EFFECT OF TRUNCATION SELECTION ON GENETIC VARIABILITY

By

DANIEL A. SØRENSEN

Ing. Agr. (Buenos Aires)
M.Sc. (Reading)

Thesis presented for the Degree of

Doctor of Philosophy

of the University of Edinburgh

in the Faculty of Science.

Institute of Animal Genetics,
University of Edinburgh.

April, 1980.

I DEDICATE THIS THESIS TO MY PARENTS.

DECLARATION

I declare that this Thesis is my own composition and that the research described herein has been done by me.

ABSTRACT

The changes of genetic parameters caused by selection are due to changes in gene frequencies and due to the generation of linkage disequilibrium. In quantitative traits, the effects associated with gene frequency changes cannot be predicted but those due to linkage disequilibrium are predictable in terms of parameters of the base population, the sign of the change depending on the type of selection applied. With directional selection the sign is negative; with disruptive selection, it is positive. These predictions however are based on models which assume an infinite number of loci (infinitesimal model). The work described in this thesis examines the validity of these predictions with models of a finite number of loci in short term selection programmes.

The first eight chapters deal with directional selection. Initially some two locus theory is developed and the results are extended to quantitative models with use of Montecarlo simulation techniques. With additive gene action predictions of selection response and of reduction in variance based on infinitesimal theory are accurate provided gene frequencies are not extreme. With dominance these predictions are inaccurate even in the first cycle of selection. In order to quantify the importance of changes of genetic parameters, the difference between observed and predicted responses to selection relative to the standard deviation of selection response is discussed for various models.

Two experiments with Drosophila designed to study changes of genetic parameters with selection are reported. Only one out of the four replicates showed evidence of negative linkage disequilibrium

and the results are interpreted in the light of models studied in earlier chapters.

It is concluded that expected changes of genetic parameters in short-term selection studies are not likely to be predicted accurately but that the generation of disequilibrium plays a fundamental role in these changes.

The last chapters deal with disruptive selection. An experiment with Drosophila is reported and the results are shown to be consistent with theoretical expectations which predict large increases in genetic parameters due to positive disequilibrium. Some further theory is developed which clarifies various aspects of the experimental results.

CONTENTS

CHAPTER Page

CHAPTER Page

CHAPTER Page

CHAPTER 1

INTRODUCTION

Introduction.

In general the aim of an animal breeding programme is to obtain high rates of genetic gain from a given amount of initial genetic variation. Accurate estimates of genetic parameters are paramount in providing information that will lead to the choice of the most efficient breeding plan and to the prediction of the expected response to selection.

As a consequence of selection, genetic variances and heritabilities change and rather little attention has been given to studying the magnitude of these changes and their impact on short-term selection predictions.

In large populations, changes of genetic parameters induced by selection come about through changes in frequencies of the genes affecting the trait and due to the generation of covariances between the frequencies of these genes, i.e. linkage disequilibrium.

The changes due to gene frequency changes are highly dependent on the distribution of gene effects and frequencies in the base population, the information on which is small in Drosophila and almost non-existent in other species. We can then do no more than lay down the conditions under which gene frequency changes will be at a minimum which basically depend on the magnitude of the proportionate effects of the genes. In other words, with large population size, for a given amount of initial genetic variation, the larger the number of loci affecting the trait, the smaller the expected change in genetic parameters due to changes in gene frequencies caused by selection. This type of argument together with

a general impression obtained from experimental evidence led people to tentatively suggest that parameters are not likely to change much during short term selection experiments and that consequently prediction of expected responses based on present heritability estimates may be valid for a period of five or more generations (i.e. Falconer, 1960; Hill, 1974).

The other way genetic parameters change is through the generation of linkage disequilibrium induced by the selection process. In a series of papers, Bulmer (1971, 1974, 1976b) developed a theory which led him to conclude that in large populations, if the character is determined by many loci, most of the changes in genetic variance in short-term selection programmes are due to the generation of linkage disequilibrium and he developed formulae that predict such changes. These formulae are functions of readily estimable parameters of the base population and are independent of the number, frequency and effects of the loci affecting the trait in question.

As was pointed out by Bulmer, his result should be considered as a limiting result in the sense that it will hold provided the number of loci is strictly infinite. In view of the theoretical and practical importance of Bulmer's results, it seems desirable to study their validity under a range of genetic models. In other words, we would like to know how accurately we can predict changes of genetic parameters when the assumption of an infinite number of loci is relaxed.

More generally, the purpose of this thesis is to attempt a better understanding of the consequences of short term selection programmes on genetic variability. The first eight chapters of this thesis deal with directional selection. Some theory is developed for the case of two locus models in order to understand the way the various parameters interact during the selection process. Attempts are made to extend the two locus results to quantitative models and considerable use is here made of Montecarlo simulation studies. At the end of this part of the thesis, experiments with Drosophila melanogaster are reported. These experiments were carried out to provide experimental evidence on the theory developed in earlier chapters.

The last chapters of the thesis deal with disruptive selection. Under this type of selection, gene frequency changes are small and most of the change in genetic parameters comes about through the generation of linkage disequilibrium. An experiment with Drosophila is described, the results of which are interpreted in the light of theoretical work based on some simple algebra and computer simulation.

CHAPTER 2

POPULATION GENETICS OF TWO LOCUS MODELS

Introduction

In this section we review some of the deterministic two-locus theory of population genetics, setting the notation to be used subsequently. We shall be dealing only with those aspects of the theory which have a bearing on the work of this thesis.

There are basically two types of models in this area: the continuous time model and the discrete generation model. We shall deal with the latter which gives a description of gene frequency changes due to selection particularly in the case of non-overlapping generations. Continuous-time models are discussed by Crow and Kimura (1970).

The basic question raised by the two locus problem concerns the interaction between linkage and selection which was probably first briefly discussed by Fisher (1930) who suggested that such an interaction might be important. The mathematical aspects of the problem were studied to a limited extent by Wright (1952) and later on expanded by Kimura (1956) using a continuous-time model, and Lewontin and Kojima (1960), and Bodmer and Parsons (1962) using a discrete-generation model. Following these papers a vast amount of literature has developed, much of which has been reviewed by Lewontin (1974) and Karlin (1975), and more recently by Hedrick *et al.* (1978).

Random Mating

Consider a large random mating population of diploid organisms with no mutation and discrete generations, and assume for simplicity that two alleles A and a, B and b, are segregating at autosomal loci A and B respectively. There are four possible chromosome types: AB, Ab, aB and ab with respective frequencies f_1, f_2, f_3 and f_4. The gene frequencies are p, (1-p) for alleles A and a and q, (1-q) for

alleles B and b respectively. As is well known,

$$f_1 = pq + D \qquad f_3 = (1-p)q - D$$
$$f_2 = p(1-q) - D \qquad f_4 = (1-p)(1-q) + D \tag{2.1}$$

such that $p = f_1 + f_2$ and $q = f_1 + f_3$, and

$$D = f_1 f_4 - f_2 f_3 \tag{2.2}$$

D is called the linkage disequilibrium parameter and can be defined in a number of ways. As defined in (2.2) it is equal to half the difference between coupling phase double heterozygote (AB/ab) over repulsion double heterozygote (Ab/aB) at the time of random union of gametes. However a more useful way of viewing D for our purposes is to define it as a covariance of gene frequency in gametes (Kojima & Lewontin, 1970; Slatkin, 1972). From this definition,

$$D = E\{(Xp-p)(Xq-q)\} \tag{2.3}$$

where Xp and Xq are the number (i.e. 0 or 1) of A and B alleles respectively. Thus from (2.3), $D = f_1 - pq$ and is equivalent to (2.2). When D = 0 the loci are independent and the gamete frequencies are given by the products of the frequencies of their constituent alleles.

In the absence of selection, recurrence equations for changes in chromosome frequencies in succeeding generations are easily derived. Let c be the recombination fraction between loci A and B. Chromosome AB, say, at generation t + 1 can be produced from the genotypes of generation t in two different ways. Firstly, it may be derived from genotypes AB/-- without recombination, where the notation, --, refers to the presence of an arbitrary allele at each locus. The probability of this event is (1-c) and the frequency of such a genotype in the population at time t is $f_1^{(t)}$. Secondly, chromosome AB may be the result of recombination between loci A and B in genotypes A-/-B with probability c. The frequency of this genotype combination is

pq, since in a large random mating population gene frequencies remain constant in all generations. Thus, with the assumption of random mating, we have:

$$f_1^{(t+1)} = (1-c)f_1^{(t)} + cpq, \text{ and similarly,}$$

$$f_2^{(t+1)} = (1-c)f_2^{(t)} + cp(1-q)$$

$$f_3^{(t+1)} = (1-c)f_3^{(t)} + c(1-p)q \qquad (2.4)$$

$$f_4^{(t+1)} = (1-c)f_4^{(t)} + c(1-p)(1-q).$$

Noting that $pq = f_1^{(t)} - D^{(t)}$, we can write,

$$f_1^{(t+1)} = f_1^{(t)} - cD^{(t)} \qquad (2.5a), \text{ and similarly}$$

$$f_2^{(t+1)} = f_2^{(t)} + cD^{(t)} \qquad (2.5b)$$

$$f_3^{(t+1)} = f_3^{(t)} + cD^{(t)} \qquad (2.5c)$$

$$f_4^{(t+1)} = f_4^{(t)} - cD^{(t)} \qquad (2.5d)$$

From (2.1) we have, $f_1^{(t)} = pq + D^{(t)}$ and $f_1^{(t+1)} = pq + D^{(t+1)}$. Substituting in (2.5a) we have the well known result:

$$D^{(t+1)} = (1-c)D^{(t)}$$

$$= (1-c)^t D^{(0)}, \text{ where } D^{(0)} \text{ is the initial linkage disequilibrium.}$$

In a large random mating population with discrete generations, D tends to zero at a rate (1-c).

Selection

Let us now consider the effect of selection. We first define an array of fitness parameters corresponding to the fitnesses of the genotypes resulting from the random union of the four different gametes, as illustrated in Table 2.1.

TABLE 2.1

Gamete Type	AB	Ab	aB	ab	
AB	W_{11}	W_{12}	W_{13}	W_{14}	
Ab	W_{21}	W_{22}	W_{23}	W_{24}	
aB	W_{31}	W_{32}	W_{33}	W_{34}	
ab	W_{41}	W_{42}	W_{43}	W_{44}	
Marginal Mean	W_1	W_2	W_3	W_4	Overall Mean: $\bar{W}$

$W_{ij}(=W_{ji})$ is the probability that an individual of the ijth genotype ($i = 1, \ldots, 4$; $j = 1, \ldots, 4$) survives from fertilization until it reproduces by mating at random. We regard W_{ij} as a measure of both survival and reproduction and thus we assume that all surviving adults have equal viability. Further, $W_{ij} = W_{ji}$ assumes no maternal effects on fitness. W_i is the expected fitness of the ith gametic phase, and is obtained as: $W_i = \sum_j W_{ij} f_j$; $\bar{W}$ is the average fitness of the population: $\bar{W} = \sum_{ij} W_{ij} f_{ij} = \sum_i W_i f_i$.

We shall consider the effect of selection at two stages: before and after recombination. From our definition of fitness the frequency of chromosome type AB amongst selected genotypes, before recombination takes place, is given by:

$$f_1^{(s,t)} = \{\frac{f_1}{\bar{W}}(f_1 W_{11} + f_2 W_{12} + f_3 W_{13} + f_4 W_{14})\}^{(t)} = (f_1 \frac{W_1}{\bar{W}})^{(t)}.$$

Similarly,

$$f_2^{(s,t)} = f_2 \frac{W_2}{\bar{W}}^{(t)}$$

$$f_3^{(s,t)} = f_3 \frac{W_3}{\bar{W}}^{(t)} \qquad (2.6)$$

$$f_4^{(s,t)} = f_4 \frac{W_4}{\bar{W}}^{(t)}$$

(In general we shall place subscripts and superscripts outside brackets when they are common to all enclosed parameters. Presence of superscript s indicates that the parameter in question is measured in the selected population. Absence of s implies that the parameter is measured before the operation of selection).

From the set of equations (2.6), we can write

$$D^{(s,t)} = (f_1 f_4 - f_2 f_3)^{(s,t)}$$

$$= \{\frac{1}{\bar{W}^2} (f_1 f_4 W_1 W_4 - f_2 f_3 W_2 W_3)\}^{(t)} \qquad (2.7)$$

We now consider the production of gametes of the selected genotypes. The frequency of gamete AB at generation t+1 in the gametic pool contributed by the selected genotypes, expressed in terms of parameters of the previous generation, before selection operated, is given by,

$$f_1^{(t+1)} = \{\frac{1}{\bar{W}}[f_1^2 W_{11} + f_1 f_2 W_{12} + f_1 f_3 W_{13} + f_1 f_4 W_{14}$$

$$- c(W_{14} f_1 f_4 - W_{23} f_2 f_3)]\}^{(t)}$$

Assuming no position effects on fitness so that $W_{14} = W_{23}$,

$$f_1^{(t+1)} = \{\frac{1}{\bar{W}}(f_1W_1 - c\,W_{14}D)\}^{(t)} \quad \text{, and similarly,}$$

$$f_2^{(t+1)} = \{\frac{1}{\bar{W}}(f_2W_2 + c\,W_{14}\,D)\}^{(t)}$$

$$f_3^{(t+1)} = \{\frac{1}{\bar{W}}(f_3W_3 + c\,W_{14}\,D)\}^{(t)} \qquad (2.8)$$

$$f_4^{(t+1)} = \{\frac{1}{\bar{W}}(f_4W_4 - c\,W_{14}\,D)\}^{(t)}$$

In general, the change in frequency of the ith gamete due to one cycle of selection is given by:

$$\Delta f_i^{(t)} = f_i^{(t)} - f_i^{(t-1)} = \{\frac{1}{\bar{W}}[f_i(W_i-\bar{W}) - k(i)\,c\,W_{14}D]\}^{(t)} \qquad (2.9),$$

where

$$k(i) = \begin{cases} 1 \text{ for } i = 1,4 \\ -1 \text{ for } i = 2,3 \end{cases} \qquad \text{(Moran, 1964)}$$

These sets of expressions are due to Lewontin and Kojima (1960). Lewontin (1964) considered the case of multiple loci, and Kimura and Ohta (1971) and Roux (1974) have given equivalent recurrence equations for a model involving multiple alleles. From the set of equations (2.8), it is readily shown that $p^{(t+1)} = f_1^{(t+1)} + f_2^{(t+1)} = \{\frac{1}{\bar{W}}(f_1W_1 + f_2W_2)\}^{(t)}$ and $q^{(t+1)} = \{\frac{1}{\bar{W}}(f_1W_1 + f_3W_3)\}^{(t)}$. Furthermore,

$$D^{(t+1)} = (f_1f_4 - f_2f_3)^{(t+1)}$$

$$= \{\frac{1}{\bar{W}^2}(f_1f_4\,W_1W_4 - f_2f_3\,W_2W_3 - c\,W_{14}\,D\,\bar{W})\}^{(t)} \qquad (2.10)$$

These recurrent equations enable us to describe the value that the parameter in question takes in terms of its value in the previous generation, before the action of selection.

From these various equations we can draw some important

conclusions:

(i) Starting at generation zero with a population in equilibrium, from (2.8) we conclude that changes in gametic frequencies (and therefore in gene frequencies) are independent of the degree of linkage between the loci involved, in the first cycle of selection. Therefore we can write $f_i^{(s,0)} = f_i^{(1)}$, independent of c.

(ii) From (2.6) and (2.8) we can write $f_i^{(s,t)} = f_i^{(t+1)}$ for $t \geqslant 0$, if $c = 0$. If $c \neq 0$ this relationship does not hold. In other words, unless there is complete linkage, the frequency of the gametes produced by the selected genotypes in the second cycle of selection depend on the recombination fraction c. Starting at generation zero (t=0), gamete frequencies are dependent on c if $t \geqslant 2$, unless $c = 0$.

(iii) In general we can write: $p^{(s,t)} = f_1^{(s,t)} + f_2^{(s,t)}$. Furthermore: $p^{(s,t)} = p^{(t+1)}$. Since $f_i^{(s,1)}$ is not dependent on c, provided $D^{(o)} = 0$, gene frequencies become dependent on c if $t \geqslant 3$.

(iv) From (2.7) and (2.10) we conclude that the amount of linkage disequilibrium generated in the gametes produced by the selected genotypes after the first cycle of selection is not dependent on c and is equal to the disequilibrium present within the chromosomes of the selected genotypes. That is, $D^{(s,o)} = D^{(1)}$, provided $D^{(o)} = 0$.

Several of these points will be pursued further in later sections.

CHAPTER 3

EFFECT OF DIRECTIONAL SELECTION ON QUANTITATIVE ADDITIVE MODELS - ONE CYCLE OF SELECTION

Introduction

In this chapter we apply the concepts developed so far in a quantitative genetic context. As a result of selection in a large population, the genotypic variance changes both due to gene frequency changes and due to the generation of linkage disequilibrium. We shall investigate the relative importance of these two processes in a first cycle of selection.

The classical theory of selection for a quantitative character assumes that there is no genetic variability in fitness (viability and fertility) and that the only selection operating is that imposed by the breeder. Given this assumption, artificial selection for a metric trait can then be considered as a case in which fitness is a function of the phenotype. Let the phenotype be determined by two additive non interacting loci, A and B, each with two alleles, A, a and B, b, and a normally distributed environmental component. We let p and q be the frequency of alleles having high value for some quantitative trait, A and B and we let a_1 and a_2 be the difference between the genotypic values of homozygotes and heterozygotes at locus A and B respectively. We shall assume that the population is in Hardy-Weinberg equilibrium. The genetic model can then be written as follows:

	AA	Aa	aa	BB	Bb	bb
Value	a_1	0	$-a_1$	a_2	0	$-a_2$
Frequency	p^2	$2p(1-p)$	$(1-p)^2$	q^2	$2q(1-q)$	$(1-q)^2$

The population mean: $M = -a_1(1-2p) - a_2(1-2q)$.

The equilibrium additive variance at each locus is:

$$Vg_A = 2p(1-p)a_1^2$$

$$Vg_B = 2q(1-q)a_2^2$$

The total equilibrium additive variance, Vg, being given by: $Vg = Vg_A + Vg_B$.

The gametic output is, as before

Gamete	AB	Ab	aB	ab
Value	$\frac{1}{2}(a_1+a_2)$	$\frac{1}{2}(a_1-a_2)$	$\frac{1}{2}(-a_1+a_2)$	$\frac{1}{2}(-a_1-a_2)$
Frequency	f_1	f_2	f_3	f_4

Due to our assumption of additivity, it follows that the gametic mean is equal to one half the genotypic mean and further due to our assumption of Hardy-Weinberg equilibrium, the total gametic variance is equal to one half the total genotypic variance, VG. It is easy to show that,

$$VG = Vg + 4a_1a_2D \qquad (3.1)$$

If there are many loci affecting the trait we have:

$$VG = 2\sum_i a_i^2 q_i(1-q_i) + 4\sum_{i<j}\sum a_i a_j D_{ij} \qquad (3.2)$$

The first term in (3.2) represents the independent contribution to the total genotypic variance of the n loci and we denote this term, the equilibrium additive variance, Vg. The second term in (3.2) reflects the contribution of covariances of allelic effects between the $\frac{1}{2}n(n-1)$ pairs of loci within gametes and following Bulmer (1971) we denote this term, joint disequilibrium and we shall symbolise it CLW. It should be clear that loci which are on different chromosomes also contribute to CLW. As was pointed out before, with random mating and no selection, D breaks down at a rate depending on the linkage

relationship of the loci involved. As D approaches zero, the second term in (3.2) tends to zero and the population is said to move towards a state of equilibrium.

Selection changes the value of VG through both its components. We shall now assume that the base population is in Hardy-Weinberg and linkage equilibrium and study the effect of one cycle of selection on each of the components of the total genotypic variance.

Changes of Total Genotypic Variance due to Changes of Gene Frequencies

Following a section in Kimura's (1958) paper, the effect of truncation selection for a metric character in a large population on gene frequency changes has been studied by Griffing (1960) and Latter (1965), and more recently by Kimura and Crow (1978). Some aspects of the general theory have been reviewed by Kempthorne (1977) and Pollak (1979).

In what follows we assume a large population in equilibrium and selection is such that a certain proportion, Q, of individuals that exceed a certain phenotypic value are saved for breeding. Mating amongst the selected group is at random. The genetic model is one in which there are many additive non interacting loci affecting the character and we shall focus our attention on one of these loci. The phenotypic variance, σ^2, is due to segregation at the rest of the loci affecting the trait plus a normally distributed environmental component. We shall first deal with the case in which the genotypic effects are small relative to the phenotypic standard deviation. This assumption is relaxed in the following section. In general we assume that the genotypic variance contributed by the locus is negligible relative to σ^2.

Genes of Small Effect

When the proportionate effects of the genes are small (i.e. a/σ - Falconer, 1960), we can ignore higher order terms in such quantities as a first approximation. In this case the change in gene frequency, Δp, per unit change in phenotype, P, is taken to be linear and can be expressed as follows (Falconer, 1960):

$$\Delta p = b_{pP} \Delta P ,$$

where b_{pP} is the linear regression of gene frequency on phenotype. Assuming normality,

$$\Delta p = i\frac{a}{\sigma}p(1-p) \qquad (3.3)$$

where i is the standardised selection differential or intensity of selection. Thus the selective value in (3.3) is approximated by the quantity ia/σ (Haldane, 1931).

The change in mean, M, resulting from gene frequency changes at this locus, noting that $\frac{dM}{dP}_{p=\frac{1}{2}} = 2a$, can be expressed as follows:

$$\frac{dM}{dP}\Delta p = \frac{i}{\sigma} 2a^2 p(1-p) \qquad (3.4)$$

and is proportional to the additive variance contributed by the locus. Furthermore, since the regression of gene frequency on phenotype is linear and d^2M/dp^2 and higher order derivatives are equal to zero, the change in mean or response to selection is symmetrical. If there are n loci affecting the character, the expected total response, R, is given by:

$$R = \frac{i}{\sigma} \sum_j 2a_j^2 p_j(1-p_j) = ih^2\sigma \qquad (3.5)$$

the usual formula of quantitative genetics.

We now look at the effect of gene frequency changes on the equilibrium additive variance, Vg, after this single cycle of selection.

The contribution from the j^{th} locus, $Vg_j^{(1)}$ is expressed as:

$$Vg_j^{(1)} = 2a_j^2(p_j+\Delta p_j)(1-p_j-\Delta p_j) \qquad (3.6)$$

Ignoring terms in Δp^2, the change in variance at this locus is given by,

$$\Delta Vg_j = 2a^2 p_j(1-p_j)(1-2p_j)\frac{ia}{\sigma} .$$

With n loci, the total change is:

$$\Delta Vg = \sum_j \Delta Vg_j = \sum_j 2a_j^2 p_j(1-p_j)(1-2p_j)a_j \frac{i}{\sigma} \qquad (3.7)$$

The general conclusion we draw from these well known results is that for a given intensity of selection and a given initial value of the equilibrium additive variance, its change will be smaller the larger the number of genes affecting the trait and as a first approximation, it will tend to zero as gene frequencies tend to 0.5.

Genes of Large Effect

When gene effects are large relative to the phenotypic standard deviation, second order terms in a/σ can no longer be ignored and as we shall see, the expected response to selection is no longer symmetrical. This problem was studied by Latter (1965) whose paper forms the basis of the discussion that follows.

Consider a large population which is normally distributed for some trait, with mean M and total phenotypic variance σ^2. Let $\bar{X}_{ij}$ denote the mean of those individuals whose genotype at a particular locus is A_iA_j and let σ^2_{ij} be their phenotypic variance arising from both segregation at other loci and from a normally distributed environmental effect.

Following Kimura (1958) we define the selective value of the ij^{th} genotype, W_{ij}, as the probability that an individual of such a genotype is selected. We assume, as we did before, that a proportion Q of those phenotypes that exceed a certain truncation point, T, are saved for breeding. With our assumption of normality and ignoring the difference between σ^2_{ij} and σ^2, we can write:

$$W_{ij} = f(\bar{X}_{ij}) = \frac{1}{\sigma\sqrt{2\pi}} \int_T^\infty \exp\{-\frac{(X-\bar{X}_{ij})^2}{2\sigma^2}\}\, dx \qquad (3.8)$$

Expanding (3.8) in a Taylor series about M, to second order terms, we obtain:

$$f(\bar{X}_{ij}) = f(M) + (\bar{X}_{ij} - M)\left.\frac{df}{d\bar{X}_{ij}}\right|_{\bar{X}_{ij}=M} + \tfrac{1}{2}(\bar{X}_{ij}-M)^2\left.\frac{d^2f}{d\bar{X}_{ij}^2}\right|_{\bar{X}_{ij}=M} + \ldots$$

Noting that $Z = iQ$, where Z is the ordinate at the point of truncation x_T, of the standard normal distribution, the selective value can be expressed as a second order approximation as follows:

$$W_{ij} = Q + \frac{Q_i}{\sigma}(\bar{X}_{ij}-M) + \tfrac{1}{2}\frac{Q_i x_T}{\sigma^2}(\bar{X}_{ij}-M)^2 \qquad (3.9)$$

The relative selective value is given by

$$\frac{W_{ij}}{\bar{W}} = \frac{Q}{\bar{W}}\{1 + \frac{i}{\sigma}(\bar{X}_{ij}-M) + \frac{ix_T}{2\sigma^2}(\bar{X}_{ij}-M)^2\} \qquad (3.10)$$

where $\bar{W}$, the proportion selected at this locus, is defined as follows, assuming random mating:

$$\bar{W} = \Sigma\Sigma_{ij} p_i p_j W_{ij}$$

Assuming two alleles per locus, we can approximate $\bar{W}$ as follows:

$$\bar{W} = Q + \frac{Q_i x_T}{2\sigma^2} 2a^2 p(1-p) \qquad (3.11)$$

to the same order of approximation.

The relative selective values of the three genotypes are expressed as follows:

$$AA : \frac{W_{11}}{\bar{W}} = \{1 + 2\frac{a}{\sigma}i(1-p) + 2(\frac{a}{\sigma})^2 i\ x_T(1-p)^2\}\frac{Q}{\bar{W}}$$

$$Aa : \frac{W_{12}}{\bar{W}} = \{1 + \frac{a}{\sigma}i(1-2p) + \tfrac{1}{2}(\frac{a}{\sigma})^2 i\ x_T(1-2p)^2\}\frac{Q}{\bar{W}}$$

$$aa : \frac{W_{22}}{\bar{W}} = \{1 - 2\frac{a}{\sigma}ip + 2(\frac{a}{\sigma})^2 i\ x_T p^2\}\frac{Q}{\bar{W}} \quad .$$

If the gene effects are large the second order terms in these expressions can be important. Latter (1965) showed that the relative selective values are poorly estimated by the first order approximation when $(\bar{X}_{ij}-M)/\sigma$ is larger than 0.5 and the proportion selected less than about 40%. In fact, his Figure 1 shows that for proportions selected less than 40%, the first order approximation underestimates the exact selective value for both positive and negative values of $(\bar{X}_{ij}-M)/\sigma$, obtained from tables of the normal distribution. Thus for an additive model, provided gene frequencies are not far from intermediate values, the effect of ignoring second order terms on gene frequency changes should not be too drastic, as we shall see shortly.

We now define $W_i/\bar{W}$ as the relative selective value of the i^{th} allele. For example for allele A,

$$\frac{W_1}{\bar{W}} = \frac{1}{\bar{W}}\{pW_{11} + (1-p)W_{12}\}$$

$$= \frac{Q}{\bar{W}}\{1 + i\frac{a}{\sigma}(1-p) + \tfrac{1}{2}i(\frac{a}{\sigma})^2\ x_T(1-p)\} \qquad (3.12)$$

The change in gene frequency, Δp, is given by:

$$\Delta p = \frac{p}{\bar{W}}(W_1 - \bar{W}) \qquad (3.13)$$

Replacing (3.11) and (3.12) in (3.13), and finally letting $Q/\bar{W} \simeq 1 - \frac{ix_T}{\sigma^2} a^2 p(1-p)$, we obtain the second order approximation for Δp (Latter, 1965):

$$\Delta p = \frac{a}{\sigma} ip(1-p) + (\frac{a}{\sigma})^2 \frac{ix_T}{2} p(1-p)(1-2p) \qquad (3.14)$$

From this result we draw the following conclusions:

(i) The expected response to selection will be poorly estimated using (3.5) unless gene frequencies are intermediate or the intensity of selection is 50%, so that $x_T = 0$. For gene frequencies less than 0.5, (3.5) will tend to overestimate the true response and the opposite holds for gene frequencies larger than 0.5.

(ii) Although the functional relationship between gene frequency and genotypic mean is linear, the change in gene frequency per unit change in phenotype is not, and therefore two way selection experiments will be asymmetrical. Given i and a/σ, this asymmetry is maximum when initial gene frequencies are $0.5 \pm 1/\sqrt{12}$. The problem of asymmetry of selection response was further discussed by Latter (1965) and more recently by Robertson (1977c) and Mäki-Tanila (1980).

We now turn to changes in the equilibrium additive variance. This change is expressed as follows:

$$\Delta Vg = \sum_j \Delta Vg_j = \sum_j 2a_j^2\{(1-2p_j)\Delta p_j - \Delta p_j^2\} \qquad (3.15)$$

For simplicity, assume all genes have the same effect and frequency.

Then the proportional change in the equilibrium additive variance relative to its value before selection, $Vg^{(0)}$, is

$$\frac{\Delta Vg}{Vg^{(0)}} = i(\frac{a}{\sigma})(1-2p) + \frac{ix_T}{2}(\frac{a}{\sigma})^2(1-2p)^2 - i^2(\frac{a}{\sigma})^2 p(1-p) \qquad (3.16)$$

In order to get some insight we now produce some numerical results. We first look at the accuracy with which expressions (3.3) and (3.14) predict gene frequency changes for different values of a/σ and initial gene frequencies. Table 3.1 corresponds to a proportion selected of 10% and Table 3.2 to 20%. In both tables $\Delta p(E)$ refers to exact changes in gene frequencies obtained by numerical integration of the normal distribution and having allowed for the fact that σ^2_{ij} is smaller than σ^2. $\Delta p(1)$ corresponds to the predictions made using (3.3) and $\Delta p(2)$ using (3.14).

The asymmetry in gene frequency changes, particularly at high values of a/σ, are clearly illustrated in these results. What is probably most striking is the high degree of accuracy with which expression (3.14) predicts gene frequency changes. The first order approximation (eq. 3.3) becomes rather poor for values of a/σ of 0.5. At gene frequencies of 0.3 or 0.7 the difference between the exact results and those predicted using the first order approximation differ by about 10% for both selection intensities. For $\frac{a}{\sigma} = 0.2$, this discrepancy decreases to about 4%.

The effect of these results on the changes of the equilibrium additive variance is presented in Table 3.3. $\Delta Vg^{(1)}$ is defined as $Vg^{(1)} - Vg^{(0)}$, where $Vg^{(1)}$ is the equilibrium additive variance after one cycle of selection obtained from the corresponding gene frequencies estimated from (3.6). $Vg^{(0)}$ refers to the value of the parameter at

TABLE 3.1: Predicted and observed changes of gene frequency at an additive locus after one cycle of selection, assuming the base population is in equilibrium. Proportion selected: 10%

$\frac{a}{\sigma}$		q	0.1	0.3	0.5	0.7	0.9
0.2	Δp(E)		0.035	0.077	0.088	0.070	0.028
	Δp(1)		0.032	0.074	0.088	0.074	0.032
	Δp(2)		0.035	0.078	0.088	0.070	0.028
0.5	Δp(E)		0.099	0.205	0.216	0.160	0.060
	Δp(1)		0.079	0.184	0.219	0.184	0.079
	Δp(2)		0.099	0.208	0.219	0.161	0.059
0.9	Δp(E)		0.205	0.382	0.376	0.253	0.086
	Δp(1)		0.142	0.332	0.395	0.332	0.142
	Δp(2)		0.208	0.408	0.395	0.255	0.077

TABLE 3.2: Proportion Selected: 20%

$\frac{a}{\sigma}$		q	0.1	0.3	0.5	0.7	0.9
0.2	Δp(E)		0.027	0.061	0.070	0.057	0.024
	Δp(1)		0.025	0.059	0.070	0.059	0.025
	Δp(2)		0.027	0.061	0.070	0.057	0.024
0.5	Δp(E)		0.073	0.160	0.176	0.135	0.052
	Δp(1)		0.063	0.147	0.175	0.147	0.063
	Δp(2)		0.074	0.159	0.175	0.135	0.052
0.9	Δp(E)		0.145	0.302	0.317	0.226	0.079
	Δp(1)		0.113	0.265	0.315	0.265	0.113
	Δp(2)		0.148	0.305	0.315	0.225	0.079

generation zero. This way of predicting $Vg^{(1)}$ is not strictly correct to first order terms since it includes a term in Δp^2 whose second order term is ignored. $\Delta Vg(E)$ is the exact result, computed from the exact gene frequency changes. Table 3.3 shows that for a proportion selected of 20%, the prediction using this approach is satisfactory for values of a/σ smaller or equal to 0.2 and reasonably accurate for values of $a/\sigma = 0.5$ unless gene frequencies are extreme.

TABLE 3.3: Values of $\Delta Vg(1)/\Delta Vg(E)$. Proportion selected: 20%. (See text for explanation).

$\frac{a}{\sigma}$	q	0.1	0.3	0.5	0.7	0.9
0.2		0.94	0.97	1.00	1.04	1.08
0.5		0.87	0.97	0.99	1.11	1.22
0.9		0.82	1.21	0.99	1.24	1.48

This way of predicting the value of the equilibrium additive variance after the first cycle of selection, assuming as we did that the population is initially in Hardy-Weinberg and linkage equilibrium seems to be operationally useful and due to its simplicity will be adopted in the comparison between predicted and observed results in the Montecarlo simulations that follow.

Throughout this section we have shown that given the selection intensity and initial gene frequencies, the change in the equilibrium additive variance is governed by the term a/σ. If there are n loci of equal effects and frequencies affecting the trait in question, a/σ

can be expressed as $\left[\frac{h^2}{2np(1-p)}\right]^{\frac{1}{2}}$. For a given amount of additive genetic variation, as the number of loci becomes very large the term a/σ becomes negligible and therefore the change in the genotypic variance due to changes of gene frequencies is likely to be small, particularly during the first few cycles of selection. (See Crow & Kimura (1970) for a more rigorous treatment of this point).

Changes of Total Genotypic Variance due to the Generation of Linkage Disequilibrium.

We now study the generation of linkage disequilibrium in the first cycle of selection and its effect on the genotypic variance of a metric trait. Having evaluated its effect on VG, we shall compare it with the effect of gene frequency changes in order to understand their relative contribution to the changes in the genotypic variance. As we did before, we assume that the base population is in Hardy-Weinberg and linkage equilibrium. We defer the general review of selection in multilocus systems for the next chapter.

Using the notation of earlier sections, the problem is reduced to obtaining an explicit expression for

$$D^{(\bar{S},\bar{O})} = D^{(\bar{1})} = (f_1f_4-f_2f_3)^{(\bar{S},\bar{O})} = (f_1f_4-f_2f_3)^{(\bar{1})} \qquad (3.17)$$

Since we assume that the gametes produced by the selected genotypes are shed into a conceptually infinite gametic pool, where gametes pair at random to form the zygotes, the covariance of gene frequencies in gametes reflects the degree of linkage disequilibrium in the chromosomes of the offspring of generation one. Due to our assumption of initial equilibrium, the frequencies of the four gametic types

produced by the genotypes of the first cycle of selection is given by the set of expressions (2.6) for t = 0. Since $D^{(0)} = 0$, (2.7) reduces to:

$$D^{(S,0)} = \{\frac{1}{\bar{W}^2}\left[f_1 f_4(W_1 W_4 - W_2 W_3)\right]\}^{(0)} \qquad (3.18)$$

Using the procedure described before to approximate the selective value of a particular genotype in a two-locus situation, it can be shown that the relative selective value of the four gametic phases is given, to second order terms, by the following set of expressions:

$$\frac{W_1}{\bar{W}} = \frac{1}{\bar{W}}\{Q + \frac{Q_i}{\sigma}\left[a_1(1-p) + a_2(1-q)\right] + \frac{Q_i x_T}{2\sigma^2}\left[a_1^2(1-p) + a_2^2(1-q) + 2a_1 a_2 f_4\right]\}$$

$$\frac{W_2}{\bar{W}} = \frac{1}{\bar{W}}\{Q + \frac{Q_i}{\sigma}\left[a_1(1-p) - a_2 q\right] + \frac{Q_i x_T}{2\sigma^2}\left[a_1^2(1-p) + a_2^2 q - 2a_1 a_2 f_3\right]\}$$

$$\frac{W_3}{\bar{W}} = \frac{1}{\bar{W}}\{Q + \frac{Q_i}{\sigma}\left[-a_1 p + a_2(1-q)\right] + \frac{Q_i x_T}{2\sigma^2}\left[a_1^2 p + a_2^2(1-q) - 2a_1 a_2 f_2\right]\}$$

$$\frac{W_4}{\bar{W}} = \frac{1}{\bar{W}}\{Q + \frac{Q_i}{\sigma}\left[-a_1 p - a_2 q\right] + \frac{Q_i x_T}{2\sigma^2}\left[a_1^2 p + a_2^2 q + 2a_1 a_2 f_1\right]\} \qquad (3.19)$$

where

$$\bar{W} = Q + \frac{Q_i x_T}{2\sigma^2}\{2a_1^2 p(1-p) + 2a_2^2 q(1-q)\}, \text{ for } D^{(0)} = 0.$$

Substituting the set of expressions (3.19) in (3.18) and finally letting $Q/\bar{W} \simeq 1 - \frac{ix}{2\sigma^2}\{2a_1^2 p(1-p) + 2a_2^2 q(1-q)\}$ to second order terms, the covariance of gene frequencies in gametes produced by the selected

parents, $D^{(1)}$, is obtained in terms of parameters of the population before selection, namely:

$$D^{(1)} = \{-i(i-x_T)\ a_1p(1-p)\ a_2q(1-q)/\sigma^2\}^{(0)} \qquad (3.20)$$

(Hill & Robertson, 1966). The quantity $-i(i-x_T)$ is always negative (unless there is no selection in which case it is zero) whether we select for high or low value of the trait and is clearly symmetrical. Directional selection then leads to a reduction of the total genotypic variance due to the generation of negative correlations between loci within individual parental gametic contributions. The value of the quantity $i(i-x_T)$ varies from 0.918 for Q = 1% to 0.637 for Q = 50%.

Let the trait of interest be determined by n additive loci. Let p_j be the frequency of the allele having high value for the trait and a_j the difference between homozygote and heterozygote at locus j. With n loci there are $\frac{1}{2}n(n-1)$ pairs of loci within each parental gamete and n(n-1) covariance terms contributing to the total variance between gametes. We can then express $D^{(1)}$ in the offspring of selected parents as follows:

$$D^{(1)} = -i(i-x_T) \sum_{j\neq k} \sum a_j p_j (1-p_j)\ a_k p_k (1-p_k)/\sigma^2$$

The total genotypic variance at generation 1, $VG^{(1)}$, is

$$VG^{(1)} = (\Sigma_j 2a_j^2 p_j(1-p_j) + \sum_{j\neq k} \sum 2a_j a_k D_{jk})^{(1)} \qquad (3.21)$$

where the first term is the equilibrium additive variance in generation 1, $Vg^{(1)}$, and the second term is a covariance of allelic effects between pairs of loci, which following Bulmer (1971) we called joint disequilibrium and we symbolised CLW. Since $CLW^{(0)} = 0$, we can write, $\Delta VG = \Delta Vg + CLW^{(1)}$. The reduction in VG due to joint disequilibrium is

$$CLW^{(1)} = -\{i(i-x_T)/\sigma^2\} \sum_{j\neq k} \sum 2a_j^2 a_k^2 p_j(1-p_j)p_k(1-p_k)$$

$$= -\{i(i-x_T)/\sigma^2\}\{\frac{Vg^{(0)2}}{2} - \Sigma_j 2a_j^2 p_j^2 (1-p_j)^2\} \ ,$$

where σ^2, $Vg^{(0)}$ and p_j are the value that the parameters take before selection. This expression can be written as (Hill, personal communication):

$$CLW^{(1)} = -\tfrac{1}{2}\frac{i(i-x_T)}{\sigma^2}\ Vg^{2(0)}(1-\frac{1+CV_v^2}{n})$$

where CV_v is the coefficient of variation of the quantities $a_j p_j(1-p_j)$. If all loci contribute equally to Vg and n is large, the reduction in variance due to joint disequilibrium is given by:

$$CLW^{(1)} = -\tfrac{1}{2}i(i-x_T)Vg^{(0)}h^{2(0)} \qquad (3.22)$$

as was shown by Bulmer (1971), independent of the number, frequency and effects of the genes involved.

The validity of this result rests on the assumption that the phenotypic distribution is normal and that the regression of allelic effects on phenotype is linear and homoscedastic. These conditions are satisfied when the phenotypic values are due to the segregation of many additive and independent loci and an independent and normally distributed environmental effect. With a finite number of genes particularly of large effect and extreme frequencies, these assumptions are unlikely to hold, introducing a considerable degree of complexity since higher order moments and higher order disequilibria may become relevant in an attempt to describe the process. It seems therefore pertinent to study the behaviour of expression (3.22) under some simple models in order to obtain some insight on its robustness to departures from the assumptions that lead to its derivation. We do this using the technique of numerical integration of the normal distribution used previously for the study of gene frequency changes.

Table 3.4 gives values of $CLW^{(1)}$, starting with a population in equilibrium, for different initial gene frequencies and number of loci. We have assumed that the initial genotypic variance is of 4 squared units, the heritability of the character is 40% and the upper 20% of the population is selected.

Proportionate effects of the genes are shown in the bottom of each gene frequency and number of loci combination. At extreme frequencies with genes of moderate to large effect, the departures from the predicted value of -0.625 using expression (3.22) are indeed quite substantial. There is a marked degree of asymmetry at extreme

TABLE 3.4: Reduction of the genotypic variance due to joint disequilibrium, after a single cycle of directional selection. It is assumed that $CLW^{(0)}=0$; $VG^{(0)}= 4$; $h^{2(0)}= 0.4$; $Q = 20\%$. The values in the table are obtained by numerical integration of the normal distribution.

Number of loci	Initial Gene Frequencies				
	0.1	0.3	0.5	0.7	0.9
10	-1.249	-0.739	-0.541	-0.381	-0.162
$\frac{a}{\sigma}$	0.47	0.31	0.28	0.31	0.47
30	-1.002	-0.717	-0.597	-0.491	-0.316
$\frac{a}{\sigma}$	0.27	0.18	0.16	0.18	0.27
80	-0.854	-0.688	-0.614	-0.546	-0.422
$\frac{a}{\sigma}$	0.17	0.11	0.10	0.11	0.17
500	-0.715	-0.652	-0.624	-0.595	-0.540
$\frac{a}{\sigma}$	0.07	0.04	0.04	0.04	0.07
$\to\infty$	-0.625	-0.625	-0.625	-0.625	-0.625
$\frac{a}{\sigma}\to$	0	0	0	0	0

gene frequencies which disappears at a slow rate with increasing number of loci. The effect of increasing the number of loci on the value of $CLW^{(1)}$ attained clearly depends on the initial gene frequencies and it is small when gene frequences are between 0.3 and 0.5. These results suggest that the model is rather sensitive to departures from the assumptions on which it rests and that higher order moments may

have to be invoked to describe the process more accurately. The validity of the approach used as a first check on the model, which as we pointed out is based on numerical integration of the normal curve has been checked by a technique suggested by Professor Alan Robertson and is described in the Appendix. This asymmetry in the generation of disequilibrium due to selection was first reported in the empirical studies of Neeley & Rawlings (1971). As we shall show in Chapter VI, this phenomenon arises as a consequence of the skewness of the genotypic distribution brought about by extreme frequencies, and for a given amount of initial genetic variation, it is accentuated, the smaller the number of loci affecting the trait.

We are now in a position to answer the following question: of the total change that takes place in the genotypic variance after the first cycle of selection, what proportion is due to changes in the equilibrium additive variance due to changes of gene frequencies and what proportion is due to the generation of joint disequilibrium. Some results based on numerical integration of the normal curve are shown in Table 3.5. The figures in the table refer to values of $\Delta CLW/\Delta VG$, where, $\Delta VG = \Delta Vg + \Delta CLW$, with $\Delta CLW = CLW^{(1)} - CLW^{(0)}$, $CLW^{(0)}$ having been assumed to be zero.

The initial genotypic variance is taken to be 4 square units, the heritability is 0.4 and Q = 20%.

At intermediate gene frequencies, when changes in the equilibrium additive variance due to gene frequency changes are at a minimum, most of the change in the genotypic variance is due to joint disequilibrium. At more extreme gene frequencies the relative contribution of each term is highly dependent upon the number of loci affecting the trait. With

TABLE 3.5: Values of joint disequilibrium, expressed as a proportion, U, of the total change in the genotypic variance after one cycle of selection. The remaining fraction, 1-U, is due to gene frequency changes.

Number of loci	Initial Gene Frequency		
	0.5	0.7	0.9
10	0.77	0.32	0.08
30	0.92	0.53	0.22
80	0.97	0.68	0.37
500	1.00	0.86	0.65
$\to\infty$	$\to$ 1	$\to$ 1	$\to$ 1

initial gene frequencies at 0.9 the number of loci has indeed to be very large before changes in the genotypic variance can be mostly attributed to joint disequilibrium. If gene frequencies are initially at low values, the effects of both terms are of opposite sign, since the equilibrium additive variance tends to increase as gene frequencies move towards intermediate values. In fact, with a model of 30 loci and initial gene frequencies of 0.3, with the same genetic parameter values as those in Table 3.5, the effects of both terms tend to cancel each other out and the total genotypic variance after one cycle of selection remains virtually unchanged.

To summarize the main points of this chapter, we can say that in large populations the overall change in the genotypic variance, in a

first cycle of selection, starting with a population in equilibrium, is very much dependent on the genetic model used. Provided gene frequencies are not extreme and the number of loci affecting the trait is not small, the largest contribution to the change in the genotypic variance comes from joint disequilibrium. At low initial frequencies, the generation of disequilibrium, and gene frequency changes have opposing effects and consequently the genotypic variance does not alter very substantially, whilst at high initial frequencies both effects act in the same direction towards reducing the genotypic variance. The predictions of changes in variance due to disequilibrium, assuming that the base population is initially in Hardy-Weinberg and linkage equilibrium, are accurate provided gene frequencies are close to intermediate values. They become less so if initial gene frequencies are at more extreme values and this is accentuated as the proportionate effects of the genes increases.

CHAPTER 4

CHANGES OF LINKAGE DISEQUILIBRIUM WITH SEVERAL CYCLES OF SELECTION - A REVIEW

Introduction.

In this section we review the effects of several generations of selection on linkage disequilibrium. We shall generally assume large populations and thus confine ourselves mostly to deterministic models.

Most of the studies of the joint effects of linkage disequilibrium and selection have been carried out by population geneticists in an attempt to understand the factors controlling the observable genetic variability in natural populations. As a consequence, a large amount of work has been concentrated on equilibrium populations and on the effect of recombination on the stability and position of the equilibria under different kinds of multilocus models. Much of this work has been reviewed by Lewontin (1974) and Karlin (1975), and more recently by Hedrick et al. (1978).

Less attention has been given to the effect of directional selection, natural or artificial, on interlocus associations, where alleles increase in frequency towards fixation. Lush stated more than thirty years ago (Lush, 1948) that selection could cause disequilibrium in the gametic array, as it "produces a minor excess of repulsion gametes as compared with what would exist if each gene had the very same frequency but no selection were practiced". According to Lush, this negative disequilibrium generated by directional selection should be very small.

Griffing (1960) investigated the effect of linkage on response to selection in large populations, assuming that gene effects were small enough that second order terms in selective advantage could be ignored. He showed that additive x additive epistasis can generate linkage

disequilibrium but his assumption caused him to ignore the disequilibrium produced by the additively acting genes.

The first derivation of changes of linkage disequilibrium due to directional selection was made by Nei (1963). Studying a two locus model, Nei showed that the value of D generated after one cycle of selection, $D^{(1)}$, starting from a population in equilibrium, was given, using the notation of previous sections by

$$D^{(1)} = f_1 f_4 (W_1 - W_2 - W_3 + W_4)/\bar{W} - \Delta p \Delta q$$

$$= f_1 f_4 \varepsilon - \Delta p \Delta q \qquad (4.1)$$

where ε is a measure of epistasis at the fitness scale and the second term is the product of gene frequency changes. Nei assumed that in the absence of epistasis at the phenotypic scale, ε vanishes and thus, replacing (3.3) in (4.1) he obtained

$$D^{(1)} = -\frac{i^2}{\sigma^2} a_1 p(1-p) a_2 q(1-q) \qquad (4.2)$$

However, we should make clear that additivity at the phenotypic level does not imply additivity at the level of fitness. In fact, it can be shown that perfect additivity in the phenotypic level leads to a value of $\varepsilon = \frac{ix_T}{\sigma^2} a_1 a_2$, to second order terms, where as before a_i is the average effect of a gene substitution at the i^{th} locus. (In the case of a dominance model, $\varepsilon = \frac{ix_T}{\sigma^2} \alpha_1 \alpha_2$, to the same order of approximation, where $\alpha_i = a_i + d_i(1-2p_i)$). Substituting this value for ε in (4.1), we obtain (3.2). Nei concluded that in a large population under selection, provided that ε is zero, the amount of disequilibrium generated by the second order effect of gene frequency changes is small

enough to be ignored.

The departure of gamete frequencies from their equilibrium value can also be measured by the following expression:

$$R = \frac{f_1 f_4}{f_2 f_3} \tag{4.3}$$

which is related to D by : $R = 1 + D/(f_2 f_3)$. Thus when $D = 0$, $R = 1$. Kimura (1965) showed that in a large population if gene frequencies are changing slowly under loose linkage and relatively weak epistatic interaction in fitness, a state is rapidly achieved in which chromosome frequencies change in such a way that R remains practically constant. He called this state quasi-linkage equilibrium. For the properties of this quantity, see Kimura (1965), Feldman and Crow (1970) and Nagylaki (1974).

If a population is initially in linkage equilibrium, for it to remain in equilibrium after selection the genes concerned must affect fitness in a multiplicative manner (i.e. $W_1 W_4 = W_2 W_3$). This can be shown by setting $D^{(1)}$ equal to zero in the following expression:

$$D^{(1)} = (f_1+\Delta f_1)(f_4+\Delta f_4) - (f_2+\Delta f_2)(f_3+\Delta f_3) \tag{4.4}$$

Substituting Δf_i by the set of expressions (2.6) in (4.4) yields $W_1 W_4 = W_2 W_3$. This was pointed out by Felsenstein (1965) who examined the qualitative effects of directional selection on linkage disequilibrium and the effects of linkage on the rate of change of gene frequencies. He showed that directional selection for an additively determined trait will immediately cause negative linkage disequilibrium. This implies an excess of gametes with both the favourable and unfavourable alleles associated thus producing eventually relatively less

extreme genotypes and therefore reducing selection response. Continuous selection produces fresh disequilibrium on each cycle, while recombination tends to break it down. With tight linkage, disequilibrium tends to accumulate and we then expect the rate of selection response to be smaller than with free recombination. In a very large population though, linkage does not affect the selection limit but only the rate of advance to that limit. Linkage disequilibrium eventually disappears when the favourable alleles become fixed.

Several of these expectations have been confirmed by Neeley and Rawlings (1971) who carried out extensive numerical studies on the effect of several cycles of selection on changes of genotypic variance under a strictly additive model. In general they found that the generation of linkage disequilibrium increases with heritability, intensity of selection and tightness of linkage, though linkage has little effect during the early generations of selection.

The most conclusive and complete study of the effect of selection on interlocus associations for a quantitative trait in infinite populations was carried out by Bulmer (1971, 1974, 1976b). Since this thesis relies heavily on Bulmer's work, we shall now review it in some detail. First we deal with the case of free recombination. We relax this assumption in the following section.

Selection Under the Infinitesimal Model.

Free Recombination.

Consider a large population in equilibrium and let P be the phenotypic value of a metric trait determined by the sum of a genotypic value, G, and an independent normally distributed environmental component, E. If we assume that G is given by the sum of an effectively infinite number of additive (non-epistatic) loci, then the phenotypic distribution will be normal. This model was first studied by Fisher (1918) and is usually referred to as the infinitesimal model. We shall now show the consequences of a first cycle of selection in the parental generation on the phenotypic variance in the offspring generation. This can be done in a variety of ways. One such way is to study as we did before, the change in the covariance between allelic effects in gametes due to selection on the phenotype. We have shown before that the covariance between allelic effects in gametes is a component of the total genotypic variance. Under the assumptions of the present model, the joint distribution of allelic effects and phenotypic values is bivariate normal and therefore the regression of one on the other is exactly linear and homoscedastic.

An alternative approach, also using regression theory, is to consider the regression of offspring on parents. Under the present model, Bulmer (1971) showed that the phenotypic values of two or more related individuals follow a multivariate normal distribution. He further proved that this result holds in the presence of linkage provided the related individuals are identical twins and offspring and one or both parents. For other types of relatives, the regression line is unaffected by linkage but the residual variance about the regression line is no longer constant (Bulmer, 1976a).

In the absence of selection, the joint regression of offspring, $P^{(1)}$, on both parents, P_m and P_f, is given by $P^{(1)} = a + bP_m + bP_f + e$, where $b = \frac{1}{2}h^2$ (Falconer, 1960). With random mating, the variance in the offspring, $VP^{(1)}$, which is equal to the variance in the parents, $VP^{(0)}$, is given by:

$$VP^{(1)} = b^2V(P_m + P_f) + V(e),$$

and therefore,

$$V(e) = VP^{(0)}(1 - \tfrac{1}{2}h^4).$$

If directional selection operates in the parental generation, so that the variance amongst the parental phenotypic values changes to $VP^{(0)}\{1 - i(i-x_T)\}$, the regression and the residual variance about the regression line is unaltered and therefore, in the offspring generation, with random mating of the selected parents, the phenotypic variance becomes:

$$VP^{(1)} = VP^{(0)} - \tfrac{1}{2}i(i-x_T)h^{4(0)}VP^{(0)} \qquad (4.5)$$

where $h^{2(0)}$ is the heritability in the base population before selection operated. Since the environmental variance is assumed to be constant, the change in the phenotypic variance is due to the change in the genotypic variance in this first cycle of selection. Therefore, from (4.5) the reduction in the genotypic variance after one cycle of selection is

$$VG^{(1)} = VG^{(0)}(1 - \tfrac{1}{2}i(i-x_T)h^{2(0)}) \qquad (4.6)$$

as (3.22).

We have shown before that this reduction in the genotypic variance is due to the generation of negative linkage disequilibrium. Bulmer (1971) showed that this was the case by studying the regression of

grandchildren on their selected grandparents, assuming that selection was relaxed in the parental generation. Following the algebra through, it is seen that the single cycle of random mating reduces the change in the variance by one half compared with its value immediately following selection. This is the rate at which D breaks down under random mating and free recombination. At the risk of being repetitive, we shall confirm Bulmer's result using an analysis of variance model, to stress the analogy between the two models.

Consider a full-sib family structure, such that $P^{(2)} = \mu + F + e$, where $P^{(2)}$ is the phenotypic value of an individual whose grandparents at generation 0 had been selected but whose parents had been chosen and mated at random. F, is the family effect, such that

$$VP^{(2)} = V(F) + V(e),$$

where V(e) is now the pooled variance within family means. V(F), the variance component between full-sib family means, estimates one half of the genotypic variance in the parental generation, $(\frac{1}{2} VG^{(1)})$ which we have shown is equal to (4.6). The variance within families, as we shall subsequently show, is unaltered and thus is an estimate of $\frac{1}{2} VG^{(0)} + VE$, where VE is the environmental variance. Putting all this together it is easily shown that

$$VG^{(2)} = VG^{(0)}(1 - \tfrac{1}{4}i(i-x_T)h^{2(0)}),$$

the reduction in variance having been halved after one cycle of random mating. In this way we confirm that in the infinitesimal model the reduction in variance due to directional selection is temporary and with free recombination, on relaxation of selection, the variance quickly reverts to its original value. There are no permanent changes

in variance due to gene frequency changes because the model assumes an infinite number of loci. We then have an expression which predicts changes of genetic variance using estimable parameters of the base population.

We must now describe the process when repeated cycles of selection are carried out. Since the genotypic values at different loci are now correlated due to the generation of joint disequilibrium, the assumption of linearity of regression of allelic effects on the phenotype and constant variance about the regression line may not strictly hold. However these assumptions will hold approximately, when the correlations between loci are small and each locus contributes a small part of the total phenotypic variation. These requirements are in line with the infinitesimal model. Consider a second cycle of selection. The joint disequilibrium in the offspring at generation 2, $CLW^{(2)}$, can be described by two components. The first one is due to the fresh disequilibrium generated in this second cycle of selection. The second component is attributable to the fact that with free recombination, half of the disequilibrium present in the offspring at generation one is preserved in the offspring at generation two. Therefore we can write

$$VG^{(2)} = VG^{(0)} + CLW^{(2)}$$

$$= VG^{(0)} - \tfrac{1}{2}i(i-x_T)VG^{(1)}h^{2(1)} + \tfrac{1}{2}CLW^{(1)},$$

and in general, at generation t + 1,

$$VG^{(t+1)} = VG^{(0)} + CLW^{(t+1)},$$

where

$$CLW^{(t+1)} = -\tfrac{1}{2}i(i-x_T)VG^{(t)}h^{2(t)} + \tfrac{1}{2}CLW^{(t)} \qquad (4.7)$$

This recurrence relationship allows us to calculate the changes in the genotypic variance in successive generations of selection. Since under directional selection, the first term in (4.7) tends to decrease in successive cycles of selection and the second term increases due to recombination, a limiting value is arrived at which can be evaluated by putting $CLW^{(t+1)} = CLW^{(t)} = CLW^*$ in (4.7). This leads to a quadratic equation which can be solved in terms of parameters of the base population, namely, $VG^{(0)}$ and $h^{2(0)}$. With an initial heritability of 50% and a proportion selected of 20%, the reduction in the genotypic variance in the first cycle of selection is of about 20%. This leads to a reduction of the observed response at generation two of 15% relative to the response predicted on the assumption of no changes of parameters due to selection.

The limiting value is achieved after about four cycles of selection and at that point, the final reduction in the genotypic variance is of the order of 25% of its original value. This shows that most of the decline in variance takes place after the first cycle of selection and that the steady state is arrived at fairly soon in the selection process. If selection is relaxed and random mating restored, the genotypic variance will soon revert to its original value.

The Presence of Linkage

It has been stated that in any generation, the joint disequilibrium, before selection, comprises two terms. The first term is due to the fresh disequilibrium generated in the parental generation, and the

second term is the proportion of the disequilibrium present in the previous offspring generation which recombination did not break down. Bulmer (1971) showed that the fresh disequilibrium is independent of linkage; the second term though is clearly dependent on the degree of linkage between the loci involved. In other words, the contribution of a pair of loci to the disequilibrium in the following generation is positively correlated with the degree of linkage between them (Bulmer, 1974). Consider a trait determined by n loci of equal effects, where n is large. Let c be the recombination fraction between a particular pair of such loci and let $\delta_{(t)}$ be the contribution from this pair to the total joint disequilibrium, $CLW^{(t)}$, in the t^{th} cycle of selection. Since all n loci have the same effect and the fresh disequilibrium produced at generation t + 1 is independent of linkage, it follows that

$$\delta_{(t+1)} = (-\tfrac{1}{2}i(i-x_T)VG^{(t)}h^{2(t)})/\tfrac{1}{2}n(n-1) + (1-c)\delta_{(t)} \quad ,$$

since there are $\frac{1}{2}n(n-1)$ pairs of loci. The limiting value of $\delta_{(t)}$, δ^*, is evaluated by putting $\delta_{(t+1)} = \delta_{(t)}$ in the above expression. Summing over pairs of loci, we obtain the limiting value of the total disequilibrium, CLW*, which is given by

$$CLW^* = -\tfrac{1}{2}i(i-x_T)h^{4*}VG^*/H \quad ,$$

where h^{2*} and VG* are the limiting values of the heritability and genotypic variance respectively and H is the harmonic mean of the recombination fractions (Bulmer, 1974). For given n, the value of H depends on the total number of chromosomes. When the number of chromosomes is large, H tends to $\frac{1}{2}$ and the system behaves as in the case of no linkage.

However, in an organism like Drosophila, Bulmer shows that H is around 0.1. Assuming an initial heritability of 50% and a proportion selected of 20%, the final reduction in variance is of about 50% of its original value, compared to the value of 25% obtained with free recombination. Hence with tight linkage the reduction in variance due to joint disequilibrium is larger and the rate of approach to the limiting value is slower than with free recombination.

As stressed by Bulmer, this theory is to be considered as a limiting result which will hold provided the number of loci is strictly infinite. With finite number of loci gene frequency changes cannot be ignored, but one can argue that for a given amount of initial genetic variation, as the number of loci increases gene frequency changes become progressively smaller and most of the change in variance during the early generations could be attributed to the generation of joint disequilibrium. A simulation study of the effects of different modes of selection on genetic variability was reported as a first check on the theory (Bulmer, 1976b). Three different types of selection were studied, namely, stabilizing, disruptive and directional selection. The metric character studied was assumed to be determined by twelve additive loci, with no dominance or epistasis. All twelve loci had equal proportionate effects on the character ($a/\sigma = 1$). Two alternative sets of simulations were undertaken. In one of them, called the mouse simulations, the twelve loci were assumed to be on different chromosomes and therefore to segregate independently. In the other set of simulations called the Drosophila simulations, the twelve loci were distributed in groups of four on three chromosomes, the recombination fraction between adjacent loci being taken as 0.1 in females

and zero in males. In most simulations, 100 individuals of each sex were selected out of a total of 500. The heritability in the base population was about 57%. As expected from theory, stabilizing selection generated negative joint disequilibrium whilst disruptive selection generated strong joint disequilibrium of the opposite sign. The limiting values of disequilibrium were in good agreement with theory. In the case of directional selection gene frequencies went to fixation very rapidly (13 and 16 generations in the mouse and Drosophila simulations respectively). This is not surprising in view of the large selection pressure at each locus. When averaged over the first five generations of selection, 62% of the total reduction in the genotypic variance was due to gene frequency changes, in both sets of simulations. Bulmer did not present results of the changes in the genotypic variance during the first few generations of selection. It is during this early stage when the effect of disequilibrium should be relatively more important as a cause of change in the genotypic variance, especially in this study,where initial gene frequencies were such that the change in the equilibrium additive variance due to gene frequency changes was minimised.

Robertson (1977b) has recently studied the response to selection in small populations using an additive model with an effectively infinite number of loci. The finiteness of the population introduces a new variable into the problem with a considerable increase in the level of complexity. The genetic variance within lines declines not only due to linkage disequilibrium but also due to genetic drift, this latter effect being accentuated by selection through an increase in the variance of family size (Robertson, 1961). Bulmer's expressions are

then likely to underestimate the decline in variance in finite populations. Robertson shows that small population size and degree of linkage interact in such a way that the limiting value of disequilibrium is only achieved in the case of free recombination or loose linkage. With tighter linkage the genetic variance declines consistently as selection proceeds, and with complete linkage the limiting value is zero. In agreement with the numerical results of the deterministic models used by Neeley and Rawlings (1971), Robertson shows that the degree of linkage has little effect on selection response during the early generations.

The implications of Bulmer's work are clearly of theoretical and practical importance. From a practical point of view, knowledge of changes of genetic parameters is essential for optimum implementation of breeding plans. These results will feature in such problems as consequences of selection on the estimation of genetic parameters and in the comparison between different kinds of selection schemes. Work on these lines has already been reported by Robertson (1977a) and Fimland (1979).

In the forthcoming chapters we shall investigate the validity of the results based on the infinitesimal model relaxing some of the assumptions on which it is based, in particular, we want to study the effect of a finite number of loci, with associated gene frequency changes. An understanding of the robustness of the model is essential before it can be applied with any benefit in the evaluation of alternative breeding programmes.

CHAPTER 5

EFFECT OF DIRECTIONAL SELECTION ON QUANTITATIVE ADDITIVE MODELS - SEVERAL CYCLES OF SELECTION

Introduction.

In this chapter we examine the effects of several cycles of selection on changes of the genotypic variance with a strictly additive model. We shall assume that the population is large enough that random drift can be ignored and therefore in the Montecarlo simulations that follow, we focus our work on short term selection response. This is in contrast with most computer simulations reported in the literature where attention was generally concentrated on the effects of small population size and degree of linkage on selection limits (i.e. Martin & Cockerham, 1960; Hill & Robertson, 1966; Robertson, 1970a).

Nei (1963) studied the effect of selection on changes in the components of the genotypic variance, ignoring the effects of linkage disequilibrium. He worked with a variety of models, including different kinds of epistasis, and the changes were strictly due to changes of gene frequencies. It was assumed in Nei's work that gene effects were small enough that second order terms in selective advantage could be ignored. Initial gene frequencies were taken as 0.5 and a/σ assumed to be 0.1 and to remain constant throughout the selection programme. Selection was followed by one generation of random mating. Nei showed that the additive variance was the component most sensitive to gene frequency changes, while the dominance x dominance component was the least affected. In general, the genetic components of variance associated with additive effects changed more rapidly than those associated with dominance.

Latter (1965) investigated the effects of genes of large proportionate effect on the expected response to selection using a single locus additive model. He showed that the expected response to selection in early generations is poorly estimated if gene effects are large and that substantial asymmetry can develop in two way selection experiments. Due to the nature of his model, effects of linkage disequilibrium on gene frequency changes were ignored.

Young (1966, 1967) examined the changes in genetic variances and heritability through Montecarlo simulations, under different genetic models. He used very large population sizes and selection was carried on for 30 generations. No theoretical predictions were made and the conclusions were basically drawn empirically from the simulation results, which indicated that the additive component of variance changed more than the other components as selection proceeded, in agreement with Nei's results. The predictions of selection response over the early generations of selection based on parameters of the base population were reasonably accurate under the strictly additive model. In general, the presence of dominance or different types of epistasis made the predictions of early response less accurate.

Wright (1977) illustrates the course of change in the genotypic variance and its components with various genetic models, assuming a heritability of 1. The different components of the variance are graphed for different gene frequencies and therefore they should be regarded as what we called, following Bulmer (1971), the equilibrium value for the component in question. The point we want to make is

that the response to selection in a particular generation depends not only on the variance of individual genes (the equilibrium component), but also on the covariance between them and therefore, for the case of a completely additive model, this disequilibrium component must be included in the description of the genotypic variance, if the latter is to reflect the response to selection at the generation in question.

We now proceed to study the theoretical consequences of directional selection on the total genotypic variance. We first develop the theory for a two locus model and then we extend it for an arbitrary number of loci.

Two Locus Models.

Changes of Gene Frequencies due to Directional Selection.

The genetic model we shall use for the two locus case has been defined before in Chapter 3 and it is reproduced below:

	AA	Aa	aa	BB	Bb	bb
Value	a_1	0	$-a_1$	a_2	0	$-a_2$
Frequency	p^2	$2p(1-p)$	$(1-p)^2$	q^2	$2q(1-q)$	$(1-q)^2$

Population Mean : $M = -a_1(1-2p) - a_2(1-2q)$

The problem of evaluating the change in gene frequency in a two locu model is easily approached by obtaining explicit expressions for the change in the frequency of the different types of gametes before (or after) recombination takes place, using the set of expressions (2.6). The relative selective advantage of the ij^{th}

genotype is given, as a second order approximation by, (Latter, 1965)

$$\frac{W_{ij}}{\bar{W}} = (1 + \frac{i}{\sigma}(\bar{X}_{ij}-M) + \frac{ix_T}{2\sigma^2}(\bar{X}_{ij}-M)^2)\frac{Q}{\bar{W}} \qquad (5.1)$$

The relative selective advantage of the $i^{\underline{th}}$ gametic phase is defined as follows,

$$\frac{W_i}{\bar{W}} = \frac{1}{\bar{W}} \sum_j W_{ij} f_j \quad , \qquad i=1,\ldots,4; \quad j=1,\ldots,4. \qquad (5.2)$$

Replacing (5.1) in (5.2) it is shown that

$$\frac{W_i}{\bar{W}} = \frac{Q}{\bar{W}}\{1 + \frac{i}{\sigma}(\bar{X}_i-M) + \frac{ix_T}{2\sigma^2}(Vw_i + (\bar{X}_i-M)^2)\} \qquad (5.3)$$

where $\bar{X}_i$ is the mean of the $i^{\underline{th}}$ gametic phase and Vw_i is the variance within the $i^{\underline{th}}$ gametic phase, given by

$$Vw_i = \sum_j (\bar{X}_{ij}-\bar{X}_i)^2 f_j \qquad (5.4)$$

Under a strictly additive model, it can be shown that $Vw_i = \frac{1}{2}VG$ for all i, where VG is the total genotypic variance defined previously (3.1). The term Vw_i appears in the second order term in (5.3) due to the fact that selection operates at the genotypic level. Vw_i does not feature in the expression for the second order approximation of the relative selective advantage of the different gametic phases if it is assumed that selection operates at the gametic level. Similarly, it can be shown that,

$$\bar{W} = Q + \frac{Qix_T}{2\sigma^2}\left[\sum_i Vw_i f_i + \sum_i (\bar{X}_i - M)^2 f_i\right] \qquad (5.6)$$

where the second term in square brackets in (5.6) is the variance between marginal means of gametic phases. Both terms in square brackets add up to the total genotypic variance contributed by the pair of loci. Hence,

$$\bar{W} = Q + \frac{Qix_T}{2\sigma^2} VG \qquad (5.7)$$

Substituting (5.3) and (5.7) in (2.9) we obtain the following set of expressions for the change in chromosome frequencies before recombination takes place, $f_i^{(s,t)}$, in terms of parameters before selection:

$$\Delta f_1^{(t)} = f_1^{(t)}\{\frac{i}{\sigma}(a_1(1-p) + a_2(1-q)) + \frac{ix_T}{2\sigma^2}(a_1^2(1-p)(1-2p) + a_2^2(1-q)(1-2q) + 2a_1a_2((1-p)(1-q) - D))\}^{(t)}$$

$$\Delta f_2^{(t)} = f_2^{(t)}\{\frac{i}{\sigma}(a_1(1-p) - a_2q) + \frac{ix_T}{2\sigma^2}(a_1^2(1-p)(1-2p) - a_2^2q(1-2q) - 2a_1a_2((1-p)q + D))\}^{(t)}$$

$$\Delta f_3^{(t)} = f_3^{(t)}\{\frac{i}{\sigma}(-a_1p + a_2(1-q)) + \frac{ix_T}{2\sigma^2}(-a_1^2(1-p)(1-2p) + a_2^2(1-q)(1-2q) - 2a_1a_2(p(1-q) + D))\}^{(t)}$$

$$\Delta f_4^{(t)} = f_4^{(t)}\{\frac{i}{\sigma}(-a_1p - a_2q) + \frac{ix_T}{2\sigma^2}(-a_1^2p(1-2p) - a_2^2q(1-2q) + 2a_1a_2(pq-D))\}^{(t)} \tag{5.8}$$

The change in frequency of allele A, Δp, is then approximated by $\Delta p = \Delta f_1 + \Delta f_2$,

$$\Delta p = \frac{i}{\sigma}(a_1p(1-p) + a_2D) + \frac{ix_T}{2\sigma^2}(a_1^2p(1-p)(1-2p) + a_2^2(1-2q)D + 2a_1a_2(1-2p)D) \tag{5.9}$$

Similarly, $\Delta q = \Delta f_1 + \Delta f_3$,

$$\Delta q = \frac{i}{\sigma}(a_2q(1-q) + a_1D) + \frac{ix_T}{2\sigma^2}(a_2^2q(1-q)(1-2q) + a_1^2(1-2p)D + 2a_1a_2(1-2q)D) \tag{5.10}$$

Expression (5.9) reduces to (3.14) obtained by Latter (1965), when the initial population is in linkage equilibrium. Selection immediately causes negative linkage disequilibrium and therefore when more than one locus is considered, D cannot be ignored after a first cycle of selection. It is clear from the above expressions that the change in gene frequency of an allele at a particular locus is due to direct selective pressure on the locus itself, and due to pressures arising from correlations with alleles at other loci.

Consider now the expected response to selection, R, from loci

A and B. Noting that $\delta M/\delta p = 2a_1$ and that $\delta M/\delta q = 2a_2$, the response due to locus A, $R_{(A)}$ and due to locus B, $R_{(B)}$ is, ignoring the second order term in (5.9) and (5.10),

$$R_{(A)} = \frac{\delta M}{\delta p}\Delta p = \frac{i}{\sigma}(2a_1^2 p(1-p) + 2a_1a_2D),$$

and similarly,

$$R_{(B)} = \frac{\delta M}{\delta q}\Delta q = \frac{i}{\sigma}(2a_2^2 q(1-q) + 2a_1a_2D).$$

Therefore the expected response due to changes in gene frequency at both loci is:

$$R = R_{(A)} + R_{(B)} = \frac{i}{\sigma}(2a_1^2 p(1-p) + 2a_2^2 q(1-q) + 4a_1a_2D) \qquad (5.11)$$

This expression is easily generalised to an arbitrary number of loci,

$$R = \frac{i}{\sigma}(\Sigma_j 2a_j^2 p_j(1-p_j) + 4\Sigma_{i<j}\Sigma a_i a_j D_{ij}) = ih^2\sigma \qquad (5.12)$$

as before (3.5). The important point we want to stress is that the joint disequilibrium generated by selection is to be regarded as part of the expected selection response or expected realised heritability. Furthermore, we can extend this argument to the case of offspring-parent regressions as estimators of heritabilities in a given population at a given time. Let ΔMo and ΔMp be the deviations of the means of the offspring and parents from the population mean M. Then, as is well known,

$$\Delta Mo = b_{O\bar{P}}\,\Delta M_P ,$$

where $b_{O\bar{P}}$ is the regression of offspring on mid-parent. Assuming normality, ΔMp is expressed as $i\sigma$, and Mo is the expected response to selection as defined above. It then follows that,

$$b_{O\bar{P}} = \frac{\Delta Mo}{\Delta Mp} = (\Sigma_j 2a_j^2 p_j(1-p_j) + 4\Sigma_{i<j} \Sigma a_i a_j D_{ij})/\sigma^2 \qquad (5.13)$$

In other words, the regression of offspring on mid-parent provides us with an unbiased means of estimating the heritability at a particular generation and therefore reflects accurately the genotypic variance available for selection response at that generation. It should be clear though, that the above argument assumes linearity of regression.

The Generation of Disequilibria with Selection.

We have pointed out that directional selection leads to negative covariances between loci within gametes and following Bulmer we have called their effect on the genotypic variance, joint disequilibrium. The purpose of this section is to show that amongst selected genotypes there are covariances both between and within gametes, the former disappearing in the offspring generation provided mating of the selected individuals is at random. As we shall subsequently show, the expressions to be derived are relevant both from a theoretical and from a practical point of view. Theoretically, it is believed that this approach leads to a clear understanding of the dynamics of the selection process. From a practical viewpoint it will be shown in later sections that these expressions feature in some methods

commonly used in animal breeding practice to estimate genetic parameters.

We first give a semi-intuitive explanation of the theory that follows. Consider a trait determined by n additive loci where $a_i^m(a_i^p)$ is the average effect of the $m^{\underline{th}}$ allele at the $i^{\underline{th}}$ locus from the maternally (paternally) derived chromosome. The genotypic value of an individual, G, can then be described by the following model,

$$G = M + \sum_{i=1}^{n} (a_i^m + a_i^p) \qquad (5.14)$$

where M is the population mean. It then follows that the variance of G is,

$$\begin{aligned} VG = {} & \sum_i (V(a_i^m) + V(a_i^p)) + 2\sum_i \mathrm{cov}(a_i^m, a_i^p) \\ & + 2 \sum_{i<j} (\mathrm{cov}(a_i^m, a_j^p) + \mathrm{cov}(a_i^p, a_j^m)) \\ & + 2 \sum_{i<j} (\mathrm{cov}(a_i^m, a_j^m) + \mathrm{cov}(a_i^p, a_j^p)) \end{aligned} \qquad (5.15)$$

There are four different kinds of terms in (5.15). The first term is the variance of alleles acting singly. We have called it the equilibrium additive variance, Vg. The second term is a covariance of allelic effects within loci between chromosomes reflecting departures from Hardy-Weinberg equilibrium and following Bulmer (1976b) we symbolise it CHW.

The second and third terms are covariances between allelic

effects at different loci, between and within chromosomes (or more generally between and within gametic contributions) respectively. We use the symbols, CLB and CLW, for this joint disequilibria between and within parental contributions respectively. Summarising, we can express (5.15) as follows:

$$VG = Vg + CHW + CLW + CLB \qquad (5.16)$$

From this account we draw the following conclusion:

If mating is strictly at random and an effectively infinite number of offspring are produced, CHW and CLB become zero in the offspring generation since by definition there are no associations between chromosomes. Therefore, in contrast to the case of CLW, values of CHW and CLB do not accumulate as selection proceeds.

We shall now study the effect of selection on these different types of covariances, before and after recombination takes place. Initially the algebra is developed for a two locus model. The results are then extended to accommodate an arbitrary number of loci.

Covariances Between and Within Gametic Contributions Amongst Selected Genotypes.

Covariance Between Loci Within Gametes ($CLW^{(s,t)}$).

The approach we follow is equivalent to the one we used in the derivation of (3.20). In this case we assume that the population we select from is initially in Hardy-Weinberg equilibrium but not necessarily in linkage equilibrium. The parents in each generation are then taken to mate strictly at random. The existence of linkage

disequilibrium in the generation prior to selection leads inevitably to more cumbersome algebra and less neat results. The final expressions, however, are amenable to clear interpretation. Normality is assumed throughout the derivation.

The covariance of gene frequencies within parental contributions in individuals of the $t^{\underline{th}}$ cycle of selection, prior to recombination, is defined, as shown in Chapter III:

$$D^{(s,t)} = (f_1 f_4 - f_2 f_3)^{(s,t)},$$

where,

$$f_i^{(s,t)} = \frac{f_i}{\bar{W}}(W_i - \bar{W})^{(t)} \qquad (i=1,\ldots,4).$$

Therefore we can write,

$$\begin{aligned} D^{(s,t)} &= \left[f_1 + \frac{f_1}{\bar{W}}(W_1-\bar{W}) \right]^{(t)} \left[f_4 + \frac{f_4}{\bar{W}}(W_4-\bar{W}) \right]^{(t)} \\ &\quad - \left[f_2 + \frac{f_2}{\bar{W}}(W_2-\bar{W}) \right]^{(t)} \left[f_3 + \frac{f_3}{\bar{W}}(W_3-\bar{W}) \right]^{(t)} \\ &= D^{(t)} + \frac{1}{\bar{W}}\{ f_1 f_4 ((W_1-\bar{W}) + (W_4-\bar{W})) - f_2 f_3 ((W_2-\bar{W}) \\ &\quad + (W_3-\bar{W})) \}^{(t)} + \frac{1}{\bar{W}^2}\{ f_1 f_4 (W_1-\bar{W})(W_4-\bar{W}) - f_2 f_3 (W_2-\bar{W})(W_3-\bar{W}) \}^{(t)} \end{aligned}$$

(5.17)

Substituting (5.3) and (5.7) in (5.17) and letting $Q/\bar{W}$ equal to $Q - \frac{ix_T}{2\sigma^2}(2a_1^2 p(1-p) + 2a_2^2 q(1-q) + 4a_1 a_2 D)$ to second order terms, we obtain:

$$D^{(s,t)} = D^{(t)} + \frac{i}{\sigma}(a_1(1-2p)D + a_2(1-2q)D)^{(t)} - \left(\frac{i(i-x_T}{\sigma^2}a_1p(1-p)\right.$$

$$\left.a_2q(1-q)\right)^{(t)} + \frac{ix_T}{2\sigma^2}((a_1(1-2p) + a_2(1-2q))^2D)^{(t)}$$

$$- \frac{i^2}{\sigma^2}(a_1^2p(1-p)D + a_2^2q(1-q)D)^{(t)} - \left(\frac{i(i+x_T)}{\sigma^2}a_1a_2D^2\right)^{(t)} \qquad (5.18)$$

In expression (5.18) we can identify two components. The first component, $D^{(t)}$, is the disequilibrium present in the offspring before selection operated. The second component, (the five terms following $D^{(t)}$) is the fresh disequilibrium generated in the $t^{\underline{th}}$ selection cycle, and we shall refer to it as $D_f^{(t)}$. Notice that $D_f^{(t)}$ is not independent of $D^{(t)}$. If the population is initially in equilibrium such that $D^{(o)} = 0$, then (5.18) reduces to (3.20). We can then write:

$$D^{(s,t)} = D^{(t)} + D_f^{(t)} \qquad (5.19)$$

and

$$CLW^{(s,t)} = 4a_1a_2D^{(s,t)} \qquad (5.20)$$

We should further point out that (5.20) shows that if $D^{(t)} \neq 0$, $D^{(s,t)}$ is not symmetrical unless gene frequencies are exactly intermediate. Selection in both directions from a population in linkage disequilibrium will lead to different values of $D_f^{(t)}$ in each direction if $p \neq 0.5$.

Covariance Between Loci Between Gametes ($CLB^{(s,t)}$)

We now investigate the generation of disequilibrium between loci on different parental contributions in genotypes of the $t^{\underline{th}}$ cycle of selection. Writing it in terms of associations of gene frequencies, we define this covariance as follows:

$$D_B^{(s,t)} = \{E(X_p-p)^{(m)}(X_q-q)^{(p)}\}^{(s,t)} \tag{5.21}$$

where X_p and X_q are the number (i.e. 0 or 1) of A and B alleles in the maternally and paternally derived chromosomes at locus A and B respectively, and p and q are their expected frequencies. Since the expected gene frequencies are the same in both sexes, the covariance of allelic effects from both chromosomes is:

$$CLB^{(s,t)} = 4a_1a_2D_B^{(s,t)}$$

The frequency of allele combination AB amongst selected individuals is seen to be (dropping superscript t):

$$f_{(AB)} = \frac{1}{\bar{W}}\{f_1(f_1W_{11} + f_3W_{13}) + f_2(f_1W_{12} + f_3W_{23})\}.$$

Following the algebra through, $D_B^{(s,t)}$ reduces to

$$D_B^{(s,t)} = D^{(s,t)} - \{\frac{1}{\bar{W}}(f_1f_4\bar{W}_{14} - f_2f_3\bar{W}_{23})\}^{(t)} \tag{5.22}$$

Assuming $W_{14} = W_{23}$, (5.22) reduces to:

$$D_B^{(s,t)} = D^{(s,t)} - \frac{W_{14}}{\bar{W}} D^{(t)}$$

$$= D^{(t)}(1 - \frac{W_{14}}{\bar{W}}) + D_f^{(t)} \tag{5.23}$$

Let us examine expression (5.23)

Under the present model the relative selective advantage of the coupling heterozygote, is given by (as a second order approximation):

$$\frac{W_{14}}{\bar{W}} = \frac{Q}{\bar{W}}\{1 + \frac{i}{\sigma}(a_1(1-2p)+a_2(1-2q)) + \frac{ix_T}{2\sigma^2}(a_1(1-2p) + a_2(1-2q))^2\} \qquad (5.24)$$

Letting $Q/\bar{W} \simeq 1 - \frac{ix_T}{2\sigma^2}(2a_1^2p(1-p) + 2a_2^2q(1-q) + 4a_1a_2D)$, it is seen that at intermediate gene frequencies, W_{14}/W is close to 1. At extreme gene frequencies, the term i/σ can be important though it will become less so as gene effects become smaller. Hence as the number of loci affecting the trait increases, the disequilibrium between loci between parental gametes will tend to become closer in value to the fresh disequilibrium within parental gametes. $D_B^{(s,t)}$, as was mentioned before, vanishes in the offspring generation (assuming large populations and random mating) and is created anew on each cycle of selection.

It is possible to get an explicit expression for (5.23) by approximating its first term and using (5.18). Carrying out the algebra it can be shown,

$$D_B^{(s,t)} = \frac{-i(i-x_T)}{\sigma^2}(a_1^2p(1-p)D + a_2^2q(1-q)D + a_1p(1-p)a_2q(1-q) + a_1a_2D^2) \qquad (5.25)$$

As in the case of the fresh disequilibrium in (5.18), the smallest order term in the above expression is $a_1p(1-p)a_2q(1-q)$. The difference between the above expression and $D_f^{(t)}$ is of order D, that is, $D_f^{(t)} - D_B^{(s,t)}$:

$$-D^{(t)}\Big[\Big(1 + \frac{i}{\sigma}\, a_1(1-2p) + a_2(1-2q)\Big) + \frac{ix_r}{2\sigma^2}\Big(a_1^2\{(1-2p)^2 - 2p(1-p)\}$$

$$+ a_2^2\{(1-2q)^2 - 2q(1-q)\} + 2a_1a_2\{(1-2p)(1-2q) - 2D\}\Big)\Big]$$

$$= -D^{(t)}\Big(1 - \frac{W_{14}}{\bar{W}}\Big).$$

If gene frequencies are low so that $W_{14}/\bar{W} > 1$, $D_f - D_B$ is <0. At high gene frequencies, $W_{14}/\bar{W} < 1$ and therefore $D_f - D_B$ is >0. Clearly, at t=0, if $D^{(0)} = 0$, $D_f^{(1)} = D_B^{(s,o)}$.

As the number of loci affecting the trait increases, the difference between both covariances will become smaller.

Covariance of allelic effects within loci between chromosomes (CHW).

This covariance due to Hardy-Weinberg departures, following expression (5.15) is defined,

$$CHW = 2\sum_i E(a_i^m \, a_i^p), \qquad (5.26)$$

where a_i^m and a_i^p are average effects of the m^{th} and p^{th} allele at the i^{th} locus. Consider our two locus model. Let D_{HW} be the departure of genotype frequencies from Hardy-Weinberg proportions caused by selection (which can be regarded as a covariance of gene frequencies within loci). Assuming two alleles per locus (m = p = 1,2) we can write:

$$f_{(AA)} - p^2_{(s,t)} = D_{HW}^{(s,t)} \qquad (5.27a)$$

$$f_{(Aa)} - 2p_{(s,t)}(1-p_{(s,t)}) = -2D_{HW}^{(s,t)} \qquad (5.27b)$$

$$f_{(aa)} - (1-p_{(s,t)})^2 = D_{HW}^{(s,t)} \qquad (5.27c)$$

where $p_{(s,t)}$ is the frequency of the plus allele at locus A and $f_{(\cdot)}$ are the genotypic frequencies. It then follows from (5.26) (dropping subscripts):

$$E(a^m, a^p) = a_1^2(p^2 + D_{HW}) + 2a_1a_2(p(1-p)-D_{HW}) + a_2^2((1-p)^2 + D_{HW})$$

$$= (a_1-a_2)^2 D_{HW}$$

$$= a^2 D_{HW} ,$$

where a is the average effect of the gene substitution at the locus. Hence, from (5.26), $CHW = 2a^2 D_{HW}$. From (5.27a) we can then write:

$$D_{HW}^{(s,t)} = \left[f_1^2 \frac{W_{11}}{\bar{W}} - f_1^{2(s,t)} \right] + \left[2f_1 f_2 \frac{W_{12}}{\bar{W}} - 2f_1^{(s,t)} f_2^{(s,t)} \right]$$

$$+ \left[f_2^2 \frac{W_{22}}{\bar{W}} - f_2^{2(s,t)} \right] \qquad (5.28)$$

In each term in square brackets, the first term reflects the value of the parameter before selection. Following the algebra through, it can be shown that the second order approximation of (5.28) is given by the following expression,

$$D_{HW}^{(s,t)} = \frac{-i(i-x_T)}{\sigma^2} \left(a_1^2 p^2 (1-p)^2 + 2a_1a_2p(1-p)D + a_2^2 D \right) \qquad (5.29)$$

Expression (5.29) tells us that the covariance of gene frequencies within the locus is always negative and depends on the covariance between it and the other locus. This is probably not surprising since we know that the change in gene frequency at a particular locus is

influenced by the existing disequilibrium with the second locus. The effect of the covariance of gene frequency within locus A, $CHW_{(A)}$ on the total genotypic variance is, from (5.29):

$$CHW_{(A)} = \frac{-i(i-x_T)}{\sigma^2}\left(2a_1^4p^2(1-p)^2 + 2a_1^2p(1-p)2a_1a_2D + 2a_2^2D^2\right)$$

For this two locus model, if gene effects and frequencies are the same at both loci, when $D^{(o)} = 0$, the three different disequilibria generated in the first cycle of selection are the same. With equality of effects and frequencies at both loci, $D_B^{(s,t)}$ and $D_{HW}^{(s,t)}$ take similar values for all t. With n loci, however, the leading term in (5.29) is of order n whereas the corresponding term in (5.25) is of order n^2, and therefore as n increases the effect of CHW on VG becomes small, relative to that coming from CLW and CLB.

Covariance Within Gametic Contributions after Recombination.

As was mentioned before, in an infinite population, under a strictly additive model, provided mating takes place at random, the variance amongst gametic values is equal to half the genotypic variance in the offspring generation. The consequence of random mating is that chromosomal values are not correlated in any way and therefore both CLB and CHW vanish in the expression of the genotypic variance which is equal to twice the variance between gametic values. We shall now show that the covariance of allelic effects within gametes at generation t+1 (after recombination) comprises two terms. The first term is a fraction approximately (1-c) of the disequilibrium present at generation t, before selection operated. The second term

is due to what we have called fresh disequilibrium generated at the $(t+1)^{\underline{th}}$ cycle of selection. The covariance of gene frequencies in gametes is,

$$D^{(t+1)} = (f_1 f_4 - f_2 f_3)^{(t+1)}.$$

Recalling the set of expressions (2.8), this can be written:

$$D^{(t+1)} = D^{(t)}(1 - \frac{W_{14}}{\bar{W}} c) + \frac{1}{\bar{W}}\{f_1 f_4((W_1-\bar{W}) + (W_4-\bar{W})) - f_2 f_3((W_2-\bar{W}) + (W_3-\bar{W}))\}^{(t)} + \frac{1}{\bar{W}^2}\{f_1 f_4 (W_1-\bar{W})(W_4-\bar{W}) - f_2 f_3 (W_2-\bar{W})(W_3-\bar{W})\}^{(t)} - cD^{(t)} \frac{W_{14}}{\bar{W}^2}\{f_1(W_1-\bar{W}) + f_2(W_2-\bar{W}) + f_3(W_3-\bar{W}) + f_4(W_4-\bar{W})\}^{(t)} \quad (5.30)$$

The last term in (5.30) clearly vanishes since it involves the expected deviation of the marginal gametic fitnesses from the mean population fitness and therefore, from (5.19),

$$D^{(t+1)} = D^{(t)}(1 - \frac{W_{14}}{\bar{W}} c) + D_f^{(t)} \quad (5.31)$$

It may be helpful to summarise at this stage the results of this section. Starting at t=0, with $D^{(o)} = 0$, selection causes three different types of covariances of gene frequencies between parts of the genome: $D^{(s,o)}$, $D_B^{(s,o)}$, $D_{HW}^{(s,o)}$.

The total reduction in the genotypic variance amongst selected individuals due to these negative correlations is:

$$CL_T^{(s,o)} = CLW^{(s,o)} + CLB^{(s,o)} + CHW^{(s,o)}.$$

After recombination, gametes pair at random and therefore, in an

infinite population, the total reduction in the offspring generation is: $CL_T^{(1)} = CLW^{(s,o)} = CLW^{(1)}$. Since at t=0 there is an equal number of coupling and repulsion heterozygotes, the degree of linkage has no effect on $CL_T^{(1)}$. After t cycles of selection:

$$CL_T^{(s,t)} = \left(CLW^{(t)} + CLW_f^{(t)}\right) + CLB^{(s,t)} + CHW^{(s,t)},$$

where $CLW_f^{(t)}$ is the fresh disequilibrium generated at the $t^{\underline{th}}$ selection cycle. The term in square brackets is the joint disequilibrium within chromosomes in selected individuals, $CLW^{(s,t)}$. After recombination and random mating, if we assume $\frac{W_{14}}{\bar{W}} \simeq 1$, $CLW^{(t+1)}$ will comprise approximately a proportion 1-c of $CLW^{(s,t)}$ and a proportion c of $CLB^{(s,t)}$ and therefore:

$$CLW^{(t+1)} \simeq (1-c)(CLW^{(t)} + CLW_f^{(t)}) + c\ CLB^{(s,t)}$$

$$\simeq (1-c)\ CLW^{(t)} + CLW_f^{(t)},$$

since under the assumption of $W_{14}/\bar{W} \simeq 1$, $CLW_f^{(t)} \simeq CLB^{(s,t)}$. This assumption will hold approximately provided gene frequencies are close to intermediate values and gene effects are small. We shall now investigate numerically the validity of these results. The technique we use to calculate what we call exact results, is described in the Appendix under the heading 'Selection Within Genotypic Classes'. We work with a model of four additive loci, with an arbitrary degree of linkage and we focus our attention on a single pair of them. We start the selection process at t=0, assuming Hardy-Weinberg and linkage equilibrium with the same initial conditions in all runs, except for the value of c. Initial gene frequencies are set to 0.25 for all loci, h^2 = 10% and Q = 20%. At t=2, when gene

frequencies are close to ½, we substitute in expressions (5.18), (5.25), (5.29) and (5.31) the values obtained from the exact results, for gene frequencies and for the disequilibrium in the offspring generation induced in the previous 2 cycles of selection. This procedure was chosen in order to avoid the asymmetry mentioned in Chapter 3 due to extreme gene frequencies. Furthermore, under additivity, at gene frequencies close to intermediate values $W_{14}/\bar{W}$ is close to 1 (0.983 for the values of q shown below). The observed and predicted results at t=3 are shown in Table 5.1, for three different values of c. The last two columns of the table show the disequilibrium within gametes and gene frequencies prior to the third cycle of selection.

TABLE 5.1: Observed and Predicted Values of Different Covariances of Gene Frequencies. See text for Explanation.

c		$D_{HW}^{(s,2)} \times 10^5$ (0)	$D^{(s,2)} \times 10^5$ (1)	$D_B^{(s,2)} \times 10^5$ (2)	$D^{(3)} \times 10^5$ (3)	$D^{(2)} \times 10^5$ (4)	$q_{(2)} \times 10^2$ (5)
0.5	OBS	-57.93	-152.85	-57.93	-105.39	-96.69	47.07
	PRED	-66.30	-161.25	-66.30	-112.90		
0.1	OBS	-56.73	-178.77	-56.73	-166.57	-124.20	47.07
	PRED	-64.74	-186.83	-64.74	-174.41		
0.0	OBS	-56.43	-185.26	-56.43	-185.26	-131.08	47.07
	PRED	-64.39	-193.23	-64.39	-193.23		

The predicted results tend to consistently overestimate the observed results. The difference between the values in columns (1) and (4) represents what we have called fresh disequilibrium, $D_f^{(t)}$. This difference is in close agreement with the value observed in column (2), illustrating that when $W_{14}/\bar{W} \simeq 1$, $D_B^{(s,t)} \simeq D_f^{(t)}$. When gene frequencies move away from intermediate values such that $W_{14}/\bar{W} \neq 1$, $D_B^{(s,t)} \neq D_f^{(t)}$, particularly in the case of the present model where gene effects are rather large. Throughout the selection process though, the value of $D_B^{(s,t)}$ is in excellent agreement with that predicted using (5.23). As predicted from (5.18), when gene frequencies are at intermediate values, the fresh disequilibrium is independent of the previously existing disequilibrium and therefore is similar for all values of c. When gene frequencies move beyond 0.5, three of the five terms involving $D_f^{(t)}$ are positive. In fact, at high gene frequencies the fresh disequilibrium becomes positive and highly dependent on the recombination fraction between the loci involved. In other words, the closer the linkage, the larger the absolute value of the disequilibrium between loci within parental contributions and therefore the higher the positive value of the fresh disequilibrium attained. The total disequilibrium within parental contributions amongst selected individuals is of course always negative and tends to zero as gene frequencies move towards fixation. We illustrate these concepts in Table 5.2, where the same model and procedures used in Table 5.1 are shown after 7 cycles of selection when gene frequencies have reached extreme values and $W_{14}/\bar{W} \simeq 0.33$. As before the predicted results overestimate the observed results, this overestimation

TABLE 5.2: Observed and Predicted Values of Disequilibria at High Gene Frequencies. See Text for Explanation.

c		$D_{HW}^{(s,7)} \times 10^5$ (0)	$D^{(s,7)} \times 10^5$ (1)	$D_B^{(s,7)} \times 10^5$ (2)	$D^{(8)} \times 10^5$ (3)	$D^{(7)} \times 10^5$ (4)	$q_{(7)} \times 10^2$ (5)
0.5	OBS	-1.98	-5.18	-1.98	-3.58	-9.83	92.46
	PRED	-5.34	-8.54	-5.34	-7.00		
0.1	OBS	-2.24	-10.54	-2.24	-9.71	-25.11	91.96
	PRED	-6.00	-13.75	-6.00	-13.00		
0.0	OBS	-2.35	-13.56	-2.35	-13.56	-33.74	91.76
	PRED	-6.10	-16.75	-6.10	-16.75		

being relatively larger at high gene frequencies. $D_f^{(7)}$ is in all cases positive and highly dependent on c. Furthermore, $D_f^{(t)}$ is very different from $D_B^{(s,t)}$, this difference being of course accentuated with tighter linkage. As before, $D_B^{(s,t)}$ is in excellent agreement with expression (5.23). Both Tables 5.1 and 5.2 show that when linkage is tight a high proportion of the already existing disequilibrium remains in each generation. In the extreme case of complete linkage, the fresh disequilibrium generated at the $t^{\underline{th}}$ cycle of selection together with the already existing disequilibrium before selection, remain in the offspring at generation t+1.

This analysis of the two locus model is here regarded as an attempt to understand the interaction of gene frequency changes and disequilibrium during directional selection. The expressions we

have arrived at do not seem to predict the course of selection with precision but they are useful in that, at least qualitatively, they highlight the way the various parameters involved interact together during the different stages of selection.

Multilocus Models.

The theory of changes of genetic parameters developed by Bulmer has the great practical advantage that it describes the process in terms of readily estimable parameters of the base population. As we have pointed out the theory is based on the assumption that the trait is determined by an effectively infinite number of loci, so that gene frequency changes can be ignored. In this section we extend the results of the previous section to an arbitrary number of loci, and we study the joint effect of gene frequency changes and generation of disequilibrium as selection proceeds.

Gene frequency changes depend on the number, effects and frequencies of the genes involved, information which on the whole is not available, particulary for the case of metric traits. Since these are variables that must be incorporated in a model which assumes a finite number of loci, the work that follows must not be interpreted as an attempt to provide expressions of direct practical application. The purpose of this work is of a different nature, namely to check the theory developed by Bulmer under a variety of genetic models and from the results obtained, arrive at some conclusions concerning the relative importance of the different forces determining changes in genetic parameters.

We shall first assume that the loci segregate independently. We relax this assumption in the next section.

Free Recombination.

Consider a trait determined by n loci of equal proportionate effects, a/σ, and frequencies. If we apply a selection intensity of i standard deviations at the $t^{\underline{th}}$ cycle of selection, the expected change of gene frequency at each locus, $\Delta p^{(t)}$, as a first order approximation, is, from (5.9):

$$\Delta p^{(t)} = \frac{i}{\sigma}\left(ap(1-p) + (n-1)aD\right) \qquad (5.32)$$

since each locus is correlated with the remaining (n-1) loci in the genome. The new gene frequency is then,

$$p^{(t+1)} = p^{(t)} + \Delta p^{(t+1)},$$

from which we obtain the equilibrium additive variance, $Vg^{(t)} = 1$ $2na^2p^{(t)}(1-p^{(t)})$. Consider now the generation of joint disequilibrium, measured before the operation of selection, in each offspring generation. Bulmer showed that under the infinitesimal model,

$$CLW^{(t+1)} = -\tfrac{1}{2}\frac{i(i-x_T)}{\sigma^2(t)} VG^{2(t)} + \tfrac{1}{2}CLW^{(t)},$$

where

$$VG^{(t)} = VG^{(o)} + CLW^{(t)}.$$

Our approximation (5.18) and (5.31) could readily be extended to accommodate an arbitrary number of loci. The resulting expression is not as readily interpreted as the one based on the infinitesimal model. An alternative approach which has been followed in this work is to attempt to find an expression, by trial and error, which can describe the process reasonably well and which takes account of the

various parameters involved. This expression which is suggested by Bulmer's result and the definition of the total genotypic variance, VG, has been found to be,

$$CLW^{(t+1)} \simeq -\tfrac{1}{2}\frac{i(i-x_T)}{\sigma^2(t)}\left[Vg^{(t)} + CLW^{(t)}\right]^2 + \tfrac{1}{2}CLW^{(t)} \qquad (5.33)$$

This is a function of well defined parameters and allows us to at least make some qualitative predictions since the disequilibrium generated is ultimately, according to (5.33), a function of gene frequencies. In order to understand its behaviour, we shall compare (5.33) and (5.31) with exact results at the end of this section. Repeated use of (5.33) allows us to predict the value of the total genotypic variance in successive generations. These results can be readily extended to predict changes in the genotypic variance due to selection for a trait determined by loci of different proportionate effects and frequencies. Assume that out of n loci affecting the character, n_1 have effect a_1 and frequency p_1, and n_2 have corresponding values of a_2 and p_2. We refer to the n_1 and n_2 loci as the type 1 and type 2 loci respectively. The expected change in frequency of each type of loci is:

$$p_1^{(t)} = \frac{i}{\sigma}\left(a_1p_1(1-p_1) + (n_1-1)a_1D_{11} + n_2a_2D_{12}\right)^{(t-1)},$$

and

$$p_2^{(t)} = \frac{i}{\sigma}\left(a_2p_2(1-p_2) + (n_2-1)a_2D_{22} + n_1a_1D_{12}\right)^{(t-1)},$$

where D_{ij} is the disequilibrium between type i and type j loci (i=j=1,2). These expressions can be used recurrently to predict

gene frequency changes from which we obtain the equilibrium additive variance $Vg^{(t)}$

$$Vg^{(t)} = Vg_1^{(t)} + Vg_2^{(t)},$$

where

$$Vg_i^{(t)} = 2n_i a_i^2 p_i(1-p_i)^{(t)}.$$

In order to predict the generation of joint disequilibrium under this model, the following covariances of allelic effects between loci (within and between types) can be readily obtained:

$$CLW_{11}^{(t+1)} = -\left[\tfrac{1}{2}\frac{i(i-x_T)}{\sigma^2} VG_1^2\right]^{(t)} (1-\frac{1}{n_1}) + \tfrac{1}{2}CLW_{11}^{(t)}$$

$$CLW_{22}^{(t+1)} = -\left[\tfrac{1}{2}\frac{i(i-x_T)}{\sigma^2} VG_2^2\right]^{(t)} (1-\frac{1}{n_2}) + \tfrac{1}{2}CLW_{22}^{(t)}$$

$$CLW_{12}^{(t+1)} = -\left[\frac{i(i-x_T)}{\sigma^2} VG_1 VG_2\right]^{(t)} + \tfrac{1}{2}CLW_{12}^{(t)},$$

where σ^2 is the total phenotypic variance given by:

$$\sigma^{2(t)} = VG_1^{(t)} + VG_2^{(t)} + VE,$$

and $\overline{VG}_i$ is the genotypic variance contributed by the $i^{\underline{th}}$ type of loci, such that

$$(VG_1 + VG_2)^{(t)}/\sigma^{2(t)} = VG^{(t)}/\sigma^{2(t)}.$$

The total joint disequilibrium, CLWT is given by,

$$CLWT^{(t+1)} = CLW_{11}^{(t+1)} + CLW_{12}^{(t+1)} + CLW_{22}^{(t+1)} \qquad (5.34)$$

The extension to an arbitrary number of types of loci is straightforward.

The Presence of Linkage.

Our analysis of the two locus model has clearly identified the complications introduced by linkage. As was shown before, the covariance of gene frequencies within parental gametic contributions in the offspring at generation t+1, contains a proportion 1-c of $D^{(s,t)}$ and a proportion c of $D_B^{(s,t)}$. From (5.21) and (5.25), this leads to:

$$D^{(t+1)} = (1 - \frac{W_{14}}{\bar{W}} c)\ D^{(t)} + D_f^{(t)},$$

the same as (5.31). For a given value of gene frequencies, the closer the linkage the larger the proportion of the previous disequilibrium which is passed on to the following generation. The fresh disequilibrium induced by the $t^{\underline{th}}$ selection cycle is little affected by the already existing disequilibrium provided gene frequencies are intermediate (see (5.18)). As was illustrated in the numerical analysis this no longer holds at more extreme gene frequencies when $D_f^{(t)}$ is highly dependent on $D^{(t)}$ and therefore on the degree of linkage. These conclusions are relevant to a model of many linked loci: pairs of loci which are closer together will contribute with different proportions to the total disequilibrium from loci far apart in the genome. From a conceptual point of view, the behaviour and understanding of a model of many linked loci is described by summing over pairs of loci in the above expression. This yields a general recurrent equation which allows for an arbitrary number of loci,

degree of linkage gene effects and frequencies. Alternatively, making some simplifying assumptions we can arrive at simpler expressions which convey a more meaningful picture and furthermore are functions of parameters which in some cases can more or less be estimated experimentally. This latter approach was taken up by Bulmer (1974) whose work forms the basis of the results that follow.

Assume n loci affect a metric trait, of equal effects and frequencies. Let $d_{ij}^{(t)}$ be the contribution to $CLW^{(t)}$ from the $ij^{\underline{th}}$ pair of loci and let c_{ij} be the recombination fraction between them. If n is large, for a given amount of genetic variation the proportionate effects, and consequently the selection pressure, at each locus is relatively small. We then may assume, following Bulmer (1974) that the contribution to the fresh disequilibrium from each pair of loci is small and more or less similar for all pairs of loci. We have shown that this is approximately true provided gene frequencies are not far from intermediate values. Therefore we can write,

$$d_{ij}^{(t+1)} = (1-C_{ij})\, d_{ij}^{(t)} - \tfrac{1}{2} \frac{i(i-x_T)}{\sigma^2(t)}\, VG^{2(t)} \,/\tfrac{1}{2}n(n-1) ,$$

since there are $\frac{1}{2}n(n-1)$ pairs of loci contributing to CLW. Since $d_{ij}^{(t)} = CLW^{(t)}/\frac{1}{2}n(n-1)$, summing over pairs of loci we obtain:

$$CLW^{(t+1)} = -\tfrac{1}{2} \frac{i(i-x_T)}{\sigma^2(t)}\, VG^{2(t)} + (1-\bar{c})\, CLW^{(t)} \qquad (5.35)$$

where $\bar{c}$ is the mean recombination fraction between the loci involved and can easily be obtained from the relation between recombination fraction and map distance. One such relationship, which assumes no interference and which will be used in this work is, $C_{ij} = \frac{1}{2}(1-e^{-2x_{ij}})$,

(Haldane, 1919), where x_{ij} is the map distance between the $ij^{\underline{th}}$ pair of loci.

If at the various loci gene frequencies are extreme or gene effects very different, the approximation (5.35) is unlikely to hold, particularly because the assumption that the fresh disequilibrium is the same for all pairs of loci is untenable. Extreme gene frequencies also lead to the additional problem of lack of normality of the genotypic distribution and as we have shown before, this may cause substantial degree of asymmetry in the generation of disequilibrium. We shall have an opportunity to study the behaviour of expression (5.35) under various models in the simulation work at the end of this chapter.

Genetic Parameters in Parental Generation.

The approach that we followed in the previous section can be used to obtain expressions for the various covariances of allelic effects between different parts of the genotype in the selected population. From (5.19) and (5.33), the joint disequilibrium within parental gametic values amongst selected individuals can be approximated by,

$$CLW^{(s,t)} \simeq -\tfrac{1}{2}\frac{i(i-x_T)}{\sigma^{2}(t)} VG^{2(t)} + CLW^{(t)} \qquad (5.36)$$

If the number of loci is large and proportionate ejfects are small, provided gene frequencies are never extreme, we may assume that W_{14}/W is close to 1 and therefore, from (5.23)

$$CLB^{(s,t)} \simeq -\tfrac{1}{2}\frac{i(i-x_T)}{\sigma^2(t)} VG^{2(t)} \qquad (5.37)$$

The effect of Hardy-Weinberg departures on the genotypic variance of selected parents can be obtained from (5.29):

$$CHW^{(s,t)} = -\frac{i(i-x_T)}{\sigma^2(t)}\left(\tfrac{1}{2}\sum_{i}^{n} Vg_i^{\,2} + \sum_{i<j}^{n\,n} Vg_i(2a_i a_j D_{ij})\right)^{(t)} \qquad (5.38)$$

where Vg_i is the equilibrium genetic variance of the $i^{\underline{th}}$ locus. If all loci have equal effects and frequencies,

$$CHW^{(s,t)} = \frac{-i(i-x_T)}{2\sigma^2(t)}\frac{Vg^{(t)}}{n}\left(V_g^{(t)} + CLW^{(t)}\right) \qquad (5.39)$$

For a given amount of genetic variation, in the first cycle of selection, assuming $CLW^{(o)} = 0$, CHW is inversely proportional to the number of loci. This should hold for later generations though eventually, the larger the selective advantage of individual loci, the faster the change in Vg and as gene frequencies move towards fixation, CHW will tend to zero. The effect of Hardy-Weinberg departures on the genotypic variance relative to the effect of CLW and CLB is of order 1/n.

We express the total genotypic variance in the selected population as:

$$VG^{(s,t)} \simeq \{Vg + CHW + CLW + CLB\}^{(s,t)}$$

As we shall see in later chapters, these expressions feature in estimates of heritability from offspring data of selected parents using intra-class correlations.

We now proceed to carry out some numerical checks on the results obtained so far, basically as a means of illustrating the limitations of the approximations which we shall use in the following sections. Three different sets of results are presented. The first one is obtained from the technique described in the Appendix (Selection within genotypic classes), and will be referred to as E. The second set of results are the outcome of repeated use of expression (5.35) and is shown under the heading I. The third set of results is generated by repeated use of (5.18) and (5.31) and we symbolise it, II. Since the model assumes two loci the heritability is taken to be 5% in order to avoid the problems of genes of very large effect. In Table 5.3 we assume that gene frequencies are initially 0.5 at both loci and c takes values of 0.5, 0.1 and 0.0. The results refer to the values of the parameters in the offspring generation, before selection.

Gene frequency changes are predicted with reasonable accuracy and as we illustrated in Chapter 3, the first order term overestimates the expected change in later generations if initial frequencies are at intermediate values. The effect of this discrepancy on the equilibrium additive variance is very small. For $c = \frac{1}{2}$ at generation 4 the value of the equilibrium additive variance predicted underestimates the true value by about 3%. The predictions of joint disequilibrium using I overestimate the true value particularly when linkage is tight. Method II is more accurate than method I but again overestimates the reduction of the genotypic variance due to disequilibrium.

TABLE 5.3:

Observed and Predicted Values of Gene Frequencies and Joint Disequilibrium. Initial Conditions: $q_o = 0.5$; $VG^{(o)} = 1.00$; $CLW^{(o)} = 0.00$; $h^{2(o)} = 5\%$; $Q = 20\%$; $a/\sigma = 0.22$. E, I and II refer to the three methods used to predict q and CLW. See text for further explanation.

		E		I		II	
t	c	$q_{(t)}$	$CLW^{(t)} \times 10^2$	$q_{(t)}$	$CLW^{(t)} \times 10^2$	$q_{(t)}$	$CLW^{(t)} \times 10^2$
	0.5	0.500	0.00	0.500	0.00	0.500	0.00
0	0.1	0.500	0.00	0.500	0.00	0.500	0.00
	0.0	0.500	0.00	0.500	0.00	0.500	0.00
	0.5	0.577	-0.91	0.578	-1.00	0.578	-1.00
1	0.1	0.577	-0.91	0.578	-1.00	0.578	-1.00
	0.0	0.577	-0.91	0.578	-1.00	0.578	-1.00
	0.5	0.649	-1.19	0.654	-1.40	0.654	-1.35
2	0.1	0.649	-1.51	0.654	-1.80	0.654	-1.68
	0.0	0.649	-1.59	0.654	-1.89	0.654	-1.77
	0.5	0.717	-1.10	0.724	-1.51	0.724	-1.32
3	0.1	0.716	-1.71	0.724	-2.39	0.724	-1.98
	0.0	0.716	-1.89	0.723	-2.68	0.724	-2.17
	0.5	0.776	-0.87	0.785	-1.30	0.785	-1.08
4	0.1	0.775	-1.58	0.785	-2.71	0.784	-1.88
	0.0	0.775	-1.84	0.784	-3.29	0.783	-2.15

Tables 5.4 and 5.5 illustrate the selection process starting with more extreme frequencies.

The predictions of gene frequency changes are again quite good; the difference between observed and predicted results is in the direction predicted from theory. Neither method I nor II allow for the initial asymmetry in disequilibrium generated when gene frequencies are not intermediate. At low initial frequencies method I is reasonably accurate, but it is considerably less so at high initial frequencies. Method II follows the changes in disequilibrium in reasonable agreement with exact results. It is worthwhile emphasizing that predictions based on the infinitesimal model, which ignore gene frequency changes are bound to break down badly if initial gene frequencies are high, since the model assumes that the joint disequilibrium always increases towards its maximum value as selection proceeds. It is clear, however, that the disequilibrium is a function of gene frequencies and as these move towards fixation, the disequilibrium tends to zero.

We now look at the results in the parental generation. Since similar comments and limitations regarding the predictions used apply to the parental generation we shall only show results for the run in Table 5.3 with $c = 0.5$. As before, E refers to exact results; results under I are obtained from recurrent use of (5.36) and (5.37). Results under II are based on recurrent use of (5.18) and (5.25). The effect of Hardy-Weinberg departures is obtained using (5.38). The results are shown in Table 5.6.

TABLE 5.4:

Observed and Predicted Values of Gene Frequencies and Joint Disequilibrium for a pair of loci. Initial Conditions: $q = 0.2$; $VG = 0.64$; $h^2 = 5\%$; $CLW = 0.0$; $Q = 20\%$; $a/\sigma = 0.27$; $c = \frac{1}{2}$.

	E		I		II	
t	q_t	CLW x 10^2	q_t	CLW x 10^2	q_t	CLW x 10^2
0	0.200	0.00	0.200	0.00	0.200	0.00
1	0.263	-0.82	0.262	-0.60	0.262	-0.60
2	0.338	-1.61	0.334	-1.30	0.334	-1.25
3	0.421	-2.22	0.413	-1.90	0.416	-1.88
4	0.509	-2.49	0.505	-2.50	0.507	-2.33

TABLE 5.5:

Observed and Predicted Values of Gene Frequencies and Disequilibrium for similar starting conditions as those in Table 5.4, except q_o are assumed to be 0.80.

	E		I		II	
t	q_t	CLW x 10^2	q_t	CLW x 10^2	q_t	CLW x 10^2
0	0.800	0.00	0.800	0.00	0.800	0.00
1	0.858	-0.36	0.861	-0.60	0.861	-0.60
2	0.898	-0.29	0.906	-0.60	0.906	-0.47
3	0.930	-0.17	0.945	-0.49	0.939	-0.27
4	0.952	-0.09	0.965	-0.28	0.960	-0.13

TABLE 5.6:

Observed and Predicted Values of Disequilibria for the Model in Table 5.3 for c = 0.5. See text for explanation.

	E				
t	$CLW^{(s,t)}$ x10^2	$CLB^{(s,t)}$ x10^2	$CLW_f^{(t)}$ x10^2	$CHW^{(s,t)}$ x10^2	$\frac{W_{14}}{\bar{W}}$
0	-0.91	-0.91	-0.91	-0.91	0.98
1	-1.59	-0.78	-0.68	-0.78	0.88
2	-1.57	-0.62	-0.38	-0.62	0.80
3	-1.27	-0.46	-0.17	-0.46	0.73

	I		II			
t	$CLW^{(s,t)}$ x10^2	$CLB^{(s,t)}$ x10^2	$CLW^{(s,t)}$ x10^2	$CLB^{(s,t)}$ x10^2	$CHW^{(s,t)}$ x10^2	$\frac{W_{14}}{\bar{W}}$
0	-1.00	-1.00	-1.00	-1.00	-0.98	0.96
1	-1.89	-0.91	-1.77	-0.91	-0.91	0.86
2	-2.18	-0.77	-1.84	-0.78	-0.78	0.78
3	-2.09	-0.60	-1.55	-0.61	-0.61	0.72

The exact results illustrate the effect of the declining value of $W_{14}/\bar{W}$ on $CLB^{(s,t)}$. Under the assumption of $W_{14}/\bar{W} = 1$, the fresh joint disequilibrium, $CLW_f^{(t)}$, should be equal to $CLB^{(s,t)}$. Notice however, that the discrepancy between $CLW_f^{(t)}$ and $CLB^{(s,t)}$ in E, is perfectly explained using (5.25). As we predicted from the theoretical analysis, the difference between $CLW_f^{(t)}$ and $CLB^{(s,t)}$ is positive and becomes larger in magnitude as gene frequencies move towards fixation. At generation 0, if $D^{(o)} = 0$, $CLB = CLW_f$. Both methods overestimate the value of $CLB^{(s,t)}$. In this particular run, the difference in the predicted value of CLB using both methods is at the $4^{\underline{th}}$ decimal place. The overestimates of the fresh disequilibrium are reflected on the predicted value of $CLW^{(s,t)}$, which is further inflated by the prediction of the already existing disequilibrium within parental contributions. Method II tends to correct for the effect of high gene frequencies and is therefore more accurate than method I. In this particular run with two loci of equal effects and frequencies, $CHW^{(s,t)} = CLB^{(s,t)}$. Both decline as gene frequencies move towards fixation.

This numerical analysis illustrates the limitations of the approximations used to predict the course of selection. Method II is more accurate than method I, but both seem useful in providing us with a means of making some predictions, at least qualitatively. The expressions used in method I are functions of parameters which can be more or less estimated experimentally. Due to this, and due to its simplicity, it will be used to describe the changes of the genotypic variance due to selection in the Montecarlo work of the next section.

Montecarlo Simulation Studies.

The Simulation Programme.

The Montecarlo approach followed in this study was that of directly simulating the processes of gamete formation, random mating, genotypic evaluation on the individual's own performance and truncation selection. Bisexual diploid organisms were simulated, their quantitative characteristic assumed to be expressed equally in both sexes. The metric trait was determined by a maximum of 30 loci, two alleles per locus, with arbitrary effects and frequencies and any degree of linkage between adjacent loci. The genetic models studied assumed additivity between loci (no epistasis) and both additive and dominant models were investigated. The mode of gene action thus specified the genotypic value of each individual and the phenotypic values were generated by adding a normally distributed random variable with zero mean and variance VE simulating environmental effects. In the directional selection studies that follow, the highest N scoring individuals of each sex out of a total of M scored were selected for breeding. The 2M individuals at generation zero were generated according to the input of gene frequencies and therefore the base population was assumed to be in Hardy-Weinberg and linkage equilibrium, any departures being due to chance. Recombination and gamete formation were carried out using an array of binary masks. Mating within selected individuals was at random, with no replacement and a constant number of offspring of each sex per family was produced each generation. The computer input can be summarised as follows:

- Number of loci
- Number of Male offspring
- Number of Female offspring
- Number of Male parents
- Number of Female parents
- Number of generations of selection
- Number of replicates
- Environmental variance
- Recombination fraction between adjacent loci
- Gene frequencies at each locus
- Additive values at each locus
- Dominance deviations at each locus

The output varied somewhat in different versions. In general, the following were printed each generation, before selection, together with the standard deviation between replicate runs for each estimate:

- Genotypic mean
- Total genotypic variance (VG)
- Equilibrium additive variance (Vg)
- Covariance of allelic effects within loci (CHW)
- Covariance of allelic effects between loci within gametes (CLW)
- Skewness of the distribution of genotypic values (g_3)

The total genotypic variance (VG) was estimated from the variance of the distribution of genotypic values. Vg was estimated from gene frequencies which were obtained from each of the n loci.

CHW was estimated by subtracting the equilibrium additive variance of each locus from the variance between genotypes within the

corresponding locus and summing over loci. CLW was obtained by difference, on the assumption that the expected value of CLB in the offspring generation was zero, i.e.:

$$CLW^{(t)} = VG^{(t)} - Vg^{(t)} - CHW^{(t)}.$$

The same estimates were printed out for the selected group of individuals, immediately after selection. In addition, the covariance of allelic effects between gametes, CLB, together with its standard deviation between replicates, was printed each generation. This estimate was calculated in the following way. An estimate of (CLW + CLB) was obtained by subtracting from VG, the pooled variance between genotypes within loci. CLW was estimated by subtracting the equilibrium additive variance from twice the variance between paternally derived gametic values across individuals (under the assumption that the variance between gametes is the same in maternal and paternal gametes). In most runs the number of individuals of each sex scored each generation was 200, the best ranking 40 of each sex being selected. This population size was chosen in order to compromise between the number of replications and the number of generations of selection for a given length of computing time. The number of replicates was seldom larger than 30, this number having been decided empirically on the basis of the results obtained in different runs. Usually not more than four cycles of selection were investigated and therefore, in this strictly short term study, given the size of population, the decline in variance within lines due to drift has been ignored.

Simulation Results.

In this section we compare the predictions made under the infinitesimal model with results obtained from Montecarlo simulations. We also produce results based on method I which takes account of gene frequency changes. In Table 5.7 we summarise the various models. These models were chosen in order to illustrate and discuss how the various parameters interact and the extent to which the previous theoretical analysis provides us with a means of explaining the results, and no strong claim is made about them reflecting possible genetic parameters of a particular character in any species. The distribution of genes in the genome, however, may give a rough indication of what can be expected in species with different numbers of chromosomes.

In the tables that follow, simulation results are headed (O); those obtained using method I, (I), and those results obtained using the infinitesimal model, (∞). The following genetic parameters, before selection, are shown below:

R : accumulated selection response

CLW : joint disequilibrium

Vg : equilibrium additive variance.

These parameters are calculated each generation in the usual way, that is:

$$R^{(t+1)} = i\, h^{2(t)} \sigma(t)$$

$$CLW^{(t+1)} = -\tfrac{1}{2} \frac{i(i-x_T)}{\sigma^2(t)} VG^{2(t)} \left(1-\frac{1}{n}\right) + (1-\bar{c}) CLW^{(t)}$$

TABLE 5.7:

Initial genetic parameters of the different models. The models are designated by the corresponding number of loci (n), the initial gene frequencies (q) and the recombination fraction between adjacent loci (c). The first three runs only differ in the number of loci. Run 4 is equivalent to Run 1, except that c is zero. Run 5 is again equivalent to Run 1 except for the linkage relationship between loci. Run 5 has 30 loci, 10 on each of 3 chromosomes and recombination fraction between loci on the same chromosome equal to 0.1. In Run 6, 5 loci of proportionate effect of 0.61 have initial frequency equal to 0.1 and 25 of proportionate effect of 0.11 have initial frequency of 0.5. The loci are assumed to recombine freely. In Run 7, initial gene frequencies at all loci are 0.2. In Run 8, initial gene frequencies at all loci are 0.8. Both Runs 7 and 8 assume free recombination.

Run No:	Run designation (n,q,c)	No. of loci(n)	Initial value of a/σ	Initial frequency, (q_o)	Linkage(c)	Initial VG	Initial h^2
1	(30,0.5,0.5)	30	0.18	0.5	0.5	15	0.5
2	(10,0.5,0.5)	10	0.32	0.5	0.5	15	0.5
3	(4,0.5,0.5)	4	0.50	0.5	0.5	15	0.5
4	(30,0.5,0.0)	30	0.18	0.5	0.0	15	0.5
5	(30,0.5,0.1)	30	0.18	0.5	(10) 0.1 (10) 0.1 (10) 0.1	15	0.5
6	(5/25,0.1/0.5,0.5)	30	(5) 0.61 (25) 0.11	(5) 0.1 (25) 0.5	0.5	15	0.5
7	(30,0.2,0.5)	30	0.23	0.2	0.5	9.6	0.5
8	(30,0.8,0.5)	30	0.23	0.8	0.5	9.6	0.5

in the case of method (I), where $VG^{(t)} = Vg^{(t)} + CLW^{(t)}$ and

$$CLW^{(t+1)} = -\tfrac{1}{2} \frac{i(i-x_T)}{\sigma^{2(t)}} VG^{2(t)} + (1-\bar{c})\ CLW^{(t-1)}$$

in the case of (∞), where $VG^{(t)} = VG^{(o)} + CLW^{(t)}$.

The results for all runs for the genetic parameters in the offspring generation, before selection, are illustrated in Tables 5.8, 5.9 and 5.10. Tables 5.11, 5.12, 5.13 and 5.14 show the genetic parameters of the runs in the parental generation immediately following selection.

Genetic Parameters before Selection.

We first discuss the results of the various models in the offspring generation, before selection. Table 5.8 shows the results of runs 1(30,0.5,0.5), 2(10,0.5,0.5) and 3(4,0.5,0.5) which illustrate the effect of varying the number of loci. For a given amount of genetic variation as the number of loci increases the disequilibrium between a single pair of loci becomes smaller, but the number of terms contributing to the total disequilibrium becomes larger. For example, if the number of loci increases from n_1 to $n_2 = n_1(1+N)$, for a given amount of genetic variance the average effect of a gene substitution, a, at a locus, decreases by a proportion $\frac{1}{\sqrt{1+N}}$ and therefore the disequilibrium between a single pair of loci decreases by a proportion $1/(1+N)$. The ratio of the joint disequilibrium with n_1 loci to that with n_2 loci is $\dfrac{n_1-1}{n_1-\dfrac{n_1}{n_2}}$. In other words, if n_1 is not too small,

TABLE 5.8: Results of Runs 1, 2, and 3. See text for explanation.

Generation		RUN 1 (20 reps.) (30, 0.5, 0.5) Vg	CLW	R	RUN 2 (50 reps.) (10, 0.5, 0.5) Vg	CLW	R	RUN 3 (50 reps.) (4, 0.5, 0.5) Vg	CLW	R
	0	14.92±0.00	0.57±0.33	0.00	14.96±0.00	0.07±0.16	0.00	14.96±0.00	0.01±0.13	0.00
0	I	15.00	0.00	0.00	15.00	0.00	0.00	15.00	0.00	0.00
	∞	15.00	0.00	0.00	15.00	0.00	0.00	15.00	0.00	0.00
	0	14.54±0.02	-2.81±0.35	3.84±0.12	14.12±0.02	-2.51±0.16	3.94±0.06	13.06±0.04	-2.00±0.14	3.87±0.04
1	I	14.76	-2.83	3.83	14.26	-2.64	3.83	13.16	-2.20	3.83
	∞	15.00	-2.93	3.83	15.00	-2.93	3.83	15.00	-2.43	3.83
	0	13.75±0.04	-3.18±0.25	7.16±0.15	12.32±0.06	-2.50±0.14	7.07±0.08	9.18±0.09	-1.53±0.14	6.76±0.05
2	I	14.17	-3.41	7.05	12.56	-3.11	7.00	9.14	-2.46	6.85
	∞	15.00	-3.57	7.08	15.00	-3.57	7.08	15.00	-3.57	7.08
	0	12.76±0.07	-3.31±0.34	10.22±0.17	9.98±0.07	-2.14±0.11	9.82±0.08	5.21±0.13	-0.47±0.07	8.82±0.06
3	I	13.33	-3.40	10.02	10.33	-2.83	9.66	5.20	-1.83	8.86
	∞	15.00	-3.72	10.19	15.00	-3.72	10.19	15.00	-3.72	10.19
	0	11.50±0.10	-2.20±0.32	12.89±0.18	7.64±0.09	-1.29±0.13	11.98±0.08	2.37±0.08	-0.11±0.03	10.00±0.04
4	I	12.27	-3.19	12.80	7.95	-2.29	11.87	2.61	-1.09	9.96
	∞	15.00	-3.75	13.28	15.00	-3.75	13.28	15.00	-3.75	13.28

Vg : equilibrium additive variance
CLW : joint disequilibrium within parental(gametic) contributions.
R : expected response to selection.

TABLE 5.9: Results of Runs 4, 5 and 6. See text for explanation.

Generation		RUN 4 (20 reps.) (30, 0.5, 0.0) Vg	CLW	R	RUN 5 (20 reps.) (30, 0.5, 0.1) Vg	CLW	R	RUN 6 (30 reps.) (5/25, 0.1/0.5, 0.5) Vg	CLW	R
0	0	14.96±0.00	-0.50±0.27	0.00	14.96±0.00	-0.06±0.19	0.00	14.91±0.08	-0.06±0.16	0.00
	I	15.00	0.00	0.00	15.00	0.00	0.00	15.00	0.00	0.00
	∞	15.00	0.00	0.00	15.00	0.00	0.00	15.00	0.00	0.00
1	0	14.66±0.01	-3.17±0.36	3.76±0.07	14.66±0.01	-2.70±0.23	3.80±0.10	21.86±0.14	-4.69±0.29	4.13±0.07
	I	14.76	-2.83	3.83	14.76	-2.83	3.83	21.22	-2.65	3.83
	∞	15.00	-2.93	3.83	15.00	-2.93	3.83	15.00	-2.93	3.83
2	0	13.95±0.04	-4.69±0.50	6.96±0.14	14.04±0.03	-3.19±0.25	6.88±0.10	27.51±0.18	-7.47±0.37	8.55±0.14
	I	14.17	-4.83	7.05	14.17	-3.74	7.05	27.56	-4.83	8.32
	∞	15.00	-5.03	7.09	15.00	-3.90	7.08	15.00	-3.57	7.08
3	0	13.02±0.09	-6.04±0.40	9.52±0.22	13.13±0.03	-3.29±0.25	9.82±0.10	30.97±0.17	-8.93±0.39	13.38±0.18
	I	13.43	-6.18	9.70	13.35	-3.92	9.95	32.03	-7.00	13.50
	∞	15.00	-6.59	9.89	15.00	-4.24	10.12	15.00	-3.72	10.19
4	0	11.82±0.14	-6.53±0.45	11.63±0.27	12.06±0.05	-2.74±0.24	12.45±0.11	30.88±0.14	-8.54±0.41	18.44±0.18
	I	12.66	-7.08	11.85	12.35	-3.78	12.62	32.00	-8.89	19.04
	∞	15.00	-7.77	12.33	15.00	-4.36	13.09	15.00	-3.75	13.28

TABLE 5.10: Results of Runs 7 and 8. See text for explanation.

Generation		RUN 7 (30 reps.) (30, 0.2, 0.5) Vg	CLW	R	RUN 8 (30 reps.) (30, 0.8, 0.5) Vg	CLW	R
0	0	9.56±0.02	-0.00±0.12	0.00	9.59±0.02	0.05±0.12	0.00
	I	9.60	0.00	0.00	9.60	0.00	0.00
	∞	9.60	0.00	0.00	9.60	0.00	0.00
1	0	11.24±0.03	-2.24±0.20	3.15±0.06	7.70±0.04	-1.53±0.11	2.88±0.05
	I	11.28	-1.81	3.07	7.60	-1.81	3.07
	∞	9.60	-1.88	3.07	9.60	-1.88	3.07
2	0	12.44±0.04	-3.33±0.14	6.22±0.08	6.11±0.05	-1.21±0.08	5.02±0.06
	I	12.64	-2.68	6.10	6.08	-1.73	5.13
	∞	9.60	-2.28	5.67	9.60	-2.28	5.67
3	0	13.32±0.04	-3.98±0.15	9.27±0.11	4.73±0.05	-0.90±0.08	6.74±0.06
	I	13.73	-3.26	9.26	4.78	-1.38	6.76
	∞	9.60	-2.38	8.16	9.60	-2.38	8.16
4	0	13.90±0.03	-3.77±0.16	12.26±0.10	3.57±0.05	-0.60±0.07	8.12±0.06
	I	14.50	-3.69	12.53	3.66	-1.03	8.09
	∞	9.60	-2.40	10.62	9.60	-2.40	10.62

TABLE 5.11: Genetic Parameters in Selected Population of Runs 1, 2 and 3. See text for explanation.

Generation		RUN 1 (20 reps.) (30, 0.5, 0.5) CLW	CLB	CLW_f	RUN 2 (50 reps.) (10, 0.5, 0.5) CLW	CLB	CLW_f	RUN 3 (4, 0.5, 0.5) CLW	CLB	CLW_f
0	0	-2.98±0.57	-2.16±0.58	-3.55	-3.13±0.24	-1.90±0.24	-3.20	-2.48±0.23	-1.46±0.26	-2.56
	I	-2.83	-2.83	-2.83	-2.64	-2.64	-2.64	-2.20	-2.20	-2.20
	∞	-2.93	-2.93	-2.93	-2.93	-2.93	-2.93	-2.93	-2.93	-2.93
1	0	-4.60±0.58	-1.49±0.55	-1.79	3.31±0.22	-1.78±0.21	-0.80	-2.18±0.15	-0.71±0.16	-0.17
	I	-4.83	-2.00	-2.00	-4.42	-1.79	-1.79	-3.56	-1.36	-1.36
	∞	-5.04	-2.11	-2.11	-5.04	-2.11	-2.11	-5.04	-2.11	-2.11
2	0	-4.49±0.48	-0.86±0.40	-1.31	-3.05±0.17	-1.11±0.16	-0.55	-0.70±0.10	-0.31±0.09	0.83
	I	-5.11	-1.70	-1.70	-4.39	-1.29	-1.29	-3.06	-0.60	-0.60
	∞	-5.50	-1.93	-1.93	-5.50	-1.93	-1.93	-5.50	-1.93	-1.93
3	0	-4.17±0.37	-0.59±0.31	-0.87	-2.04±0.13	-0.48±0.13	0.10	-0.03±0.08	-0.23±0.08	0.44
	I	-4.90	-1.49	-1.49	-3.72	-0.88	-0.88	-2.01	-0.18	-0.18
	∞	-5.61	-1.89	-1.89	-5.61	-1.89	-1.89	-5.61	-1.89	-1.89

CLW : joint disequilibrium within parental (gametic) contributions.

CLB : joint disequilibrium between parental (gametic) contributions.

CLW_f : fresh disequilibrium within parental (gametic) contributions.

TABLE 5.12: Genetic Parameters in Selected Population of Runs 4, 5 and 6.

Generation		RUN 4 (20 reps.) (30, 0.5, 0.0) CLW	CLB	CLW_f	RUN 5 (20 reps.) (30, 0.5, 0.1) CLW	CLB	CLW_f	RUN 6 (30 reps.) (5/25, 0.1/0.5, 0.5) CLW	CLB	CLW_f
0	0	-3.20±0.34	-2.64±0.26	-2.70	-2.81±0.41	-2.80±0.42	-2.87	-5.26±0.47	-4.22±0.52	-5.32
	I	-2.83	-2.83	-2.83	-2.83	-2.83	-2.83	-2.65	-2.65	-2.65
	∞	-2.93	-2.93	-2.93	-2.93	-2.93	-2.93	-2.93	-2.93	-2.93
1	0	-4.85±0.40	-1.29±0.31	-1.63	-4.70±0.34	-1.48±0.42	-2.01	-10.73±0.78	-3.67±0.67	-6.04
	I	-4.83	-2.00	-2.00	-4.83	-2.00	-2.00	-6.16	-3.51	-3.51
	∞	-5.03	-2.11	-2.11	-5.03	-2.11	-2.11	-5.04	-2.11	-2.11
2	0	-5.87±0.39	-1.43±0.35	-1.18	-4.58±0.36	-1.62±0.40	-1.39	-13.60±0.63	-4.26±0.62	-6.13
	I	-6.18	-1.36	-1.36	-5.35	-1.62	-1.62	-9.41	-4.58	-4.58
	∞	-6.59	-1.56	-1.56	-5.74	-1.84	-1.84	-5.50	-1.93	-1.93
3	0	-6.68±0.44	-0.65±0.21	-0.64	-3.62±0.32	-1.29±0.28	-0.33	-14.23±0.53	-3.36±0.58	-5.30
	I	-7.08	-0.89	-0.89	-5.29	-1.38	-1.38	-12.39	-5.39	-5.39
	∞	-7.77	-1.18	-1.18	-6.00	-1.76	-1.76	-5.61	-1.89	-1.89

TABLE 5.13: Genetic Parameters in Selected Population of Runs 7 and 8.

Generation		RUN 7 (30 reps.) (30, 0.2, 0.5) CLW	CLB	CLW_f	RUN 8 (30 reps.) (30, 0.8, 0.5) CLW	CLB	CLW_f
0	0	-2.86±0.25	-1.64±0.28	-2.86	-1.49±0.17	-1.15±0.15	-1.44
	I	-1.81	-1.81	-1.81	-1.81	-1.81	-1.81
	∞	-1.88	-1.88	-1.88	-1.88	-1.88	-1.88
1	0	-4.78±0.24	-1.59±0.28	-2.53	-1.58±0.14	-0.71±0.13	-0.05
	I	-3.59	-1.78	-1.78	-2.64	-0.82	-0.82
	∞	-3.22	-1.34	-1.34	-3.22	-1.34	-1.34
2	0	-5.25±0.28	-2.10±0.26	-1.92	-1.00±0.14	-0.62±0.11	0.21
	I	-4.60	-1.92	-1.92	-2.24	-0.51	-0.51
	∞	-3.52	-1.24	-1.24	-3.52	-1.24	-1.24
3	0	-5.86±0.29	-1.84±0.29	-1.88	-0.80±0.08	-0.35±0.08	0.11
	I	-5.32	-2.06	-2.06	-1.71	-0.34	-0.34
	∞	-3.59	-1.21	-1.21	-3.59	-1.21	-1.21

TABLE 5.14: Observed (O) and Predicted (P) reductions of the Genotypic Variance due to Departures from Hardy-Weinberg Equilibrium, among selected individuals ($CHW^{(s,t)}$).

Run Number		1	2	3	4	5	6	7	8
Model Designation		(30,0.5,0.5)	(10,0.5,0.5)	(4,0.5,0.5)	(30,0.5,0.0)	(30,0.5,0.1)	(5/25,0.1/0.5,0.5)	(30,0.2,0.5)	(30,0.8,0.5)
Generation									
0	O	-0.48±0.10	-0.41±0.07	-0.79±0.09	-0.19±0.06	-0.29±0.07	-0.64±0.14	-0.16±0.04	-0.12±0.03
	P	-0.10	-0.59	-0.73	-0.10	-0.10	-0.28	-0.06	-0.06
1	O	-0.58±0.08	-0.39±0.07	-0.47±0.09	-0.37±0.10	-0.43±0.10	-1.75±0.24	-0.39±0.06	-0.11±0.03
	P	-0.09	-0.24	-0.54	-0.09	-0.09	-0.58	-0.07	-0.04
2	O	-0.37±0.11	-0.18±0.06	-0.26±0.04	-0.36±0.12	-0.16±0.08	-1.32±0.32	-0.37±0.08	-0.09±0.01
	P	-0.08	-0.19	-0.27	-0.07	-0.08	-0.89	-0.08	-0.02
3	O	-0.43±0.09	-0.17±0.04	-0.07±0.02	-0.21±0.05	-0.34±0.09	-1.76±0.31	-0.41±0.08	-0.06±0.02
	P	-0.07	-0.13	-0.09	-0.06	-0.07	-0.93	-0.09	-0.02

there is very little effect of increasing the number of loci on the value of joint disequilibrium. When n_2 tends to infinity this ratio tends to $(n_1-1)/n_1$, indicating that given n_1 the maximum possible value that the joint disequilibrium can take, for the same amount of genetic variation, is a proportion $1 + \frac{1}{n_1}$ of its value with n_1 loci. This is illustrated in the results of Table 5.8 corresponding to the first generation of selection. The expected ratio of $CLW^{(1)}$ for the model of four loci to that of 30 loci is 0.78; the observed ratio is 0.71. The expected and observed ratios for 10 and 30 loci are 0.93 and 0.90 respectively. Observed and predicted results would not be in such close agreement if initial gene frequencies were extreme due to the problem of asymmetry mentioned before and to be discussed in the next chapter.

Since gene frequencies are initially at intermediate values we expect the genotypic variance to decline due to both gene frequency changes and due to the generation of linkage disequilibrium. With few loci of large effect, gene frequencies move quickly towards fixation and therefore the amount of disequilibrium quickly tends to zero. At high frequencies both (I) and (∞) overestimate $CLW^{(t)}$ as expected, particularly (∞) which is bound to breakdown badly when few genes are segregating. As the number of loci increases, observed and predicted results are in closer agreement. Gene frequency changes are predicted with reasonable accuracy but in disagreement with theory, the predictions are an underestimate. This is probably due to two reasons. Firstly the overestimation of $CLW^{(t)}$ tends to reduce the predicted value of Δp (see 5.32) and, secondly, the effect

of drift will tend to increase the decline in heterozygosity.

The observed reduction in the total genotypic variance due to the effect of joint disequilibrium, relative to its value at generation 0, for runs 1, 2 and 3 is illustrated in Table 5.15. The figures in the table show the effect of varying the number of loci on the relative contribution of joint disequilibrium and gene frequency changes on the change in the genotypic variance during the course of selection.

TABLE 5.15:

Observed reduction of total genotypic variance due to joint disequilibrium relative to its value at generation zero. The complementary fraction is the relative reduction due to changes in gene frequency (ignoring a negligible reduction due to Hardy-Weinberg disequilibrium).

t	Run 1 (30;0.5;0.5)	Run 2 (10;0.5;0.5)	Run 3 (4;0.5;0.5)
1	0.88	0.72	0.50
2	0.76	0.49	0.21
3	0.63	0.30	0.05
4	0.44	0.15	0.01

Table 5.8 also shows the observed and predicted response to selection in the three runs. Observed and predicted responses using method (I) tend to be in good agreement even in the case of run 3 where the total genotypic variance is underestimated due to an overestimation of $CLW^{(t)}$. It must be emphasized that the agreement is

strictly illusory particularly in the case of run 3 in later generations. Our prediction of selection response is based on the assumption of linearity of offspring-parent regression. This assumption does not hold during the later stages in run 3. After three cycles of selection, gene frequencies have reached a value of around 0.9. With high extreme gene frequencies the distribution of genotypic values is negatively skewed. Since the environmental distribution is assumed to be normal, this leads to non-linearity of offspring-parent regression due to inequality of ratios of third to second moments for the genotypic and environmental distributions and we would expect a higher response to selection downwards than upwards (Robertson, 1977c). The skewness of the distribution of genotypic values amongst the offspring of the 3rd generation is -0.80±0.04. The ratio of the total observed genotypic variance to total observed phenotypic variance in generation 3 is 0.24 whereas the observed realized heritability is 0.20 ± 0.01. The predictions made using method (I) yield a value for the ratio VG/VP of 0.18. Observed and predicted standardised selection differentials are 1.32 and 1.40 respectively. Therefore, the underestimation of the prediction of the heritability using (I) is compensated by an overestimation in the selection differential and hence observed and predicted responses to selection at generation 4 are in reasonable agreement (1.18 and 1.10 respectively). It is interesting to note that even though we have a situation of a few genes of large effect at extreme frequencies, with additive gene action the non-linearity of selection response does not seem to be very serious. For the model of 30 loci, the skewness of the distribution of genotypic values in the 3rd generation offspring

is -0.02 ± 0.04. Observed and predicted standardised selection differentials are 1.38 and 1.40 respectively and the ratio of observed VG/VP at generation 3 is 0.38 compared with a realised heritability of 0.39 ± 0.02, which agrees with the predicted value of 0.40 using method (I). Gene frequencies for run 1 at generation 3 reached a value of around 0.7.

The effect of linkage is illustrated comparing runs 1(30,0.5,0.5), 4(30,0.5,0.0) and 5(30,0.5,0.1). Up until generation 2 the equilibrium additive variance and the response to selection should be unaffected by linkage and any differences amongst the runs are probably due to sampling. From generation 2 onwards we expect the response to be reduced with tight linkage but as it is clear from the results the effect in the early generations is indeed very small, even for the case of complete linkage. Observed realized heritabilities are shown in Table 5.16 illustrating this point further.

TABLE 5.16:

Observed realized heritabilities at each generation for different values of linkage. $\bar{c}$ refers to the average recombination fraction between loci.

	Run 1	Run 5	Run 4
	(30,0.5,0.5)	(30,0.5,0.1)	(30,0.5,0.0)
t $\bar{c}$	0.5	0.38	0.00
0	0.50 ± 0.02	0.50 ± 0.01	0.52 ± 0.01
1	0.45 ± 0.02	0.44 ± 0.02	0.44 ± 0.01
2	0.43 ± 0.02	0.41 ± 0.01	0.37 ± 0.02
3	0.40 ± 0.02	0.38 ± 0.01	0.33 ± 0.01

Observed and predicted results of $CLW^{(t)}$ are in general in good agreement, but as expected the predicted values tend to be an overestimate. The response to selection is predicted reasonably accurately with method (I) and in general, with these runs starting at intermediate gene frequencies the results from the infinitesimal model follow the course of selection in a satisfactory manner, since overestimation of CLW partly compensates for the fact that the reduction in the total genotypic variance due to gene frequency changes is ignored.

Run 6 (5/25,0.1/0.5,0.5) simulates a model in which the character is determined by several genes of small effect at intermediate frequencies and some loci of large effect at low initial frequencies. In this run, about 67% of the total genetic variation at generation zero is contributed by the five major loci. This rather extreme situation will serve us to illustrate several points which were discussed earlier. From the point of view of the amount of joint disequilibrium generated after a first cycle of selection we would expect both methods (I) and (∞) to grossly underestimate the true value due to the asymmetry caused by loci at low initial frequencies, in this case, the phenomenon being accentuated due to the rather large proportionate effect of such loci. In subsequent cycles of selection as these loci move towards intermediate values the amount of disequilibrium should rise steeply. With this model, method (I) should be considerably more accurate than (∞) because the crucial factor causing the changes of disequilibrium as selection proceeds is the change in gene frequencies.

Predictions of selection response in the first generation are likely to be underestimated because of the positively skewed distribution of genotypic values due to the major loci at low frequencies. The skewness should quickly tend to disappear as the frequency of the major loci move away from extreme values.

These points are illustrated in Table 5.9, where the results are in good agreement with our verbal predictions. Notice that the predicted value of the equilibrium genetic variance is in reasonable agreement with observed results but the joint disequilibrium predicted using (I) and especially (∞) is an underestimate. The selection response predicted using (I) is again misleadingly accurate (except at generation 1), this being due to basically the fact that the overestimate of the predicted total genotypic variance is more or less balanced out by higher than expected observed realized heritabilities in view of the non-linearity problem. The observed standardised selection differentials in the first and last cycles of selection for this run were 1.38 and 1.39 respectively.

Table 5.17 gives the values of the skewness of the distribution of genotypic values in the offspring generation, the observed realised heritabilities at each generation and the observed ratios of total genotypic to total phenotypic variance as selection proceeds.

The table illustrates the fact that as the skewness tends to disappear observed and predicted results (h^2 and VG/VP) tend to agree more closely. It is interesting to notice that gene frequency changes and the generation of disequilibrium act together in such a way that the realized heritability changes little during the course

TABLE 5.17:

Observed parameter estimates for Run 6(5/25,0.1/0.4,0.5). g_3 refers to the skewness of the genotypic distribution in each offspring generation. h^2 is the single generation realized heritability.

t	g_3	h^2	VG/VP
0	0.44 ± 0.02	0.54 ± 0.01	0.50
1	0.27 ± 0.03	0.55 ± 0.01	0.52
2	0.16 ± 0.02	0.57 ± 0.01	0.57
3	0.04 ± 0.02	0.60 ± 0.01	0.59

of selection. Indeed, a very large experiment would be required to detect such a change. Upon relaxation of selection, if 75% of the disequilibrium breaks down after several cycles of random mating the heritability would rise to about 68%.

A similar model as the one in Run 6 but with all loci initially at intermediate frequencies would not produce the initial asymmetry in selection response or in the generation of joint disequilibrium. We would predict such a model to behave in a manner analogous to run 1(30,0.5,0.5), but with gene frequencies of the major loci moving rapidly towards extreme values we would find lack of agreement between observed and predicted values of CLW in later generations. This overestimate of CLW would more or less balance out the smaller than predicted realized heritability and again the predicted response

using (I) should be reasonably accurate. These verbal predictions were confirmed running such a model on the computer.

Runs 7(30,0.2,0.5) and 8(30,0.8,0.5) in Table 5.10 illustrate the problem of asymmetry in the generation of disequilibrium. As discussed before, both methods (I) and (∞) underestimate the value of CLW when gene frequencies are at low values and overestimate CLW at high initial frequencies. Due to the non-linearity of offspring-midparent regression the response is underestimated at low frequencies and overestimated at high frequencies, as discussed before. The predictions made under the infinitesimal model are reasonable when gene frequencies are below 0.5, but much less so at the other extreme situation, as expected. In general, the infinitesimal model breaks down badly if initial gene frequencies are higher than 0.5 because contrary to what it predicts, the amount of disequilibrium generated becomes smaller as selection proceeds. As we illustrate in the following section, this is due to the fact that the fresh disequilibrium eventually becomes positive as predicted from our theoretical analysis of the two locus model. It is interesting to notice that in run 7(30,0.2,0.5), gene frequency changes and the generation of joint disequilibrium result in an almost constant value for the observed (and predicted) realized heritability. In fact the observed realized heritabilities in each cycle of selection are as follows: 0.52 ± 0.01; 0.50 ± 0.01; 0.50 ± 0.01; 0.50 ± 0.01. The equilibrium additive variance however changes from its value of 9.57 ± 0.01 at t=0 to 13.91 ± 0.03 at t=4 and therefore if linkage equilibrium were restored, the realized heritability corresponding to that value of $Vg^{(4)}$ would be about 60%.

Genetic Parameters in Parental Generation.

Tables 5.11 to 5.14 show the observed and predicted values of disequilibria for the various runs. The previous analysis of the two locus model allows us to make the following verbal predictions:

(i) At low gene frequencies, we expect CLW_f < CLB and the opposite should hold at high gene frequencies.

(ii) At intermediate gene frequencies, CLB and CLW_f should take similar values.

(iii) As gene frequencies reach extreme high values, we expect the fresh disequilibrium to become positive. This phenomenon is not allowed for in the predictions made under methods (I) or (∞).

(iv) With increasing number of loci, the discrepancy between CLW_f and CLB should tend to be smaller.

Point (i) is illustrated in runs 7(30,0.2,0.5) and 8(30,0.8,0.5) in table 5.13. In run 7 as gene frequencies move towards intermediate values, CLB and CLW_f tend to be in closer agreement. Table 5.11 shows that with larger numbers of loci, the discrepancy between CLW_f and CLB as selection proceeds is reduced. In both runs 2(10,0.5,0.5) and 3(4,0.5,0.5) CLW_f > CLB, and CLW_f does indeed become positive at the last stages of selection, particularly in run 3 where gene frequencies have reached high extreme values. In run 8(30,0.8,0.5), which assumes high initial frequencies, CLW_f becomes positive very quickly though the absolute value reached is smaller than in the case of run 3(4,0.5,0.5) presumably due to the smaller proportionate effect of the loci. Gene frequencies at the 3rd parental generation in run

8(30,0.8,0.5) have reached a value of about 0.94.

With very tight linkage we expect CLW_f to move relatively more quickly towards positive values. A comparison between runs 1(30,0.5,0.5) and 4(30,0.5,0.0) shows that this effect is indeed very small. The values of CLB for run 1(30,0.5,0.5) at generations 2 and 3 are rather small but standard errors are large. In fact, another run of the same model using a different random number generator produced the following results for CLB and CLW_f respectively for generations 0 to 3: (-2.55 ± 0.41; -3.10); (-1.78 ± 0.40; -1.46); (-1.30 ± 0.30; -1.33); (-1.30 ± 0.34; -1.21).

The effect of Hardy-Weinberg departures on the genotypic variance amongst selected parents is shown in Table 5.14. From (5.38) and (5.39) we expect CHW to be smaller at extreme gene frequencies and larger at intermediate frequencies. Furthermore, for a given amount of genetic variation, CHW should increase with decreasing number of loci. These points are illustrated in the different runs in Table 5.14. It will be noticed, however, that unless the number of loci is small, the value of CHW predicted using (5.38) underestimates the true value. This is partly due to the finiteness of the population. It can be shown that in a population of size N, with random mating and no selection the expected frequency of the three genotypes at single loci is as follows:

AA	Aa	aa
$\dfrac{p^2 - \frac{p}{2N}}{1 - \frac{1}{2N}}$	$\dfrac{2p(1-p)}{1 - \frac{1}{2N}}$	$\dfrac{(1-p)^2 - \frac{(1-p)}{2N}}{1 - \frac{1}{2N}}$

In other words, in finite populations we expect an excess of heterozygotes and this implies a negative value for D_{HW}. With n loci, the covariance of gene frequencies within loci due to finite size, D^*_{HW}, is:

$$D^*_{HW} = \sum_{i=1}^{n} f(A_iA_i) - p_i^2 = -\frac{\Sigma p_i(1-p_i)}{2N-1} \qquad (5.40)$$

In our previous discussion of genetic parameters in the offspring generation we ignored the reduction in VG due to departures from Hardy-Weinberg equilibrium due to the fact that this reduction was virtually negligible. However, it should be mentioned that the value of CHW in the offspring generation was consistently negative in all runs and in good agreement with (5.40) above.

Discussion.

In this chapter we have attempted to describe the changes in variance due to selection under a variety of genetic models and at the same time see to what extent the infinitesimal model gives a good prediction of such a change. We have seen that provided gene frequencies are initially at more or less intermediate values and gene effects are small, the infinitesimal model is in good agreement with observed results, at least in the first few generations of selection studied here. In models in which gene frequencies are extreme and/or gene effects large, predictions are in poorer agreement, particularly at high frequencies.

Method (I) was derived more or less empirically and we have shown that it predicts the course of selection with reasonable accuracy

provided again, that gene frequencies are not extreme. Under this method, the effect of gene frequency changes on joint disequilibrium is allowed for and has been clearly illustrated in run 6(5/25,0.1/0.5,0.5) where loci had widely different selective values. At extreme gene frequencies however, the disequilibrium generated and the realised heritability are estimated with little accuracy but as we have seen, these two biases are of opposite sign and therefore, in general, the response to selection is predicted with misleading accuracy.

It might be thought that the models used are rather restricted in that only one intensity of selection and one value of heritability were used. The results of the different runs though, were shown simply to illustrate the theory developed in earlier sections and to show how the various parameters interact. It should be clear, however, that with smaller heritabilities and or lower selection intensities the relative reduction in variance is smaller and furthermore, the predictions made under the infinitesimal model are more accurate. This is illustrated in Table 5.18 where the predicted and observed values of joint disequilibrium and response to selection are shown for a run equivalent to run 1(30,0.5, 0.5) but with $h^{2(0)} = 0.30$.

If h^2 is further reduced to 20% and 50% are selected each generation (100/200 of each sex) the difference between method I and (∞) is even smaller and predicted and observed results for all genetic parameters studied here agree remarkably well. In fact, in such a model, the observed response at generation 4 is 2.61 ± 0.10 and the predicted results using (I) and (∞) are 2.60 and 2.61 respectively.

TABLE 5.18:

Observed and predicted results of joint disequilibrium (CLW) and selection response (R) for a run equivalent to run 1(30,0.5,0.5) but with initial heritability of 0.30. Observed results are the average of 30 replicates.

t		CLW	R
	0	-0.08 ± 0.04	0.00
0	I	0.00	0.00
	∞	0.00	0.00
	0	-0.44 ± 0.06	1.46 ± 0.04
1	I	-0.42	1.47
	∞	-0.43	1.47
	0	-0.49 ± 0.06	2.74 ± 0.05
2	I	-0.55	2.78
	∞	-0.57	2.79
	0	-0.50 ± 0.06	3.98 ± 0.05
3	I	-0.58	4.01
	∞	-0.61	4.06
	0	-0.42 ± 0.06	5.13 ± 0.05
4	I	-0.57	5.19
	∞	-0.62	5.33

Gene frequencies in this run moved from 0.50 to about 0.58 in the four cycles of selection and the relative change in the total genotypic variance due to joint disequilibrium relative to its value at generation zero is: 0.90; 0.74; 0.72 and 0.64 for generations 1 to 4. In this run and the one in Table 5.18, the agreement of observed and predicted response is not a spurious one; this agreement is a reflection of,(i) accuracy of prediction of gene frequency changes and the generation of joint disequilibrium and (ii) lack of asymmetry of selection response.

Having devoted considerable effort in attempting to understand the dynamics of gene frequency changes and the generation of joint disequilibrium, it is natural to ask how important are these changes from a practical point of view. Some idea of this can be obtained from the results in Tables 5.8 to 5.10. In order to be more precise though, we can ask how important are these changes in variance due to selection relative to variation of response to selection that we observe in replicated selection experiments of short duration.

Over the last ten years a considerable body of theory has been developed to describe the variance among replicated selected lines in short-term experiments (see Hill, 1977, 1980 for references). This theory is approximate in that it assumes no changes of variances as selection proceeds and therefore it will hold better, the shorter the duration of the selection process and the larger the number of genes affecting the trait. Without intending to review this theory, it should be enough to say that the variation in response between different lines sampled from the same base population is basically due to sampling of individuals chosen as parents, which produces

variance between lines in mean due to binomial sampling of genes (genetic drift) and variance between replicates of the within line variance, the latter being mostly due to linkage disequilibrium provided the number of loci is larger than the population size and having an effect of the same order of magnitude as the former (Avery & Hill, 1977; Hill, 1977).

In order to get a feel for the relevance of changes of genotypic variance as selection proceeds, we now compare the empirical standard deviation of the response to selection with the difference between the observed response (R(0)) and the one predicted on the assumption of no changes of genetic parameters. We also include, out of interest, predictions of the standard deviation of response based on formulae by Hill, though it must be clear that these predictions are not definite since the number of replicates used in these simulations are not in general large enough to study second order moments. The results are shown in Table 5.19. For simplicity runs 4 and 5, which involve linkage, have been omitted; no appreciable difference in variance of response or mean response could be detected between runs 1(30,0.5,0.5) and 5(30,0.5,0.1). Run 4(30,0.5,0.0), with complete linkage, showed larger variance of response as predicted from theory. We also include in Table 5.19 a measure of the extent to which the predictions of selection response made under the infinitesimal model are an improvement over those made ignoring any changes of genetic parameters. The measure of this is the ratio of the absolute value of $R_{(\infty)} - R_{(0)}/(tR_{(1)} - \Sigma R_i)$, where $R_{(\infty)}$ and $R_{(0)}$ are defined before, $R_{(1)}$ is the observed response at t=1 and ΣR_i is the observed cumulative response. If this ratio

TABLE 5.19: $SDR_{(O)}$: observed standard deviation of selection response. $SDR_{(P)}$: predicted standard deviation of R based on the following formula due to Avery & Hill, 1977.

$V(R) \simeq \frac{th^2VP}{N}(1-h^2(1-0.2-p)) + ti^2h^4VP(1-\frac{1}{2}h^2)2r^2$, where $2r^2 = 0$ for t=1; $2r^2 = \frac{1}{N}$ for t=2 and as a rough approximation, $2r^2 = 2/3N$ for t>2 (Avery & Hill, 1979). $V = tR_{(1)} - \Sigma R_i$; $W = \frac{R_{(\infty)} - R_{(O)}}{V}$

Run 9: 30 loci; $q^{(o)} = 0.5$, $VG^{(o)} = 15$, $h^{2(O)} = 0.2$, c = 0.5 for all loci. 100/200 selected in each sex. 16 replicates. See Text for Explanation.

t		Run 1 (30,0.5,0.5)	Run 2 (10,0.5,0.5)	Run 3 (4,0.5,0.5)	Run 6 (5/25,0.1/0.5,0.5)	Run 7 (30,0.2,0.5)	Run 8 (30,0.8,0.5)	Run 9
1	SDR(O)	0.54	0.39	0.30	0.40	0.31	0.30	0.20
	SDR(P)	0.36	0.36	0.36	0.36	0.29	0.29	0.17
	V	-	-	-	-	-	-	-
	W	-	-	-	-	-	-	-
2	SDR(O)	0.65	0.55	0.38	0.74	0.46	0.34	0.23
	SDR(P)	0.73	0.73	0.73	0.73	0.58	0.58	0.25
	V	0.52	0.81	0.97	-0.29	0.08	0.74	0.06
	W	0.15	0.01	0.31	5.00	6.87	0.88	0.33
3	SDR(O)	0.76	0.54	0.41	1.02	0.58	0.31	0.35
	SDR(P)	0.81	0.81	0.81	0.81	0.65	0.65	0.30
	V	1.31	2.00	2.78	-1.00	0.18	1.90	0.08
	W	0.02	0.19	0.49	3.19	6.17	0.75	0.25
4	SDR(O)	0.81	0.59	0.24	1.10	0.55	0.35	0.40
	SDR(P)	0.94	0.94	0.94	0.94	0.75	0.75	0.35
	V	2.45	3.77	5.40	-1.93	0.33	3.40	0.15
	W	0.16	0.34	0.61	2.67	4.97	0.74	0.00

is smaller than 1, the predictions made under the infinitesimal model are better than those made on the assumption that genetic parameters do not change, the opposite holding if the ratio is larger than 1.

A general glance at the table shows that once again the results are very much model dependent. In runs 1(30,0.5,0.5) and 2(10,0.5, 0.5), the changes of genetic parameters have a more or less clear effect on selection response at generation 3. In run 3(4,0.5,0.5) where genetic parameters change rather dramatically, these changes are likely to be detected at generation 2. In all these three runs, W shows that the predictions made under the infinitesimal model improve our prediction of selection response. In run 6(5/25,0.1/0.5,0.5) where the variance of response is considerably increased due to the presence of loci of large effect, changes of genetic parameters are more difficult to detect. The negative value of V indicates that contrary to the prediction made under the infinitesimal model, the genotypic variance increases as selection proceeds due to the effect of the major loci moving towards intermediate values. The large value of W gives further indication of the lack of accuracy of these predictions with this model. Run 7(30,0.2,0.5) shows a similar general picture. In run 8(30,0.8,0.5), when gene frequencies are very extreme, the variance of response is reduced and the effect of parameter changes on response is more easily detectable, though as in runs 2(10,0.5,0.5) and 3(4,0.5,0.5), most of the changes are due to gene frequency changes. Run 9 illustrates an interesting situation, in that with low selection intensity and genes of small effect, although predicted and observed results agree remarkably well,

the changes of genetic parameters are so small relative to the standard deviation of response that they are not likely to be easily detected.

The figures in the table show that, in general, the prediction of the variance of response is in reasonable agreement with observed results except in the models where gene frequencies move quickly towards fixation or, in the case of run 6, where major loci move towards intermediate frequencies. The simple expression based on the reduction in variance in the absence of selection, $t\ VG^{(0)}/N$, gives results similar to those obtained using Hill's approximation, as suggested by Hill (1977) and Robertson (1977b).

When population size is small and many loci of small effect are being selected, most changes in the equilibrium additive variance will be due to inbreeding. Computer runs made with population size of about 10 (5 males and 5 females) and selection intensities of 50% have shown this to be the case in agreement with theory (Robertson, 1960). The correlation of gene frequencies as selection proceeds however, does not seem to be affected by the size of population though the amount of joint disequilibrium generated decreases with smaller population size, presumably due to higher changes in heterozygosity. The computer results suggest that the following expressions can be used to describe the process:

$$Vg^{(t)} = Vg^{(o)}\left(1-\frac{1}{2N}\right)^{(t)} \quad ,$$

and assuming no linkage,

$$CLW^{(t)} = -\tfrac{1}{2}\,\frac{i(i-x_T)}{\sigma^2(t)}\ VG^{2(t)} + \tfrac{1}{2}\ CLW^{(t)} \quad ,$$

as before, where

$$VG^{(t)} = Vg^{(t)} + CLW^{(t)}.$$

The selection response at generation t being,

$$R^{(t)} = \sum_{j=0}^{t-1} i\ VG_j/\sigma_j.$$

These expressions are checked in table 5.20 where a model equivalent to run 1(30;0.5;0.5) was used, except that 5 out of 10 individuals in each sex were selected as parents in each generation. The predicted results agree fairly well with observed results.

TABLE 5.20:

Observed and predicted genetic parameters when N is small. The model assumes 30 unlinked loci, $h^{2(o)} = 0.50$; $VG^{(0)} = 15$ 5/10 selected in each sex. O: observed results (40 replicates), P: predicted results.

t		Vg	CLW	R
1	O	13.83 ± 0.05	-2.79 ± 0.57	2.09 ± 0.04
	P	14.25	-2.31	2.03
2	O	13.18 ± 0.06	-2.21 ± 0.65	3.61 ± 0.05
	P	13.53	-2.78	3.70
3	O	12.44 ± 0.06	-2.47 ± 0.53	5.07 ± 0.05
	P	12.86	-2.77	5.25
4	O	11.77 ± 0.09	-2.15 ± 0.54	6.46 ± 0.06
	P	12.20	-2.63	6.71

Table 5.21 shows the relationship between the reduction in variance and the variance of response to selection under this model.

TABLE 5.21:

SDR(O) : observed standard deviation of response.

$V = t\,R_{(1)} - \Sigma R_i$, $W' = \frac{R_{(P)} - R_{(O)}}{V}$, where $R_{(P)}$ and $R_{(O)}$ are predicted and observed responses respectively.

t		
2	SDR(O)	1.37
	V	0.57
	W'	0.19
3	SDR(O)	1.63
	V	1.20
	W'	0.15
4	SDR(O)	1.85
	V	1.90
	W'	0.13

Under this model, at least four or five replicates would be required to detect changes of genetic parameters at generation 4 and many more to detect changes in early generations. The W' values indicate that our predictions reflect the course of selection fairly accurately.

The above analysis shows that it is not easy to make very general statements about the importance of changes of genetic parameters

during short term selection programmes, because as we have shown the results are very much dependent on the model used. Provided gene effects are not large and gene frequencies not extreme, with reasonably high h^2 and population size, changes of genetic parameters will be considerably larger than the standard deviation of selection response after 3 or 4 cycles of selection. In these cases, the predictions made under the infinitesimal model will be fairly accurate. With low initial frequencies, as illustrated in runs 6 and 7, the effects of gene frequency changes and the generation of joint disequilibrium tend to oppose each other and the total genotypic variance changes very little during the first few cycles of selection. The detection of changes in the variance under such a model would require relaxation of selection which should expose the changes brought about by the effect of gene frequencies moving towards intermediate values.

Probably the most general important feature of this work is that it highlights the fact that the reduction in variance of quantitative traits due to the generation of joint disequilibrium should not be omitted in short-term selection studies.

CHAPTER 6

EFFECT OF DIRECTIONAL SELECTION ON QUANTITATIVE DOMINANT MODELS

Introduction

The purpose of this chapter is to study the approximations made under the infinitesimal model when the character is affected by a finite number of loci exhibiting non-additive gene action. As we did in the previous chapter, we shall concentrate attention upon short term responses and we shall ignore problems of drift and inbreeding depression.

If a character is assumed to be determined by an effectively infinite number of independent loci, the regression of offspring on parent is linear and homoscedastic whether or not there is dominance. In this case the expected response to selection for high or low value of the trait is perfectly symmetrical and the amount of disequilibrium generated is the same in each direction (Bulmer, 1971). With a finite number of loci, the presence of dominance can cause the regression to be non-linear and the variability about the regression line to be no longer constant. As is well known, the expected response to up and down selection of the same intensity is no longer symmetrical even if gene effects are small (i.e. Kojima, 1961) and furthermore, as we shall show, the amount of joint disequilibrium generated in each direction is different in magnitude. In other words, expression (3.22) which we have used previously involving second order terms in a/σ is no longer accurate in the presence of dominance. In fact as we have shown in previous chapters, this expression is only accurate enough with additive models at intermediate frequencies. Extreme frequencies cause significant departures between observed and predicted results, this being accentuated with larger gene effects.

In order to have a theoretical framework against which we can interpret the simulation results, we shall first briefly review the asymmetry in mean response when selection operates on a dominant model. We shall then study the changes in the genotypic variance due to gene frequency changes and due to the generation of joint disequilibrium.

Asymmetry in Response in a First Cycle of Selection

This problem has been recently studied by Robertson (1977c) and Mäki-Tanila (1980). Following Robertson we can study the relationship between progeny values and parental values at two different stages: (i) the change in gene frequency in the parents for a given phenotypic change, and (ii) the functional relationship between offspring mean and gene frequency. Consider the following single locus model,

	A_1A_1	A_1A_2	A_2A_2
Frequency	p^2	$2p(1-p)$	$(1-p)^2$
Value	a	d	-a

Assuming normality it can be shown that the change in frequency of the A_1 allele after one cycle of selection can be approximated by the following expression:

$$\Delta p = \frac{i}{\sigma} \alpha p(1-p) + \frac{ix_T}{2\sigma^2} (\alpha^2 p(1-p)(1-2p) - 4adp^2(1-p)^2) \qquad (6.1)$$

where α is the average effect of a gene substitution at the locus, equal to $a + d(1-2p)$ (Falconer, 1960). This expression reduces to the one obtained by Latter (1965) when $d = 0$. Notice that the second order

term vanishes when the intensity of selection is 50% or, with complete dominance (d = a), when initial gene frequencies are equal to $1 - 1/\sqrt{2}$ (i.e. about 1/3). At this value of gene frequencies, the genotypic distribution is symmetrical with the two genotypic values equally frequent.

With this model, the population mean, M, is $M(p) = -a(1-2p) + 2dp(1-p)$. If the gene frequency changes from p to $p + \Delta p$, we can expand in a Taylor series to get,

$$M(p+\Delta p) = M(p) + \Delta p\, M'(p) + \tfrac{1}{2}\Delta p^2 M''(p) + \ldots$$

$$= M(p) + 2\Delta p\alpha - 2\Delta p^2 d \qquad (6.2)$$

This shows that the relationship between gene frequency and mean, when both sexes are selected, is not linear and therefore the response to selection is always asymmetrical. If dominance is complete, the second order term in (6.1) is negative (unless gene frequencies are smaller than $1-1/\sqrt{2}$) and so is the term in Δp^2 in (6.2). Prediction of selection response using the linear term in (6.1) will tend to overestimate the expected response. When selection is for a recessive, the second order term in (6.1) is positive (unless initial frequencies are very low or very high) and so is the term in Δp^2 in (6.2). In this case the prediction of selection response based on the linear term in (6.1) will underestimate the expected response.

If selection operates in one sex only, males say, the mean can be written,

$$M(pm) = -a(1-pm-p) + d(pm+p-2pm\,p)$$

where pm is the frequency in males and p is the frequency in females. Therefore,

$$\frac{dM(pm)}{dpm} = a + d(1-2p) \ ,$$

$$\frac{d^2M(pm)}{dpm^2} = 0 \ .$$

Since the second derivative is zero, the expected response to selection in the first generation will be symmetrical, or equivalently, the regression of offspring on single parent will be linear, when the second order term in (6.1) is zero.

We shall now proceed to show some numerical results in order to illustrate the consequences of the presence of dominance on gene frequency changes. Table 6.1 shows observed and predicted values of gene frequency changes for different initial gene frequencies and proportionate effects. The exact results, $\Delta p(E)$, are obtained from the technique described in the Appendix. $\Delta p(1)$ refers to predictions made using the first order term in (6.1), whilst $\Delta p(2)$ corresponds to results obtained using (6.1).

When gene effects are small the predictions made using $\Delta p(2)$ are in excellent agreement with exact results, provided gene frequencies are not too high. In agreement with theoretical expectations, the first order term tends to overestimate the change in gene frequency when dominance is complete, though observed and predicted results agree fairly well at low initial frequencies. At higher values of $\frac{\alpha}{\sigma}$ observed and predicted results are in poor agreement, particularly the predictions made using the first order term. At intermediate gene frequencies, $\Delta p(1)$ overestimates the observed results by almost 40%.

TABLE 6.1: Observed and predicted values of Δp for a single locus model with complete dominance. $\Delta p(E)$: exact results obtained from numerical integration; $\Delta p(1)$: predicted values using first order approximation; $\Delta p(2)$: predicted values using expression (6.1).

$a = \alpha$		Initial Gene Frequency				
		0.1	0.3	0.5	0.7	0.9
0.20	$\Delta p(E)$	0.027	0.058	0.063	0.044	0.009
	$\Delta p(1)$	0.025	0.059	0.070	0.059	0.025
	$\Delta p(2)$	0.026	0.058	0.064	0.045	0.004
0.50	$\Delta p(E)$	0.071	0.138	0.134	0.068	0.010
	$\Delta p(1)$	0.063	0.147	0.182	0.147	0.063
	$\Delta p(2)$	0.070	0.146	0.146	0.062	-

These results illustrate the fact that predictions of expected selection response, in the presence of dominance, are likely to be significantly less accurate than for completely additive models. We shall pursue this point further in the simulation work that follows.

The Two Locus Dominant Model.

Gene frequency changes.

Consider the following model of two loci, A and B, with two alleles at each locus,

	A_1A_1	A_1A_2	A_2A_2	B_1B_1	B_1B_2	B_2B_2
Frequency	p^2	$2p(1-p)$	$(1-p)^2$	q^2	$2q(1-q)$	$(1-q)^2$
Value	a_1	d_1	$-a_1$	a_2	d_2	$-a_2$

The population mean, M, is:

$$M = -a_1(1-2p) + 2d_1p(1-p) - a_2(1-2q) + 2d_2q(1-q) .$$

With Hardy-Weinberg equilibrium, the total genotypic variance, VG, under this model, can be shown to be

$$VG = 2p(1-p)\alpha_1^2 + 2q(1-q)\alpha_2^2 + (2p(1-p)d_1)^2 + (2q(1-q)d_2)^2$$

$$+ 4\alpha_1\alpha_2 D + 8d_1d_2D^2 \qquad (6.3)$$

D is the disequilibrium parameter, as before and σ_i (i=1,2) is the average effect of a gene substitution at the i^{th} locus. The first two terms represent the equilibrium additive variance, Vg; the second two terms represent the dominance variance, VD. The fifth term in (6.3) is the covariance of average effects of genes between loci within parental contributions, CLW, and the last term can be shown to be equal to twice the covariance of dominance deviations between loci. We can then write,

$$VG = Vg + VD + CLW + Cdd \qquad (6.4)$$

The change in frequency of allele A_1, say, is given by,

$$\Delta p = \frac{1}{\bar{W}}\{f_1(W_1-\bar{W}) + f_2(W_2-\bar{W})\} \tag{6.5}$$

where, as before,

$$W_i = Q\{1 + \frac{i}{\sigma}(\bar{X}_i-M) + \frac{ix_T}{2\sigma^2}(Vw_i + (\bar{X}_i-M)^2)\} \tag{6.6}$$

$$\bar{W} = Q(1 + \frac{ix_T}{2\sigma^2}\{\Sigma Vw_i + \Sigma(\bar{X}_i-M)^2 f_i\}) = Q(1 + \frac{ix_T}{2\sigma^2} VG) \tag{6.7}$$

The variance within gametic phases, Vw_i, can be shown to be given by the following expressions,

$$\begin{aligned}
Vw_1 &= (a_1-d_1)^2p(1-p) + (a_2-d_2)^2q(1-q) + 2D(a_1-d_1)(a_2-d_2)\\
Vw_2 &= (a_1-d_1)^2p(1-p) + (a_2+d_2)^2q(1-q) + 2D(a_1-d_1)(a_2+d_2)\\
Vw_3 &= (a_1+d_1)^2p(1-p) + (a_2-d_2)^2q(1-q) + 2D(a_1+d_1)(a_2-d_2)\\
Vw_4 &= (a_1+d_1)^2p(1-p) + (a_2+d_2)^2q(1-q) + 2D(a_1+d_1)(a_2+d_2)
\end{aligned} \tag{6.8}$$

All these expressions reduce to $\frac{1}{2}(Vg + CLW)$ if $d_1 = d_2 = 0$. Using (6.8), (6.7) and (6.6) and substituting in (6.5), it can be shown that the second order approximation for the change in gene frequency of the two locus model with dominance is,

$$\begin{aligned}
\Delta p = \frac{i}{\sigma}(\alpha_1 p(1-p) + \alpha_2 D) + \frac{ix_T}{2\sigma^2}(&p(1-p)(1-2p)\alpha_1^2 - 4a_1d_1p^2(1-p)^2\\
&+ \alpha_2^2(1-2q)D - 4a_2d_2q(1-q)D + 2\alpha_1\alpha_2(1-3p)D - 8d_1d_2pD^2\\
&+ 2D\{(a_1-d_1)(a_2+d_2)p - 2d_2(a_1-d_1)(pq+D)\})
\end{aligned} \tag{6.9}$$

This expression reduces to (6.1) when D = 0 and further it reduces to (5.9) when $d_1 = d_2 = 0$. Ignoring the second order term in (6.9) and linearising the relationship between genetic mean and gene frequency, the response to selection from loci A and B is seen to be,

$$R = \frac{i}{\sigma}(2p(1-p)\alpha_1^2 + 2q(1-q)\alpha_2^2 + 4\alpha_1\alpha_2 D) = ih^2\sigma \qquad (6.10)$$

It is clear that this is strictly an asymptotic result which should hold provided gene effects are very small. This restriction is much more severe in the presence of dominance than under complete additivity as is illustrated in the simulation work at the end of this chapter.

The generation of joint disequilibrium.

In this section we deal with problems of asymmetry in the generation of joint disequilibrium with dominance models. Our approach is to show that third order moments in a/σ are required to explain the reduction in variance due to joint disequilibrium and that the asymptotic value given by the expression derived on the basis of the infinitesimal model is attained very slowly. We shall assume that the population is initially in Hardy-Weinberg and linkage equilibrium. The model we use is the two locus dominance model defined in the previous section.

From the third order Taylor series expansion of expression (3.8) it can be shown that the relative probability of selection of the ij^{th} genotype is given by:

$$\frac{W_{ij}}{\bar{W}} = \frac{Q}{\bar{W}}\{1 + \frac{i}{\sigma}(\bar{X}_{ij}-M) + \frac{ix_T}{2\sigma^2}(\bar{X}_{ij}-M)^2 + \frac{i(x_T^2-1)}{6\sigma^3}(\bar{X}_{ij}-M)^3\} \qquad (6.11)$$

Replacing (6.11) in (5.2) which describes the relative selective advantage of the i^{th} gametic phase, $W_i/\bar{W}$, it can be shown that, to third order terms:

$$\frac{W_i}{\bar{W}} = \frac{Q}{\bar{W}}\{1+ \frac{i}{\sigma}(\bar{X}_i-M) + \frac{ix_T}{2\sigma^2}(Vw_i + (\bar{X}_i-M)^2 + \frac{i(x_T^2-1)}{6\sigma^3}(\mu_{i(3)}$$

$$+ 3(\bar{X}_i-M)\ Vw_i + (\bar{X}_i-M)^3)\} \qquad (6.12)$$

where $\bar{X}_i$ is the mean of the i^{th} gametic phase; Vw_i is given by (6.8) and $\mu_{i(3)}$ is the third moment from the mean of the i^{th} gametic phase, defined as follows:

$$\mu_{i(3)} = \sum_j (\bar{X}_{ij}-\bar{X}_i)^3 f_j.$$

Assuming initial linkage equilibrium, the disequilibrium parameter of selected individuals and their offspring, $D^{(1)}$, is

$$D^{(1)} = \frac{f_1 f_4}{\bar{W}}(W_1W_4 - W_2W_3) \qquad (6.13)$$

Substituting (6.12) in (6.13) we obtain,

$$D^{(1)} = \frac{i^2}{\sigma^2}\{(\bar{X}_1-M)(\bar{X}_4-M) - (\bar{X}_2-M)(\bar{X}_3-M)\}f_1f_4 + \frac{ix_T}{2\sigma^2}\{(\bar{X}_1-M)^2$$

$$- (\bar{X}_2-M)^2 - (\bar{X}_3-M)^2 + (\bar{X}_4-M)^2\}f_1f_4 + \frac{i^2x_T}{2\sigma^3}\{(\bar{X}_1-M)(\bar{X}_4-M)^2$$

$$+ (\bar{X}_1-M)^2(\bar{X}_4-M) - (\bar{X}_2-M)(X_3-M)^2 - (\bar{X}_2-M)^2(\bar{X}_3-M) + Vw_1(\bar{X}_4-M)$$

$$- Vw_2(\bar{X}_3-M) - Vw_3(\bar{X}_2-M) + Vw_4(\bar{X}_1-M)\}f_1f_4 + \frac{i(x_T^2-1)}{6\sigma^3}\mu_{1(3)}$$

$$- \mu_{2(3)} - \mu_{3(3)} + \mu_{4(3)} + 3(Vw_1(\bar{X}_1-M) - Vw_2(\bar{X}_2-M) - Vw_3(\bar{X}_3-M)$$

$$+ Vw_4(\bar{X}_4-M)) + (\bar{X}_1-M)^3 - (\bar{X}_2-M)^3 - (\bar{X}_3-M)^3 + (\bar{X}_4-M)^3\} f_1f_4 \qquad (6.14)$$

Expression (6.14) as it stands is clearly not very informative. The terms of order $(\frac{a}{\sigma})^2$ can be shown to equal

$$\frac{-i(i-x_T)}{\sigma^2}\alpha_1 p(1-p)\alpha_2 q(1-q),$$

of the same form as (3.20), obtained by Hill and Robertson (1966) who assumed additive gene action.

A little insight into the third order term can be obtained by assuming that dominance is complete ($d_i = a_i$) and that gene effects and frequencies are the same at both loci. Under these assumptions, (6.14) reduces to

$$D^{(1)} = \frac{-i(i-x_T)}{\sigma^2}\alpha^2p^2(1-p)^2 + \frac{ix_T(x_T-i)-1}{\sigma^3}\,8a^3p^2(1-p)^2$$

$$\{(1-p)(1-2p) - p\} \qquad (6.15)$$

The third order term in (6.15) vanishes when initial gene frequencies are equal to $1-1/\sqrt{2}$, in other words, when the genotypic distribution is perfectly symmetrical. For other values of p, the sign of this

term depends critically on the selection intensity, in marked contrast with the first term in (6.15). When the highest 20% are selected, if gene frequencies are low, the absolute value of $D^{(1)}$ using the second order approximation is underestimated whilst the opposite holds if gene frequencies are higher than $1-1/\sqrt{2}$. With a completely recessive model such that d = -a, the third order term can be shown to equal

$$\frac{ix_T(x_T-i) - 1}{\sigma^3} 8a^3p^2\{p(1-2p) + (1-p)\} ,$$

which vanishes when $p = 1/\sqrt{2}$. If one locus shows complete dominance and the other complete recessivity, the third order term becomes

$\frac{ix_T(x_T-i) - 1}{\sigma^3} 8a^3p(1-p)(1-2p)$. In other words it vanishes when the genotypic distribution is symmetrical. In the absence of dominance, with a completely additive model of equal effects and frequencies, the third order approximation for the disequilibrium parameter generated after a single cycle of selection is given by the following expression

$$D^{(1)} = \frac{-i(i-x_T)}{\sigma^2} a^2p^2(1-p)^2 + \frac{ix_T(x_T-i) - 1}{\sigma^3} a^3p^2(1-2p)$$

When gene frequencies are initially smaller than 0.5, the second term is negative and the opposite holds if gene frequencies are higher than 0.5. We therefore expect larger absolute values of disequilibrium at low frequencies than at high frequencies. At intermediate frequencies, when the genotypic distribution is symmetrical, the second term vanishes and we expect the second order approximation to describe the process with good accuracy. The

problems of asymmetry of disequilibrium with an additive model, raised in Chapters III and V are formally explained in terms of the above expression.

We now produce some numerical results to illustrate some of the points of this section. We assume a model of complete dominance and we study the effect of varying the number of loci affecting the trait and the initial gene frequencies. The equilibrium additive variance and the total phenotypic variance are fixed for all the number of loci and gene frequency combinations, this being achieved by varying the environmental variance for different values of the dominance variance. The results are shown in Table (6.2) where we compare observed (O) and predicted (P) values using expression (6.15). Observed results are obtained by numerical integration as described in the Appendix.

The results show that unless gene frequencies are in the vicinity of $1 - 1/\sqrt{2}$, expression (3.22) is indeed very inaccurate. Furthermore, the asymptotic value is attained very slowly indeed. The predictions made using (6.15) are in excellent agreement with exact results provided gene frequencies are not much higher than 0.5. At high gene frequencies, particularly with small number of loci, the predicted values are very inaccurate. This is probably not entirely surprising since with complete dominance and very high frequencies the genotypic distribution should be markedly irregular and probably higher order moments should be invoked to describe the process more accurately.

The reason for the lack of agreement between observed and predicted results at high gene frequencies can be perhaps clearly

TABLE 6.2:

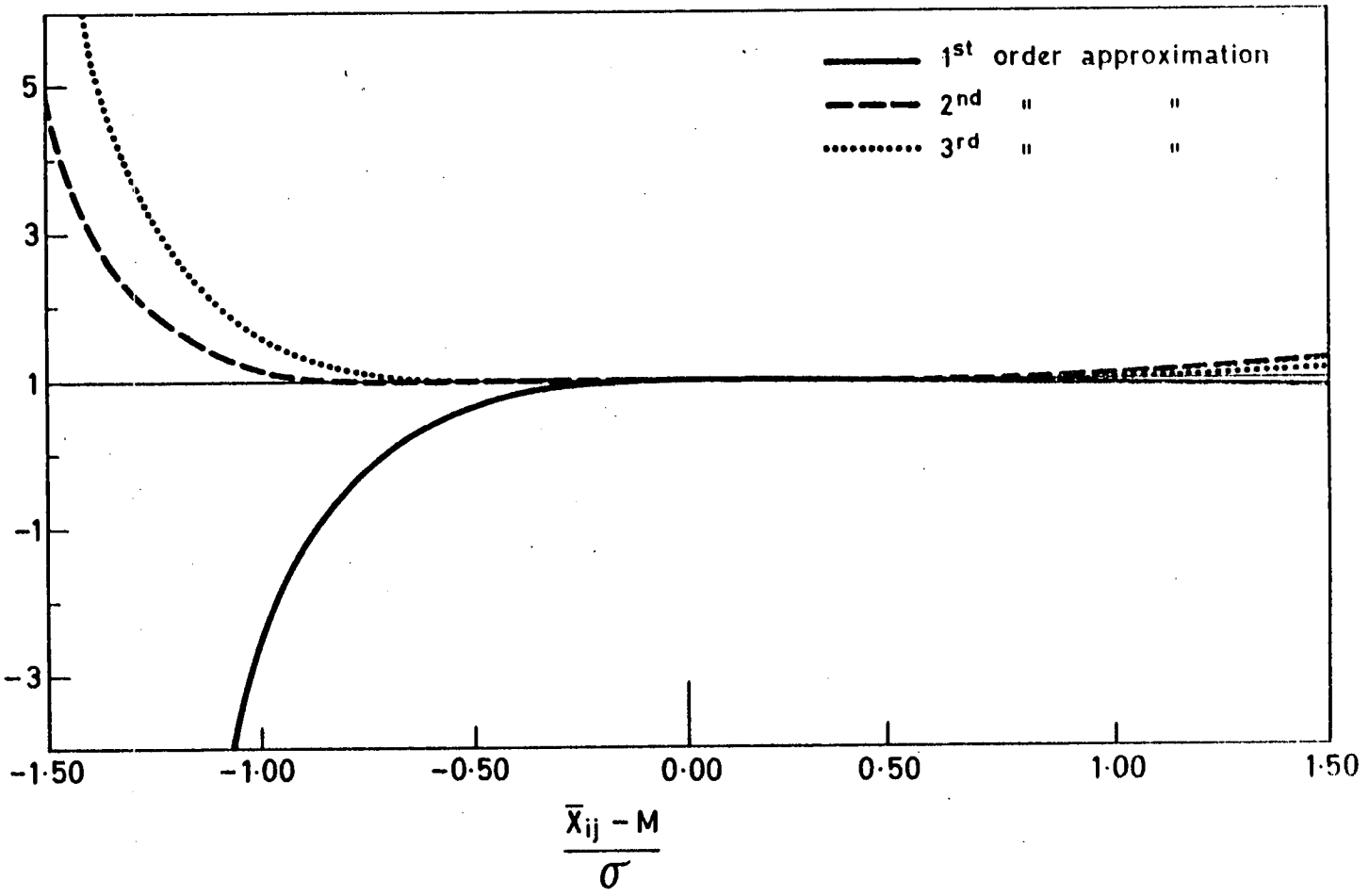
Observed (O) and predicted (P) values of joint disequilibrium generated in a first cycle of selection with initial Hardy-Weinberg and linkage equilibrium. We assume a completely dominant model with $Vg = 4$, $\sigma^2 = 40$ and the top 20% are selected. The asymptotic value of joint disequilibrium for n (number of loci) tending to infinity is calculated from: $CLW = -\frac{1}{2}\frac{(i(i-x_T)}{\sigma^2} Vg^2 = -0.156$.

		Initial Gene Frequencies				
n		0.1	0.3	0.5	0.7	0.9
	O	-0.206	-0.139	-0.091	-0.034	-0.000
10	P	-0.206	-0.143	-0.090	0.020	0.861
	(α/σ)	0.24	0.15	0.14	0.15	0.24
	O	-0.190	-0.149	-0.119	-0.072	-0.000
30	P	-0.190	-0.151	-0.120	-0.054	0.404
	(α/σ)	0.14	0.09	0.08	0.09	0.14
	O	-0.176	-0.154	-0.137	-0.106	-0.010
100	P	-0.176	-0.155	-0.137	-0.101	0.146
	(α/σ)	0.07	0.05	0.04	0.05	0.07
	O	-0.156	-0.156	-0.156	-0.151	-0.128
10,000	P	-0.156	-0.156	-0.155	-0.151	-0.126
	(α/σ)	0.01	0.00	0.00	0.00	0.01

understood by studying the accuracy of the selective values obtained from the first, second and third order expansion of the normal distribution, for different standardised deviations from the population mean. This is illustrated in Figure 6.1. For values of $(\bar{X}_{ij}-M)/\sigma$ larger than 1, the third order approximation overestimates the exact selective value by about 8%. The bias is relatively more severe for negative deviations from the mean. At $(\bar{X}_{ij}-M)/\sigma = -1$, the overestimation is of about 54%. This has drastic consequences in the prediction of changes in gamete frequencies and the amount of disequilibrium generated. For example, to take an extreme situation, with 10 loci and gene frequencies of 0.9, the deviation of the genotypic value of the double recessive is of -4.67 standard deviations which leads to a very high overestimate of its selective value. Changes in the gamete frequencies are predicted very inaccurately, so much so that the predicted disequilibrium parameter takes a positive value.

The general conclusion that we draw from this theoretical analysis is that predictions of joint disequilibrium generated by selection based on the infinitesimal model are unlikely to hold in the presence of dominance, unless the number of loci affecting the trait is assumed to be very large. In subsequent cycles of selection, predictions of joint disequilibrium are likely to become more inaccurate.

FIGURE 6.1: Selective values (W_{ij}) approximated using first, second and third order expansions of the normal distribution, expressed as a proportion of selective values obtained from numerical integration of the normal distribution. (Proportion selected: 20%).



Simulation Studies.

The Computer Programme and Genetic Models.

The basic structure of the programme is the same as in the case of the additive models, the minor alteration now being the incorporation in the programme of an arbitrary value of the dominance deviation, d, at each locus.

The partitioning of the total genotypic variance in the offspring generation merits a brief comment. We are interested in estimating the equilibrium additive variance and the degree of joint disequilibrium generated, since these are the two components involved in the prediction of the expected response to selection assuming that the values of offspring and parent follow a bivariate normal distribution (expression 6.10). From a practical point of view we want to know the efficiency with which this expression predicts the expected response to selection, how accurately we can predict the way its components change as selection proceeds and provide satisfactory explanations in case observed and predicted results are not in good agreement.

In the offspring generation, before selection operates, we calculate the gene frequency at each locus. Assuming Hardy-Weinberg equilibrium, we obtain an estimate of the average effect of a gene substitution at each locus and this gives us a means of obtaining an estimate of the equilibrium additive variance, Vg. The assumption of Hardy-Weinberg equilibrium is a reasonable one since selected individuals mate at random and as was mentioned in the previous chapter, departures from equilibrium due to finite population size are very small, of the order of 1/2N. The joint disequilibrium, CLW, is

estimated by subtracting the equilibrium additive variance from the variance of breeding values between individuals.

This approach was checked in a different version of the programme, relaxing the assumption of Hardy-Weinberg equilibrium. Genotype frequencies were obtained for each locus and average effects of alleles obtained by least squares (i.e. Kempthorne, 1957). The variance of breeding values obtained in this way was compared to the one obtained before and the results were virtually unchanged.

The genetic parameters of the models studied are summarized in Table 6.3. In all cases, the highest 40 out of 200 individuals of each sex were selected and all runs assume 30 loci with free recombination. When d = a, dominance is complete; when d = -a there is complete recessivity and when d = 0 there is no dominance. The runs are identified by their gene frequency and degree of dominance. For example, run 4 is designated 0.5/d = a, -a, implying that the model assumes initial gene frequencies of 0.5, 15 loci show complete dominance and the remaining 15 loci show complete recessivity. The parameters in the models were chosen to illustrate and check the theory that has been developed in the earlier sections of this chapter. Thus, we have a range of gene frequencies which will indicate the problems of asymmetry in mean and in the generation of disequilibrium as selection proceeds.

TABLE 6.3: Initial Genetic Parameters of the Models Studied

Run	Designation of Run	Gene Frequency	Proportionate Effect	Degree of Dominance	Vg	h^2
1	0.1/d=a	0.1	0.27	d=a	17.50	0.40
2	0.293/d=a	0.293	0.20	d=a	24.85	0.50
3	0.5/d=a	0.5	0.18	d=a	15.00	0.50
4	0.5/d=a_1-a	0.5	0.18	15 loci,d=a 15 loci,d=-a	15.00	0.50
5	0.5/d=a,0	0.5	0.18	15 loci,d=a 15 loci,d=0	15.00	0.50
6	0.7/d=a	0.7	0.14	d=a	4.54	0.24

Results and Discussion

The results for the first cycle of selection for the different models are shown in Table 6.4.

In contrast with the predictions of selection response and joint disequilibrium, the change in gene frequency, which is reflected in the value of the equilibrium additive variance, is predicted fairly accurately in all models. The lack of agreement between observed and predicted values of joint disequilibrium and selection response is as expected from theoretical considerations. At very low initial gene frequencies we expect, in the case of a completely dominant model, to underestimate the value of joint disequilibrium attained after a first cycle of selection. This is illustrated in model 0.1/d=a. The observed value of joint disequilibrium when selection was for the low value of the character of the same proportion was -2.26 ± 0.21. When initial gene frequencies are higher than $1-1/\sqrt{2}$, the predicted joint disequilibrium should be an overestimate. This is shown in models 0.5/d=a and 0.7/d=a. In the former, the predicted result overestimates the observed result by a factor of 2 and in the latter by a factor of almost 9. When the same proportion is selected for the low value of the character, the observed joint disequilibrium at generation 1 is of -4.66 ± 0.31 and -1.01 ± 0.09 for models 0.5/d=a and 0.7/d=a respectively. At intermediate gene frequencies, when only half of the loci show complete dominance (model 0.5/d=a,0) the discrepancy between observed and predicted results is smaller. In the case of no directional dominance with intermediate gene frequencies, the change

TABLE 6.4: Observed (O) and Predicted (P) values of the genetic parameters at generation 1 after a first cycle of selection. The base population is assumed to be in Hardy-Weinberg and linkage equilibrium. The predicted values for the different parameters are obtained from the following expressions

$$R = ih^{2(O)}\sigma^{(O)}$$

$$CLW = -\tfrac{1}{2}\frac{i(i-x_T)}{\sigma^{2(O)}}\ Vg^{2(O)}$$

$$Vg = 2\sum_i p_{(1)i}(1-p_{(1)i})\ \alpha^2_{(1)i}, \text{ where } p_{(1)} = p_{(0)} + \Delta p, \ \Delta p$$

being predicted using (6.1).

$g_{(3)}$ refers to the skewness of the genotypic distribution at generation zero.

Observed results are the average of 30 replicates.

Model		Vg	CLW	R	g(3)
0.1/d=a	O	20.67±0.07	-3.82±0.36	3.86±0.09	0.30±0.02
	P	20.61	-2.75	3.72	
0.293/ d=a	O	22.75±0.05	-4.33±0.36	4.84±0.09	0.02±0.02
	P	23.02	-4.82	4.92	
0.5/d=a	O	11.46±0.07	-1.46±0.20	3.35±0.08	-0.21±0.02
	P	11.52	-2.93	3.83	
0.5/d=a, -a	O	15.34±0.07	-2.98±0.24	3.95±0.09	0.01±0.02
	P	15.05	-2.93	3.83	
0.5/d=a,0	O	13.14±0.05	-2.15±0.20	3.60±0.09	-0.14±0.02
	P	13.16	-2.93	3.83	
0.7/d=a	O	3.30±0.04	-0.05±0.09	1.31±0.05	-0.46±0.02
	P	3.31	-0.43	1.46	

in gene frequency per unit change in phenotype is linear and so is the functional relationship between gene frequency and genotypic mean. Furthermore, the genotypic distribution is symmetrical and we therefore expect good agreement between observed and predicted results of both, selection response and joint disequilibrium. This is confirmed by the results of model 0.5/d=a,-a. In model 0.293/d=a, the genotypic distribution is symmetrical, the second order term in Δp is zero but the functional relationship between gene frequency and mean is not linear. We expect symmetry in the generation of joint disequilibrium and slightly higher observed than predicted response to selection. The small discrepancies observed are probably due to sampling.

The differences between observed and predicted genotypic means after a first cycle of selection are all consistent with theoretical expectations. The heritability used in the prediction of response corresponds to the expected value of the estimate that one would obtain by regressing offspring means on selected parents, assuming the relationship to be linear, when in fact it is not. From (6.2) the expected response is given by $2\Sigma_i(\Delta p_i \alpha_i - \Delta p_i^2 d_i)$, which, when evaluated to second order terms gives us an indication of the degree of bias to be expected using the linear term only. With the exception of model 0.5/d=a,-a, the expected response to a first cycle of selection is not linear but the departures from linearity are small in models 0.1/d=a and 0.293/d=a. Of the models in Table 6.4 those which show the highest degree of asymmetry are 0.5/d=a and 0.7/d=a as expected from theory.

In subsequent cycles of selection the situation becomes considerably more complicated due to the presence of joint disequilibrium. If selection is carried out in both directions the different amount of disequilibrium generated in each direction contributes to the degree of asymmetry of selection response. If we assume a completely dominant model with initial gene frequencies larger than $1 - 1/\sqrt{2}$, from expression (6.15) we expect the reduction in variance due to disequilibrium to be larger in the low line than in the high line. This is also related to the fact that the equilibrium additive variance is a function of $p(1-p)^3$, which is larger at low values of p. It is difficult and probably not very meaningful to try to separate this source of asymmetry from the one discussed previously solely in terms of gene frequencies on the assumption of linkage equilibrium. However, in a simple minded way, ignoring disequilibrium, we expect the single generation realised heritabilities to decrease as gene frequencies move towards high values and to increase as they move towards $1 - 1/\sqrt{2}$. The effect of the different degree of disequilibrium generated in both directions is, under this model, to reduce the expected degree of asymmetry of selection response.

Table 6.5 shows the observed single generation realized heritabilities, joint disequilibria and correlation of gene frequencies for selection in the high and low direction for models 0.1/d=a, 0.5/d=a, 0.7/d=a. The most conspicuous feature of the results is the degree of asymmetry of response to selection and joint disequilibrium between the high and low selected replicates within each model. In model 0.1/d=a, the equilibrium additive variance increases

TABLE 6.5: Montecarlo simulation results of single generation realized heritabilities, joint disequilibria and correlation of gene frequencies within chromosomes for 3 cycles of selection. (H) refers to the value of the parameters estimated in the high selection and (L) in the low selection. The values of h^2 at the top of the table correspond to the ratio of the equilibrium additive variance to total phenotypic variance in the conceptually infinite base population. The values in brackets correspond to the ratio of joint disequilibrium to equilibrium additive variance in the generation in question. The standard error of each heritability estimate is 0.01.

MODEL	0.1/d=a h^2=0.40				0.5/d=a h^2=0.50				0.7/d=a h^2=0.24			
t	$h^2_{(H)}$	$h^2_{(L)}$	$CLW_{(H)}$	$CLW_{(L)}$	$h^2_{(H)}$	$h^2_{(L)}$	$CLW_{(H)}$	$CLW_{(L)}$	$h^2_{(H)}$	$h^2_{(L)}$	$CLW_{(H)}$	$CLW_{(L)}$
0	0.41	0.38	0.00±0.26	-0.02±0.23	0.46	0.55	0.07±0.14	-0.34±0.23	0.21	0.28	0.03±0.05	0.03±0.07
1	0.39	0.29	-3.82±0.37 (-0.185)	-2.26±0.23 (-0.175)	0.36	0.55	-1.46±0.19 (-0.128)	-4.66±0.31 (-0.250)	0.16	0.32	-0.09±0.06 (-0.027)	-1.02±0.09 (-0.153)
2	0.38	0.24	-5.07±0.37 (-0.224)	-1.52±0.16 (-0.161)	0.29	0.56	-1.20±0.15 (-0.137)	-5.98±0.31 (-0.282)	0.11	0.35	-0.04±0.06 (-0.018)	-1.78±0.22 (-0.200)
3	0.37	0.18	-5.57±0.39 (-0.237)	-0.58±0.18 (-0.091)	0.24	0.58	-0.84±0.07 (-0.126)	-7.50±0.37 (-0.331)	0.09	0.41	-0.02±0.06 (-0.009)	-2.22±0.22 (-0.198)

in the up selection from 17.5 to 23.6 as gene frequencies move towards $1 - 1/\sqrt{2}$. This effect is counteracted by the sharp increase in joint disequilibrium and as a result of these two factors, the single generation realized heritability is virtually unchanged. This is a situation where observed and predicted responses to short term selection would be in excellent agreement, though a rather substantial change in genetic parameters is taking place during the course of selection. In the low lines of this model both selection response and the reduction in variance due to joint disequilibrium are considerably smaller than in the high lines. In this case, the equilibrium additive variance falls abruptly from its original value of 17.5 to a value of 4.0 squared units after 4 cycles of selection.

In the other two models where gene frequencies are initially higher than $1 - 1/\sqrt{2}$, the equilibrium additive variance increases in the low lines and decreases in the high lines. As expected from our theoretical analysis, the amount of disequilibrium generated is much larger in the low selected replicates, this effect being considerably accentuated by the increase in the equilibrium additive variance. Model 0.7/d=a is a good example of the low rate of response obtained with favourable dominant loci at high frequencies.

The values in brackets in Table 6.5 correspond to the correlation of gene frequencies in each generation. This is a parameter which is less affected by gene frequency changes than CLW. The difference between the correlation of gene frequencies in the high and low lines within each model is relatively smaller than the covariance of allelic effects but the asymmetry persists.

The conclusion to be drawn from the theoretical analysis and the simulation study is that the pattern of selection response and the degree of joint disequilibrium generated are very much dependent on the genetic model. This model dependence seems to be more accentuated than in the case of additive models. A general statement can be made though: in the presence of directional dominance, predictions of short-term response and particularly of joint disequilibrium from base population parameter estimates are not likely to be very precise.

CHAPTER 7

EFFECT OF SELECTION ON HERITABILITY ESTIMATED FROM INTRA-CLASS CORRELATIONS

Introduction

When records are available on two generations, heritability can be estimated using regression of offspring on parent or maximum-likelihood procedures (Thompson, 1976) and the estimates are not affected by selection of parents. If records are available on one generation only, heritabilities are usually estimated from intra-class correlation among sibs obtained from an analysis of variance. As is well known, selection of parents introduces a bias in this estimator (Reeve, 1953; Morley, 1955; Brown & Turner, 1968; Robertson, 1977a; Ponzoni & James, 1978). The expected value of the heritability estimate from intra-class correlations on selected data is easily derived from procedures that were first developed by Pearson (1903) and later on extended by Cochran (1951), Finney (1956) and Tallis (1961) to mention a few.

Let X and Y be two random variables which follow a bivariate normal distribution. It then follows that the regression of one on the other is linear and homoscedastic. Assume that truncation selection is practised on X. Let r, Var(X) and Var(Y) be the correlation between X and Y, the variance of X and the variance of Y respectively before selection operates. After selection, a proportion r^2 of the variance of Y which is associated with X will be reduced by a fraction $(1-i(i-x_T))$, and the remaining fraction $(1-r^2)$ will be unaffected since it is independent of X. We can then write for the variance of Y in the selected population $(Var(Y)^{(S)})$:

$$\mathrm{Var}(Y)^{(S)} = \mathrm{Var}(Y)(1 - r^2 + (1 - i(i-x_T))r^2)$$

$$= \mathrm{Var}(Y)(1 - i(i-x_T)r^2) \qquad (7.1)$$

In a genetic context, let X be the phenotypic value and Y the genotypic value. The correlation between X and Y before selection is $\sqrt{h^{2(0)}}$, where $h^{2(0)}$ is the heritability at time 0 and therefore (7.1) can be written in terms of parameters before the operation of selection as follows:

$$VG^{(S,0)} = VG^{(0)}(1 - i(i-x_T)h^{2(0)}) \qquad (7.2)$$

where $VG^{(0)}$ and $VG^{(S)}$ are the variance of genotypic values before and after selection respectively.

Assuming initial linkage equilibrium, we have shown before that, after a first cycle of selection,

$$VG^{(S,0)} = Vg^{(S,0)} + CHW^{(S,0)} + CLW^{(S,0)} + CLB^{(S,0)} \qquad (7.3)$$

where, $CLW^{(S,0)} = CLB^{(S,0)} = -\frac{1}{2}i(i-x_T)VG^{(0)}h^{2(0)}$. Under the assumption of an infinite number of loci, $Vg^{(S,0)} = Vg^{(0)}$ and we ignore CHW relative to CLW and CLB, and therefore (7.3) reduces to (7.2). In other words, as shown by Bulmer (1971), the reduction in the genotypic variance caused by selection is due to the generation of joint disequilibrium.

Consider a full-sib family structure in which both males and females have been selected and mated at random. The variance component within families is not affected in the first cycle of selection. The variance component between full-sibs estimates one

half of the genotypic variance between chosen families and the sum of both components of variance estimates twice the gametic contribution of the selected genotypes to the offspring generation. From (7.2), the expected heritability is given by,

$$2 \times t_c = h^{2(O)} \frac{(1 - i(i-x_T)h^{2(O)})}{(1 - \frac{1}{2}i(i-x_T)h^{4(O)})} \qquad (7.4)$$

as obtained by Reeve (1953), where t_c is the intra-class correlation between full-sib families.

According to (7.4), this estimator of $h^{2(O)}$ is biased downwards, the bias being due to the generation of covariances within and between chromosomes induced by selection. Some aspects of this problem have recently been discussed by Robertson (1977a)and Ponzoni and James (1978). It is important to emphasize however that the assumptions that lead to the derivation of (7.4) imply a model of an infinite number of loci and that the base population is in Hardy-Weinberg and linkage equilibrium. Furthermore, this result is strictly valid for a single generation of selection and allowance for the bias, if at all possible (see Robertson, 1977a),should strictly not be extended beyond the first selection cycle as it has been inadvertently either suggested or carried out in the literature (Rahnefeld et al., 1963; Brown & Turner, 1968; Katz & Enfield, 1977). With finite number of loci, expression (7.4) is of questionable accuracy since it ignores gene frequency changes due to selection. If gene frequencies are not initially at intermediate values, we have shown before that the generation of disequilibrium is not symmetrical. Assuming additivity,gene frequency changes are

relatively larger if they depart from intermediate values and therefore (7.4) or (7.2) are even less accurate.

The purpose of this chapter is to investigate both theoretically and with Montecarlo methods, the problems raised by the estimation of heritability from intra-class correlations with selected data using additive models with a finite number of loci. The results in this chapter may also be relevant in the evaluation of selection programmes involving some kind of family selection. Before studying the effects of selection, we first investigate the effect of correlations between different parts of the genotype in the parental generation on intra-class correlations.

Random Mating - Effect of Disequilibria on intra-class correlations.

We first review the effect of linkage disequilibrium in the parental generation on the components of genetic variance. The disequilibrium could have arisen by chance, or selection or any other reason in the past history of the population. At time t we assume that the parental population mates at random to produce a very large number of offspring which constitute generation t+1. For simplicity, we assume a two locus additive model as described in Chapter 5 and a full-sib family structure where Var(BFS) and Var(WFS) denote the variance between and within full-sib families respectively and VE is the environmental variance assumed constant generation to generation. Avery and Hill (1979), who worked with considerably more sophisticated models, showed that

$$\begin{aligned} \text{Var(WFS)}^{(t+1)} &= \tfrac{1}{2}\text{Vg}^{(t)} + 2a_1a_2D^{(t)}(1-2c) + \text{VE} \\ \text{Var(BFS)}^{(t+1)} &= \tfrac{1}{2}\text{Vg}^{(t)} + 2a_1a_2D^{(t)} \end{aligned} \qquad (7.5)$$

It is clear from these expressions that if loci are unlinked there is no contribution from pairs of loci to the variance within full-sib families. The intra-class correlation estimated from (7.5) adequately describes the ratio of the genotypic to total phenotypic variance in the population at time t if c = 0. For any other value of c, the estimate is biased. For example, for c = $\frac{1}{2}$ and using notation of earlier chapters, we have, omitting subscript t,

$$2 \times t_c = \frac{Vg + CLW}{Vg + \frac{1}{2}CLW + VE} \qquad (7.6)$$

This result also holds for the case of heritability estimates based on intra-class correlations between half-sib families. With a half-sib family structure, the variance between full-sib families is partitioned into two independent components: the variance between half-sibs (Var(HS)) and the variance between full-sibs within half-sibs (Var(FS/HS)), where, in the case of our model assumptions,

$$Var(HS)^{(t+1)} = \tfrac{1}{4}Vg^{(t)} + a_1a_2D^{(t)}$$

$$Var(FS/HS)^{(t+1)} = \tfrac{1}{4}Vg^{(t)} + a_1a_2D^{(t)}$$

We now consider the effect of covariances of gene frequencies within loci, due to departures from Hardy-Weinberg equilibrium (D_{HW}), covariances between loci between chromosomes (D_B) and covariances between loci within chromosomes (D) on heritability estimates based on intra-class correlations between sibs. We assume that all these covariances or disequilibria are present in the parental generation and that they could have arisen by chance or non-random mating. The

parents mate at random to produce a conceptually infinite number of offspring and an analysis of variance is performed in the offspring generation.

Since we are working with an additive model, we can work with single genotypes and partition the total variance contributed by their gametes into two independent components: the variance between gametic means and the variance between gametes within genotypes. The former is one half Var(BFS) and the latter is one half Var(WFS). This approach leads to considerable algebraic simplicity.

A complete specification of the genotypic frequencies of the two locus model involves, in addition to the disequilibria involving pairs of genes, disequilibria among groups of three and four genes. Weir (1979) has referred to these various disequilibria as digenic, trigenic and quadrigenic disequilibria. We shall now briefly sketch the analysis that shows the intuitively obvious result that when we are dealing with second order moments such as variances and covariances, trigenic and quadrigenic disequilibria cancel out and we are only left with disequilibria involving pairs of loci, such as D, D_B, D_{HW}.

Consider the usual case of two loci, A and B, with alleles $A_i (i=1,2)$ and $B_j (j=1,2)$ respectively. Genotypes are formed by the union of maternal gametes, $A_i B_j$, and paternal gametes, $A_k B_\ell$, and have frequencies $P^{k\ell(p)}_{ij(m)} = P^{ij(p)}_{k\ell(m)}$. In addition to the pairwise disequilibria which has been referred to throughout this work, we must define trigenic and quadrigenic disequilibria.

With two alleles at each of two loci, there are two independent trigenic disequilibria. One involves alleles at the maternal (paternal) gamete in loci A and B and the paternal (maternal) allele

at locus A. In Weir's notation, this is symbolised, $D^{k.}_{ij}$ (= $D^{ij}_{k.}$). The other trigenic disequilibria involves alleles at the maternal (paternal) gamete in loci A and B and the paternal (maternal) allele at locus B. This is, $D^{.\ell}_{ij}$ (= $D^{ij}_{.\ell}$). With two alleles per locus there are 8 terms in each type which of course, add up to zero. It is easily shown that, $D^{1.}_{11} = -D^{2.}_{11} = -D^{1.}_{12} = D^{2.}_{12} = -D^{1.}_{21} = D^{2.}_{21} = D^{1.}_{22} = -D^{2.}_{22} = T^{A(p)}_{AB(m)}$ (= $T^{AB(p)}_{A(m)}$), referring to the fact that the trigenic disequilibria involves both loci in the maternal gamete and locus A in the paternal gamete. The same applies to the other set which we symbolise $T^{B(p)}_{AB(m)}$ (= $T^{AB(p)}_{B(m)}$).

With two alleles at each of two loci there is only one independent quadrigenic disequilibrium which we symbolise $Q^{AB(p)}_{AB(m)}$.

Table 7.1 shows the gametic output of the ten genotypes. We can write, following Weir (1979), the frequency of each genotype in terms of its constituent gene frequencies and various functions of the disequilibria involved. The gametic means, pooled by the corresponding genotype frequencies lead to one half the variance component between full-sib families and the pooled variance between gametes within genotypes leads to one half the variance within full-sib families. Following the algebra through, trigenic and quadrigenic disequilibria are seen to cancel out and the variance components are expressed in terms of second order moments.

As an illustration, consider terms involving products of allelic effects, a_1 and a_2, at loci A and B respectively. These terms only appear in the variance between gametes within the double heterozygotes. The genotypic frequencies of the coupling and repulsion

TABLE 7.1. Gametic Output of the Two Locus Additive Model.

Genotype	Frequency	Gametic Output				Mean	Variance Between Gametes Within Genotypes
		AB $\frac{1}{2}(a_1+a_2)$	Ab $\frac{1}{2}(a_1-a_2)$	aB $\frac{1}{2}(a_2-a_1)$	ab $-\frac{1}{2}(a_1+a_2)$		
AB/AB	P^{11}_{11}	1				$\frac{1}{2}(a_1+a_2)$	0
AB/Ab	P^{12}_{11}	$\frac{1}{2}$	$\frac{1}{2}$			$\frac{1}{2}a_1$	$\frac{1}{4}a_2^2$
AB/aB	P^{21}_{11}	$\frac{1}{2}$		$\frac{1}{2}$		$\frac{1}{2}a_2$	$\frac{1}{4}a_1^2$
AB/ab	P^{22}_{11}	$\frac{1}{2}(1-c)$	$\frac{1}{2}c$	$\frac{1}{2}c$	$\frac{1}{2}(1-c)$	0	$(\frac{1}{2}(a_1+a_2))^2 - ca_1a_2$
Ab/Ab	P^{12}_{12}		1			$\frac{1}{2}(a_1-a_2)$	0
Ab/aB	P^{21}_{12}	$\frac{1}{2}c$	$\frac{1}{2}(1-c)$	$\frac{1}{2}(1-c)$	$\frac{1}{2}c$	0	$(\frac{1}{2}(a_1-a_2))^2 + ca_1a_2$
Ab/ab	P^{22}_{12}		$\frac{1}{2}$		$\frac{1}{2}$	$-\frac{1}{2}a_2$	$\frac{1}{4}a_1^2$
aB/aB	P^{21}_{21}			1		$-\frac{1}{2}(a_1-a_2)$	0
aB/ab	P^{22}_{21}			$\frac{1}{2}$	$\frac{1}{2}$	$-\frac{1}{2}a_1$	$\frac{1}{4}a_2^2$
ab/ab	P^{22}_{22}				1	$-\frac{1}{2}(a_1+a_2)$	0

double heterozygotes can be written as follows: (we drop subscript m and superscript p)

$$f(AB/ab) = P_{11}^{22} = P_{22}^{11} = p(1-p)q(1-q) - (1-q)T_{AB}^{A} + qT_{AB}^{A} - (1-p)T_{AB}^{B}$$

$$+ pT_{AB}^{B} - q(1-q)D_{HW(A)} - p(1-p)D_{HW(B)} + D_{HW(A)}\ D_{HW(B)}$$

$$+ (1-p)(1-q)D + pqD - (1-p)qD_{B} - p(1-q)D_{B} + (D)^{2} + (D_{B})^{2} + QD_{11}^{22} .$$

$$f(Ab/aB) = P_{12}^{21} = P_{21}^{12} = p(1-p)q(1-q) + qT_{AB}^{A} - (1-q)T_{AB}^{A} - (1-p)T_{AB}^{B}$$

$$+ pT_{AB}^{B} - p(1-p)D_{HW(B)} - q(1-q)D_{HW(A)} + D_{HW(A)}\ D_{HW(B)} - p(1-q)D$$

$$- (1-p)qD + pqD_{B} + (1-p)(1-q)D_{B} + (D)^{2} + (D_{B})^{2} + QD_{12}^{21} .$$

From Table 7.1 after some simplification, terms involving a_1a_2 are of the form $\frac{1}{2}a_1a_2D(1-2c) - \frac{1}{2}a_1a_2D_B(1-2c)$ and all other higher order disequilibria cancel out. Following the algebra through, the following partitioning of twice the gametic variance is obtained,

$$\text{Var(WFS)}^{(t+1)} = \tfrac{1}{2}(Vg^{(t)} - CHW^{(t)}) + \tfrac{1}{2}CLW^{(t)}(1-2c) - \tfrac{1}{2}CLB^{(t)}(1-2c) + V_E$$

$$\text{Var(BFS)}^{(t+1)} = \tfrac{1}{2}(Vg^{(t)} + CHW^{(t)}) + \tfrac{1}{2}CLW^{(t)} + \tfrac{1}{2}CLB^{(t)}$$

$$\text{Total} : \quad Vg^{(t)} + CLW^{(t)}(1-c) + CLB^{(t)}c + V_E \qquad (7.8)$$

The total phenotypic variance in the parental generation can be shown to be given by,

$$VP^{(t)} = (Vg + CHW + CLW + CLB + V_E)^{(t)}$$

This quantity can differ considerably from the estimate given by (7.8).

Comparison of Var(BFS) with(7.7)shows clearly that the former estimates one half of the genotypic variance between chosen sires. The total variance is an estimate of the parental contribution to the next generation,(assuming random union of gametes) and therefore the effect due to Hardy-Weinberg disequilibrium does not come into it. With free recombination, CLW and CLB does not affect the variance within families but they have opposite effects on it if $c \neq \frac{1}{2}$.

In a large population, one cycle of random mating causes both CLB and CHW to vanish and therefore, assuming $c = \frac{1}{2}$ the heritability estimate reduces to (7.6).

Selection of Parents - Effect on Intra-class Correlations.

Theory.

In this section we assume that in the parental generation at generation t, before selection, there are no covariances between chromosomes (i.e. CLB = CHW = 0) but there may be linkage disequilibrium. Truncation selection is practised among the parents and they mate at random to produce a conceptually infinite number of offspring. Thus, at generation t, before parents are selected, the genotypic variance is given by: $VG^{(t)} = Vg^{(t)} + CLW^{(t)}$. After selection, the genotypic variance becomes:

$$VG^{(s,t)} = Vg^{(s,t)} + CLW^{(t)} + CLW_f{}^{(t)} + CLB^{(s,t)} + CHW^{(s,t)},$$

where $Vg^{(s,t)} = Vg^{(t+1)}$ reflects gene frequency changes; $CLW^{(t)}$ +

$CLW_f^{(t)} = CLW^{(s,t)}$ is the total allelic covariance within chromosomes, $CLW_f^{(t)}$ reflecting the fresh disequilibrium generated in the $t^{\underline{th}}$ cycle and $CLB^{(s,t)} + CHW^{(s,t)}$ are the total allelic covariances between chromosomes generated by this $t^{\underline{th}}$ selection cycle.

An analysis of variance is performed by which the total variance among the offspring values is partitioned within and between full-sib progeny groups. The necessary parameters are shown in Table 7.2. As we did before, we assume additive gene action and we can then partition the total variance between gametes into the previous two independent components. The variance between gametic means can be here regarded as the variance between half-sibs or equivalently, as half the variance between full-sibs. After considerable algebra we arrive at the following expressions:

$$\text{Var(WFS)}^{(t+1)} = \tfrac{1}{2}(Vg^{(s,t)} - CHW^{(s,t)}) + \tfrac{1}{2}CLW^{(t)}(1-2c)\frac{W_{14}}{\bar{W}} + V_E$$

$$\text{Var(BFS)}^{(t+1)} = \tfrac{1}{2}(Vg^{(s,t)} + CHW^{(s,t)}) + \tfrac{1}{2}CLW^{(t)}\frac{W_{14}}{\bar{W}} + \tfrac{1}{2}CLW_f^{(t)} + \tfrac{1}{2}CLB^{(s,t)}$$

$$\text{Total} = Vg^{(s,t)} + CLW^{(t)}(1-c)\frac{W_{14}}{\bar{W}} + \tfrac{1}{2}CLW_f^{(t)} + \tfrac{1}{2}CLB^{(s,t)} + V_E \qquad (7.9)$$

A comparison of (7.9) and (7.8) is interesting. Notice that the parental chromosomes are uncorrelated before selection operated and therefore there is no effect of $CLB^{(s,t)}$ on Var(WFS). Selection of parents immediately leads to the generation of new sets of disequilibria, which adds on to the already existing disequilibrium within chromosomes. This is reflected in the variance component between full-sib families which includes all the disequilibria present

TABLE 7.2: Gametic output of the two locus additive model. The frequencies shown refer to the frequencies of the different genotypes after selection operates, in terms of parameters before selection.

Genotype	Frequency		Gametic Output				Mean	Variance between gametes within genotypes
			AB $\frac{1}{2}(a_1+a_2)$	Ab $\frac{1}{2}(a_1+a_2)$	aB $\frac{1}{2}(a_2-a_1)$	ab $-\frac{1}{2}(a_1+a_2)$		
AB/AB	f_1^2	$\frac{W_{11}}{\bar{W}}$	1				$\frac{1}{2}(a_1+a_2)$	0
AB/Ab	$2f_1f_2$	$\frac{W_{12}}{\bar{W}}$	$\frac{1}{2}$	$\frac{1}{2}$			$\frac{1}{2}a_1$	$\frac{1}{4}a_2^2$
AB/aB	$2f_1f_3$	$\frac{W_{13}}{\bar{W}}$	$\frac{1}{2}$		$\frac{1}{2}$		$\frac{1}{2}a_2$	$\frac{1}{4}a_1^2$
AB/ab	$2f_1f_4$	$\frac{W_{14}}{\bar{W}}$	$\frac{1}{2}(1-c)$	$\frac{1}{2}c$	$\frac{1}{2}c$	$\frac{1}{2}(1-c)$	0	$(\frac{1}{2}(a_1+a_2))^2 - ca_1a_2$
Ab/Ab	f_2^2	$\frac{W_{22}}{\bar{W}}$		1			$\frac{1}{2}(a_1-a_2)$	0
Ab/aB	$2f_2f_3$	$\frac{W_{23}}{\bar{W}}$	$\frac{1}{2}c$	$\frac{1}{2}(1-c)$	$\frac{1}{2}(1-c)$	$\frac{1}{2}c$	0	$(\frac{1}{2}(a_1-a_2))^2 + ca_1a_2$
Ab/ab	$2f_2f_4$	$\frac{W_{24}}{\bar{W}}$		$\frac{1}{2}$		$\frac{1}{2}$	$-\frac{1}{2}a_2$	$\frac{1}{4}a_1^2$
aB/aB	f_3^2	$\frac{W_{33}}{\bar{W}}$			1		$-\frac{1}{2}(a_1-a_2)$	0
aB/ab	$2f_3f_4$	$\frac{W_{34}}{\bar{W}}$			$\frac{1}{2}$	$\frac{1}{2}$	$-\frac{1}{2}a_1$	$\frac{1}{4}a_2^2$
ab/ab	f_4^2	$\frac{W_{44}}{\bar{W}}$				1	$-\frac{1}{2}(a_1+a_2)$	0

immediately following selection.

Notice that twice the variance between full-sibs minus the true genotypic variance in the previous unselected generation is equal to (assuming $\frac{W_{14}}{\bar{W}} = 1$),

$$(Vg^{(s,t)} - Vg^{(t)}) + CHW^{(s,t)} + CLW_f^{(t)} + CLB^{(s,t)}.$$

In other words, the estimate obtained from the variance between full-sibs is considerably biased downwards, particularly with large number of loci.

If the population is initially at equilibrium at t = 0, the variance component between full-sib groups at t = 1 is given (ignoring $CHW^{(s,0)}$:

$$Var(BFS)^{(1)} = \tfrac{1}{2}Vg^{(s,0)} + CLW_f^{(t)} \qquad (7.10)$$

(since in the first cycle of selection, CLW_f is equal to CLB). From (5.36) twice the covariance between full-sib groups becomes,

$$2Var(BFS)^{(1)} = Vg^{(s,0)} - \frac{i(i-x_T)}{\sigma^2(0)}Vg^{2(0)}, \text{ which reduces to (7.2)}$$

if gene frequency changes are ignored.

In general, the variance components within and between full-sib groups assuming a model of an infinite number of loci can be shown to be given by:

$$Var(BFS)^{(t+1)} = \tfrac{1}{2}VG^{(0)} + \tfrac{1}{2}i(i-x_T)\frac{VG^{2(t)}}{VP^{(t)}} + \tfrac{1}{2}CLW^{(t)}$$

$$Var(WFS)^{(t+1)} = \tfrac{1}{2}VG^{(0)} + \tfrac{1}{2}CLW^{(t)}(1-2\bar{c}) + V_E \qquad (7.11)$$

where, as before, $\bar{c}$ is the average recombination fraction and $CLW^{(t)}$ is given by,

$$CLW^{(t)} = -\tfrac{1}{2}\frac{i(i-x_T)}{VP^{(t)}} VG^{2(t)} + (1-\bar{c})CLW^{(t-1)}$$

We now produce Montecarlo simulation results in order to illustrate the concepts developed in this section.

Montecarlo Simulation Studies.

Table 7.3 shows the simulation results for the various models. The details of the parameters of these models are given in Chapter V. The observed results are obtained by an analysis of variance of the Montecarlo simulation results, averaged over replicates. Those headed P, are obtained by replacing the observed results, each generation, in the set of expressions (7.9), assuming $W_{14}/\bar{W} = 1$. Comparison between these two sets of results can be taken as a check on the algebra that lead to the derivation of (7.9). Results headed P_∞ are obtained from (7.11) and ignore gene frequency changes, the reduction in variance being solely due to the generation of linkage disequilibrium. The table also shows values of single generation observed and predicted realized heritabilities, the latter, using results of the infinitesimal model. Comparison between the estimate based on the intra-class correlation at generation (t+1) with the realized heritability at generation t gives an indication of the degree of bias introduced by selection of parents. In all the models studied, the bias is seen to be of considerable magnitude.

Table 7.3 shows that even though the predictions of the separate variance components based on the infinitesimal model are not strikingly accurate, the resultant intra-class correlation gives a rough indication of its decline as selection proceeds and the degree of bias to be expected. Unless models are rather extreme, such as (4, 0.5, 0.5)

TABLE 7.3: Observed (O) and Predicted (P, P_∞) genetic parameters for six additive genetic models. The P values are obtained by substituting the observed results in equations (7.9). The P_∞ values are obtained using the results of the infinitesimal model. t_c is the intra-class correlation. h^2_r is the single generation realised heritability. The variance components at generation 0 are the expected values obtained from the corresponding genetic model. The variance components at generation t refers to the analysis of variance of the offspring data of generation t generated by parents of generation s, t-1. The bias in the intra-class correlation is reflected by comparing 2 x t_c at generation t+1 with h^2_r at generation t. n = number of loci; q = initial frequency; c = recombination fraction between adjacent loci.

Model (n,q,c)		30, 0.5, 0.5 (20 reps)				4, 0.5, 0.5 (50 reps)			
t		Var(WFS)	Var(BFS)	2 x t_c	h^2_r	Var(WFS)	Var(BFS)	2 x t_c	h^2_r
0		22.50	7.50	0.50	0.50	22.50	7.50	0.50	0.50
	0	22.78±0.51	4.47±0.62	0.32	0.45	21.93±0.26	4.24±0.22	0.32	0.41
1	P_∞	22.50	4.57	0.34	0.45	22.50	4.57	0.34	0.45
	P	22.43	4.46	0.33		21.89	4.13	0.32	
	0	22.76±0.64	4.11±0.60	0.30	0.43	19.71±0.24	2.81±0.19	0.24	0.32
2	P_∞	22.50	3.93	0.30	0.43	22.50	3.93	0.30	0.43
	P	22.12	3.50	0.27		19.80	2.89	0.25	
	0	22.09±0.72	3.06±0.41	0.24	0.40	17.24±0.19	2.15±0.13	0.21	0.21
3	P_∞	22.50	3.78	0.29	0.43	22.50	3.78	0.29	0.43
	P	21.47	3.42	0.27		17.75	1.98	0.20	
	0	20.96±0.63	3.68±0.54	0.30		16.37±0.16	1.01±0.12	0.12	-
4	P_∞	22.50	3.75	0.29	-	22.50	3.75	0.29	
	P	20.91	3.11	0.26		16.22	1.03	0.12	

TABLE 7.3 (Continued): The standard error of 2 x t is approximately 0.04 and of h^2_r is about 0.01.

Model		30, 0.5, 0.0				30, 0.5, 0.1			
t		Var(WFS)	Var(BFS)	2 x t_c	h^2_r	Var(WFS)	Var(BFS)	2 x t_c	h^2_r
0		22.50	7.50	0.50	0.50	22.50	7.50	0.50	0.50
	0	22.28±0.49	4.33±0.36	0.32	0.44	23.36±0.43	3.92±0.30	0.28	0.44
1	P_∞	22.50	4.57	0.34	0.45	22.50	4.57	0.34	0.45
	P	22.39	4.27	0.32		22.44	4.35	0.32	
	0	19.91±0.48	3.66±0.35	0.30	0.37	21.38±0.44	3.69±0.28	0.29	0.41
2	P_∞	21.04	3.93	0.31	0.40	22.15	3.93	0.30	0.43
	P	20.55	3.69	0.30		21.87	3.68	0.29	
	0	19.75±0.42	2.53±0.25	0.22	0.33	21.50±0.34	3.34±0.31	0.26	0.38
3	P_∞	19.99	3.43	0.29	0.36	22.03	3.70	0.30	0.42
	P	19.31	2.65	0.24		21.23	3.36	0.27	
	0	17.95±0.33	1.97±0.26	0.20	-	20.23±0.38	3.58±0.29	0.30	-
4	P_∞	19.21	3.03	0.27		21.99	3.62	0.28	
	P	17.97	2.02	0.20		20.79	3.39	0.28	

TABLE 7.3 (Continued):

Model		(5/25, 0.1/0.5, 0.5)				(30, 0.2, 0.5)			
t		Var(WFS)	Var(BFS)	2 x t_c	h^2_r	Var(WFS)	Var(BFS)	2 x t_c	h^2_r
0		22.50	7.50	0.50	0.50	14.40	4.80	0.50	0.50
	0	26.02±0.39	5.76±0.44	0.36	0.56	15.54±0.22	3.26±0.22	0.34	0.50
1	P_∞	22.50	4.57	0.34	0.45	14.40	2.93	0.34	0.45
	P	26.18	5.81	0.36		15.27	3.26	0.35	
	0	29.04±0.52	5.84±0.56	0.33	0.57	15.70±0.22	2.71±0.16	0.30	0.50
2	P_∞	22.50	3.93	0.30	0.43	14.40	2.52	0.30	0.43
	P	29.57	5.62	0.32		16.00	2.83	0.30	
	0	30.92±0.66	5.92±0.43	0.32	0.60	16.18±0.23	2.70±0.16	0.28	0.50
3	P_∞	22.50	3.78	0.29	0.43	14.40	2.42	0.29	0.43
	P	31.10	5.86	0.32		16.44	2.79	0.29	
	0	30.82±0.52	6.23±0.53	0.33	-	16.73±0.22	2.74±0.18	0.28	-
4	P_∞	22.50	3.75	0.29		14.40	2.40	0.29	
	P	31.25	5.69	0.31		16.73	2.87	0.29	

which has relatively large gene effects, most of the decline in the intra-class correlation takes place during the first cycle of selection. This suggests that the predictors available in the literature (i.e. Reeve, 1953; Robertson, 1977a)which are a description of a unique cycle of selection are likely to provide a useful guide to the bias of estimates of heritabilities based on intra-class correlations in short term selection programmes.

Discussion and Conclusions.

Selection bias in the estimation of heritability by intra-class correlation between sibs has been well established by studies that assumed one cycle of selection on models of an infinite number of loci. In this work we have attempted an understanding of the problem with models of finite number of additive loci. As we pointed out in early chapters, prediction of the course of selection with such models involves expressions which assume knowledge of the numbers, frequency and effects of genes affecting the trait and are therefore of no direct practical application. These expressions however, provide us with a means of understanding the way the various genetic parameters interact in the selection process and the consequences of introducing a family structure into the model.

We have shown that the variance component within families is not affected by the presence of disequilibrium provided there is free recombination. If this is not the case, the disequilibrium already present before selection operates slightly reduces this component of variance. The fresh disequilibrium generated in the

new selection cycle only affects the variance component between sibs which is further reduced by half the disequilibrium present prior to selection. Both variance components are dependent on gene frequency changes and the sign of this effect will depend on the initial gene frequencies. The predictions made under the infinitesimal model are reasonably useful in providing a rough guide to the changes and degree of bias of the intra-class correlation during short-term selection studies.

Differences in heritability estimated by intra-class correlation and regression methods have been reported in various studies. Higher values for daughter-dam regression estimates for milk yield than those obtained from half-sibs were published by Van Vleck and Bradford (1965) and in one of two sets of data by Butcher and Freeman (1969). More recently, further evidence of higher daughter-dam regressions than intra-class correlations estimates was provided by Dymnicki *et al*. (1975). Van Vleck and Bradford (1965) suggested that the difference between both methods of estimation could be explained by a large genetic maternal effect though this seems to have been disproved by work of Lee and Henderson (1969) who showed that genetic maternal effects were of little importance in milk production. Syrstad(1966) suggested that environmental covariance between daughter and dam in the same herd could be a cause of discrepancy between both methods of estimation.

The problem is clearly not settled but it is interesting in that it could provide experimental evidence on the existence and magnitude of the negative joint disequilibrium, presumably generated by selection.

Purser (1980) has provided further evidence of lower estimates obtained from half-sib correlations than those obtained from realised response for various characters in sheep. The intra-class correlations were pooled over 7 or 8 generations of selection and were compared with estimates obtained from regressions of cumulative response on cumulative selection differential. Further estimates were obtained from a random bred control line.

For the characters studied (cannon bone length and medullation index) the largest estimate was the one obtained from the unselected control, followed by the realised heritability estimate. In agreement with theory, the smallest value was consistently obtained from the intra-class correlation estimate. The data published by Purser (1980) provide good experimental evidence of the existence of linkage disequilibrium generated by directional selection. The work with Drosophila reported in the next chapter was specifically set up in an attempt to provide further experimental evidence on this point.

CHAPTER 8

EFFECT OF DIRECTIONAL SELECTION ON GENETIC VARIABILITY

- EXPERIMENTS WITH *DROSOPHILA*

Introduction.

One of the most fundamental concepts in quantitative genetics is that of the additive genetic variance of a population and its related parameter, the heritability, knowledge of which allows us to predict the immediate response of the population to selection pressure. As a consequence of selection, the additive genetic variance and the heritability themselves change, and therefore strictly speaking, prediction of selection response based on present estimates of heritability are only valid for one cycle of selection.

These changes of genetic parameters due to selection are a consequence of changes of the frequencies of the genes affecting the trait and due to the generation of joint disequilibrium generated by selection. The magnitude of the changes of frequencies of genes depend on the number, effect, initial frequencies and linkage relationships between the loci involved, information which on the whole is not available to us. We are, therefore, unable to predict the changes in heritability brought about by selection coming from this source and all we can do is to say that if gene effects are very small relative to the phenotypic standard deviation of the trait, changes in their frequencies are not likely to be important during the early cycles of selection.

The other source of change of genetic parameters comes about through the generation of joint disequilibrium. Bulmer (1971, 1974) has developed a theory which allows us to predict the magnitude of its effect in terms of parameters of the base population, before the operation of selection. This theory is based on a model which assumes an infinite number of loci and therefore ignores the

changes of genetic parameters due to changes of gene frequencies. In Chapters 5 and 6, we have studied the consequences of introducing a finite number of loci under a range of genetic models, on the predictions of changes of genotypic variance based on Bulmer's theory. We have shown that, in the absence of dominance, the predictions of the generation of joint disequilibrium are reasonably accurate, provided gene frequencies are not extreme. However, the prediction of changes in the genotypic variance can depend critically on the gene frequency distribution in the base population. In fact, we have shown that with low initial frequencies, the effect of gene frequency changes and the effect of joint disequilibrium tend to cancel each other out and consequently, the genotypic variance remains virtually stable during the first four generations of selection.

How important then are the changes of genetic parameters likely to be in selection programmes of short duration? From an operational point of view, an answer to this question can be obtained by comparing predicted responses to selection, based on estimates of parameters of the base population, and observed responses (see Wright, 1977 for a review), though clearly this type of information does not tell us anything about the causes determining the agreement or otherwise between these results. Lack of agreement does not imply that substantial changes are taking place. Estimates of base population parameters from small samples may be highly variable due to sampling; the regression of offspring on parent may be non linear and therefore the expected responses to selection in the up and down direction will be asymmetrical (Robertson, 1977c); there may be substantial

maternal effects affecting the trait (Falconer, 1963); there may be natural selection opposing artificial selection (Clayton & Robertson, 1957a)or there may be problems of scale (Robertson, 1970c). On the other hand, agreement between observed and predicted responses should not necessarily lead us to conclude that no genetic changes are taking place, since we have shown that the effect of gene frequency changes can be opposed by the effect of joint disequilibrium and consequently, genetic parameters remain fairly stable, at least during the early cycles of selection.

The experiments reported in this chapter were designed to study the effect of short term directional selection on changes of the heritability, in particular, those changes associated with the generation of joint disequilibrium. Two experiments of different designs were performed, each one run with two replicates. Essentially both experiments involved a few cycles of selection followed by a period of relaxation. Heritability estimates were obtained during the period of selection and at the end of the period of relaxation. On the basis of the theory developed by Bulmer (1971) we anticipate an increase of heritability at the end of the period of random mating due to the breakdown of negative joint disequilibrium generated during the early cycles of selection.

In the first experiment the direction of selection was reversed each generation in an attempt to minimise changes of genetic parameters due to gene frequency changes and avoid complications introduced by scale effects. The amount of joint disequilibrium should accumulate however, regardless of the direction of selection. With

a finite number of loci, the reversion of the direction of selection may have a small effect on the amount of disequilibrium produced each generation, but this effect is not likely to be important unless gene effects are very large.

In the second experiment selection was practised for high value of the trait and changes of parameters are then due to both the permanent effects of gene frequency changes and due to the temporary effects of the generation of disequilibrium. The permanent effects on the heritability should be reflected in the estimate obtained at the end of the period of relaxation, during which a large proportion of the temporary effects should disappear.

Material and Methods.

Lines were derived from the Dahomey population which has been kept in cages in this laboratory since 1969.

The character measured was the sum of the abdominal bristles on the fourth and fifth segments in males and fifth and sixth segments in females.

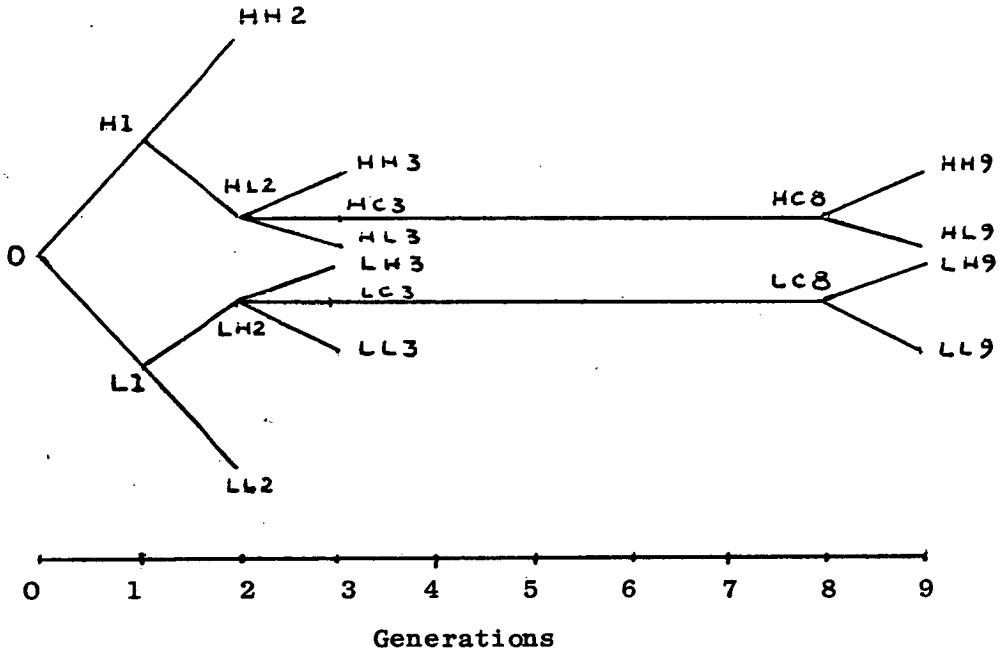
Flies were reared in standard Edinburgh agar-molasses killed yeast medium in which drops of live yeast had been added. All cultures were kept at 25^{o}C.

Two experiments were carried out which we shall designate experiment 1 and experiment 2, each one being run with two replicates, a and b.

Experiment 1(a) was carried out in half-pint milk bottles. Eggs were sampled from the cage population with several bottles. When the adults emerged, 150 males and 150 females were scored and

they constituted generation zero. Two way selection was practised by selecting the highest (H) 30 males and 30 females and the lowest (L) 30 males and 30 females. These two groups were introduced, each one in a separate bottle to mate for 48 hours. On the third day, flies were transferred into a fresh bottle and allowed to lay eggs for 10 hours. The adults emerging from these bottles constituted generation 1. Those derived from individuals selected for the high value of the character were designated H1; those derived from individuals selected for low value of the character were designated L1. From H1, two way selection was practised once again by selecting the 30 highest and 30 lowest of each sex out of a total of 150 scored from each sex. A similar procedure was followed in L1, and therefore, at generation 2, four sets of 300 flies in each set were scored and designated as follows. The two way selection originated from H1, yielded HH2 and HL2. The two way selection originated from L1, yielded LH2 and LL2. From HL2, three lines were derived. Selection of extremes and random mating within extremes lead to HH3 and HL3. At the same time, a random sample of 30 males and 30 females were chosen from HL2 and this procedure constituted the first cycle of random mating. The offspring of the first cycle of random mating was designated HC3. Similarly, from LH2 we generated LH3, LL3 and LC3. Random mating was continued for 6 cycles. At generation 8, two way selection was practised from HC8 and LC8, and the four lines were designated HH9, HL9, LH9, LL9. The design of the experiment and designation of the lines are shown in Figure 8.1.

FIGURE 8.1: Design of Experiment 1

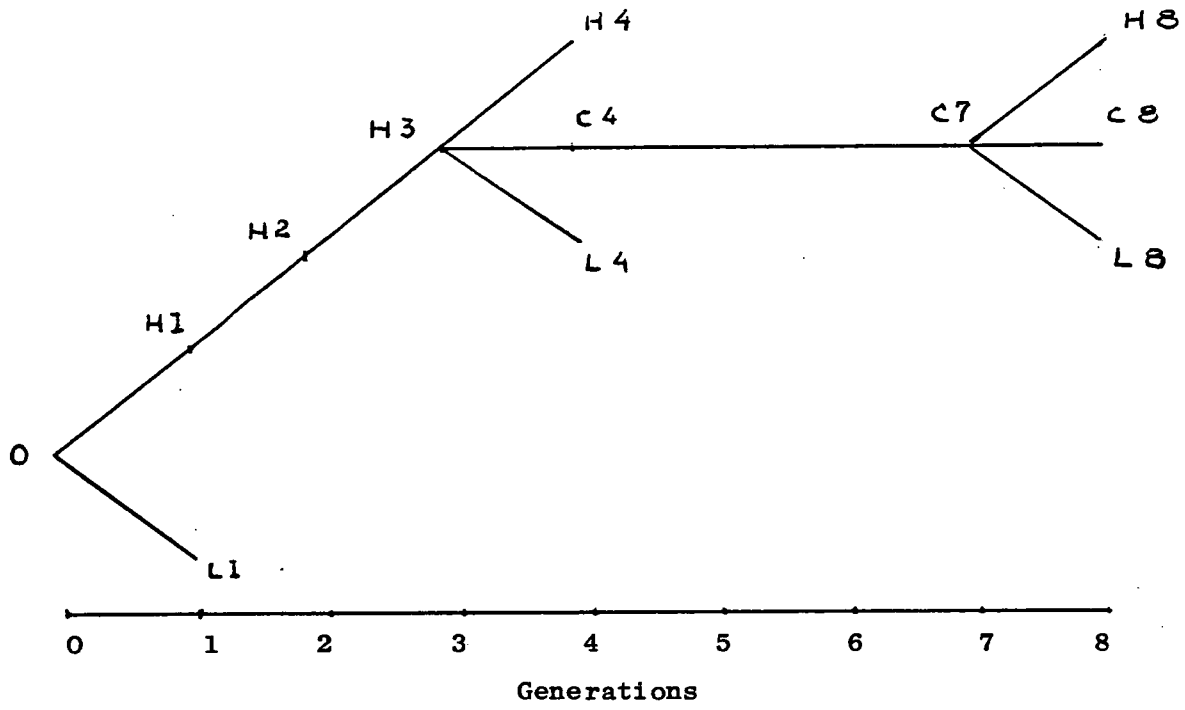


Experiment 1(b) was essentially similar to experiment 1(a), the difference being that flies were kept in vials with the exception of the period comprising the 6 cycles of random mating, when flies were kept in bottles. While flies were kept in vials, each of the 30 full-sib families contributed 5 males and 5 females to those scored the next generation. Selection was always carried out on the basis of the scores of individual flies.

In experiment 2, eggs were sampled from the cage using several bottles. 30 emerging flies of each sex were sampled from the bottles and mated in individual vials, one male and one female per

vial. Ten days later, when the offspring of these flies emerged, 5 males and 5 females were randomly chosen from each of the 30 families. These 300 flies were scored and constituted generation zero. Two way selection was practised by selecting the highest (H) and lowest (L) 10% at each extreme. The selected flies, chosen on the basis of their individual scores were mated at random and their offspring constituted H1 and L1. Line L1 was discarded. From H1, 300 flies were scored, 10 males and 10 females from each of the 15 full-sib families and the top 10% selected and mated at random. This procedure was repeated and lead to H2 and H3. At generation 3, from H3, three lines were started. Two way selection lead to H4 and L4, whereas random sampling of one male and one female from each full-sib family from H3 and subsequent random mating produced C4, which constituted the offspring of the first cycle of random mating in vials. In addition to these three lines, flies were also sampled from H3 and relaxed in bottles under crowded conditions to investigate the possible regression of the mean during the period of random mating. At generation 7, two way selection was practised, together with random sampling of 7C and this constituted H8, L8 and C8. These three lines were derived from the line that was relaxed in vials. Both experiment 2a and experiment 2b were run in the same way. The design and designation of the lines are summarised in Figure 8.2.

FIGURE 8.2: Design of Experiment 2.



Results.

Estimates of Parameters of the Base Population.

Base population parameter estimates are shown below. The means and phenotypic variances are obtained from over 1,200 observations in each sex. The heritability estimate based on offspring mid-parent regression (with parents selected at both extremes of the distribution) is obtained by pooling 7 independent estimates. The full-sib estimate is obtained by pooling two independent estimates.

Males		Females	
Mean	Phenotypic variance	Mean	Phenotypic variance
33.61	7.84	38.95	9.11

	Heritability
Offspring mid-parent regression	0.35 ± 0.02
Twice Intraclass correlation between full-sibs	0.38 ± 0.07

	Total Variance
Vg	3.13 ± 0.21
Within fly variance	3.98 ± 0.16
Not accounted for	1.37

The 'developmental error' variance was estimated from the mean squared difference between scores in both segments. The estimate was of about 4 squared units leaving only about 1.4 units to account for true environmental and other non-additive genetic components. Notice however, the close agreement between the heritability estimate based on offspring mid-parent regressions and the one obtained from intra-class correlations between full sibs, suggesting that common environmental variance and non-additive genetic variance are not important sources of variation in this experiment.

Results of Experiment 1.

Means and variances for each sex at generation zero, based on 150 observations, are shown below for each replicate.

Replicate		Males	Females
1(a)	Mean	33.81	38.42
	Phenotypic variance	8.73	12.81
1(b)	Mean	32.91	38.53
	Phenotypic variance	7.53	8.83

Table 8.1 shows the difference between the high and low lines and the sum of the standardised selection differentials of the high and low lines during the course of the selection programme for both replicates of experiment 1.

The standardised selection differentials are smaller than the expected value of 2.8 and they are consistently smaller in experiment 1(b) than in the other replicate. The smaller selection differentials in 1(b) reflect the degree to which spare flies had to be used due to the occurrence of unsuccessful matings. In 1(a), all selected flies were introduced in a bottle and therefore no record of individual flies were kept.

Table 8.2 shows the single generation realized heritabilities, the genotypic variance and twice the intra-class correlation between full-sibs (in the case of experiment 1(b)) during the selection programme. The single generation realized heritabilties were obtained from the ratio of the divergence over the sum of the selection differentials. The genotypic variance was estimated by multiplying

TABLE 8.1:

Sum of intensities of selection for high and low selected lines (i) and the divergence, for both replicates ($\bar{X}_H - \bar{X}_L$). S.E. of divergence obtained from Hill (1972a).

Generation	Experiment 1(a)		Experiment 1(b)	
	i	$(\bar{X}_H - \bar{X}_L)$	i	$(\bar{X}_H - \bar{X}_L)$
H1/L1	2.71	2.69±0.38	2.38	2.05±0.34
HH2/HL2	2.68	3.84±0.38	2.64	2.46±0.34
LH2/LL2	2.71	1.83±0.30	2.54	1.94±0.31
HH3/HL3	2.75	4.47±0.36	2.59	2.61±0.33
LH3/LL3	2.75	3.25±0.35	2.61	2.52±0.32
HH9/HL9	2.67	3.52±0.37	2.51	2.03±0.31
LH9/LL9	2.70	2.72±0.34	2.54	2.71±0.34

TABLE 8.2:

Single generation realized heritabilities (h^2), genotypic variance (VG) and twice the intraclass correlation between full-sib families ($2xt_c$) for both replicates, obtained from analysis of variance corresponding to the generation shown in brackets next to the estimate. VG estimated from the product of the realized heritability times the phenotypic variance, the latter estimated from the analysis of variance in the case of experiment 1(b). Standard errors of h^2 obtained from Hill (1972a). St errors of VG obtained from the square root of σ^4 Var($\hat{h}^2$) + h^4 Var($\hat{\sigma}^2$), assuming $\hat{h}^2$ and $\hat{\sigma}^2$ are uncorrelated. Standard errors of 2tc obtained from Fisher (1941).

Generation	Experiment 1(a) VG	Experiment 1(a) h^2	Experiment 1(b) VG	Experiment 1(b) h^2	Experiment 1(b) $2xt_c$
0	3.23±0.51	0.30±0.04	2.54±0.39	0.31±0.04	0.24±0.13(H1) 0.25±0.13(L1)
H1	4.47±0.53	0.47±0.04	2.83±0.46	0.36±0.05	0.35±0.15(HH2) 0.48±0.15(HL2)
L1	1.83±0.32	0.26±0.04	1.93±0.35	0.25±0.04	0.06±0.08(LH2) 0.30±0.14(LL2)
HL2	4.78±0.52	0.56±0.04	2.66±0.71	0.35±0.05	0.29±0.14(HH3) 0.37±0.15(HL3)
LH2	3.38±0.43	0.41±0.04	2.49±0.42	0.34±0.05	0.49±0.15(LH3) 0.59±0.14(LL3)
HC8	3.97±0.49	0.44±0.04	2.22±0.40	0.31±0.05	0.54±0.15(HH9) 0.31±0.14(HL9)
LC8	2.92±0.41	0.35±0.04	3.09±0.41	0.38±0.04	0.27±0.14(LH9) 0.54±0.15(LL9)

the estimate of the realized heritability by the contemporaneous estimate of the phenotypic variance.

The critical comparisons that should provide evidence for the build up of disequilibrium during selection are between the estimate at generation 0 and both estimates at HL2 and LH2. The standard errors are not small but neither replicate show any decrease in the genotypic variance or the heritability. Furthermore, there are no signs of an increase in genetic parameters during the period of random mating.

Another source of evidence for the build up of disequilibrium should come from a comparison between the realized heritabilities and twice the intra-class correlations. We expect the latter to be smaller due to the negative bias introduced by selection in the parental generation. With the possible exception of the estimates of generation H1 and L1, a general glance at the table shows no indication of any detectable effect of disequilibrium. The equivocal nature of these results, together with information coming from Montecarlo simulation studies stimulated the development of experiment 2.

Results of Experiment 2.

Estimates of means and variances from each replicate at generation zero, based on 150 observations for each sex are shown below.

Table 8.3 shows the standardised selection differentials and the response to selection for both replicates.

TABLE 8.3:

Standardised selection differentials (i) and selection response (R) for experiment 2. The response is obtained by the difference in mean between one generation and the mean of the preceding generation. Standard errors of response obtained from Hill (1972b) assuming parameters do not change.

Generation	Experiment 2(a)		Experiment 2(b)	
	i	R	i	R
LL	-1.60	-2.32±0.40	-1.83	-2.28±0.40
H1	1.74	1.20±0.40	1.52	1.90±0.40
H2	1.70	2.54±0.40	1.87	2.12±0.40
H3	1.67	0.76±0.40	1.70	2.45±0.40
H4	1.75	1.78±0.40	1.78	2.93±0.40
L4	-1.68	-0.82±0.40	-1.62	-2.54±0.40
H8	1.58	0.71±0.40	1.62	6.72±0.40
L8	-1.72	-1.70±0.40	-1.50	-5.68±0.40

Replicate		Males	Females
2(a)	Mean	33.16	39.28
	Phenotypic variance	8.24	7.29
2(b)	Mean	33.19	38.78
	Phenotypic variance	6.92	10.37

The observed selection intensities are, in general, somewhat smaller than the expected value of 1.75 obtained from normal tables. The response to selection is smaller and more erratic in experiment 2(a) than 2(b).

Table 8.4 shows estimates of single generation realized heritabilities and genotypic variances obtained from the product of the realized heritabilities and their contemporaneous phenotypic variances estimated from the analysis of variance, and Table 8.5 shows estimates of heritabilities obtained from intra-class correlations between full sibs and their corresponding variance component between families for both replicates of experiment 2.

The general picture is one of remarkable disagreement between both replicates. In the case of replicate (a), the single generation realized heritabilities show an irregular pattern during the first four cycles of selection. This replicate does not show an increase in genetic parameters during relaxation, suggesting that any effect of the breakdown of disequilibrium, if any, was too small to be detected. The results of the variance components between families are consistent with those obtained for the realized heritabilities and the genotypic variance in that the pattern of change

TABLE 8.4:

Experiment 2: Estimates of single generation realized heritabilities (h^2) and genotypic variances (VG), obtained as in Table 8.2. Standard error of h^2 from Hill (1972b).

Generation	Experiment 2(a)		Experiment 2(b)	
	VG	h^2	VG	h^2
O	2.96±0.46	0.38±0.05	3.63±0.52	0.42±0.04
H1	4.34±0.75	0.52±0.08	3.44±0.78	0.38±0.06
H2	1.24±0.55	0.16±0.07	4.48±1.04	0.46±0.10
H3	2.28±0.40	0.26±0.04	5.83±0.61	0.45±0.03
C7	2.08±0.36	0.26±0.04	7.66±0.74	0.59±0.03

TABLE 8.5:

Experiment 2: Estimates of heritability based on intra-class correlation between full sib families ($2t_c$) and corresponding variance components between families (σ_b^2). Standard errors of t_c from Fisher (1941); standard error of $\hat{\sigma}_b^2$ obtained from the square root of :

$$\frac{2}{n^2}\left\{\frac{(MSB)^2}{n_f+1} + \frac{(MSW)^2}{n_f(n-1)+2}\right\}, \text{ where}$$

n: number of observations per family

n_f: number of families

MSB: mean square between families

MSW: mean square within families

Generation	Experiment 2(a)		Experiment 2(b)	
	$2t_c$	σ_b^2	$2t_c$	σ_b^2
H1	0.11±0.11	0.51±0.08	0.25±0.17	1.15±0.10
L1	0.27±0.17	0.93±0.09	0.23±0.17	0.91±0.09
H2	0.40±0.20	1.52±0.11	0.38±0.20	1.85±0.12
H3	0.20±0.15	0.88±0.09	0.33±0.19	2.14±0.13
H4	0.21±0.15	0.96±0.09	0.34±0.19	2.38±0.14
L4	0.10±0.10	0.40±0.07	0.07±0.08	0.27±0.06
C7	0.11±0.10	0.45±0.07	0.56±0.20	3.61±0.16
H8	0.11±0.10	0.49±0.08	0.29±0.18	2.19±0.13
L8	0.21±0.15	0.88±0.09	0.23±0.17	1.17±0.10
Cumulative Response on cumulative selection differential (gens 0-H4)	0.36±0.03		0.41±0.03	

is irregular and the value at generation C7 has not increased relative to its value at generation H3.

The single generation realized heritabilities in replicate (b) remain fairly stable during the first four cycles of selection. During the period of random mating, there is a marked increase in the genotypic variance and in the heritability and this result is substantiated by a considerable increase in the variance component between full-sib families at the end of the period of relaxation relative to its value at generation H3. In marked contrast with replicate (a), this replicate showed a considerable increase in the variance components between and within families during the cycles of selection.

During the period of relaxation, flies were sampled from the vials in each replicate, from which the mean was estimated. The data (see Table 8.6) clearly show that there is no regression of the mean during the relaxation of selection.

At generation 7 flies were sampled from the line relaxed in bottles and reared in vials for one generation. At generation 8, C8, the mean was estimated for both replicates. These means do not differ from those obtained from the lines relaxed in vials, suggesting that there must be rather weak evolutionary forces holding the mean in its original position and further, that natural selection does not seem to oppose artificial selection in this short term selection experiment. The results are summarised in Table 8.6.

TABLE 8.6:

Means for both replicates during the period of relaxation. The last value corresponds to the estimate of the mean from the lines relaxed in bottles.

Generation	Replicate (a)	Replicate (b)
H3	40.94 ± 0.17	42.40 ± 0.20
C4	41.32 ± 0.41	42.35 ± 0.51
C5	40.95 ± 0.38	41.72 ± 0.56
C6	40.32 ± 0.38	41.52 ± 0.52
C7	40.94 ± 0.40	42.18 ± 0.51
C8	41.55 ± 0.38	42.32 ± 0.54
C8 (relaxed in bottles)	41.02 ± 0.42	42.83 ± 0.42

Discussion.

Before discussing the results we shall briefly justify the two different designs of experiments 1 and 2.

Experiment 1 was designed in an attempt to keep gene frequency changes to a minimum. It will be noticed that the first cycle of divergent selection followed by a second cycle of divergent selection starting from the H1 and L1 lines is equivalent to a process in which selection is practised in one direction and in the following generation, the direction of selection is reversed (see Figure 8.1). The heritability estimate at generation 0, obtained from the divergence of H1 and L1 should in principle, be compared with the estimates at

HL2 and LH2, these estimates having been obtained from the divergence of HH3/HL3 and LH3/LL3 respectively. Gene frequencies at generation zero and at both HL2 and LH2 should be more or less similar but the two cycles of selection should have caused disequilibrium and this should lead to a reduction of the heritability at HL2 and LH2. We also expect an increase in the heritability estimate at generations HC8 and LC8 after the various cycles of random mating, on the assumption that the reduction at HL2 and LH2 was due to the generation of linkage disequilibrium.

In marked contrast with experiment 1, experiment 2 was designed in a straightforward fashion and gene frequency changes were not controlled. In fact, a higher selection intensity was applied (10% rather than 20%) and one way selection for high abdominal scores was practised for three generations. Heritability was accurately estimated at generation zero, at generation 3 and finally at generation 7, after the four cycles of random mating. A comparison of the estimate at generation 3 with the estimate at generation 7, should provide evidence for the effect of the breakdown of disequilibrium which was generated during the three cycles of selection. Furthermore, the difference between the estimate at generation 7 and at generation zero should provide some idea of the effect of gene frequency changes during this short term study.

In experiment 1, the results do not suggest that selection has generated disequilibrium of any considerable magnitude. Before embarking on a description of a genetic model that could account for these results, it is important to notice that the heritability estimates of the base population obtained from both replicates were

indeed very low. In fact, the experiment was designed on the assumption that the heritability for the sum of the abdominal scores in this population was 50%. Our estimates turned out to be considerably smaller in both replicates. In a large population where the heritability is 30%, one cycle of selection of the intensity we used in experiment 1 (20%) is expected to reduce the heritability to about 27%. The effect is very small and not likely to be detected. This was realized at the time but notwithstanding we decided to continue with the experiment to see "what happened". It turned out that basically, "nothing happened".

The results of experiment 2, however, merit some speculation in terms of a model that may lead to rather substantial changes in genetic parameters, particularly after relaxation of selection, in one replicate and none in the other. First of all we want to point out that it is unlikely that the increase in heritability in replicate (b) is due to the elimination of lethal factors. The design we used minimised any effect of natural selection operating between families and furthermore, the lack of change in the mean in flies reared in crowded bottles during the period of random mating suggest that the effect of natural selection opposing artificial selection must have been very weak in both replicates.

The model that we suggest could account reasonably well for this set of results is one in which the character in the base population is affected by a few loci of large effect at extreme frequencies and several minor loci at intermediate frequencies. A model of natural selection for an intermediate optimum value of the quantitative trait, together with uniform mutation rates involving

two alleles per locus leads to an equilibrium configuration of the kind represented by this model (Latter, 1960). How is such a model likely to respond when submitted to selection for high value of the character followed by a period of relaxation, as in the case of experiment 2?

First of all there is the effect of gene frequency change on the genotypic variance. The plus alleles at very high frequencies do not contribute much initially to the equilibrium additive variance and their contribution becomes even smaller as upward selection proceeds. However, those plus alleles initially at very low frequency will make substantial contributions to the equilibrium additive variance as they quickly move towards intermediate values. The minor genes at intermediate frequencies are not likely to have an impact on the equilibrium additive variance of any real importance, particularly during the early generations of selection. Therefore, the changes in the genotypic variance arising from overall gene frequency changes are likely to be positive due to the overall increase in the equilibrium additive variance.

Secondly, there is the effect of disequilibrium. We have shown in earlier chapters that extreme low initial frequencies, particularly when proportionate effects of the genes are large, leads to larger reductions in the genotypic variance than predicted on the basis of the infinitesimal model. If the number and effects of the loci at low initial frequencies are the same as those at high frequency no asymmetry should develop in a first cycle of selection in either the amount of disequilibrium generated or the selection response. In a second cycle of selection, the immediate differential change in gene frequency of loci at both extremes of the gene frequency

range, will lead to substantial reductions in the genotypic variance due to relatively large generation of disequilibrium contributed by those loci at low initial frequencies. On top of this, we have the effect of gene frequencies moving towards intermediate values of the major loci, with associated increase in the amount of disequilibrium generated. We then expect considerable reductions in the genotypic variance coming from this source. Overall, the genotypic variance is not likely to change very substantially during the early generations of selection and the direction of the change will largely depend on the genetic parameters of the model.

This type of model is also likely to lead to considerable degree of variation between replicates. This will depend on the size of the initial sample and on how extreme the frequency of the major loci are likely to be in the base population. As James (1971) pointed out, if a trait is affected by loci of large effect where the favourable allele is rare, initial samples of moderate size are likely to generate more variation between replicates than samples of very small or very large size, because in the latter case, a large proportion of the samples will either have, or have not,included the favourable alleles, whereas in the former, appreciable proportions would include and fail to include them.

Our conjecture will be that in replicate (b), more loci of large effect were picked up than in replicate (a). An important question in this respect is, given that in the base population there are rare loci of large effect, how many of those are likely to have been missed in replicate (a) in order to explain the considerable difference of behaviour compared with replicate (b). In order to

get some idea of this and to further check the behaviour of this model, we resorted to Montecarlo simulation. The model we used was one in which 30 loci were distributed in equal numbers in three chromosomes, the recombination fraction between loci on the same chromosome being 0.1. We sampled from a population which assumed 24 loci at initial frequency of 0.4; 4 adjacent loci in chromosome 1 at frequency 0.05 and 2 adjacent loci in chromosome 3 at frequency 0.95. Each of the 24 minor loci had an effect of 0.13 standard deviations whereas the 6 loci at extreme frequencies had an effect of about 0.60 standard deviations. The genotypic variance in the equilibrium base population was 3.8 square units and the phenotypic variance was 10 square units. These parameters are similar to the estimates we obtained from our Drosophila experiment 2. Together with this model we ran others which assumed that 2 and 3 of the 4 favourable major alleles were completely absent, and finally we ran a model in which the genetic variation was due to 30 loci of equal effects and initial frequency of 0.4. The parameters of the various models, together with the model designation are summarised in Table 8.7. All models have about the same initial equilibrium additive variance and heritability and the highest 15 out of 150 scored in each sex were selected and mated at random each generation.

The results are shown in Table 8.8 and 8.9. Table 8.8 shows the equilibrium additive variance, the amount of joint disequilibrium and the realized heritabilities. Table 8.9 shows the intraclass correlations and the variance components between and within full-sib families.

TABLE 8.7:

Model Designation	Number of +loci at frequency 0.05	Number of +loci at frequency 0.95	Number of minor loci at frequency 0.4
4/24	4	2	24
2/24	2	2	24
1/24	1	2	24
0/30	0	0	30

The results of model 4/24, including the increase in heritability predicted during random mating agree closely with those of replicate (b). The variance components between and within families however, do not increase in the simulation as they do in the Drosophila experiments. This, however, is probably due to a scale problem in the sense that, as discussed by Robertson (1970c) for the case of his sternopleural lines, in our case, the scale we are using to measure abdominal bristle scores may not be the one in which the effect of a gene substitution is constant as selection proceeds.

The results of replicate 2(a) are reasonably compatible with those of model 1/24 or 0/30, that is, it is not likely that the number of major loci sampled initially is larger than 1. These two models lead to rather small reductions in variance due to disequilibrium and after 4 cycles of selection, the equilibrium additive variance is slightly smaller than it was originally. Consequently, heritability estimates at generation 4 are smaller than the estimate obtained from the base population and after relaxation, the break-

TABLE 8.8: Equilibrium additive variance (Vg), joint disequilibrium (CLW) and single generation realized heritabilities (h^2) for the models of Table 8.4 in 4 cycles of directional selection. The results are average of 25 replicates. The S.E. of h^2 is about 0.01.

Model Generation	4/24			2/24			1/24			0/30		
	Vg	CLW	h^2	Vg	CLW	h^2	Vg	CLW	h^2	Vg	CLW	h^2
0	3.76±0.04	0.08±0.07	0.39	3.65±0.02	0.02±0.05	0.38	3.66±0.03	-0.03±0.05	0.37	3.74±0.00	0.05±0.05	0.39
1	4.75±0.09	-0.74±0.15	0.42	4.16±0.06	-0.63±0.12	0.38	3.61±0.06	-0.50±0.05	0.33	3.82±0.00	-0.37±0.11	0.35
2	6.13±0.14	-1.63±0.18	0.42	5.03±0.10	-1.08±0.15	0.41	4.07±0.10	-0.93±0.10	0.33	3.74±0.01	-0.60±0.12	0.32
3	7.08±0.14	-2.62±0.24	0.40	5.35±0.10	-1.18±0.15	0.41	4.13±0.07	-0.82±0.12	0.31	3.53±0.03	-0.71±0.10	0.28
4	7.20±0.17	-2.65±0.23	-	4.94±0.11	-1.35±0.18	-	3.73±0.12	-0.62±0.12	-	3.31±0.03	-0.67±0.07	-
4 cycles of random mating			0.56			0.43			0.37			0.34

TABLE 8.9: Variance components within (σ^2_w) and between full sib families (σ^2_b) together with heritability estimates based on intra-class correlations ($2xt_c$) for the various models. The S.E. of σ^2_w and σ^2_b are about 0.2. The S.E. of t_c is about 0.01.

Model Generation	4/24			2/24			1/24			0/30		
	σ^2_w	σ^2_b	$2xt_c$	σ^2_w	σ^2_b	$2xt_c$	σ^2_w	σ^2_b	$2xt_c$	σ^2_w	σ^2_b	$2xt_c$
1	8.8	1.5	0.30	8.4	1.4	0.29	8.1	1.3	0.26	8.3	1.4	0.28
2	9.2	1.6	0.30	9.1	1.2	0.23	8.3	1.2	0.25	8.2	1.2	0.25
3	9.2	1.5	0.28	9.2	1.5	0.28	8.3	1.1	0.23	7.9	1.1	0.24
4	9.2	1.5	0.27	8.8	1.2	0.23	8.3	1.2	0.24	7.9	1.0	0.22

down of the relatively small amount of joint disequilibrium generated during the selection process is not likely to be detected. The results of experiment 2(a) conform reasonably well with these expectations.

How are the models in Table 8.7 likely to react in the case of the experimental design of experiment 1? Montecarlo simulation results are shown in Table 8.10. The design of experiment 1 was simulated by selecting up in one generation followed by a second cycle of reverse selection, this pattern of selection being continued for 4 generations. The initial heritability was assumed to be 30% and 30 out of 150 scored in each sex were selected as parents each generation.

In all models, the equilibrium additive variance changes very little during the 4 cycles of intermittent selection. Model 0/30 produces, as expected, relatively larger amount of disequilibrium but the effect on heritability estimates is small. In all the other models, the genes of large effect are kept at low frequencies and therefore their contribution to the reduction in the genotypic variance due to disequilibrium is only trivial. The results show clearly that the increase in heritability during relaxation, even if all the disequilibrium broke down would be very small and it would require a very large experiment to detect such a change.

What is the evidence for this type of model available in the literature? All the evidence we have is rather indirect and circumstantial. A model of genes of large effect at extreme frequencies has been postulated by Clayton et al. (1957b) and Sen and Robertson

TABLE 8.10:

Models of table 8.4 submitted to alternate cycles of high and low selection for 4 generations. The parameters shown are the amount of disequilibrium generated (CLW) and the single generation realized heritabilities (h^2). S.E. of h^2 about 0.01.

Model	4/24		2/24		1/24		0/30	
Generation	CLW	h^2	CLW	h^2	CLW	h^2	CLW	h^2
0	0.06±0.04	0.31	-0.09±0.03	0.29	0.04±0.07	0.29	-0.01±0.04	0.31
1	-0.49±0.11	0.28	-0.49±0.11	0.27	-0.31±0.13	0.26	-0.57±0.10	0.30
2	-0.50±0.07	0.26	-0.45±0.11	0.26	-0.51±0.11	0.26	-0.74±0.10	0.28
3	-0.44±0.10	0.29	-0.62±0.11	0.26	-0.65±0.11	0.27	-0.82±0.11	0.27
4	-0.44±0.10	-	-0.50±0.10	-	-0.54±0.12	-	-0.81±0.14	-

(1964) to account for the observations of correlated response in sternopleural bristle number on selection for abdominal bristle number. Frankham et al. (1968) showed that several of their lines showed periods of rapid response associated with increases in variance and further, the crosses involving one of their lines with others gave rapid response to selection in contrast with the response of the crosses not involving this particular line. They pointed out that these observations can be reconciled by the presence of major loci at low frequency in the base population. Evidence of this model was also provided by Robertson (quoted by James, 1971), and more recently by Yoo (1980), who suggested it as a model that could account for the large variation between replicates that he observed amongst his lines.

What conclusions can be drawn from this work? In agreement with the results arrived at in Chapters 5 and 6, we believe that, even in the case of short term selection studies, changes of genetic parameters are rather dependent on the underlying genetic model, that is, the distribution of gene frequencies and effects in the base population. Predictions of the generation of joint disequilibrium may be in some cases reasonably accurate, but we are not in a position to predict changes in the genotypic variance, unless we have some idea of the likely rate of gene frequency change during selection, as would be the case of a population resulting from a cross between highly inbred lines, or as shown in Chapter 5, when population size is small enough that most of the changes in the equilibrium additive variance are likely to be due to genetic drift.

The response to 4 cycles of selection in experiment 2 was 6.88 and 9.29 units for replicates (a) and (b) respectively, with an average of 8.08. The usual prediction, $tih^{2(0)}\sigma^{(0)}$, gives an expected total response of 8.12 units. Predicted and observed results are in good agreement but if one tentatively accepts the model we have proposed to explain the results, it is clear that in this case at least, an explanation based on the general idea that parameters have not changed, although operationally correct, may be misleading.

CHAPTER 9

EFFECT OF DISRUPTIVE SELECTION ON GENETIC VARIABILITY - MONTECARLO SIMULATION STUDIES

Introduction

Under natural conditions, variation in the environment over the range occupied by an interbreeding population may lead to differences in the value of the optimum phenotype favoured by selection. This type of selection was termed centrifugal selection by Simpson (1944), while Mather (1955) proposed the term disruptive selection. Mather's term is usually associated with the situation where those individuals at both extremes of the distribution survive and the intermediates do not.

A considerable body of literature on experimental results of selection for such phenotypic deviants has grown over the years, much of which has been reviewed by Thoday (1972). Several researchers have reported considerable increases in the genetic components of variance of quantitative traits as an outcome of disruptive selection (Thoday, 1959; Millicent & Thoday, 1961; Gibson & Thoday, 1963; Scharloo, 1964; Scharloo *et al.*, 1967; Barker & Cummins, 1969) as well as genetic diversity at enzyme loci (Powell, 1971; McDonald & Ayala, 1974). This type of selection may also lead to an increase of non-genetic components of variance since, assuming that there is genetic control of sensitivity to environmental factors, selection of extremes should result in selection of the most sensitive individuals.

The changes in the genetic components of variance are due to changes in gene frequency and due to the generation of positive linkage disequilibrium amongst the loci affecting the trait. Robertson (1956) using a single locus model, showed that in very large populations disruptive selection leads to stable intermediate gene frequencies,

though the change in frequency per cycle of selection is likely to be very slow. The existence of positive linkage disequilibrium was shown experimentally by Thoday and Boam (1959) but it is only recently that we have had a theoretical framework which allows us to understand its quantitative effects on the genetic variance of metric traits (Bulmer, 1971). Bulmer worked with a model of an infinite number of loci and showed that the changes in the genotypic variance caused by disruptive selection were due to exclusively the generation of positive linkage disequilibrium and developed formulae which predict such changes. These results were checked by computer simulation studies (Bulmer, 1976b)and it was found that observed and predicted values of disequilibria in equilibrium populations were in good agreement.

The purpose of this section is to extend Bulmer's results to an additive model with a finite number of loci with particular emphasis on experiments of short term duration. This work was stimulated by the results of a replicated disruptive selection experiment with Drosophila which is reported in the following chapter.

Changes in the Genotypic Variance Caused by Disruptive Selection.

In this section we study the effect of disruptive selection on changes of the genotypic variance. We shall deal with additive models of the type described in Chapter 3. We first deal with the change in gene frequencies and its effect on the equilibrium additive variance and in the proceeding section we study the generation of joint disequilibrium.

Changes in gene frequencies

We now assume that the metric trait is normally distributed and that a certain constant and equal proportion are saved for breeding at each generation. The truncation point at each end of the distribution corresponding to the proportion selected is M+T and M-T. The probability of selection of the $ij^{\underline{th}}$ genotype is:

$$W_{ij} = \frac{1}{\sigma\sqrt{2\pi}} \left\{ \int_{-\infty}^{-T} \exp\left(\frac{-(X-\bar{X}_{ij})^2}{2\sigma^2} \right) dx + \int_{T}^{\infty} \exp\left(\frac{-(X-\bar{X}_{ij})^2}{2\sigma^2} \right) dx \right\} \qquad (9.1)$$

As we did in the case of directional selection, we expand (9.1) in a Taylor series about the population mean, M, which after some manipulation yields the following second order approximation,

$$W_{ij} = 2Q + \frac{Qix_T}{\sigma^2} (\bar{X}_{ij}-M)^2 \qquad (9.2)$$

where, as before, Q is the proportion selected at each extreme of the distribution, x_T, is the point of truncation in standard deviation units corresponding to Q and i is the intensity of selection. The probability of selection of the $i^{\underline{th}}$ gametic phase is easily shown to be equal to,

$$W_i = 2Q + \frac{Qix_T}{\sigma^2}(Vw_i + (\bar{X}_i-M)^2) \qquad (9.3)$$

where Vw_i is the variance within the $i^{\underline{th}}$ gametic phase and $\bar{X}_i$ its mean. The mean fitness is given as a second order approximation by,

$$\bar{W} = 2Q + \frac{Qix_T}{\sigma^2}(2a_1^2p(1-p) + 2a_2^2q(1-q) + 4a_1a_2D) \qquad (9.4)$$

The change in the frequency of the plus allele at locus A is,

$$\Delta p = \frac{1}{\bar{W}}(f_1(W_1-\bar{W}) + f_2(W_2-\bar{W})) \qquad (9.5)$$

Substituting (9.3) and (9.4) in (9.5) and letting $\frac{2Q}{\bar{W}} = 1 - \frac{ix_T}{\sigma^2} VG$ to second order terms, we obtain:

$$\Delta p = \frac{ix_T}{2\sigma^2}(a_1^2p(1-p)(1-2p) + a_2^2(1-2q)D + 2a_1a_2(1-2p)D) \qquad (9.6)$$

which reduces to the expression obtained by Robertson (1956) if gene frequencies between loci A and B are uncorrelated. From (9.6) we can draw the following important conclusions:

(i) In large populations, gene frequency changes under disruptive selection are likely to be small if gene effects are not large and they tend to a stable intermediate equilibrium value and,

(ii) In laboratory experiments of relatively short term duration changes in the equilibrium additive variance due to changes of gene frequencies caused by disruptive selection are not likely to be detected. In fact, most of the changes in the equilibrium additive variance are likely to be due to genetic drift.

The Generation of Linkage Disequilibrium.

We now derive a second order approximation for the covariance of gene frequencies within gametes generated by disruptive selection for the two locus additive model. We first study the disequilibrium in the parental generation, before recombination takes place and we then extend the result to the offspring generation, allowing for recombination.

At generation t, amongst selected genotypes, we have:

$$D^{(s,t)} = (f_1 f_4 - f_2 f_3)^{(s,t)},$$

where

$$f_i^{(s,t)} = f_i^{(t)} + \{\frac{f_i}{\bar{W}}(W_i - \bar{W})\}^{(t)}, \quad i = 1,\ldots,4.$$

Using (9.3) and (9.4) and following the algebra through, the disequilibrium in the parental generation is given by,

$$D^{(s,t)} = D^{(t)} + \{\frac{ix_T}{\sigma^2} a_1 p(1-p) a_2 q(1-q)\}^{(t)} - \frac{ix_T}{2\sigma^2}\{2a_1 a_2 D^2 - (a_1(1-2p) + a_2(1-2q))^2 D\}^{(t)} \qquad (9.7)$$

This can be written,

$$D^{(s,t)} = D^{(t)} + D_f^{(t)},$$

where $D_f^{(t)}$ is the fresh disequilibrium generated in the $t^{\underline{th}}$ selection cycle.

When the population is initially in linkage equilibrium, (9.7) reduces to,

$$D^{(s,o)} = D^{(1)} = \frac{ix_T}{\sigma^2} a_1 p(1-p) a_2 q(1-q) \qquad (9.8)$$

For a given intensity of selection the initial generation of positive linkage disequilibrium is maximum when gene frequencies are intermediate. In contrast with the case of directional selection, $D^{(s,t)}$ is rather sensitive to the amount of selection applied. This is illustrated in Table 9.1 where values of ix_T are shown for different proportions selected. For comparison we also produce values of $i(i-x_T)$ corresponding to the directional selection situation.

TABLE 9.1:

Values of ix_T and $i(i-x_T)$ for different proportions selected. In the case of disruptive selection (ix_T), Q corresponds to the total proportion selected at both ends of the distribution. (For example, Q = 20% implies that 10% are selected at each extreme). For directional selection ($i(i-x_T)$), Q = 20% implies that 20% are selected at one extreme of the distribution.

	Q			
	1%	10%	20%	50%
ix_T	7.449	3.393	2.249	0.857
$i(i-x_T)$	0.918	0.821	0.781	0.637

The figures in the table also illustrate the fact that for the same proportion selected, the amount of disequilibrium generated by disruptive selection is considerably larger than the disequilibrium generated by directional selection, particularly for high selection intensities. For example, for a total proportion selected of 20%

disruptive selection generates almost three times more disequilibrium (of the opposite sign) than directional selection.

In the t+1 cycle of selection, the covariance of gene frequencies within gametes in the offspring generation is easily derived, since, as was shown before,

$$D^{(t+1)} = D^{(t)}(1 - \frac{W_{14}}{\bar{W}} c) + D_f^{(t+1)} \qquad (9.9)$$

The important point to notice is that, with close linkage, a larger proportion of the previously existing disequilibrium is passed on to the next generation. Since, with disruptive selection, D is positive, from (9.6) we conclude that, the tighter the linkage the higher the change in gene frequency at a given locus. This result is intuitively obvious, since it is clear that with disruptive selection, both extremes are favoured and once we have generated such combinations, we do not want to break them down. Similar results were arrived at empirically by Maynard Smith (1979) using deterministic simulations.

We now produce some numerical results to illustrate some of the conclusions we have drawn from this analysis. Table 9.2 shows the course of 10 generations of disruptive selection in an infinite population, for a two locus additive model. The results are obtained by numerical integration.

The results clearly indicate that gene frequencies do not change by a very substantial amount after 10 generations of selection even though gene effects are quite large. In agreement with theory the change in gene frequency is towards intermediate values and is larger

TABLE 9.2:

Gene frequencies (q) and linkage disequilibrium (D) in 10 generations of disruptive selection (10% at each extreme), for a two locus additive model. Gene effects and frequencies are the same at both loci and recombination fraction (c) is 0.5 and 0.01. Proportionate effects are 0.34 phenotypic standard deviations. Initial gene frequencies are set at 0.35.

	$q \times 10^2$		$D \times 10^5$	
t	c = 0.5	c = 0.01	c = 0.5	c = 0.01
0	35.00	35.00	0.00	0.00
2	36.77	36.77	210.29	269.48
4	38.66	38.84	287.64	553.53
6	40.39	41.01	319.81	833.80
8	41.90	43.07	335.62	1,094.00
10	43.20	44.86	344.70	1,324.04

with tight linkage but the difference is small. There are considerable differences in the amount of disequilibrium generated with the two degrees of linkage. This merely reflects the fact that in large populations undergoing disruptive selection, the tighter the linkage the smaller the relative 'loss' of the favourable combinations through recombination with the consequent increase in frequency of the coupling heterozygote over the repulsion heterozygote in successive generations.

These results can be extended to allow for an arbitrary number of loci. The simplest possible approach is to ignore the changes in the equilibrium additive variance on the grounds that in a very large population, gene frequency changes due to disruptive selection are small. It then follows that we can describe the process using the results of the infinitesimal model proposed by Bulmer. Thus, using the notation of earlier chapters, we can write, assuming free recombination and following Bulmer (1971):

$$CLW^{(t+1)} = \frac{ix_T}{\sigma^{2(t)}} VG^{2(t)} + \tfrac{1}{2}CLW^{(t-1)} \qquad (9.10)$$

where

$$VG^{(t)} = Vg^{(0)} + CLW^{(t)} \qquad (9.11)$$

where $Vg^{(0)}$ is the equilibrium additive variance in the base population. The validity of this approach is checked in the simulation work that follows.

Simulation Studies.

The simulation programme used in this work was developed from the one used to study directional selection. A subroutine which selects the lowest scoring males and females was incorporated into the programme. The N highest (H) and lowest (L) males (m) and females (f) out of a total of M individuals scored from each sex were selected and mated in the following way during t cycles of selection.

$\frac{1}{2}$ Nm (H) x $\frac{1}{2}$ Nf (H)

$\frac{1}{2}$ Nm (H) x $\frac{1}{2}$ Nf (L)

$\frac{1}{2}$ Nm (L) x $\frac{1}{2}$ Nf (H)

$\frac{1}{2}$ Nm (L) x $\frac{1}{2}$ Nf (L)

Each pair of mates contributed the same number of offspring to the next generation, 25% of which came from each of the four types of mating. The choice of which of the highest or lowest selected individual should contribute to a particular type of mating was completely at random. Under this type of mating regime the expected phenotypic correlation between mates amongst selected individuals is zero.

Table 9.3 summarises the genetic parameters of the various models studied. All these models assume additivity between and within the 30 loci which are uniformly distributed along the chromosome and 20 out of 200 individuals of each sex are selected at each extreme of the distribution.

TABLE 9.3: Summary of the input of the various models.

Model designation (q/c)	Rec. fraction (c)	Proportionate effect (a/σ)	Heritability (h^2)	Initial frequencies (q)
0.5/0.5	0.5	0.18	0.50	0.5
0.2/0.5	0.5	0.23	0.50	0.2
0.8/0.5	0.5	0.23	0.50	0.8
0.5/0.0	0.0	0.18	0.50	0.5
0.5/0.01	0.01	0.18	0.50	0.5
0.5/0.1	0.1	0.18	0.50	0.5
0.2;0.5/0.5	0.5	5 loci 0.46 25 loci 0.11	0.50	5 loci 0.2 25 loci 0.5

The first column in Table 9.3 shows the way the models are designated. The figure or figures at the left of the slash represent the initial gene frequencies. The figure at the right of the slash represents the recombination fraction which is the same for all adjacent loci.

Results and Discussion.

The results for the models with free recombination are shown in Table 9.4 and 9.5. The genetic parameters shown during the four cycles of disruptive selection are the total genotypic variance (VG), the amount of joint disequilibrium (CLW) and the realized heritability in the high (h^2_H) and low (h^2_L) selected fraction of the population. These realized heritabilities were calculated in the following way. For example, h^2_H is obtained by dividing the deviation of the mean of the offspring of the H x H matings from the mean of the preceding unselected population by the deviation of the mean of the H x H selected individuals from their contemporary (unselected) mean. These estimates are a description of the available genotypic variance for selection at a particular generation at each end of the distribution.

In both Tables 9.4 and 9.5, the difference between $VG^{(t)}$ and $CLW^{(t)}$ is an estimate of the equilibrium additive variance (ignoring a small effect due to departures from Hardy-Weinberg equilibrium). It will be noticed that in model 0.5/0.5, this difference is practically constant during the four cycles of selection, reflecting the fact that gene frequency changes during this period are minimal. In model 0.2; 0.5/0.5, where five loci have relatively large effect,

TABLE 9.4: Observed (O) and Predicted (P) values of genetic parameters after four cycles of selection, for models 0.5/0.5 and 0.2; 0.5/0.5. The predicted results are obtained from (9.10) and (9.11).

VG: total genotypic variance = Vg + CLW

h_H^2: single generation realized heritability in the high extreme

h_L^2: single generation realized heritability in the low extreme

Observed results are the average of 30 replicates. The standard errors of h^2 are 0.01.

		Model 0.5/0.5				Model 0.2; 0.5/0.5			
		VG	CLW	h_H^2	h_L^2	VG	CLW	h_H^2	h_L^2
0		14.79±0.16	-0.11±0.15	0.53	0.48	9.68±0.12	0.23±0.10	0.55	0.49
1	O	23.00±0.49	8.26±0.48	0.68	0.70	15.16±0.31	5.37±0.28	0.75	0.65
	P	23.43	8.43			14.90	5.36		
2	O	39.97±0.91	24.94±0.91	0.77	0.74	26.08±0.77	15.22±0.74	0.78	0.73
	P	35.30	20.30			22.44	12.91		
3	O	58.06±0.97	43.04±0.97	0.80	0.80	37.56±1.16	26.15±1.11	0.81	0.77
	P	53.02	38.02			33.70	24.18		
4	O	78.58±1.30	63.94±1.31	-	-	49.36±1.29	37.60±1.24	-	-
	P	80.50	65.50			51.17	41.64		

TABLE 9.5: Observed (O) and Predicted (P) values of genetic parameters for models 0.2/0.5 and 0.8/0.5. Observed results are the average of 30 replicates.

t		Model 0.2/0.5				Model 0.8/0.5			
		VG	CLW	h_H^2	h_L^2	VG	CLW	h_H^2	h_L^2
0		9.74±0.11	0.20±0.11	0.53	0.50	9.68±0.14	0.15±0.14	0.48	0.54
1	O	15.15±0.37	5.58±0.37	0.73	0.66	14.99±0.37	5.43±0.35	0.68	0.74
	P	15.00	5.40			15.00	5.40		
2	O	25.63±0.71	15.63±0.70	0.80	0.73	25.89±0.64	15.93±0.62	0.74	0.80
	P	22.59	12.99			22.59	12.99		
3	O	37.46±1.04	27.29±1.02	0.82	0.77	38.36±0.80	28.19±0.80	0.78	0.81
	P	33.93	24.33			33.93	24.33		
4	O	51.62±1.42	41.36±1.37	-	-	50.61±1.12	40.41±1.11		
	P	51.52	41.92			51.52	41.92		

there is a small increase in Vg from about 9.6 squared units to 11.7 at generation 4, reflecting the change in gene frequency of the major loci towards intermediate values as selection proceeds. The results in Table 9.6 show that even though initial gene frequencies are rather extreme, predicted results are in good agreement with observed results. In both runs, gene frequencies move slowly towards intermediate values. In run 0.2/0.5, the equilibrium additive variance at generations 0 and 4 was 9.6 and 10.0 respectively and the increase of the genotypic mean was of 1.72±0.11 units. In run 0.8/0.5, the value of Vg at generations 0 and 4 was of 9.6 and 10.0 squared units and the decrease in the genotypic mean was 1.56±0.15 units. The small degree of asymmetry in the observed realized heritabilities in both runs is as expected from theoretical considerations, this asymmetry tending to decline in later generations.

The predictions of joint disequilibrium are in good agreement with observed results. The increase in the genotypic variance due to the generation of joint disequilibrium is reflected in the increasing values of h^2 in both directions. This merely says that as selection proceeds and the genotypic means of the high and low matings move towards opposite extremes, the phenotype of an individual becomes a more accurate predictor of its genotype. This phenomenon of course is likely to cause departures from normality and will affect the estimates of heritability in that these will be different for different intensities of selection. The regression of offspring on parent, although it is symmetrical, is no longer linear. We shall discuss in more detail the development of the lack of linearity of the offspring

parent regression with disruptive selection in the next chapter.

Table 9.6 shows the simulation results of the amount of joint disequilibrium generated as selection proceeds with different degrees of linkage. Gene frequencies are initially at intermediate values in all runs.

TABLE 9.6:

Montecarlo simulation results of joint disequilibrium for various degrees of linkage (c) between adjacent loci. Average of 30 replicates in all runs. Proportion selected: 20/200 in each sex at each extreme.

	c			
t	0.5	0.1	0.01	0.00
0	-0.11±0.15	0.11±0.21	-0.12±0.15	-0.10±0.16
1	8.26±0.48	8.37±0.55	8.56±0.49	8.35±0.46
2	24.94±0.91	25.06±1.17	25.05±1.02	24.84±0.83
3	43.04±0.97	44.33±1.97	42.17±1.80	41.35±1.70
4	63.94±1.31	62.81±2.22	59.87±2.92	55.58±2.74

Up until generation two there is no detectable effect of linkage on the degree of joint disequilibrium generated. In later generations contrary to theoretical expectations based on deterministic models the tighter the linkage the smaller the amount of disequilibrium produced. This result is a consequence of the finiteness of the population. The maximum amount of disequilibrium

produced is when the selected individuals at each extreme of the distribution are fixed for either the plus or minus alleles. Consider the extreme case of no recombination. In this situation, we can do no better than fix the best gamete out of the initial sample. The probability of obtaining the desirable allele at all loci in the initial sample of gametes will depend on the gene frequency, the number of loci and the sample size. Provided the population size is not too small, the critical parameters are the number of loci and the gene frequencies. The larger the number of loci the higher the required initial gene frequency in the base population to have a given probability of obtaining the extreme gamete in the initial sample (Robertson, 1970a).

For a relatively small number of loci, provided the initial frequencies at all loci are not small, we will expect little effect of the degree of linkage on the amount of disequilibrium generated throughout the selection process, because we are likely to have picked the extreme gamete in the initial sample. If selection intensities are high enough that we can select our extremes from the H x H and L x L matings exclusively, we are likely to fix all the plus and minus alleles at both ends of the distribution and therefore CLW will reach its maximum possible value. On the other hand, with large number of loci, we have a very small probability of selecting the best possible gamete in the initial sample and if linkage is complete we cannot generate it through recombination as selection proceeds. We therefore expect a smaller degree of divergence between the mean of the high and low extremes and considerably less disequilibrium at fixation than in the case of free recombination.

These points are illustrated in Table 9.7. Models involving two different numbers of loci (4 and 20) and different degrees of linkage between adjacent loci were run for 20 generations of selection. As expected, with a small number of loci fixation is reached in a few generations and linkage has no detectable effect on either the amount of joint disequilibrium generated or on the number of generations required to reach the maximum possible divergence. With 20 loci however, linkage has a substantial effect on both parameters though these effects are small during the first few cycles of selection. Equivalent models to the ones presented in Table 5.8 with 50/200 selected at each extreme showed similar patterns though the difference was less marked. It is interesting to point out that in the case of the infinitesimal model, with large (infinite) population size, the selection intensities that we have been using in these simulations would lead to an increase of joint disequilibrium without bound. This is due to the fact that with intense selection we pick our extreme phenotypes from the extreme matings and theoretically the divergence does not reach a maximum possible value (Bulmer, 1976b).

These results have important consequences on estimators of heritability from populations which have undergone disruptive selection. In the case of intra-class correlations, as was shown before, all the fresh disequilibrium takes place between full-sib families and the within family component is not affected if loci recombine freely, because a proportion approximately equal to (1-2c) of the already existing disequilibrium takes place within full-sib families. This result was derived using a deterministic model and it does not hold in the case of

TABLE 9.7:

Montecarlo simulation results of joint disequilibrium (CLW) and ratio of observed mean of H x H matings to maximum possible value ($\bar{X}_H/\bar{X}_{MAX}$) assuming fixation of plus allele at all loci for 3 values of recombination fraction (c) between adjacent loci. In all runs, the highest and lowest 20 out of 200 are selected. Initial gene frequencies are 0.4 and the equilibrium additive variance is 9.6 for all models. The results of the models involving 4 loci are the average of 15 replicates; those of 20 loci are based on 7 replicates.

		4 loci		20 loci	
t	c	CLW	$\frac{\bar{X}_H}{\bar{X}_{MAX}}$		$\frac{\bar{X}_H}{\bar{X}_{MAX}}$
	0.50	-0.13±0.15	0.40	0.04±0.18	0.40
0	0.05	-0.10±0.14	0.40	-0.26±0.18	0.40
	0.00	0.10±0.14	0.40	-0.16±0.18	0.40
	0.50	12.62±0.71	0.76	16.02±1.12	0.54
2	0.05	13.14±0.84	0.77	13.48±1.13	0.54
	0.00	12.85±0.72	0.76	14.14±1.18	0.54
	0.50	26.31±0.52	0.95	59.83±2.92	0.68
5	0.05	26.65±0.52	0.95	44.43±5.59	0.66
	0.00	26.36±0.64	0.94	42.40±3.08	0.66
	0.50	30.19±0.10	1.00	108.89±5.39	0.82
10	0.05	30.38±0.20	1.00	86.94±6.42	0.77
	0.00	30.38±0.20	1.00	53.98±5.75	0.70
	0.50	30.10±0.05	1.00	176.70±4.77	0.97
20	0.05	30.19±0.09	1.00	135.41±7.94	0.87
	0.00	30.21±0.10	1.00	59.32±6.73	0.71

disruptive selection if populations are of finite size. Our simulation results suggest that the degree of linkage only affects the between family component through its effect on the amount of disequilibrium generated, but the within family component is not affected by it. The genetic component of the variance within families tends to zero as the extreme gametes tend to fixation.

Table 9.8 shows the variance components within and between full-sib families, the intraclass correlation and the single generation realized heritability of the high selected extremes for the models of Table 9.7. The corresponding estimate of h^2 for the low end of the distribution did not differ from that of the high and is omitted. For brevity, only the results corresponding to complete linkage and free recombination are shown in the table.

The intra-class correlation at generation t is obtained from an analysis of the phenotypic variance of the offspring of generation t+1. In other words, we are estimating the genotypic variance amongst individuals selected at both ends of the distribution at generation t. For comparison we include the single generation realized heritabilities in the high direction. This estimate at generation t is based on regressing the deviation of the mean of the offspring of the H x H matings at generation t+1 from the unselected mean at generation t, on the selection differential. The most conspicuous feature of the results is the large bias upwards of the intraclass correlation as an estimator of the heritability due to the effect of the joint disequilibrium on the between family component. It is also clear that the within family component is not affected by the degree

TABLE 9.8:

Variance component between full-sib families (σ^2_b), within full-sib families (σ^2_w) and intra-class correlation (t) for the models of Table 9.7. The environmental variance in both models is of 9.6 square units. The standard errors of t and of the single generation realized heritabilities (h^2_r) are 0.01. $h^{2(0)}$ = base population heritability.

t	c	σ^2_w	σ^2_b	t	h^2_r
		4 loci $h^{2(0)} = 0.50$			
1	0.5	13.54±0.28	10.50±0.36	0.44	0.74
	0.0	13.44±0.39	10.12±0.47	0.43	0.74
2	0.5	13.01±0.23	20.24±0.62	0.60	0.75
	0.0	13.71±0.39	19.83±0.75	0.59	0.78
5	0.5	10.44±0.15	36.51±1.09	0.78	0.77
	0.0	10.66±0.25	36.09±0.94	0.77	0.74
10	0.5	9.81±0.22	40.98±0.64	0.81	0.75
	0.0	10.02±0.22	40.93±0.63	0.80	0.75
20	0.5	9.98±0.23	41.29±0.46	0.81	0.77
	0.0	10.06±0.25	41.30±0.45	0.80	0.77
		20 loci $h^{2(0)} = 0.50$			
1	0.5	14.30±0.37	12.50±0.47	0.47	0.67
	0.0	13.46±0.54	10.39±0.66	0.43	0.69
2	0.5	14.60±0.41	21.75±1.06	0.60	0.77
	0.0	15.53±0.56	19.56±1.51	0.55	0.75
5	0.5	13.08±0.84	71.51±3.77	0.84	0.81
	0.0	14.47±0.61	50.54±3.84	0.77	0.77
10	0.5	11.03±0.56	115.39±5.70	0.91	0.81
	0.0	9.52±0.71	61.04±5.78	0.86	0.77
20	0.5	9.86±0.51	191.50±0.76	0.95	0.86
	0.0	9.53±0.58	68.24±6.41	0.87	0.77

of linkage and that its genetic component tends to zero as extreme gametes reach fixation.

Conclusions

The main findings of this work can be summarised as follows:

(i) Using a two locus additive model, we have shown that under disruptive selection, gene frequency changes are of order $(a/\sigma)^2$ and that in large populations, gene frequencies tend to move towards stable intermediate values. These results are in agreement with those obtained previously by Robertson (1956) who worked with single locus models. In large populations, changes in gene frequency increase with tight linkage.

(ii) Recurrence equations are developed for the amount of disequilibrium generated with disruptive selection under a two locus additive model. In contrast with the case of directional selection, the generation of disequilibrium is rather sensitive to the intensity of selection applied and it is perfectly symmetrical. In other words, the same amount of positive disequilibrium is generated with models whose initial frequencies are equidistant from 0.5.

(iii) Contrary to predictions based on deterministic models, we have shown that with populations of finite size, the tighter the linkage the smaller the amount of disequilibrium produced. This effect however is not important during the early cycles of disruptive selection.

(iv) We suggest that in populations undergoing disruptive selection, the difference between heritability estimates based on single generation realized heritabilities and on intra-class correlation between sibs provides evidence of the existence of joint disequilibrium generated during selection. Problems of lack of linearity of offspring parent regressions which develop as a consequence of disruptive selection are mentioned and will be discussed in the next chapter.

As was mentioned at the beginning of this chapter, this work was stimulated by a disruptive selection experiment carried on with Drosophila. Some of these theoretical results however may be relevant to the question posed initially by Fisher (1930) and more recently by Maynard Smith (1978), namely: what selective forces maintain sexual reproduction and genetic recombination in nature? In a recent paper, Maynard Smith (1979) concluded that both normalising and disruptive selection are forces that tend to reduce recombination. We have shown that in the case of disruptive selection, this is true for infinite populations. With finite population size however, this result does not seem to hold, at least for the rather extreme model of selection studied here. Further work on this area may help towards the elucidation of this challenging problem.

CHAPTER 10

EFFECT OF DISRUPTIVE SELECTION ON GENETIC VARIABILITY - EXPERIMENTS WITH *DROSOPHILA*

Introduction.

The consequences of disruptive selection on the genotypic variance of a metric trait have been well established over the years. Robertson (1956) working with a single locus model, showed that selection of extremes in large populations will cause gene frequencies to move towards stable intermediate values, but the rate of change per generation is likely to be very slow. Mather (1941, 1943) argued that the higher fitness of metric intermediates will lead to the build up of repulsion linkages. Mather's argument in a disruptive selection context implies the build up of coupling linkages or positive linkage disequilibrium. Assuming an additive model, both gene frequencies moving towards intermediate values and the generation of positive linkage disequilibrium will lead to an increase of the genotypic variance.

Substantial increases in the additive genetic variance of metric traits in disruptive selection experiments have been reported by various workers (Thoday, 1959; Millicent & Thoday, 1961; Gibson & Thoday, 1963; Scharloo, 1964; Scharloo et al., 1967; Barker & Cummins, 1969) and Thoday and Boam (1959) have provided evidence for the maintenance of coupling linkage disequilibrium.

It is only recently, however, that we have had a theoretical framework which allows us to quantify the effect of disruptive selection on the genotypic variance of a metric trait (Bulmer, 1971). Bulmer worked with a model of an infinite number of loci and showed that the changes in the genotypic variance caused by disruptive selection were due to exclusively the generation of positive linkage

disequilibrium and developed formulae which predict such changes. The consequences of introducing a finite number of loci into the model were reported in the previous chapter.

In view of the rather inconclusive results of the directional selection experiments with Drosophila described in Chapter 8 aimed at studying the generation of joint disequilibrium due to selection, it was decided to perform experiments on disruptive selection. Selection of extremes leads to relatively larger generation of joint disequilibrium and therefore its effects on the genotypic variance are more likely to be detected experimentally. The present short term experiment with Drosophila melanogaster was designed as a check on the theory developed by Bulmer.

Basically, the experiment consisted of carrying out three cycles of disruptive selection, followed by a period of relaxation. Heritability was estimated during the period of selection and at the end of the period of random mating. On the assumption of Bulmer's theory we anticipate an increase in the heritability during selection, due to the build up of positive linkage disequilibrium followed by a decline at the end of the period of random mating, presumably due to the breakdown of the joint disequilibrium.

Material and Methods.

The lines were derived from the Dahomey population. This population originated from a large sample (numbers unknown) of flies collected in West Africa in 1969. Since then, a number of cage populations have been maintained in this laboratory from which samples

were taken to originate the lines.

The character measured was the sum of the abdominal bristles on the fourth and fifth segments in males and fifth and sixth segments in females.

The experiment was run with two replicates. In replicate 1 eggs were sampled from the population cage in four half-pint milk bottles. When the adults emerged, 160 males and 160 virgin females, sampled in equal numbers from each bottle were scored and constituted generation 0. The highest (H) and lowest (L) 16 males and 16 females were selected and mated in individual vials in the following way:

Number of full-sib families	Males x Females	Number of offspring contributed by each mating pair
8	H x H	5
8	H x L	5
8	L x H	5
8	L x L	5

Within each type of mating, flies were paired at random and the choice of which flies within each extreme should be mated with high or low partners was also random. The expected phenotypic correlation between mates is therefore, 0. At generation 1, 5 males and 5 females from each full-sib family (vial) were chosen at random from those first emerging (from the first 36 hrs of emergence). The males and the virgin females were aged in vials for three days and after scoring they were mated as described above, with several spare

matings kept until the hatching of the larvae to replace unsuccessful matings. At generation 3, after scoring, one male and one female were chosen at random from each of the 32 families and introduced into a half-pint milk bottle for random mating. After 24 hrs all flies were shaken off into another bottle without etherization and allowed to lay eggs for about 8 hours. This random mating procedure was continued for 7 generations. At generation 10, after 7 cycles of random mating, and at generation 11, after 8 cycles of random mating, 160 flies of each sex were sampled and scored from the half-pint milk bottles and the extremes selected and mated as described above following which the replicate was discontinued.

Replicate 2 differed slightly from replicate 1 in that flies that contributed to generation zero were themselves reared in vials. Further, the cycles of random mating were carried on in vials rather than in bottles, each family contributing one male and one female to the next generation and this procedure was continued for 4 generations rather than 7. The replicates were not run contemporaneously.

The flies were reared in standard Edinburgh agar-molasses-killed yeast medium in which drops of live yeast had been added. The cultures were kept in a room at constant temperature (25^{o}C) and lit continuously for 24 hours.

In both replicates, heritability was estimated from the regression of offspring on the selected mid-parental values, at generation 0, 1 and 2, and at the end of the period of relaxation.

Results.

Estimates of various base population parameters from each replicate are shown in Table 10.1. The first cycle of selection provided estimates of heritabi lity of the base population based on offspring-midparent regressions.

TABLE 10.1:

Parameter estimates from each replicate

$\bar{X}_m$: mean of males

$\bar{X}_f$: mean of females

$VP_{(m)}$: Phenotypic variance (males)

$VP_{(f)}$: Phenotypic variance (females)

(1) : heritability estimates based on offspring-midparent regressions.

Means and phenotypic variances based on 160 observations.

	$\bar{X}_m$	$\bar{X}_f$	$VP_{(m)}$	$VP_{(f)}$	h^2 (1)
Replicate 1	33.48	38.89	8.19	7.32	0.38 ± 0.07
Replicate 2	35.61	40.21	7.52	9.02	0.35 ± 0.08

The heritability estimates of both replicates are in reasonable agreement with those reported in Chapter 8. The means of replicate 2 are significantly higher than those of replicate 1 and this may be a consequence of the fact that at generation zero, flies in replicate 2 were reared in vials whilst those of replicate 1 were reared in bottles.

Tables 10.2 and 10.3 show the means of the offspring of the different mating types for both replicates and Figure 10.1, shows the means of the offsprings of the H x H and L x L matings expressed as deviations from their contemporary means. The results show that there is good agreement between the responses of the H x H and L x L matings in both replicates and no signs of asymmetry are suggested by the data. In both replicates, the means of the H x L and L x H are very similar and the difference between their average and the contemporary mean does not differ significantly from zero, suggesting that neither sex-linkage nor dominance gene action are important in this character.

In replicate 1 the overall mean increases from generation 3 to generation 6 and it remains at a value between 38 and 39 bristles until the end of the experiment. As no controls were used it is difficult to assess whether the change is due to an environmental trend, due to natural selection acting against the low deviants or due to drift. Since in this replicate the flies were kept in bottles during the period of relaxation, there is no control over the effective population size and any bottlenecks during this period can cause substantial changes in both mean and variance. No signs of trends in the overall mean are present in replicate 2.

Figures 10.2 and 10.3 show the frequency distribution of the total number of bristles of individual females for various periods of the selection programme for replicate 1 and 2 respectively (the distribution in males follow similar patterns and are omitted). We have also included the frequency distribution of the offspring of the

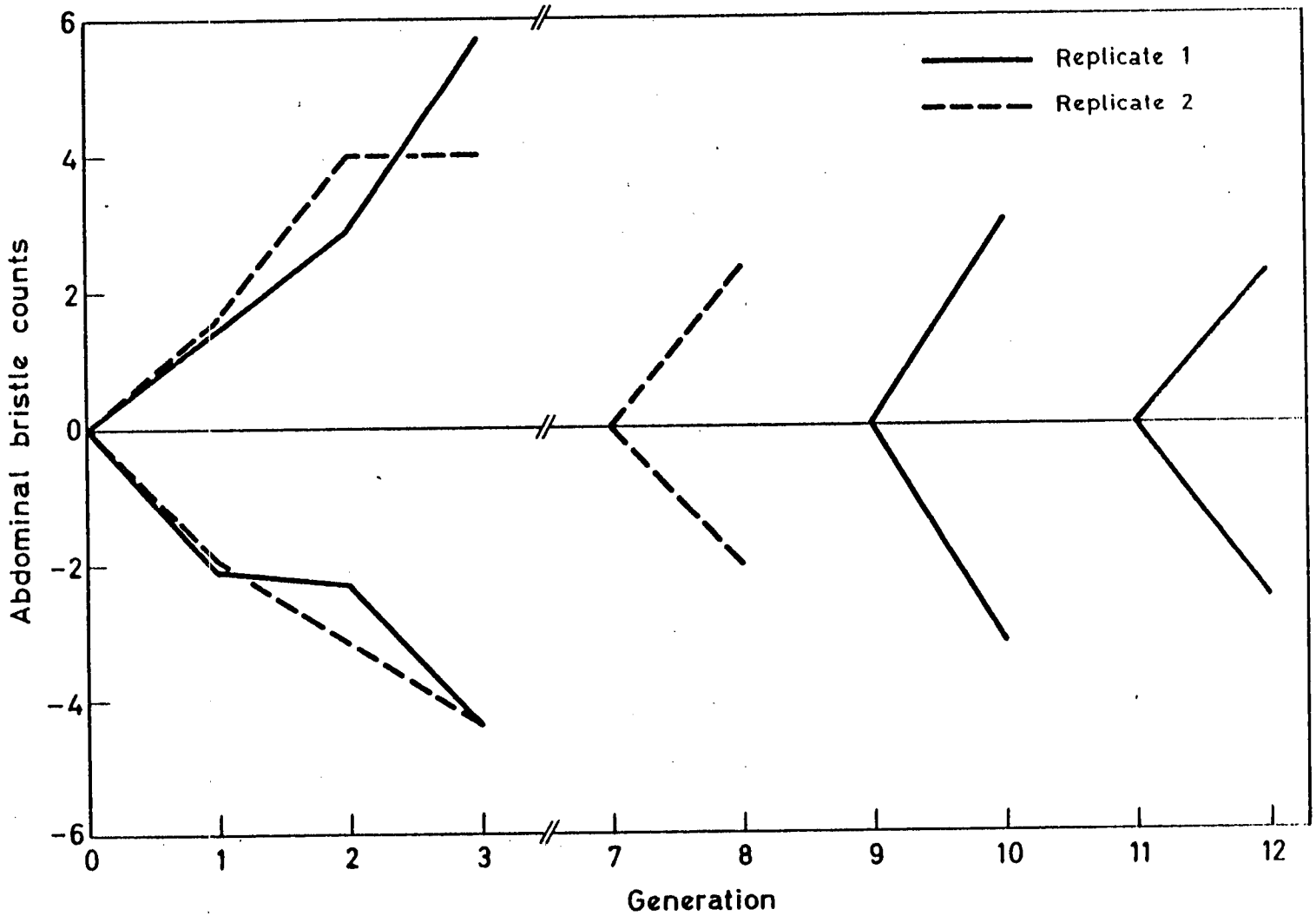
TABLE 10.2: Means of the offspring of the different type of matings in replicate 1

Generation	H x H	H x L	L x H	L x L	Overall Mean
1	37.2±0.4	35.7±0.4	36.8±0.4	33.8±0.4	35.8±0.2
2	38.4±0.4	35.2±0.4	35.3±0.4	33.3±0.4	35.6±0.2
3	41.7±0.6	34.9±0.6	35.7±0.6	31.6±0.6	36.0±0.3
6					39.1±0.3
9					39.0±0.2
10	41.0±0.5	37.8±0.5	38.3±0.5	34.8±0.5	38.0±0.2
12	41.6±0.5	39.2±0.5	39.7±0.5	36.8±0.5	39.3±0.2

TABLE 10.3: Means of the offspring of the different types of mating in replicate 2.

Generation	H x H	H x L	L x H	L x L	Overall Mean
1	38.5±0.3	37.3±0.3	37.2±0.3	35.0±0.3	37.0±0.2
2	42.1±0.5	37.2±0.5	38.0±0.5	34.9±0.5	38.1±0.3
3	41.3±0.5	36.9±0.5	38.1±0.5	32.9±0.5	37.3±0.3
8	39.6±0.4	36.7±0.4	37.5±0.4	35.3±0.4	37.3±0.2

FIGURE 10.1: Means of abdominal bristle scores of H x H and L x L offspring expressed as deviations from contemporary mean. (Average of males and females).



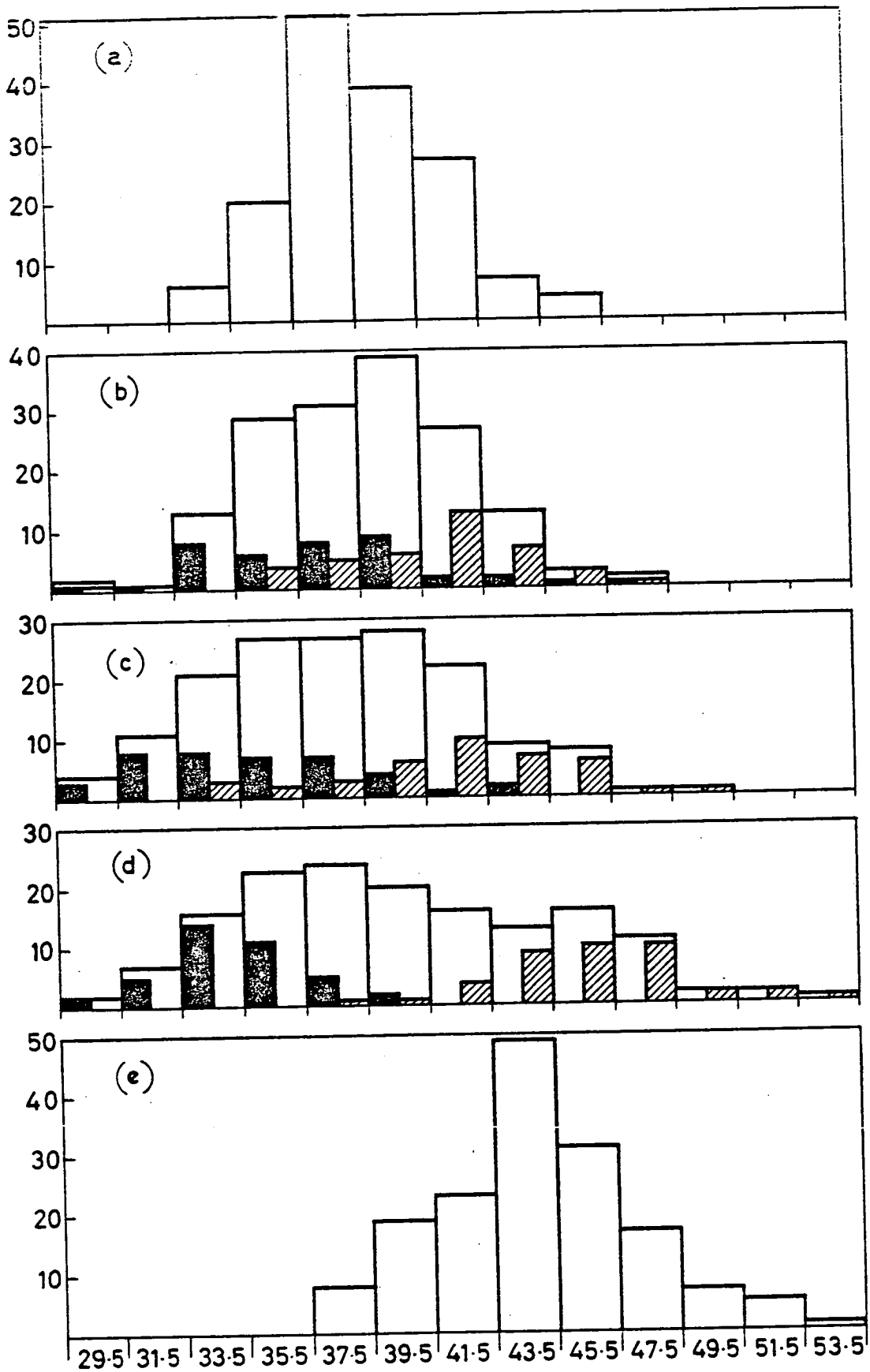


FIGURE 10.2: Replicate 1. Frequency distribution of abdominal bristle scores in females at generation zero (a) during the three cycles of selection (b,c,d) and at the end of the period of relaxation (e). The solid columns refer to the offspring of the L x L matings and the cross hatched columns to the offspring of the H x H matings.

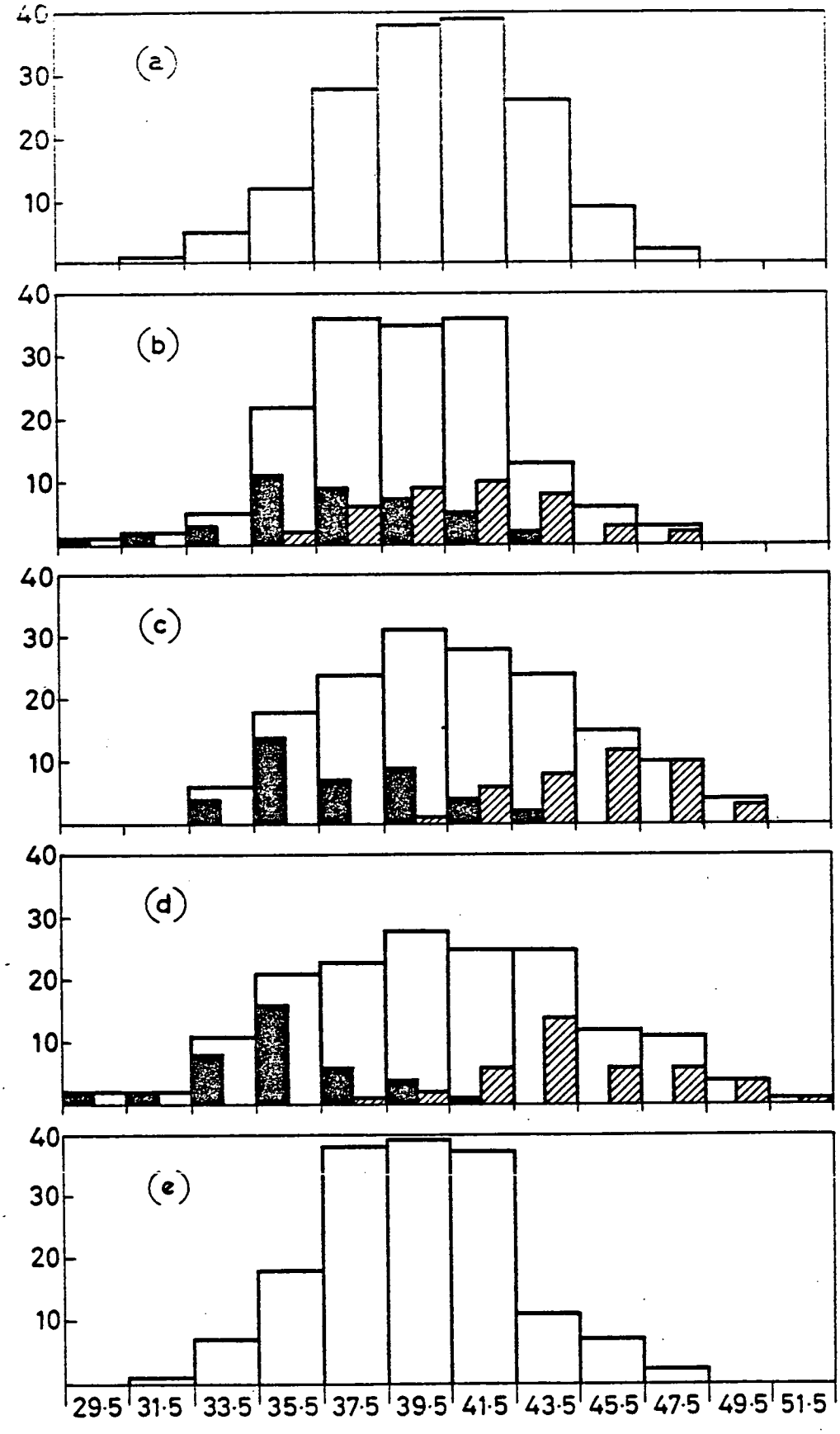


FIGURE 10.3:

Replicate 2. Frequency distribution of abdominal bristle scores in females at generation zero (a) during the three cycles of selection (b,c,d) and at the end of the period of relaxation (e). The solid columns refer to the offspring of the L x L matings and the cross hatched columns to the offspring of the H x H matings.

H x H and L x L matings. As expected, as selection proceeds the overall distribution becomes more platykurtic and the H x H and L x L distributions tend to move apart. After three cycles of selection there is very little overlap between the H x H and L x L distributions and the result is consistent in both replicates. This result is also illustrated in Table 10.4 (a and b) where we show the observed (and expected) proportions of flies selected at each extreme of the distribution coming from each of the four types of mating in the previous generation. It is clear that as the two extremes become differentiated, the phenotype of an individual becomes a better predictor of its breeding value and therefore the choice at each end of the distribution is more accurate. This is another way of saying that this type of selection causes substantial amounts of positive linkage disequilibrium with consequent increases in the heritability.

The difference between observed and predicted contributions (numbers in brackets in the table) partly give an indication of the degree of unsuccessful matings. These were low and non-fertile matings were not associated with any given mating type.

Tables 10.5 (a and b) show estimates of offspring-midparent regressions and estimates of intraclass correlations obtained from analysis of variance between full-sibs during different stages of the selection programme in both replicates. The positive build up of linkage disequilibrium is clearly demonstrated in the estimates of heritability by offspring-midparent regressions. As was shown in the previous chapter, disruptive selection in the parental generation causes the estimates of heritability from intra-class

TABLE 10.4a: Replicate 1.

Generation	H x H	H x L(L x H)	L x L
Proportion of high extremes selected from the offspring of each type of mating.			
1	0.41 (0.47)	0.41 (0.38)	0.18 (0.15)
2	0.84 (0.84)	0.16 (0.16)	0.00 (0.00)
Proportion of low extremes selected from the offspring of each type of mating.			
1	0.09 (0.09)	0.25 (0.13)	0.66 (0.780
2	0.06 (0.06)	0.22 (0.22)	0.72 (0.72)

TABLE 10.4b: Replicate 2.

Generation	H x H	H x L(L x H)	L x L
Proportion of high flies selected from the offspring of each type of mating.			
1	0.62 (0.62)*	0.34 (0.34)	0.04 (0.04)
2	0.88 (0.88)	0.09 (0.09)	0.03 (0.03)
Proportion of low flies selected from the offspring of each type of mating.			
1	0.07 (0.07)	0.34 (0.24)	0.59 (0.69)
2	0.00 (0.00)	0.31 (0.31)	0.69 (0.69)

*In both tables, the numbers in brackets refer to the expected proportion based on the selected mating. Due to infertility spare matings were occasionally used and this effect is partly reflected in the difference between the number in brackets and the number directly on top of it.

TABLE 10.5a:

Estimates of heritability obtained from offspring midparent regressions and intraclass correlations between full-sib families. The intraclass correlation at generation t is obtained from the analysis of the offspring at generation t whose parents were selected in the previous generation.

Generation	$b_{O\bar{P}}$	2 x t_c
0	0.38±0.07	0.77±0.12
1	0.54±0.09	1.00±0.11
2	0.74±0.07	1.25±0.13
9	0.47±0.04	0.61±0.13
11	0.47±0.06	0.66±0.14

TABLE 10.5b: Replicate 2.

Generation	$b_{O\bar{P}}$	2 x t_c
		(0.24±0.14) [1]
0	0.35±0.08	0.88±0.12
1	0.72±0.07	1.03±0.11
2	0.64±0.07	1.09±0.11
		(0.46±0.13) [1]
7	0.44±0.07	0.77±0.12

[1] The structure of the data in this replicate allowed estimation of the intraclass correlation before the operation of selection, both at generation 0 and at the end of the period of relaxation. In both tables, $b_{O\bar{P}}$ refers to the offspring-midparent regression and t_c is the intraclass correlation between full-sibs.

correlations to be biased upwards and this is clearly reflected in the results of both replicates.

During the period of relaxation, the disequilibrium breaks down at a rate approximately equal to $(1-\bar{c})$ (or slightly less if account is taken of the finiteness of the population) and this is again reflected in the $b_{O\bar{P}}$ of both replicates and the relatively smaller degree of bias of the intra-class correlation. The decline in variance during the period of relaxation is also illustrated in Figures 10.2(e) and 10.3(e), where the frequency distribution of total counts are shown at the end of the experiment for both replicates.

The expected decline in variance in replicate 1 during the 6 cycles of random mating due to the effect of genetic drift, using estimates of the ratio of effective to actual numbers reported by Crow and Morton (1955) is of the order of 6%. In replicate 2, where flies were kept in vials during the four generations of relaxation and each family contributed with equal members to the next generation, the expected decline in variance due to drift is of about 2%. In both replicates the observed decline in variance is well in excess of the expected decline due to drift alone.

The mean squared difference between both segments can be regarded as one of the components of the environmental variance, namely, that one due to the effects of local accidents of development which prevent perfect replication of the same phenotype under the same environmental conditions. This within fly variance was calculated in each generation in both replicates and it remained

virtually unchanged throughout the selection programme at a value of about 4 square units.

Table 10.6 shows observed (O) and predicted (P) values of offspring-midparent regressions. Observed results are obtained by pooling the estimates obtained from both replicates whenever these estimates were obtained from the same cycle of selection. Predicted results are based on the ratio of genotypic to phenotypic variance obtained using Bulmer's predictions with a model which assumes 30 additive loci of equal effects distributed on 3 chromosomes, the recombination fraction between adjacent loci being 0.1. The equilibrium additive variance at generation zero was assumed to be 3.7 and the phenotypic variance, 10. The mapping function used to obtain the mean recombination fraction was, $c = \frac{1}{4}(1-e^{-2x})$ for loci on the same chromosome and $c = \frac{1}{2}$ for loci on different chromosomes.

The data tend to suggest that predicted results tend to underestimate the estimates based on offspring midparent regressions, particularly during the period when disruptive selection is operating. This effect seems to disappear after the cycles of random mating.

Discussion.

The purpose of this work has been to provide an experimental check on the theory developed by Bulmer (1971, 1974). The results obtained are consistent with the expectation that disruptive selection causes positive linkage disequilibrium which leads to an increase in the heritability of the metric trait. Evidence for the existence of joint disequilibrium is provided by the reduction in heritability during the period of random mating and also by the substantial bias in the intra-class correlations.

TABLE 10.6.

Observed (O) and predicted (P) heritability estimates during the disruptive selection experiment.

$h^2(O)$: pooled offspring mid-parent regressions.

$h^2(P)$: predicted results based on Bulmer's theory, from the ratio of the genotypic to phenotypic variance.

Generation (Selection)	$h^2(O)$	$h^2(P)$
0	0.37±0.05	(0.37)
1	0.65±0.06	0.47
2	0.68±0.05	0.55
Number of cycles of random mating		
4	0.44±0.07*	0.50
7	0.47±0.04**	0.43
8	0.47±0.05**	0.41

* obtained from replicate 2 only.

** obtained from replicate 1 only.

The results however seem to suggest that observed estimates of heritability are larger than the predicted ratio of genotypic to phenotypic variance. If this is a real phenomenon, two possible reasons may account for it. First, the predicted results may be model dependent and therefore different combinations of the number, frequencies, effects and recombination values of the loci affecting the trait, for the same initial genetic parameters may yield predictions in closer agreement with observed results. Alternatively, it is possible that our estimates of the changes in heritability based on the regressions of offspring on selected parents may be biased due to departures from normality generated by this type of selection as is clearly illustrated in Figures 10.2 and 10.3.

In the previous chapter, we have shown that, particularly in the early cycles of selection, the predictions of the generation of joint disequilibrium based on the infinitesimal model are in good agreement with Montecarlo simulation results. This provides some evidence against the model dependence argument. Further evidence is provided in Table 10.7 where Montecarlo simulation results are shown for three genetic models involving different degrees of linkage, proportionate effects of the loci involved and initial gene frequencies for the same initial values of genotypic and phenotypic variance of the model used in Table 10.6, taken from the Drosophila experiment. The simulation programme assumes the same mating structure as the Drosophila experiment, with 16 males and females selected at each end of the distribution. The first model is based on 30 loci of equal effects and frequencies with free recombination.

We symbolise it (30,0.5,0.5). The second model is equivalent to the first but the 30 loci are distributed in 3 chromosomes with 10 loci on each chromosome and recombination value between adjacent loci is 0.1. We symbolise it (30;0.5;0.1). The third model which assumes free recombination, is based on five loci at initial frequency of 0.1 and proportionate effects, a/σ, of 0.45 and 25 loci at initial frequency of 0.4 and proportionate effect 0.13. We symbolise it (5/25,0.1/0.4,0.5).

The results in Table 10.7 show that observed and predicted values of disequilibrium in this short term selection study are in good agreement and that there is no clear difference among models.

Table 10.8 provides evidence which supports the suggestion that the lack of agreement between observed and predicted results shown in Table 10.6 is due to a problem of non-linearity of the offspring parent regression.

The simulation results show that the ratio of genotypic to phenotypic variance is in very close agreement with the results predicted on the basis of the infinitesimal model and furthermore, the realized heritabilities in all models are remarkably similar to the estimates of heritability based on offspring mid-parent regressions of the Drosophila experiment.

The discrepancy between the ratio of genotypic to phenotypic variance and the realized heritabilities obtained from the selection of extreme deviants can be explained in the following way. The first cycle of disruptive selection produces considerable divergence between the means of the offspring of the H x H and L x L matings.

TABLE 10.7:

Montecarlo simulation results (O) and predicted results (P) of joint disequilibrium for three genetic models with equal equilibrium additive variance and phenotypic variance at generation zero. Predicted results based on Bulmer's theory. Observed results are the average of 10 replicates.

Generation		Model (30,0.5,0.5)	(30,0.5,0.1)	(5/25;0.1/0.4,0.5)
0	O	-0.04 ± 0.09	0.06 ± 0.08	-0.14 ± 0.09
	P	0.00	0.00	0.00
1	O	1.30 ± 0.14	1.74 ± 0.18	1.54 ± 0.18
	P	1.61	1.61	1.61
2	O	4.20 ± 0.39	5.08 ± 0.41	3.93 ± 0.41
	P	3.64	3.76	3.64
3	O	8.04 ± 0.71	8.49 ± 0.66	7.81 ± 0.75
	P	6.36	6.92	6.36
4	O	11.12 ± 0.80	12.00 ± 0.92	10.55 ± 1.24
	P	10.25	11.64	10.25

TABLE 10.8:

Montecarlo simulation results of the ratio of the genotypic to phenotypic variance (VG/VP) and realized heritabilities of the H x H matings (h^2_H). The corresponding realized heritabilities of the L x L matings are similar and are omited. $\frac{VG}{VP}(P)$ are the predicted results based on Bulmer's theory. The standard errors of h^2_H are about 0.03. σ_b^2 and σ_w^2 are the variance components between and within families respectively.

	Model						
Generation of selection.	30,0.5,0.5		30,0.5,0.1		5/25,0.1/0.4,0.5		
	$\frac{VG}{VP}$	h^2_H	$\frac{VG}{VP}$	h^2_H	$\frac{VG}{VP}$	h^2_H	$\frac{VG}{VP}(P)$
0	0.38	0.35	0.37	0.38	0.38	0.38	0.38
1	0.45	0.60	0.44	0.58	0.46	0.56	0.47
2	0.56	0.67	0.56	0.67	0.58	0.67	0.56
3	0.66	0.70	0.66	0.72	0.67	0.69	0.64
4	0.73	0.75	0.72	0.77	0.73	0.77	0.71
Generation	σ_w^2	σ_b^2	σ_w^2	σ_b^2	σ_w^2	σ_b^2	
1	8.03	3.31	8.10	3.44	8.02	2.95	
2	8.08	6.40	8.12	7.36	8.21	5.97	
3	7.72	10.81	8.00	11.51	7.78	10.63	
4	7.94	15.51	8.03	16.39	8.01	15.12	

In fact, this cycle of selection generates what we can consider to be three subpopulations, originated from the offspring of the H x H, L x L and LH(HL) matings. The amount of disequilibrium generated is a description of the genetic situation in the whole population. However, in a second cycle of selection, in the high extreme say, the proportion of the selected individuals which are selected from the offspring of the H x H matings will generate offspring whose mean will tend to regress towards the mean of the subpopulation they were selected from rather than to the overall population mean. This clearly causes a higher heritability than the one we would obtain if there were no genetic differentiation among subpopulations and all the mated individuals generated offspring whose mean would tend to regress towards the (single) population mean. When the gametes at each end of the distribution reached fixation, all the variance within the high, low and their combination is environmental. It follows that at this stage, very intense selection will lead to smaller realized heritabilities than those obtained from less intense selection provided that, in both cases, the extremes are all chosen from the extreme genotypes. This is because in both cases, the selection response is the same, regardless of the selection intensity, but the selection differential is smaller with less intense selection. This point is illustrated with Montecarlo simulation results in which a model of four loci reaches fixation after a few cycles of intense disruptive selection. At the end of the selection programme, the realized heritability is obtained by selecting the highest 50 out of 200 of each sex, or the highest 20 out of 200 of

each sex. In the first case, the estimate is 0.95 ± 0.01 and in the second case is 0.75 ± 0.01.

More generally, it can be shown that the relationship between the conditional offspring means and the parental values is of a double sigmoid type, the single sigmoid relationship for values higher than the mean being a mirror image of the one for values smaller than the mean. It then follows that the response to selection of the same intensity at each end of the distribution is symmetrical but depends on the selection pressure applied. The ratio of the genotypic to phenotypic variance, as predicted from results based on the infinitesimal model is a linear description of the expected response to selection in a situation where the selection forces per se lead inevitably to non-linear relationships. Our experimental results would have been in closer agreement with predicted results based on infinitesimal model theory had we estimated the heritability from regressions of offspring on non-selected, randomly mated parents and fitted (incorrectly) a linear regression equation through the data.

Could we get more insight by studying the effects of disruptive selection from a different point of view? The experimental design we have used allows us to follow the experiment as if it were a two way selection experiment. As was mentioned earlier, after a first cycle of selection and mating, one can consider the whole population as being composed of a mixture of various normally distributed sub-populations, corresponding to the offspring of the H x H, H x L

(and L x H) and L x L matings, in a ratio 1 : 2 : 1. We have designated these three subpopulations, H, HL and L. The mean of the H x H mating expressed as a deviation from the overall mean is $ih^{2(0)}\sigma^{(0)}$ and likewise, the mean of the L x L is $-ih^{2(0)}\sigma^{(0)}$. The H x L (and L x H) matings yield a distribution with zero mean and the phenotypic variance within each of the three distributions is $\sigma^{2(0)}(1 - \frac{1}{2}i(i-x_T)h^{4(0)})$. It then follows that the overall distributions, has zero mean and variance equal to

$$\sigma^{2(1)} = \sigma^{2(0)}(1 - \tfrac{1}{2}i(i-x_T)h^{4(0)}) + \tfrac{1}{2}i^2h^4\sigma^{2(0)}$$

$$= \sigma^{2(0)}(1 + \tfrac{1}{2}ix_T h^{4(0)}),$$

as obtained by Bulmer (1971). We now apply a second cycle of selection, and those individuals exceeding a truncation point x_T (in units of $\sqrt{\sigma^{2(1)}}$) standard units from the overall mean in the mixed distribution at each extreme are saved for breeding. Assuming normality, it is possible to obtain the proportion of individuals which belong to each subpopulation, whose heritability has been reduced from $h_w^{2(0)}$ to $h_w^{2(1)}$=

$$= \frac{VG_w^{(0)}(1 - \frac{1}{2}i(i-x_T)h^{2(0)})}{\sigma_w^{2(0)}(1 - \frac{1}{2}i(i-x_T)h^{4(0)})}$$

the subscript w referring to the within population parameters. In this second cycle of selection, the top 10% comprises about 23.1% of subpopulation H, 7.7% of HL and 1.7% of L. The expected mean of each subpopulation in the second cycle of selection is then given by:

$$\bar{X}_{H(2)} = i_{H(2)}\, h_w^{2(1)} \sqrt{\sigma_w^{2(1)}} + \bar{X}_{H(1)}$$

$$\bar{X}_{HL(2)} = i_{HL(2)}\, h_w^{2(1)} \sqrt{\sigma_w^{2(1)}} + 0.0$$

$$\bar{X}_{L(2)} = i_{L(2)}\, h_w^{2(1)} \sqrt{\sigma_w^{2(1)}} + \bar{X}_{L(1)} ,$$

where, for example, $i_{H(2)}$ refers to the selection differential at the second cycle of selection within the H subpopulation corresponding to a proportion selected of 23.1%. The overall genetic mean of the top 10% selected is therefore,

$$\tfrac{1}{4}(0.231)\, \bar{X}_{H(2)} + \tfrac{1}{2}(0.77)\, \bar{X}_{HL(2)} + \tfrac{1}{4}(0.17)\, \bar{X}_{L(2)} = 3.52.$$

This exercise can be repeated for a second cycle of selection, but bearing in mind that the total reduction in variance within subpopulations is comprised of the reduction generated in the new selection cycle plus a proportion $(1-\bar{c})$ of the reduction incurred in the previous cycle of selection. This procedure is similar to the one used by Robertson (1970b)though he ignored changes in variance.

Table 10.9 shows four sets of results corresponding to the divergence between the means of the offspring of the H x H and L x L matings in the first three cycles of selection. The predictions based on the infinitesimal model which ignore the subdivision of the overall population into separate subpopulations tend to underestimate the observed divergence and this is due to the non-linearity of response which we discussed previously. When attempts are made to allow for the lack of distributional uniformity, the divergence

TABLE 10.9:

Divergence between extreme matings obtained from the Drosophila experiment, $D_{(O)}$; from predictions based on the infinitesimal model ignoring departures from overall normality, $D_{(I)}$; from similar predictions acknowledging the existence of subpopulations; $D_{(IM)}$; from Montecarlo simulations of model (30,0.5,0.1), $D_{(MC)}$.

Generation	$D_{(O)}$	$D_{(I)}$	$D_{(IM)}$	$D_{(MC)}$
1	3.5 ± 0.3	4.2	4.2	4.0 ± 0.2
2	6.2 ± 0.4	5.4	7.0	6.7 ± 0.2
3	9.3 ± 0.5	7.2	12.1	9.1 ± 0.2

is overestimated, particularly at generation 3. This is probably due to the lack of normality associated with the finite number of loci, which develops rather soon in disruptive selection of high intensity in both the overall distribution and within subpopulations, making the predictions based on this approach of questionable validity. The Montecarlo simulation results are in good agreement with observed results.

When we discussed the changes of variance in the context of directional selection, we illustrated the difficulties which arise in making reasonably accurate predictions of selection response due to the problem introduced by changes of gene frequencies. These changes not only had an effect on the amount of disequilibrium

generated but also could cause problems of lack of linearity of offspring parent regressions. With an infinite number of loci, however, these problems are virtually overcome, and the amount of non-linearity introduced after a first cycle of directional selection is negligible (Bulmer, personal communication). In the case of disruptive selection, particularly in experiments of short term duration, gene frequency changes are very small but the lack of linearity of offspring parent regression arises due to the type of distribution which develops as selection starts operating. In both types of selection, however, predictions of changes in variance are likely to be more accurate when selection intensities are low, but for quite different reasons in each case.

CHAPTER 11

SUMMARY AND CONCLUSIONS

Summary and Conclusions.

I. Directional Selection.

At the beginning of this thesis we asked the question, how accurately can the predictions based on the infinitesimal model theory describe the changes of genetic parameters induced by directional selection. The answer to this question is clearly dependent on the genetic model we have in mind. In this thesis we have concentrated attention on additive and dominant models. Assuming additivity between and within loci, the short term predictions of expected response to selection allowing for the changes caused by the build up of disequilibrium and ignoring those due to gene frequency changes are in good agreement with observed results provided gene frequencies are not far from intermediate values and gene effects are not large. This is hardly a surprising result since we are basically stating the conditions under which gene frequency changes are minimised and furthermore, assumptions of normality are not grossly violated. Extreme gene frequencies and/or the presence of loci of large effect will restrict the validity of the predictions not only of the amount of disequilibrium generated, but more generally of short term selection response.

In agreement with other reports in the literature, we showed that linkage, even if it is very tight, has little effect on selection response during the first four cycles of selection.

If the population size is small enough that most of the changes in the equilibrium additive variance are due to drift, we have been

able to show that reasonably accurate predictions of expected response can be obtained from estimable genetic parameters. These results may be useful in the case of laboratory experiments with Drosophila in situations where the effective population size is of the order of 10.

As to the importance of these changes and their effect on the accuracy of the prediction of expected selection response, we obtained a quantitative answer by comparing the difference between observed and predicted responses (the latter obtained from base population parameters assuming that these remain unchanged during selection) with the standard deviation of response derived from Montecarlo simulations. The results again depend on the underlying gene frequency distribution and effects and the size of the experiment. With reasonably large population size (of the order of 60 or more), provided gene frequencies are not initially at low values, changes of genetic parameters become relevant after three or four cycles of selection. With low selection intensities and low heritabilities, predictions of joint disequilibrium are very accurate but its effect on changes of genetic parameters is small and not likely to be detected even if population size is very large.

The presence of dominance introduces more serious complications, even when the number of loci is as large as 30 and gene effects relatively small ($\alpha/\sigma \simeq 0.20$). Independently of the problem of changes of genetic parameters, we are faced with the non-linearity of offspring parent regressions and the consequent asymmetry of immediate selection response. Predictions of the generation of joint disequilibrium and expected selection response based on infinitesimal

model theory are inaccurate and two way selection experiments show considerable asymmetry in the amount of disequilibrium generated.

Two different experiments with Drosophila melanogaster each one run with two replicates were carried out in an attempt to study experimentally the changes of genetic parameters during selection. The results were rather equivocal in that only one out of the four replicates showed significant evidence of any build up of disequilibrium during selection. An interesting feature of the results of this replicate, however, was the fact that no significant changes of genetic parameters were apparent during the four cycles of directional selection but upon four generations of random mating a considerable increase in the immediate response to selection was achieved. This result was reconciled in terms of a model in which the trait (abdominal bristle number in this case) was determined by several loci of small effect at intermediate frequencies and few loci of large effect at extreme frequencies. With this model, the increase in genotypic variance due to the permanent effects of gene frequencies moving towards intermediate values is partly compensated by the reduction in the genotypic variance caused by the temporary effect of joint disequilibrium, and consequently, genetic parameters remain fairly stable during the early cycles of selection. As selection is relaxed and disequilibrium breaks down, the permanent effects are unmasked and the realized heritability increases.

This result probably points to the moral of this work. We are still unable to make accurate predictions of expected short term

responses to selection from present base population parameter estimates unless we have some idea of the underlying frequency distribution and effects of the genes affecting the trait in question or unless the effective population size is very small. However we have clearly shown that the reduction in variance due to the generation of joint disequilibrium plays a major role in the selection process and it should not be omitted in short term selection studies.

II. Disruptive Selection.

Selection of extreme deviants followed by random mating of the selected individuals is known to lead to small changes of gene frequency, particularly in experiments of short duration. Most of the changes of the genotypic variance are due to the generation of linkage disequilibrium which if selection intensity is high, should lead to considerable increase of genetic parameters and therefore its effects are likely to be detected experimentally with little ambiguity. The experiment carried out with Drosophila, reported in Chapter 10 was set up in an effort to provide evidence for the generation of linkage disequilibrium on a quantitative trait.

The character measured was abdominal bristle scores. The experiment was run with two replicates and in each one three generations of disruptive selection lead to conspicuous increases of the heritability estimated by offspring mid-parent regressions, the parents being selected at both ends of the distribution. A large proportion of the increment in genetic parameters disappeared after several cycles of selection, this result being consistent with the expectation

that the increments observed were temporary and due to the generation of positive linkage disequilibrium. A second source of evidence on the build up of disequilibrium during disruptive selection was provided by contemporary estimates of heritability based on intra-class correlations between sibs which, in agreement with work reported in Chapter 7 lead to positively biased estimates.

This experiment stimulated the theoretical work on disruptive selection reported in Chapter 9, which aided in the interpretation of several aspects of the results obtained with Drosophila. In particular, we clearly showed that disruptive selection leads immediately to non-linear relationships between offspring and parents and therefore estimates of genetic parameters obtained by fitting linear regression models to the data during the course of selection must be interpreted with some qualifications.

Another interesting outcome of the theoretical work was the results of the interaction between small population size and the degree of linkage. In marked contrast with the case of deterministic models we showed that the tighter the linkage the smaller the amount of disequilibrium generated, but this effect is small during the first four or five generations of disruptive selection.

APPENDIX

In this Appendix we briefly describe the two methods based on deterministic models which have been used in this thesis to obtain what we have termed "Exact Results" (Chapters 3, 5, 6) for changes of gene and gamete frequencies and various types of disequilibria. For the purpose of the description that follows, we refer to these methods as,

(i) Method based on numerical integration of the normal density function - Method I.

(ii) Method based on selection within genotypic classes - Method II.

Some numerical examples are presented at the end of this section.

Method I.

This method has been widely used by quantitative geneticists (i.e. Griffing, 1960; Latter, 1965) and was probably first considered by Fisher (1918). The conceptual framework on which it is based assumes that the metric trait is determined by many additive (non epistatic) loci, and a normally distributed environmental component, such that the distribution of phenotypic values is normal, with mean M and variance σ^2. Attention is focussed on one or two loci, say, such that individuals of the $i^{\underline{th}}$ genotype have mean $\bar{X}_i$ and their variance in the population is σ_i^2. The variance contributed by the locus or pair of loci is $\sigma_*^2 = \sigma^2 - \sigma_i^2$. Once the genetic model is specified in terms of the number, frequencies and effects of the genes involved and the size of the environmental

variance (VE) truncation selection is practised and those individuals which exceed a certain value, T, are saved for breeding. With two loci we have 10 genotypic distributions (many of which have the same mean) with the same variance, σ_i^2. Given T, we find the truncation point in the underlying distribution for the $i^{\underline{th}}$ genotype, T_i, and the selective value of the genotype is obtained by integrating the normal curve, from T_i to infinity. Using standard population genetics theory we can then find the frequency of the various genotypes, gametes and genes after selection has operated.

It must be noticed that an important assumption in this technique is that we concentrate on a pair of loci say, and we assume that the distribution of the various genotypes is normal, this assumption arising from the fact that the environmental component follows a normal distribution and that there are many more other loci which are still segregating. The question then arises: what are the consequences of reducing the total number of loci to a small number (10, say) whose frequencies are extreme and whose effects are not small? How does this affect the results obtained in terms of the predictions of changes of gene and gamete frequencies and the amount of disequilibria generated by selection? Attempts to answer these questions lead to the development of Method II - Selection Within Genotypic Classes. This method was kindly suggested to me by Professor Alan Robertson.

Method II.

This technique was developed in an effort to understand the validity of the results obtained using Method I, when gene frequencies are extreme and gene effects are large. The basis of the method is as follows. The choice of the total number of loci and type of gene action, immediately specifies the number of genotypic classes, each genotypic class being defined in terms of its mean and variance. Consider an additive model of 4 loci, with each plus allele at each locus having an effect of +1 on the trait, and each minus allele an effect of -1. There are then 9 possible genotypic values which range from -4 to +4 and we refer to these as genotypic classes. Out of the 4 loci, we focus our attention on two of them. For example, individuals carrying genotype AB/Ab (which has genotypic value 1) at this pair of loci, may be segregating for any other of the possible 10 genotypes at the other 2 loci and therefore these individuals can assume genotypic values ranging from (1-2) to (1+2), the frequency within each class being determined by the initial gamete frequencies in the population.

The important point to notice is that the variance within each class is environmental since by definition, all the genotypes belonging to a class have the same genotypic value. Each genotypic class therefore is normally distributed, with mean $\bar{X}_i$ and variance VE.

Truncation selection is practised at the level of the phenotypic distribution. We can obtain the truncation point at the level of the various genotypic classes and calculate the selective value of each class either from normal tables or by numerical integration of

the normal density function. Knowledge of the selective value of the various genotypic classes and the frequency of each genotype within each class leads to the frequency of the various genotypes amongst selected individuals.

We shall now compare the results obtained from both methods using three models, all of which involve a total number of 4 additive loci with the same gene effects and frequencies. The parameters of the models are summarised in Table A.1. All models assume initial equilibrium.

Table A.2 shows the observed gene frequencies and the amount of linkage disequilibrium (covariance of gene frequencies within gametes) at generation 1, after a first cycle of selection.

As expected, with model 0.5/0.10, both methods give very similar results. What is rather surprising, however, is that method I, based on normality, seems to give results in good agreement with those obtained using method II even when gene frequencies are extreme and/or gene effects rather large (models 0.1/0.10 and 0.1/0.40).

Kempthorne (1977) suggested that the validity of method I should be investigated. We believe that this has been done through the analysis described in this Appendix.

TABLE A.1:

Initial gene frequencies (q), Proportionate effects at each locus (a/σ), initial genotypic variance (VG) and heritability (h^2) for the three models.

Model designation (q/h^2)	q	a/σ	VG	h^2
0.5/0.10	0.5	0.22	2.00	0.10
0.1/0.10	0.1	0.37	0.72	0.10
0.1/0.40	0.1	0.75	0.72	0.40

TABLE A.2:

Gene frequencies (p) and linkage disequilibrium (D) for the various models after one cycle of selection, obtained by methods I and II.

Model	0.5/0.10		0.1/0.10		0.1/0.40	
	$px10^2$	$Dx10^5$	$px10^2$	$Dx10^5$	$px10^2$	$Dx10^5$
Method I	57.83	-23.85	15.26	-17.02	21.45	-106.68
Method II	57.83	-23.86	15.24	-16.76	21.18	-95.19

ACKNOWLEDGEMENTS

This Thesis was done under the supervision of Dr W.G. Hill. Above all it is a pleasure to express my gratitude to him for the time spent in clarifying misconceptions and engaged in invaluable discussion.

I am also very pleased to thank:
Professor A. Robertson for discussing various aspects of this work,
Professor D.S. Falconer for providing facilities at the Institute of Animal Genetics,
Dr P.J. Avery for early suggestions,
Asko Maki-Tanila for much stimulating discussion,
Jenny Smith, Anne Douglas, Maureen Edwards and Marjorie McEwan for computational assistance,
Mr E.D. Roberts for drawing the figures,
Angela Aldridge for technical assistance, and
Jackie Bogie for her excellent and patient typing.

While I was a student in Reading, Dr D.E. Daykin spent many hours with various computing problems which have had a bearing on this work. I wish to take this opportunity to record my gratitude to him.

Last, but not least, I want to thank my parents for their systematic support over the years.

REFERENCES

AVERY, P.J. and HILL, W.G. (1977). Variability in genetic parameters among small populations. Genet.Res. 29, 193-213.

AVERY, P.J. and HILL, W.G. (1979). Variance in quantitative traits due to linked dominant genes and variance in heterozygosity in small populations. Genetics 91, 817-844.

BARKER, J.S.F. and CUMMINS, L.J. (1969). Disruptive selection for sternopleural bristle number in Drosophila melanogaster. Genetics 61, 697-712.

BODMER, W.F. and PARSONS, P.A. (1962). Linkage and recombination in evolution. Adv.Genet. 11, 1-99.

BROWN, G.H. and TURNER, H.N. (L968). Response to selection in Australian Merino sheep. II. Estimates of phenotypic and genetic parameters for some production traits on Merino ewes and an analysis of the possible effects of selection on them. Aust.J.Agric.Res. 19, 303-322.

BULMER, M.G. (1971). The effect of selection on genetic variability. Amer.Natur. 105, 201-211.

BULMER, M.G. (1974). Linkage Disequilibrium and genetic variability. Genet.Res. 23, 281-289.

BULMER, M.G. (1976a). Regressions between relatives. Genet.Res. 23, 281-289.

BULMER, M.G. (1976b). The effect of selection on genetic variability: A simulation study. Genet.Res. 28, 101-117.

BUTCHER, D.F. and FREEMAN, A.E. (1969). Estimation of heritability and repeatability of milk and milk fat production with selection effects removed. J.Dairy Sci. 52, 1259-1267.

CLAYTON, G.A. and ROBERTSON, A. (1957a). An experimental check on quantitative genetical theory. II. The long-term effects of selection. J.Genetics 55, 152-170.

CLAYTON, G.A., KNIGHT, G.R., MORRIS, J.A. and ROBERTSON, A. (1957b). An experimental check on quantitative genetical theory. III. Correlated responses. J.Genetics 55, 171-180.

COCHRAN, W.G. (1951). Improvement by means of selection. Proc. Second Berkeley Symp.Math.Stat.Prob., 449-470.

CROW, J.F. and MORTON, N.E. (1955). Measurement of gene frequency drift in small populations. Evolution 9, 202-214.

CROW, J.F. and KIMURA, M. (1970). An introduction to population genetics theory. Harper & Row, New York, Evanston and London.

DYMNICKI, E., LEDERER, J. and AVERDUNK, G. (1975). Heritability estimates for milk yield by daughter-dam regression. Züchtungs Kunde 47, 285-290.

FALCONER, D.S. (1960). An Introduction to Quantitative Genetics. Oliver & Boyd, Edinburgh.

FALCONER, D.S. (1963). Maternal effects and selection response. In: Genetics Today. Proc.XI Int.Cong.Gen., 763-773.

FELDMAN, M.W. and CROW, J.F. (1970). On quasi-linkage equilibrium and the fundamental theorem of natural selection. Theoret. Pop.Biol. 1, 371-391.

FELSENSTEIN, J. (1965). The effect of linkage on directional selection. Genetics 42, 349-363.

FIMLAND, E. (1979). The effect of selection on additive genetic parameters. Z.Tierz.Züchtbiol. 96, 120-134.

FINNEY, D.J. (1956). The consequences of selection for a variate subject to errors of measurement. Bull.Inst.Int.Statistics 24, 1-10.

FISHER, R.A. (1918). The correlation between relatives on the supposition of Mendelian inheritance. Trans.Roy.Soc.Edin. 52, 399-433.

FISHER, R.A. (1930). The genetical theory of natural selection. Oxford University Press.

FISHER, R.A. (1941). Statistical methods for research workers. Oliver & Boyd, Edinburgh.

FRANKHAM, R., JONES, L.P. and BARKER, J.S.F. (1968). The effects of population size and selection intensity in selection for a quantitative character in Drosophila. III. Analyses of the lines. Genet.Res. 12, 267-283.

GIBSON, J.B. and THODAY, J.M. (1963). Effects of disruptive selection. VIII. Imposed quasi-random mating. Heredity 18, 513-524.

GRIFFING, B. (1960). Theoretical consequences of truncation selection based on individual phenotype. Aust.J.biol.Sci. 13, 307-343.

HALDANE, J.B.S. (1919). The combination of linkage values and the calculation of distance between loci of linked factors. J. Genet. 8, 299-309.

HALDANE, J.B.S. (1931). A mathematical theory of natural and artificial selection. VII. Selection intensity as a function of mortality rate. Proc.Camb.Phil.Soc. 27, 131-136.

HEDRICK, P., JAIN, S. and HOLDEN, L. (1978). Multilocus systems in evolution. In: Evolutionary Biology, Vol. 11. eds. M.K. Hecht, W.C. Steere and B. Wallace. Plenum Publ. Corp.

HILL, W.G. (1972a). Estimation of realised heritabilities from selection experiments. I. Divergent selection. Biometrics 28, 747-765.

HILL, W.G. (1972b). Estimation of realised heritabilities from selection experiments. II. Selection in one direction. Biometrics 28, 767-780.

HILL, W.G. (1974). Heritabilities: Estimation problems and the present state of information. 1st Wld Cong. on Genetics Applied to Livestock Prod., Madrid, Spain.

HILL, W.G. (1977). Variation in response to selection. In: Proc. Int.Conf.Quant.Gen., Ames, Iowa.

HILL, W.G. (1980). Design of quantitative genetic selection experiments. In: Symposium of selection experiments with laboratory animals. ed. A. Robertson.

HILL, W.G. and ROBERTSON, A. (1966). The effect of linkage on limits to artificial selection. Genet.Res. 8, 269-294.

JAMES, J.W. (1971). The founder effect and response to artificial selection. Genet.Res. 16, 241-250.

KARLIN, S. (1975). General two-locus selection models: Some objectives, results and interpretations. Theor.Pop.Biol. 7, 364-398.

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

KEMPTHORNE, O. (1957). An Introduction to Genetic Statistics. Wiley & Sons, New York.

KEMPTHORNE, O. (1977). Status of Quantitative Genetics Theory. Proc.Int.Conf.Quant.Gen., Ames, Iowa. 719-760.

KIMURA, M. (1956). A model of a genetic system which leads to closer linkage by natural selection. Evolution 10, 278-287.

KIMURA, M. (1958). On the change of population fitness by natural selection. Heredity 12, 145-167.

KIMURA, M. (1965). Attainment of quasi-linkage equilibrium when gene frequencies are changing by natural selection. Genetics 52, 875-890.

KIMURA, M. and OHTA, T. (1971). Theoretical Aspects of Population Genetics. Princeton Univ.Press, Princeton, N.J.

KIMURA, M. and CROW, J.F. (1978). Effect of overall phenotypic selection on genetic change at individual loci. Proc.Natl. Acad.Sci. 75, 6168-6171.

KOJIMA, K.I. (1961). Effects of dominance and size of population on response to mass selection. Genet.Res. 2, 177-188.

KOJIMA, K.I. and LEWONTIN, R.C. (1970). Evolutionary significance of linkage and epistasis. In: Mathematical Topics in Population Genetics. ed. K.Kojima, Springer-Verlag, Berlin-Heidelberg-New York.

LATTER, B.D.H. (1960). Natural selection for an intermediate optimum. Aust.J.Biol.Sci. 13, 30-35.

LATTER, B.D.H. (1965). The response to artificial selection due to autosomal genes of large effect. I. Changes in gene frequency at an additive locus. Aust.J.biol.Sci. 18, 585-598.

LEE, A.J. and HENDERSON, C.R. (1969). Components of genetic variance in milk yield. J.Dairy Sci. 52, 780-789.

LEWONTIN, R.C. (1964). The interaction of selection and linkage. I. General considerations; heterotic models. Genetics 49, 49-67.

LEWONTIN, R.C. (1974). The Genetic Basis of Evolutionary Change. Columbia Univ.Press, New York.

LEWONTIN, R.C. and KOJIMA, K. (1960). The evolutionary dynamics of complex polymorphisms. Evolution 14, 450-472.

LUSH, J.L. (1948). The Genetics of Populations. Mimmeograph.

MÄKI-TANILA, A. (1980). The validity of the heritability concept in quantitative genetics. Ph.D. Thesis, Univ. of Edinburgh (in preparation).

MARTIN, F.G. and COCKERHAM, C.C. (1960). High speed selection studies. In: Biometrical Genetics. ed. O. Kempthorne. Pergamon Press, 35-45.

MATHER, K. (1941). Variation and selection of polygenic characters. J.Genet. 41, 159-193.

MATHER, K. (1943). Polygenic inheritance and natural selection. Biol.Rev. 18, 32-64.

MATHER, K. (1955). Polymorphism as an outcome of disruptive selection. Evolution 9, 52-61.

MAYNARD SMITH, J. (1978). The Evolution of Sex. Cambridge Univ. Press.

MAYNARD SMITH, J. (1979). The effects of normalizing and disruptive selection on genes for recombination. Genet.Res. 33, 121-128.

McDONALD, J.F. and AYALA, F.J. (1974). Genetic response to environmental heterogeneity. Nature 250, 572-574.

MILLICENT, E. and THODAY, J.M. (1961). Effects of disruptive selection. IV. Gene flow and divergence. Heredity 16, 199-217.

MORAN, P.A.P. (1964). On the non existance of adaptive topographies. Ann.Human Genetics 27, 383-393.

MORLEY, F.H.W. (1955). Selection for economic characters in Australian Merino sheep. V. Further estimates of phenotypic and genetic parameters. Aust.J.Agric.Res. 6, 77-90.

NAGYLAKI, T. (1974). Quasi-linkage equilibrium and the evolution of two-locus systems. Proc.Natl.Acad.Sci. US 71, 526-530.

NEELEY, D.L.R. and RAWLINGS, J.O. (1971). Disequilibria and genotypic variance in a recurrent selection system for an additive genetic model. Inst. of Statistics. Mimmeograph series No. 729, Raleigh, North Carolina.

NEI, M. (1963). Effects of selection on the components of genetic variance. In: Statistical Genetics and Plant Breeding. ed. W.D. Hanson and H.F. Robinson. Natl.Acad.Sci., Natl.Res. Council Publ.No. 982, Washington, 501-515.

PEARSON, K. (1903). On the influence of natural selection on the variability and correlation of organs. Phil.Trans.Roy.Soc. Lond. A200, 1-66.

POLLAK, E. (1979). Some models of genetic selection. Biometrics 35, 119-138.

PONZONI, R.W. and JAMES, J.W. (1978). Possible biases in heritability estimates from intraclass correlation. Theor.Appl. Gen. 53, 25-27.

POWELL, J.R. (1971). Genetic polymorphisms in a varied environment. Science 174, 1035-1036.

PURSER, A.F. (1980). Comparison of expected and realised responses in three sheep selection experiments. In: Symposium of selection experiments with laboratory animals. ed. A. Robertson.

RAHNEFELD, G.W., BOYLAN, W.J., COMSTOCK, R.E. and SINGH, M. (1963). Mass selection for post-weaning growth in mice. Genetics 48, 1567-1583.

REEVE, E.C.R. (1953). Studies in quantitative inheritance. III. Heritability and genetic correlation in progeny tests using different mating systems. J.Genet. 51, 520-542.

ROBERTSON, A. (1956). The effect of selection against extreme deviants based on deviation or on homozygosis. J.Genet. 54, 236-248.

ROBERTSON, A. (1960). A theory of limits in artificial selection. Proc.Roy.Soc.B 153, 234-249.

ROBERTSON, A. (1961). Inbreeding in artificial selection programmes. Genet.Res. 2, 189-194.

ROBERTSON, A. (1970a). A theory of limits in artificial selection with many linked loci. In: Mathematical Topics in Population Genetics (ed. Kojima).

ROBERTSON, A. (1970b). A note on disruptive selection experiments in Drosophila. Am.Nat. 104, 561-569.

ROBERTSON, A. (1970c). The state of quantitative genetics in relation to the real world. Proc. 19th Nat.Breed.Roundtable, Kansas City.

ROBERTSON, A. (1977a). The effect of selection on the estimation of genetic parameters. Z. Tierz. Züchtbiol. 94, 131-135.

ROBERTSON, A. (1977b). Artificial selection with a large number of linked loci. Proc.Int.Conf.Quant.Gen., Ames, Iowa.

ROBERTSON, A. (1977c). The non-linearity of offspring parent regression. Proc.Int.Conf.Quant.Gen., Ames, Iowa, 297-306.

ROUX, C.Z. (1974). Hardy-Weinberg equilibria in random mating populations. Theor.Pop.Biol. 5, 393-416.

SCHARLOO, W. (1964). The effect of disruptive and stabilizing selection on the expression of a cubitus interruptus mutant in Drosophila. Genetics 50, 553-562.

SCHARLOO, W., HOOGMOED, M.S. and TER KUILE, A. (1967). Stabilizing and disruptive selection on a mutant character in Drosophila. I. The phenotypic variance and its components. Genetics 56, 709-726.

SEN, B.K. and ROBERTSON, A. (1964). An experimental examination of methods for the simultaneous selection of two characters,using Drosophila melanogaster. Genetics 50, 199-209.

SIMPSON, G.G. (1944). Tempo and mode in evolution. Columbia Univ.Press, New York.

SLATKIN, M. (1972). On treating the chromosome as the unit of selection. Genetics 72, 157-168.

SYRSTAD, O. (1966). Studies on dairy herd records. IV. Estimates of phenotypic and genetic parameters. Acta Agric.Scand. 16, 79-96.

TALLIS, G.M. (1961). The moment generating function of the truncated multi-normal distribution. J.Roy.Stat.Soc. B 23, 223-229.

THODAY, J.M. (1959). Effects of disruptive selection. I. Genetic flexibility. Heredity 13, 187-203.

THODAY, J.M. (1972). Review Lecture: Disruptive selection. Proc. R.Soc.Lond. B 182, 109-143.

THODAY, J.M. and BOAM, T.B. (1959). Effects of disruptive selection. II. Polymorphism and divergence without isolation. Heredity 13, 205-218.

THOMPSON, R. (1976). Design of experiments to estimate heritability when observations are available on parents and offspring. Biometrics 32, 283-304.

VAN VLECK, L.D. and BRADFORD, G.E. (1965). Comparison of h^2 estimates from daughter-dam regressions and paternal half-sib correlation. J.Dairy Sci. 48, 1372-1375.

WEIR, B.S. (1979). Inferences about linkage disequilibrium. Biometrics 35, 235-254.

WRIGHT, S. (1952). The Genetics of Quantiative Variability. In: Quantitative Inheritance. ed. E.C.R. Reeve and C.H. Waddington London, HMSO.

WRIGHT, S. (1977). Evolution and the genetics of populations. Vol. 3, Experimental results and evolutionary deductions. Univ. Chicago Press.

YOO, B.H. (1980). Long term selection for a quantitative character in large replicate populations of Drosophila. 1. Response to selection. Genet.Res. 35, 1-17.

YOUNG, S.S.Y. (1966). Computer simulation of directional selection in large populations. I. The programme, the additive and dominance models. Genetics 53, 189-205.

YOUNG, S.S.Y. (1967). Computer simulation of directional selection in large populations. II. The additive x additive and mixed models. Genetics 56, 73-87.