew several random samples of 5, 10, 20, 30, 40 and 50 individuals from a large (4,688) poultry flock. They selected the best 20 percent from each of these groups of individuals for body weight at eight weeks, and compared the observed and expected selection differentials. They observed a slight skewness of the observed distribution, but found that, although the agreement with expected values was not good, sampling errors could explain most of the discrepancies. They illustrated the influence of sample size and selection intensity on the ratio of selection differentials from a finite sample and an infinite population. For any given population size, the ratio was closest to unity when moderate proportions (.2 to .8) were selected. They concluded that, "Compared with selection intensity as a factor influencing selection pressure, size of population is only a secondary force."

The expected selection differentials which correspond to the proportions selected ($b = 1/2, 1/4, 1/6, 1/8$) and the total progeny population sizes which were simulated in this study are:

<u>N</u>	<u>1/2</u>	<u>1/4</u>	<u>1/6</u>	<u>1/8</u>
4	0.66			
8		1.13		
12			1.37	
16	0.76			1.52
32	0.78	1.24		
48			1.46	
> 50	0.80	1.27	1.50	1.65

For $N > 50$, the approximate spacing ratio of the proportions selected is 6:3:2, i.e., $(z/.25 - z/.5)$ is about twice $(z/.167 - z/.25)$ and three times $(z/.125 - z/.167)$. This is the main reason why means of populations with 1/2 selection consistently were significantly inferior to the means of populations selected more intensely under most gene models, whereas the differences between other levels of selection rarely were statistically significant.

Departures from normality in the distributions of the phenotypes simulated in this study may be expressed in terms of the differences between achieved and expected, or predicted selection differentials. The quantity $(z/b - \Delta P/\sigma_p)$ expresses this difference in phenotypic standard deviations, or standardized units. In any given population under selection, this quantity was calculated in each of 30 generations, or fewer if complete fixation occurred first. The standard error of these differences replicated in time measures the significance of departures from normality. The mean differences between expected and observed selection differentials for populations under the nine different gene models, are given in Tables 19-21, Appendix B. From the 144 populations which were studied, only two were found in which the mean observed selection differential over the entire selection period exceeded the expected value, and neither of these differences was larger

than the standard error. The two differences occurred in relatively small populations (24, 48). The mean selection differential which usually was achieved in all of the other populations consistently was smaller than the expected value, although the difference was statistically significant in only five of the populations. In any given generation, some of the differences were positive and some were negative, but, in general, the expected differential was larger. This was especially true in later generations.

Table 19-21 show that the difference between expected and observed differentials consistently was larger for populations under intense selection than for those of the same size with moderate culling levels. This is best illustrated for populations of 64 individuals, where one may compare three different intensities of selection. The probable reason for the influence of selection is that the phenotypic values are distributed differently from the normal case to an extent which is observable. The observed distribution has fewer extremes than the normal curve, so that heavy culling has less effect than indicated by the expected selection differential.

Under duplicate and complementary gene action, four of the five populations in which the differences between expected and observed selection differentials were statistically significant (Table 20) had two things in common which may

responsible. These populations had no environmental variation and they were associated with gene models in which the maximum fixation states, i.e., the homozygous genotypes of highest merit, were not distinguishable from some other genotypes. It appears that these populations probably had strongly discontinuous phenotypic distributions in the later generations of selection for three reasons. First, only 40 loci were simulated, and a high proportion of these were fixed after several generations of selection. Second, there was no environmental variation to smooth the discontinuity resulting from discrete segregations of a few genes. Third, selection was unable to finish the job of fixing the loci which were still segregating because dominance hindered the discrimination between homozygous and heterozygous genotypes. Therefore, the mean difference between expected and observed selection differentials in these populations was weighted heavily by the observations over many later generations in which the phenotypic distribution was distinctly abnormal. The errors associated with these differences were also somewhat larger than in other cases.

Under models such as the additive one, where the maximum genotype is readily discernible, prolongation of fixation does not occur as it does when dominance is involved. Therefore, the mean differences are based on observations

from fewer generations, most of which come from phenotypic distributions which are reasonably normal.

The same phenomenon which was noted for the complementary and duplicate cases, was apparent in populations with complete dominance which had no environmental variation, although the differences were not statistically significant (Table 19).

The four populations for each model, which lacked environmental variation were the ones associated with sizes $N = 24$, $N = 48$ (1/6 selection) and the two cases for $N = 128$. The last three of these cases involved differences between expected and observed selection differentials that were larger than their standard errors under the complete dominance, duplicate factor and complementary factor models. The population of size 24 probably conformed more closely to the normal distribution at the point of truncation because the selection intensity was moderate.

The fifth statistically significant difference occurred in a population under intense selection (Table 20, duplicate model, $N = 64$, 1/8 selection). Thus, the point of truncation was near one extreme of the distribution, where it is most likely to be anormal.

Populations under intense selection without environmental variation in the optimum number model (Table 20) also showed differences between expected and observed selection

differentials which were larger than the standard errors. In this model, too, the homozygous maximums are indistinguishable from a heterozygous genotype, so that anormality of the phenotypic distribution may result when only a few loci are segregating in later generations.

Estimation of Realized Heritability from Selection Experiments

The usual linear prediction equation, $\hat{\Delta\mu} = h^2\Delta P$, may be perceived, from another point of view, as a means of estimating heritability from the results of selection already imposed upon a population. This estimate of heritability commonly is called realized heritability, since it is based on the ratio of the response realized from selection to the selection differential, i.e., $h^2 = \hat{\Delta G}/\Delta P$, where $\hat{\Delta G}$ is the phenotypic estimate of the genetic gain from one generation to the next.

Lush (1949) has summarized some of the classical examples of selection experiments, and pointed out that the only real difference in principle between the selection method of estimating heritability (h^2) and the observation of offspring on mid-parent ($b_{\overline{OP}}$) in a non-experimental population, is that epistatic variance may contribute much to $b_{\overline{OP}}$, but can contribute little to h^2 , except in the first two generations of selection. If the data from the

first two generations are discarded, this method comes nearer than any other to excluding epistatic variance from the estimate of heritability in the narrow sense. In experimental populations, the difference between progress in the first and second generations and progress in later generations may be the best indication of the importance of epistasis. These postulations of Lush are largely supported by the results found by Griffing (1960a).

Unfortunately, precise interpretation of results from selection experiments has been obstructed by lack of experimental designs that permit estimation of genetic response independent of environmental trends, inbreeding effects and maternal effects, and by the difficulties encountered in measuring the selection applied for a given trait when there is natural or deliberate selection for other traits, or when there is selection for family or progeny performance in addition to mass selection. However, the ratio of response to selection differential is of interest whether it provides a valid estimate of heritability or does not. It provides a useful description of the effectiveness of selection, which allows comparison of different experiments to be made even when the intensity of selection is not the same.

Falconer (1960) has pointed out that a good part of the effects of natural selection can be accounted for by

weighting the selection differential, i.e., by weighting the deviations of the parents according to the number of their offspring that are measured.

Selection in two directions eliminates some environmental bias from estimates of response, but the response may not be symmetrical in the two directions. In experiments with selection in only one direction, genetic change may occur between generations independent of selection, and epistasis or environmental correlation of offspring and parent may contribute to the regression of offspring on parent, but not to permanent response from selection. Dickerson (1959) has discussed the use of control populations to reduce variation due to changes in environment.

Another problem in evaluating the response to selection arises from the variability of genetic means. The two major causes of this variation are sampling variation, which depends on the number of individuals measured, and environmental change. In practice, the best measure of the average response per generation is obtained from the slope of a regression line fitted to the generation means, the assumption being made that the true response is constant over the period. The variation between generation means appears as error variation about the regression line, and the usual standard error of the estimate of response is based upon it.

Prout (1962) has derived a mathematical expression of the error variance of the estimate of realized heritability from the response to one generation of selection, which takes into account the number of selected parents, i.e., it contains a component due to genetic drift. He assumed that truncation selection is practiced on a random mating population with equal numbers of male and female parents, and found the variance of the estimate of h^2 which is conditional upon the selection differential being constant. He gave the error variance of \hat{h}^2 as $[(h^2[1-h^2]\sigma_{p0}^2/n) + (\sigma_{p1}^2/N)]/(\Delta P)^2$, where σ_{pi}^2 is the phenotypic variance in the total population in generation i , n is the number of parents, N is the number of offspring and ΔP is the selection differential. In practice, estimates would have to be substituted in the equation. The estimate of the variance of h^2 then is biased, i.e., $E[h^2(1-h^2)] = h^2(1-h^2) - \sigma_{\hat{h}^2}^2$. However, a multiplicative correction factor may be used. It is $(\Delta P/\hat{\sigma}_{p0}^2)^2 n / [(\Delta P/\hat{\sigma}_{p0}^2)^2 n - 1]$. Obviously the correction factor will be important only if the numerator is small.

Prout's formula may be simplified to obtain crude estimates which can be compared to results observed in the populations which were simulated in this study. If one takes the additive case, the average phenotypic variance observed over the entire period of selection was approximately 32, and the average estimate of heritability was about .3.

Then, if one assumes $\sigma_{p0}^2 = \sigma_{p1}^2$ and $(\Delta P)^2 = (z/b)^2 \sigma_{p0}^2$ (or the equivalent of z/b for small populations), Prout's formula becomes $(.21/n + 1/N)/(z/b)^2$, without the correction factor. Progeny populations of 16, 64 and 256 with 1/2, 1/4 and 1/8 selection, respectively, broadly represent the different populations which were simulated. If one refers to these as populations 1, 2 and 3, the respective error variances are approximately .155, .017 and .004. Thus, the standard errors for populations 1, 2 and 3 are about .4, .15 and .06, respectively. These values may be compared with the estimates of the standard error of realized heritability for simulated populations which are given in Tables 22-24. The approximate values calculated from the simplification of Prout's formula correspond rather well to the empirical values for the additive model, and, perhaps, apply even better to the average values over all models.

The empirical standard errors were derived from the difference (D) between the estimate of realized heritability, $\hat{\Delta G}/\Delta P$ and the true value, $\Delta G/\Delta P$, in each of the 30 generations in which a population was selected. In the simulated populations, the selection differentials are measured without error, and environmental trends and maternal effects do not exist. Therefore, the difference, D, measures the sampling variation from generation to

generation, which is based on environmental deviations, weighted by the selection differential which is known exactly in each generation. That is, $D = (\hat{\Delta}G - \Delta G) / \Delta P = [(P_{i+1} - G_{i+1}) - (P_i - G_i)] / \Delta P$, where P_i and G_i are the phenotypic and genotypic means in generation i .

Tables 22-24 show the standard errors of the heritability estimates for all nine gene models with populations ranging in size from 16 to 256. Obviously, the errors are strongly related to population size under any gene model because of the effect of random drift on the change in the mean. They appear to be somewhat smaller for a given population size when the gene model is such that the regression of genotypic mean on generation number is negative, i.e., when random drift and inbreeding depression are more powerful than selection (overdominance, Table 22 and DxD, Table 24).

The size of the errors also is related to selection intensity. This may be observed by studying the values for populations of 32 which have 1/2 and 1/4 selection but equal environmental variation (σ_E^2), and for populations of 64 which have 1/2 and 1/8 selection but equal σ_E^2 . The populations which are selected the most intensely have the largest selection differentials and, thus, have errors of smaller magnitude associated with the estimates of heritability, because the selection differential, ΔP , is in the denominator

of D.

Environmental variation, of course, is related to the magnitude of the errors because larger environmental deviations, $E_i = (P_i - G_i)$, cause more variability in the quantities $(P_{i+1} - G_{i+1})$ and $(P_i - G_i)$. Obviously, when no environmental variation is included there are no errors of estimation of h^2 .

Epistasis, in general, does not appear to influence the magnitude of the errors over a period as long as 30 generations (Table 22 versus Tables 23-24). Lush's (1949) postulation that the difference between progress in the first two generations and progress in later generations might be the best indication of epistasis, appears to be correct in some populations for some types of epistasis. In populations with rather intense selection (1/8, 1/6) and tight linkage (.005, .05) under the duplicate, complementary or AxD gene model, progress in the first two generations usually was somewhat larger than in later generations. With the AxA model, the same was true for populations with tight linkage (.005). However, with the optimum number and DxD gene models, there was little apparent difference between the amount of progress made in the first two generations and the amount made in subsequent generations. This result was expected for the DxD case, but it is not clear why populations with optimum number gene action did not advance in

mean value more rapidly in the early generations, except that the useful variance was nearly all σ_{AA}^2 instead of σ_A^2 at gene frequencies near .5, whereas the additive portion is somewhat larger at other frequencies. Perhaps the fact that the original population of completely heterozygous parents was already at maximum genetic value was partially responsible for this phenomenon.

The number of individuals in a population associated with a given magnitude of error of estimating heritability from the response to selection appears to be considerably smaller than the number required to achieve an estimate of comparable accuracy from the conventional parent-offspring regression analysis. Values derived by Searle (1962), when compared with those in this study, indicate that 3-5 times as many parent-offspring pairs are required to achieve standard errors of similar magnitude, the ratio being somewhat larger for smaller predetermined standard errors.

SUMMARY AND CONCLUSIONS

A major deficiency of quantitative genetics is the void between the mathematician working with simple genetic models and the experimenter working with organisms of extreme genetic complexity. A mathematical description of the complex genetic situations results in equations too cumbersome for solution and simplifying assumptions usually lead to large departures from reality.

This study was undertaken as a feasible approach to increasing understanding of genetic selection. The technique was based on simulation of the processes of genetics through the use of repetitive sequences involving random numbers generated by a high-speed computer. This type of approach, termed the Monte Carlo technique, enabled a study of the joint effects of dominance, epistasis, linkage and environmental variation upon the progress of finite genetic populations under selection.

Unisexual diploid individuals were simulated and the quantitative characteristics were assumed to be expressed in both sexes. Equal numbers of selected parents of each sex were mated at random by sampling with replacement. The populations were assumed to be free from the effects of mutation and natural selection over a period of 30 non-overlapping generations. Forty loci were equally spaced over eight chromosomes, with two alleles per locus and

equal genetic effects for all loci. The recombination frequency was uniform for adjacent loci on the same chromosome and cross-over interference was not simulated. Epistasis was restricted to sequential pair-wise interactions of loci. Nine gene models were simulated. These included additive, complete dominance and overdominance cases, optimum number, duplicate factor and complementary factor classical epistatic models, and three models which involve only one component of epistatic variance at gene frequency of one-half. Parent populations were varied in size from 8 to 32, progeny populations from 16 to 256, degree of truncation selection from $1/2$ to $1/8$, linkage from .005 to free recombination and environmental variation from zero to an amount which was three times the expected genotypic variation in the initial progeny population. The initial parent population was completely heterozygous with random association of coupling and repulsion linkage phases.

The progress of the genotypic mean of the progeny population was recorded for 30 generations or until complete fixation occurred for each of 16 runs, or parameter sets, associated with each of the nine models. The content of each of the parameter sets was derived from the orthogonal arrays of a $1/16$ fractional replication of a 4^4 factorial plan involving population size, selection intensity, environmental variation and linkage as the four factors.

Specific interactions are not estimable from this plan, but it is appropriate for an investigation such as this one, which is primarily of the screening type, designed to pick out the most important factors. However, the conclusions drawn in this discussion must be treated with some caution. They are subject to misinterpretation, especially those concerning the main effects of population size and environmental variation, which are confounded with interactions which may not be negligible.

An analysis of variance was made of the results averaged over each five consecutive generations. The basic parameter sets were repeated using different random starts in the computer. This procedure produced an error term which allowed for more precision in testing differences than could be obtained using the mean square for pooled interactions as error. This procedure also allowed for testing the significance of the pooled interactions. In nearly all cases, this test was statistically significant at the probability level $\alpha = .01$. Multiple range procedures were used to determine which differences between levels of a particular factor were responsible for a statistically significant mean square.

In general, the effects of population size on the mean were small in relation to the force of selection in populations without dominance variation of some type, but the smaller populations showed more inbreeding depression when

complete dominance, overdominance, complementary or dominance-by-dominance gene models were used. Results with the complementary factor and complete dominance models were similar because the genotypic values are similar and the epistatic variance for the complementary situation is negligible for gene frequencies above .7. Indications of the existence of an interaction between selection and environment were observed in small populations with complete dominance, overdominance or complementary gene action. A negative regression of genetic mean on generation number was observed for population of all sizes under mass selection with overdominance and dominance-by-dominance models. In duplicate factor and additive-by-dominance cases, population size had little effect on genetic progress despite the presence of dominance variation, because genetic drift and selection had mutual effects on the mean with both models, and inbreeding "uplift" occurred in the latter model.

The mean frequency of unfixed alleles was not strongly related to population size after several generations of selection because a high proportion of loci were fixed in most cases. Differences in the rate of fixation due to population size were larger when the effects of random drift and selection upon fixation were strongly antithetical (e.g., overdominance and DxD models) than when they were somewhat mutual in effect (e.g., optimum number, duplicate factor and

AxA models).

Robertson's (1961) theory that the inbreeding effect is larger than the amount calculated from population size when both selection intensity and heritability are high received tentative confirmation.

Ample indication was given that the mean time to allelic fixation which is implied by the panmictic index, a function of population size, overestimates the actual time. The values derived by Knox (1962) probably are more nearly correct.

In populations under the dominance-by-dominance model, which never contain additive or additive-by-additive variance at any gene frequencies, the approach to fixation was almost entirely due to random drift. The average proportion of total fixation which occurred after ten or more generations of selection corresponded almost exactly, regardless of the amount of linkage involved, to values derived from tables of mean absorption time given by Knox (1962). However, after only five generations of selection, the amount of fixation in populations with tight linkage was only 65 percent of that observed in populations with more recombination.

Current formulae for relating the change in gene frequency to change in the mean of non-random mating populations under selection generally overestimated the change in

small populations because terms for the effects of dominance or epistasis are not independent of drastic changes which sometimes occur in finite populations.

Comparison of the observed results from simulated populations involving dual epistasis to results predicted from the equation developed by Griffing (1960a) for arbitrary dominance, epistasis and linkage showed that the predicted contribution to change in the mean attributed to additive-by-additive variance, generally was far too large over several generations of selection. Random genetic drift and selection appeared to have considerable influence in changing the genetic parameters quickly, so that prediction over a period of several generations was quite erratic. However, the magnitude of the discrepancies noted between predicted means and those observed in Monte Carlo results probably is larger than it would be in an applied breeding situation because of restrictions in the mechanics of simulation.

In general, selection was rather effective in advancing the genetic mean of small populations under all models of gene action in which the genotype of highest merit is homozygous. However, selection was ineffective, or at least weaker than random drift, in small populations under mass selection for a character which involves only a heterozygous genotype as optimum. In these cases (overdominance,

DxD) the regression of genetic merit on generation number was negative for populations associated with all parameter sets. These results are consistent with the general theory that progress from selection is proportional to the amount of additive variance in the population.

Major differences in genetic means were produced by different intensities of selection under models which involved a single peak of genetic merit or one maximum fixation state (additive, complete dominance, overdominance, complementary factors). The presence of large amounts of dominance variance intensified differences due to selection.

Minor differences were observed over a few generations when two homozygous peaks of genetic merit existed (optimum number, AxA), but no important differences in means resulted from different levels of selection when multiple goals existed (Duplicate factors, AxD, DxD).

Every statistically significant comparison of a pair of levels of selection involved the inferior performance of populations from which one-half were selected as parents compared to the performance of populations selected more intensely.

Under the additive-by-additive (AxA) conditional epistatic model, genetic progress was even more rapid than that achieved in the simple additive situation. This

phenomenon was explained by the fact that the AxA model contains two peaks of genetic merit instead of one, and by the fact that the AxA model has only AxA gene action or AxA variance at gene frequency of one-half, but has considerable amounts of additive variance at other gene frequencies. Populations under this model provided some evidence against the hypotheses of Dempster (1955) and Robertson (1960) that the ultimate limit for $1/2$ selection should be higher than the limit for more intense selection. However, the simulated populations were limited to 40 loci. This limitation probably contributed to large changes in the additive genetic variance as a proportion of total variance, which were not allowed for in the theory of limits.

In most cases, selection intensity appeared to have little relation to the mean frequency of unfixed genes after several generations of selection because of fixation at many loci.

In the complete dominance cases, the mean gene frequencies of small populations were remarkably close to the theoretical values after several generations of selection, in view of the fact that random drift was involved. The theoretical values were derived by translating the intensity of selection into the approximate coefficient of selection against the unfavorable allele and calculating the time required for selection to increase the gene frequency by a

given amount.

Mass selection for an overdominant character resulted in less fixation than in corresponding populations under selection for homozygous maximums. Selection for genotypes which are not necessarily heterozygous, but which have intermediate numbers of one type of allele (e.g., A and B) at each pair of interacting loci (optimum number model), is only a minor contributor to fixation in relation to random drift.

Under the duplicate factor model, no fixation of inferior epistatic combinations occurred in any population at any level of selection.

The amount of simulated environmental variation was limited to a function of the genotypic variance associated with only 40 segregating loci. This caused the magnitude of the average gene effects, relative to other sources of variation, to be somewhat larger in simulated populations than it probably is for most actual metric characters. Consequently, the dynamics of changing parameters also differed from real populations.

As one would expect, differences in the amounts of simulated environmental variation between populations generally were important in affecting the relative progress of the mean only when the mean was changing rapidly because of selection. This occurred in most populations under the

additive, complete dominance and complementary factor models over relatively long periods of selection, in populations under the overdominance model after many generations of selection and in other populations over relatively short periods of time. In the overdominance case, populations without environmental variation were able to achieve an equilibrium in 25-30 generations, whereas the others could not.

For most epistatic models, it appeared that large amounts of environmental variance were detrimental to the progress of the mean for short periods of time after segregation of the original heterozygous genotypes, but that the damage to the potential progress of the population over a long period of selection was of little consequence. However, under the optimum number model, populations with environmental variation continued to advance in merit after 20 generations of selection, whereas those without it tended to reach plateaus of genetic merit. These plateaus were permanent, because of fixation, when selection intensity was high, but a population with $1/2$ selection and tight linkage (.005 recombination) developed a slightly fluctuating plateau even though 20-50 percent of the original genotypic variance was available. The plateau was postulated to be caused by the combined effects of epistasis, tight linkage and random drift opposing weak selection.

In most cases, the level of environmental variation had a negligible effect on the differences in the mean proportions of loci fixed and in the mean gene frequency of unfixed loci.

With complete dominance, the results from simulated populations showed that inbreeding, not linkage disequilibrium, was responsible for most of the error in estimating the total genotypic variance by means of the standard formulas based on gene frequencies from conceptually random mating populations in linkage equilibrium. However, in the overdominance situation, the ratio of estimated to observed genotypic variation was much smaller for populations with tight linkage than for those with free recombination, indicating that linkage disequilibrium due to selection was the source of considerable bias.

The effect of different levels of linkage on the differences observed among means of simulated populations was statistically significant in only one case. That occurred during the first five generations under the dominance-by-dominance model (DxD). However, in both the overdominance and DxD situations, it appeared that tight linkage (.005 recombination) was somewhat more effective than levels which involved more recombination, in hindering depression of the mean after a long period of selection. This delaying action probably was caused by an increase of linkage disequilibrium

due to selection for a genetic intermediate.

In populations with optimum number or complementary factor gene action, tight linkage appeared to slow genetic progress for a few generations, but the effect on potential progress of the population mean appeared to be important only for the complementary factor situation. The relative merit of the AAbb and aaBB fixation states is a basic difference between the two models.

The mean proportion of loci fixed and the mean gene frequency of unfixed loci essentially were unaffected by different levels of linkage except in the first few generations in a few populations.

Departures from normality in the distributions of the phenotypes simulated in this study were expressed in terms of the mean difference between achieved and predicted selection differentials over 30 generations. Although the observed selection differentials exceeded the expected ones occasionally in a given generation, the mean difference over 30 generations was positive in only two small populations out of the 144 which were studied, and these differences were smaller than their standard errors. The mean selection differentials observed in all other populations were smaller than the expected values, although the difference was statistically significant in only five of the populations. The differences were larger for populations

under intense selection, probably because the distribution of phenotypes had fewer extremes than the normal case, so that heavy culling had less effect than expected. Nearly all of the large differences between achieved and predicted selection differentials were found in populations which had no environmental variance and were associated with gene models involving dominance to a high degree (complete dominance, duplicate factors, complementary factors). These two common factors, plus the fixation of a high proportion of loci, probably resulted in strongly discontinuous phenotypic distributions after many generations of selection. In populations with less dominance, complete fixation was accomplished more readily so that the mean selection differential was based on fewer generations in which an abnormality of the phenotypic distribution existed.

Empirical standard errors were derived from the difference between the estimate of realized heritability and the "true" value, dependent on a particular set of genetic segregations, from the response to selection in each of 30 consecutive generations. The differences measured the sampling variation from generation to generation which were based on environmental deviations, weighted by the selection differential, which was variable but known in each generation. These differences were free from environmental trends, maternal effects and errors of measurement of the

selection differentials. However, the differences change over time because the value of heritability is changing, too. Therefore, the derived errors are only an average result. Despite this fact, the standard errors of the estimates of realized heritability closely resembled those derived from approximations of the formula given by Prout (1962) for random mating finite populations with equal numbers of male and female parents and constant selection differentials.

The empirical standard errors were strongly related to population size under all nine gene models, although the errors for a given population size were smaller in populations which had negative regressions of genotypic mean on generation number (e.g., overdominance and DxD models).

Populations of a given size which were selected more intensely had larger selection differentials and smaller errors associated with the estimates of realized heritability.

Epistasis, in general, did not appear to influence the magnitude of the errors over a period of selection as long as 30 generations. Populations with rather intense selection ($1/8$, $1/6$) and tight linkage (.005, .05) under the duplicate, complementary, AxA and AxD gene models produced results which support Lush's (1949) postulation that the difference between response in the first two generations and

response in later generations might be the best indication of epistasis. Results from other populations under these models and from all other epistatic models did not support the hypothesis.

The number of individuals in a population necessary for a given magnitude of error of estimating heritability from the response to selection appears to be considerably smaller than the number required to achieve an estimate of comparable accuracy from the conventional parent-offspring regression analysis.

The most general conclusion which can be drawn from the results of this study is that they fit the existing theory rather well in all but a few minor cases. However, some clarification of the nature of existing problems may have been achieved.

Some suggestions, based upon limitations of the present study, as to the methodology and choice of parameters which may be useful in designing future investigations are as follows:

- (1) More loci should be simulated, if this can be done economically.
- (2) Populations of more extreme size should be tried. For example, full-sibbing results should be useful, and populations as large as 100 might eliminate most of the effect of random drift,

which was still somewhat evident in populations of size 32 in this study.

- (3) Selection intensities which are unequal in the sexes might be informative.
- (4) The amount of environmental variation which is simulated should be enlarged, especially if no more than 40 loci are used. This is necessary in order to slow fixation and to achieve average gene effects which are small relative to other sources of variation. The populations with no environmental variation produced peculiar results, chiefly in connection with the discontinuity of the phenotypic distribution.
- (5) Linkage which allows as much as .005 recombination appeared to have little effect on the progress of the mean under long-time selection. Perhaps smaller recombination values should be studied. If linkage relations between loci other than adjacent loci could be included, and non-uniform recombination values were studied, the results might be more informative in relation to the epistasis involved and to actual biological populations.
- (6) Interactions among the factors of population size, selection intensity, environmental variation and

linkage must be studied before definite conclusions can be drawn about their main effects.

- (7) More two-locus interactions, and, possibly, more complex epistatic situations, should be simulated.
- (8) The conditional epistatic models which were studied each have peculiarities which are interesting, but possibly unreal. None of them corresponds to the individual component of variance it represents except at gene frequency of one-half. However, it seems possible that some of the changes in the partition of variance which occur with changing gene frequencies in these models, may resemble the changes which occur when such variance is present in actual populations under selection, but this has not been clearly demonstrated.
- (9) Populations under the complementary factor model produced results which differed little from those of populations under complete dominance. However, the similarities are not necessarily due to common causes.
- (10) The optimum number and duplicate factor models appear to be the most promising for studying epistasis, the former because of its flexibility in more complex epistatic situations, the latter

because epistatic variation increases as a proportion of the total genotypic variance as gene frequency increases.

In conclusion, it seems that the major usefulness of this preliminary study will not be in the direct application of the inferences which have been made, but in clarifying thoughts and definitions of problems which should be investigated.

LITERATURE CITED

- Bailey, N. T. J. 1961. Introduction to the mathematical theory of genetic linkage. London, The Clarendon Press.
- Baker, L. H. and R. E. Comstock. 1961. Linkage and heterozygosity in finite populations. Unpublished paper presented at a Symposium on Statistical Genetics and Plant Breeding, Raleigh, North Carolina, Apr. 1961. Mimeo. Johnston, Iowa, Hy-Line Poultry Farms. 1961.
- Barker, J. S. F. 1958a. Simulation of genetic systems by automatic digital computers. III. Selection between alleles at an autosomal locus. Australian Journal of Biological Sciences 11: 603-612.
- Barker, J. S. F. 1958b. Simulation of genetic systems by automatic digital computers. IV. Selection between alleles at a sex-linked locus. Australian Journal of Biological Sciences 11: 613-625.
- Bartlett, M. S. and J. B. S. Haldane. 1934. The theory of inbreeding in autotetraploids. Journal of Genetics 29: 175-180.
- Bernstein, F. 1930. Fortgesetzte Untersuchungen aus der Theorie der Blutgruppen. Zeitschrift für induktive Abstammungs- und Vererbungslehre 56: 233-273.
- Bohidar, N. R. 1960. The role of sex-linked genes in quantitative inheritance. Unpublished Ph. D. thesis. Ames, Iowa, Library, Iowa State University of Science and Technology.
- Cochran, W. G. 1951. Improvement by means of selection. Berkeley Symposium on Mathematical Statistics and Probability Proc. 2: 449-470.
- Cockerham, C. C. 1954. An extension of the concept of partitioning hereditary variance for analysis of covariances among relatives when epistasis is present. Genetics 39: 859-882.
- Cockerham, C. C. 1956. Effects of linkage on the covariance between relatives. Genetics 41: 138-141.
- Cockerham, C. C. 1959. Partitions of hereditary variance for various genetic models. Genetics 44: 1141-1148.

- Comstock, R. E. 1960. Dominance, genotype-environment interaction and homeostasis. In Kempthorne, O., ed. *Biometrical genetics*. pp. 3-9. New York, N. Y., Pergamon Press.
- Crow, J. F. 1954. Breeding structure of populations. 2. Effective population number. In Kempthorne, O., T. A. Bancroft, J. W. Gowen and J. L. Lush, eds. *Statistics and mathematics in biology*. pp. 543-556. Ames, Iowa, Iowa State College Press.
- Crow, J. F. and M. Kimura. 1955. Some genetic problems in natural populations. *Berkeley Symposium on Mathematical Statistics and Probability Proc.* 3: 1-22.
- Dempster, E. R. 1955. Genetic models in relation to animal breeding problems. *Biometrics* 11: 535-536.
- Dempster, E. R. and I. M. Lerner. 1950. Heritability of threshold characters. *Genetics* 35: 212-236.
- Dickerson, G. E. 1959. Techniques for research in quantitative animal genetics. In Terrill, C. E., ed. *Techniques and procedures in animal production research*. pp. 56-105. Beltsville, Maryland, American Society of Animal Production.
- Duncan, D. B. 1955. Multiple range and multiple F-tests. *Biometrics* 11: 1-42.
- Eisenhart, C. 1939. The interpretation of certain regression methods and their uses in biological and industrial research. *Annals of Mathematical Statistics* 10: 162-186.
- Falconer, D. W. 1960. *Introduction to quantitative genetics*. New York, N. Y., The Ronald Press Company.
- Fisher, R. A. 1918. The correlation between relatives on the supposition of Mendelian inheritance. *Transactions of the Royal Society, Edinburgh* 52, Part 2: 399-433.
- Fisher, R. A. 1922. On the dominance ratio. *Proceedings of the Royal Society, Edinburgh* 42: 321-341.
- Fisher, R. A. 1930. *The genetical theory of natural selection*. London, The Clarendon Press.
- Fisher, R. A. 1941. Average excess and average effect of a gene substitution. *Annals of Eugenics (London)* 11: 53-63.

- Fisher, R. A. 1949. The theory of inbreeding. London, Oliver and Boyd, Ltd.
- Fisher, R. A. 1960. The design of experiments. 7th ed. New York, Hafner Publishing Company, Inc.
- Fisher, R. A. and F. Yates. 1943. Statistical tables. 2nd ed. Edinburgh, Oliver and Boyd, Ltd.
- Fraser, A. S. 1957a. Simulation of genetic systems by automatic digital computers. I. Introduction. Australian Journal of Biological Sciences 10: 484-491.
- Fraser, A. S. 1957b. Simulation of genetic systems by automatic digital computers. II. Effects of linkage on rates of advance under selection. Australian Journal of Biological Sciences 10: 492-499.
- Fraser, A. S. 1960a. Simulation of genetic systems by automatic digital computers. V. Linkage, dominance and epistasis. In Kempthorne, O., ed. Biometrical genetics. pp. 70-83. New York, N. Y., Pergamon Press.
- Fraser, A. S. 1960b. Simulation of genetic systems by automatic digital computers. VI. Epistasis. Australian Journal of Biological Sciences 13: 150-162.
- Fraser, A. S. 1960c. Simulation of genetic systems by automatic digital computers. VII. Effects of reproductive rate and intensity of selection on genetic structure. Australian Journal of Biological Sciences 13: 344-350.
- Greenberger, M. 1961. Notes on a new pseudo-random number generator. Journal of the Association of Computing Machinery 8: 163-167.
- Griffing, B. 1960a. Theoretical consequences of truncation selection based on the individual phenotype. Australian Journal of Biological Sciences 13: 307-343.
- Griffing, B. 1960b. Accommodation of linkage in mass selection theory. Australian Journal of Biological Sciences 13: 501-526.
- Haldane, J. B. S. 1923. A mathematical theory of natural and artificial selection. I. Transactions of the Cambridge Philosophical Society 23: 19-42.

- Haldane, J. B. S. 1926. A mathematical theory of natural and artificial selection. IV. Proceedings of the Cambridge Philosophical Society 23: 607-615.
- Horner, T. W. 1956. The components of variance in symmetrical random mating populations with the frequency of the more favorable allele the same at all loci. Iowa State College Journal of Science 31: 67-77.
- Horner, T. W., R. E. Comstock and H. F. Robinson. 1955. Non-allelic gene interactions and the interpretation of quantitative genetic data. North Carolina Agricultural Experiment Station Technical Bulletin 118.
- Harter, H. L. 1960. Critical values for Duncan's new multiple range test. Biometrics 16: 671-685.
- Jennings, H. S. 1914. Formulae for the results of inbreeding. Genetics 1: 53-89.
- Kempthorne, O. 1952. Design and analysis of experiments. New York, N. Y., John Wiley and Sons, Inc.
- Kempthorne, O. 1954. The correlation between relatives in a random mating population. Proceedings of the Royal Society, London, B, 143: 103-113.
- Kempthorne, O. 1957. An introduction to genetic statistics. New York, N. Y., John Wiley and Sons, Inc.
- Kimura, M. 1958. On the change of population fitness by natural selection. Heredity 12: 145-167.
- King, J. W. B., J. H. Watson and G. B. Young. 1959. Genotype-environment interactions in the wintering of lambs. Journal of Agricultural Science 53: 156-171.
- Knox, S. R. 1962. A study of a random-mating population of fixed size. Unpublished paper presented at the annual meeting of the American Statistical Association, Minneapolis, Minnesota, Sept. 1962. (Dittoed) Jackson, Mississippi, Department of Mathematics, Millsaps College, 1962.
- Kojima, K. 1961. Effects of dominance and size of population on response to mass selection. Genetical Research 2: 177-188.

- Lehmer, D. H. 1951. Mathematical methods in large scale computing units. *Annals of the Computer Laboratory, Harvard University* 26: 141-145.
- Lerner, I. M. 1954. Genetic homeostasis. London, Oliver and Boyd, Ltd.
- Li, C. C. 1955. Population genetics. Chicago, The University of Chicago Press.
- Lush, J. L. 1945. Animal breeding plans. 3rd ed. Ames, Iowa, Iowa State College Press.
- Lush, J. L. 1946. Chance as a cause of changes in gene frequency. *American Naturalist* 80: 318-342.
- Lush, J. L. 1947. Family merit and individual merit as bases for selection. I. *American Naturalist* 81: 241-261.
- Lush, J. L. 1948. The genetics of populations. (Mimeo.) Ames, Iowa, Department of Animal Science, Iowa State University of Science and Technology.
- Lush, J. L. 1949. Heritability of quantitative characters in farm animals. *Proceedings of the International Congress of Genetics* 8: 356-375.
- Lush, J. L. 1954. Breeding structure of populations. 1. General considerations. In Kempthorne, O., T. A. Bancroft, J. W. Gowen, and J. L. Lush, eds. *Statistics and mathematics in biology*. pp. 537-541. Ames, Iowa, Iowa State College Press.
- Malécot, G. 1948. *Les mathématiques de l'hérédité*. Paris, Masson et Cie.
- Martin, F. G., Jr. and C. C. Cockerham. 1960. High speed selection studies. In Kempthorne, O., ed. *Biometrical genetics*. pp. 35-45. New York, N. Y., Pergamon Press.
- Mason, R. W., R. Bogart and H. Krueger. 1958. Heritability estimates as affected by sex, genetic-environmental interaction and type of selection practiced. *Proceedings of the Western Section of the American Society of Animal Production* 9, Part 45: 1-6.

- Mather, K. 1943. Polygenic inheritance and natural selection. *Biological Review* 18: 32-64.
- Meyer, H. A., ed. 1956. *Symposium on Monte Carlo methods*. New York, N. Y., John Wiley and Sons, Inc.
- Nordskog, A. W. and A. J. Wyatt. 1952. Genetic improvement as related to size of breeding operations. *Poultry Science* 31: 1062-1066.
- Pearl, R. 1917. Studies on inbreeding. *American Naturalist* 51: 545-559.
- Perotti, Juliette M. 1943. Mean improvement in a normal variate under direct and indirect selection. Unpublished M.S. thesis. Ames, Iowa, Library, Iowa State University of Science and Technology.
- Prout, T. 1962. The error variance of the heritability estimate. *Biometrics* 18: 404-407.
- Reeve, E. C. R. and F. W. Robertson. 1953. Analysis of environmental variability in quantitative inheritance. *Nature (London)* 171: 874-875.
- Rendel, J. M. and A. Robertson. 1950. Estimation of genetic gain in milk yield by selection in closed herds of dairy cattle. *Journal of Genetics* 50: 1-8.
- Robertson, A. 1952. The effect of inbreeding on the variation due to recessive genes. *Genetics* 37: 189-207.
- Robertson, A. 1955a. Prediction equations in quantitative genetics. *Biometrics* 11: 95-98.
- Robertson, A. 1955b. The explorations of ceilings and plateaus on a theoretical and practical basis. *Poultry Breeder's Roundtable Proc.* 4: 57-77.
- Robertson A. 1960. A theory of limits in artificial selection. *Royal Society, Edinburgh, Proc. B*, 153: 234-249.
- Robertson A. 1961. Inbreeding in artificial selection programmes. *Genetical Research* 2: 189-194.
- Robertson, F. W. and E. C. R. Reeve. 1952a. Heterozygosity, environmental variation and heterosis. *Nature(London)* 170: 286.

- Robertson, F. W. and E. C. R. Reeve. 1952b. Studies on quantitative inheritance. I. The effects of selection of wing and thorax length in Drosophila melanogaster. Journal of Genetics 50: 414-448.
- Robinson, H. F., C. C. Cockerham and R. H. Moll. 1960. Studies on estimation of dominance variance and effects of linkage bias. In Kempthorne, O., ed. Biometrical genetics. pp. 171-177. New York, N. Y., Pergamon Press.
- Robinson, H. F. and R. E. Comstock. 1955. Analysis of the genetic variability in corn with reference to probable effects of selection. Cold Spring Harbor Symposia on Quantitative Biology 20: 127-136.
- Rotenberg, A. 1960. A new pseudo-random number generator. Journal of the Association of Computing Machinery 7: 75-77.
- Searle, S. R. 1962. Number of paired observations needed for estimating repeatability and heritability from regression analysis. Journal of Animal Science 21: 426-427.
- Schnell, F. W. 1961. The covariance between relatives in the presence of linkage. Unpublished paper presented at a Symposium on Statistical Genetics and Plant Breeding, Raleigh, North Carolina, March, 1961. Mimeo. Basse bei Neustadt am Rubenberge, Germany., Zweigstelle Scharnhorst, Max Planck-Institut fur Zuchtungsforshung. 1961.
- Watterson, G. A. 1962. Some theoretical aspects of diffusion theory in population genetics. Annals of Mathematical Statistics 33: 939-957.
- Winsor, C. P. 1946. Which regression? Biometrics 2: 101-109.
- Wright, S. 1918. On the nature of size factors. Genetics 3: 367-374.
- Wright, S. 1920. The relative importance of heredity and environment in determining the piebald pattern of guinea pigs. Proceedings of the National Academy of Sciences 6: 320-332.
- Wright, S. 1921a. Systems of mating. II. The effects of inbreeding on the genetic composition of a population. Genetics 6: 124-143.

- Wright, S. 1921b. Systems of mating. IV. The effects of selection. *Genetics* 6: 162-178.
- Wright, S. 1922. Coefficients of inbreeding and relationship. *American Naturalist* 56: 330-338.
- Wright, S. 1931. Evolution in Mendelian populations. *Genetics* 16: 97-159.
- Wright, S. 1935. The analysis of variance and the correlation between relatives with respect to deviations from an optimum *Journal of Genetics* 30: 243-256.
- Wright, S. 1945. The differential equations of the distribution of gene frequencies. *Proceedings of the National Academy of Sciences* 31: 382-389.
- Wright, S. 1951. The genetic structure of populations. *Annals of Eugenics* 15: 323-354.

ACKNOWLEDGMENTS

The author wishes to express his indebtedness to Drs. L. N. Hazel, Oscar Kempthorne and D. L. Harris for valuable suggestions which they contributed during the preparation of the manuscript. Kenneth Rowe's assistance in programming the computer is also highly appreciated. The work was supported by grant number NSF-G19218 from the National Science Foundation.

APPENDIX A: FIGURES

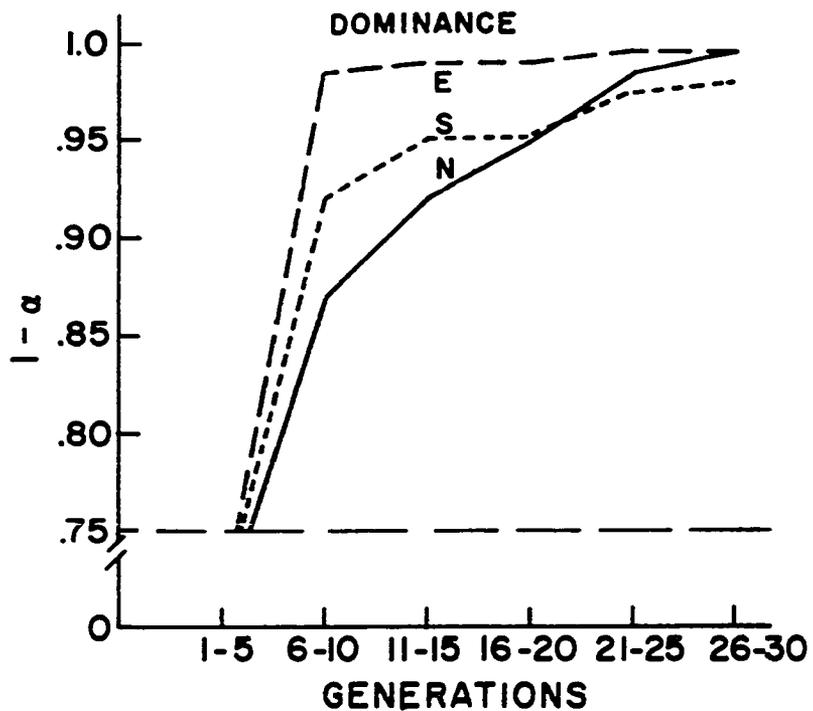
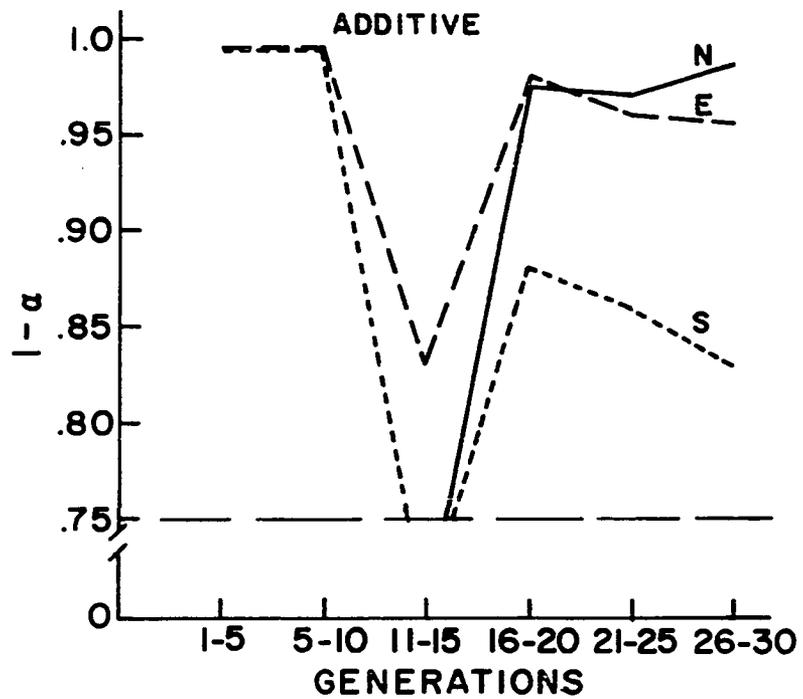


Figure 1. Significance of differences in the mean due to population size (N), selection intensity (S), and environmental variation (E)

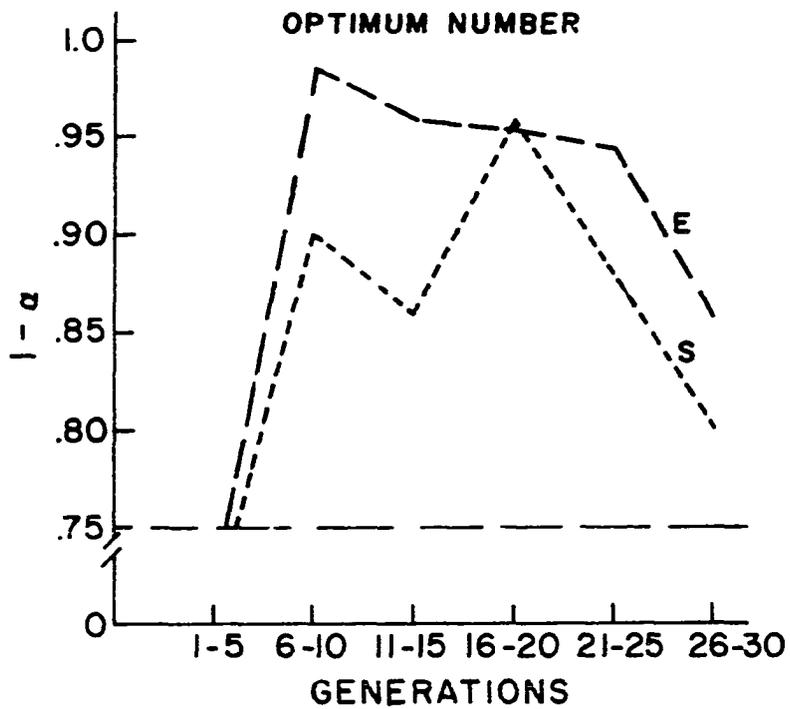
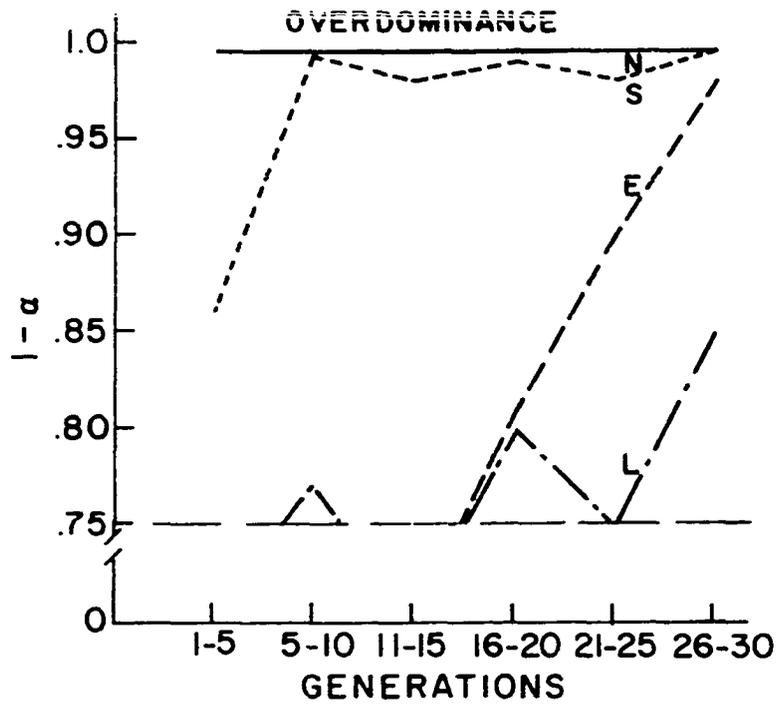


Figure 2. Significance of differences in the mean due to population size (N), selection intensity (S), environmental variation (E) and linkage (L)

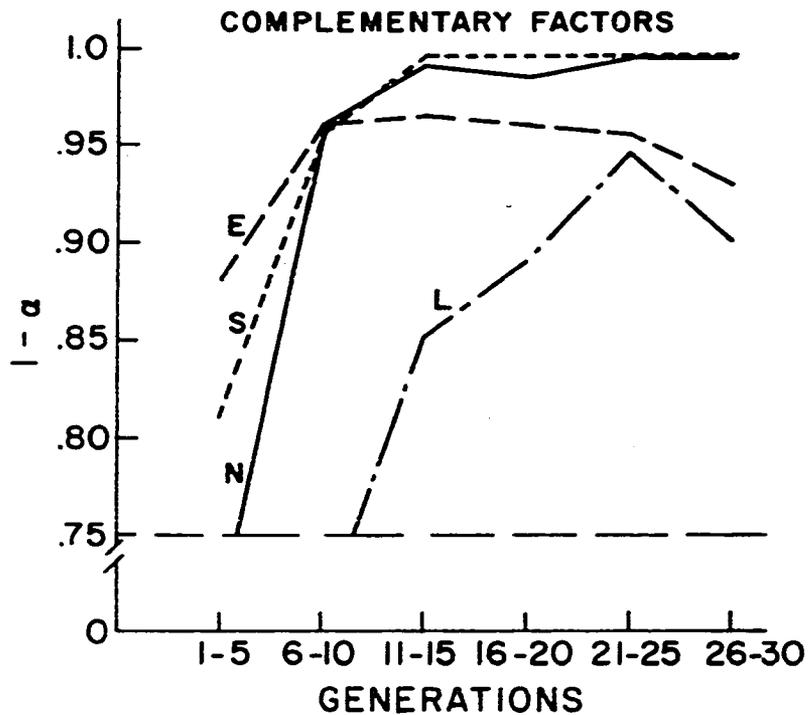
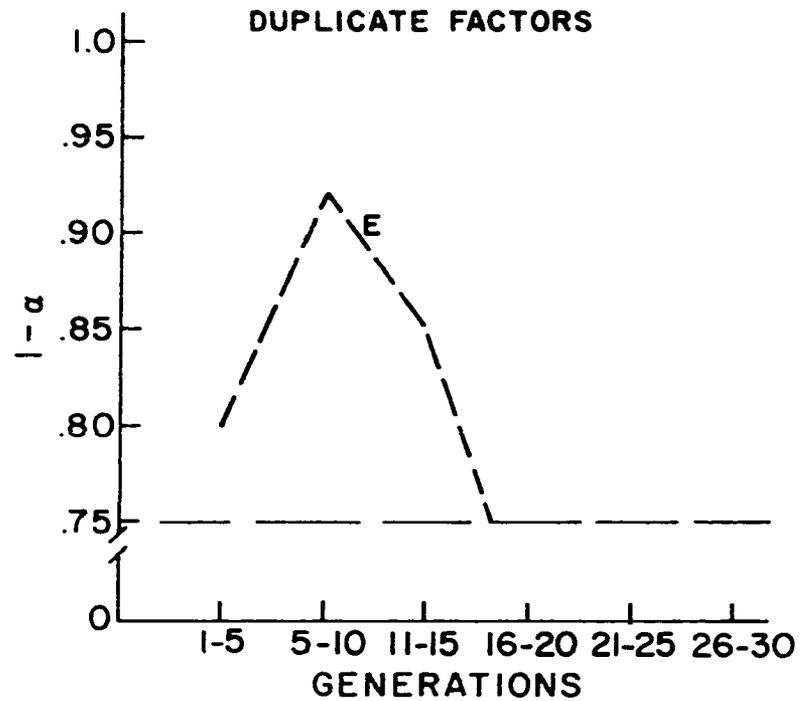


Figure 3. Significance of differences in the mean due to population size (N), selection intensity (S), environmental variation (E) and linkage (L)

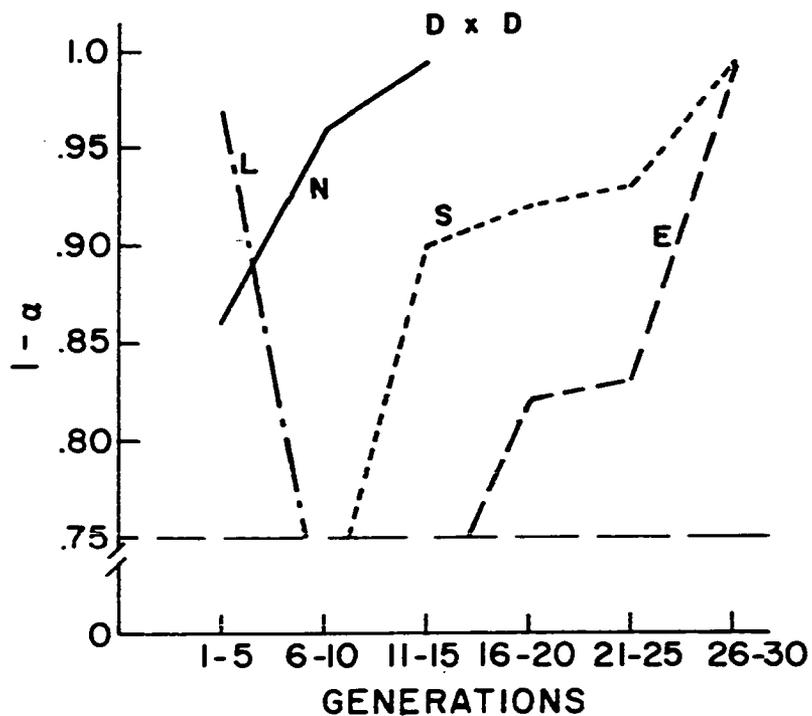
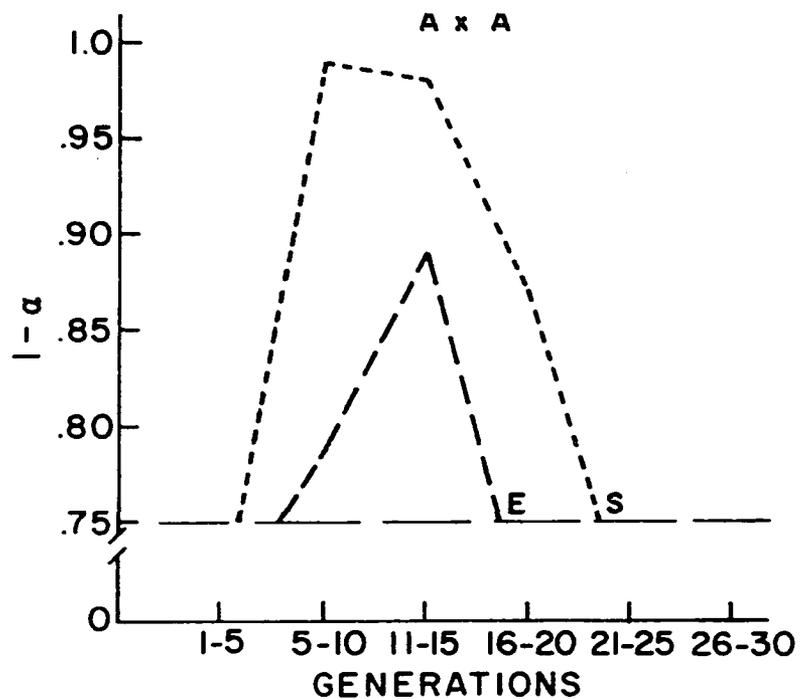


Figure 4. Significance of differences in the mean due to population size (N), selection intensity (S), environmental variation (E) and linkage (L)

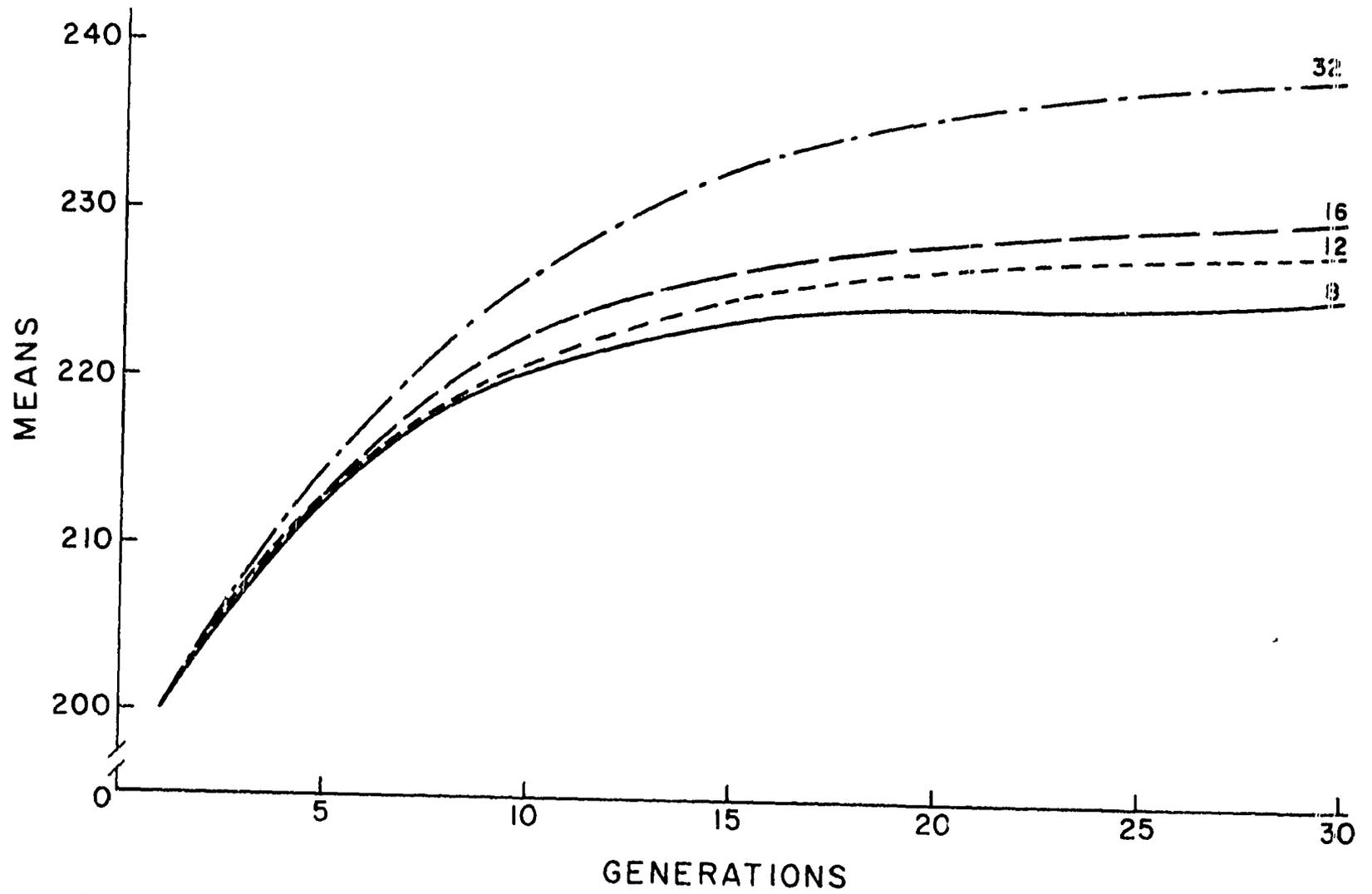


Figure 5. Mean genetic progress by level of population size with additive gene action

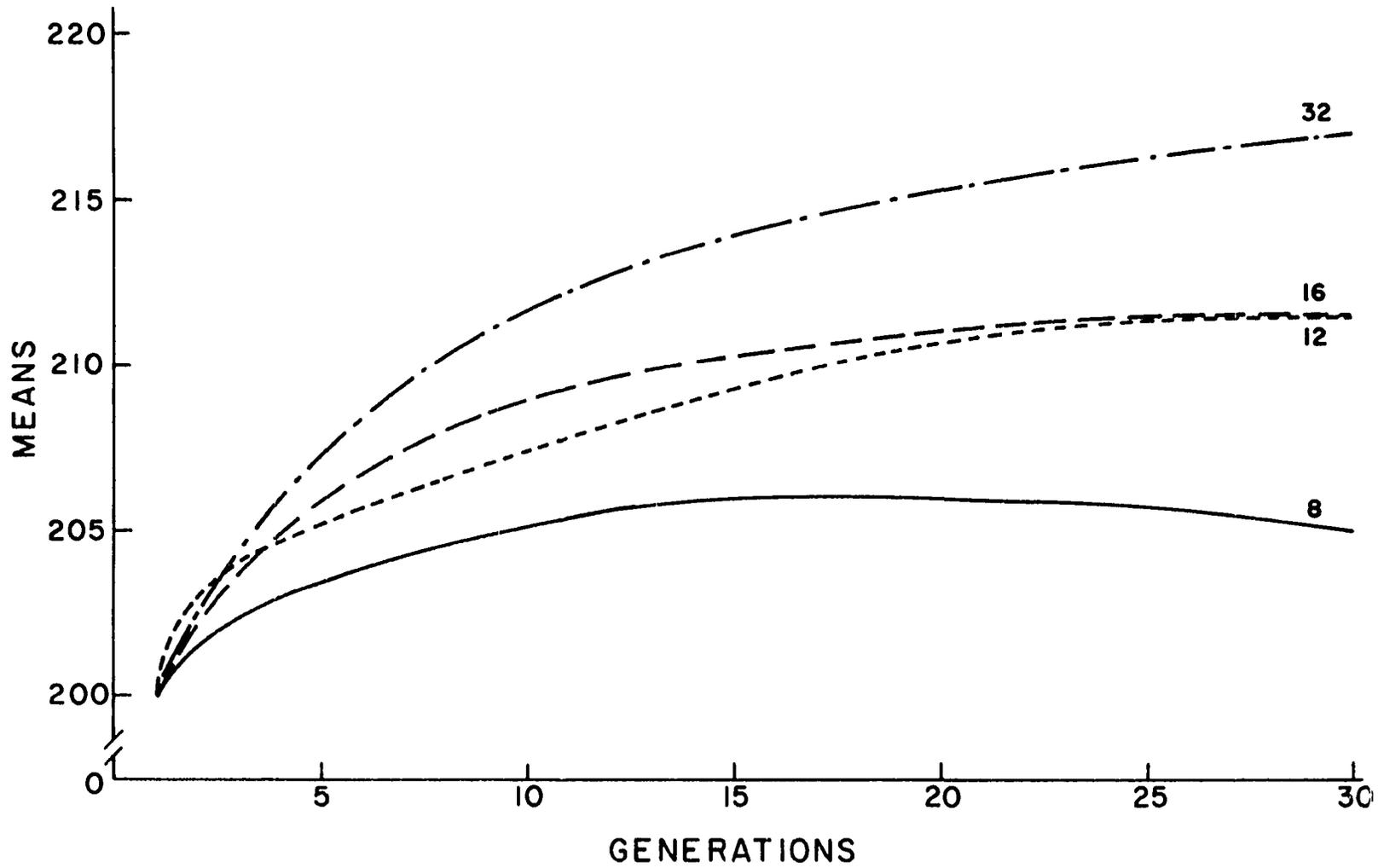


Figure 6. Mean genetic progress by level of population size with complete dominance

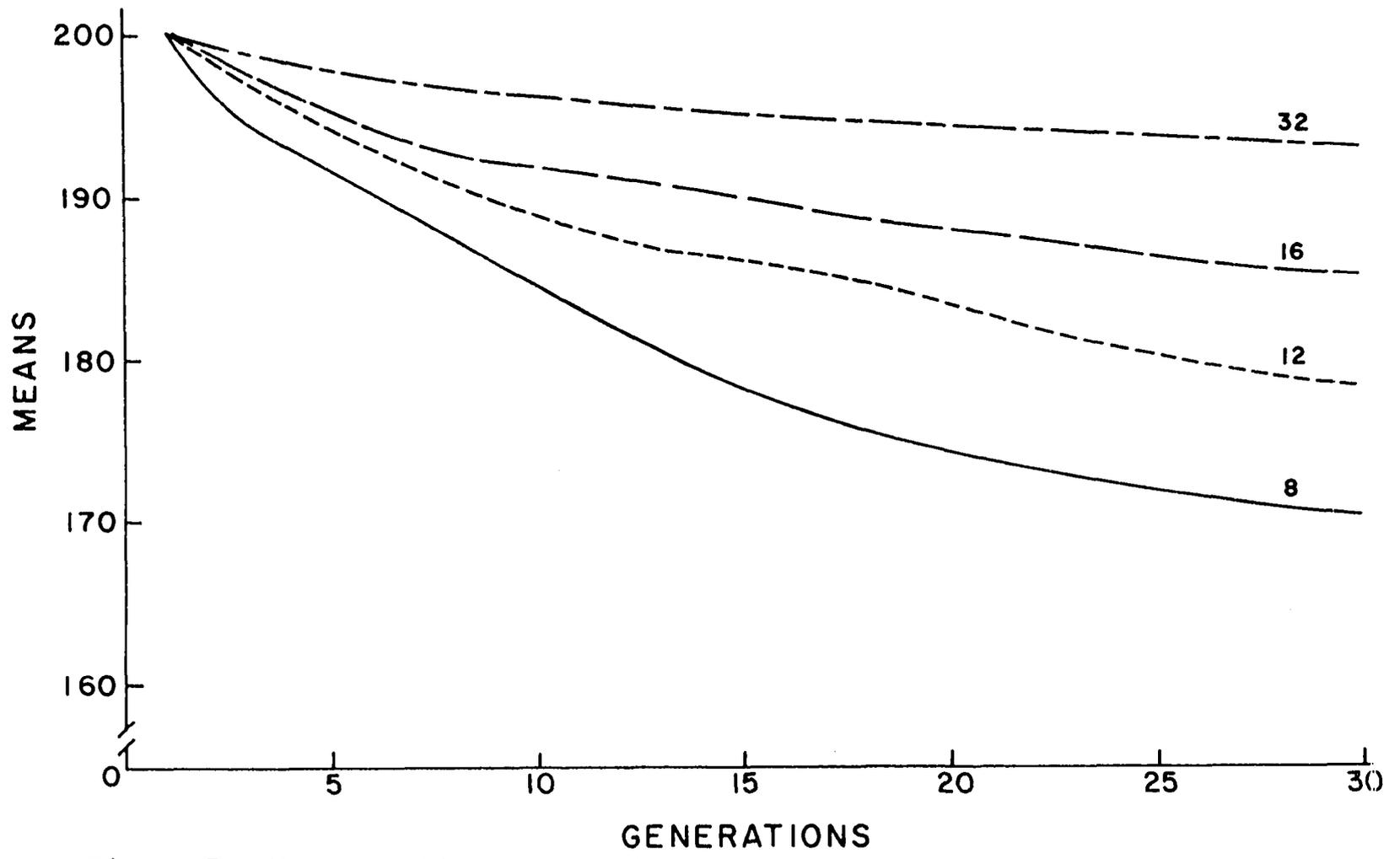


Figure 7. Mean genetic progress by level of population size with over-dominance

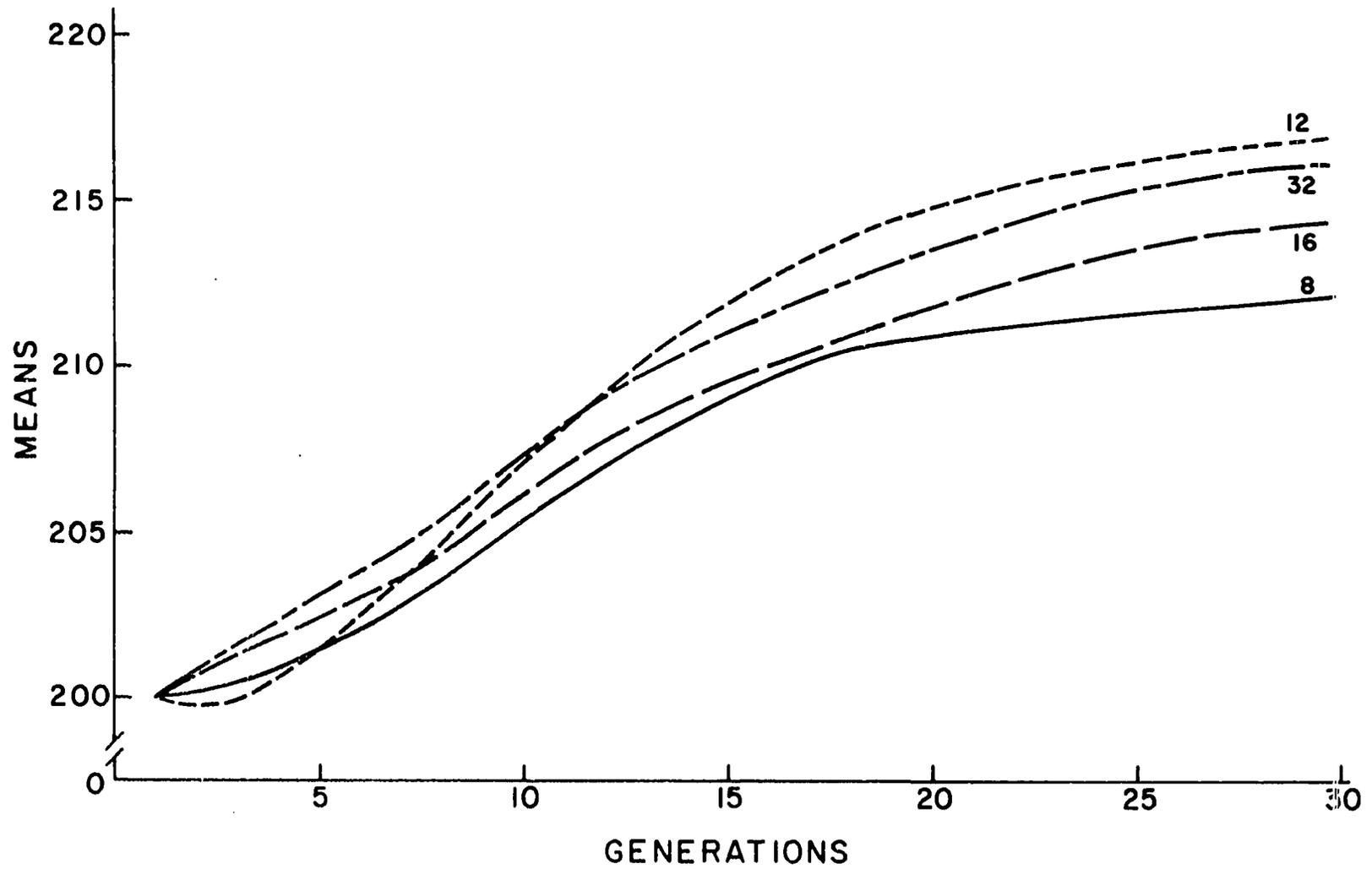


Figure 8. Mean genetic progress by level of population size with optimum number of genes intermediate

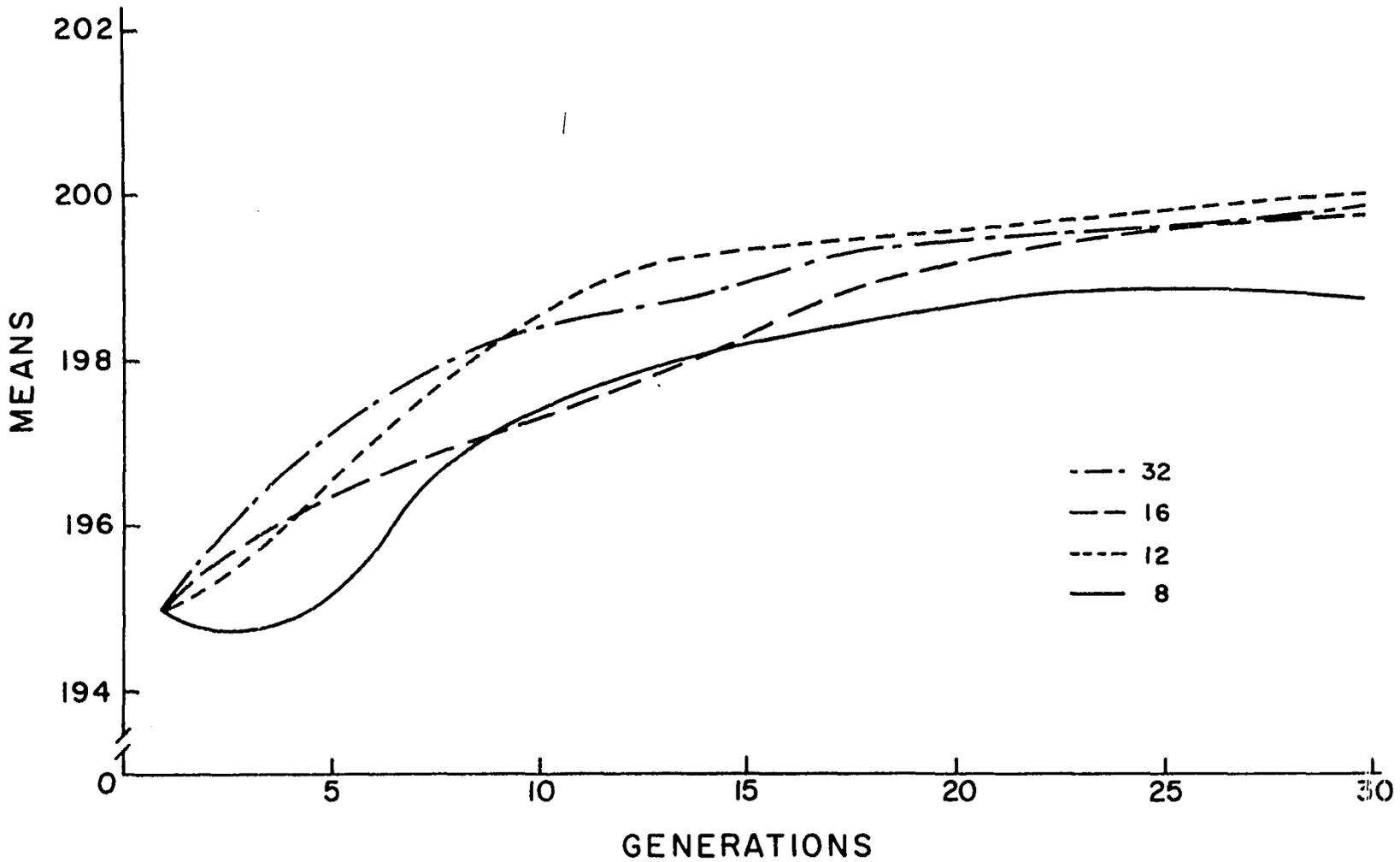


Figure 9. Mean genetic progress by level of population size with duplicate gene action

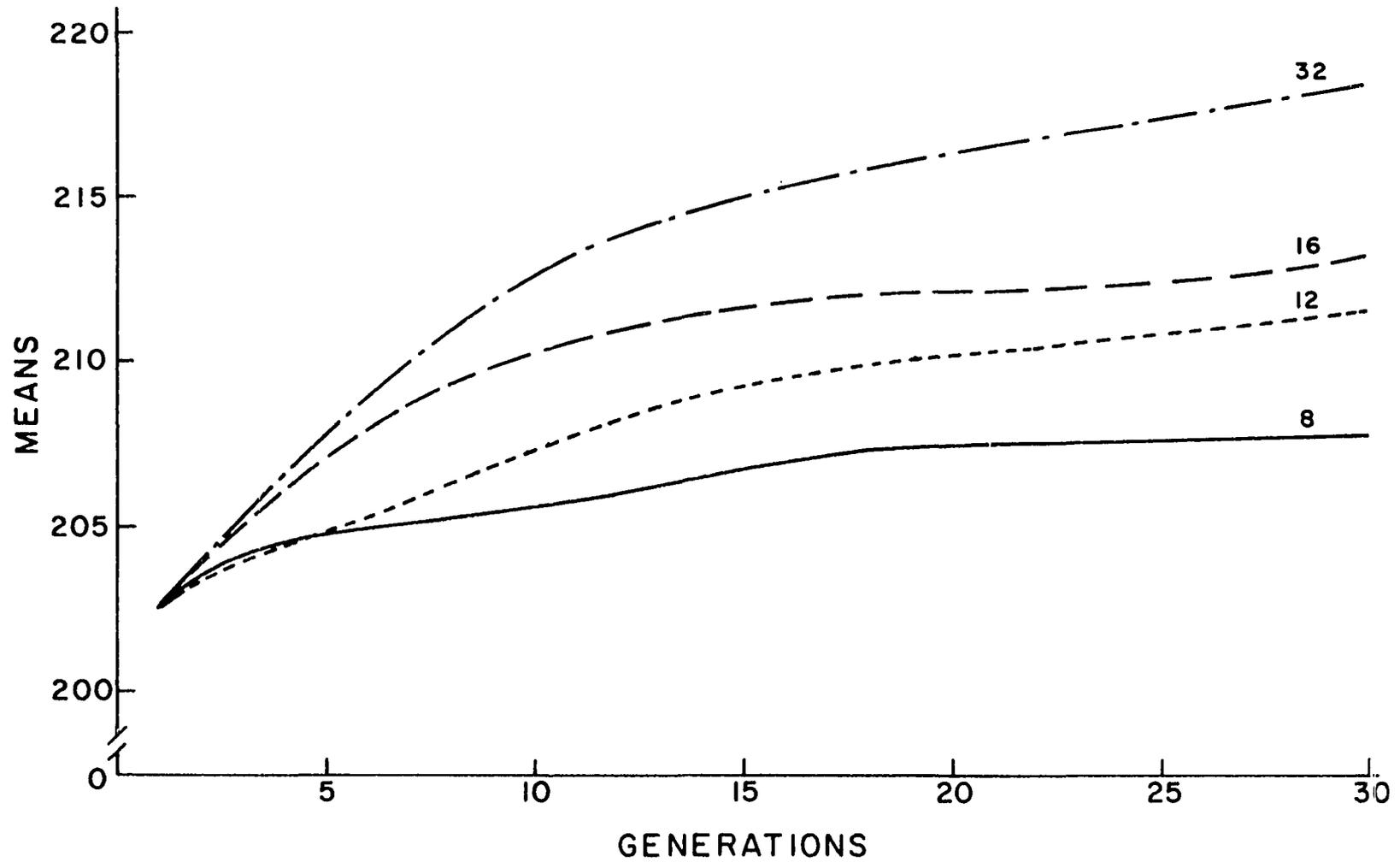


Figure 10. Mean genetic progress by level of population size with complementary gene action

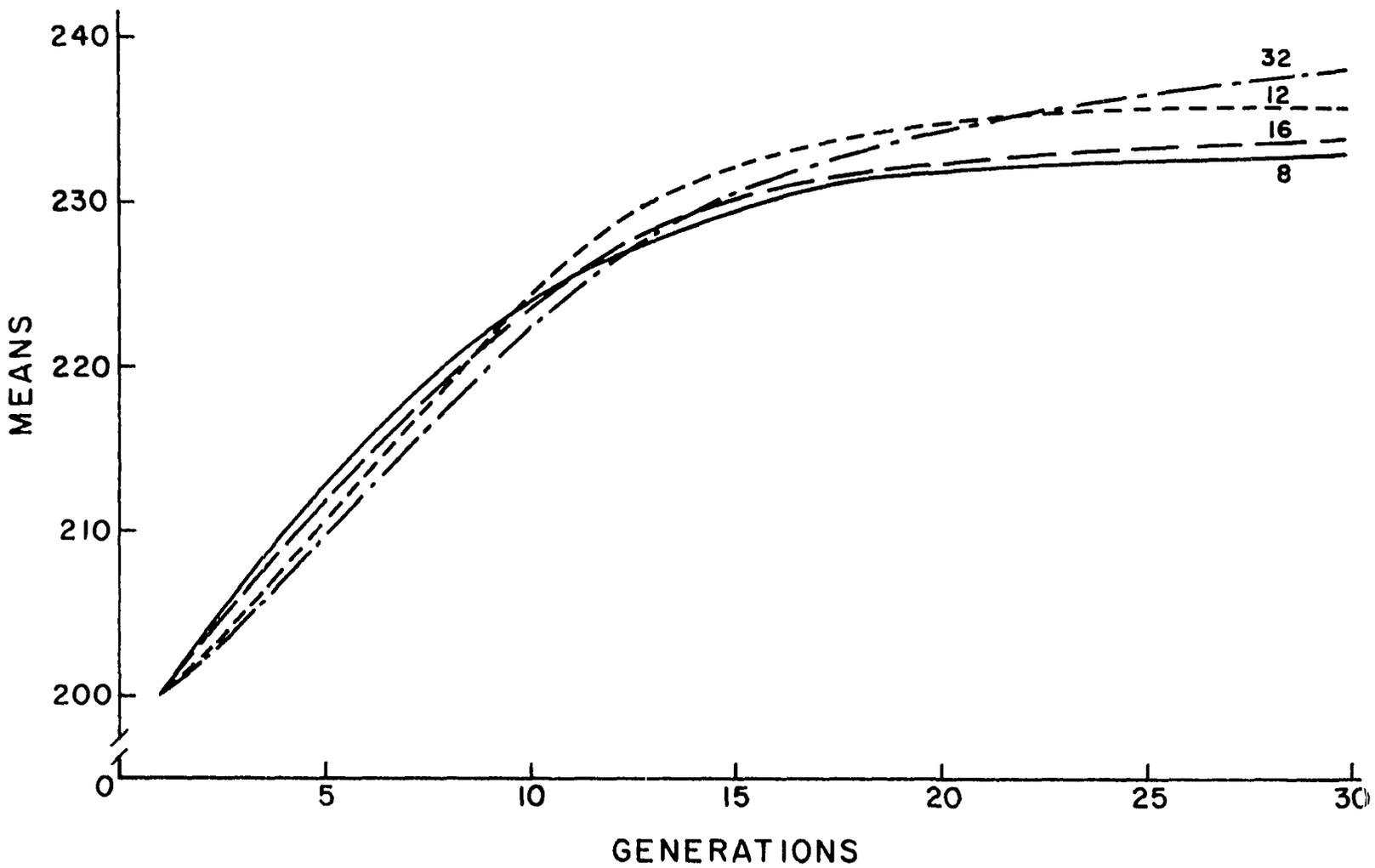


Figure 11. Mean genetic progress by level of population size with additive-by-additive interaction

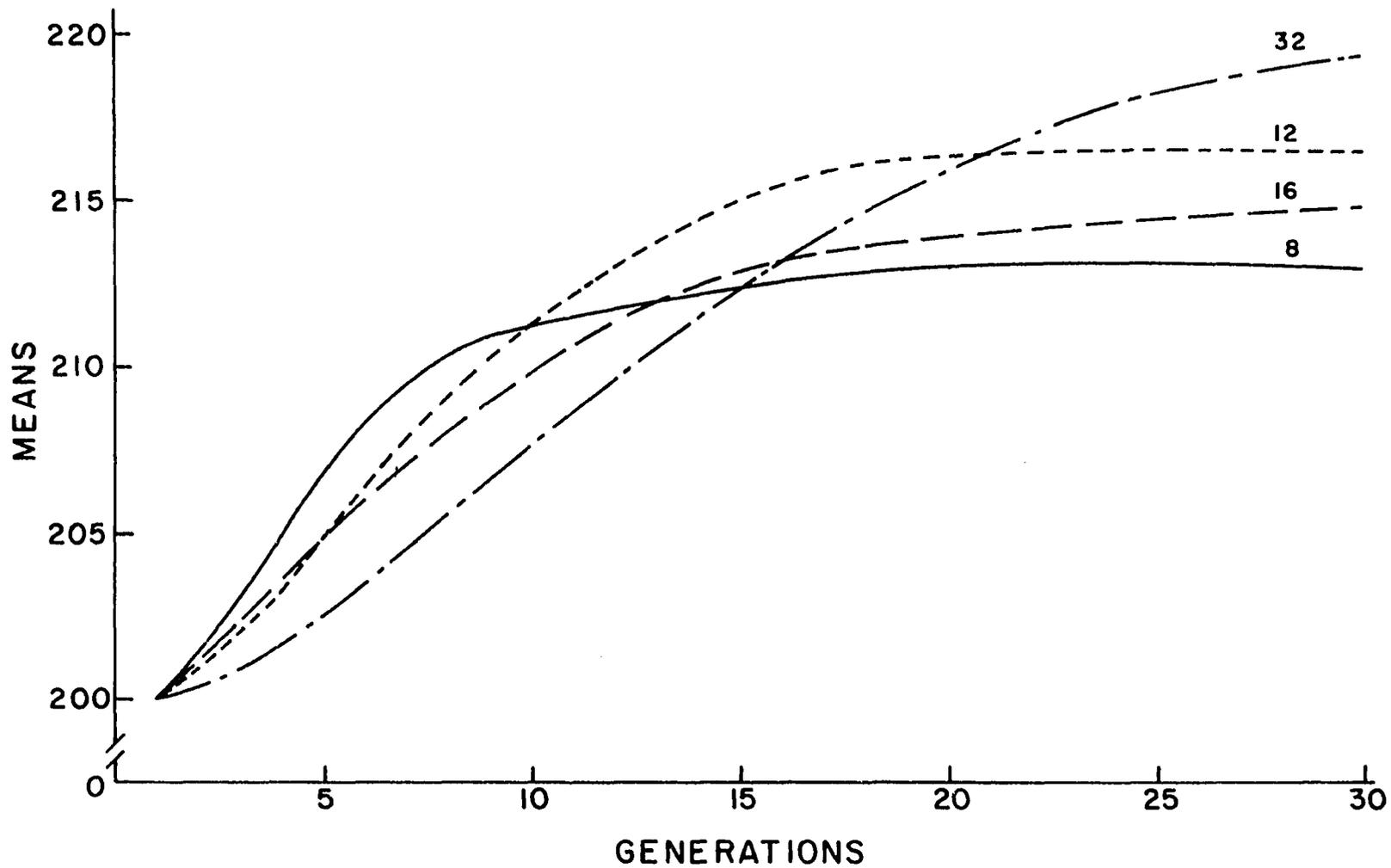


Figure 12. Mean genetic progress by level of population size with additive-by-dominance interaction

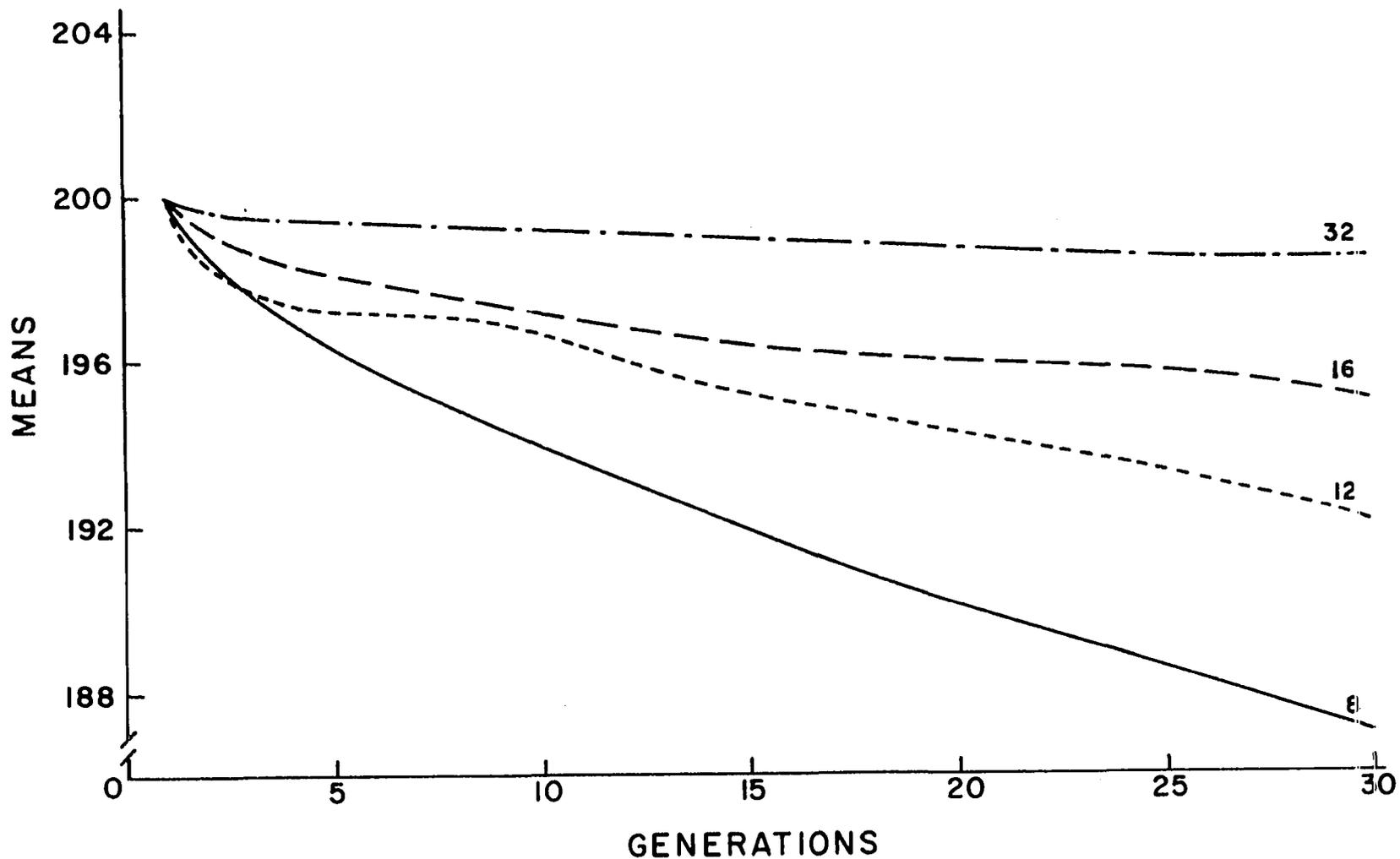


Figure 13. Mean genetic progress by level of population size with dominance-by-dominance interaction

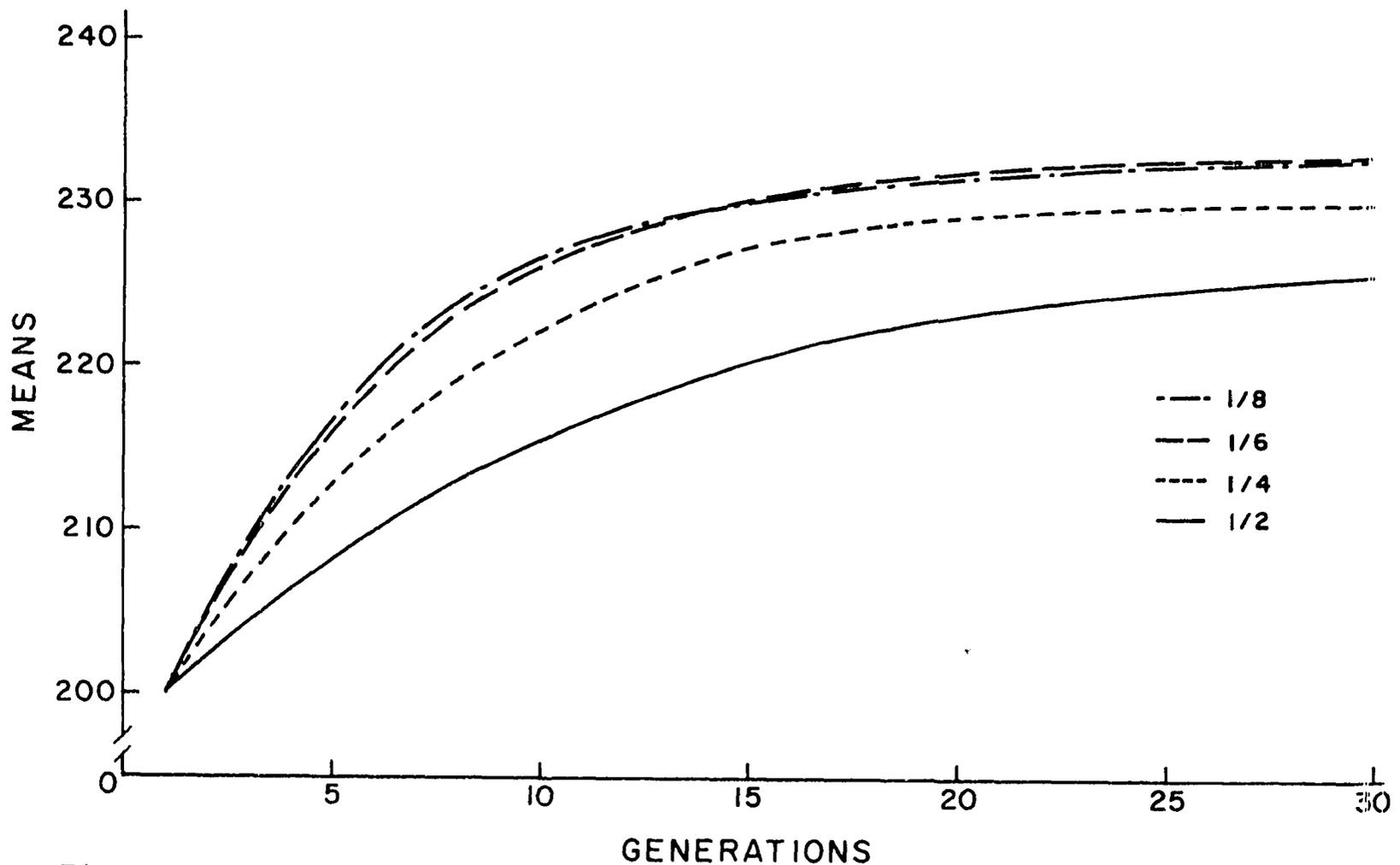


Figure 14. Mean genetic progress by level of selection intensity with additive gene action

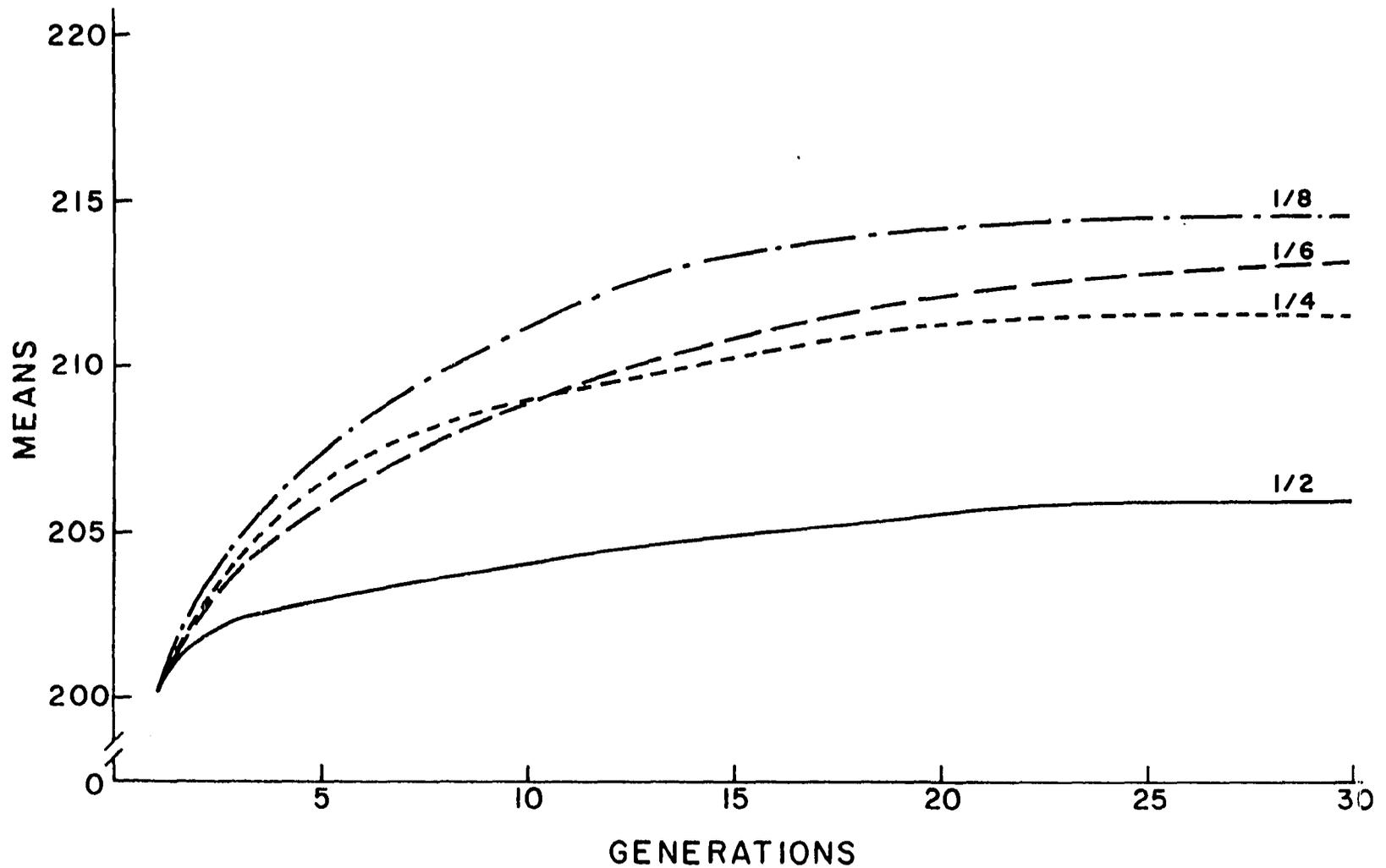


Figure 15. Mean genetic progress by level of selection intensity with complete dominance

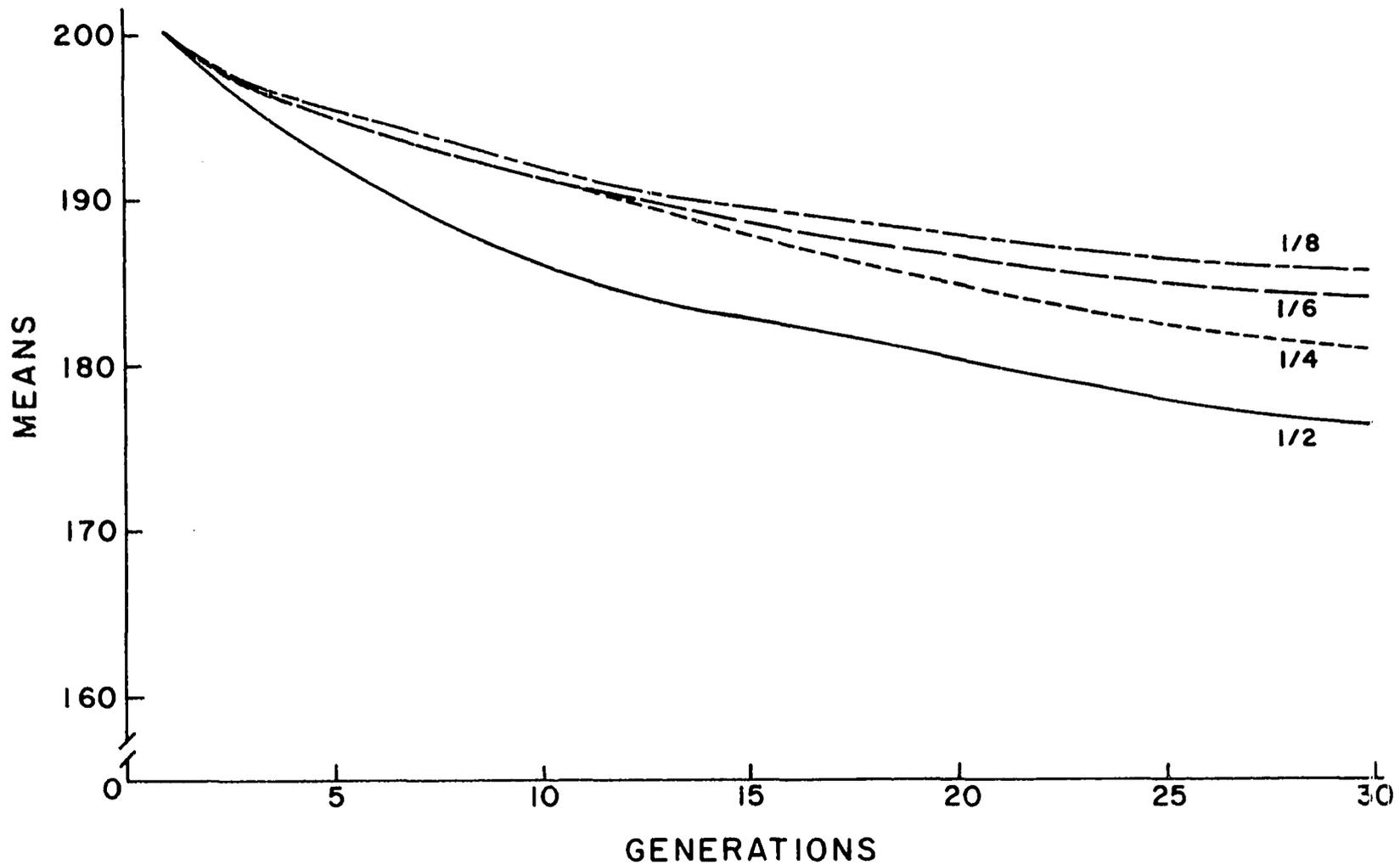


Figure 16. Mean genetic progress by level of selection intensity with overdominance

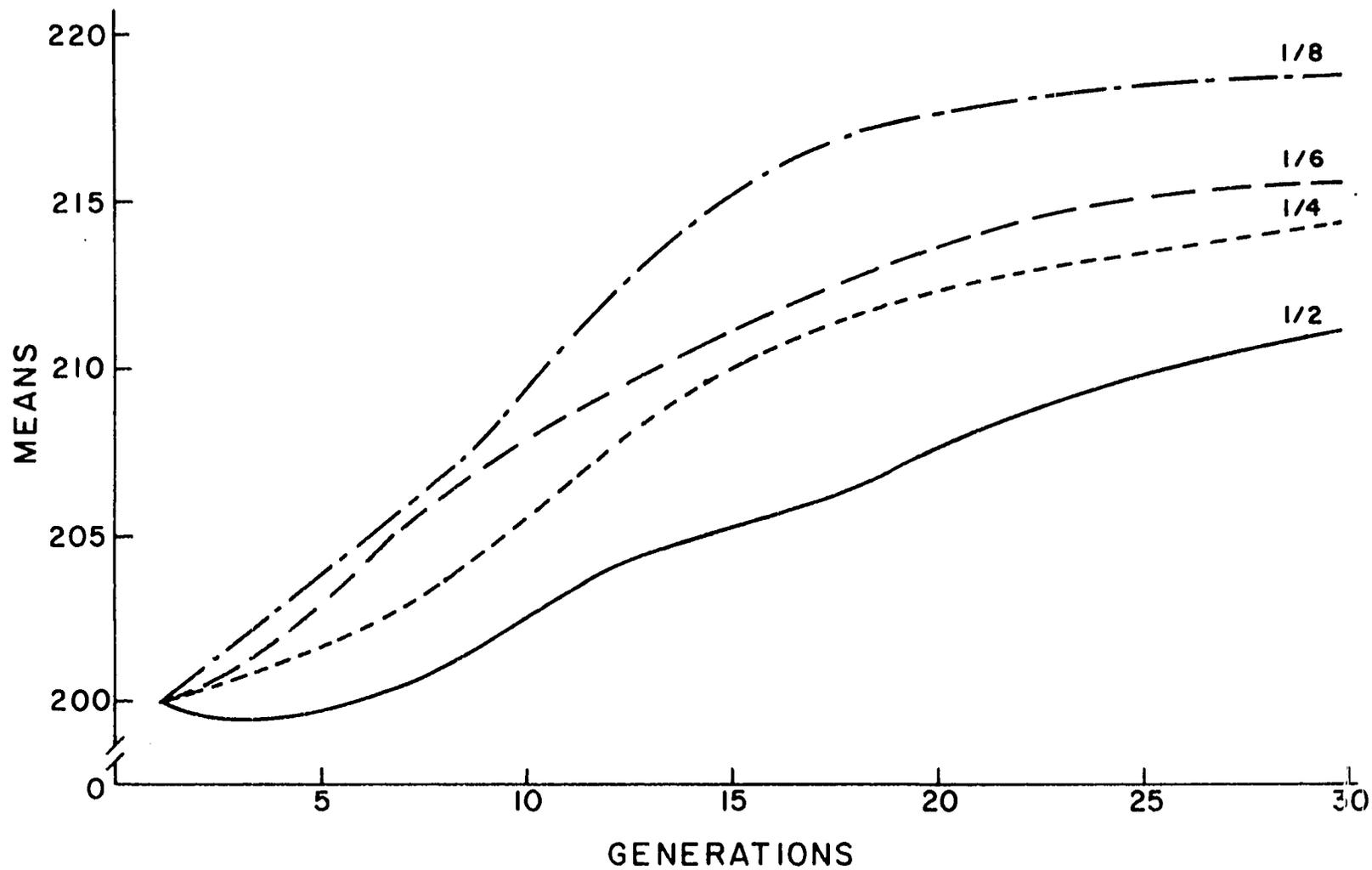


Figure 17. Mean genetic progress by level of selection intensity with optimum number of genes intermediate

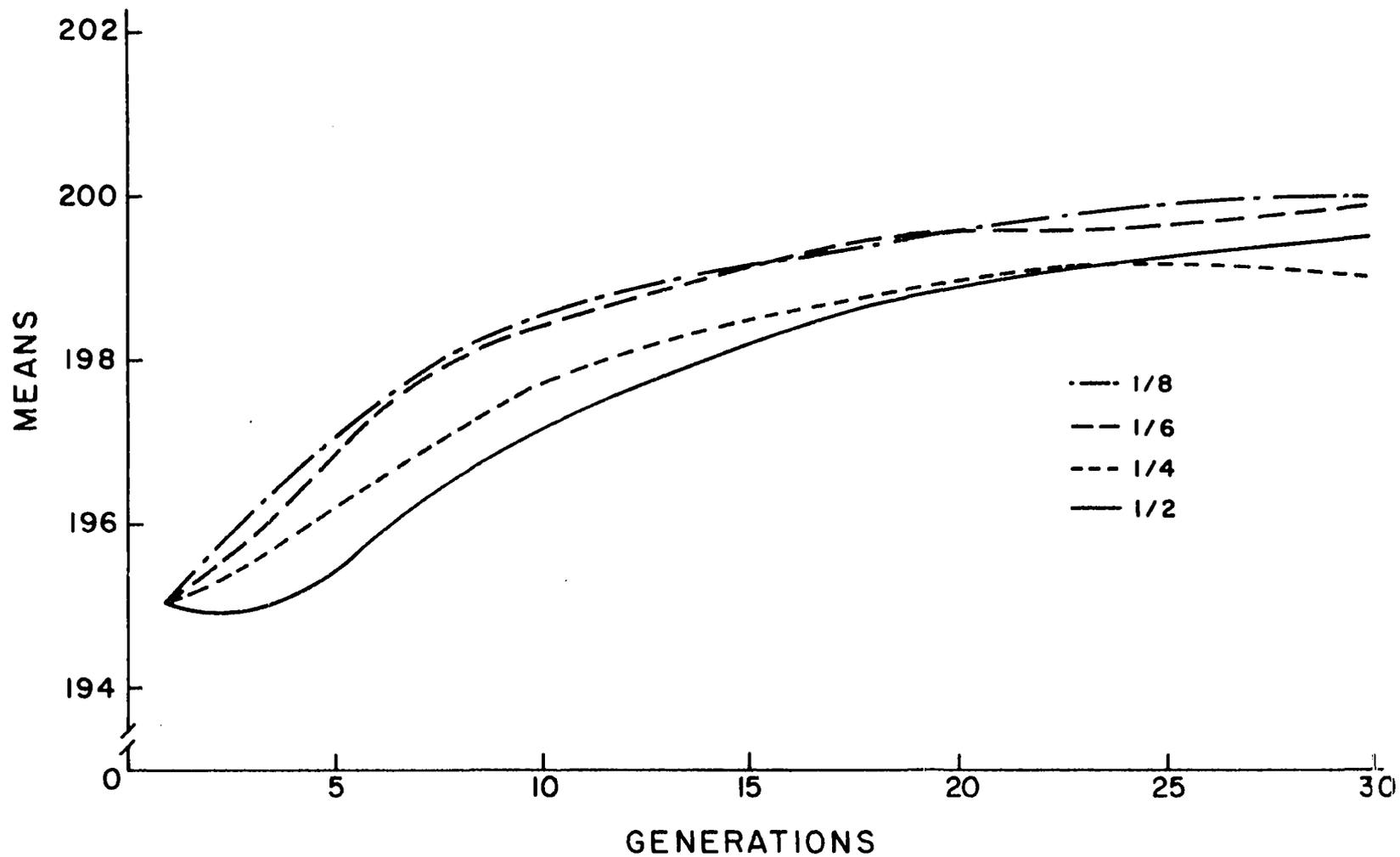


Figure 18. Mean genetic progress by level of selection intensity with duplicate gene action

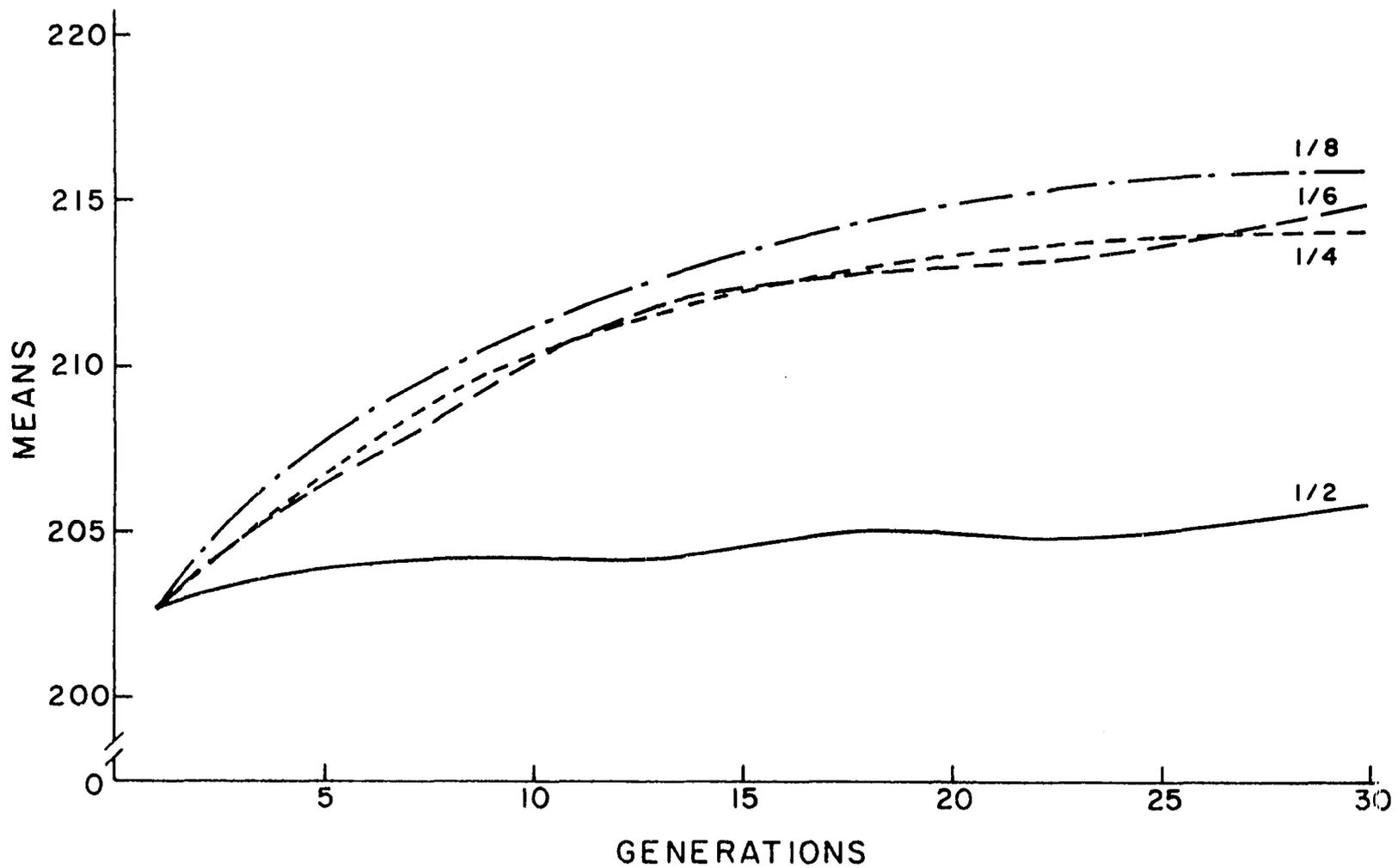


Figure 19. Mean genetic progress by level of selection intensity with complementary gene action

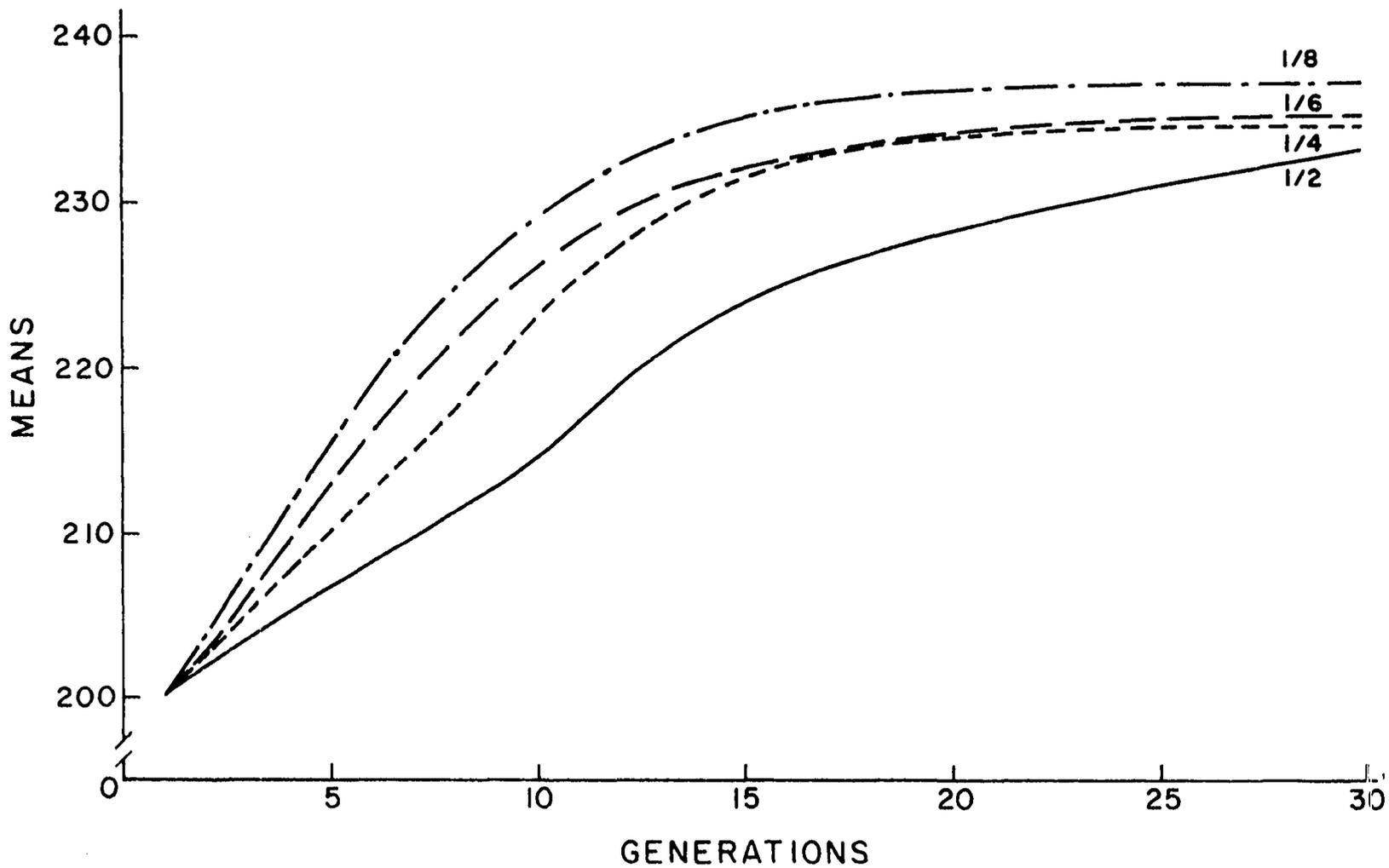


Figure 20. Mean genetic progress by level of selection intensity with additive-by-additive interaction

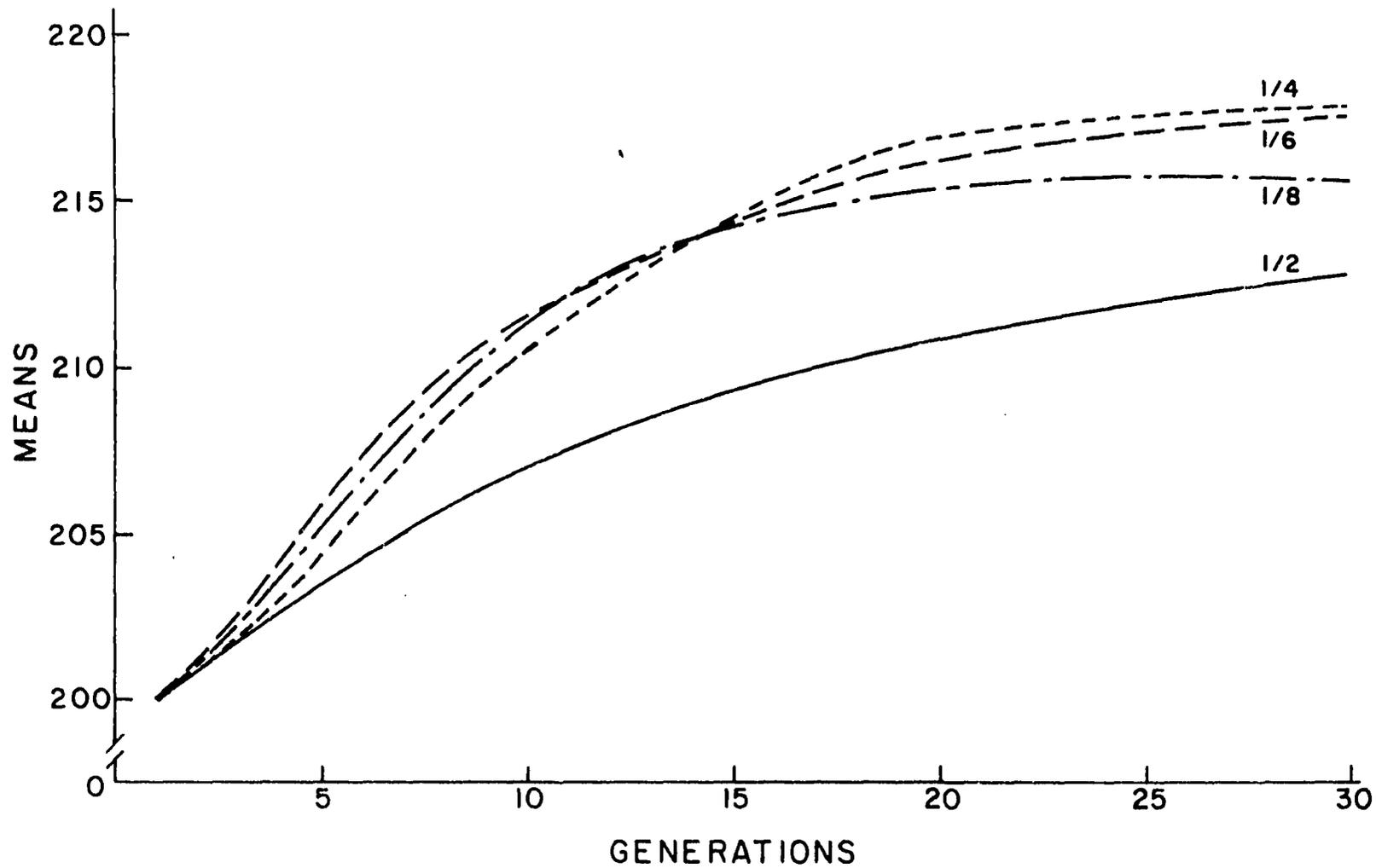


Figure 21. Mean genetic progress by level of selection intensity with additive-by-dominance interaction

γ

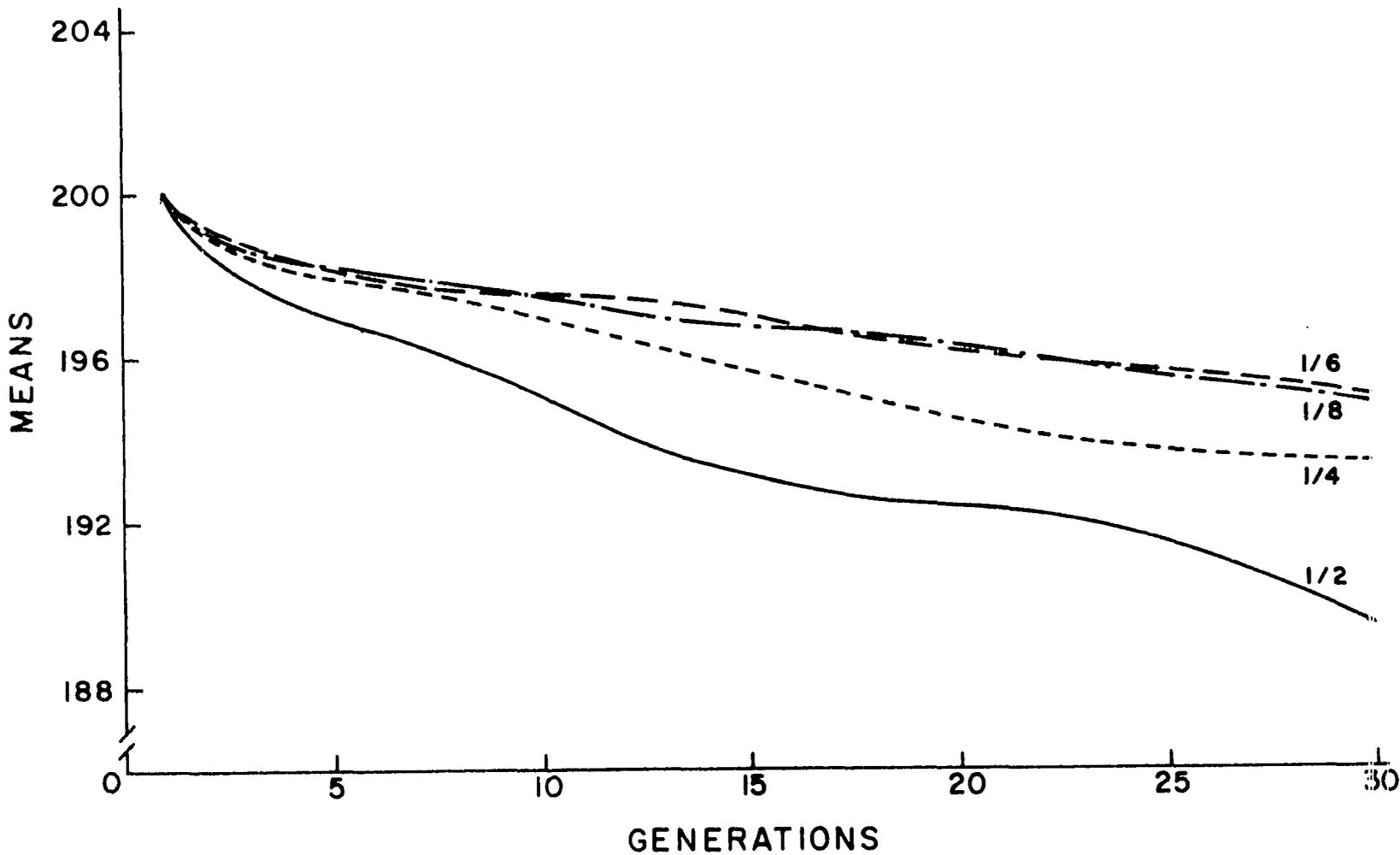


Figure 22. Mean genetic progress by level of selection intensity with dominance-by-dominance interaction

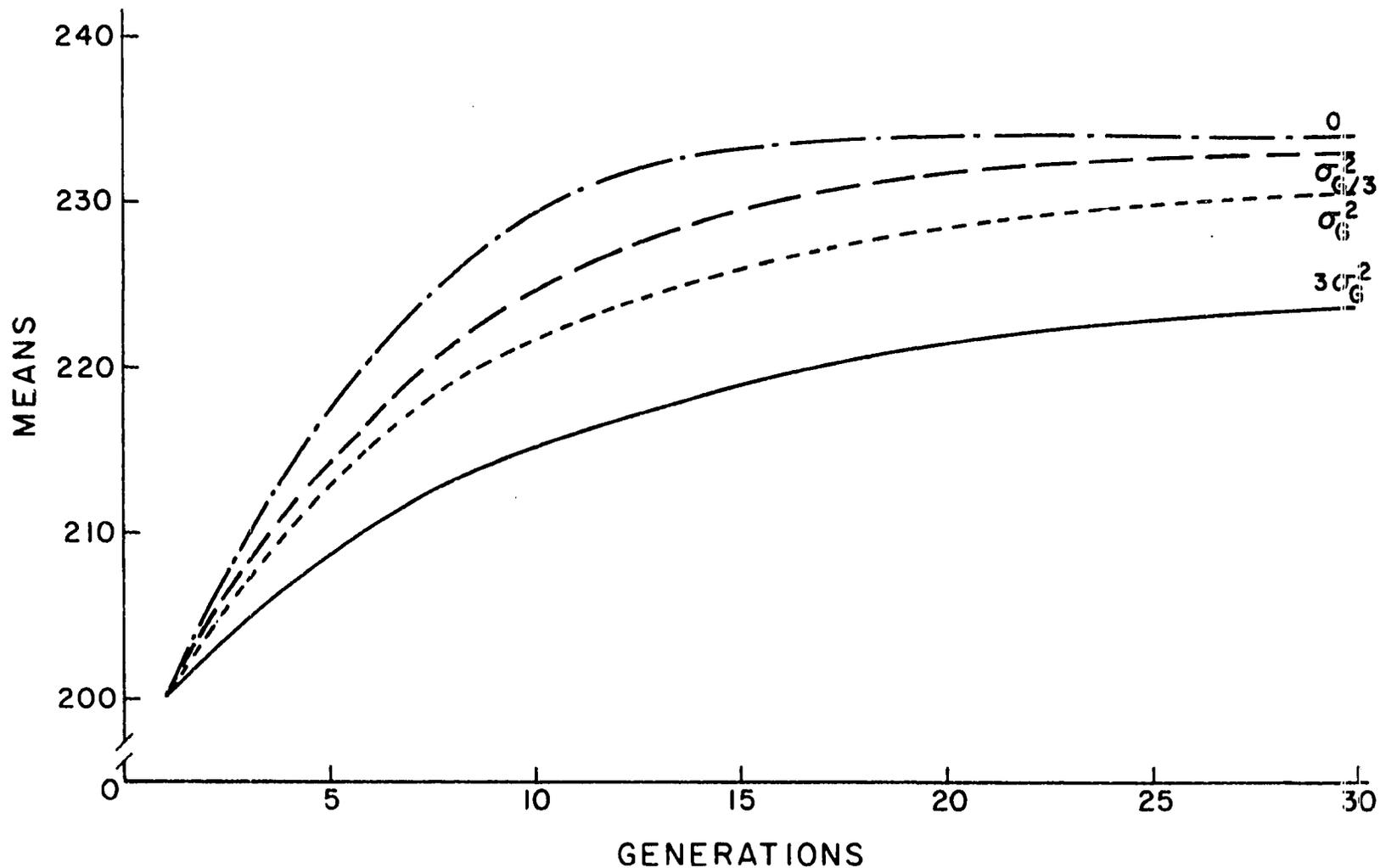


Figure 23. Mean genetic progress by level of environmental variation with additive gene action

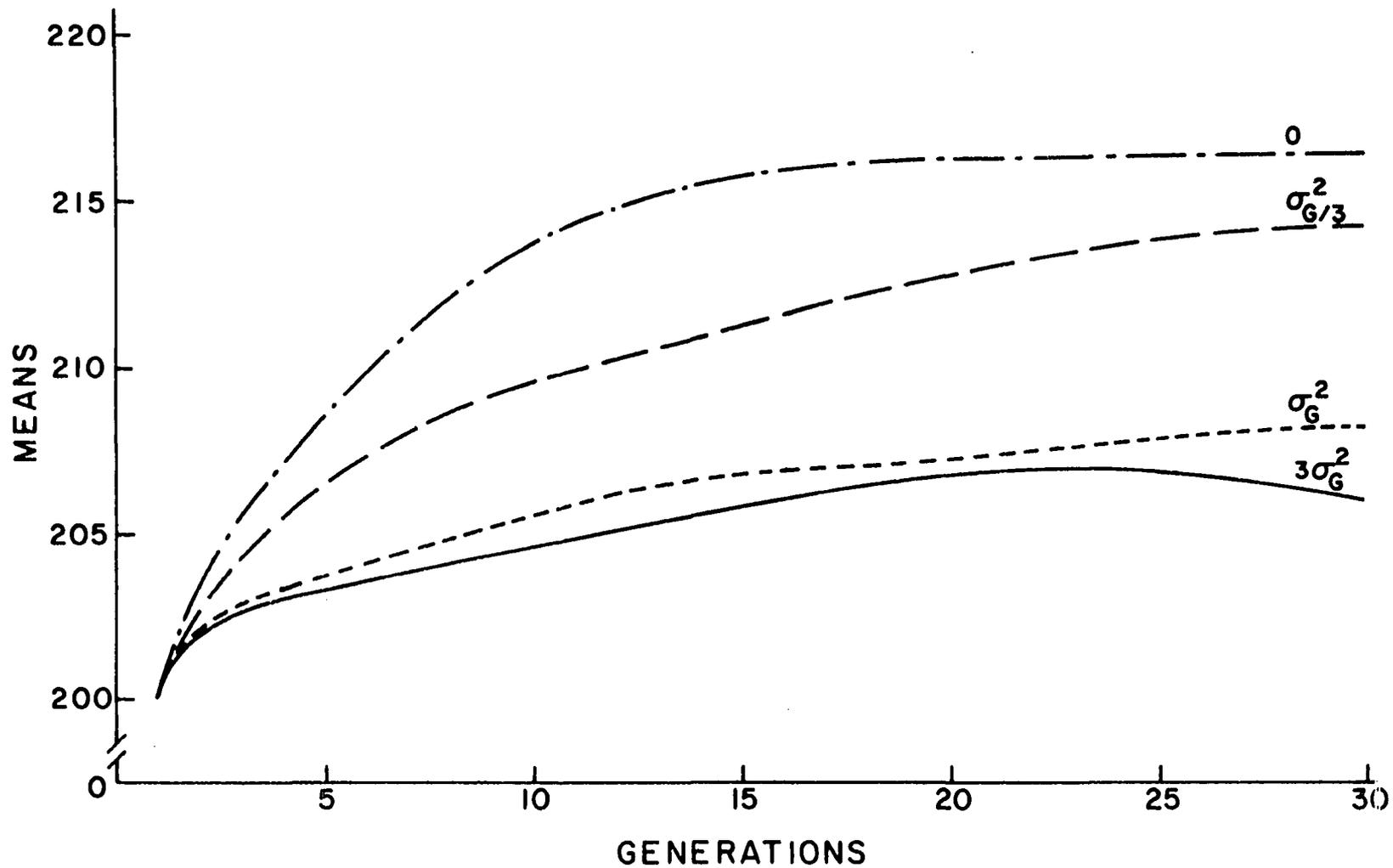


Figure 24. Mean genetic progress by level of environmental variation with complete dominance

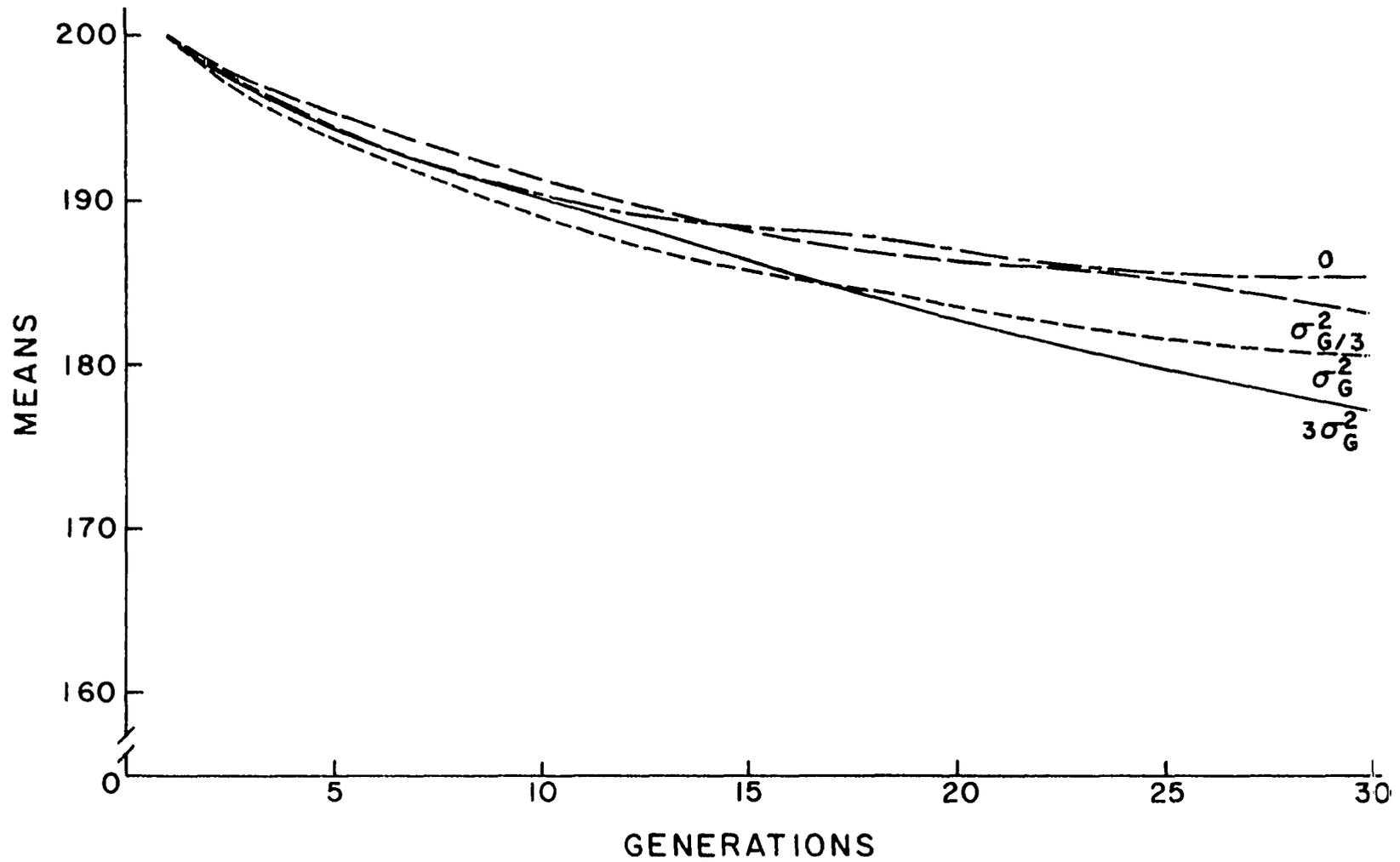


Figure 25. Mean genetic progress by level of environmental variation with overdominance

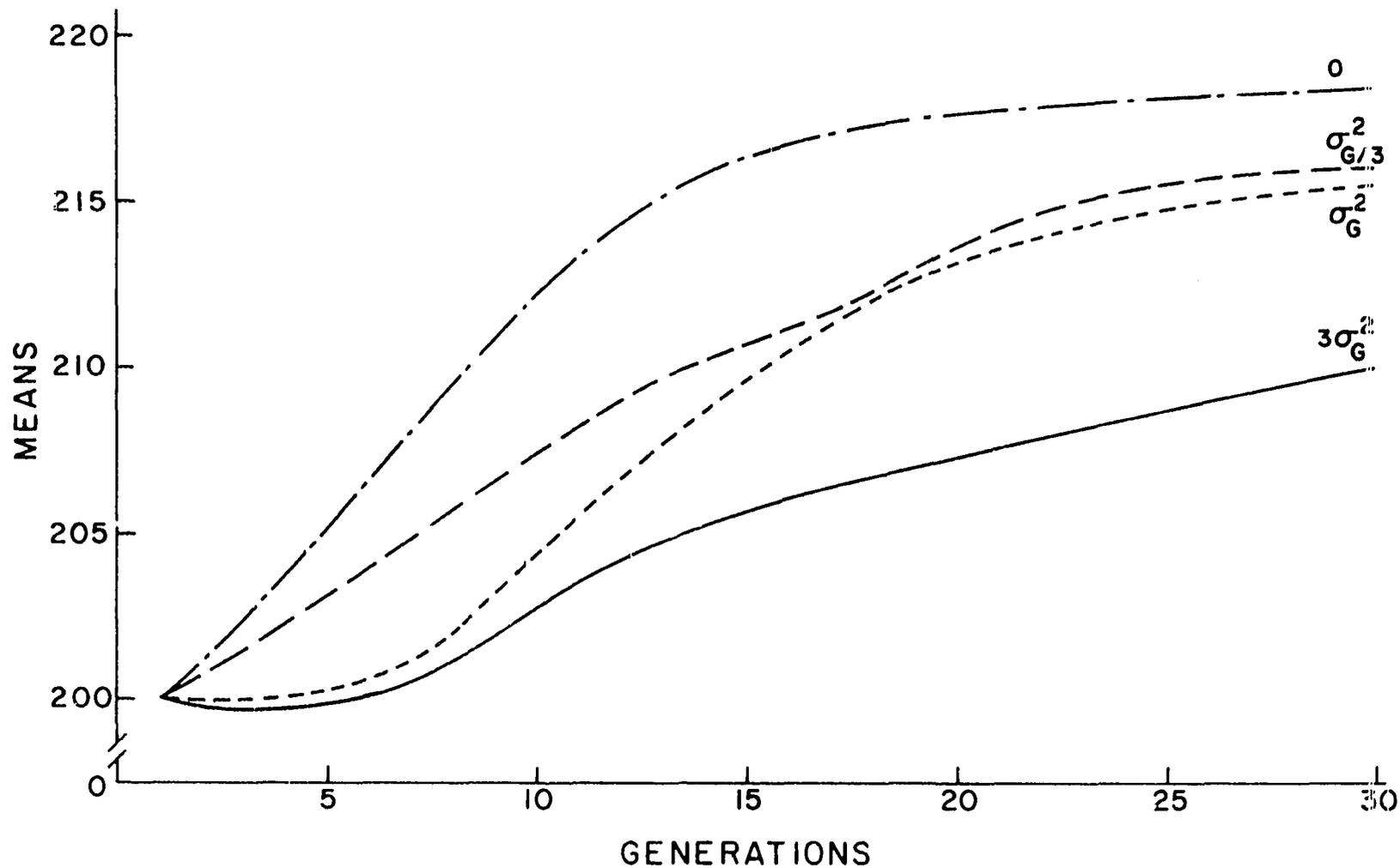


Figure 26. Mean genetic progress by level of environmental variation with optimum number of genes intermediate

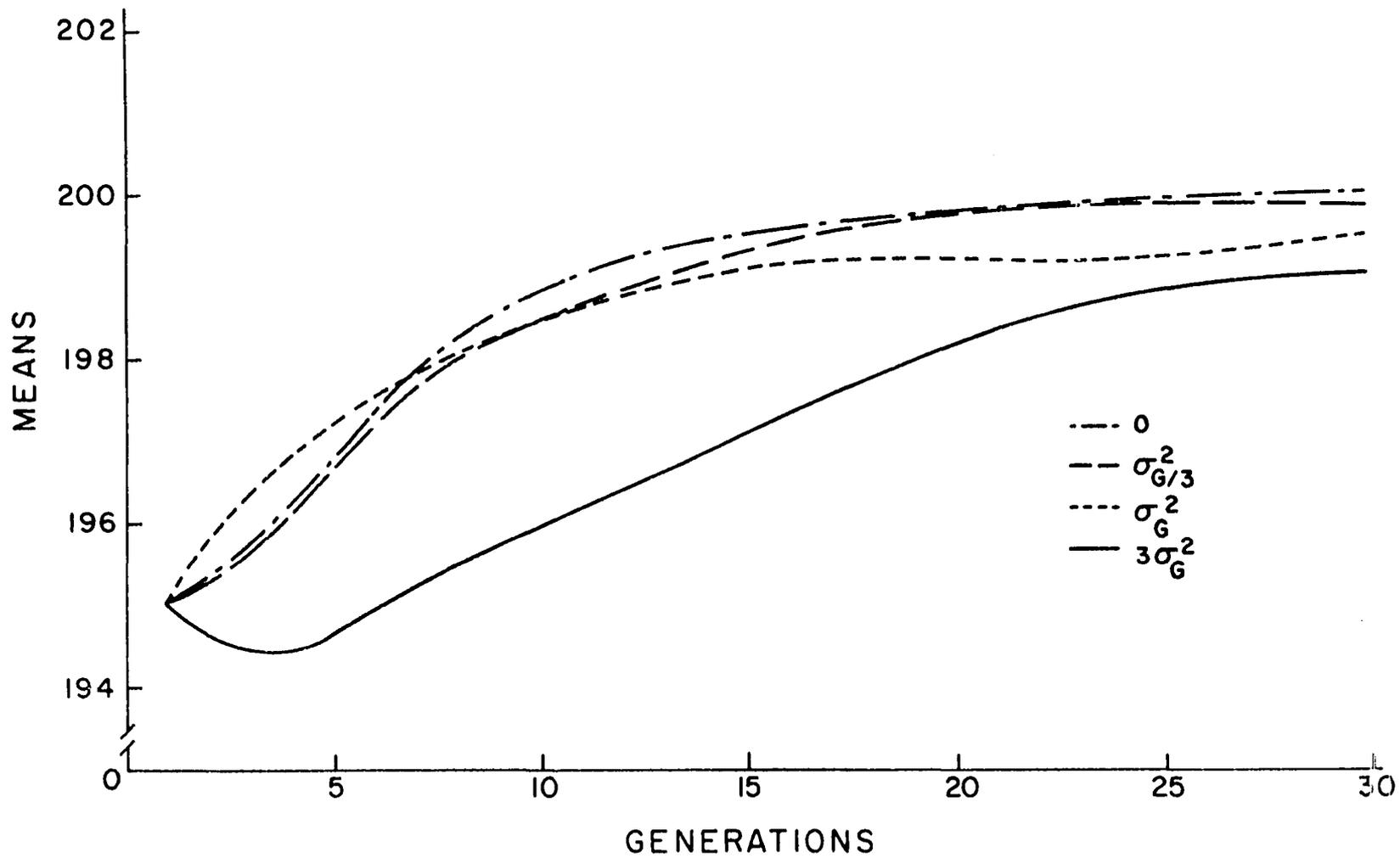


Figure 27. Mean genetic progress by level of environmental variation with duplicate gene action

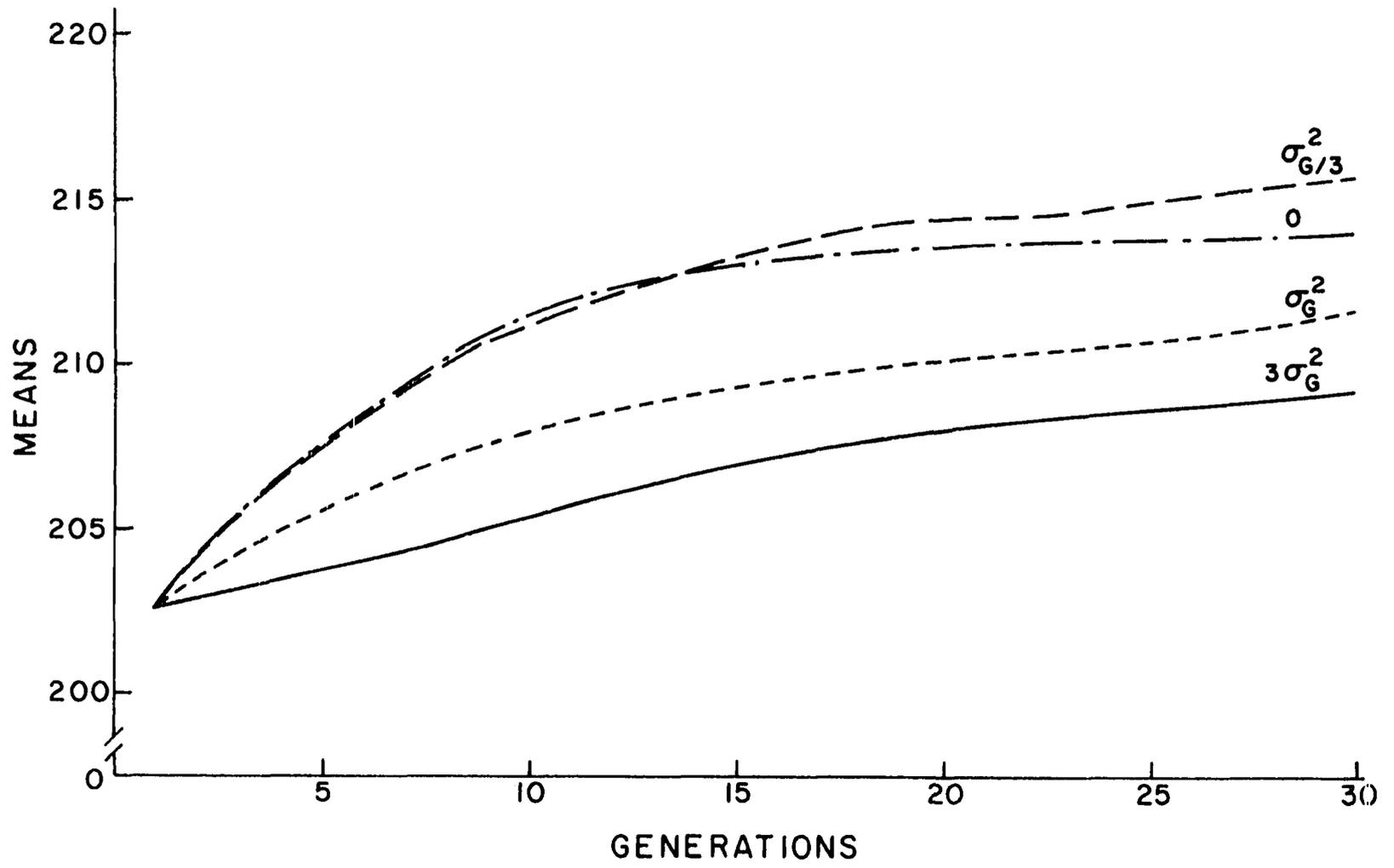


Figure 28. Mean genetic progress by level of environmental variation with complementary gene action

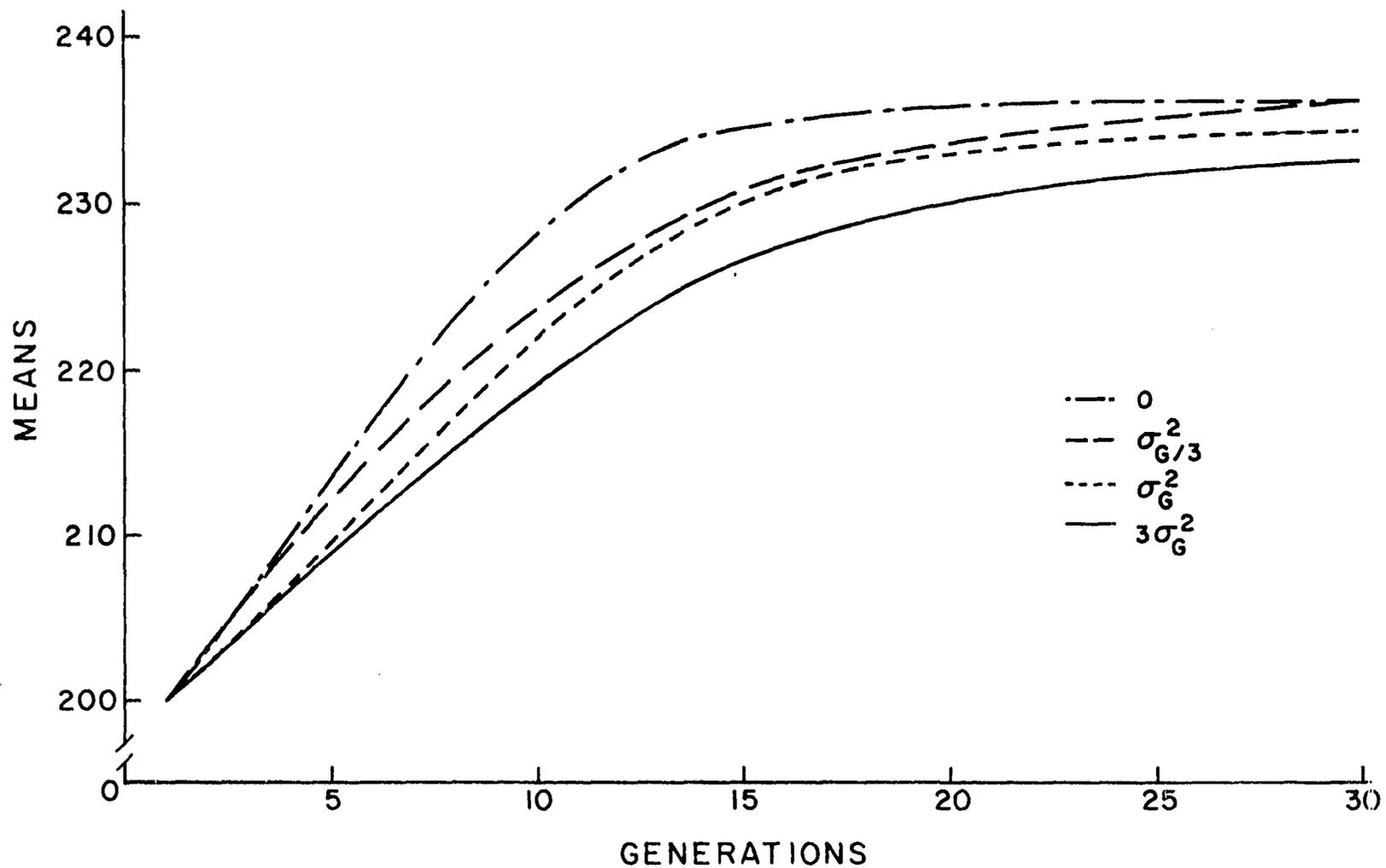


Figure 29. Mean genetic progress by level of environmental variation with additive-by-additive interaction

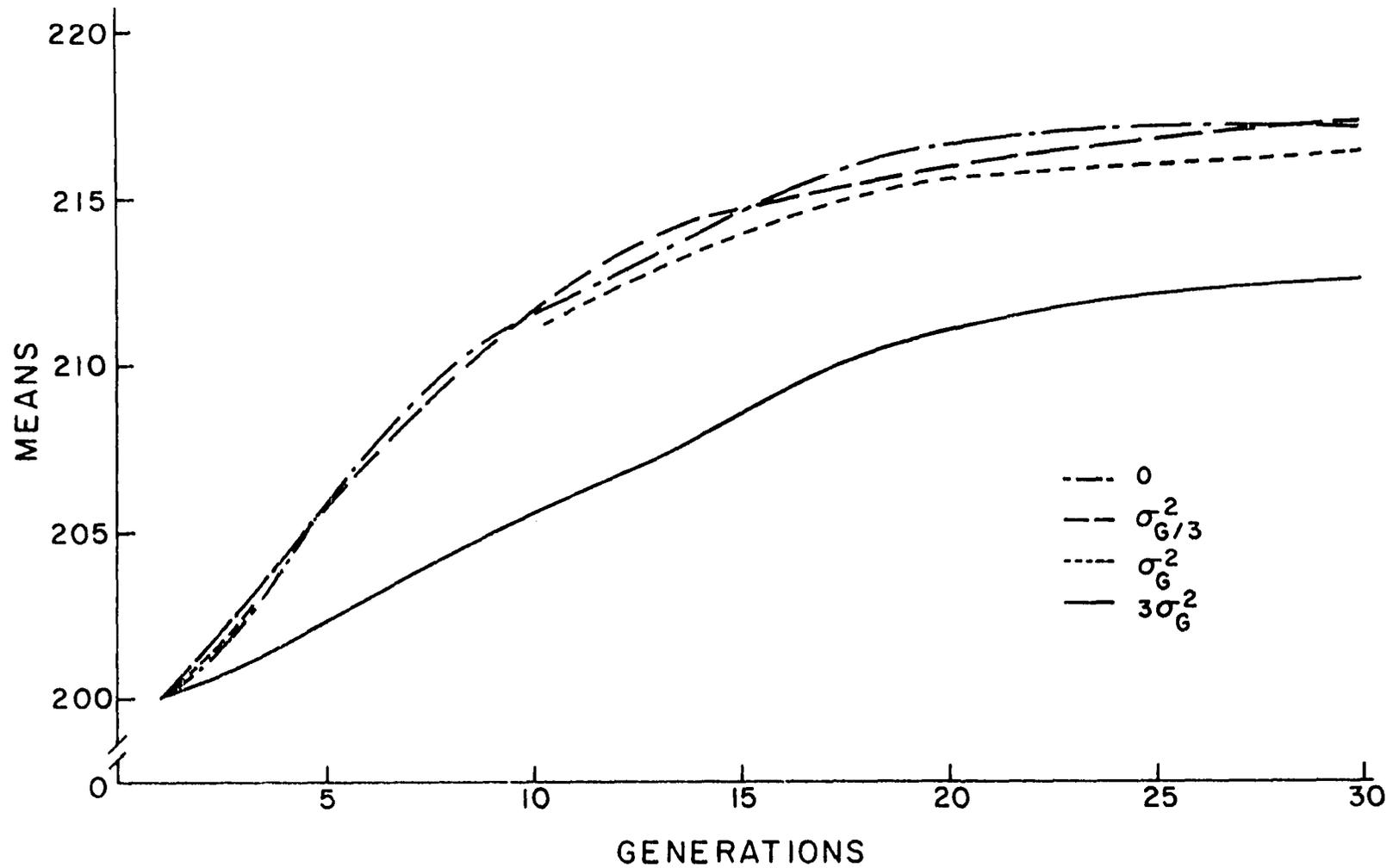


Figure 30. Mean genetic progress by level of environmental variation with additive-by-dominance interaction

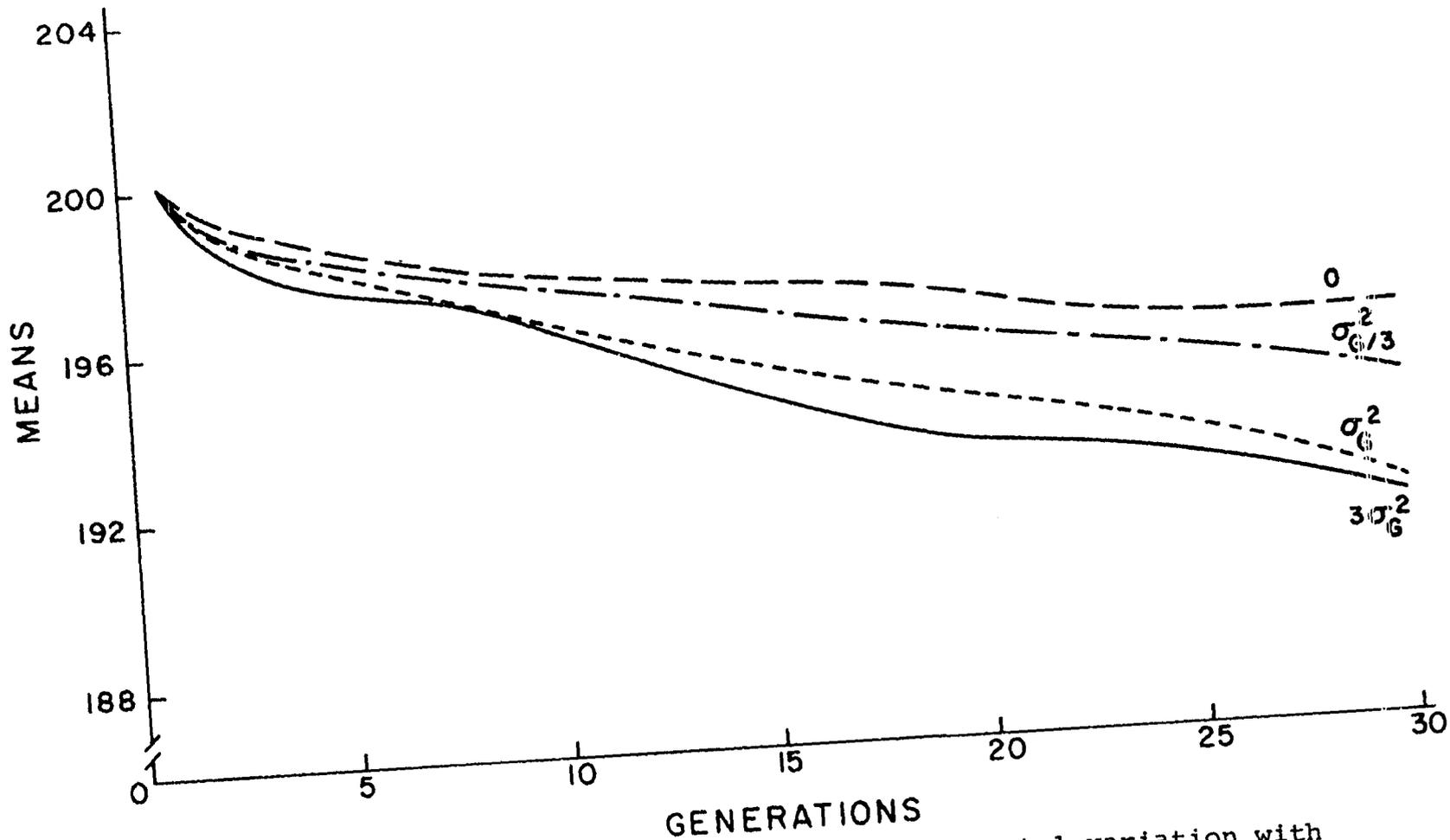


Figure 31. Mean genetic progress by level of environmental variation with dominance-by-dominance interaction

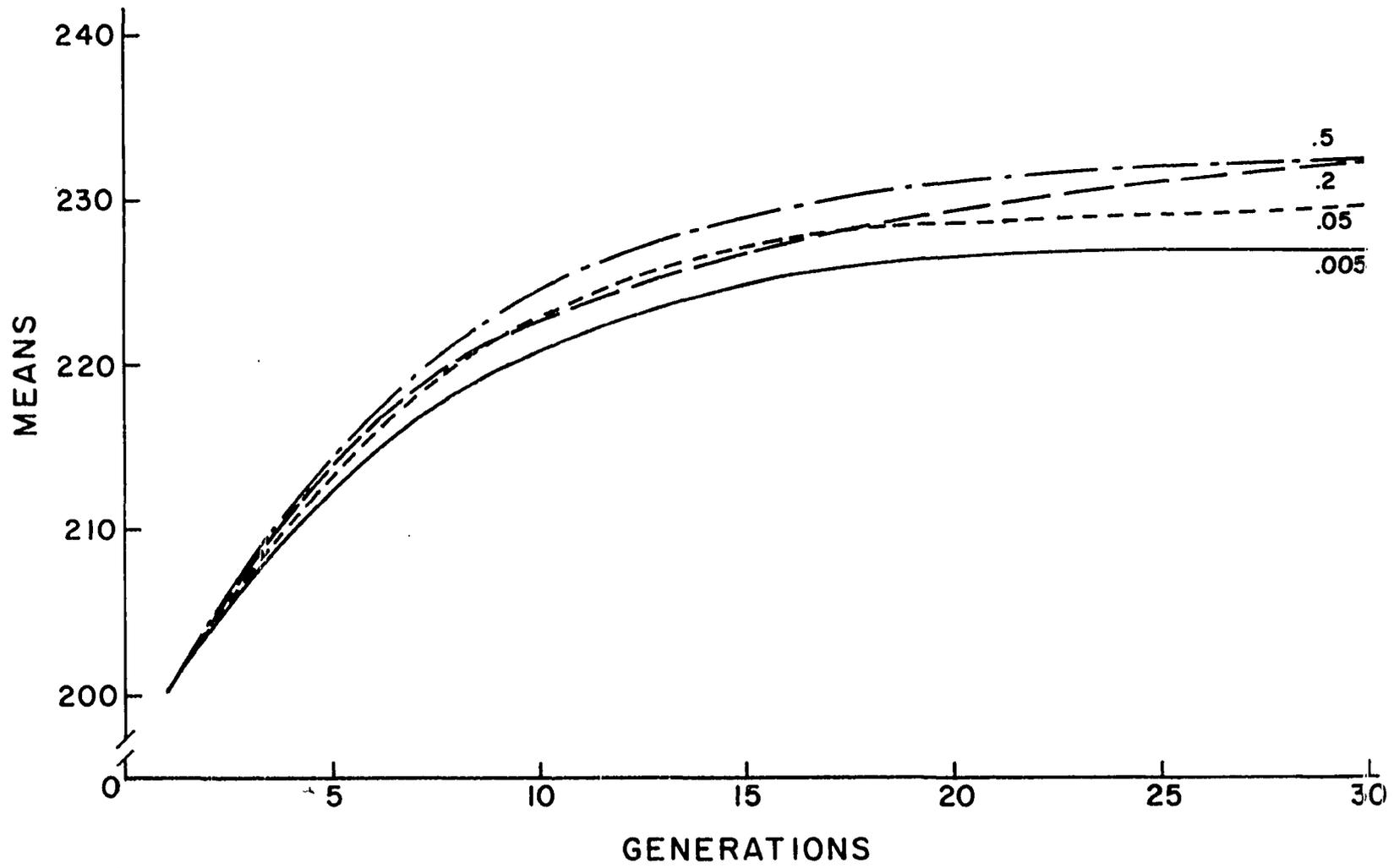


Figure 32. Mean genetic progress by level of linkage with additive gene action

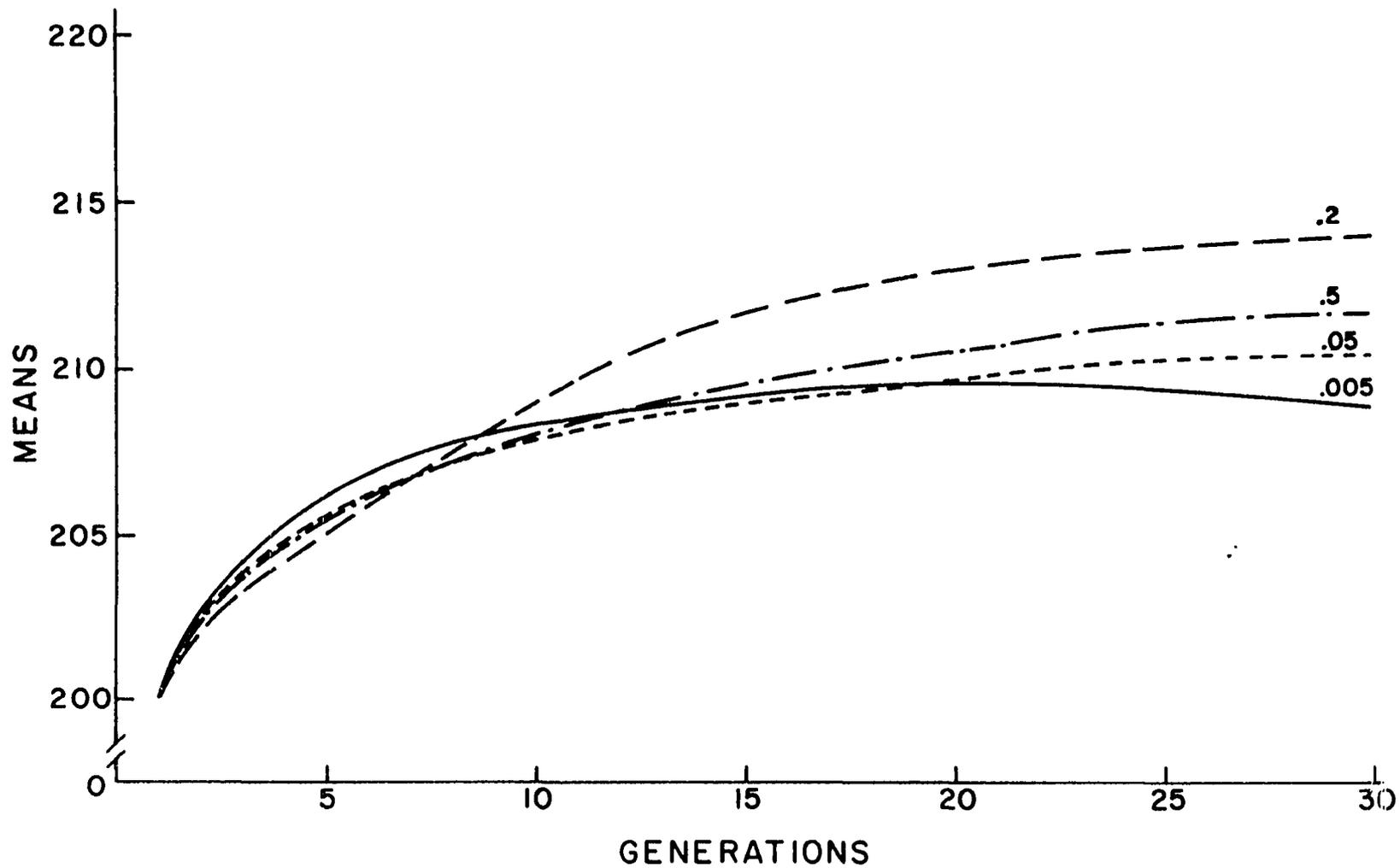


Figure 33. Mean genetic progress by level of linkage with complete dominance

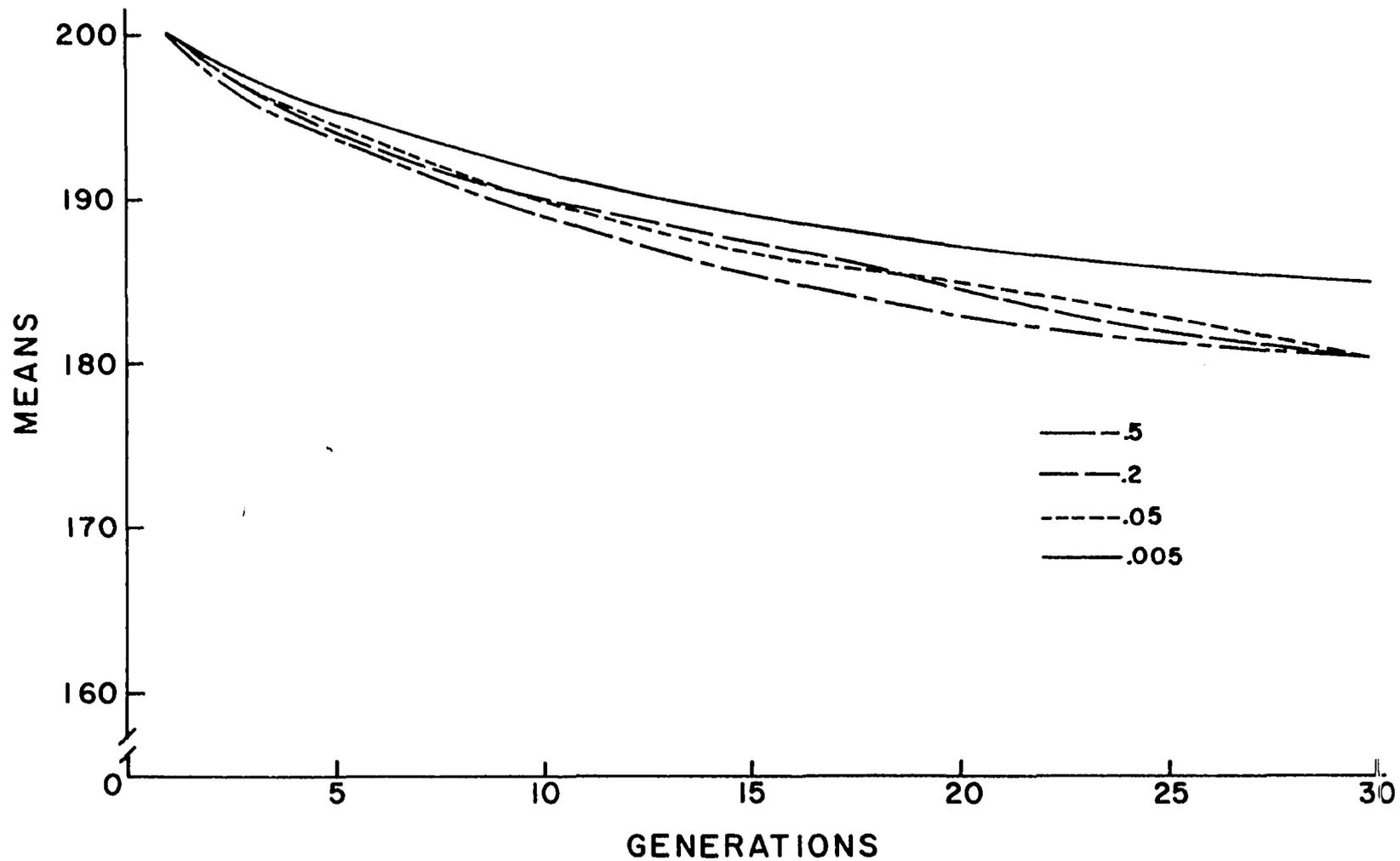


Figure 34. Mean genetic progress by level of linkage with overdominance

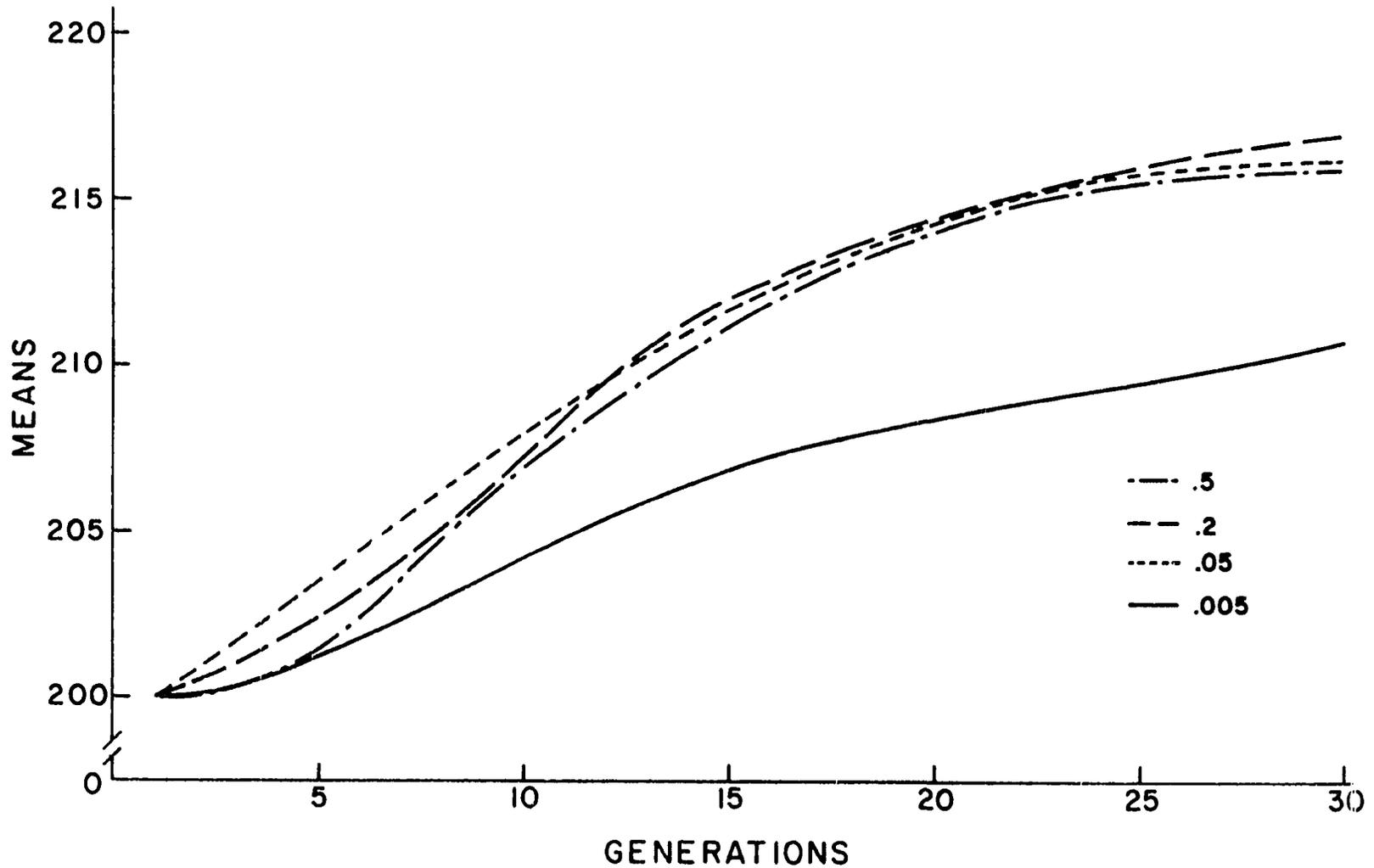


Figure 35. Mean genetic progress by level of linkage with optimum number of genes intermediate

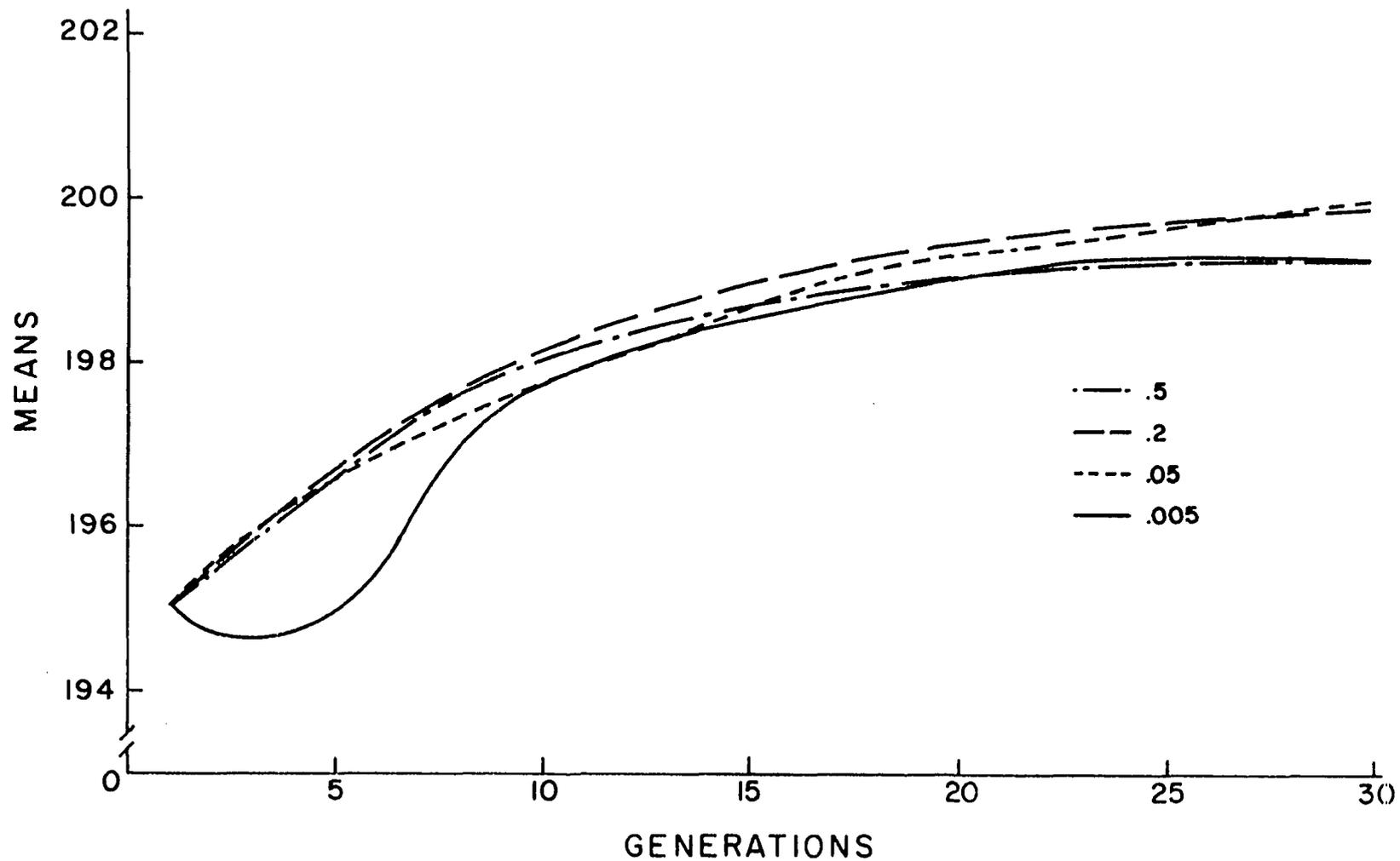


Figure 36. Mean genetic progress by level of linkage with duplicate gene action

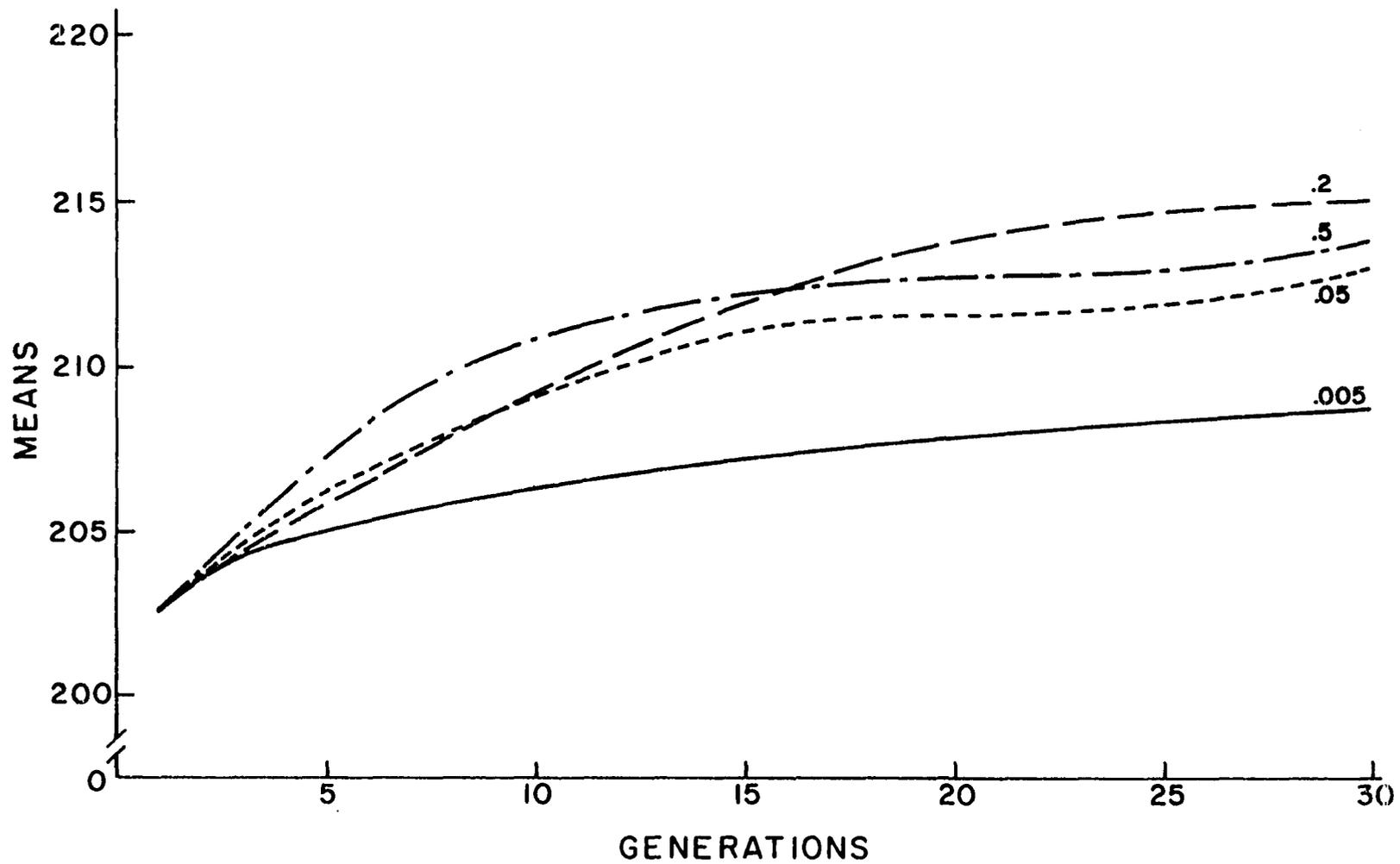


Figure 37. Mean genetic progress by level of linkage with complementary gene action

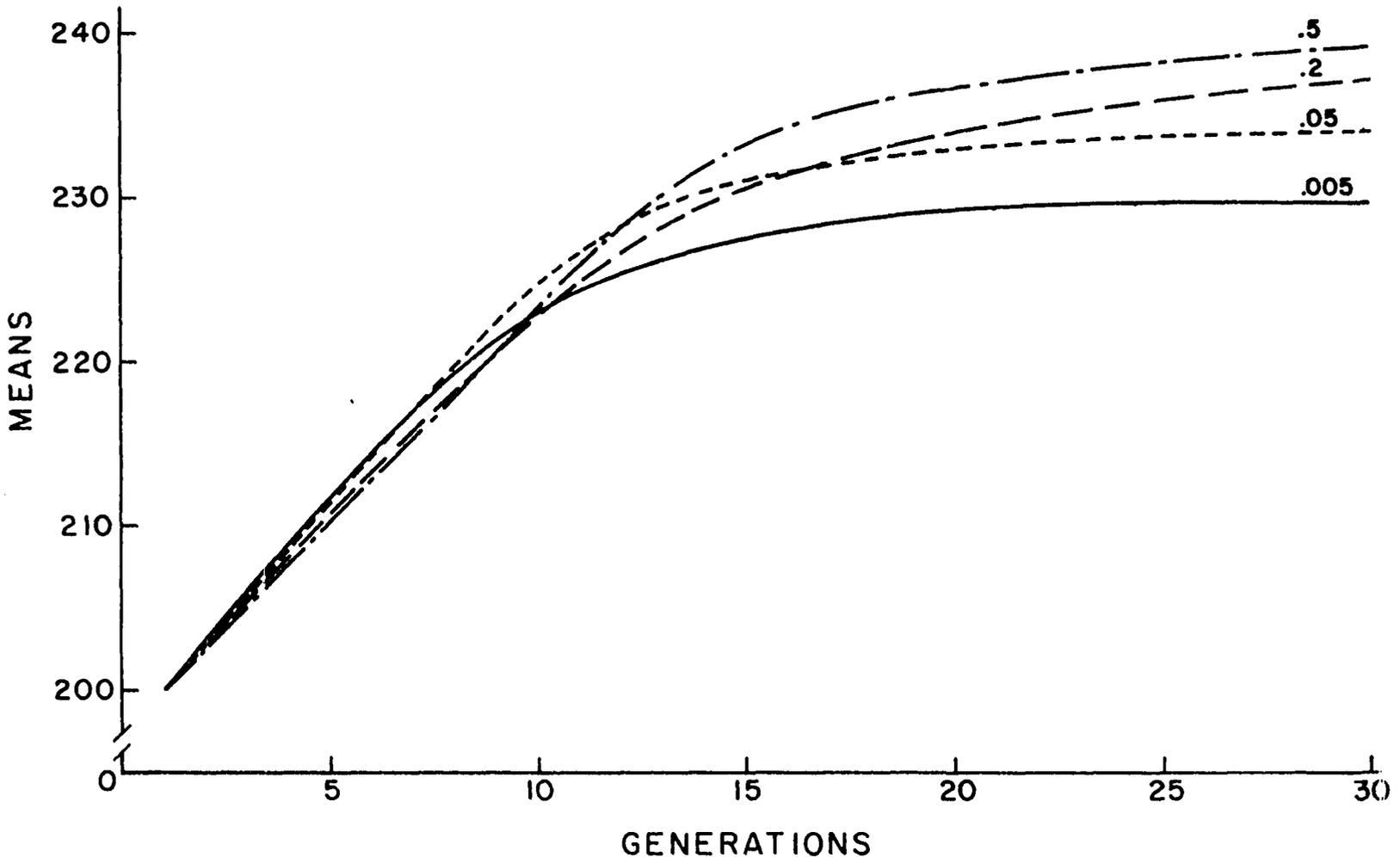


Figure 38. Mean genetic progress by level of linkage with additive-by-additive interaction

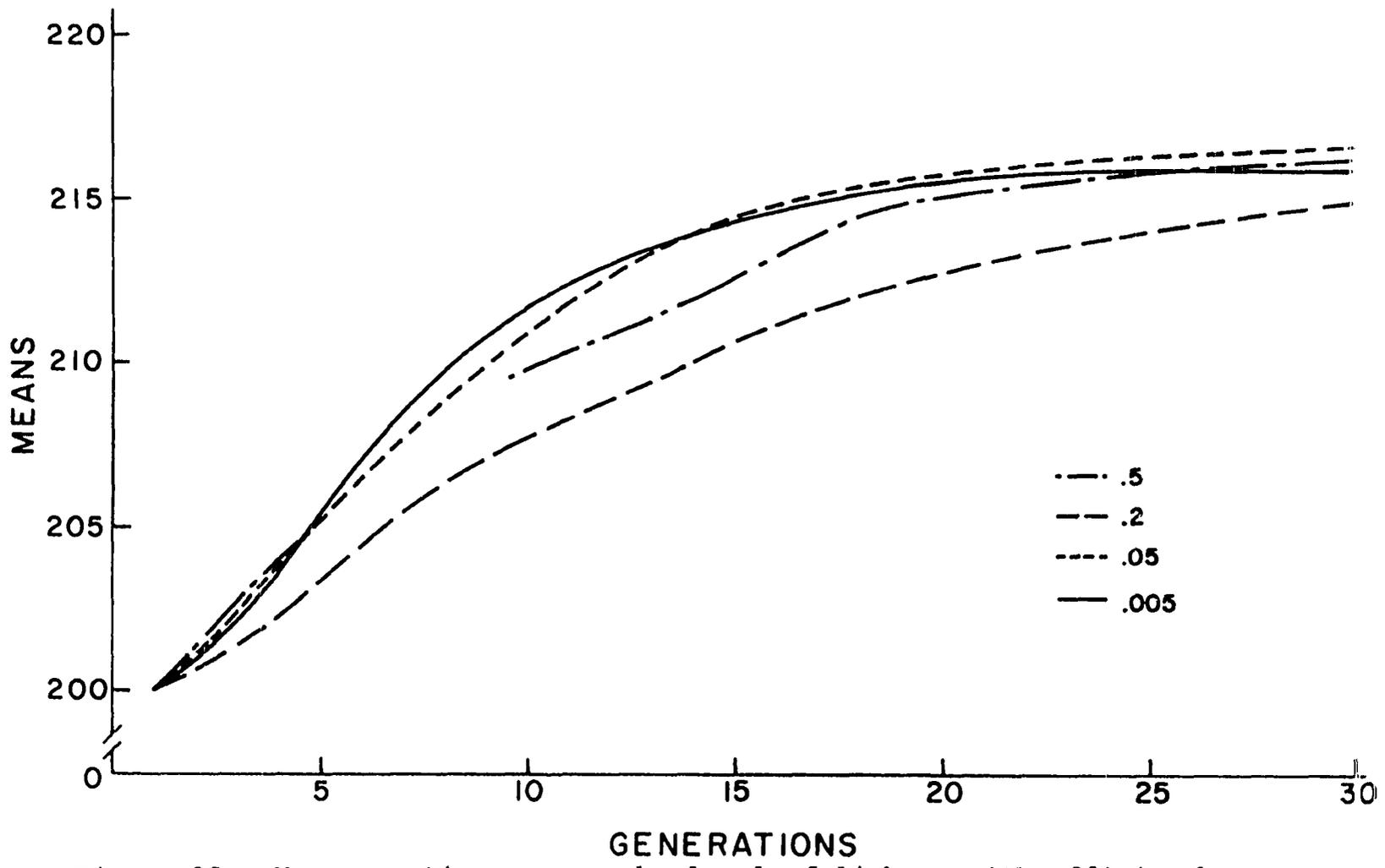


Figure 39. Mean genetic progress by level of linkage with additive-by-dominance interaction

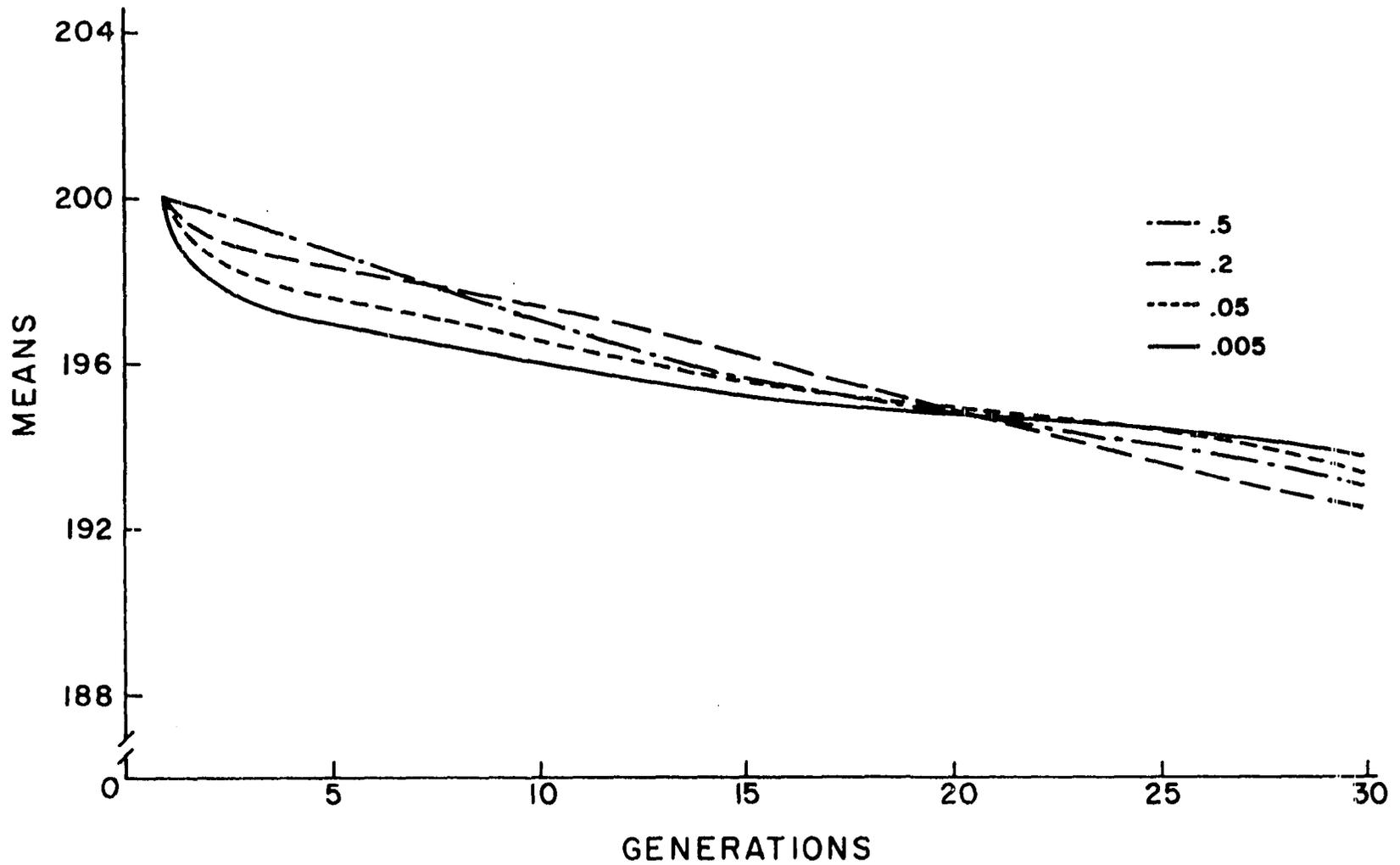


Figure 40. Mean genetic progress by level of linkage with dominance-by-dominance interaction

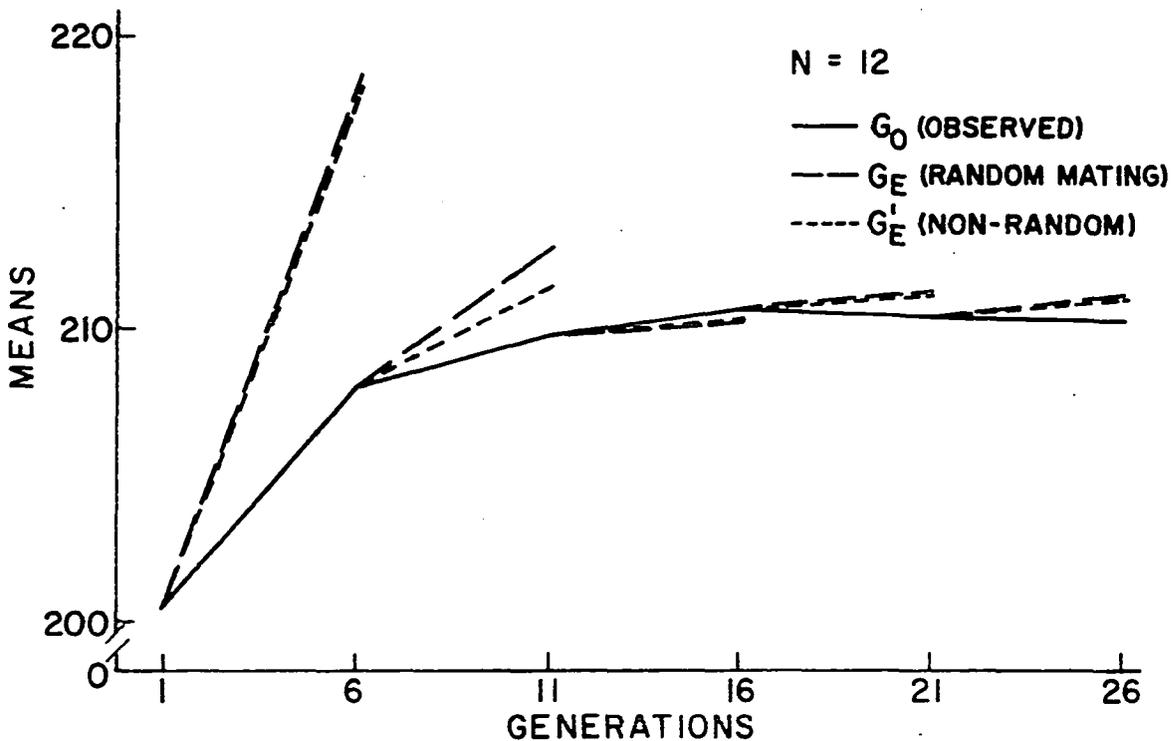
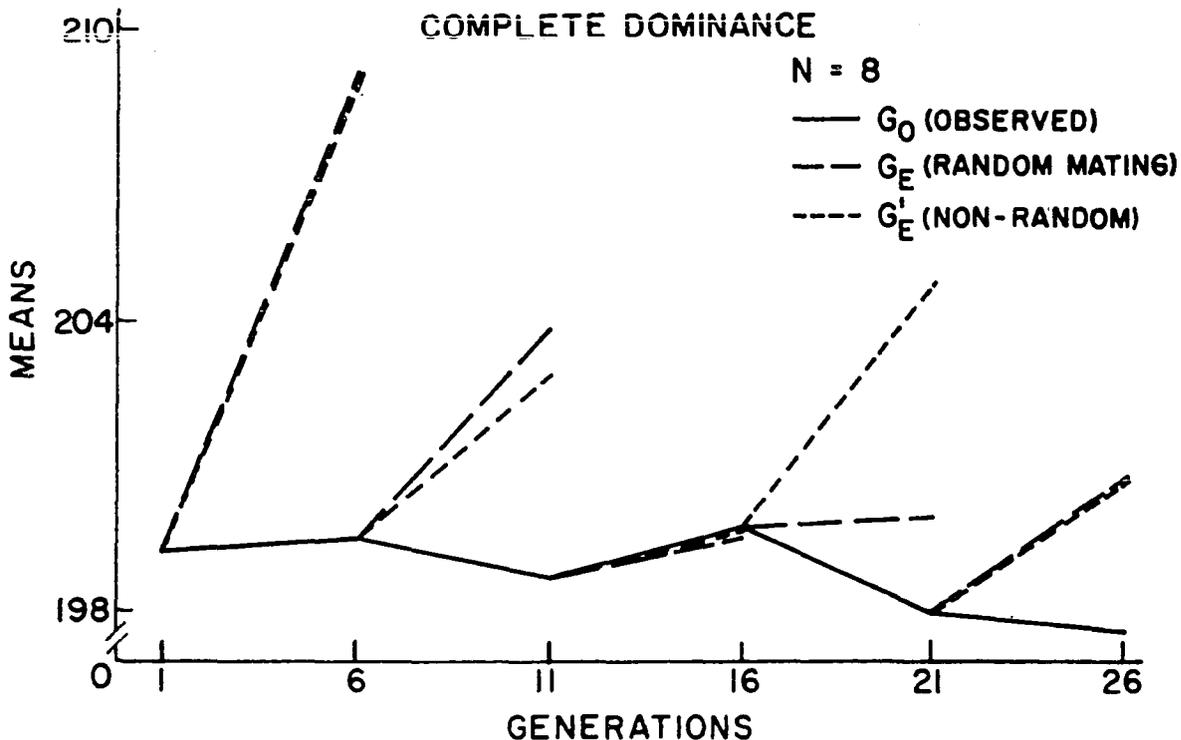


Figure 41. Relation of change in mean (G_0) to that expected from changes in gene frequency (G_E, G'_E) with complete dominance and N = 8, 12

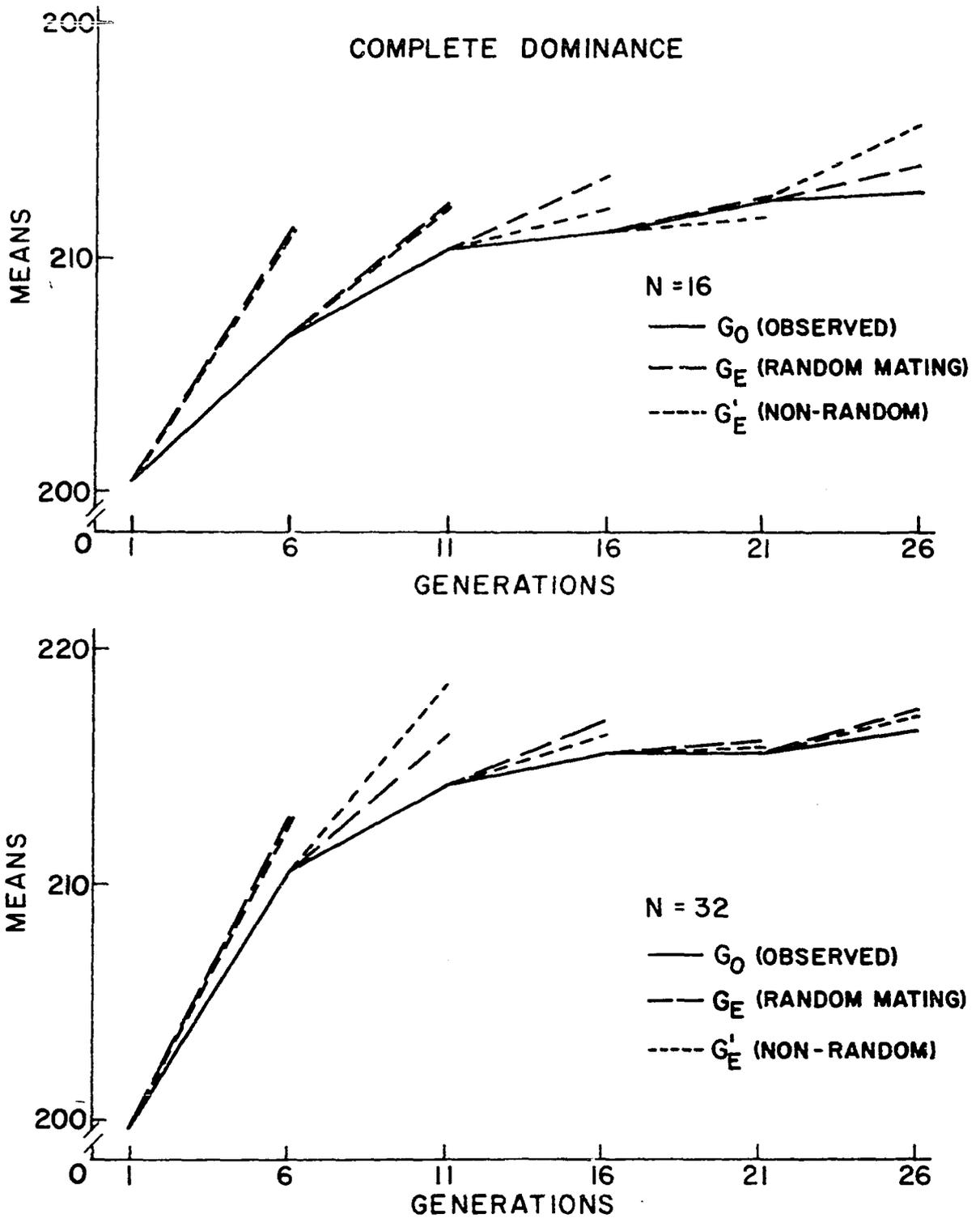


Figure 42. Relation of change in mean (G_0) to that expected from changes in gene frequency (G_E, G'_E) with complete dominance and $N = 16, 32$

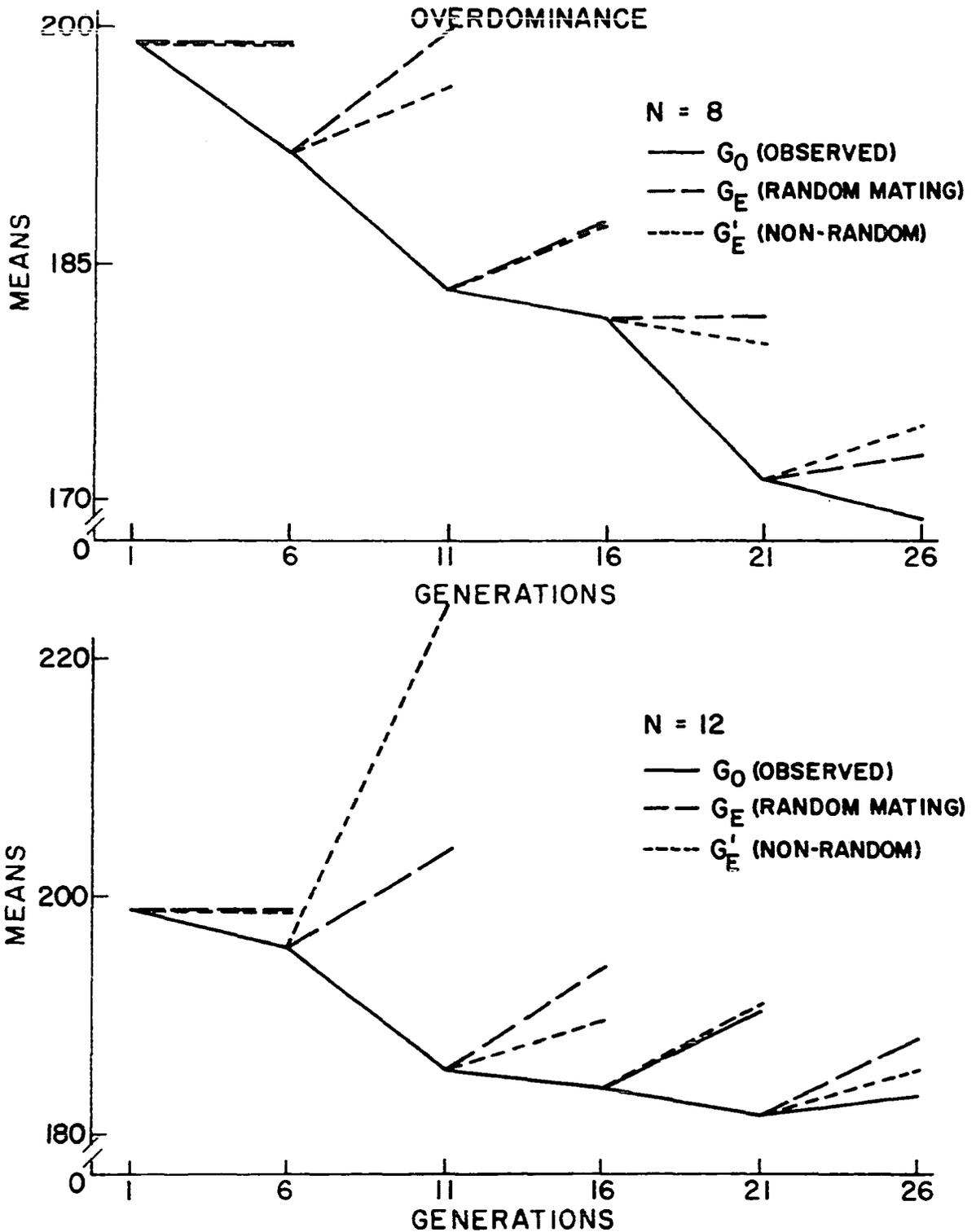


Figure 43. Relation of change in mean (G_0) to that expected from changes in gene frequency (G_E, G'_E) with overdominance and $N = 8, 12$

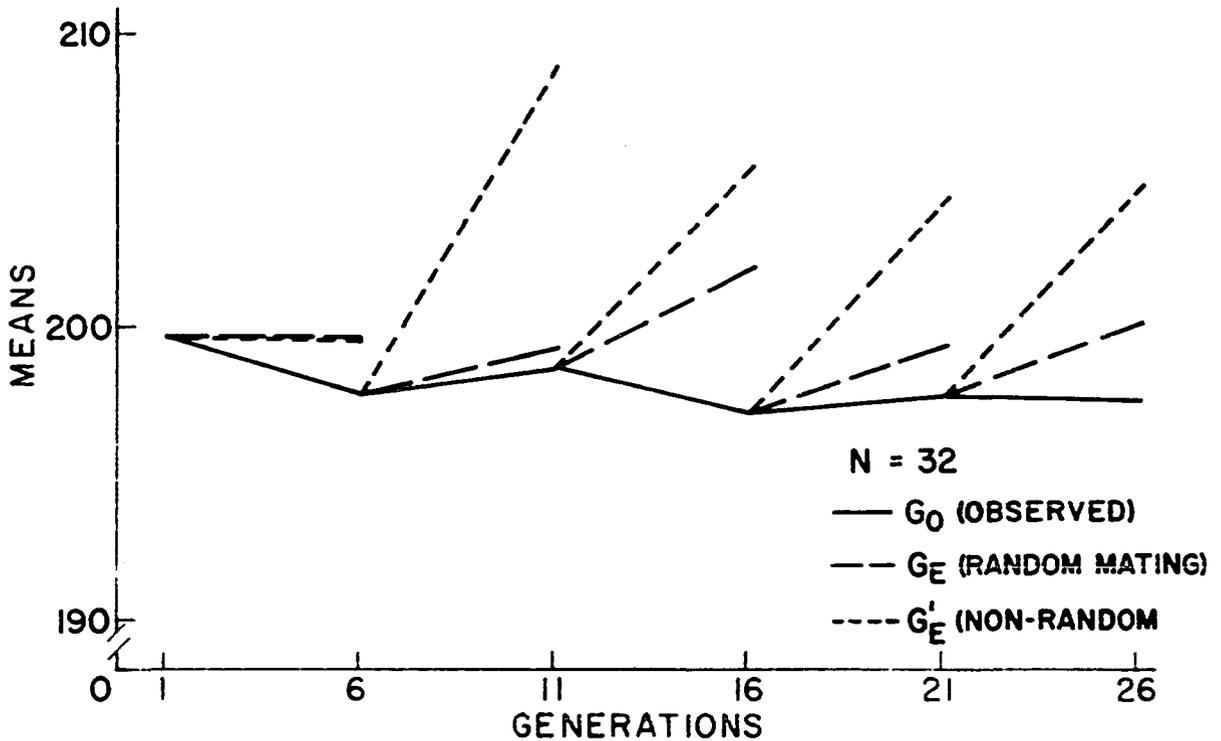
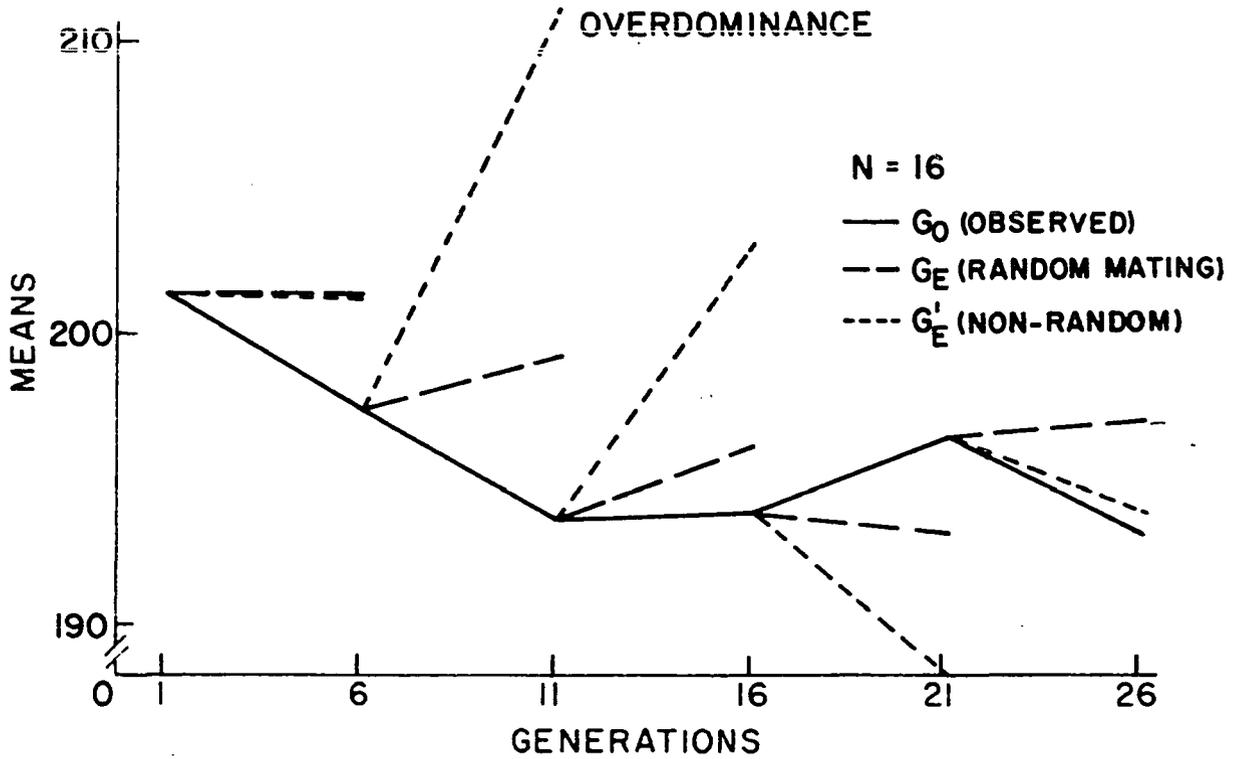


Figure 44. Relation of change in mean (G_0) to that expected from changes in gene frequency (G_E, G'_E) with overdominance and $N = 16, 32$

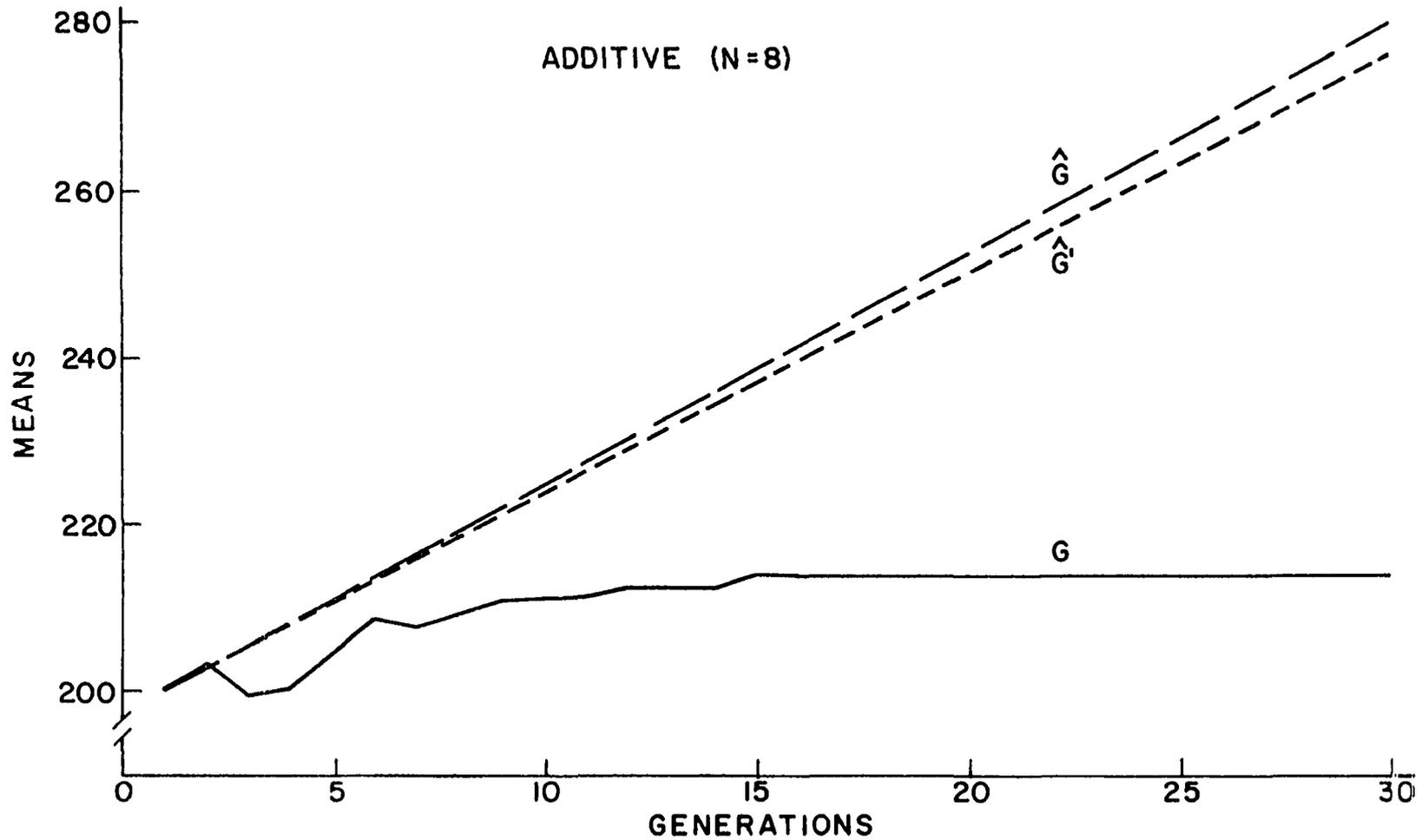


Figure 45. Relation of the mean (G) of a small population to that predicted from parameters of simulation (\hat{G}) and from parameters observed in generation 1 (\hat{G}') (additive gene action)

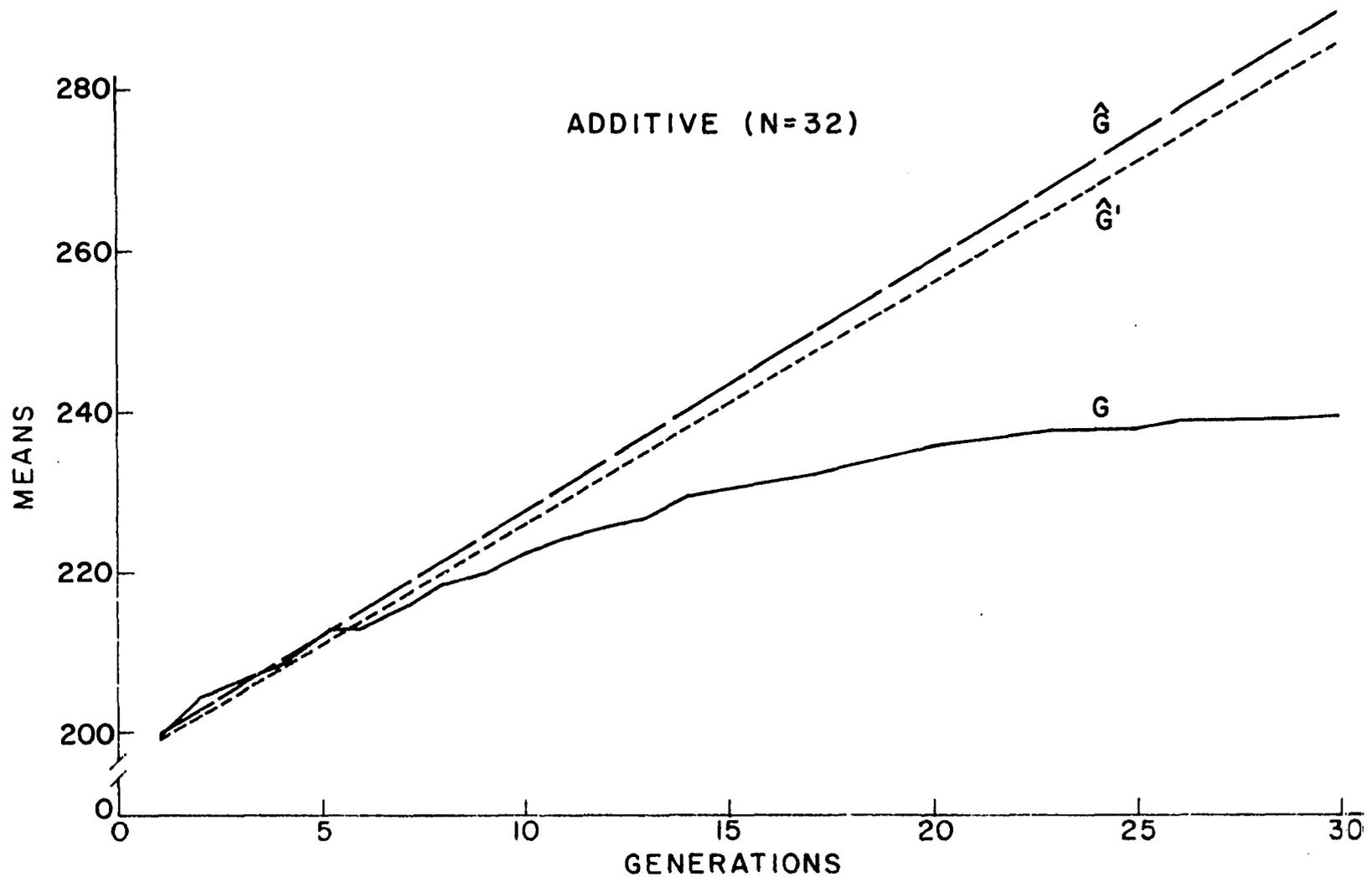


Figure 46. Relation of the mean (G) of a "large" population to that predicted from parameters of simulation (\hat{G}) and from parameters observed in generation 1 (\hat{G}') (additive gene action)

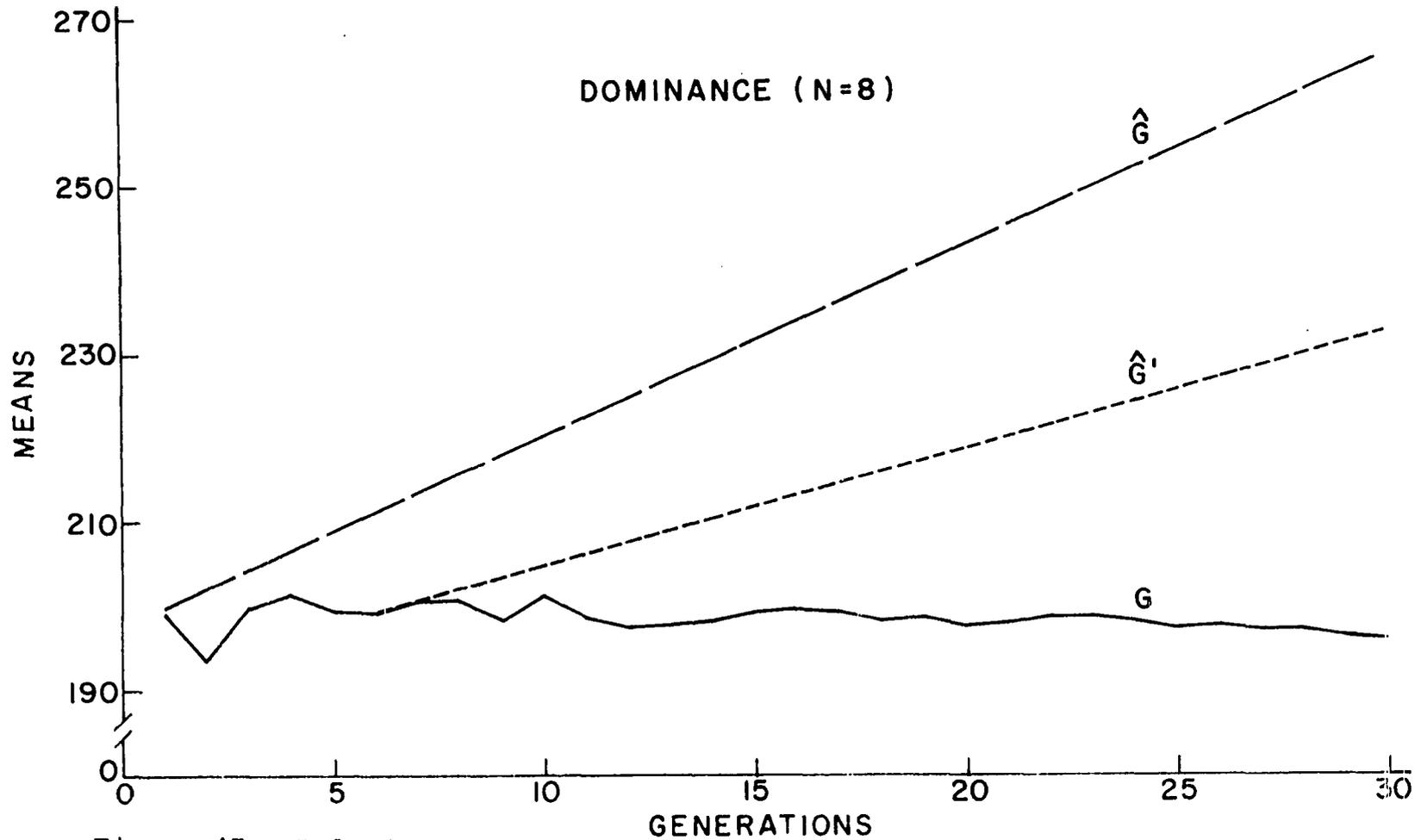


Figure 47. Relation of the mean (G) of a small population to that predicted from parameters of simulation (\hat{G}) and from parameters observed in generation 6 (\hat{G}') (complete dominance)

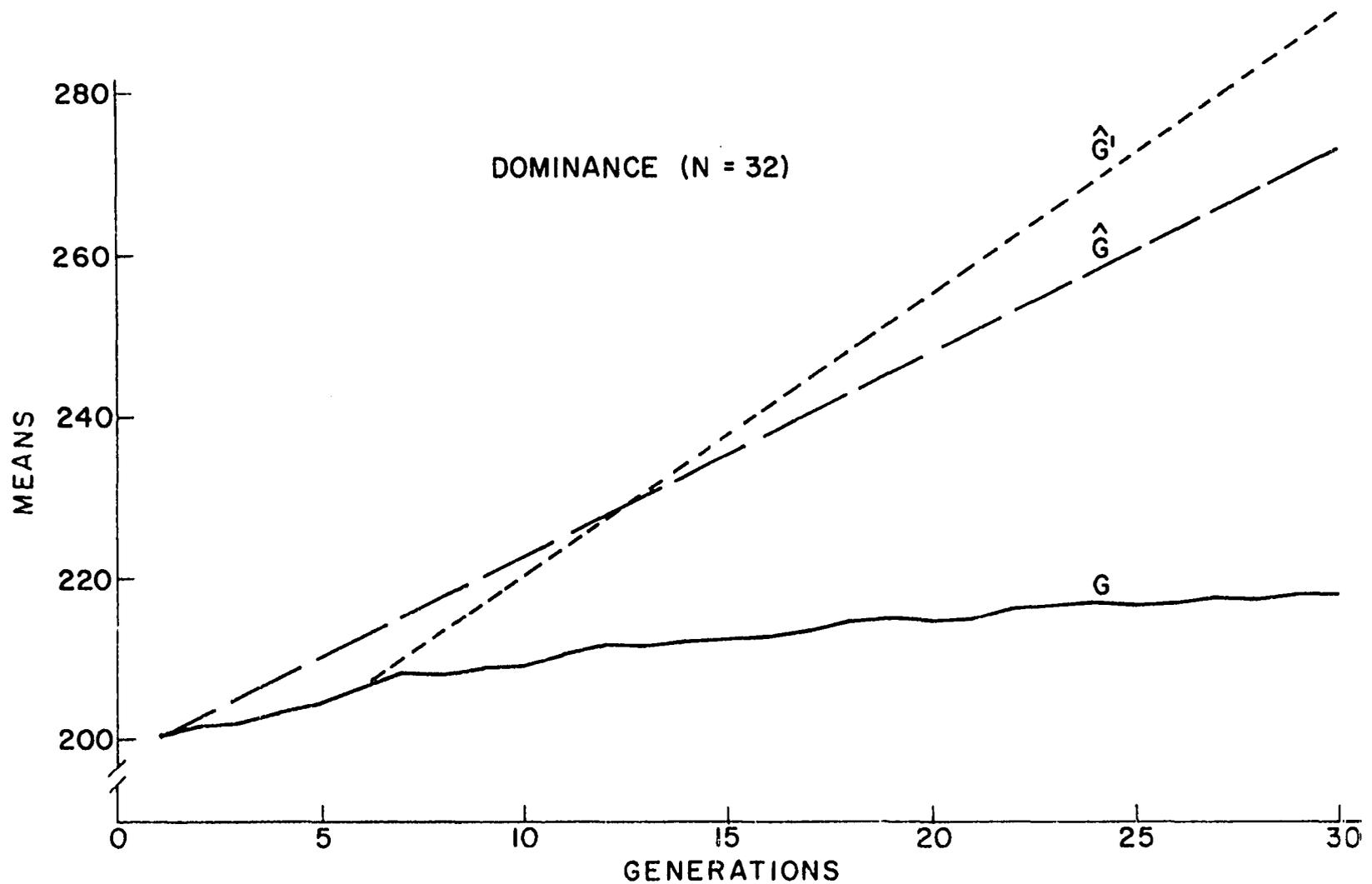


Figure 48. Relation of the mean (G) of a "large" population to that predicted from parameters of simulation (\hat{G}) and from parameters observed in generation 6 (\hat{G}') (complete dominance)

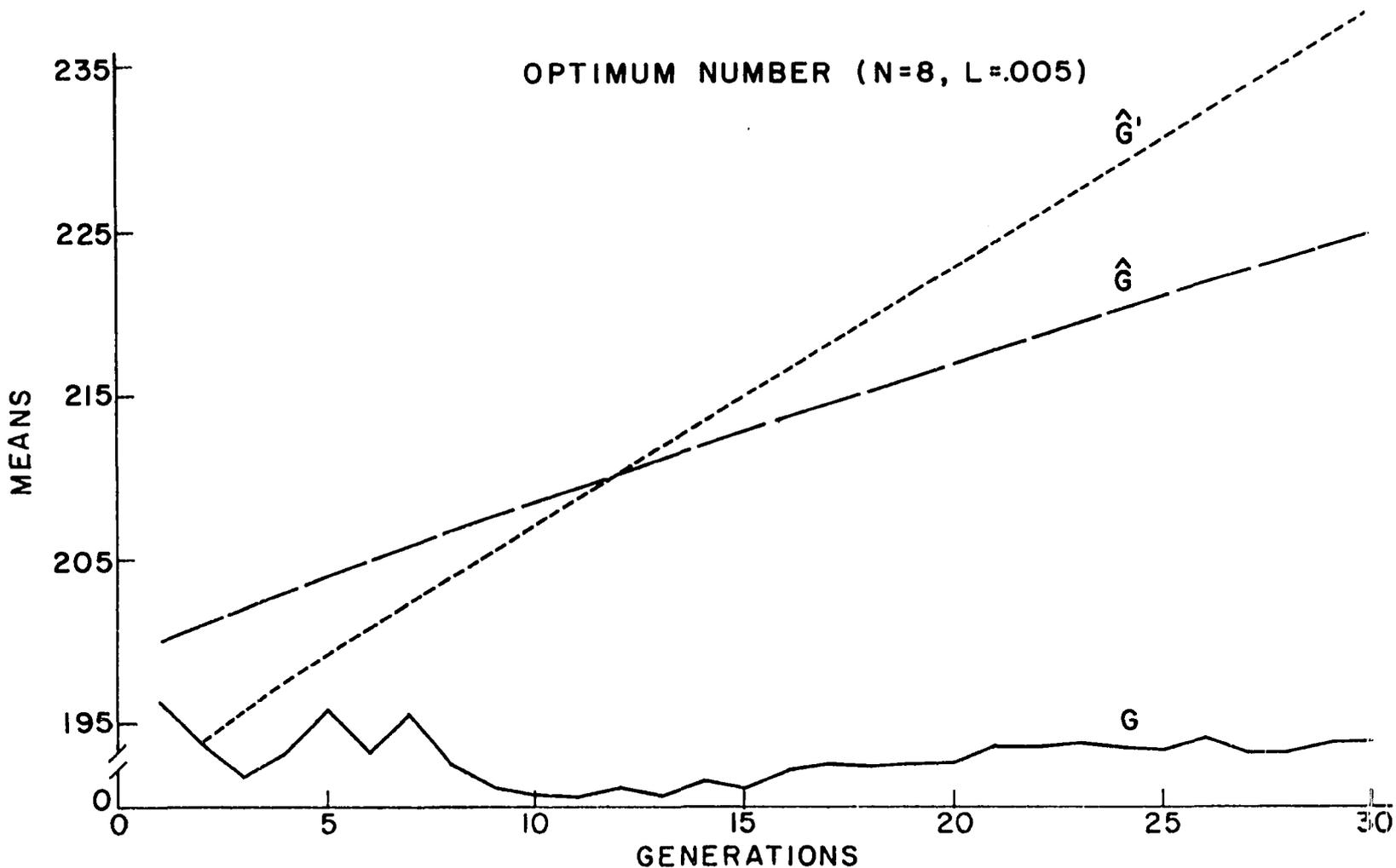


Figure 49. Relation of the mean (G) of a small population to that predicted from parameters of simulation (\hat{G}) and from parameters observed in generation 2 (\hat{G}') (optimum number model, tight linkage)

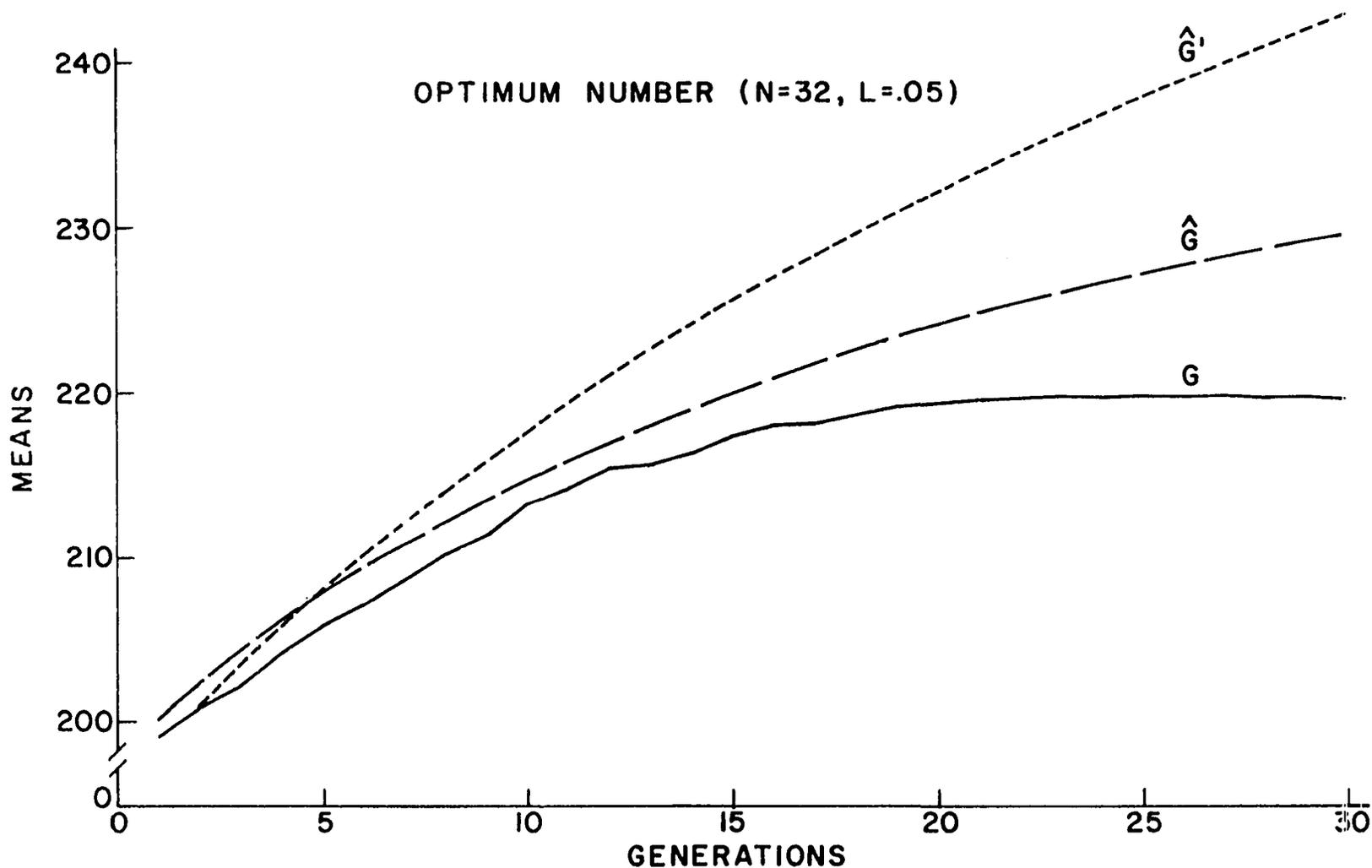


Figure 50. Relation of the mean (G) of a "large" population to that predicted from parameters of simulation (\hat{G}) and from parameters observed in generation 2 (\hat{G}') (optimum number model, tight linkage)

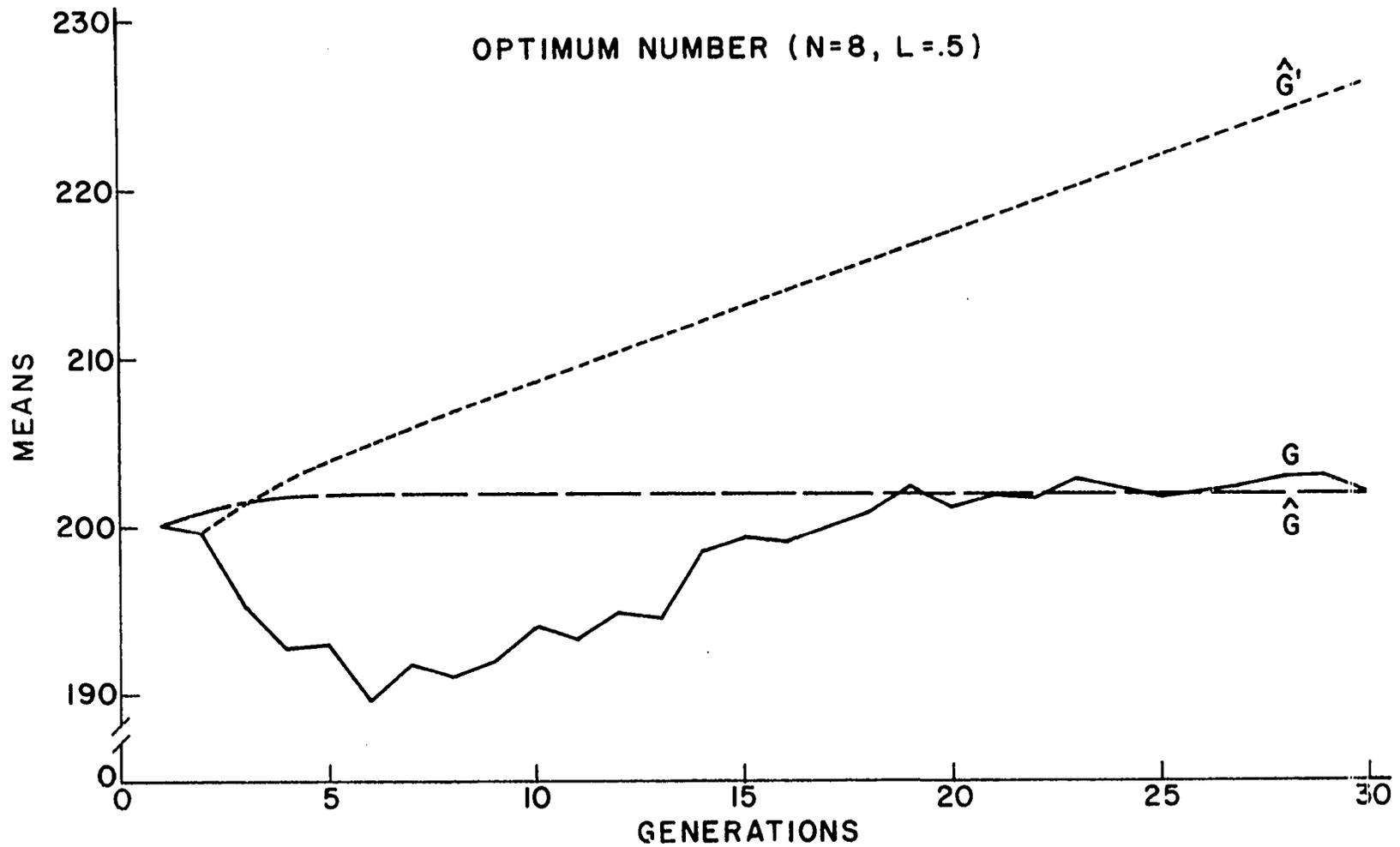


Figure 51. Relation of the mean (G) of a small population to that predicted from parameters of simulation (\hat{G}) and from parameters observed in generation 2 (\hat{G}') (optimum number model, no linkage)

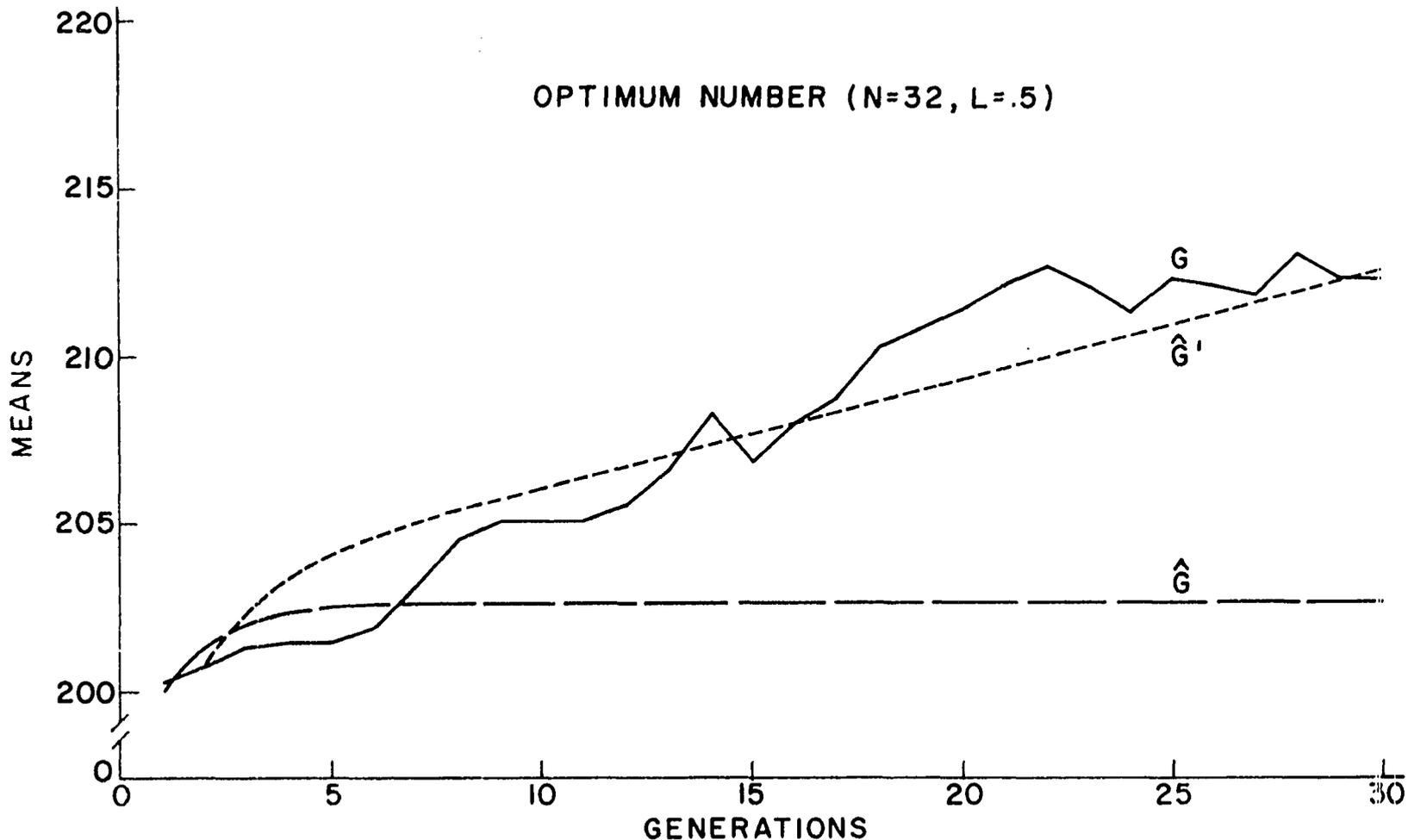


Figure 52. Relation of the mean (G) of a "large" population to that predicted from parameters of simulation (\hat{G}) and from parameters observed in generation 2 (\hat{G}') (optimum number model, no linkage)

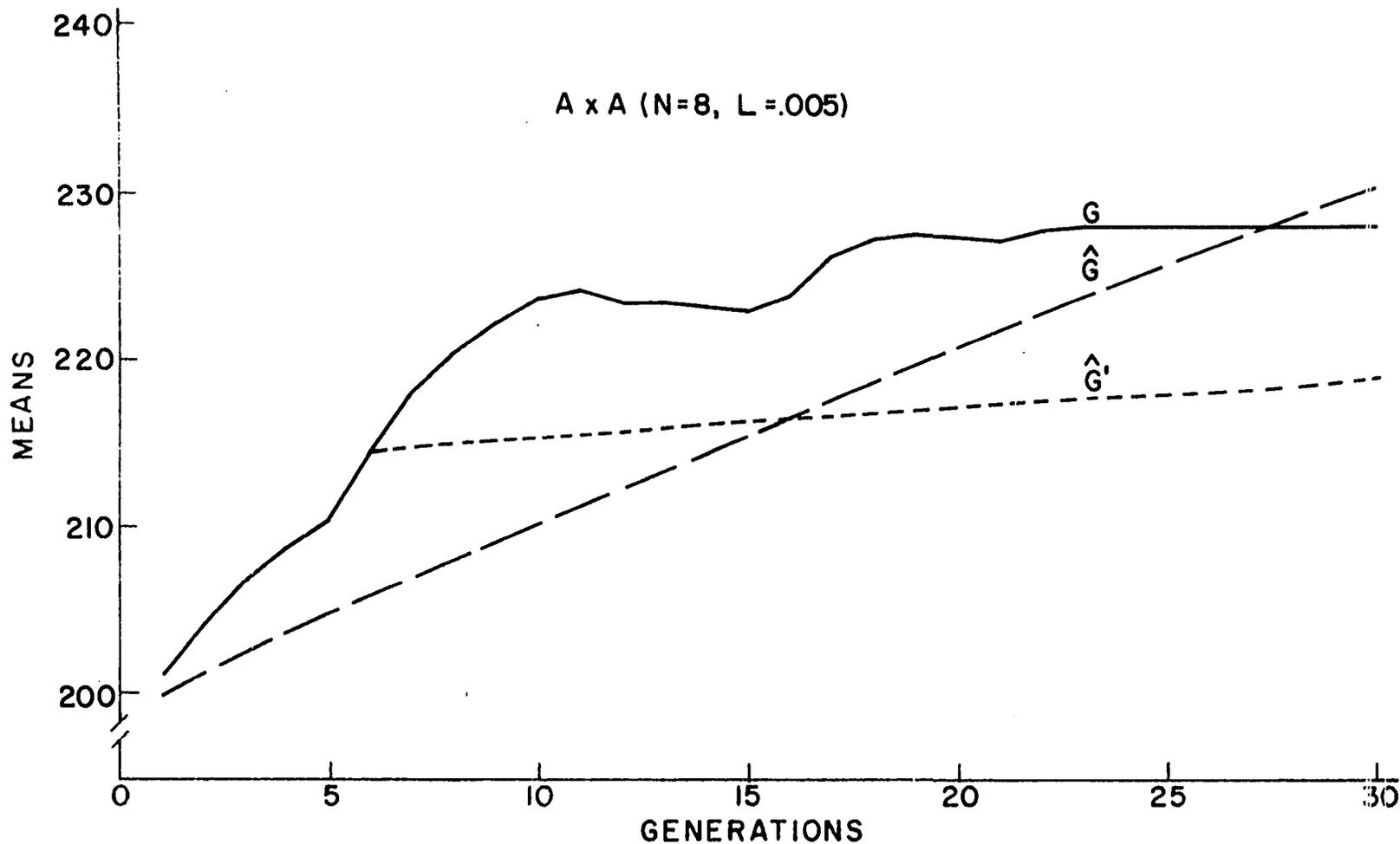


Figure 53. Relation of the mean (G) of a small population to that predicted from parameters of simulation (\hat{G}) and from parameters observed in generation 6 (\hat{G}') (additive-by-additive model, tight linkage)

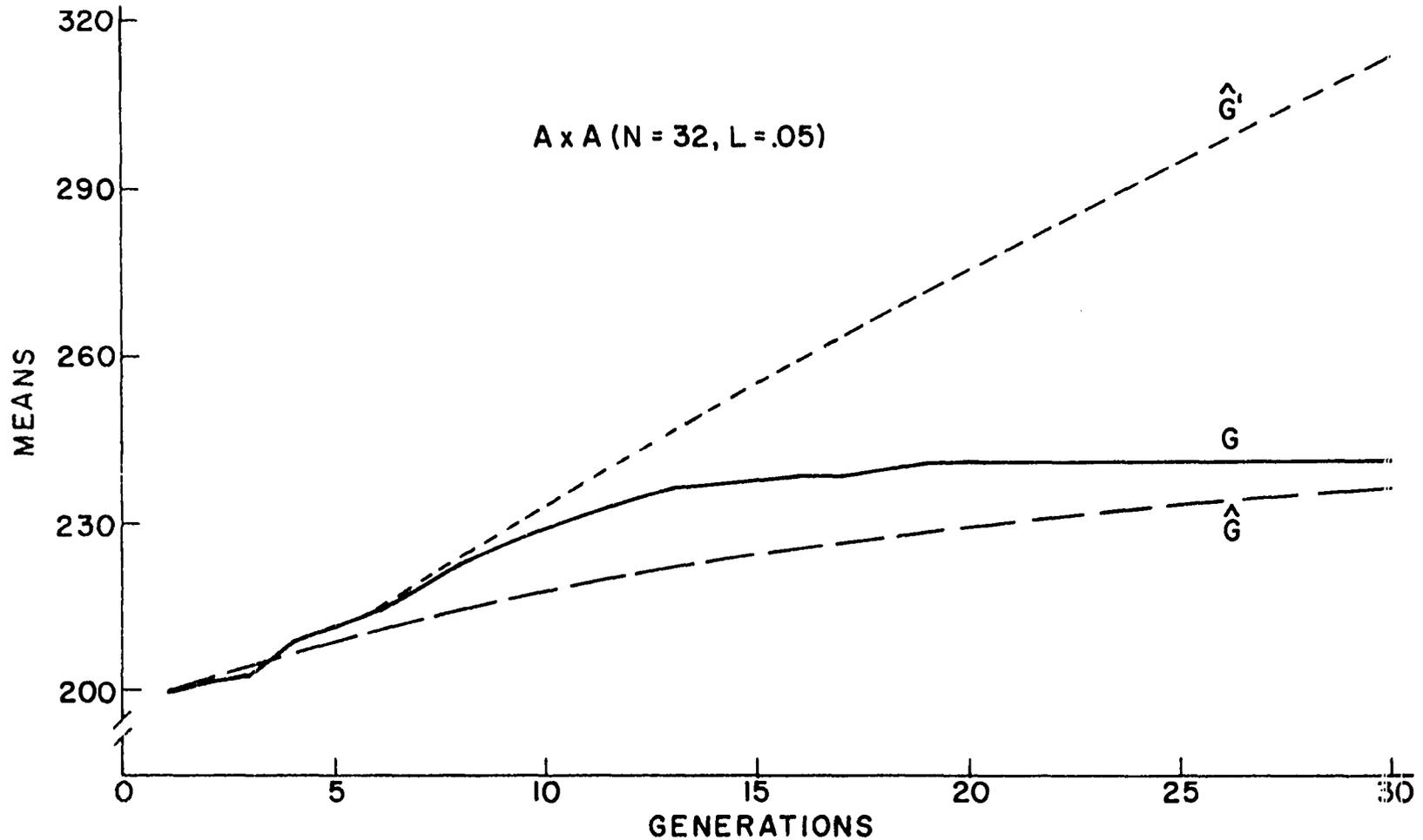


Figure 54. Relation of the mean (G) of a "large" population to that predicted from parameters of simulation (\hat{G}) and from parameters observed in generation 6 (\hat{G}') (additive-by-additive model, tight linkage)

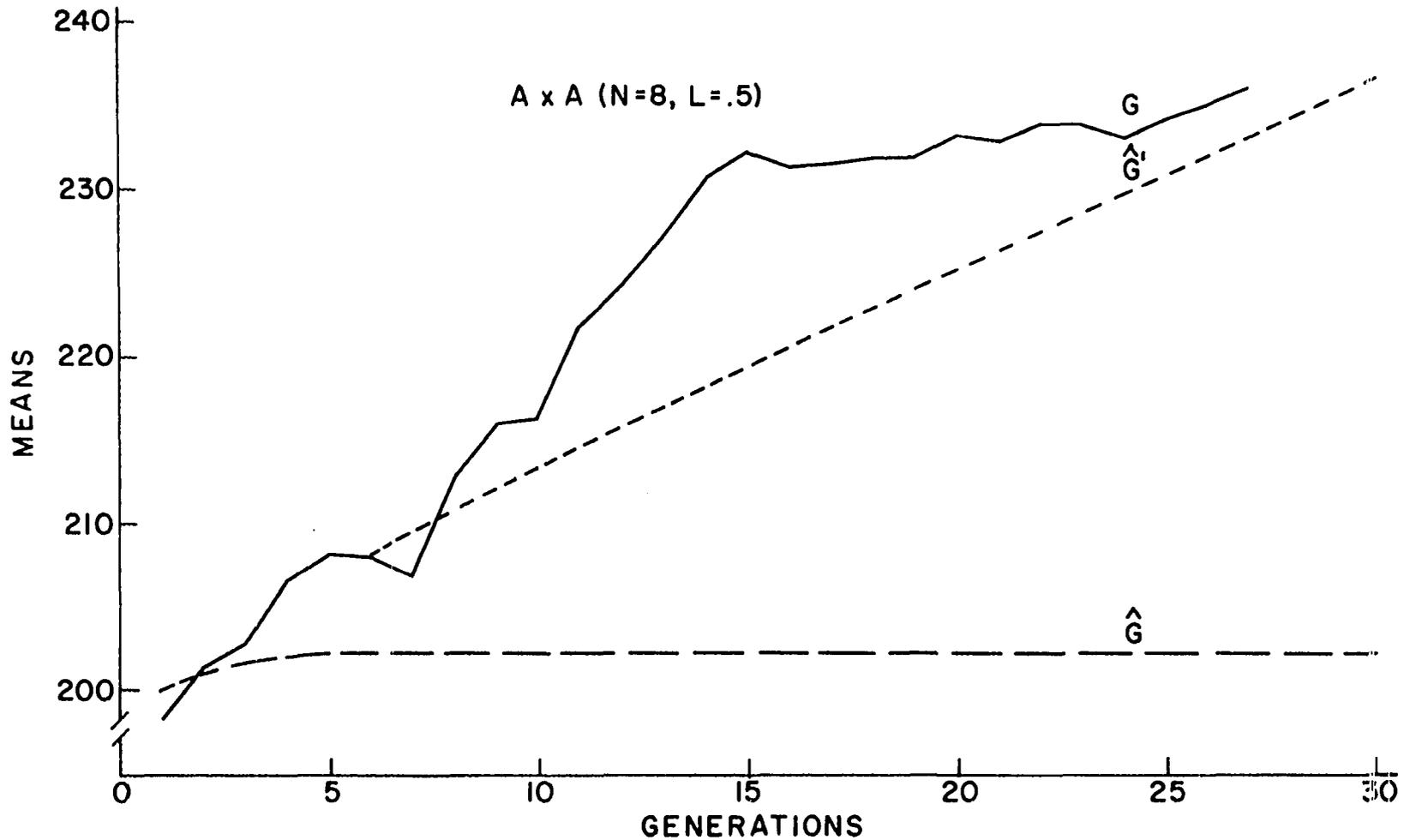


Figure 55. Relation of the mean (G) of a small population to that predicted from parameters of simulation (\hat{G}) and from parameters observed in generation 6 (\hat{G}') (additive-by-additive model, no linkage)

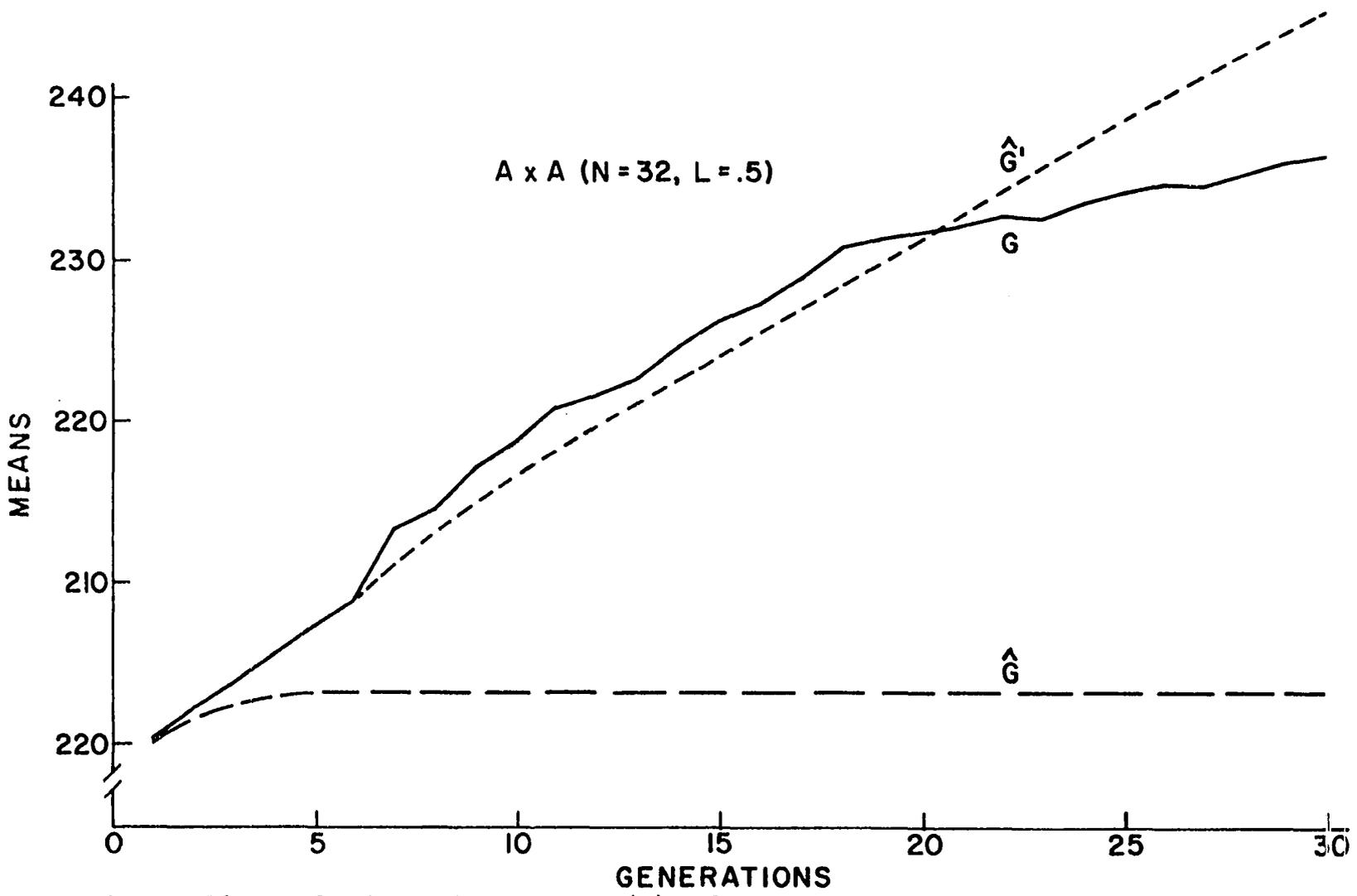


Figure 56. Relation of the mean (G) of a "large" population to that predicted from parameters of simulation (\hat{G}) and from parameters observed in generation 6 (\hat{G}') (additive-by-additive model, no linkage)

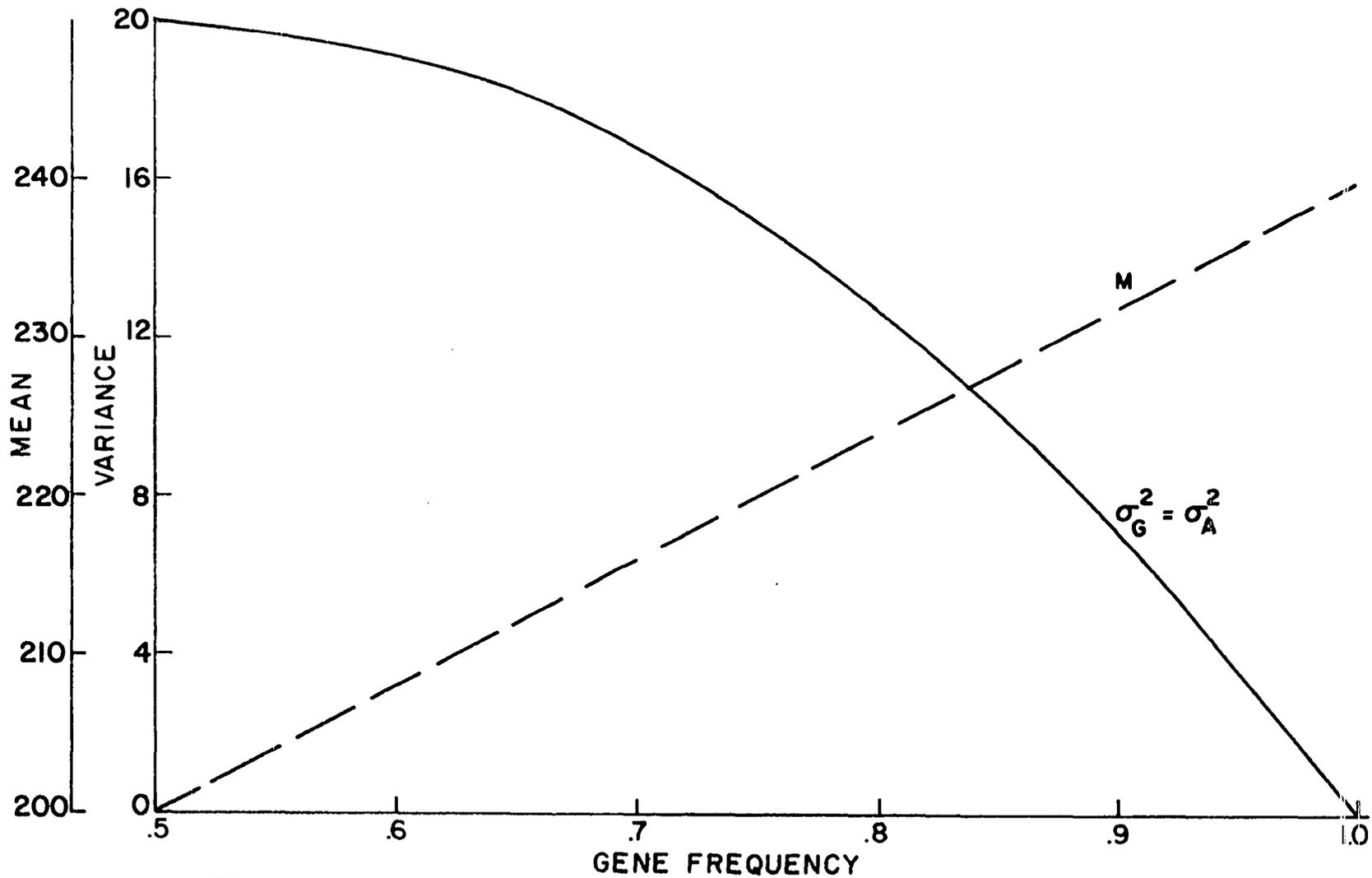


Figure 57. Changes in the mean (M) and the genotypic variance when the change in gene frequency is the same at all loci (additive model)

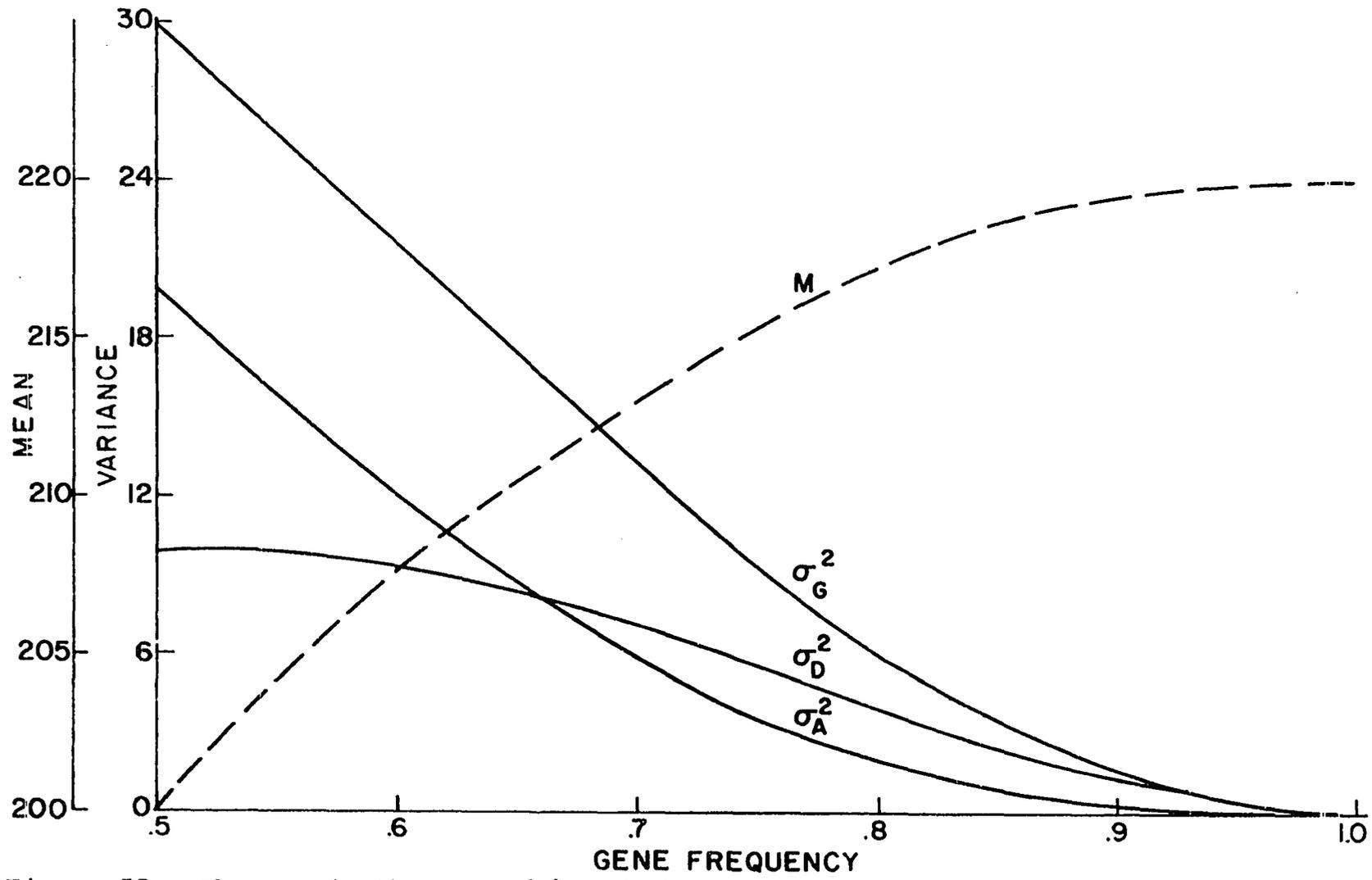


Figure 58. Changes in the mean (M) and the components of genotypic variance when the change in gene frequency is the same at all loci (complete dominance)

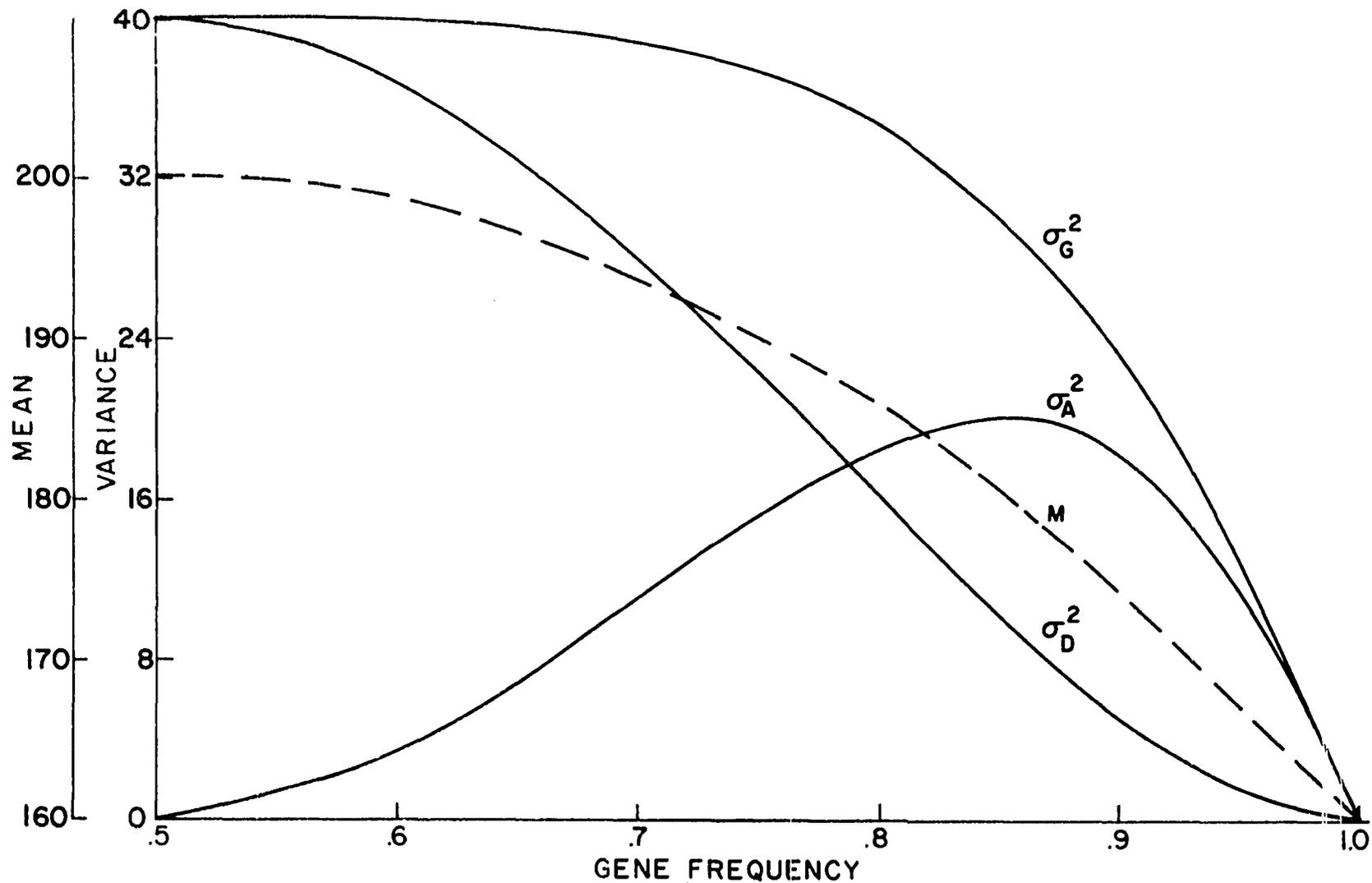


Figure 59. Changes in the mean (M) and the components of genotypic variance when change in gene frequency is the same at all loci (overdominance)

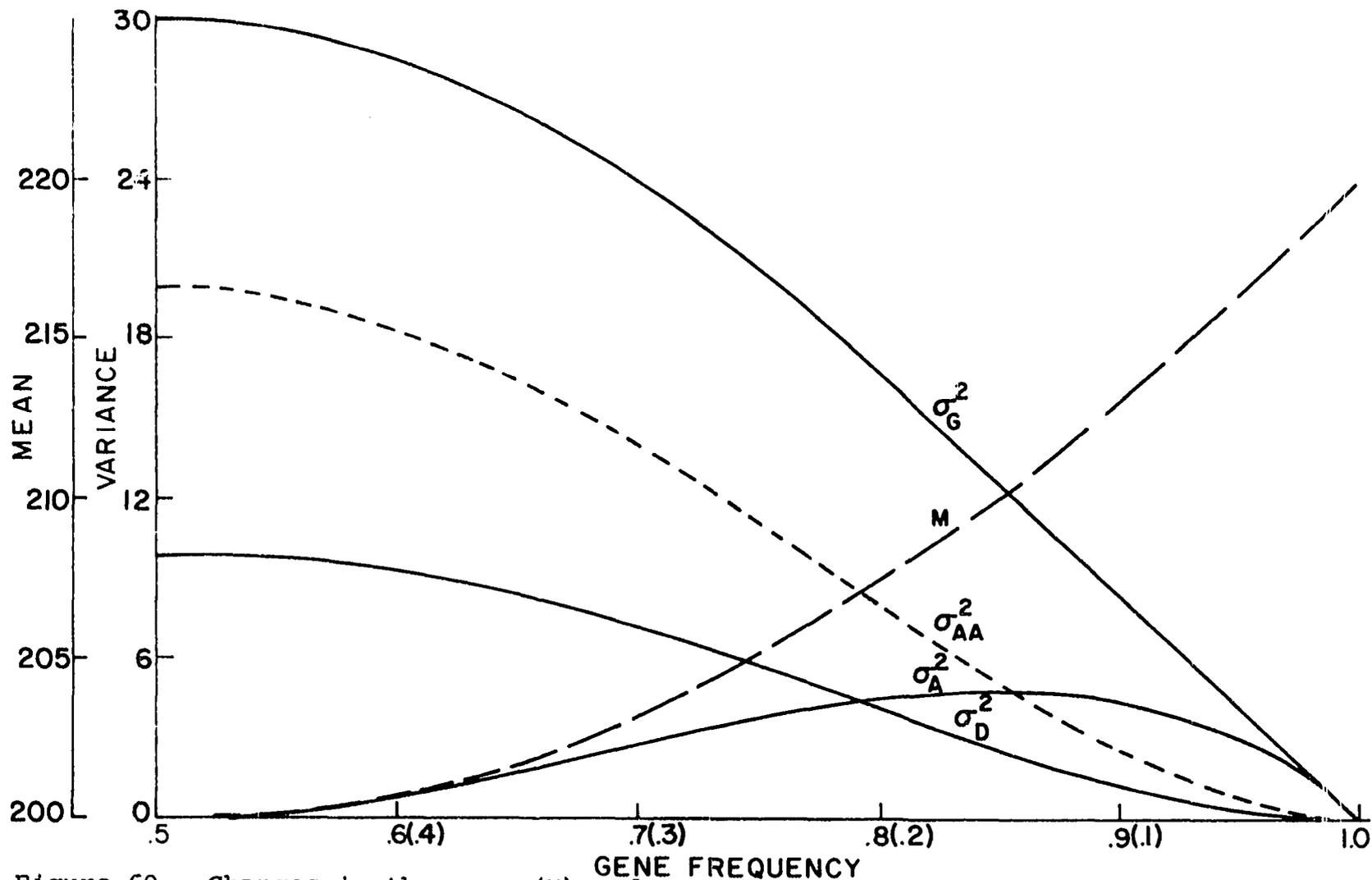


Figure 60. Changes in the mean (M) and variance when changes in the frequencies of two interacting genes are of equal magnitude but opposite in direction (optimum number model)

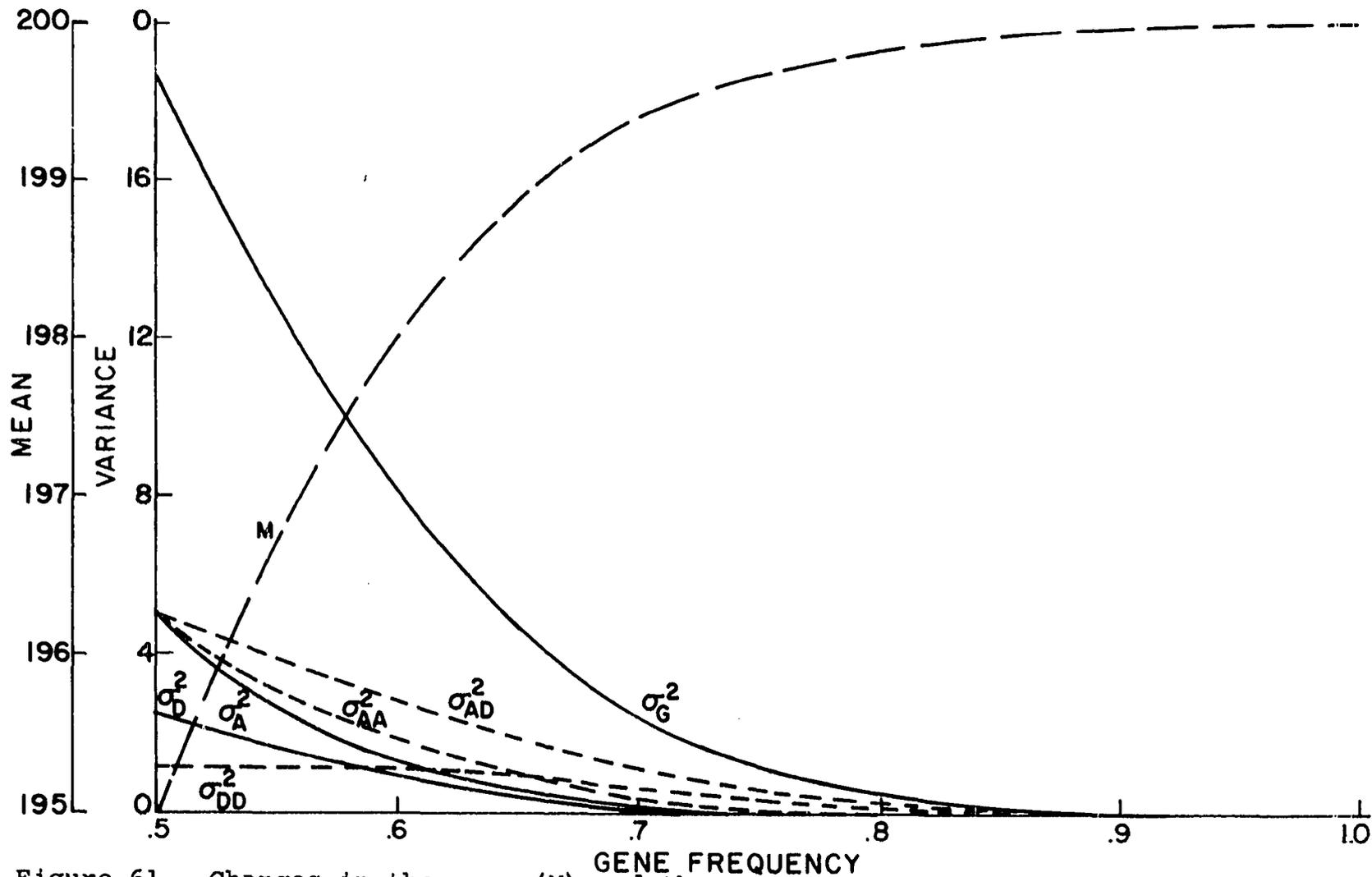


Figure 61. Changes in the mean (M) and the components of genotypic variance when the change in gene frequency is the same at all loci (duplicate factor model)

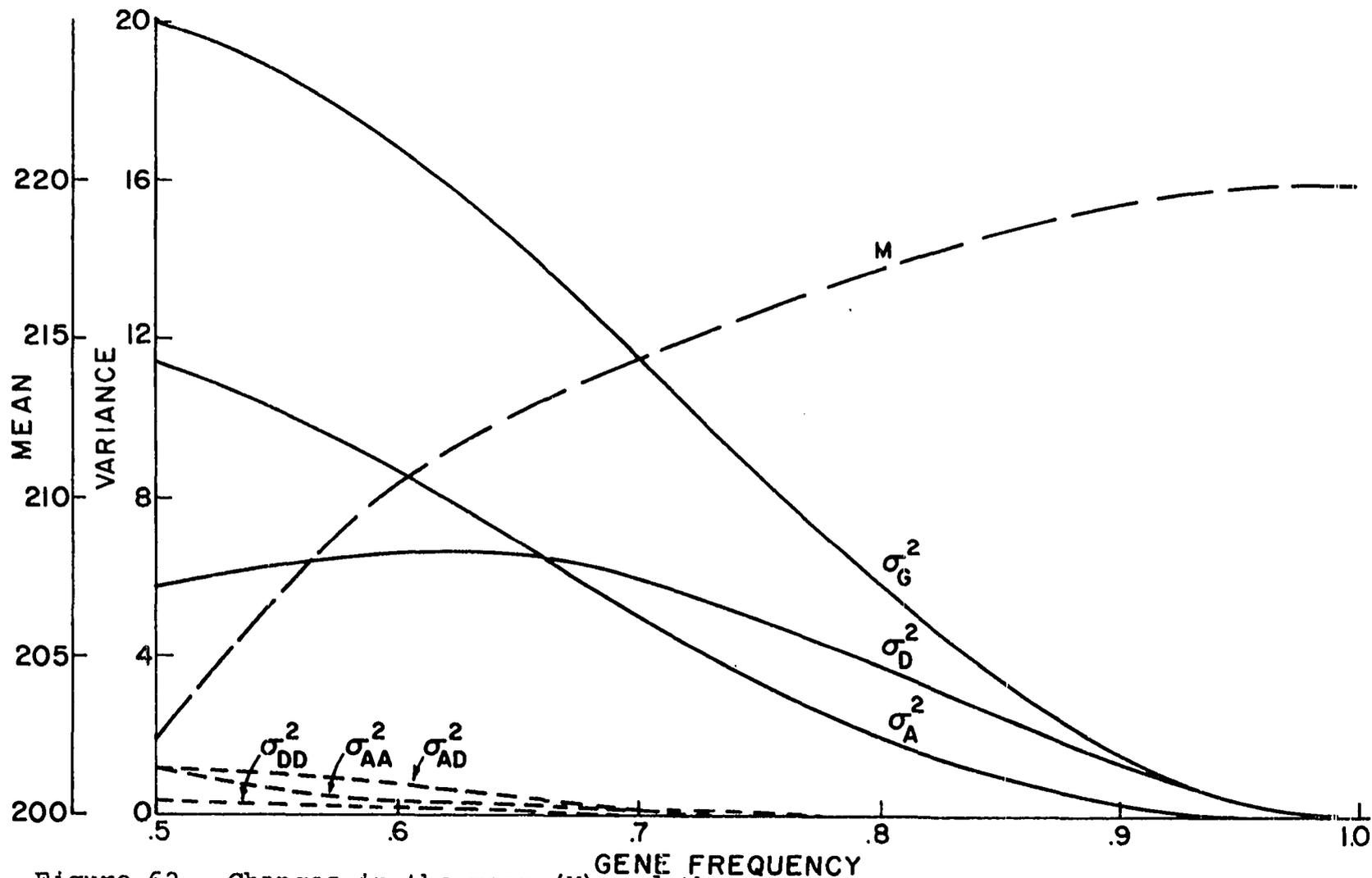


Figure 62. Changes in the mean (M) and the components of genotypic variance when the change in gene frequency is the same at all loci (complementary factor model)

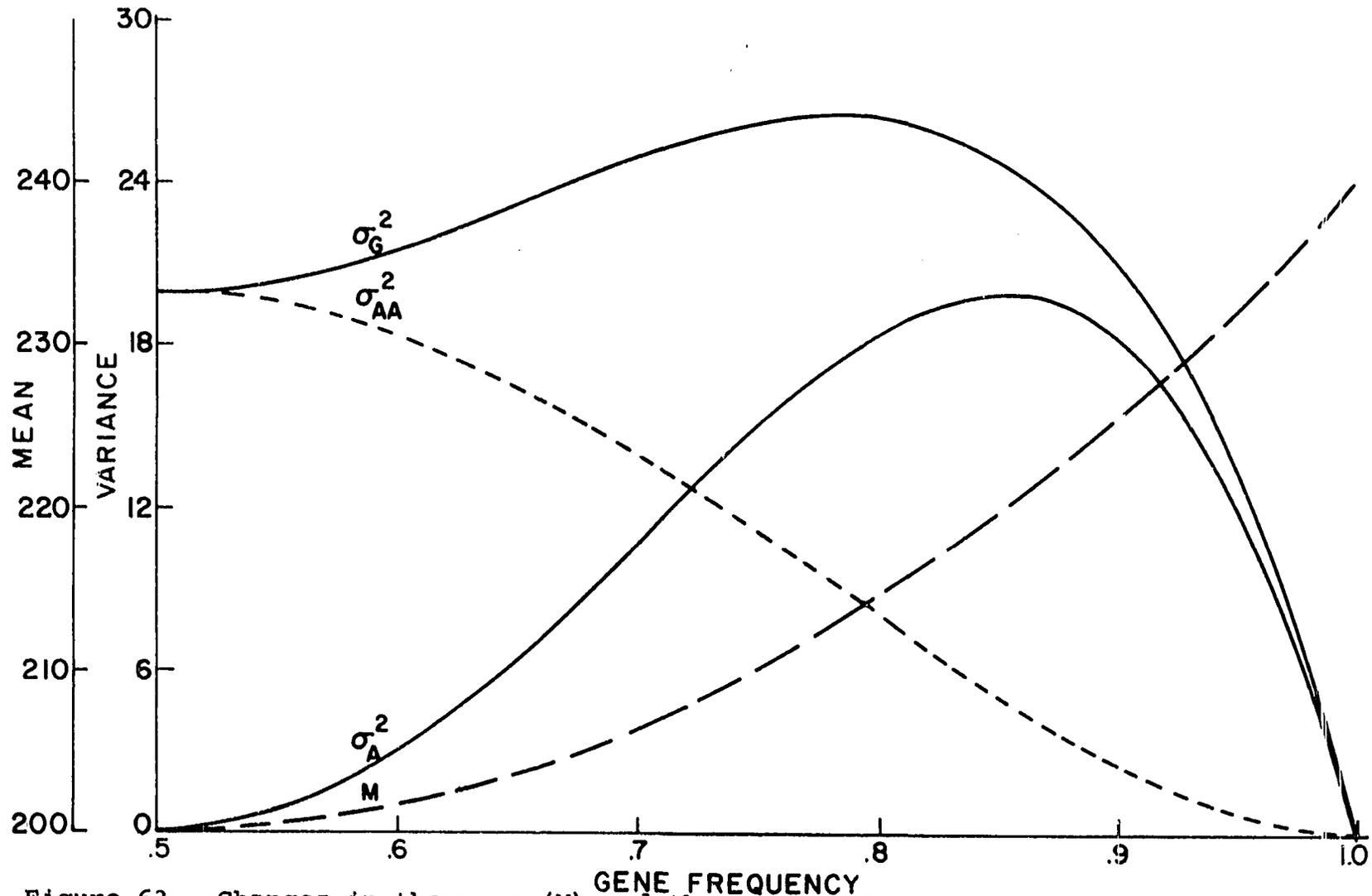


Figure 63. Changes in the mean (M) and the components of genotypic variance when the change in gene frequency is the same at all loci (additive-by-additive model)

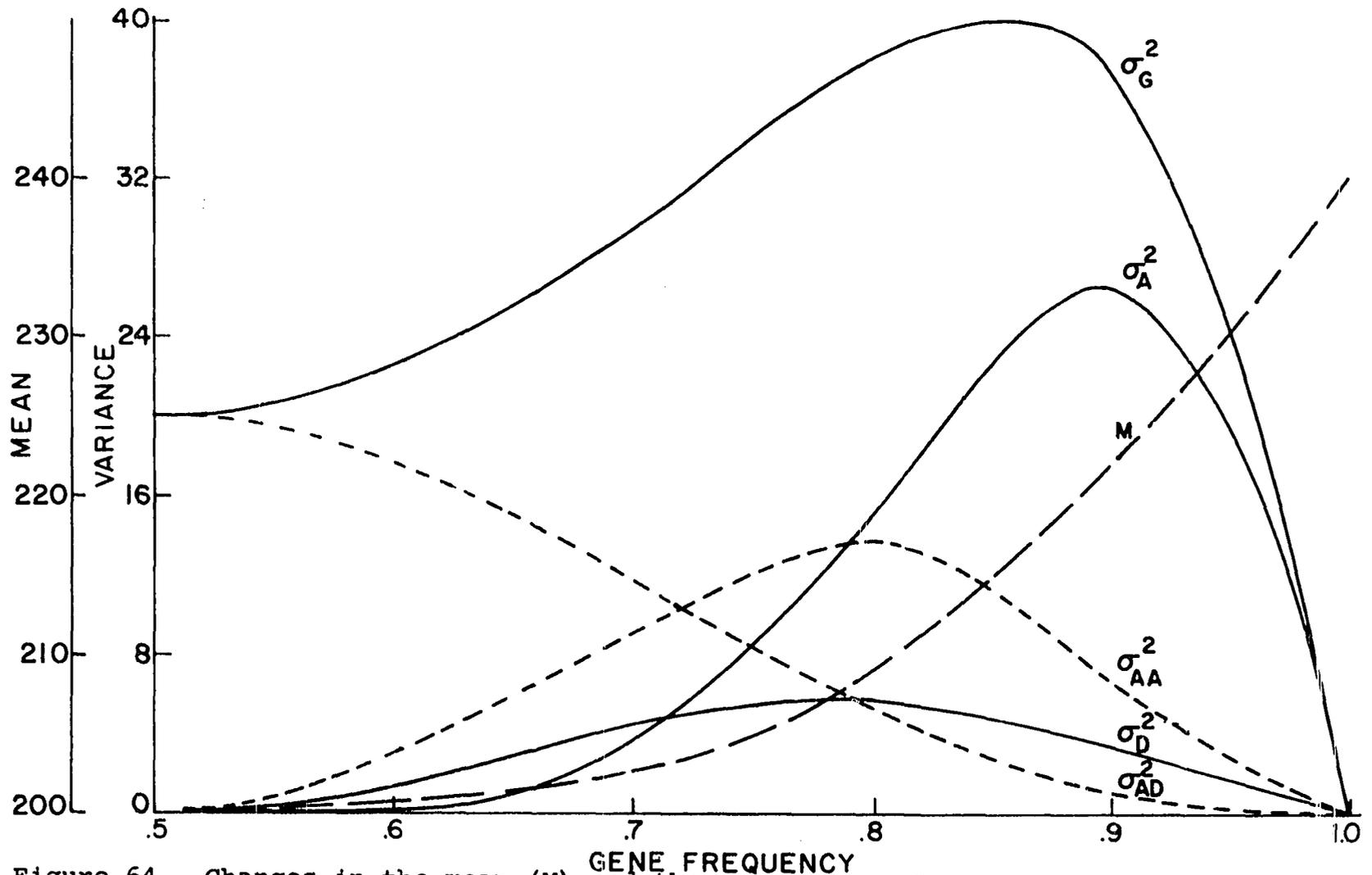


Figure 64. Changes in the mean (M) and the components of genotypic variance when the change in gene frequency is the same at all loci (additive-by-dominance model)

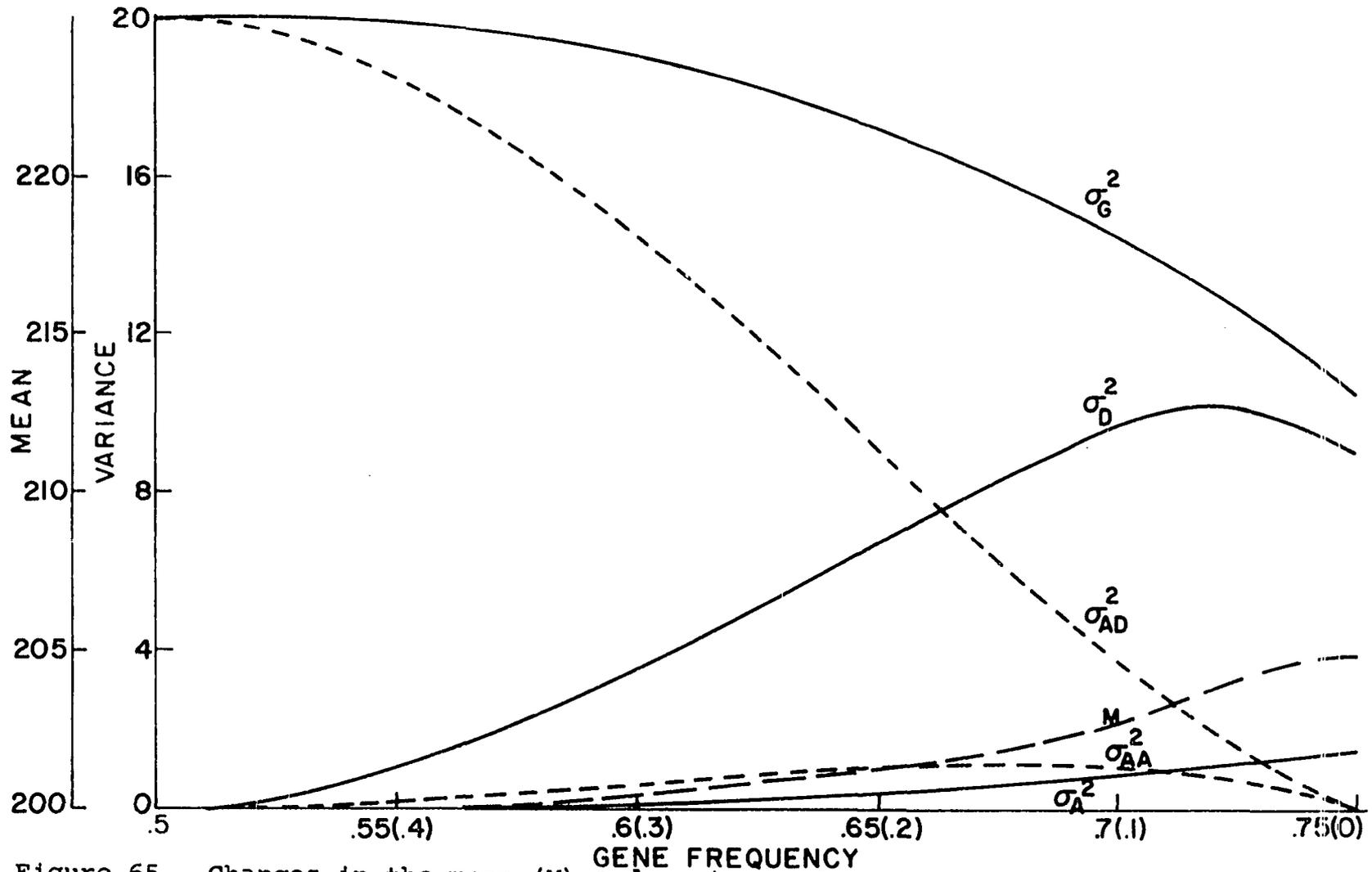


Figure 65. Changes in the mean (M) and variance when the change in one gene frequency is half that of an interacting gene but opposite in direction (additive-by-dominance model)

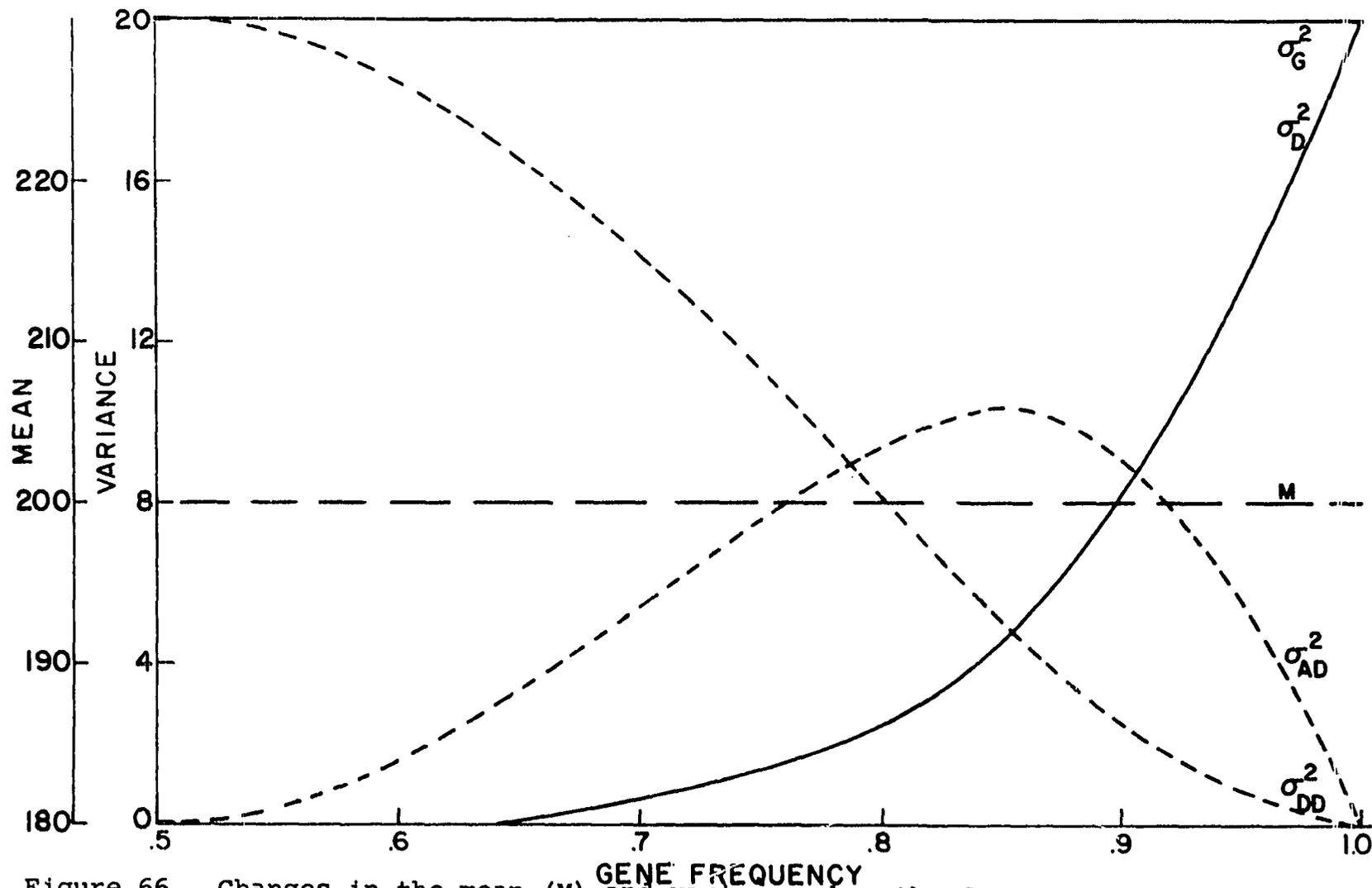


Figure 66. Changes in the mean (M) and variance when the frequency of one gene changes and that of an interacting gene remains constant at .5 (dominance-by-dominance model)

APPENDIX B: TABLES

Table 1. Mean proportion of loci fixed and mean gene frequency of unfixed loci classified by population size (complete dominance)

Generation number	Frequency			Frequency			
	Fraction fixed A	Fraction fixed a	of unfixed A's	Fraction fixed A	Fraction fixed a	of unfixed A's	
		N = 8			N = 12		
0	0	0	.500	0	0	.500	
5	.181	.050	.582	.106	0	.615	
10	.475	.112	.644	.356	.058	.619	
15	.600	.138	.661	.512	.067	.660	
20	.675	.162	.645	.638	.067	.656	
25	.750	.175	.613	.700	.083	.657	
		N = 16			N = 32		
0	0	0	.500	0	0	.500	
5	.050	0	.636	0	0	.645	
10	.225	.050	.688	.075	0	.738	
15	.394	.062	.699	.212	0	.779	
20	.544	.100	.705	.419	0	.784	
25	.638	.100	.619	.550	0	.814	

Table 2. Mean proportion of loci fixed and mean gene frequency of unfixed loci classified by population size (overdominance)

Genera- tion number	Fraction fixed		Frequency of unfixed	Fraction fixed		Frequency of unfixed
	A	a	A's	A	a	A's
	N = 8			N = 12		
0	0	0	.500	0	0	.500
5	.067	.050	.494	0	.050	.506
10	.150	.131	.547	.069	.074	.480
15	.269	.219	.499	.119	.138	.498
20	.325	.288	.491	.169	.175	.473
25	.356	.312	.515	.206	.225	.504
	N = 16			N = 32		
0	0	0	.500	0	0	.500
5	.050	0	.503	0	0	.494
10	.088	.033	.508	.025	.025	.511
15	.108	.056	.517	.025	.025	.504
20	.150	.088	.493	.025	.050	.493
25	.156	.125	.490	.025	.050	.486

Table 3. Mean proportion of loci fixed and mean gene frequency of unfixed loci classified by population size (optimum number of genes)

Genera- tion number	Fraction fixed		Frequency of unfixed		Frequency of unfixed	
	A	a	A's	A	a	A's
	N = 8			N = 12		
0	0	0	.500	0	0	.500
5	.094	.133	.499	.044	.113	.509
10	.306	.331	.503	.156	.181	.532
15	.412	.381	.542	.294	.319	.544
20	.475	.456	.393	.356	.375	.533
25	.481	.469	.563	.406	.412	.520
	N = 16			N = 32		
0	0	0	.500	0	0	.500
5	.025	.025	.488	0	0	.502
10	.106	.112	.495	.025	.025	.511
15	.256	.269	.480	.094	.094	.493
20	.300	.388	.521	.219	.194	.479
25	.440	.431	.532	.288	.312	.520

Table 4. Mean proportion of loci fixed and mean gene frequency of unfixed loci classified by population size (duplicate factors)

Generation number	Fraction fixed A	Fraction fixed a	Frequency of unfixed A's	Fraction fixed A	Fraction fixed a	Frequency of unfixed A's
	N = 8			N = 12		
0	0	0	.500	0	0	.500
5	.150	.112	.574	.088	.050	.567
10	.425	.244	.601	.300	.144	.570
15	.631	.275	.512	.625	.181	.544
20	.650	.275	.485	.688	.181	.493
25	.681	.275	.554	.731	.219	.562
	N = 16			N = 32		
0	0	0	.500	0	0	.500
5	.038	.025	.570	0	0	.575
10	.188	.067	.555	.058	.025	.590
15	.312	.119	.572	.131	.038	.571
20	.394	.181	.583	.212	.044	.556
25	.538	.194	.566	.319	.094	.545

Table 5. Mean proportion of loci fixed and mean gene frequency of unfixed loci classified by population size (complementary factors)

Genera- tion number	Fraction fixed		Frequency of unfixed	Fraction fixed		Frequency of unfixed
	A	a	A's	A	a	A's
	N = 8			N = 12		
0	0	0	.500	0	0	.500
5	.131	.067	.617	.062	.050	.605
10	.500	.133	.584	.288	.062	.635
15	.656	.138	.644	.475	.067	.621
20	.738	.144	.646	.588	.081	.659
25	.769	.175	.615	.656	.088	.703
	N = 16			N = 32		
0	0	0	.500	0	0	.500
5	.050	0	.615	0	0	.620
10	.225	.025	.654	.100	0	.718
15	.306	.067	.713	.188	0	.741
20	.462	.117	.752	.312	0	.779
25	.581	.125	.751	.444	0	.793

Table 6. Mean proportion of loci fixed and mean gene frequency of unfixed loci classified by population size (AxA gene action)

Generation number	Fraction fixed		Frequency of unfixed	Fraction fixed		Frequency of unfixed
	A	a	A's	A	a	A's
	N = 8			N = 12		
0	0	0	.500	0	0	.500
5	.212	.262	.471	.075	.083	.508
10	.394	.469	.622	.350	.375	.431
15	.438	.506	.349	.450	.481	.688
20	.456	.513	.531	.494	.481	.903
25	.475	.513	.625	.519	.481	--
	N = 16			N = 32		
0	0	0	.500	0	0	.500
5	.069	.042	.554	.025	.025	.476
10	.431	.188	.533	.175	.212	.417
15	.600	.262	.771	.281	.375	.416
20	.656	.269	.497	.369	.475	.474
25	.681	.281	.417	.381	.512	.668

Table 7. Mean proportion of loci fixed and mean gene frequency of unfixed loci classified by population size (AxD gene action)

Generation number	Fraction fixed		Frequency of unfixed	Fraction fixed		Frequency of unfixed
	A	a	A's	A	a	A's
	N = 8			N = 12		
0	0	0	.500	0	0	.500
5	.244	.125	.530	.138	.044	.548
10	.425	.231	.551	.419	.144	.484
15	.519	.294	.611	.562	.225	.543
20	.544	.300	.679	.600	.250	.623
25	.569	.300	.600	.612	.275	.695
	N = 16			N = 32		
0	0	0	.500	0	0	.500
5	.233	.108	.509	.025	.025	.517
10	.338	.162	.486	.100	.050	.520
15	.438	.238	.619	.206	.088	.545
20	.519	.269	.630	.338	.138	.564
25	.550	.281	.613	.400	.167	.529

Table 8. Mean proportion of loci fixed and mean gene frequency of unfixed loci classified by population size (DxD gene action)

Generation number	N = 8			N = 12		
	Fraction fixed A	fixed a	Frequency of unfixed A's	Fraction fixed A	fixed a	Frequency of unfixed A's
0	0	0	.500	0	0	.500
5	.150	.081	.516	.050	.050	.478
10	.250	.169	.507	.119	.106	.483
15	.319	.250	.519	.194	.194	.522
20	.350	.300	.501	.250	.212	.491
25	.388	.338	.531	.4	.244	.485
	N = 16			N = 32		
0	0	0	.500	0	0	.500
5	.025	.075	.475	0	0	.506
10	.094	.144	.495	.033	.031	.506
15	.162	.194	.474	.042	.088	.504
20	.200	.250	.477	.056	.112	.498
25	.219	.288	.507	.088	.144	.520

Table 9. Mean proportion of loci fixed and mean gene frequency of unfixed loci classified by selection intensity (complete dominance)

Generation number	Fraction fixed		Frequency of unfixed A's	Fraction fixed		Frequency of unfixed A's
	A	a		A	a	
	$s = 1/2$			$s = 1/4$		
0	0	0	.500	0	0	.500
5	.172	.050	.586	.133	.050	.629
10	.325	.133	.633	.238	.075	.666
15	.412	.158	.659	.419	.100	.709
20	.531	.192	.632	.550	.117	.711
25	.600	.192	.653	.662	.133	.731
	$s = 1/6$			$s = 1/8$		
0	0	0	.500	0	0	.500
5	.067	0	.619	.150	0	.645
10	.238	.038	.695	.331	.025	.696
15	.388	.038	.724	.500	.025	.707
20	.512	.038	.711	.681	.050	.735
25	.619	.062	.690	.756	.050	.632

Table 10. Mean proportion of loci fixed and mean gene frequency of unfixed loci classified by selection intensity (overdominance)

Generation number	Frequency of unfixed A's			Frequency of unfixed A's		
	Fraction fixed A	Fraction fixed a	Fraction fixed A's	Fraction fixed A	Fraction fixed a	Fraction fixed A's
	$s = 1/2$			$s = 1/4$		
0	0	0	.500	0	0	.500
5	.075	.050	.493	.025	.025	.530
10	.125	.105	.504	.058	.038	.522
15	.175	.150	.505	.142	.125	.524
20	.194	.175	.476	.225	.208	.513
25	.231	.200	.481	.250	.252	.522
	$s = 1/6$			$s = 1/8$		
0	0	0	.500	0	0	.500
5	.025	.062	.497	0	.050	.476
10	.138	.083	.502	.062	.067	.518
15	.225	.100	.489	.108	.133	.501
20	.275	.133	.491	.150	.175	.470
25	.275	.156	.497	.183	.200	.495

Table 11. Mean proportion of loci fixed and mean gene frequency of unfixed loci classified by selection intensity (optimum number of genes)

Genera- tion number	Frequency			Frequency		
	Fraction fixed A	Fraction fixed a	of unfixed A's	Fraction fixed A	Fraction fixed a	of unfixed A's
	$s = 1/2$			$s = 1/4$		
0	0	0	.500	0	0	.500
5	.075	.067	.503	.050	.025	.497
10	.138	.233	.491	.175	.112	.516
15	.219	.306	.540	.281	.256	.495
20	.306	.375	.574	.325	.331	.491
25	.381	.406	.545	.425	.431	.535
	$s = 1/6$			$s = 1/8$		
0	0	0	.500	0	0	.500
5	.042	.062	.499	.075	.100	.499
10	.175	.194	.502	.138	.208	.533
15	.281	.250	.498	.275	.300	.515
20	.356	.325	.480	.362	.381	.403
25	.384	.369	.499	.425	.419	.539

Table 12. Mean proportion of loci fixed and mean gene frequency of unfixed loci classified by selection intensity (duplicate factors)

Generation number	Frequency of unfixed A's			Frequency of unfixed A's		
	Fraction fixed A	Fraction fixed a	Fraction fixed A's	Fraction fixed A	Fraction fixed a	Fraction fixed A's
	$s = 1/2$			$s = 1/4$		
0	0	0	.500	0	0	.500
5	.075	.050	.562	.083	.044	.566
10	.256	.119	.585	.244	.106	.561
15	.488	.150	.564	.350	.140	.550
20	.525	.181	.587	.431	.181	.502
25	.588	.200	.562	.469	.206	.541
	$s = 1/6$			$s = 1/8$		
0	0	0	.500	0	0	.500
5	.075	.050	.566	.133	.075	.591
10	.292	.131	.566	.238	.125	.604
15	.425	.162	.534	.438	.138	.587
20	.469	.175	.539	.519	.175	.570
25	.538	.188	.581	.675	.194	.545

Table 13. Mean proportion of loci fixed and mean gene frequency of unfixed loci classified by selection intensity (complementary factors)

Generation number	Frequency			Frequency		
	Fraction fixed A	fixed a	of unfixed A's	Fraction fixed A	fixed a	of unfixed A's
	$s = 1/2$			$s = 1/4$		
0	0	0	.500	0	0	.500
5	.033	.050	.568	.112	.025	.631
10	.150	.075	.584	.331	.044	.656
15	.269	.150	.627	.488	.050	.678
20	.425	.250	.678	.650	.069	.673
25	.488	.258	.613	.781	.081	.774
	$s = 1/6$			$s = 1/8$		
0	0	0	.500	0	0	.500
5	.125	.012	.601	.100	.012	.657
10	.317	.019	.669	.458	.019	.682
15	.381	.044	.718	.488	.031	.696
20	.444	.044	.723	.581	.038	.761
25	.519	.044	.699	.662	.038	.776

Table 14. Mean proportion of loci fixed and mean gene frequency of unfixed loci classified by selection intensity (AxA gene action)

Generation number	Fraction fixed		Frequency of unfixed	Fraction fixed		Frequency of unfixed
	A	a	A's	A	a	A's
	$s = 1/2$			$s = 1/4$		
0	0	0	.500	0	0	.500
5	.067	.175	.500	.088	.088	.492
10	.206	.200	.495	.325	.369	.466
15	.388	.312	.442	.394	.488	.730
20	.438	.381	.450	.488	.488	.792
25	.475	.425	.547	.512	.488	--
	$s = 1/6$			$s = 1/8$		
0	0	0	.500	0	0	.500
5	.142	.175	.481	.158	.133	.536
10	.325	.394	.535	.494	.281	.506
15	.412	.450	.534	.575	.375	.416
20	.444	.481	.682	.606	.388	.917
25	.456	.488	.736	.612	.388	--

Table 15. Mean proportion of loci fixed and mean gene frequency of unfixed loci classified by selection intensity (AxD gene action)

Generation number	Frequency of unfixed A's			Frequency of unfixed A's		
	Fraction fixed A	Fraction fixed a	Fraction fixed A	Fraction fixed a	Fraction fixed A	Fraction fixed a
	$S = 1/2$			$S = 1/4$		
0	0	0	.500	0	0	.500
5	.175	.062	.523	.181	.069	.520
10	.256	.125	.532	.306	.119	.529
15	.394	.206	.598	.450	.200	.583
20	.494	.231	.632	.488	.212	.671
25	.538	.269	.594	.550	.225	.623
	$S = 1/6$			$S = 1/8$		
0	0	0	.500	0	0	.500
5	.208	.069	.547	.183	.117	.513
10	.325	.175	.495	.300	.192	.486
15	.438	.206	.579	.444	.231	.557
20	.519	.238	.603	.500	.275	.591
25	.538	.250	.611	.506	.281	.608

Table 16. Mean proportion of loci fixed and mean gene frequency of unfixed loci classified by selection intensity (DxD gene action)

Genera- tion number	Frequency of unfixed A's		Frequency of unfixed A's			
	Fraction fixed A	Fraction fixed a	Fraction fixed A	Fraction fixed a		
	$s = 1/2$		$s = 1/4$			
0	0	0	0	0	.500	
5	.121	.075	.497	.050	.062	.490
10	.150	.125	.478	.081	.094	.517
15	.200	.206	.417	.144	.169	.516
20	.225	.250	.510	.194	.225	.514
25	.244	.288	.550	.231	.250	.491
	$s = 1/6$		$s = 1/8$			
0	0	0	.500	0	0	.500
5	.067	.112	.486	.100	.038	.502
10	.167	.144	.506	.131	.088	.491
15	.217	.206	.486	.200	.144	.500
20	.225	.225	.491	.238	.175	.452
25	.225	.262	.513	.288	.212	.489

Table 17. Relation of the mean in generation 30 (G_{30}) to that predicted from parameters of simulation (\hat{G}_{30}) and from parameters observed in generation one (\hat{G}'_{30}) (additive gene action)

Popula- tion size	Selec- tion in- tensity	Environ- mental variation	Linkage	G_{30}	\hat{G}_{30}	\hat{G}'_{30}
8	1/2	σ_G^2	.5	214.88	269.54	282.16
8	1/4	$3\sigma_G^2$.005	214.00	280.16	276.58
8	1/6	0	.2	230.00	389.89	359.50
8	1/8	$\sigma_G^2/3$.05	228.00	385.14	413.10
12	1/2	0	.005	222.00	300.11	303.94
12	1/4	$\sigma_G^2/3$.5	228.00	339.84	337.60
12	1/6	σ_G^2	.05	232.19	337.58	414.93
12	1/8	$3\sigma_G^2$.2	224.97	306.89	306.42
16	1/2	$3\sigma_G^2$.05	214.81	250.58	241.34
16	1/4	σ_G^2	.2	236.00	316.58	316.77
16	1/6	$\sigma_G^2/3$.005	232.01	368.49	351.09
16	1/8	0	.5	240.00	413.73	387.12
32	1/2	$\sigma_G^2/3$.2	239.25	289.61	286.25
32	1/4	0	.05	240.00	364.84	357.71
32	1/6	$3\sigma_G^2$.5	237.36	297.27	309.63
32	1/8	σ_G^2	.005	238.00	351.15	368.51

Table 18. Relation of the mean in generation 30 (G_{30}) to that predicted from parameters of simulation (\hat{G}_{30}) and from parameters observed in generation six (\hat{G}'_{30}) (complete dominance)

Popula- tion size	Selec- tion in- tensity	Environ- mental variation	Linkage	G_{30}	\hat{G}_{30}	\hat{G}'_{30}
8	1/2	σ_G^2	.5	191.88	256.78	246.33
8	1/4	$3\sigma_G^2$.005	196.00	265.42	232.72
8	1/6	0	.2	216.00	355.03	317.18
8	1/8	$\sigma_G^2/3$.05	215.06	351.15	319.63
12	1/2	0	.005	211.00	281.78	243.75
12	1/4	$\sigma_G^2/3$.5	213.33	314.20	315.95
12	1/6	σ_G^2	.05	212.56	312.29	297.92
12	1/8	$3\sigma_G^2$.2	211.19	287.35	301.40
16	1/2	$3\sigma_G^2$.05	200.12	241.30	249.09
16	1/4	σ_G^2	.2	214.62	295.18	311.29
16	1/6	$\sigma_G^2/3$.005	211.02	337.58	342.24
16	1/8	0	.5	220.00	374.52	406.85
32	1/2	$\sigma_G^2/3$.2	218.25	273.20	290.65
32	1/4	0	.05	219.84	334.62	367.90
32	1/6	$3\sigma_G^2$.5	217.99	279.40	295.40
32	1/8	σ_G^2	.005	217.07	323.42	366.17

Table 19. The mean difference between expected and observed selection differentials in standard deviation units for non-epistatic models

Total size of population	Selection intensity	Expected differential ^a	Additive	Complete dominance	Over-dominance
16	1/2	.758	.01±.05	.01±.06	.02±.06
24	1/2	.772	.05±.07	.05±.09	.03±.03 ^b
32	1/2	.780	.01±.03	.01±.03	.02±.05
32	1/4	1.236	.01±.08	.07±.09	.04±.09
48	1/4	1.245	.03±.08	.03±.09	.05±.07
48	1/6	1.464	.02±.09	.62±.44	.05±.11
64	1/2	.798	.01±.04	.02±.02	.03±.03
64	1/4	1.271	.04±.06	.04±.06	.06±.07
64	1/8	1.648	.05±.11	.10±.13	.12±.14
72	1/6	1.500	.05±.08	.07±.08	.08±.09
96	1/6	1.500	.03±.13	.08±.07	.06±.10
96	1/8	1.648	.06±.10	.04±.07	.03±.10
128	1/4	1.271	.02±.07	.42±.29	.08±.05
128	1/8	1.648	.01±.02	.73±.48	.11±.07
192	1/6	1.500	.03±.05	.01±.04	.03±.04
256	1/8	1.648	.03±.06	.04±.06	.05±.08

^aCalculated from Table 20, Fisher and Yates (1943) for $N < 50$.

^bNegative estimate of difference (i.e., observed $>$ expected).

Table 20. The mean difference between expected and observed selection differentials in standard deviation units for classical epistatic models

Total size of population	Selection intensity	Expected differential ^a	Optimum number	Duplicate factors	Complementary factors
16	1/2	.758	.01±.05	.03±.06	.01±.04
24	1/2	.772	.03±.06	.14±.19	.03±.06
32	1/2	.780	.01±.04	.01±.05	.02±.02
32	1/4	1.236	.02±.10	.04±.09	.03±.13
48	1/4	1.245	.04±.09	.08±.11	.04±.07
48	1/6	1.464	.23±.17	.82±.37 ^b	.59±.43
64	1/2	.798	.01±.04	.02±.05	.02±.04
64	1/4	1.271	.06±.07	.06±.11	.04±.08
64	1/8	1.648	.11±.16	.26±.12 ^b	.06±.09
72	1/6	1.500	.07±.08	.06±.10	.08±.05
96	1/6	1.500	.13±.10	.10±.10	.06±.09
96	1/8	1.648	.07±.09	.05±.12	.10±.20
128	1/4	1.271	.43±.35	.83±.24 ^c	.16±.14
128	1/8	1.648	.49±.35	1.07±.30 ^c	.79±.14 ^c
192	1/6	1.500	.03±.05	.02±.06	.02±.03
256	1/8	1.648	.07±.06	.08±.07	.04±.04

^aCalculated from Table 20, Fisher and Yates (1943) for $N < 50$.

^bStatistically significant ($P < .05$).

^cStatistically significant ($P < .01$).

Table 21. The mean difference between expected and observed selection differentials in standard deviation units for conditional epistatic models

Total size of population	Selection intensity	Expected differential ^a	AxA	AxD	DxD
16	1/2	.758	.02±.07	.03±.06	.01±.06
24	1/2	.772	.05±.13	.04±.06	.03±.05
32	1/2	.780	.01±.03	.01±.04	.01±.04
32	1/4	1.236	.03±.07	.04±.07	.01±.10
48	1/4	1.245	.03±.08 ^b	.02±.10	.05±.07
48	1/6	1.464	.10±.15	.12±.13	.03±.13
64	1/2	.798	.02±.04	.02±.04	.02±.03
64	1/4	1.271	.02±.07	.04±.06	.03±.08
64	1/8	1.648	.10±.12	.07±.17	.10±.10
72	1/6	1.500	.02±.09	.03±.07	.03±.10
96	1/6	1.500	.01±.11	.03±.10	.03±.08
96	1/8	1.648	.05±.12	.06±.10	.02±.10
128	1/4	1.271	.03±.30	.07±.06	.05±.04
128	1/8	1.648	.01±.07	.18±.12	.03±.06
192	1/6	1.500	.03±.07	.02±.04	.02±.06
256	1/8	1.648	.02±.07	.04±.06	.02±.05

^aCalculated from Table 20, Fisher and Yates (1943) for $N < 50$.

^bNegative estimate of difference (i.e., observed > expected).

Table 22. Monte Carlo estimates of the standard errors of estimation of realized heritability from selection experiments (non-epistatic gene action)

Total size of population	Selection intensity	Environmental variation ^a	Additive	Complete dominance	Over-dominance
16	1/2	σ_G^2	.55	.44	.32
32	1/2	$3\sigma_G^2$.28	.27	.28
32	1/4	$3\sigma_G^2$.28	.24	.20
48	1/4	$\sigma_G^2/3$.19	.12	.08
64	1/2	$\sigma_G^2/3$.16	.16	.10
64	1/4	σ_G^2	.15	.12	.11
64	1/8	$\sigma_G^2/3$.08	.08	.05
72	1/6	σ_G^2	.09	.08	.07
96	1/6	$\sigma_G^2/3$.09	.06	.03
96	1/8	$3\sigma_G^2$.08	.10	.09
192	1/6	$3\sigma_G^2$.07	.06	.05
256	1/8	σ_G^2	.05	.05	.03

^aGiven in terms of the expected genotypic variance among the initial generation of offspring.

Table 23. Monte Carlo estimates of the standard errors of estimation of realized heritability from selection experiments (classical epistatic models)

Total size of population	Selection intensity	Environmental variation ^a	Optimum number	Duplicate factors	Complementary factors
16	1/2	σ_G^2	.57	.31	.12
32	1/2	$3\sigma_G^2$.33	.29	.27
32	1/4	$3\sigma_G^2$.22	.15	.25
48	1/4	$\sigma_G^2/3$.18	.16	.07
64	1/2	$\sigma_G^2/3$.16	.17	.21
64	1/4	σ_G^2	.14	.15	.09
64	1/8	$\sigma_G^2/3$.09	.09	.12
72	1/6	σ_G^2	.10	.08	.10
96	1/6	$\sigma_G^2/3$.07	.09	.07
96	1/8	$3\sigma_G^2$.08	.10	.12
192	1/6	$3\sigma_G^2$.07	.07	.07
256	1/8	σ_G^2	.04	.05	.06

^aGiven in terms of the expected genotypic variance among the initial generation of offspring.

Table 24. Monte Carlo estimates of the standard errors of estimation of realized heritability from selection experiments (conditional epistatic models)

Total size of population	Selection intensity	Environmental variation ^a	AxA	AxD	DxD
16	1/2	σ_G^2	.35	.15	.45
32	1/2	$3\sigma_G^2$.33	.30	.35
32	1/4	$3\sigma_G^2$.25	.15	.15
48	1/4	$\sigma_G^2/3$.07	.18	.10
64	1/2	$\sigma_G^2/3$.14	.15	.12
64	1/4	σ_G^2	.15	.14	.10
64	1/8	$\sigma_G^2/3$.07	.12	.05
72	1/6	σ_G^2	.12	.08	.09
96	1/6	$\sigma_G^2/3$.04	.07	.05
96	1/8	$3\sigma_G^2$.09	.08	.08
192	1/6	$3\sigma_G^2$.08	.08	.07
256	1/8	σ_G^2	.04	.02	.03

^aGiven in terms of the expected genotypic variance among the initial generation of offspring.

Table 25. Parameter sets included in the 1/16 fractional replication of the 4⁴ factorial plan

Set	Parent population size	Selection intensity	Environmental variation ^a	Linkage ^b
1	8	1/2	σ_G^2	.5
2	8	1/4	$3\sigma_G^2$.005
3	8	1/6	0	.2
4	8	1/8	$\sigma_G^2/3$.05
5	12	1/2	0	.005
6	12	1/4	$\sigma_G^2/3$.5
7	12	1/6	σ_G^2	.05
8	12	1/8	$3\sigma_G^2$.2
9	16	1/2	$3\sigma_G^2$.05
10	16	1/4	σ_G^2	.2
11	16	1/6	$\sigma_G^2/3$.005
12	16	1/8	0	.5
13	32	1/2	$\sigma_G^2/3$.2
14	32	1/4	0	.05
15	32	1/6	$3\sigma_G^2$.5
16	32	1/8	σ_G^2	.005

^aGiven in terms of the expected genotypic variance among the initial generation of offspring.

^bGiven in terms of recombination frequency between adjacent loci.

Table 26. Genetic models simulated

Model	Genotypic values			Mean (expected values for 40 loci)	Variance	
	AA	Aa	aa			
Additive	BB:	12	11	10	200	20
	Bb:	11	10	9		
	bb:	10	9	8		
Complete dominance	BB:	11	11	9	200	30
	Bb:	11	11	9		
	bb:	9	9	7		
Overdominance	BB:	8	10	8	200	40
	Bb:	10	12	10		
	bb:	8	10	8		
Optimum number	BB:	7	10	11	200	30
	Bb:	10	11	10		
	bb:	11	10	7		
Duplicate factors	BB:	10	10	10	195	18.75
	Bb:	10	10	10		
	bb:	10	10	6		
Complementary factors	BB:	11	11	9	202.5	20
	Bb:	11	11	9		
	bb:	9	9	9		
Additive-by-additive initial variance	BB:	12	10	8	200	20
	Bb:	10	10	10		
	bb:	8	10	12		
Additive-by-dominance initial variance	BB:	12	9	10	200	20
	Bb:	9	10	11		
	bb:	10	11	8		
Dominance-by-dominance initial variance	BB:	9	11	9	200	20
	Bb:	11	9	11		
	bb:	9	11	9		