THE EFFECT OF INITIAL LINKAGE DISEQUILIBRIUM AND NATURAL SELECTION IN POOLED POPULATIONS

EXPERIMENTS WITH *TRIBOLIUM* AND SIMULATION

P. STAM

575.115/.116.087.1:595.767.29

P. STAM

THE EFFECT OF INITIAL LINKAGE DISEQUILIBRIUM AND NATURAL SELECTION IN POOLED POPULATIONS

EXPERIMENTS WITH *TRIBOLIUM* AND SIMULATION

PROEFSCHRIFT

TER VERKRIJGING VAN DE GRAAD
VAN DOCTOR IN DE LANDBOUWWETENSCHAPPEN,
OP GEZAG VAN DE RECTOR MAGNIFICUS, PROF. DR. IR.
H. A. LENIGER, HOOGLERAAR IN DE TECHNOLOGIE,
IN HET OPENBAAR TE VERDEDIGEN
OP WOENSDAG 15 NOVEMBER 1972 DES NAMIDDAGS TE VIER UUR
IN DE AULA
VAN DE LANDBOUWHOGESCHOOL TE WAGENINGEN

H. VEENMAN & ZONEN N.V. – WAGENINGEN – 1972

STELLINGEN

I

Naast simulatie bieden ook experimenten de mogelijkheid om de kwantitatieve aspecten van associatieve overdominantie te bestuderen.

II

Wanneer het materiaal dat de veredelaar introduceert in een kruisbevruchtende populatie een relatief hoge inteeltcoëfficiënt heeft, bestaat de mogelijkheid dat de hiermee beoogde vergroting van de genetische variabiliteit dier populatie in enkele generaties wordt gereduceerd tot het oorspronkelijke niveau.

III

Aangezien de berekening van het heterozygotieverlies door inteelt of kleine populatieomvang voor de veredelaar het karakter heeft van een risicoberekening, ware bij deze berekening behalve de te verwachten fractie ook de variatie in de fractie homozygote loci te betrekken.

IV

Het dient te worden betwijfeld of modellen van ecologische systemen, welke voldoende isomorf zijn met de werkelijkheid om een voorspellende waarde te bezitten, hanteerbaar zullen zijn in hun gebruik als voorspeller.

V

Indien het niet constant zijn van selektiecoëfficiënten wordt afgeleid uit populatiestellingen in een aantal opeenvolgende generaties, waarbij slechts één telling per generatie is uitgevoerd, mag dit niet worden gebruikt ter ondersteuning van de hypothese dat de selektiecoëfficiënten frekwentie-afhankelijk zijn.

VI

In een Poisson proces met intensiteit λ, waarin een willekeurig gekozen interval met lengte L wordt beschouwd, heeft de wachttijd, x, gerekend vanaf een willekeurig punt in dit interval tot de eerstvolgende aankomst, waarbij het rechter uiteinde van het interval in ieder geval als aankomst wordt aangemerkt, de volgende kansdichtheid:

$$f(x) = \frac{1}{L}\left\{ \lambda(L-x) + 1 \right\} e^{\lambda x}, 0 \leq x \leq L \ .$$

Deze kansvariabele doet zich voor wanneer men de afstand beschouwt waarover de homologe chromosomen autozygoot zijn, gegeven dat de allelen van een bepaald locus autozygoot zijn.

VII

Het door Lewontin en White geconstateerde feit dat de door hen onderzochte *Moraba* populaties een zodanige frekwentieverdeling van karyotypen hebben dat een kleine verschuiving hierin nauwelijks invloed zou hebben op de gemiddelde populatiefitness, is een artefact voortkomend uit de door hen ten onrechte toegepaste methode voor het schatten van fitnesses.

R. C. LEWONTIN en M. J. D. WHITE (1960). Evolution 14. 116–129.

VIII

De door Gowe, Robertson en Latter gegeven uitdrukking voor effectieve populatieomvang (N_e),

$$\frac{1}{N_e} = \frac{3}{16\,N_m} + \frac{1}{16\,N_f},$$

waarin N_m en N_f de aantallen mannelijke, respectievelijk vrouwelijke ouders zijn, heeft betrekking op de variantie in genfrekwentieverschuiving in vergelijking met een eenhuizige populatie van $N_m + N_f$ individuen uit welks gametenverzameling de nakomelingen worden geloot. Zij mag daarom niet worden gebruikt voor de berekening van de kans op autozygotie in de nakomelingschap.

A. S. GOWE, A. ROBERTSON en B. D. H. LATTER (1959). Poultry Science 38, p 464.

IX

Indien de gangbare methode van „vermenigvuldigen met de hand", zoals die op de basisschool wordt onderwezen, werd vervangen door de methode waarbij de bewerkingen vermenigvuldigen en optellen geheel gescheiden in plaats van afwisselend worden uitgevoerd, zou dit in de toekomst wellicht veler rekenvaardigheid verhogen.

Proefschrift van P. STAM
Wageningen, 15 november 1972

VOORWOORD

Gaarne wil ik hier mijn dank tot uitdrukking brengen jegens degenen die hebben bijgedragen tot het tot stand komen van dit proefschrift.

In de eerste plaats dank ik mijn ouders die mij in de gelegenheid stelden te gaan studeren.

Hooggeleerde Van der Veen, de grote mate van vrijheid die U mij heeft gelaten tijdens het onderzoek en Uw vele inspirerende suggesties heb ik op hoge prijs gesteld. Ik hoop ook in de toekomst mijn voordeel te kunnen doen met Uw adviezen.

Hooggeleerde De Wit, de gesprekken die ik met U mocht hebben over het simuleren van mijn experimenten waren vruchtbaar. Zij waren helaas weinig talrijk.

Een woord van dank gaat uit naar A. P. M. den Nijs die als student aan het onderzoek meewerkte, en naar ir. C. de Jonge die mij ten alle tijde behulpzaam was bij het verzamelen van simulatieresultaten.

De personeelsleden van de afdeling erfelijkheidsleer dank ik voor hun vriendelijke medewerking, in het bijzonder de Heer K. Knoop, die de afbeeldingen verzorgde en Mej. Henriet Boelema, die het manuscript typte.

Jou, Trieneke, ben ik dankbaar voor de steun die je me tijdens het onderzoek wist te geven.

CONTENTS

1. INTRODUCTION AND OUTLINE

The large number of polymorphic loci found in natural populations, e.g. *Drosophila*, mice and man, have frequently been discussed by population geneticists. Various models have been proposed to explain the existence of so many genetic polymorphisms in nature (a current estimate of the proportion of polymorphic loci in *Drosophila* is 0.3; KIMURA, 1971). Initially these models were based on some type of equilibrating selection mechanism acting upon each polymorphic locus. Recently, KIMURA suggested that many polymorphic loci might be selectively neutral; he started work on a theory which should liberate us from 'pan selectionism' (KIMURA, 1968, 1971). These neutral loci then may be under apparent equilibrating selection pressure as a result of their association with closely linked loci which are under real selection pressure (OHTA and KIMURA, 1970). This so called 'associative overdominance' then is a result of overall linkage disequilibrium between the neutral loci and the non-neutral loci. In a natural population this deviation from linkage equilibrium is due to the finite size of the population.

An implication of this model is the following. When a laboratory population is initiated with small samples from a relatively large population (or different populations) the 'amount' of overall linkage disequilibrium is greatly amplified, resulting in amplified apparent selection coefficients at the selectively neutral loci in the new population. This point is discussed in chapter 2 (section 2.5) (see also KIMURA, 1971).

With neutral alleles which show associative overdominance, the general features of the models for equilibrating selection remain the same as with the previously used selection models, be it that selection now is restricted to the interspersed non-neutral loci.

Chapter 2 discusses a number of models which have been proposed for the maintenance of genetic polymorphisms; the general conditions are discussed which should be satisfied by the selection models (i.e. fairly high selection coefficients and a not too heavy segregational load), as well as the arguments which have been used for their plausibility from the evolutionary point of view (i.e. the principle of minimizing the segregational load). This rather extensive discussion has been made in order to arrive at a realistic general operational model which covers the experimental results reported in literature and which also describes and explains my own experimental findings.

The rationale of my experiments (chapter 3) is as follows. *Tribolium castaneum* HERBST was chosen since it is a genetically well known and experimentally convenient object. The marker locus 'black' has been chosen for investigation of the selective forces acting upon it, since all three genotypes can phenotypically be distinguished. Now, if the black stock is mixed with the wild type

stock (together with their F_1) the observed shifts in gene frequency of the *b* allele during generations can a priori be ascribed to:

1. Frequency dependent selection at the marker locus (i.e. selection dependent on the frequencies of the marker genotypes),
2. Non-frequency dependent selection at the marker locus itself,
3. Indirect selection at the marker locus via linked fitness loci and/or fitness loci on other chromosomes (both groups of loci contribute to initial linkage disequilibrium and therefore to differential fitness of the marker genotypes).

On the basis of the experiments to be described in chapter 3 these three situations can be distinguished. Of course, when introducing a mutant stock (mutant genotype black) into a wild type population, it must be expected that the mutant genotype is at a selective disadvantage, since in general laboratory practice the construction of such a mutant stock involves a certain amount of inbreeding, that is, a certain amount of homozygosity in its genetic background. The experiments will now prove this to be the case and rule out the other two possibilities (1 and 2 above). The basic feature of these experiments is that mutant and wild type populations and their F_1 are mixed in a number of different genotype frequencies (representing Weinberg-Hardy frequencies) and that a classical F_2 population is used to evaluate the relative importance of fitness loci linked resp. non-linked to the marker locus (the F_2 population will be in approximate linkage equilibrium for non-linked loci). The initial linkage disequilibrium (and therefore the apparent selection at the neutral marker locus) in these experiments is not so much due to the small samples from which the pooled populations are started (cf. KIMURA, 1971) as to the different genotypic backgrounds of the founder stocks.

The foundation of a population from two or more populations with different coadapted gene pools can be described as the intake of 'foreign' genetic material by a receiving population. This process does not only occur when populations of different origin are used to initiate a laboratory stock, but also plays a role in the events following introgressive hybridization in nature, and in the introduction of 'fresh' genetic material into a breeding stock. Analysis of this process requires study of the role of selection and recombination. As will be demonstrated in chapter 3, information on the effects of initial linkage disequilibrium and selection in pooled populations can be inferred from the changes in gene frequency of a selectively neutral marker locus.

In chapter 4 a simulation model is presented (based on FRASERS's technique of binary representation of genotypes) by which the fate of newly introduced genetic material in a population can be studied theoretically. Simulation then may be helpful in a better understanding of a process which is interesting not only from the evolutionary point of view but also with regard to the introduction of 'fresh' genetic material into animal and plant breeding stocks.

2. THE MAINTENANCE OF GENETIC POLYMORPHISMS

Many polymorphisms have been observed in natural populations. Especially the work of HUBBY and LEWONTIN (1966), which revealed the existence of a large amount of genetic variability with respect to protein structure in *Drosophila* populations, has stimulated the discussion on the mechanisms through which polymorphisms can be maintained in natural populations. In this discussion the central question is how the forces acting upon a population, i.e. random drift, selection, mutation and migration balance each other such that the polymorphisms are maintained.

This chapter discusses overdominance, frequency dependent selection and neutral alleles as mechanisms for the maintenance of polymorphisms. Overdominance at many loci needs some further exploration since a number of models for the joint effect of many overdominant loci (i.e. multiplicative effects and additive effects) on fitness implies too high a load and/or too small selection coefficients to counteract drift. I have constructed a third model (diminishing returns) which also leads to too low selection coefficients. Finally KING's (1966) threshold model proves, upon further examination, to be the most satisfactory in all respects. In section 2.5 a synthesis will be made by using the concept of associative overdominance. In experiments the consequences of this mechanism may be easily mistaken to be the outcome of frequency dependent selection. An experiment will be proposed which can lead to unambiguous interpretation. The results of such experiments will be described in chapter 3.

2.1. DEFINITIONS

Fitness

The term fitness is used as a synonym for reproductive capacity. This may apply to either populations, genotypes or individuals. In a population with overlapping generations the growth is continuous and may be expressed as

$$dy/dt = m \cdot y, \tag{2.1}$$

where y is the population size and m is the net growth rate (net effect of birth and death rate). The parameter m, the Malthusian parameter, measures the reproductive capacity or fitness of the population. When m is a constant, Eq. 2.1 yields

$$y_t = y_o e^{mt}, \tag{2.2}$$

where y_o is the value of y at time $t = 0$.

When generations are non-overlapping, i.e. when after reproduction the parental population is replaced by the offspring generation, the population growth is discontinuous and is expressed as

$$y_{t+1} = w \cdot y_t \; (t = 0, 1, 2,..), \tag{2.3}$$

where w is the Wrightian fitness of the population, i.e. the mean number of offspring per individual. When w is constant,

$$y_t = y_o \cdot w^t \; (t = 0, 1, 2, \ldots) \tag{2.4}$$

The relation between w and m is seen by equating expressions (2.2) and (2.4) (for integer values of t):

$$w = e^m,$$

or

$$m = {}^e\log w.$$

Throughout this study the term fitness is used in the Wrightian sense.

When applied to genotypes or individuals, fitness again is defined as the mean number, or rather the expected number of offspring. Consequently, as a measure of the fitness of a genotype its mean number of offspring is used. Total fitness includes the components viability and fertility. Viability is defined as the probability of survival from zygote to the reproductive stage. Fertility is defined as the expected number of offspring once the adult stage has been reached, i.e. the conditional expectation of the number of offspring of a zygote, the condition being that the reproductive stage is reached. Denoting viability by v and fertility by f, the total or net fitness (w) becomes

$$w = v \cdot f.$$

Selection

Selection occurs when some individuals leave more offspring than others. Since the production of offspring is subject to both variation in genetic factors and non-genetic random effects, selection in the genetical sense is defined as 'non random differential reproduction of genotypes' (LERNER, 1958). This definition shows that genotypic selection refers to differences in reproductive capacity as far as these differences correspond to differences in genotypic constitution. Selection is said to favour those genotypes which, on an average, leave more offspring than others.

Relative fitnesses and selection coefficients

When three genotypes at a locus, AA, Aa and aa, say, have fitnesses w_2, w_1 and w_0 respectively, then the ratios w_2/w_1 and w_0/w_1 are called the relative fitnesses of the genotypes AA and aa respectively. (Of course, any genotype may be used as a reference for relative fitnesses.) The selection coefficients s_2 and s_0 of the genotypes AA and aa are defined by

$$1 - s_2 = w_2/w_1 \text{ and } 1 - s_0 = w_0/w_1$$

Thus, the selection coefficient s_2 measures the average selective advantage (or disadvantage, when $s_2 < 0$) of the genotype Aa over AA. The relative fitnesses of the genotypes AA, Aa and aa in terms of s_2 and s_0 are $1-s_2$, 1 and $1-s_0$,

respectively. Changes in genotype and gene frequencies are more easily described in terms of selection coefficients than in terms of absolute fitnesses.

Genetic load

The genetic load of a population is defined as 'the amount by which the mean fitness of the population is depressed for genetic reasons below that of the genotype with maximum fitness' (WALLACE, 1968). It is mostly expressed as the ratio of this difference and the maximum fitness:

$$L = (w_{max} - \bar{w}) / w_{max} = 1 - \bar{w}/w_{max}. \quad (2.5)$$

The load of a population which is due to the occurrence of deleterious mutants is known as the mutational load, which will not be discussed in this chapter. The segregational load of a population is the load which is due to segregation at one or more loci. Segregational load has as a reference the maximum fitness which may be realized in the population. The load of a monomorphic population depends on whether monomorphism is regarded as a limiting case of polymorphism or not. When regarding a monomorphic population as a limiting case of a polymorphic population, it must be regarded as still segregating and therefore it has a segregational load, with the same reference fitness as a polymorphic population. When however it is not regarded as a limiting case of a polymorphic one, a monomorphic population does not segregate and therefore its segregational load, i.e. the load due to segregation does not exist. Therefore, it should be stressed that a population for which the segregational load is not defined, can not be considered to have a load equal to zero.

The two alternative ways of considering monomorphic populations and the corresponding loads are shown in Table 2.I. The difficulty with segregational load is that it has, contrary to the proportion of non-survivors, no immediate biological significance. In Table 2.I it is seen that the ratio's of proportions of non-survivors of the three populations (i.e. 0.625:0.700:0.529) give a more realistic picture of the actual differences between the populations than the

TABLE 2.I. Mean viabilities and genetic load of monomorphic and polymorphic populations. The fertility component of fitness is not considered. Viabilities of the genotypes *AA*, *Aa* and *aa* are 0.375, 0.6 and 0.3, respectively. $\bar{v}$: mean viability; $1-\bar{v}$: proportion of non-survivors; L_s: segregational load; a: the two monomorphic populations are considered here as limiting cases of polymorphic populations; b: the two monomorphic populations are considered here as non-segregating populations.

	monomorphic *AA*	monomorphic *aa*	polymorphic 16/49 *AA*, 24/49 *Aa*, 9/49 *aa*
$\bar{v}$	0.375	0.300	0.471
$1-\bar{v}$	0.625	0.700	0.529
L_s a	0.375	0.500	0.214
L_s b	–	–	0.214

ratio's of their genetic loads (i.e. 0.375:0.500:0.214). On the other hand however, the proportion of non-survivors is not a genetic measure, whilst the load (cf. the definition) expresses a reduction in fitness which indeed is due to genetic factors. For this reason, the genetic load is the most convenient measure in comparing Mendelian populations. Note that the load can be interpreted as the proportion of non-survivors, provided that the genotype with maximum viability survives with probability unity (see Eq. 2.5). (Introducing also fertility differences into the above considerations does not alter their validity since differences in total fitness can be considered as differences in 'genetic mortality'.)

Hard selection and soft selection

As pointed out by WALLACE (1968), natural selection, as far as realised by differential viability, may have two (not mutually exclusive) modes: hard and/or soft selection. (Hard selection should not be confused with intense selection.) These two types of selective forces are illustrated in Fig. 2.1, which shows the distribution of genotypes of two populations, *A* and *B*. In order to show the difference between the two modes of selection, it is assumed that there are no non-genetic random effects. With hard selection all individuals survive which have a 'genotypic value' larger than a fixed truncation value, *t*. The proportion of the population of zygotes that survives varies with the distribution of genotypes in the population. With soft selection the population size is reduced to a given number, irrespective of the distribution of genotypes.

Soft selection is likely to occur in those populations which are kept at a level which the environment can sustain, e.g. when there is predation or crowding. In a computer model constructed by WILLS, CRENSHAW and VITALE

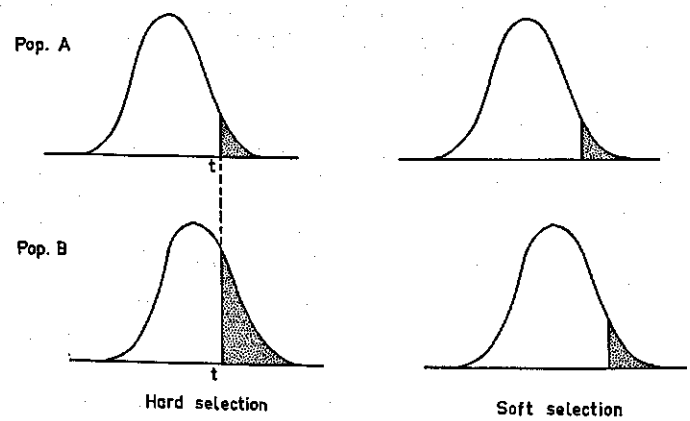


FIG. 2.1. Hard selection and soft selection (after WALLACE, 1968). The populations *A* and *B* have different distributions of genotypes. With hard selection all genotypes survive which are to the right of the fixed truncation value (*t*). With soft selection a fixed proportion of the population of zygotes survives, irrespective of the genotype distribution. Dotted area: survivors.

(1969), selection was simulated on the basis of soft selection only. This seems to be unrealistic because, if soft selection were the only operative selection force, it implies that the fitness of a particular inferior genotype will be larger when its frequency in the population is higher. This might be true for the situation where selection results from predation only; when however the inferiority of the particular genotype is due to e.g. its susceptability to a disease, its survival probability must be understood as hard selection mainly. Therefore, the assumption that soft selection is the main component of natural selection must be an oversimplification.

2.2. Overdominance

When the fitnesses w_2, w_1 and w_0 of the genotypes at a locus with two alleles (*AA*, *Aa* and *aa*) satisfy the condition

$$w_2 < w_1 > w_0$$

(i.e. the locus is overdominant), then, in the ideal situation (no random drift, no mutation and no migration), a stable equilibrium will be reached at which the frequency of the allele *A* is given by

$$p_e = s_o/(s_o + s_2),$$

where s_2 and s_0 are the selection coefficients of the genotypes *AA* and *aa*, respectively. Natural populations however are not infinitely large and, therefore, the selection pressure at an overdominant locus should be large enough to withstand the force of random drift. (Mutation and migration are not considered for the moment.) This means that selection coefficients must have a certain minimum value in order to ensure the maintenance of the polymorphism over many generations. Strictly speaking, without mutation and immigration a polymorphism can not be permanent in a finite population, even with equilibriating selection (overdominance). However, for intermediate equilibrium gene frequencies (i.e. in the range 0.2—0.8) overdominance will increase the mean time until fixation (see Robertson, 1962; Kimura, 1964). As Robertson (l.c.) has shown, an effective population size of the order of 1000 is required to maintain a polymorphism when the selective advantage of a heterozygote over either homozygotes equals 0.005, provided a mutation rate of 10^{-5} from one allele to the other (and vice versa). Because 10^{-5} probably is too high an estimate of the mutation rate and many natural populations are effectively smaller than 1000, selection coefficients probably should be of the order of 0.01 to maintain non-transient polymorphisms.

When many polymorphisms occur in a population, all of which are maintained by overdominance and recurrent mutation, the joint effect of many overdominant loci on fitness becomes the central question. In the following, four theoretical relationships between number of heterozygous loci and fitness, as well as their consequences with respect to the mean population fitness and

the average selection pressure at individual loci are discussed. The calculations for models *A*, *B* and *C* are based on the following assumptions:

1. Differential viability is the only component of selection. This assumption, i.e. the absence of genotypic fertility differences is no essential limitation.
2. All loci have two alleles.
3. All loci are considered overdominant with equal selective advantage of the heterozygote over either homozygote (which implies gene frequencies 0.5).
4. All loci are not linked and not only the probability for any individual to be homozygous at any locus equals 0.5, but also the probability for an individual to be homozygous at any two loci equals 0.25, etc.

Although the latter assumption is a gross oversimplification, the calculations below clearly show the impacts of the different models for multiple gene action. The total number of polymorphic overdominant loci will be denoted by N.

A. Multiplicative effects (see e.g. LEWONTIN and HUBBY, 1966; SVED et al., 1967).

With multiplicative effects the relation between viability (v) and number of homozygous loci (n) becomes

$$v_n = v_o (1 - s)^n, \tag{2.6}$$

where s is the selection coefficient of single gene homozygotes and v_o is the viability of the complete multiple heterozygote ($n = 0$). The mean viability of the population ($\bar{v}$) is given by

$$\bar{v} = v_0 \sum_{k=0}^{N} \binom{N}{k} \left(\frac{1}{2}\right)^N (1 - s)^k,$$

which reduces via

$$\bar{v} = v_0 \left(\frac{1}{2}\right)^N \left\{1 + (1 - s)\right\}^N$$

to

$$\bar{v} = v_0 (1 - s/2)^N. \tag{2.7}$$

The ratio $\bar{v}/v_o$ thus equals $(1-s/2)^N$ and the load, as defined by Eq. 2.5 becomes

$$L = 1 - \bar{v}/v_0 = 1 - (1 - s/2)^N \tag{2.8}$$

In Table 2.IIa I calculated the load of the hypothetical population with the multiplicative model for gene effects with various values of s and N. Table 2.IIa shows that even small selection coefficients impose a heavy load on the population when the number of overdominant loci is of the order of 1000. Many plant and invertebrate animal species produce zygotes in abundance and may be able to carry a heavy load; for most vertebratae however a load of the order of 0.9 must be considered as being too heavy to be realistic.

TABLE 2.IIa. The load of the hypothetical population with multiplicative effects for different values of s and N; N: total number of overdominant loci, s: selection coefficient of single gene homozygotes.

N	$s = 0.01$	$s = 0.005$	$s = 0.002$
1000	0.994	0.919	0.629
701	0.970	0.828	0.500
500	0.919	0.715	0.391
276	0.750	0.500	0.239
200	0.634	0.395	0.180
130	0.500	0.278	0.121
100	0.395	0.220	0.094

SVED, REED and BODMER (1967) proposed a modification of the model with multiplicative effects: they assumed multiplicative effects for a large range of genotypes with asymptotic approach of fitness to a plateau. Their idea was as follows: When N equals 1000, say, the variance of the number of homozygous loci per individual equals 250, and individuals with more than 600 or less than 400 heterozygous loci, say, are so rare that they contribute but very little to the mean fitness of the population. Mean population fitness is hardly changed when the genotypes with more than 600 heterozygous loci are assigned equal fitnesses, i.e. the upper limit of fitness. With this model the reference fitness (w_{max}) is reduced considerably and therefore also the load of the population is lowered to an acceptable level. Since on physiological grounds there can not be a truncation value of number of heterozygous loci beyond which all fitnesses are equal, SVED et al. proposed an asymptotical approach to the limiting value.

B. Additive effects

The concept of additive effects has been forwarded by a.o. MILKMAN (1967) who suggests that these effects result in a smaller load than multiplicative effects. It seems worthwhile to formally work out the consequences of this model.

With additive gene effects there is a linear relationship between viability (v) and number of homozygous loci (n):

$$v_n = v_0 - c \cdot n \; (v_n \geq o), \tag{2.9}$$

where c is a constant. The value of c depends on v_o (viability of the complete multiple heterozygote) and v_N (viability of the complete multiple homozygote) such that

$$v_N = v_0 - c.N, \text{ or } c = (v_0 - \bar{v}_N) / N.$$

Then Eq. 2.9 becomes

$$v_n = v_o - n \, (v_o - v_N)/N \tag{2.10}$$

Now it will be shown that with this model, the load of the hypothetical population is reduced considerably but that at the same time the selection coefficients at

individual loci become very small. Since the relationship between viability and number of homozygous loci is linear in n (cf. Eq. 2.10), the mean population viability ($\bar{v}$) can be obtained by substituting $E(n) = N/2$ into Eq. 2.10; this yields

$$\bar{v} = v_{\frac{1}{2}N} = (v_o + v_N)/2 \tag{2.11}$$

The load then, by Eq. 2.5 becomes

$$L = \frac{1}{2}\left(1 - \frac{v_N}{v_0}\right) \leqslant 0{\cdot}5 \tag{2.12}$$

The average selective advantage of a heterozygote over a homozygote at any locus (s) is given by

$$s = 1 - \bar{v}_{HO}/\bar{v}_{HE}, \tag{2.13}$$

where $\bar{v}_{HO}$ is the mean viability of individuals which are homozygous at a given locus; $\bar{v}_{HE}$ is defined analogously. Now (from Eq. 2.10)

$$\bar{v}_{HO} = v_o - \bar{n}_{HO}\,(v_o - v_N)\,/\,N, \tag{2.14}$$

and

$$\bar{v}_{HE} = v_o - \bar{n}_{HE}\,(v_o - v_N)\,/\,N, \tag{2.15}$$

where $\bar{n}_{HO}$ is the mean number of homozygous loci of individuals which are homozygous at a given locus and $\bar{n}_{HE}$ is the analogous number for individuals which are heterozygous at a given locus. Further one has

$$\bar{n}_{HO} = 1 + (N - 1)\,/\,2 = (N + 1)\,/\,2, \tag{2.16}$$

and

$$\bar{n}_{HE} = (N - 1)\,/\,2. \tag{2.17}$$

Substitution of Eq. 2.16 and Eq. 2.17 into Eq. 2.14 and Eq. 2.15 and finally into Eq. 2.13 yields

$$s = 1 - \frac{N(v_o + v_N) - (v_o - v_N)}{N(v_o + v_N) + (v_o - v_N)} \tag{2.18}$$

Assuming $v_{\frac{1}{2}N} = 0$, i.e. lethality of the complete multiple homozygote, we have (by Eq. 2.12)

$$L = 0.5,$$

and (by Eq. 2.18)

$$s = 1 - (N-1)/(N+1).$$

Values of s for different values of N, assuming $v_{\frac{1}{2}N} = 0$, are tabulated in Table 2.IIb.

These calculations show that with an additive model for multiple gene effects, the load is reduced to an acceptable level (cf. Eq. 2.12), but that at

TABLE 2.IIb. Selection coefficients (s) of single gene homozygotes in the hypothetical population under the model of additive gene effects for various values of N (the total number of overdominant loci) and for $v_N = 0$ (see text). The load in all cases equals 0.5.

N	s	N	s
1000	0.002	200	0.010
500	0.004	100	0.020
400	0.005		

the same time selection coefficients for approximately 1000 overdominant loci become very small. Accepting for the moment a load of 0.5 and a value of s of 0.005 as reasonable, a comparison of Tables 2.IIa and 2.IIb shows that the multiplicative model then allows the existence of 276 polymorphisms in the hypothetical population, whilst the additive model allows 400.

C. Diminishing returns

I propose to investigate a third mathematical model for multiple gene action with an asymptotic approach to the maximum viability as the number of heterozygous loci tends to its maximum. With this model the effect of increased heterozygosity decreases as the total number of heterozygous loci increases. This is essentially a model with 'diminishing returns' which is expressed by a relation of the type

$$v_{n+1} = v_n + t\,(c - v_n), \tag{2.19}$$

or

$$v_{n+1} - v_n = t\,(c - v_n)\;(n = 0, 1, 2,.., N), \tag{2.20}$$

where v_n stands for viability, and n, in contrast to models A and B now stands for the number of *heterozygous* loci; t and c are constants. Since the increase in viability per heterozygous locus, i.e. $t(c-v_n)$, must be positive and decreasing for all $n<N$ as v_n increases, we have the following conditions:

$$t > 0$$

and $\quad c \geqslant v_N.$

Setting $n=N-1$ in Eq. 2.20 one obtains

$$t = (v_N - v_{N-1}) \,/\, (c - v_{N-1}),$$

which is always less than or equal to unity because $c \geqslant v_N$
So we arrive at the conditions

$$\left.\begin{aligned} 0 < t \leqslant 1 \\ c \geqslant v_N. \end{aligned}\right\}$$

or, writing $\quad c = k.v_N,$

$$\left.\begin{aligned} 0 < t \leqslant 1 \\ k \geqslant 1. \end{aligned}\right\}$$

Assuming lethality of the complete multiple homozygote and non-lethality of all other genotypes (i.e. $v_0=0$ and $v_n>0$ for $n>0$), the solution of the recurrence relation (Eq. 2.19) is as follows:

$$v_n = k \cdot v_N \{1 - (1 - t)^n\} \tag{2.21}$$

For given values of t and N the corresponding value of k is found by writing $n=N$ in Eq. 2.21, yielding

$$k = 1 / \{1 - (1 - t)^N\} \tag{2.22}$$

Substitution of Eq. 2.22 into Eq. 2.21 yields

$$v_n = v_N \cdot \frac{1 - (1 - t)^n}{1 - (1 - t)^N} \tag{2.23}$$

This relation between viability and number of heterozygous loci is shown graphically in Fig. 2.2 for $N=1000$ and different values of t. Note that the graphs of Fig. 2.2 have no common upper limit if they were allowed to continue beyond $n=N$; this is because the limiting value of v, if n was allowed to increase to infinity, equals $k \cdot v_N$ which by Eq. 2.22 depends on both N and t. As a measure for the shape of the graphs, i.e. their curvature, one could use the parameter m, defined by

$$v_{\frac{1}{2}N} = m.v_N.$$

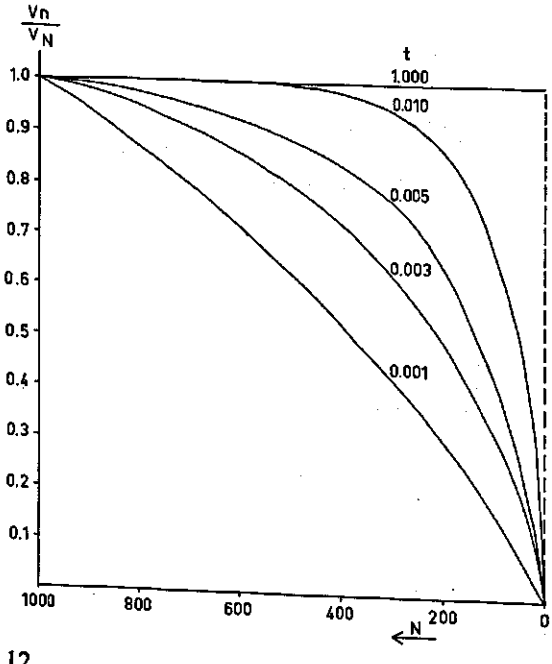


FIG. 2.2. The relation between viability (expressed as v_n/v_N) and number of heterozygous loci (n) according to Eq. 2.32 for $N=1000$ and different values of t. For the sake of completeness $t=1$ has been added (all viabilities being equal, except the complete multiple homozygote, having viability zero). Note that $t=0$ is non-existant. The direction in which n is plotted is reversed in order to make the model comparable with models A and B (cf. Fig. 2.7).

Values of m close to unity correspond to curves for which an increase in number of heterozygous loci beyond $\frac{1}{2}N$ has hardly any effect (cf. t=0.010 in Fig. 2.2).
Writing

$$m = v_{\frac{1}{2}N}/v_N,$$

one obtains by substitution of $n=\frac{1}{2}N$ into Eq. 2.23:

$$m = \frac{1 - (1 - t)^{\frac{1}{2}N}}{1 - (1 - t)^N} \tag{2.24}$$

which shows that m depends on both t and N.

Since the shape of the graph of v as a function of n is the most essential feature of the model, it is reasonable to start further considerations on the model from a given value of m. For given m and N, the corresponding value of t is found by noting that

$$1 - (1 - t)^N = \{1 - (1 - t)^{\frac{1}{2}N}\} \cdot \{1 + (1 - t)^{\frac{1}{2}N}\},$$

so that one obtains from Eq. 2.24:

$$(1 - t)^{\frac{1}{2}N} = (1 - m) / m$$

or

$$\log (1 - t) = \frac{2}{N} \log \left(\frac{1 - m}{m} \right). \tag{2.25}$$

A table for t can then be constructed. From Eq. 2.25 it is seen that, since $t<1$, m should satisfy the condition $0.5<m<1$.

The continuous analogon of Eq. 2.20, i.e.

$$dv/dn = t(c - v),$$

leads to

$$v_n = v_N \cdot \frac{1 - e^{-tn}}{1 - e^{-tN}} \; (o \leqslant n \leqslant N), \tag{2.26}$$

which for small values of t is equivalent to Eq. 2.23. Thus, large values of N (for which the discrete relationship (Eq. 2.20) may be approximated by its continuous analogon), should correspond to small values of t. Equation 2.25 shows that for large N (1000, say) t indeed will be very small in general.

Denoting for the moment Eq. 2.26 by $v_n=f(n)$, the mean population viability ($\bar{v}$) can be approximated by the series

$$\bar{v} = f(\bar{n}) + \frac{1}{2} f''(\bar{n}) . \text{ var } (n) + \ldots \tag{2.27}$$

Neglecting terms containing t^2 and higher orders of t, $\bar{v}$ can without serious error be approximated by

$$\bar{v} = f(\bar{n}) = v_{\frac{1}{2}N},$$

or

$$\bar{v} = m \cdot v_N \tag{2.28}$$

The load then becomes

$$L = 1 - \bar{v}/v_N = 1 - m \leq 0.5. \tag{2.29}$$

Defining the average selective advantage (s) of a heterozygote over either homozygote in the same way as with model B (Eq. 2.13):

$$s = 1 - \bar{v}_{HO}/\bar{v}_{HE},$$

and also using Eqs. 2.16 and 2.17 and writing

$$\bar{v}_{HO} = v_{\frac{1}{2}(N-1)} \text{ and } \bar{v}_{HE} = v_{\frac{1}{2}(N+1)}$$

one obtains

$$s = 1 - \frac{1 - (1 - t)^{\frac{1}{2}(N-1)}}{1 - (1 - t)^{\frac{1}{2}(N+1)}} \tag{2.30}$$

For a given set of values for $L = 1 - m$ (Eq. 2.29) and N, the corresponding values of t and s have been calculated according to the expressions Eq. 2.25 and Eq. 2.30 and have been tabulated in Table 2.IIc. It is seen that for values of $N \geqslant 500$ the selection coefficients become very small, perhaps too small to ensure the maintenance of stable polymorphisms.

TABLE 2.IIc. Values of t and s (selection coefficient of single gene homozygotes) with the model with diminishing returns for different values of $L = 1-m$ and N. The parameters t and m are discussed in the text; N: total number of overdominant loci. The values of L (load) = $1-m$ are chosen, t is calculated from Eq. 2.25, s from Eq. 2.30.

	$L = 0.4$ $m = 0.6$	$L = 0.3$ $m = 0.7$	$L = 0.2$ $m = 0.8$	$L = 0.1$ $m = 0.9$
$N = 100$	0.00808 (t)	0.01680	0.02734	0.04299
	0.0161 (s)	0.0125	0.0092	0.0059
$N = 200$	0.00405	0.00844	0.01760	0.02173
	0.0081	0.0063	0.0040	0.0027
$N = 500$	0.00160	0.00338	0.00552	0.00875
	0.0032	0.0026	0.0018	0.0011
$N = 1000$	0.00080	0.00168	0.00277	0.00437
	0.0016	0.0013	0.0009	0.0007

In conclusion:

The models discussed so far are rather unsatisfactory. Model *A* must be regarded as unrealistic because it imposes a too heavy load on the population; with models *B* and *C* the load is reduced considerably, but at the same time very small selection coefficients are implied. They probably are too small to maintain balanced polymorphisms in a finite population. Assuming for the moment that a selection coefficient $\geqslant 0.01$ is required and that the load should not exceed 0.5, it is seen from Table 2.IIa, 2.IIb and 2.IIc that none of the models *A*, *B* and *C* can account for the maintenance of 500 or more polymorphisms in the hypothetical population.

D. Threshold model (see KING, 1967)

The models *A*, *B* and *C* do not account for the variation which is brought about by non-genetic random effects. Random effects have been accounted for in the threshold model proposed by KING (1967), which makes it a very elegant model for natural selection. The relation between number of homozygous loci and viability obtained by KING is essentially the same as the one proposed by SVED, REED and BODMER (1967) discussed earlier.

In KING's model the non-genetic factors which affect fitness are normally distributed. An individual is assigned a hypothetical value, the 'survival factor parameter', which in fact is a random variable. The mean of the 'survival factor parameter' depends on the individual's genotype: it is a linear function of its number of homozygous loci. The variance of the 'survival factor parameter' is interpreted as the variance due to random effects. The survival probability is interpreted as the probability that the survival factor parameter takes a value less than *t*, a threshold value. Thus, the survival probability as a function of the number of homozygous loci is identical with the cumulative normal distribution (see Fig. 2.3).

As KING demonstrates with a set of numerical calculations, the model accomodates the maintenance of many balanced polymorphisms (of the order of 1000) through an average selective advantage of heterozygotes over homo-

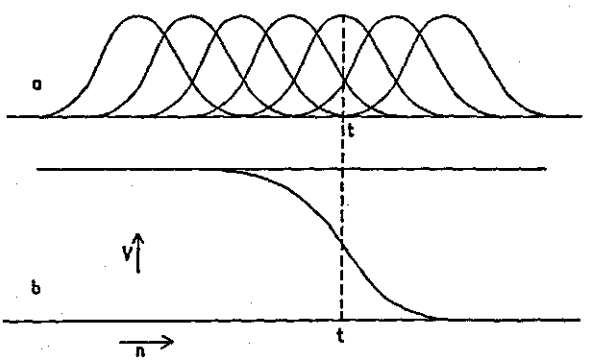


FIG. 2.3. (From KING, 1967). a. Probability distributions of the 'survival factor parameter' for different genotypes. b. Viability (v) as a function of the number of homozygous loci (n). A linnear relationship is assumed between n and the mean value of the survival factor parameter.

zygotes as large as 0.01 without imposing a heavy load on the population ($L \approx 0.5$).

The essential feature of this model which makes it preferable to models *A*, *B* and *C* is its property that selection coefficients of a realistic magnitude (e.g. 0.01) do not impose a heavy load. This justifies a further formal exploration into the merits of the threshold model. This is most easily done by comparing it with the additive model (*B*) (see Fig. 2.4). Since for large values of N, the variance of the number of homozygous loci per individual is relatively small as compared to the range of this number, the curve *D* in Fig. 2.4 can be assumed to be linear for the range in which most of the genotypes actually occur. Rare genotypes contribute little to the mean and variance of viability in the population. Thus, as an approximation one can write (see model *B*):

$$s = 1 - \frac{f(\frac{1}{2}(N+1))}{f(\frac{1}{2}(N-1))} = 1 - \frac{f(\frac{1}{2}(N-1)+1)}{f(\frac{1}{2}(N-1))}$$

where f denotes the function which relates viability to number of homozygous loci: $v = f(n)$.
Writing

$$f(\tfrac{1}{2}(N-1)+1) = f(\tfrac{1}{2}(N-1)) - \Delta v,$$

we have

$$s = \Delta v \,/\, f(\tfrac{1}{2}(N-1)),$$

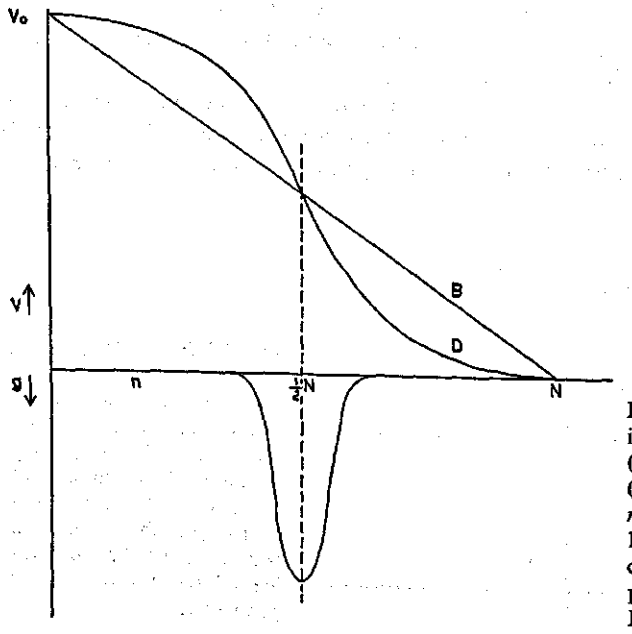


FIG. 2.4. Graphical comparison of the additive model (*B*) and the threshold model (*D*).
n: number of homozygous loci; v: viability; g: frequency distribution of n in the population.
For explanation see text.

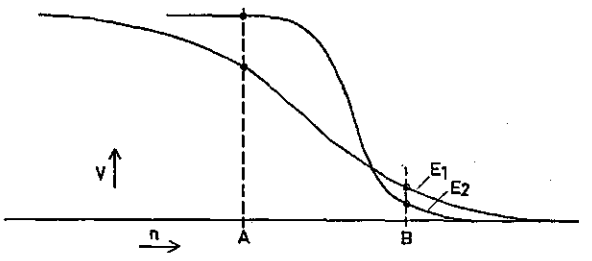


FIG. 2.5. Viability (v) as a function of the number of homozygous loci (n) in two different environments according to the threshold model. E_1: heterogeneous environment; E_2: homogeneous environment. A and B are defined genotypes; the ratio v_A/v_B differs for the two environments.

which approximately equals $\Delta v/\bar{v}$, provided Δv is small as compared with $\bar{v}$. The quantity Δv represents the amount by which v increases when the number of homozygous loci decreases from $\frac{1}{2}(N-1)+1$ to $\frac{1}{2}(N-1)$ and thus can be interpreted as the tangent to $v=f(n)$ in the point $\{\frac{1}{2}N, f(\frac{1}{2}N)\}$. Having found $s \approx \Delta v/\bar{v}$, one immediately sees from Fig. 2.4 that model D implies larger selection coefficients than model B, whilst the mean viability and therefore the load ($L=1-\bar{v}/v_{max}$) is in both cases approximately equal.

An elegant feature of the threshold model is seen when the relative viabilities of two defined genotypes (A and B) are compared in two different environments. In Fig. 2.5 the relationship between number of homozygous loci and viability is schematically shown for two environments, E_1 and E_2; E_1 is heterogeneous with large variance of random effects, E_2 is homogeneous with smaller environmental variance. The ratio of the two viabilities (v_B/v_A) is changed when the genotypes A and B are removed from E_1 to E_2. This point will be extended later on (chapter 3). This feature of the model, which may be referred to as genotype by environment interaction, also explains any inbreeding depression curve. The form and shape of such a curve depends on the distance between the threshold value t and the mean number of homozygous loci in the non-inbred population and the variance of random effects, respectively. These may differ for different species and different environments, which, as KING (l.c.) points out, results in different inbreeding depression curves.

As far as the viability component of fitness is concerned, KING's model seems to be correct. However, when regarding total fitness, defined as the expected number of offspring of a zygote, this can no longer be treated in terms of probabilities only. Curves for total fitness as a function of the number of homozygous loci which are of the same form as the curve of Fig. 2.3 are obtained only if the expected number of offspring of adults approaches a maximum as the number of heterozygous loci tends to its maximum. This means that KING's model, in which the form of this curve is essential, only is correct if there is an upper limit to fertility, which is reached asymptotically. Since on physiological grounds there must be an upper limit to fertility, the relation between number of homozygous loci and net fitness may very well be of the form proposed by KING.

This 'generalized threshold model', applying to total fitness, also accounts for soft selection as can be seen from Fig. 2.6. Here the relation between total

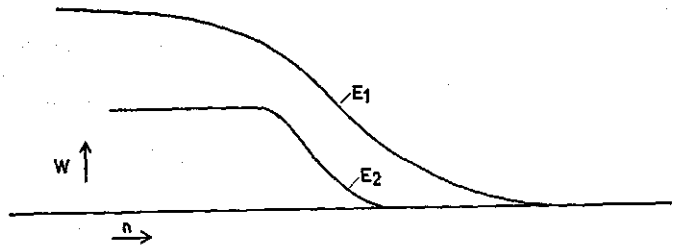


FIG. 2.6. Soft selection with the generalized threshold model. w: total fitness; n: number of homozygous loci. E_1 and E_2 are two different environments. In E_2 soft selection is an important component of natural selection as compared with E_1. In E_2 there is also less variation in random effects.

fitness (w) and number of homozygous loci (n) is shown schematically for two environments, E_1 and E_2. In E_2 there is a smaller environmental variance than in E_1 and besides the maximum fitness is lower than in E_1. The latter can be exemplified by overcrowding or a higher rate of predation, which can be understood as soft selection (cf. page 6).

Finally a graphical comparison of the models A, B, C and D is given in Fig. 2.7. As argued on page 16, the selection coefficient s may be approximated by

$$s = \Delta v \,/\, f\left(\tfrac{1}{2}(N-1)\right) \approx \Delta v/\bar{v},$$

in which Δv is the slope of the curve at $n=\frac{1}{2}N$. The graphs immediately show that only model D can account for both relatively large selection coefficients and a relatively small load.

2.3. FREQUENCY DEPENDENT SELECTION

Frequency dependent selection implies that the fitness of a genotype is in one way or another related to its relative frequency (and possibly that of other genotypes). If frequency dependent selection plays a role in the maintenance of genetic polymorphisms it should be of an equilibrating type, i.e. its mechanism should cause the gene frequency to move towards some stable equilibrium value. (Theoretically many other types of frequency dependent selection can exist, e.g. types leading to fixation of an allele or with unstable equilibrium gene frequencies.)

On the basis of their experiments KOJIMA and YARBROUGH (1967), KOJIMA and TOBARI (1969) and TOBARI and KOJIMA (1967) suggest that at the loci (or chromosome inversions) under consideration the less frequent genotype is favoured. When this mechanism acts upon a given locus, gene frequency will reach a stable equilibrium at which all genotypes have equal fitnesses (are in a sense equally favoured). In an equilibrium population of this type all relative

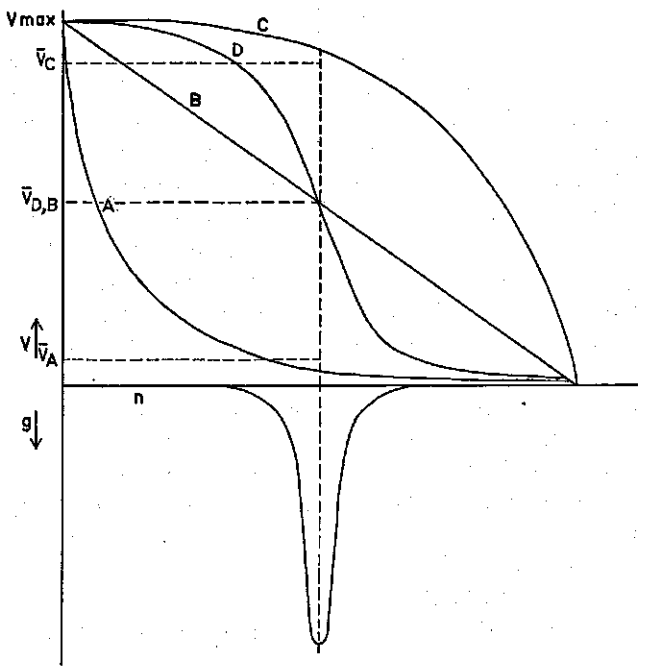


FIG. 2.7. A graphical comparison of the models *A*, *B*, *C* and *D*. *v*: viability; *n*: number of homozygous loci; *N*: total number of overdominant loci; *g*: frequency distribution of *n* in the population.
A: multiplicative model
B: additive model
C: 'deminishing returns' model
D: threshold model
$\bar{v}_A$, etc: mean viability with model *A*, etc.
See further text.

fitnesses equal unity and, consequently, there is no segregational load, though the population is polymorphic. Strictly speaking, the load only is defined when there is differential fitness in the population and thus the load of a population as mentioned above is not defined. However, when considering the equilibrium population as a limiting case of a non-equilibrium population (cf. also section 2.1), the segregational load equals zero. The absence of 'substantial' segregation load in a population which is in stable equilibrium as a result of frequency dependent selection has been used by KOJIMA (see KOJIMA and YARBROUGH, 1967 and KOJIMA, 1971) as an argument in favour of frequency dependent selection as a major cause of the existence of many polymorphisms in natural populations. It is then tacitly assumed that the absence of segregational load is an optimal situation from the evolutionary point of view. KOJIMA (1971) has made a set of numerical calculations which are to demonstrate that also with finite population size (which will cause deviations from the exact equilibrium frequency) the expected load is in general less with frequency dependent selection than with overdominance. However, the utility of such calculations is doubtful because they are based on relative fitnesses, and therefore do not adequately describe the mean population fitness. From the evolutionary point of view, in particular when considering competition on the population level, mean population fitness rather than segregational load is of interest. In theory it is possible that when fitness is measured on the relative scale, that is with the maximum fitness of the population as a reference, the

mean population fitness is maximized by natural selection, but that mean population fitness when measured on the absolute scale does not change (or even decreases). An illustration of this theoretical possibility is given in Table 2.III where viabilities are frequency dependent such that less frequent genotypes have higher fitness. This example shows that the equilibrium and non-equilibrium populations may have equal mean absolute fitnesses, whilst paradoxically the segregational load, calculated on the basis of relative fitnesses would indicate that the equilibrium population has the highest mean fitness. Thus, frequency dependent selection may be a mechanism through which polymorphisms are maintained, but the fact that it implies absence of segregational load in equilibrium populations does not in itself make it a more likely mechanism.

A difficulty with frequency dependent selection is that it can hardly be detected in a reliable way from population data (population censuses in successive generations), as has been pointed out by PROUT (1969). He demonstrates that deviations from the estimation model may mimic frequency dependency of fitnesses (see also page 28); in these cases the frequency dependency of fitnesses is merely a statistical artefact. The estimation of fitnesses from population data essentially is a maximum likelihood procedure (see PROUT; 1965, 1969; DUMOUCHEL and ANDERSON, 1968; STAM, 1971), based on the following assumptions:

1. There is random mating,
2. Selection is completed at the time of census,
3. Fitnesses are constant over generations.

Deviations from one or more of these assumptions may result in non-convergence of the likelihood or negative fitness estimates, as demonstrated by PROUT (l.c.). So non-convergence of the likelihood or negative fitness estimates

TABLE 2.III. A theoretical example of frequency dependent selection: the less frequent genotypes have the higher fitnesses. Populations I and III are not in equilibrium; population II is the equilibrium population. It is seen that the three populations have equal mean absolute viabilities, whilst the equilibrium population has the maximum mean relative viability.

	genotype	relative frequency	absolute viability	mean absolute viability	relative viability	mean relative viability	load
	AA	0.81	0.60		0.750		
pop. I	*Aa*	0.18	0.70	0.62	0.875	0.785	0.215
	aa	0.01	0.80		1.000		
	AA	0.25	0.62		1.000		
pop. II	*Aa*	0.50	0.62	0.62	1.000	1.000	0.000
	aa	0.25	0.62		10000		
	AA	0.01	0.80		1.000		
pop. III	*Aa*	0.18	0.70	0.62	0.875	0.785	0.215
	aa	0.81	0.60		0.750		

can not be considered as proof for non-constancy of fitnesses and certainly not for frequency dependency of fitnesses. (Frequency dependent selection in a non-equilibrium population will result in changing fitnesses over generations.) Frequency dependent selection, therefore, can to my opinion not be reliably demonstrated from comparisons between genotype frequencies of successive generations, but must be inferred from successive counts during the life span of one generation. Several examples of the latter approach can be found in literature (SOKAL and KARTEN, 1964; KOJIMA, 1969).

2.4. NEUTRAL ALLELES

It has been suggested by KIMURA (1968) that many protein polymorphisms as found in e.g. *Drosophila* and man are selectively neutral. If this is true, the polymorphism is the outcome of mutation, random drift and migration. With neutral alleles one expects to find different sets of alleles in isolated sub-populations. However, in general the same sets of alleles are found, and from this ROBERTSON (1968) concluded that most polymorphisms are maintained by selection.

Using the concept of identity by descent, KIMURA and CROW (1964) derived the expression

$$n_e = 4\,N_e u + 1\,, \tag{2.31}$$

in which n_e is the 'effective number of alleles', N_e is the variance effective population number and u is the mutation rate. By definition

$$n_e = 1\,/\,\Sigma\,p_i^2\,,$$

where p_i is the frequency of the i-th allele. Equation 2.31 is based on the assumption that every (neutral) mutation of the allele leads to a new, not pre-existing allele.

As shown by KIMURA (1971), a value of $n_e = 1.1$ roughly corresponds to an average fraction 0.3 of all enzyme loci investigated being polymorphic, which is in good agreement with the observations. (In this calculation a population in which the frequency of one allele is greater than 0.95 is supposed to be classified as a monomorphic one.)

It has long been known (for a recent proof see MARUYAMA, 1970) that conspicuous divergence of subpopulations is possible only when the mean number of gametes which each isolate exchanges with other isolates is less than one per generation. MARUYAMA (1970) derived the expression (based on the so called 'island model')

$$n_e = 4N_e\,nu + \frac{nu}{m} + 1, \tag{2.32}$$

where N_e is the variance effective number of each isolate, n is the number of isolates, u is the mutation rate, m is the rate at which each colony exchanges gametes with all other colonies and n_e is the effective number of alleles. When com-

pared with Eq. 2.31 one sees that the increase in effective number of alleles due to subdivision equals $\frac{n\,u}{m}$, which for reasonable values of m (e.g. $m > \frac{2}{N_e}$) becomes very small. As an example consider first a single closed panmictic population of effective size 25,000. Let the mutation rate be 10^{-6}, then, according to Eq. 2.31, $n_e = 1.1$. Now let the population be subdivided into 25 isolates each of effective size 1000 and let $m = 0.002$ (i.e. an average of 4 gametes per generation are exchanged between the isolates). Then, according to Eq. 2.32, $n_e \approx 1.11$.

These considerations show that a group of sub-populations of a mobile species like a *Drosophila* species effectively behaves as a single panmictic population, and therefore an effective number of alleles of 1.1 can be considered as realistic (cf. KIMURA and OHTA, 1971).

Another possible explanation for the existence of polymorphisms with neutral alleles is supported by so called 'associative overdominance'. Associative overdominance is generated by linkage disequilibrium between an overdominant locus and a neutral locus: when not in linkage equilibrium, the neutral locus behaves as if it were overdominant (OHTA and KIMURA, 1970). In finite populations there will be linkage disequilibrium as a result of the finite sample of gametes which contributes to the next generation and thus associative overdominance is generated.

2.5. SYNTHESIS AND HYPOTHESIS

Among the mechanisms by which genetic polymorphisms can be maintained in finite populations, overdominance has received most attention. As discussed in section 2.2, overdominance with multiplicative gene effects on fitness can most likely be ruled out because it drastically reduces mean population fitness, even with a moderate number of overdominant loci. The models with additive effects and my own model with 'diminishing returns' in their turn are unsatisfactory because they do not accomodate realistic values of selection coefficients. The threshold model proposed by KING does not suffer from these 'disadvantages' and besides it is very attractive because of its generality.

The analysis of migration models (see MARUYAMA, 1970) has shown that for populations which are divided into sub-populations and which have a reasonable migration rate, the force of random drift is of far less importance than previously assumed. As a consequence of this, KIMURA (1970) suggested that many polymorphisms in fact may be selectively neutral.

The experiments of VANN (1966) and of SVED and AYALA (1970) have demonstrated the existence of overdominance at the chromosome level in *Drosophila*, that is overall overdominance when blocks of loci are considered.

Now, when combining the results of OHTA and KIMURA (1970) on the development of associative overdominance with the observation on overdominance at the chromosome level, I propose the following hypothesis: Part of the observed polymorphic loci are truly overdominant (or may behave as overdominant loci because of very close linkage of two dominant loci in repulsion phase) but most

of the polymorphic loci, which are in fact selectively neutral, show associative overdominance as a result of overall linkage disequilibrium in finite populations.

On the basis of this hypothesis some comments can be made on the experimental results obtained by KOJIMA and associates. (In connection with my own experiments this point is of special interest.) An important corollary of the above hypothesis is that linkage disequilibrium may result in spurious frequency dependency of fitnesses, as will be discussed below.

Suppose that small samples from a polymorphic base population are taken to establish the two homozygous marker strains and that the marker locus is fitness-neutral. With the individuals from these homozygous marker strains populations are initiated at different levels of marker gene frequency. The expected pattern of fitnesses of the marker genotypes in these populations corresponds to that of overdominance if the marker locus is linked with an overdominant locus (or chromosome segment). This will initially cause the gene frequency at the marker locus to shift in the direction of its apparent equilibrium value. However, if the population is of a reasonable size, the linkage disequilibrium decreases as a result of recombination and the apparent fitnesses of the marker genotypes will converge to a common value as the population breeds. As the fitnesses of the marker genotypes approach each other, the apparent selection pressure at the marker locus decreases (see Fig. 2.8). Thus, the observed changes in marker gene frequency coincide with changes in fitness of the marker genotypes. Knowing little about the linkage disequilibrium in the initial populations, one might conclude that fitnesses are frequency dependent (i.e. fitnesses are functions of the genotype frequencies). In the situation considered here however, the changes in fitness are not *caused* by the changes in genotype frequen-

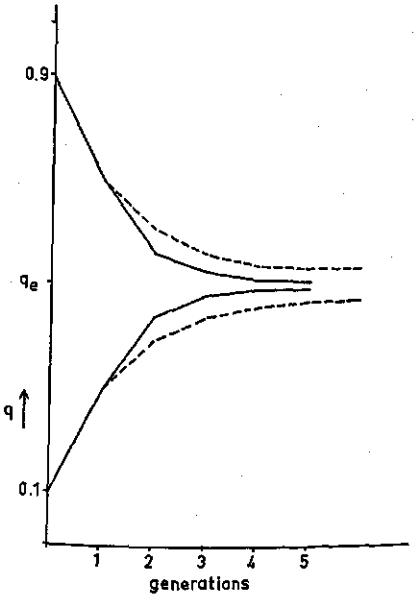


FIG. 2.8. Changes in gene frequency (q) with constant selection coefficients (solid curves) and decreasing selection coefficients (dotted curves) for an overdominant locus with intermediate equilibrium gene frequency (q_e). The dotted curves may easily be mistaken to be the outcome of frequency dependent selection. For explanation see text.

cies but the two simply *coincide*, and therefore there is no true frequency dependence of fitnesses. The basic shortcoming of experiments as described above is that changes in fitness can not a priori be ascribed to the changes in gene frequency when little is known about linkage disequilibrium in the initial populations. To check whether fitnesses are really functions of gene frequency, the experiment should be started with a range of initial marker gene frequencies instead of two extreme values, since, if for the set of gene frequency shifts during the first generation interval a unique set of fitness parameters can be found, fitnesses are obviously not frequency dependent. Only if no unique set of parameters can be found fitnesses are indeed related to gene frequency. In my own experiments (see chapter 3) a unique set of fitness parameters was found.

The effect of linkage disequilibrium in the initial populations on the results of this type of experiments has been recognized by KOJIMA (1971) but he failed to conclude that the experimental results, if indeed caused by associative overdominance, do not justify his speculation that real frequency dependent selection is a major force in the maintenance of genetic polymorphisms in natural populations.

3. EXPERIMENTS AND DISCUSSION OF RESULTS

3.1. MATERIAL AND GENERAL METHODS

Wild type and mutant stocks of the flour beetle *Tribolium castaneum* HERBST (*Coleoptera*, family *Tenebrionidae*) were kindly provided by Dr. A. SOKOLOFF, San Bernardino, California, U.S.A. The experiments were started with a wild type strain (to be denoted by 'CAW') and a homozygous black strain ('CAB'). Black is an autosomal mutant affecting body colour with intermediate expression in the heterozygote. The wild type body colour of *T. castaneum* is red-rust; the gene *b* causes the formation of a dark pigment. The heterozygote is referred to as 'bronze' (SOKOLOFF, 1966).

Populations were kept in 'population vials' (diameter 4 cm, height 8 cm) on a standard medium consisting of fine sifted whole wheat flour to which 5% dried yeast was added. The medium was heated to 60°C for 16 hours before it was stored for later use. The vials were closed with a ring-shaped lid into which a fine meshed polyether cloth was mounted for air circulation. Cultures were placed in a room at 31°C and ± 70% relative humidity. Under these conditions, with moderate population density, the developmental period from egg to adult averaged 30 days. The removal of adults and pupae from the medium was performed with sieves. Virgin females were collected by sexing the pupae and rearing the sexes separately. In the experiments generations were non-overlapping. A population was initiated by 200 virgin adults (sex ratio 1:1) in different proportions from CAW, CAB and their reciprocal F_1's. These were put in a 'mating vial' (diameter 5 cm, height 6 cm) containing 20 g medium and were allowed to mate and lay eggs for 2 days. Then the medium (with eggs) was transferred to a population vial. This moment marks the beginning of a new generation. After 35 days a random sample of approximately 500 sexually mature adults was taken from the population vial and transferred to a mating vial (now containing 30 g medium). After a 2 day stay in the mating vial, the adults were removed from the medium and counted and classified; at the same time the medium (with eggs) was transferred to a population vial. Then the cycle was repeated. For further experimental details see under 3.2 (Population Experiments) and 3.4 (Fertility Experiments).

3.2. POPULATION EXPERIMENTS

In the present experiments population size was kept approximately constant and therefore population density factors were not taken into account. The total number of individuals from which the stocks CAW and CAB were started in our laboratory did not exceed 50 per stock, which are relatively small samples. The stocks CAW and CAB have not been intercrossed to obtain 'isogenic' black

and wild type strains. When it turned out that CAW and CAB had markedly different genotypic backgrounds, attention focused on the background structures.

Heterozygous individuals were obtained from the reciprocal crosses between CAW and CAB. Adults of the marker genotypes (+/+, +/*b* and *b*/*b*) each with equal numbers of males and females were mixed in Weinberg-Hardy frequencies and placed into a mating vial. The total size of the initial parental populations was 200; in each of the subsequent generations ± 500 adults were sampled to be the parents of the next generation. Five initial relative frequencies of the *b* allele were chosen, viz: 0.07, 0.3, 0.5, 0.7 and 0.9. With each initial gene frequency four replicate populations were started simultaneously. These $5 \times 4 = 20$ populations were maintained for 8 generations in the way described sub 3.1.

The results of the population experiments are summarized in Table 3.I and Fig. 3.1, where the frequency of the *b* allele is plotted for the successive generations.

The first generation interval will be considered first. The shifts in gene frequency from generation 0 to generation 1 indicate that selection pressure against the *b* allele must be considerable. In order to obtain an impression of the magnitude of the selection coefficients during the first generation interval, the follow-

TABLE 3.I. Data (frequencies of *b*-allele) of the population experiment. Roman numerals indicate replicates.

gene-rations	$q_o = 0.07$				$q_o = 0.3$				$q_o = 0.5$			
	I	II	III	IV	I	II	III	IV	I	II	III	IV
0	0.06	0.07	0.07	0.07	0.30	0.30	0.30	0.30	0.50	0.50	0.50	0.50
1	0.04	0.05	0.04	0.05	0.19	0.21	0.19	0.19	0.34	0.33	0.34	0.34
2	0.03	0.03	0.03	0.04	0.13	0.15	0.14	0.15	0.26	0.26	0.27	0.26
3	0.03	0.03	0.02	0.03	0.10	0.12	0.11	0.15	0.24	0.23	0.26	0.22
4	0.03	0.03	0.01	0.03	0.08	0.10	0.10	0.12	0.21	0.23	0.25	0.20
5	0.02	0.02	0.01	0.03	0.09	0.09	0.11	0.14	0.19	0.21	0.22	0.18
6	0.02	0.03	0.01	0.03	0.08	0.09	0.10	0.13	0.19	0.19	0.21	0.17
7	0.02	0.02	0.02	0.03	0.08	0.09	0.11	0.12	0.18	0.19	0.19	0.16
8	0.02	0.01	0.02	0.02	0.08	0.09	0.11	0.11	0.19	0.18	0.19	0.16

	$q_o = 0.70$				$q_o = 0.90$			
	I	II	III	IV	I	II	III	IV
0	0.70	0.70	0.70	0.70	0.90	0.90	0.90	0.90
1	0.47	0.48	0.50	0.46	0.67	0.66	0.69	0.66
2	0.39	0.37	0.40	0.39	0.59	0.56	0.63	0.57
3	0.35	0.35	0.36	0.38	0.58	0.52	0.59	0.54
4	0.32	0.28	0.35	0.39	0.52	0.49	0.57	0.53
5	0.31	0.28	0.32	0.36	0.48	0.48	0.55	0.52
6	0.29	0.26	0.35	0.33	0.45	0.46	0.54	0.50
7	0.26	0.25	0.31	0.32	0.45	0.49	0.56	0.53
8	0.28	0.26	0.33	0.31	0.44	0.48	0.54	0.50

gene-rations	$q_o = 0.60$		
	I	II	III
0	0.60	0.60	0.60
1	0.60	0.59	0.60
2	0.60	0.58	0.59
3	0.59	0.58	0.59
	$q_o = 0.80$		
0	0.80	0.80	0.80
1	0.80	0.79	0.78
2	0.79	0.80	0.77
3	0.78	0.79	0.77

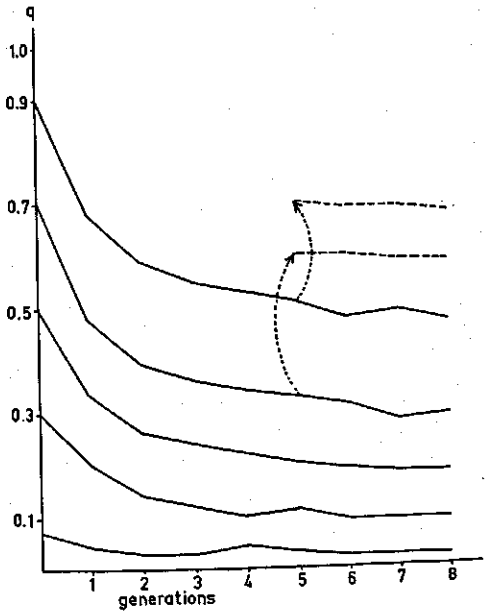


FIG. 3.1. Frequency of b allele (q) in the different sets of populations. Each solid curve represents the average of 4 replicates. Dotted curves represent populations readjusted to $q_0'=0.8$ and $q_0'=0.6$ (see text).

ing approach is sometimes useful, though in our case not correct, as it will be assumed for the moment that selection acts through differential viability only. Then the relative viabilities of the genotypes $+/+$ and b/b can be estimated by the methods presented by PROUT (1969) and STAM (1971). The relative viability estimates for the populations initiated at $q_0=0.3$, 0.5, 0.7 and 0.9 are shown in Table 3.II. The estimates for the population with $q_0=0.07$ were not calculated because the number of b/b individuals in generation 1 was too low for efficient estimation. Although the estimation model used here is not correct because genotypic fertility differences probably played an important role (see sub 3.4), it is clear that selection coefficients must be considerably large in order to give rise to the observed shifts in gene frequencies. When this estimation model is used, one might conclude from Table 3.II that selection coefficients are frequency

TABLE 3.II. Estimates of relative viabilities of genotypes $+/+$ and b/b (v_{++}/v_{+b} and v_{bb}/v_{+b}), calculated from the shifts in genotype frequencies during the first generation interval. The estimation model is based on the assumptions of random mating and absence of genotypic fertility differences (see PROUT, 1969). q_o: frequency of b-allele among parents.

q_o	$\frac{v_{++}}{v_{+b}}$	$\frac{v_{bb}}{v_{+b}}$	q_o	$\frac{v_{++}}{v_{+b}}$	$\frac{v_{bb}}{v_{+b}}$
0.3	1.83	0.66	0.7	2.71	0.41
0.5	2.10	0.56	0.9	3.96	0.21

dependent. As pointed out by PROUT (1965) this may be an artefact caused by an incorrect estimation model (see also ch. 2, p. 20). As an alternative approach to the order of magnitude of the selection pressure during the first generation interval, one can assume absence of genotypic viability differences and ascribe the shifts in gene frequency to parental fertility differences only. The direction in which q is changed is the same for all initial gene frequencies, i.e. 'unidirectional' selection against the b allele. This suggests that the genotypes $+/+$ and $+/b$ are superior in fertility to b/b (sub 3.3. it will be discussed why partial dominance for fitness of the marker genotypes in the parental generation must be considered unlikely). With this model, that is selection acting only through a reduced fertility of b/b, a least square procedure was applied in order to obtain the selection coefficient of b/b which fitted best to the observations (Fig. 3.1), yielding $s=0.85$ with an excellent fit. Although this procedure too might be incorrect, it once again shows the appreciable selection pressure against the b allele during the first generation interval. Because with this model a unique set of fitness parameters can be found for all initial gene frequencies, the data do not allow frequency dependent fertilities. Later on (3.4.) it will be shown that viability differences only play a minor role in the change in gene frequency; so frequency dependent selection is less probable.

Including now also the second and further generations, the following points are obvious from Fig. 3.1:

1. The b allele is selected against.
2. Selection pressure against the b allele decreases as the populations breed.
3. After about 8 generations the gene frequencies apparently reach equilibria, the equilibrium frequency being dependent on the initial gene frequency.

It is clear that the strains CAW and CAB have different genotypic backgrounds: if these were similar, then selection would act through the b locus only and the b allele would become either lost or fixed in all populations, or, in the case of overdominance or equilibrating frequency dependent selection, it would be maintained at some intermediate level, which should be the same for all initial gene frequencies. Thus frequency dependent selection is excluded. Fig. 3.1 suggests that the ultimate gene frequency in a population is determined by its 'history'. This point was checked as follows. In the 5th generation virgin adults were collected from the populations with $q_0=0.9$ and $q_0=0.7$ ($\bar{q}_5=0.51$ and $\bar{q}_5=0.32$, respectively); with these adults new initial parental populations were composed (3 replicates of 200 individuals) such that the marker genotypes were in Weinberg-Hardy frequencies with $q_0'=0.8$ and $q_0'=0.6$, respectively. Thus, the gene frequencies of the populations with $q_0=0.9$ and $q_0=0.7$ were artificially set to values close to their initial values (the exact initial values could not be reconstructed because the sample of adults contained too few b/b individuals). The gene frequencies of these populations is given by the dotted curves in Fig. 3.1. The results are unambiguous: the gene frequency in these 're-adjusted' populations remains nearly constant as would be the case if the black locus were fitness-neutral.

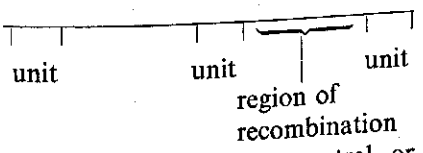
The simplest explanation for these observations is as follows. The black locus itself is fitness-neutral and selection acts through another locus, *A-a*, say. The initial populations are not in linkage equilibrium, CAW carrying the favoured allele, *A*, say. As the populations breed the apparent selection against the *b* allele becomes weaker and weaker because the association between the alleles *a* and *b* is broken down on the way to linkage equilibrium. With this hypothesis a set of fitness parameters and a recombination fraction might be fitted to the observations. (Note that this hypothesis implies an appreciable selection coefficient at the *A-a* locus.) It would be more satisfactory if also a more general explanation could be given, viz. in terms of homozygosity of chromosome segments (in which crossing over can take place) rather than a specific locus (or e.g. a small inversion). For this reason another hypothesis, which accounts for the joint action of many fitness loci, has been constructed.

3.3. Hypotheses

In chapter 2 it was discussed how selectively neutral loci can stay polymorphic in a population as the result of their association with selectively non-neutral loci (associative overdominance). Although with this model the number of overdominant loci per unit chromosome length is less than with the models which assign a selective force to each polymorphic locus, the model with neutral loci still implies overdominance at the chromosomal level. Overdominance at the chromosomal level is the basic assumption in the following approach to the explanation of the experimental results.

When a hypothesis which accomodates multiple gene action on fitness is formulated, a relationship between number of heterozygous loci and fitness has to be assumed. Several types of this relationship have been discussed in chapter 2. For our present purpose the assumption suffices that fitness is a monotonically increasing function of the number of heterozygous loci. In formulating the possible causes of the fitness difference between the original marker genotypes (CAW, CAB and their F_1) the following assumptions will be made.

1. Each chromosome can be considered as a number of successive units of intrachromosomal recombination, (i.e. recombination can only take place between units) as diagrammed below:



2. The marker locus is fitness-neutral, or rather the unit which contains the marker locus is fitness-neutral.
3. The other units may or may not contain a fitness locus and, if so, they show overdominance for fitness.

4. Since no general statements can be made about the distribution of fitness loci over chromosomes, it will be assumed that these follow a uniform distribution in the sense that each unit has equal probability to contain an overdominant locus (or block of loci).
5. Fitness heterozygotes at any unit have equal average advantage over either homozygote.
6. The strains CAW and CAB have been derived from the same base population.

In the following a chromosome unit which has originated from the wild type population (CAW)is denoted by a 0; a unit originated from the marker strain (CAB) by a 1. With this notation the parental strains and their F_1 are represented as in table 3.III. It should be noted that the 'state' of a digit (0 or 1) does not correspond to an alternative of allelic units but only indicates which of the parental strains the unit originates from. For our purpose we now only need to assign probabilities of homozygosity to each of the classes 0/0, 0/1 and 1/1 (0/1 and 1/0 are equivalent). In the following these will be denoted by X_{00}, X_{10} and X_{11}, respectively. Thus X_{00} stands for the probability of homozygosity for any two allelic units from the wild type strain. This implies that the average total number of homozygous fitness loci of an individual from CAW is proportional to X_{00}.

Now, since fitness was assumed to be a monotonic function of the total number of homozygous overdominant fitness loci, the difference in fitness between the two parental stains can be expressed in terms of X_{00} and X_{11} (viz. $X_{00} \neq X_{11}$). Since CAB clearly is the least fit of the two strains,

$$X_{00} < X_{11} ,$$

which corresponds to a higher inbreeding coefficient of CAB, or at least an excess of homozygosity in CAB. The following two hypotheses (A and B) are based on random and non-random distribution of the excess of homozygosity in CAB over the chromosomes:

A. randomness of the excess of homozygosity in CAB referring to $X_{00} < X_{11}$ for all chromosomes.

TABLE 3.III. Binary representation of genomes of the original marker genotypes; $+/b$: marker locus.

CAW	$\frac{...0000+0000...}{...0000+0000...}$,	$\frac{...00000...}{...00000...}$,	etc.
F_1	$\frac{...0000+0000...}{...1111\ b\ 1111...}$,	$\frac{...00000...}{...11111...}$,	etc.
CAB	$\frac{...1111b111...}{...1111b111...}$,	$\frac{...11111...}{...11111...}$,	etc.

B. non-randomness of the excess of homozygosity in CAB, referring to $X_{00}<X_{11}$ for marker chromosomes only, the other chromosomes bearing no fitness contrast between the strains.

Hypothesis A

The strains CAW and CAB differ with respect to their inbreeding coefficents. With inbreeding the mean number of homozygous loci of the individuals increases, thus reducing the mean population fitness. When both parental strains have been inbred this is expressed as

$$X_{00} > X_{10} < X_{11} \tag{3.1}$$

The expected mean number of homozygous loci of hybrid individuals, obtained by crossing two 'randomly inbred' populations (derived from the same base population) equals that of individuals in a non-inbred population. Thus, when only one of the parental strains has been 'inbred', CAB being the inbred parent, we have

$$X_{00} = X_{10} < X_{11} \tag{3.2}$$

When the fitnesses of the parental strains and their F_1 are compared, expression (3.1) corresponds to net overdominance and expression (3.2) corresponds to net dominance. As discussed earlier (sub 3.2) the apparent 'unidirectional' selection at the *b* locus suggests dominance rather than overdominance.

Hypothesis B

The strains CAW and CAB differ with respect to the degree of homozygosity of the marker chromosomes only. This possibility becomes apparent when the history of a mutant stock is considered. A mutant stock often is established starting from a single homozygous mutant individual, observed in a wild type population, and therefore two randomly sampled mutant alleles in the mutant stock have a high probability to be identical by descent. This also holds for alleles at loci which are closely linked to the marker locus. A mutant strain which differs from the wild type population with respect to the marker locus only, will only be effectively obtained by repeated backcrossing to a large wild type population. As an illustration consider the situation that a heterozygous mutant $(+/m)$ is backcrossed a number of times to the wild type population. Let further an allele on the marker chromosome be at a distance such that the recombination fraction between this locus and the marker locus is r. Then, after n backcrosses, the probability (P) that the marker chromosome still carries a copy of the allele which originally occurred on this chromosome is given by

$$P = (1 - r)^n.$$

For unlinked loci with $n=10$ this becomes

$$P = (0.5)^{10} \approx 0.00098.$$

For a locus closely linked to the marker locus with $r=0.02$, say, however

$$P = (0.98)^{10} \approx 0.82.$$

The high probabilities of being identical by descent of alleles on loci closely linked to the marker locus lead to a high degree of homozygosity at the marker chromosome in the region around the marker locus. Hypothesis *B* is justified when it is assumed that CAB has been derived from CAW or a related population without having been backcrossed often enough to ensure approximately random differences between the stocks with respect to the marker chromosome. These conditions being fulfilled, our attention can then be restricted to the marker chromosome. In terms of X_{00}, X_{10} and X_{11} the hypothesis can be formulated as

$$X_{00} = X_{10} < X_{11}$$

which corresponds to net dominance when the fitnesses of the parental strains and their hybrid are compared.

The hypotheses *A* and *B* are not mutually exclusive: CAB may carry an excess of homozygosity on the marker chromosome while both or one of the parental strains may have been 'inbred' to some extend. Under both hypotheses the genotypic background of the marker genotypes in the population experiments gradually become similar as the populations breed, resulting in a deminishing apparent selection pressure at the marker locus. The level of the ultimate (neutral) equilibrium gene frequency is determined by the input ratios of the initially differing backgrounds and the rate at which these backgrounds converge to the same ultimate general background. In order to determine to what extend hypotheses *A* and *B* are true, the following consideration is useful. *Tribolium castaneum* has 10 chromosome pairs (SOKOLOFF, 1966); with the assumption of uniformly distributed fitness loci, each chromosome contributes an amount proportional to its length to the reduction in fitness of an inbred population, provided inbreeding depression increases linearly with the number of homozygous loci (which under KING's model (ch.2) holds for a large range of genotypes). Supposing that the marker chromosomes represent one tenth of the genome, the marker chromosome contributes a proportion 0.1 to the inbreeding depression of an inbred population.

Now consider under both hypotheses the fitness of the marker genotypes in an F_2-generation obtained by random mating among F_1 individuals. Since under both hypotheses the marker genotypes in an F_2 will not differ systematically with respect to their genotypic backgrounds on the nonmarker chromosomes, the systematic differential fitness of these marker genotypes in an F_2 is due to the different make-up of their marker chromosomes only. Consequently, if the differences between the original marker genotypes (CAW, CAB and F_1) are due to 'random inbreeding', these differences will be reduced considerably when the marker genotypes of an F_2 are compared, because the proportion of the genome which contributes to these differences is reduced by at least 90%.

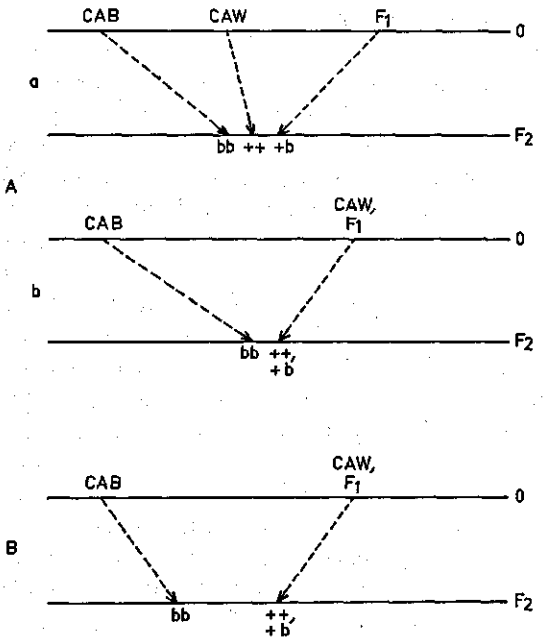


FIG. 3.2. The expected pattern of fitness differences between the original marker genotypes (0) and the marker genotypes of an F_2 under hypotheses *A* (random inbreeding of both or one of the parental strains) and *B* (only the marker chromosome bearing a fitness contrast between the two parental strains).
a: hypothesis *A* with both parental strains being inbred
b: hypothesis *A* with one parental strain (i.c. CAB) being inbred.

If however the differences between the original marker genotypes are due to a different make-up of the marker chromosomes only, these differences will among the marker genotypes in an F_2 still be appreciable (though less, as a result of crossing over at F_1 meiosis). The pattern of fitness differences among the original marker genotypes (CAW, CAB and F_1) and the marker genotypes in an F_2 under the two hypotheses is schematically shown in Fig. 3.2.

From Fig. 3.1 it is seen that the slopes of the curves show a gradual decline after the first generation interval, which means that the fitness differences between the marker genotypes in the populations only gradually diminish. Although this point may be interpreted as an indication that the marker chromosome plays an important role in the pattern of fitness differences it may in the case of a random mating population not be used as an argument against hypothesis *A*, because also under hypothesis *A* in such a population (where *all* chromosomes contribute to linkage disequilibrium), there is a gradual approach to equal fitnesses of the marker genotypes in contrast to the abrupt shifts found in an F_2.

As pointed out above, the information which is required to decide to which extend the hypotheses *A* and *B* are true (*A* and *B* are not mutually exclusive) can best be obtained from experiments in which the fitness differences among the original marker genotypes are compared with the corresponding differences among the marker genotypes of an F_2. The experiments described sub 3.4 (F_2 experiment) and sub 3.5 (Fertility experiments) have been designed for this purpose.

3.4. F_2 EXPERIMENT

In order to compare the fitness differences among the marker genotypes of an F_2 with the corresponding differences among the original marker genotypes, the following experiment was carried out. The original marker strains were crossed in both directions to obtain F_1 individuals ($+/b$). These were allowed to mate and lay eggs in order to obtain an F_2 generation. Population density in this experiment was approximately the same as in the population experiments (section 3.2).

If differential viability is an important component of differential total fitness among the F_2 marker genotypes, then this will result in a deviation from the expected 1:2:1 ratio among F_2 adults. Table 3.IV gives the observed numbers, which in each of the four replicates closely fit to the expected numbers. This means that there are no significant viability differences among the F_2 marker genotypes.

From a second set of 4 duplicate F_2 populations the pupae were sexed, virgin adults were sampled, mixed in the exact 1:2:1 ratio (100 adults of each sex per replicate), and were allowed to mate and lay eggs for a 2 day period to produce the next generation (F_3). Thus, this part of the experiment is strictly comparable with the population experiment started with marker gene frequency $q_0=0.5$.

Since no viability differences among the F_2 marker genotypes are present, no viability differences are expected in the F_3 generation neither. A shift in the gene frequency in the interval from F_2 to F_3 then can be compared with the observed shift in the population experiment with $q_0=0.5$. The difference between these shifts is a measure for the reduction in overall fitness differences between the marker genotypes of an F_2 as compared with those of the original marker genotypes. The observed numbers of F_3 adult marker genotypes and gene frequen-

TABLE 3.IV. Population counts of 4 replicate F_2 populations. *o*: observed number; *e*: expected number on the basis of the 1:2:1 ratio. *N*: sample size. The homogeneity chi-square (6 d.f.) is 1.926 + 0.027 + 0.883 + 0.22 – 1.270 = 1.587 ($0.95<P<0.99$). *q*: frequency of the *b* allele; $\bar{q}$: unweighted mean of *q*.

Genotype		+/+	+/*b*	*b*/*b*	*N*	*q*	X^2	*P*
Replicate								
I	*o*	369	676	355	1400	0.495	1.926	$0.30<P<0.50$
	e	350	700	350				
II	*o*	243	492	243	978	0.500	0.027	$0.95<P<0.99$
	e	244	489	244				
III	*o*	236	448	240	924	0.502	0.883	$0.50<P<0.70$
	e	231	462	231				
IV	*o*	207	411	204	822	0.498	0.022	$0.95<P<0.99$
	e	205.5	411	205.5				
Total	*o*	1055	2027	1042	4124	$\bar{q}=0.499$	1.270	$0.50<P<0.70$
	e	1031	2062	1031				

TABLE 3.V. Population counts in the F_3 generation. Chi-square values represent values in the goodness -of- fit test to the Weinberg-Hardy frequencies. N: sample size; q: frequency of the b allele; $\bar{q}$: unweighted mean of q.

Genotype	+/+	+/b	b/b	N	q	X_1^2	P
Replicate							
I	190	283	114	587	0.435	0.221	$0.5<P<0.7$
II	243	430	180	853	0.463	0.159	$0.5<P<0.7$
III	284	393	145	822	0.415	0.198	$0.5<P<0.7$
IV	272	416	166	854	0.438	0.094	$0.7<P<0.8$
					$\bar{q} = 0.438$		

cies are given in Table 3.V. In none of the 4 replicates there is a significant deviation from the Weinberg-Hardy frequencies. This means that gene frequency shifts due to differential fertilities among the 3 marker genotypes are not widely different for male and female parents. A small difference between gene frequencies among male and female gametes gives only a small heterozygote excess. As an example, compare the two following situations (denoting gene frequencies among male and female gametes by q_m and q_f, respectively): A: $q_m=q_f=0.44$ and B: $q_m=0.5$; $q_f=0.38$. In both cases the gene frequency or in the next generation equals 0.44. The relative frequencies of the three genotypes are 0.1936, 0.4928, 0.3136 and 0.1900, 0.5000, 0.3100 in the cases A and B, respectively. This shows that the difference between q_m and q_f in case B is hardly detectable from the genotype frequencies in the next generation.

It is further seen from Table 3.V that there is a marked change in frequency of the b allele: from 0.5 in F_2 to an (unweighted) mean of 0.438 in F_3. The corresponding change in the population experiment with $q_0=0.5$ was from 0.5 to an average of 0.34 (see Fig. 3.3).

It is clear that there are still appreciable fertility differences among the marker genotypes of an F_2; however, these are considerably smaller than the differences among the original marker genotypes.

From these experiments the following conclusions can be drawn:

1. Differential viability among marker genotypes, if any exists, is negligible with respect to the changing gene frequencies of the population experiments.
2. Comparison of the results of the F_2 experiment with those of the population experiment with $q_0=0.5$ shows that both the marker chromosome and the other chromosomes contribute to the difference in fitness between the strains CAW and CAB.

The second conclusion means that neither of the hypotheses A and B can be rejected on the basis of this experiment.

As an approach to the relative importance of the marker chromosome and non-marker chromosomes with respect to the fitness contrast between the strains CAW and CAB, the following consideration is useful. A selection coefficient (s) can be calculated for the observed changes in gene frequencies, i.e. the

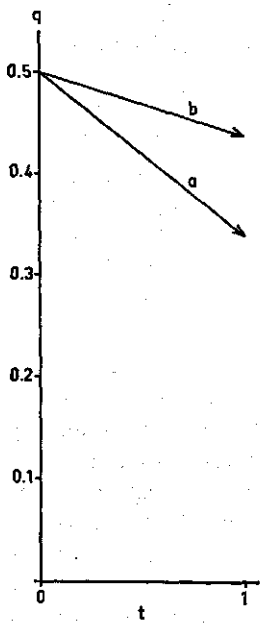


FIG. 3.3. Comparison of changes of frequency of the *b* allele in the population experiment with $q_0=0.5$(a) and the F_2 experiment (b). *q*: frequency of the *b* allele; *t*: generations.

change from 0.5 to 0.34 in the population experiment with $q_0=0.5$ and the change from 0.5 to 0.44 in the F_2 experiment. Assuming equal selection coefficients for males and females for the moment, one can calculate *s* for the cases of dominance (i.e. a relative fitness set 1, 1, 1-*s*) and no dominance (a relative fitness set 1, $1-\frac{1}{2}s$, 1-*s*). For the case of dominance the values of *s* in the population experiment and the F_2 experiment are 0.90 and 0.42, respectively. For the case of no dominance these are 0.78 and 0.38, respectively. Assuming further a linear relationship between fitness and number of homozygous loci, one sees that in either of the alternative situations (i.e. dominance and no dominance) approximately half of the reduction in fitness of the original marker genotypes with reduced fitness is due to homozygosity on the marker chromosome. This means that the marker chromosome on the one hand and the nine non-marker chromosomes on the other hand contribute equal proportions to the fitness difference between CAW and CAB. This shows the relative great importance of the marker chromosome with respect to the fitness contrast between the strains CAW and CAB.

3.5. FERTILITY EXPERIMENTS

Since the F_2 experiment described in section 3.4 does not provide information on the *pattern* of fertility differences among marker genotypes (i.e. dominance and partial dominance can not be distinguished), the following more detailed fertility experiments were carried out.

3.5.1. *Methods*

Virgin females of a given marker genotype were mated to males of different genotypes, according to a diallel cross scheme:

♀♀ \ ♂♂	*b/b*	+/*b*	+/+
b/b			
+/*b*			
+/+			

For each of the nine combinations 10 males and 10 females were allowed to mate in a 3 cm diameter mating vial during 2 days. All males and females were taken about 5 days after emergence from the pupae. After the two days of mating, single females were placed in an 'egg laying tower' (see Fig. 3.4). A tower consisted of 2 glass tubes, fitted into each other. The large (bottom) tube contained $\pm$ 2 g of medium; the bottom of the smaller (top) one was a piece of polyether cloth. Single females were put (without etherizing) into the top tube so that eggs passed through the bottom sieve into the medium. After 2 days of egg laying the inner (top) tube with the female was removed and put on top of a second bottom tube for another 2 day period of egg laying. In this way 2 two-day periods of egg laying were recorded. The tube containing the medium with eggs was kept under the usual environmental conditions. After all adults in these tubes had emerged, they were counted and classified by genotype. In this way three experiments, to be denoted by E_1, E_2 and E_3 were carried out. These experiments differed with respect to the source of the parents:

E_1. Parents collected from CAW, CAB and a hybrid population which was obtained by reciprocal mass crosses.

E_2. Parents collected from an F_2 population which was obtained by random mating (in mass) of hybrid individuals.

E_3. Parents collected from the populations started with $q_0=0.7$ in the 4th generation ($q_4 \approx 0.37$).

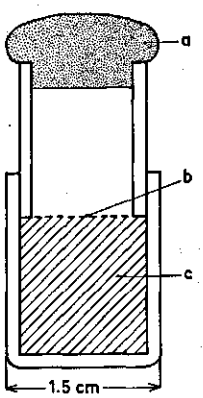


FIG. 3.4. An egg laying tower as used in the fertility experiments. a: polyether stopper; b: polyether sieve; c: medium

TABLE 3.VI. Results of fertility experiments E_1, E_2 and E_3. The body of the tables contains the mean number of adult offspring per female. The right hand columns and bottom rows contain the unweighted means of rows and columns respectively.

♂♂ / ♀♀	*bb*	*+b*	*++*	
bb	7.7	6.4	7.1	7.1
+b	12.1	25.8	11.3	16.4
++	11.6	12.4	11.9	12.0
	10.5	14.9	10.1	11.8

E_1 (foundation stocks)

♂♂ / ♀♀	*bb*	*+b*	*++*	
bb	5.5	5.8	7.4	6.2
+b	6.3	9.6	5.8	7.2
++	7.5	7.8	17.3	10.9
	6.4	7.7	10.2	8.1

E_2 (F_2 marker genotypes)

♂♂ / ♀♀	*bb*	*+b*	*++*	
bb	8.3	12.8	5.2	8.8
+b	17.0	12.4	13.0	14.1
++	19.0	16.1	16.0	17.0
	15.1	13.8	11.4	13.3

E_3 (population expt. gen. 4)

It should be noted that the conditions of mating and egg laying in these fertility experiments are of course quite different from the conditions in the population experiments; the results of the fertility tests may therefore not automatically be extrapolated to the population experiments.

3.5.2. *Results*

The results of the fertility tests are summarized in Table 3.VI, which gives the mean number of adult offspring per female during the 4 day stay in the 'towers'. Analysis of the segregation ratio's showed that in none of the single female progenies these ratio's differed significantly from the expected values, nor did the pooled numbers (pooled over 10 females). So it can be concluded that there was no differential viability within progenies, at least not related to the marker genotypes. Inspection of the estimates of variance within each of the $3\times3\times3=27$ combinations showed an approximately linear relationship between mean and variance, as is expected when the offspring number per female follows a Poisson distribution. Therefore, the square root transformation of the data was applied in order to obtain variance estimates which are independent of the mean, so that an analysis of variance could be applied.

As a first approach to the analysis of the data, the components of variance were estimated, according to the model:

$$y_{hk}=\mu+\alpha_h+e_{hk}\ (h=1,..,9;\ k=1,..,\ 10),$$

where y_{hk} is the number of offspring of the kth female in combination h,
α_h is the random effect of the hth combination,
e_{hk} is the random error in observation y_{hk}, and
μ is the overall mean,

Although this model probably is not correct (there may e.g. be a correlation between the three α_h's within a given maternal or paternal genotype), the estimates s_A^2 and s_e^2 of var (α) (the between combinations component) and var (e) (the within combinations component) respectively, in the three experiments give an impression of the overall differences within each of the experiments. The estimates are given in table 3.VII. This table shows that the overall differences,

TABLE 3.VII. Estimates of components of variance in the fertility experiments E_1, E_2 and E_3, according to the model $y_{hk} = \mu + \alpha_h + e_{hk}$ (see text). The estimates for E_3 were obtained from the data of the second egg laying period only. In E_1 and E_2 the data of both egg laying periods (i.e. the sums) were used for the estimation. s_A^2: estimate of var(α); s_e^2: estimate of var(e); $s^2 = s_A^2 + s_e^2$ (total variance).

	s_A^2	s_e^2	s^2
E_1	0.54	1.22	1.76
E_2	0.25	1.21	1.47
E_3	0.25	1.71	1.96

measured as s_A^2 are in experiment E_2 considerably smaller than in experiment E_1: s_A^2 drops from 0.54 in E_1 to 0.26 in E_2; the estimated variances of random effects within combinations (s_e^2) are equal in the two experiments. This evidently is a result of the more homogeneous backgrounds of the parents in E_2 as compared with E_1. The next step in the analysis of the data concerned the detection of interactions between paternal and maternal genotypes. Analysis of variance was applied with the model

$$y_{ijk} = \mu + A_i + B_j + AB_{ij} + e_{ijk} \; (i = 1, 2, 3; j = 1, 2, 3; k = 1,..., 10),$$

where y_{ijk} is the progeny number of the kth female of genotype i, mated to males of genotype j,
μ is the overall mean number of progeny,
A_i is the fixed effect of the ith maternal genotype,
B_j is the fixed effect of the jth paternal genotype,
AB_{ij} is the interaction of maternal genotype i and paternal genotype j, and
e_{ijk} is the random error in observation y_{ijk} (normally distributed with mean zero).

The results of this analysis of variance are summarized in Table 3.VIII. The significant interactions of parental genotypes in experiments E_1 and E_2 imply that the magnitude and even the direction of the differences (cf. Table 3.VI) of offspring production of the maternal genotypes $+/+$, $+/b$ and b/b is not independent of the genotype of the male tester. This raises the following question on the measure to be used in the estimation of fertilities. On the one hand one

TABLE 3.VIII. P values in the tests of significance of maternal and paternal effects and interactions in the fertility experiments (E_1, E_2 and E_3).

	maternal effect	paternal effect	interaction
E_1	$P<0.005$	$0.100<P$	$0.025<P<0.050$
E_2	$0.025<P<0.050$	$0.050<P<0.100$	$0.025<P<0.050$
E_3	$0.010<P<0.025$	$0.100<P$	$0.100<P$

might use the number of offspring produced when marker genotypes are mated inter se (i.e. the diagonal entries of the diallel tables in Table 3.VI); on the other hand the number of offspring produced when females of a given genotype are mated to different male genotypes is a measure of the overall fertility. The overall fertility of a female of given genotype might be measured as its expected number of offspring when tested against a random sample of male populations in which the genotypes occur in Weinberg-Hardy frequencies and among which the gene frequency is uniformly distributed. This of course is a very arbitrary way of measuring fertility but it nevertheless gives an impression of the overall behaviour. Suppose that for a given female genotype the mean numbers of offspring when mated to males of genotypes b/b, $+/b$ and $+/+$ are α, β and γ, respectively. Then the expected mean number of offspring, measured as outlined above, equals

$$\int_0^1 \{\alpha p^2 + \beta \cdot 2p(1-p) + \gamma(1-p)^2\}\, dp$$

which reduces to

$$(\alpha + \beta + \gamma)/3 \tag{3.3}$$

It thus turns out that when fertility is measured this way, the values in the right hand columns of Table 3.VI can be used. The two alternative measurements of fertility, i.e. the mean number of offspring per female when genotypes are mated inter se (i.e. the diagonal entries of Table 3.VI) and the measurement according to (3.3) are shown in Table 3.IX. Although the absolute values of the two measurements differ within the same experiment, the *pattern* of differences between experiments is the same for the two measurements: a change from overdominance in experiment E_1 to partial dominance in experiments E_2 and E_3.

Before attaching conclusions to the data of the fertility experiments, it is useful to consider the interactions of parental genotypes in some more detail. As has been mentioned earlier, there are no significant viability differences within progenies; however, viability differences between progenies of different parental combinations might be a cause of interaction. This type of interaction however

TABLE 3.IX. The two alternative measurements of female fertility in the experiments E_1, E_2 and E_3.
OD: overdominance; PD: partial dominance. The two measurements are discussed in the text. Data taken from Table 3.VI, i.e. without the square root transformation.

	Inter se	Overall		Inter se	Overall		Inter se	Overall
	E_1			E_2			E_3	
bb	7.7	7.1	*bb*	5.5	6.2	*bb*	8.3	8.8
$+b$	25.8	16.4	$+b$	9.6	7.2	$+b$	12.4	14.1
$++$	11.9	12.0	$++$	17.3	10.9	$++$	16.0	17.0
	OD	OD		PD	PD		PD	PD

probably does not play an important role, as can be seen by inspection of the first row of experiment E_1 (Table 3.VI). The b/b females produce approximately equal numbers of offspring when mated to b/b, $+/b$ or $+/+$ males. Since b/b (CAB) is the least fit genotype, viability differences between progenies would result in less adult offspring from the matings $b/b \times b/b$. In other words, if there are viability differences between progenies, we expect to find these among progenies of b/b females.

The data suggest an entirely different type of interaction, which can be described as follows: the number of offspring (Z_{ij}) produced by a pair with female genotype i and male genotype j is determined by the capacity of producing effective gametes of that parent which sets the lower limit. E.g. a female which potentially produces 20 eggs, can only produce a number of 10 fertilized eggs if its mate produces only 10 effective gametes. Denoting female and male potencies by α_i and β_j respectively, the maximum number of offspring (Z_{ij}) equals min (α_i, β_j). With this model for interaction the arrays $\{\alpha\}$ and $\{\beta\}$ for experiment E_1 become as follows:

α/β	12.1	25.8	11.3
7.7	7.7	7.7	7.7
25.8	12.1	25.8	11.3
12.4	12.1	12.4	11.3

These figures are virtually identical to the actual data for E_1 presented in Table 3.VI. Because this type of interaction under-estimates the capacity in producing offspring when mean numbers of offspring per female ore used (a few males with low capacity will decrease the mean number of offspring per female), the values of Table 3.VI were 'adjusted' using only those of the 10 values per parental combination which were above the median. These 'adjusted' data, together with the corresponding (approximate) values of α_i, β_j and $Z_{ij} = \min(\alpha_i, \beta_j)$ are shown in Table 3.X. It is seen from this table that with this model for interaction, using the 'adjusted' data there is the same pattern of differences between the experiments as with the two previously used measurements of fertility: a change from overdominance in experiment E_1 to (partial) dominance in experiments E_2 and E_3.

It should be noted that the interaction as described above, possibly plays a less important role under population conditions. In the fertility experiments adults were allowed to mate only for a limited period. Under population conditions however, mating is not limited to a short interval, and in particular a low capacity of males may thus be counterbalanced by multiple matings.

The results of the fertility experiments can be summarized as follows:

1. There is a marked decrease in overall differences between parental combinations from experiment E_1 (original marker genotypes) to experiment E_2 (marker genotypes of an F_2) (see the variances in Table 3.VII). In fact the decrease is to the level of overall differences in E_3 (population with q_0=0.7 in

TABLE 3.X. Adjusted mean numbers of offspring per female (A) and corresponding values of α_i, β_j and $z_{ij} = \min(\alpha_i, \beta_j)$ (B) for experiments E_1, E_2 and E_3.
*: this value is less reliable since it is based on only three observations. For explanation see text.

A				B				
	♂				β			
♀	*bb*	*+b*	++	α	16	30	16	
bb	12.0	9.8	12.0	12	12	12	12	
+b	17.0	30.0	15.8	30	16	30	16	E_1
++	16.6	18.0	17.6	16	16	16	16	
	♂				β			
♀	*bb*	*+b*	++	α	10	10	24	
bb	9.6	8.6	9.0	10	10	10	10	
+b	10.0	13.8	9.0	10	10	10	10	E_2
++	12.8	10.8	24.0	24	10	10	24	
	♂				β			
♀	*bb*	*+b*	++	α	14	18	18	
bb	9.0	12.8	3.3*	10	10	10	10	
+b	12.6	12.2	13.8	12	12	12	12	E_3
++	14.8	18.4	18.4	18	14	18	18	

the 4th generation). This decrease must be caused by the much greater resemblance between genotypic backgrounds of the F_2. As a whole it can be taken as an indication that the non-marker chromosomes played a role in the fertility patterns in these experiments. This is in accordance with hypothesis *A*.

2. When the fertility estimates, based on either inter se crosses or on testing against a set of different tester populations or on the model for interaction described earlier, are compared for experiments E_1 and E_2, there is a shift from apparent overdominance in E_1 to apparent partial dominance in E_2. The overdominance in E_1 is in accordance with hypothesis *A* for the case that both parental strains have been 'inbred' to some extend. The partial dominance in E_2 is in accordance with hypothesis *B*, as will be discussed below.
3. The overall differences between parental combinations in experiment E_3 are hardly significant (cf. Table 3.VIII). This is in accordance with the general hypothesis (both *A* and *B*) that the genotypic backgrounds of the marker genotypes gradually become similar as the populations breed.

From these 3 points the conclusion is that neither of the hypotheses *A* and *B* can be rejected, i.e. it is not excluded that both marker and non-marker chromosomes play a role in the fitness difference between the stocks CAW and CAB.

The partial dominance in experiment E_2 is indeed explained by hypothesis *B*, and not by *A*, as is seen from the following considerations (see Table 3.III; remember that with hypothesis *B*, $X_{00}=X_{10}<X_{11}$). Part of the + gametes, produced by a hybrid will carry units from CAB (1), as a result of recombination within the marker chromosome. For the same reason, part of the *b* gametes will carry units from CAW (0). Now it can be intuitively seen that when gametes

unite at random, the +/+ zygotes will, on an average, carry less 1/1 units than the +/*b* zygotes, which in their turn will carry less 1/1 units than *b*/*b* zygotes. This can also be seen from the following more formal treatment. Consider a single unit at a distance from the marker carrying unit such that the recombination fraction is *r*. Writing $P_{++}(11)$ for the probability that 2 randomly sampled + gametes each carry a copy of a unit from CAB, we have

$$P_{++}(11) = r^2,$$
$$P_{+b}(11) = r(1-r) \text{ and}$$
$$P_{bb}(11) = (1-r)^2,$$

which shows that for any r ($\leqslant 0.5$):

$$P_{bb}(11) \geqslant P_{+b}(11) \geqslant P_{++}(11) \tag{3.4}$$

With hypothesis *B*, i.e. $X_{00}=X_{10}<X_{11}$, the fitness of an individual is a monotonically decreasing function of its number of 1/1 unit pairs only (see also ch. 4). Thus expression (3.4) implies

$$w_{bb} < w_{+b} < w_{++}$$

(where w stands for fitness), i.e. partial (or no) dominance.

3.6. Discussion

The parental populations CAW and CAB used in the experiments may have been in approximate overall linkage equilibrium. The results of the population experiments however indicate that their genotypic backgrounds markedly differ and that the marker locus (*b*) is almost fitness neutral. When two marker strains with different genotypic backgrounds are used to initiate a new population, the latter will exhibit linkage disequilibrium. Whether or not the apparent selection at the marker locus is ascribed to a single locus, the phenomenon of decreasing apparent selection pressure can be explained as the result of the approach to linkage equilibrium. The F_2 experiment and the fertility tests were carried out to determine the relative importance of the linkage disequilibrium generated by marker chromosomes and non-marker chromosomes. The results indicated that neither of the two hypotheses (i.e. random or non random distribution of 'extra homozygosity' in CAB over chromosomes) can be rejected. With respect to the history of the strain CAB it now becomes probable that CAB has not been backcrossed often enough to the wild type population from which it has been extracted to ensure only random differences between the backgrounds of the marker strain and the wild type strain (indicated by the role of the marker chromosome). Secondly, the role of the non-marker chromosomes indicates that both parental strains have been 'inbred' to some extend which may be due to the small number of individuals the strains were established from in this laboratory (as mentioned earlier, these numbers were about 50).

It should be emphasized here again that the results of the fertility experiments

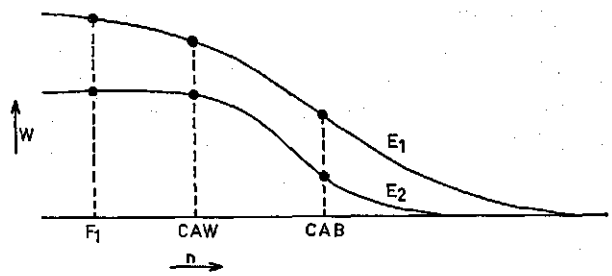


FIG. 3.5. Relation between fitness (w) and number of homozygous loci (n) in two environments, according to King's threshold model. E_1: heterogeneous environment, uncrowded (fertility experiments); E_2: homogeneous environment, crowded (population experiments).

may not automatically be extrapolated to the population experiments because of the widely different environmental conditions: in the fertility experiments 'population' density during mating and egg laying was extremely low as compared with the population experiments. Crowding is likely to impose soft selection on a population (cf. ch. 2), resulting in a reduced fitness of all genotypes, especially of the better adapted ones. Besides, the variance of non genetic factors probably has been different in the two experiments: in the fertility experiments every female was reared seperately in a vial, whilst in the F_2 experiment and in the population experiments all females of a population shared the same environment, i.e. the same vial. The effects of different environmental conditions of these experiments might have been as illustrated in Fig. 3.5. In this figure the possible relations between number of homozygous loci and fitness, according to KING's threshold model (see ch. 2) are given for the two environmental conditions. As is seen from Fig. 3.5. the overdominance in the fertility experiments (E_1) might be changed to dominance in the conditions of the population experiments (E_2). The shift of fitness differences as proposed by Fig. 3.5 is of course a guess, though not a random guess because the 'unidirectional' selection at the *b* locus observed in the population experiments also suggest dominance rather than overdominance.

The experiments described in this chapter have shown that linkage disequilibrium caused by both marker and non-marker chromosomes have played a role in the process of changing marker gene frequencies in the populations. Linkage disequilibrium generated by non-marker chromosomes however, is of secondary interest because after some generations it contributes but very little to overall linkage disequilibrium (the term overall linkage disequilibrium here applies to disequilibrium between the marker locus and any other locus).

In the next chapter a more detailed study of the process of approach to linkage equilibrium with selection is presented for the model of hypothesis *B* (an excess of homozygosity on marker chromosomes only in CAB).

4. COMPUTER SIMULATION

In this chapter a computer model for simulation of the population experiments is presented and the results are given of a number of simulation runs.

4.1. METHOD OF SIMULATION OF GENETIC SYSTEMS

The simulation of a genetic system often involves the computation of genotype and gene frequencies at many loci simultaneously. For realistic numbers of loci on a single chromosome the number of variables soon becomes of a magnitude which is far beyond the memory capacity of most computer systems. Here the features of binary computer systems become very useful and can be exploited in an elegant way. The idea is as follows. An allele at a given locus at which two alleles occur is either in 'state' A or in 'state' a. Since all information in a digital computer is stored binary, the alleles A and a can be represented by 'bits' which are in state '1' or '0'. Since a single word in a binary computer consists of many bits, each of which can be in state '1' or '0', a word can represent a chromosome region, or, in general, any set of alleles. Thus, a single variable listed in the program, which is represented by a word, can represent a number of alleles which corresponds to the number of bits in a word. In this way the amount of core required is considerably reduced, so that computers with relatively small capacity can be used. A pair of words then represents a pair of homologous chromosome regions or, in general, a genotype defined at many loci (see Table 4.I). This approach has been developed by FRASER and BURNELL (1970) to construct computer programs for the simulation of genetic systems which involve the manipulation of a set of chromosomes (or alleles). Then, by application of so called 'logic functions', which are operations in Boolean algebra, mostly being available as assembly routines of the computer system, the state of a locus (heterozygous or homozygous) can be determined and the process of recombination can be simulated. The logic operations NOT, AND, OR and EOR and their applications are shown in Table 4.II (OR is not applied here). In the C.D.C. computer system, used by the author, these bit-by-bit operations are executed at a FORTRAN statement of the form: K=AND(L,M).

TABLE 4.I. Binary representation of genotypes.

Genotype	Binary notation
.. *A B C* ..	.. 1 1 1 ..
.. *A B C* ..	.. 1 1 1 ..
.. *A b c* ..	.. 1 0 0 ..
.. *a b C* ..	.. 0 0 1 ..
.. *A B c* ..	.. 1 1 0 ..
.. *a B C* ..	.. 0 1 1 ..

From Table 4.II it is seen that:

1. the symbols 1 appearing in AND (K,L) represent the loci at which the individual is homozygous dominant (capitals denote dominant alleles),
2. the symbols 1 appearing in EOR(K,L) represent the loci at which the individual is heterozygous,
3. the symbols 1 appearing in AND (NOT(K), NOT(L)) represent the loci at which the individual is homozygous recessive.

The number of heterozygous loci is obtained as the 'digital sum' of EOR(K,L), that is the total number of symbols 1 appearing in EOR(K,L). The simulation of recombination is shown in part *c* of Table 4.II. It is assumed that recombination occurs between the second and the third locus, as indicated. A so called 'mask' (M) is generated which contains the symbols 1 from the left end up to the point of exchange and the symbols 0 elsewhere. It is seen that the string P=AND(M,K) is a duplicate of the string K from the left end to the point of exchange. Similarly, the string Q=AND(NOT(M),L) is a duplicate of L from the point of exchange to the right end. The arithmatic sum (R) of P and Q represents the chromosome region generated by a cross over.

TABLE 4.II. a. Logic operations NOT, AND, OR and EOR (Exclusive OR).
b. Binary representation of a genotype and application of logic operations to determine the states at different loci.
c. Application of logic operations in the simulation of crossing over. For explanation see text.

a.

x	y	NOT(x)	AND(x, y)	OR(x, y)	EOR(x, y)
1	1	0	1	1	0
1	0	0	1	1	1
0	1	1	0	1	1
0	0	1	0	0	0

b. Genotype Binary representation

ABcDeF 1 1 0 1 0 1 = K NK = NOT(K) = 0 0 1 0 1 0
aBcDEF 0 1 0 1 1 1 = L NL = NOT(L) = 1 0 1 0 0 0

K = 1 1 0 1 0 1
L = 0 1 0 1 1 1
AND(K, L) = 0 1 0 1 0 1

K = 1 1 0 1 0 1
L = 0 1 0 1 1 1
EOR(K, L) = 1 0 0 0 1 0

NK = 0 0 1 0 1 0
NL = 1 0 1 0 0 0
AND(NK, NL) = 0 0 1 0 0 0

c. Mask = M = 1 1 0 0 0 0

M = 1 1 0 0 0 0
K = 1 1 0 1 0 1
P = AND(M, K) = 1 1 0 0 0 0

NM = NOT(M) = 0 0 1 1 1 1

NM = 0 0 1 1 1 1
L = 0 1 0 1 1 1
Q = AND(NM, L) = 0 0 0 1 1 1
P = 1 1 0 0 0 0 +
R = 1 1 0 1 1 1
(*ABcDEF*)

Because only a finite number of individuals can be represented by their binary analogues, the size of the population to be simulated by means of the above techniques is finite. Consequently, the use of the FRASER technique is restricted to stochastic simulation.

4.2. SIMULATION OF THE POPULATION EXPERIMENTS

In chapter 3 it has been argued that the strain CAB probably carries an 'excess' of homozygous loci on the marker chromosome (hypothesis *B*). This should be understood as follows: the average frequency at which homozygous loci occur in a region near the marker locus in individuals from CAB exceeds the corresponding frequency in individuals from CAW. Or, in other words, the probability that a locus in a region near the marker locus is homozygous is in CAB greater than in CAW. The fate of such a chromosome region with an excess of homozygosity, when introduced into a population which has not such an excess of homozygosity, will now be studied by means of computer simulation. For this purpose I adopted the FRASER technique as outlined in the previous section.

As in chapter 3, the term *segment* is used for the part of the marker chromosome which carries an excess of homozygosity in CAB. The term *unit* denotes a unit of intrachromosomal recombination. So the segment under consideration consists of many units (cf. page 29).

In addition to the assumptions mentioned in section 3.3 (page 29) it will now be assumed that the probability of recombination between any pair of successive units is equal (i.e. the units are at equal map distances).

4.2.1. *Formalisation of the genetic process*

Using binary notation of genotypes, the initial populations, composed with adults from CAW, CAB and hybrid individuals are represented in Table 4.III. For convenience it has been assumed that the regions at either side of the marker locus do not behave differently, so that for a study of the behaviour of the whole segment around the marker locus our attention can be restricted to one side of the marker locus. The marker locus is not terminal or subterminal (SOKOLOFF, 1966). The units of recombination, as described in chapter 3, are denoted by the digits 0 or 1. The 'state' of a digit (0 or 1) does in my application of the FRASER technique not refer to an alternative of alleles which is present, but indicates which of the parental strains it originated from: units denoted by a 0 are from CAW, those denoted by a 1 from CAB.

With our model (see Ch. 3) we have

$$X_{00} = X_{10} < X_{11},$$

where X_{00}, X_{10} and X_{11} are the probabilities of homozygosity for any two (by assumption) allelic units from CAW, the hybrid and CAB, respectively. Denoting the relative frequencies of unit pairs 0/0, 0/1 and 1/1 of an individual by

P_{00}, P_{10} and P_{11} respectively, and the relative frequency of homozygous loci of an individual by y, we have

$$y = P_{00}X_{00} + P_{10}X_{10} + P_{11}X_{11},$$

or, since

$$X_{00} = X_{10} \text{ and } P_{00} + P_{10} = 1 - P_{11},$$
$$y = X_{00} + P_{11}(X_{11} - X_{00}) \tag{4.1}$$

In Eq. 4.1 X_{00} and X_{11} are constants; only the value of P_{11} is allowed to vary from individual to individual. Thus, assuming a relationship (φ) between fitness (w) and number of homozygous loci (y): $w=\varphi(y)$, knowledge of X_{00}, X_{11} and P_{11} suffices in calculating w:

$$w = \varphi\{X_{00} + P_{11}(X_{11} - X_{00})\} = f(P_{11}) \tag{4.2}$$

Denoting the mean value of P_{11} for the three marker genotypes by $P_{11}(++)$, $P_{11}(+b)$ and $P_{11}(bb)$ respectively, we have in the initial populations (mixtures of CAW, hybrid and CAB):

$$P_{11}(++) = 0$$
$$P_{11}(+b) = 0$$
$$P_{11}(bb) = 1$$

The process we are interested in, is the change of the variables $P_{11}(++)$, $P_{11}(+b)$ and $P_{11}(bb)$ as the populations breed. The rate at which these variables change in successive generations depends on the mean number of cross overs in the chromosome segment under consideration and the intensity of selection (i.e. the function φ) Besides, the initial composition of the population plays an important role. In the absence of selection the ultimate frequency of 1/1 unit pairs would be equal for the three marker genotypes (i.e. $P_{11}(++)=P_{11}(+b)=P_{11}(bb)$ because recombination ultimately will result in linkage equilibrium between the marker locus and any of the units under consideration, at least in an infinitely large population. Without selection, the expected ultimate value of $P_{11}(..)$ (.. stands for $++$, $+b$ or bb) equals q^2, the square of the frequency of 1 units. Selection however favours the 0 units (fitness was assumed to be a strictly decreasing function of y) and therefore the actual mean ultimate value of $P_{11}(..)$ will be less than q^2. (Actually, the limiting value of $P_{11}(..)$ is zero because every pair of allelic units is treated as a single dominant locus.)

Now the question arises what type of simulation has to be applied, deterministic or stochastic? Since we are interested in the expected values of $P_{11}(..)$ and the frequencies of the marker alleles rather than in their probability distributions, the a priori choice is a deterministic simulation. Deterministic simulation however requires the knowledge of the complete frequency distributions of the number of 0/0, 1/0 and 1/1 unit pairs for each of the three marker genotypes in every generation. This is seen as follows. Suppose that the chromosome segment under consideration consists of n units of recombination; then per marker genotype there are more than 3^n possible constitutions, because not only at each unit

the constitutions 0/0, 1/0 or 1/1 may be realized, but for multiple heterozygotes the repulsion and coupling phase should also be distinguished. The frequency distribution of these genotypes can be computed from the frequency distribution of the 2×2^n types of gametes in the previous generation (in a gamete with given marker allele at each unit the constitution 0 or 1 may be realized). This distribution in its turn is determined by the distributions of the number of 0/0, 1/0 and 1/1 unit pairs among the adults. Thus, deterministic simulation would require the manipulation of an array of more than 3×3^n variables. It is clear that realistic values of n (of the order of 100) can not be simulated deterministically; so it was decided to follow the stochastic approach. Thus a situation arises where stochastic simulation is applied not because one is interested in the stochastic effects of the process, but because stochastic simulation is the only feasible method. In this situation the advantages of the FRASER technique (i.e. reduction of the required memory capacity of the computer system) are obvious.

In the stochastic simulation a finite set of chromosome pairs is manipulated. The variables which are of interest, i.e. the mean value of $P_{11}(..)$ and the frequency of the marker allele are recorded from this sample of chromosome pairs. The mean values of these variables are then estimated from the results of a number of simulation runs. The constitution of the chromosomes during the process is recorded by the binary make-up of the 'words' which represent these chromosomes.

4.2.2. *The simulation program*

The simulation program was written in FORTRAN IV and was run on a C.D.C. 3200 computer. A flow sheet of the program is given in Fig. 4.1. A chromosome was represented by a string of 5 'words' and an additional word for identification of the marker allele. A single word contains 24 bits in the C.D.C.; a number of 20 bits (out of the 24) was used to represent the units of recombination to obtain a total of 100 bits per chromosome. An additional 6th word was used for the marker locus.

The initial set of marker chromosome pairs was generated according to the notation of Table 4.III. Chromosomes carrying the + allele were represented by a string of 6 variables, 5 of which were assigned the value $2^{20}-1$ (in binary notation this is a string of 20 1's); the 6th variable was assigned an alternative of 2 arbitrary values. The corresponding 5 variables representing a chromosome with the *b* allele were assigned the value 0.

The simulation of one generation cycle consisted of the following steps:

1. Two individuals were sampled at random from the population. This mimics random mating but does not account for dioecy since males and females were not distinguished. This discrepancy probably is of minor importance (cf. SVED, 1968).

2. From each of the two sampled parents one gamete was generated; the pair of gametes represented a zygote. It was assumed that the occurrence of cross

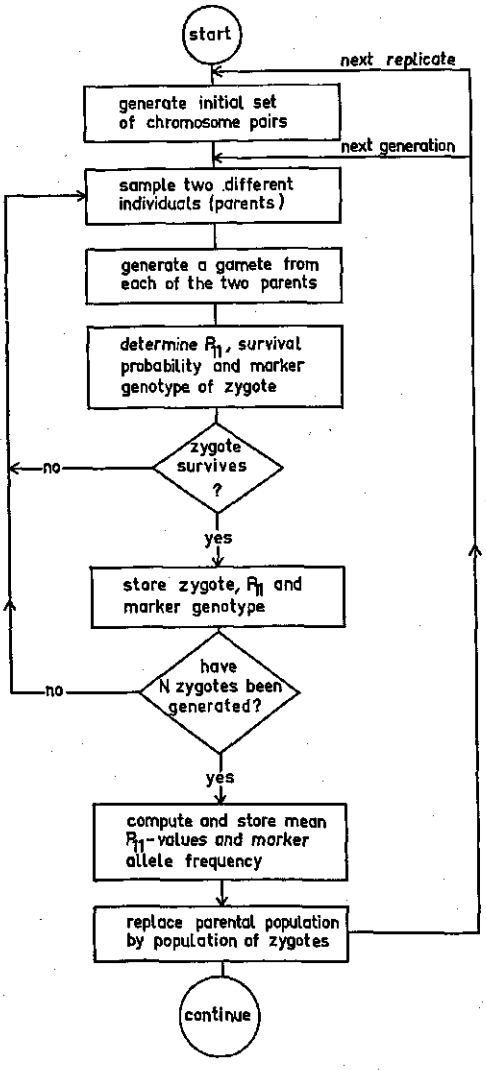


FIG. 4.1. Flow diagram of the simulation program. N: population size. For explanation see text.

TABLE 4.III. Binary representation of the genotypes in the initial populations. p: frequency of the + allele ($p+q=1$).

Genotype	Frequency
$\frac{+0000...}{+0000...}$	p^2
$\frac{+0000...}{b1111...}$	$2pq$
$\frac{b1111...}{b1111...}$	q^2

overs involving a particular chromatid is a Poisson process, so that the distance between successive cross overs is exponentially distributed with probability density function:

$$f(x) = \lambda\, \bar{e}^{\lambda x}\ (x \geq 0),$$

where λ is the intensity of the process (i.e. the probability of an event in an interval of unit length). Thus, for simulation of crossing over the sites of cross overs can be obtained by successive sampling from the exponential distribution function:

$$F(x) = 1 - \bar{e}^{\lambda x}$$

In this way the sites of successive cross overs were determined until a cross over occurred outside the chromosome segment under consideration. After determination of cross over sites the corresponding masks (cf. section 4.1) were applied to a randomly chosen member of the pair of homologous chromosomes, which mimics random assortment of chromosomes.

3. The zygote was scanned for its marker genotype and its number of 1/1 unit pairs, being the 'digital sum' (see sub 4.1) of AND (K,L), where K and L are the homologous chromosomes. This digital sum was divided by 100 (the total number of units in the segment) to obtain the P_{11} value. The relation between number of homozygous loci (y) and fitness (w) was assumed to be linear (as discussed in chapter 2, according to KING's threshold model, a linear relationship will hold for a wide range of genotypes):

$$w = w_{max} - c.y\,, \tag{4.3}$$

where w_{max} is the maximum fitness and c is a constant. Since only relative fitnesses are relevant in the change of genotype frequencies over generations and because y is a linear function of P_{11} (cf. Eq. 4.1), Eq. 4.3 can arbitrarily be written as

$$w = 1 - s.P_{11}\,, \tag{4.4}$$

and w can be interpreted as the survival probability. Notice that the selective advantage of CAW over CAB is

$$1 - (w_{CAB} / w_{CAW}) = s\,.$$

A value of $s = 0.8$ was chosen in order to realize the considerable shifts in gene frequency during the first generation interval, as was indicated by the population experiments (Ch. 3).

After determination of the survival probability (w) of the zygote, it was 'tested' for survival by means of a pseudo random number (r) uniformly distributed on [0,1] : $r \leqslant w$ resulted in a surviving zygote, $r > w$ in a non-surviving one.

If surviving, the zygote, its P_{11} value and marker genotype were stored; if non-surviving a new pair of parents was sampled from the parental population and the procedure of gametogenesis and 'tossing' for survival was repeated (steps 1–3).

The above 3 steps were repeated until the required number of survivors had been generated.

4. From the stored P_{11} values and marker genotypes, the mean P_{11} value per marker genotype and frequency of the marker allele among the zygotic population was computed (and stored for later use in the computation of mean P_{11} values per marker genotype per generation from the replicate runs).

5. The parental population was replaced by the population of zygotes.

The procedure of sampling the parents of the next generation not only accounts for a random choice of the parents (random mating) but also accounts for random variation in number of offspring. In the absence of selection the number of offspring approximately is Poisson-distributed because the probability of being sampled as a parent is small with a reasonable population size.

The present author's method for determination of the sites of cross overs deviates somewhat from the method used in simulation studies on recombination processes by WILLS et al. (1970) and SVED (1969): these authors used a sample value from a pre-defined Poisson distribution for the number of cross overs in the chromosome segment under consideration and then determined the sites of the cross overs by successive sampling from a discrete uniform distribution on [1, n] where n is the total number of possible cross over sites. This method and the present author's one are essentially equivalent; the latter is somewhat less laborious because the sites of the cross overs need not to be ordered afterwards.

4.2.3. *Results of simulation and discussion*

The simulation program was run with the following initial populations representing approximate Weinberg-Hardy frequencies of the marker genotypes:

	+/+	+/*b*	*b*/*b*	*N*	q_0
I	2	16	32	50	0.8
II	12	26	12	50	0.5
III	32	16	2	50	0.2

The parameter of the Poisson-process of the occurrence of cross overs (see section 4.2.2) was set equal to 0.005. This choice is of course arbitrary. A map distance between successive units which equals 0.5 centimorgan implies that the mean number of chiasmata in which any one of the two sister chromatids is involved equals unity.

Now, the purpose of the simulation is not so much to find a set of parameters which closely fit to the results of the population experiments (in fact, only the marker chromosomes are considered in the simulation), as to make a model which essentially gives the same results and through which the results are readily understood.

Each of the initial populations, I, II and III was run for 15 generations. The data of 20 replicate runs were averaged. The results are shown in Figs. 4.2–4.5.

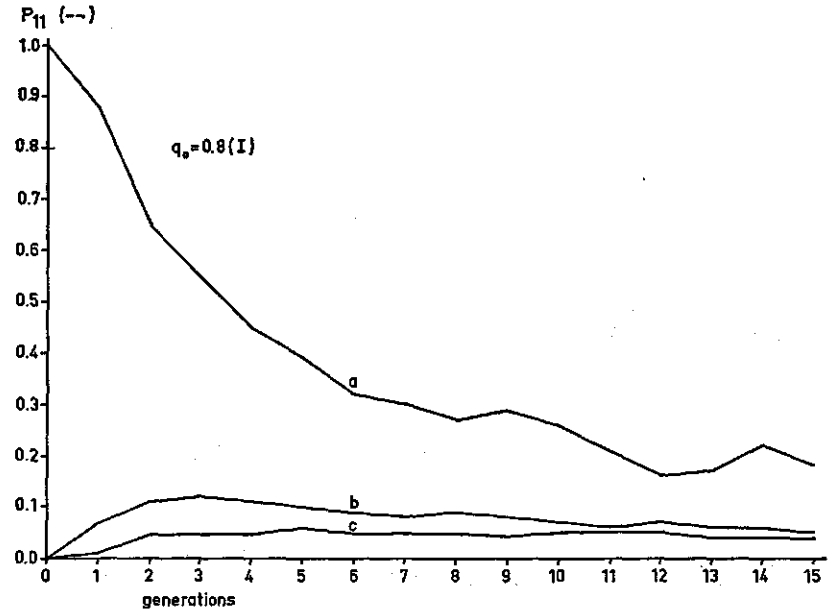


FIG. 4.2. Graphs of $P_{11}(..)$ values in the simulation. Initial gene frequency $q_0=0.8$ (population I).
a: $P_{11}(bb)$, b: $P_{11}(+b)$, c: $P_{11}(++)$.

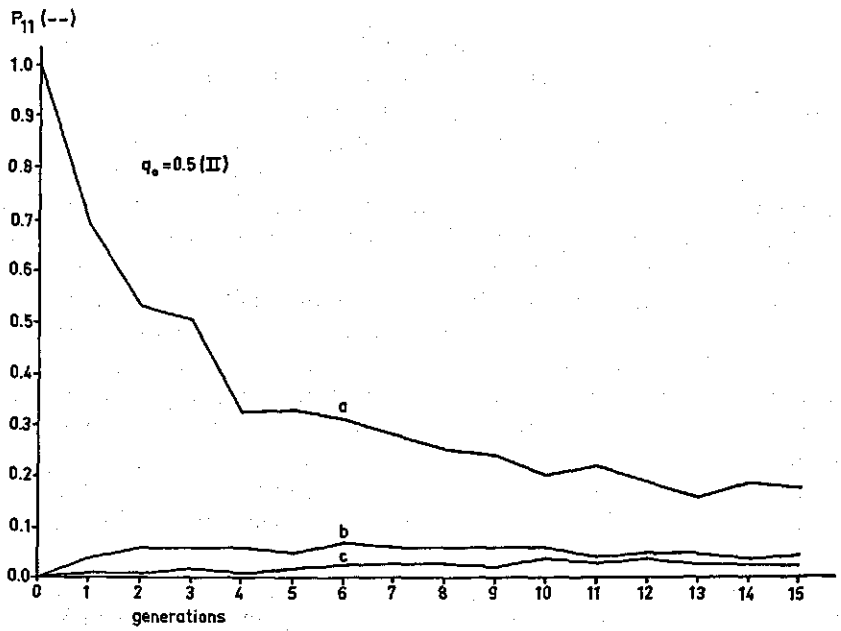


FIG. 4.3. Graphs of $P_{11}(..)$ values in the simulation. Initial gene frequency $q_0=0.5$ (population II). See further Fig. 4.2.

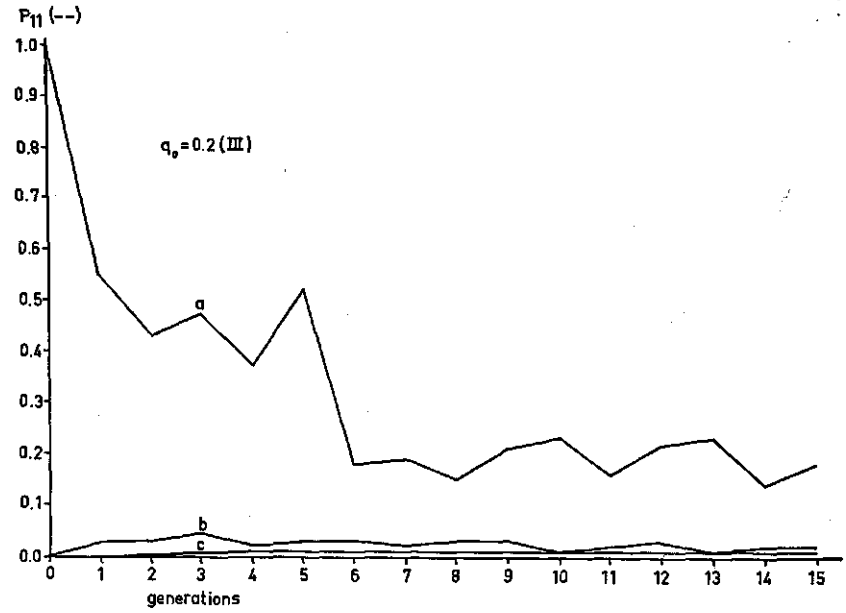


FIG. 4.4. Graphs of $P_{11}(..)$ values in the simulation. Initial gene frequency $q_0=0.2$ (population III). See further Fig. 4.2.

As is seen from these figures, there is a slight difference in the rate of decrease of the mean $P_{11}(bb)$ values between the populations during the first few generations. This is expected since with a higher value of q_0 (i.e. population I) fewer units from CAW ('0 units') will initially be taken up by b/b. This can also be seen from the $P_{11}(++)$ and $P_{11}(+b)$ graphs: these reach a maximum after a few generations, the maximum value depending on the initial composition of the population; The more CAB units there are in the initial population, the more of these are initially taken up by the marker genotypes $+/b$ and $+/+$.

It is further seen that the mean value of $P_{11}(bb)$ in population III (Fig. 4.4) fluctuates more over generations than the $P_{11}(bb)$ values in populations I and II (Figs. 4.2 and 4.3) .This is caused by the smaller sample sizes in population III from which the mean $P_{11}(bb)$ values are obtained: there are, especially during the first few generations, considerably less b/b individuals in population III than in populations I and II.

After about some 15 generations the difference between $P_{11}(bb)$ on the one hand and $P_{11}(+b)$ and $P_{11}(++)$ on the other hand is still appreciable on the scale of Figs. 4.2–4.4;
however, this difference obviously is too small to bring about a further significant change in the frequency of the b allele, as can be seen from Fig. 4.5.

In order to compare the results of the simulation as given in Fig. 4.5 with the results of the population experiments (Fig. 3.1), a correction has to be made.

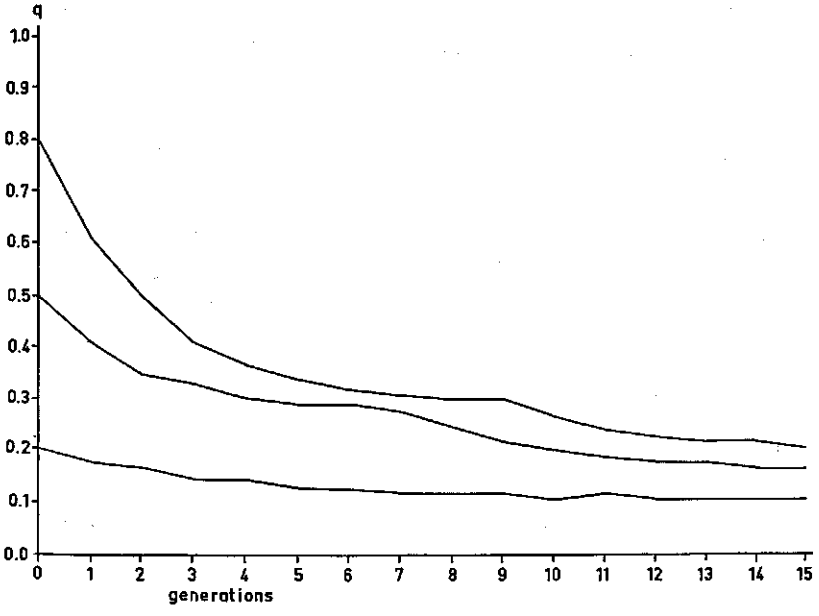


FIG. 4.5. Frequency of *b* allele (q) in the simulated population. Each curve represents the average of 20 replicate populations.

Denoting the differential fitness which changes the gene frequency from q to q' by the operator S, one can write $q'=S(q)$. When population counts are made among adults, it makes no difference when S is ascribed to either differential fertility among parents (in case of no viability differences) or to differential viability among the offspring (in case of no fertility differences). In our case S changes from one generation to the next as a result of recombination between the marker locus and fitness loci. For the change in gene frequency during the first generation interval of the population experiment we have

$$q_1 = S_0\,(q_0)\,,$$

where S_0 corresponds to differential fertility among parents (in the population experiments only differential fertility played a role, as shown in chapter 3). For the simulation runs however, where selection acts through differential viability of zygotes (cf. the description of the program, section 4.2.2), we have

$$q_1 = S_1\,(q_0),$$

where S_1 corresponds to differential viability among the first offspring generation (see Fig. 4.6). Thus, in order to make the results of the population experiments strictly comparable with the results of the simulation runs, one has to correct the initial gene frequency for the simulation runs and to shift the generation index one generation forward, such that the first change in gene frequency

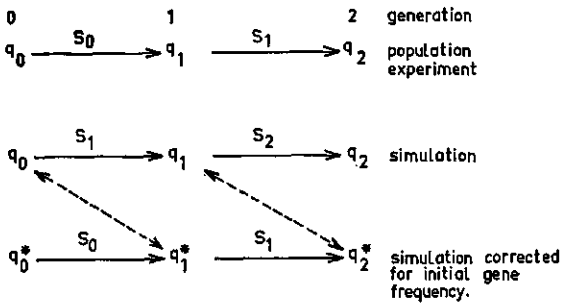


FIG. 4.6. Diagram for the changes in gene frequency in the population experiments and the simulation runs. q_0, q_1, etc.: gene frequency among adults in successive generations. S_0, S_1, etc.: selection operators which change the gene frequency from one generation to the next. Dotted arrows connect equal gene frequencies: $q_1^* = S_0(q_0^*) = q_0$. For explanation see text.

is brought about by S_0 instead of S_1. This is indicated in Fig. 4.6: we have to calculate q_0^* such that

$$q_1^* = S_0(q_0^*) = q_0.$$

Since in the simulation runs selection during the first generation interval only acts through a reduced viability of the b/b genotype, the corrected initial gene frequency (q_0^*) is readily found from

$$q_0 = q_1^* = \frac{q_0^*(1 - sq_0^*)}{1 - s(q_0^*)^2}, \tag{4.5}$$

where s is the selection coefficient of b/b. Solving Eq. 4.5 for q_0^* yields

$$q_0^* = \frac{1 \pm \sqrt{1 - 4s(1-q_1^*)\,q_1^*}}{2s(1-q_1^*)}$$

of which the smaller root satisfies. The value of s used in the simulation runs is 0.8. The corrected initial gene frequency (q_0^*) calculated from the above formula and the corrected graphs, together with the results of the population experiments are shown in Fig. 4.7. From this figure it is seen that the gene frequency in the simulation runs continues to drop over a longer period than in the population experiments, especially for the populations with the higher initial gene frequencies (i.e. $q_0 = 0.9$ in the population experiment and $q_0 = 0.94$ in the simulation). This can be understood as follows. As has been argued in chapter 3, the marker chromosomes and the joint non-marker chromosomes contribute approximately equally to the reduced fitness of CAB. In the simulation however, only the marker chromosomes were considered. This means that in the population experiments the apparent selection pressure against the b allele decreases quicker with generations than in the simulation runs, because the approach to linkage equilibrium between marker locus and any other locus is faster for loci on non-marker chromosomes than for loci on the marker chromosome. As an approach to the pattern of this difference, assume for the moment that during all generations selection acts through a reduced fitness of the b/b genotype only (i.e. complete dominance). Then, after a few generations the apparent selection coeffi-

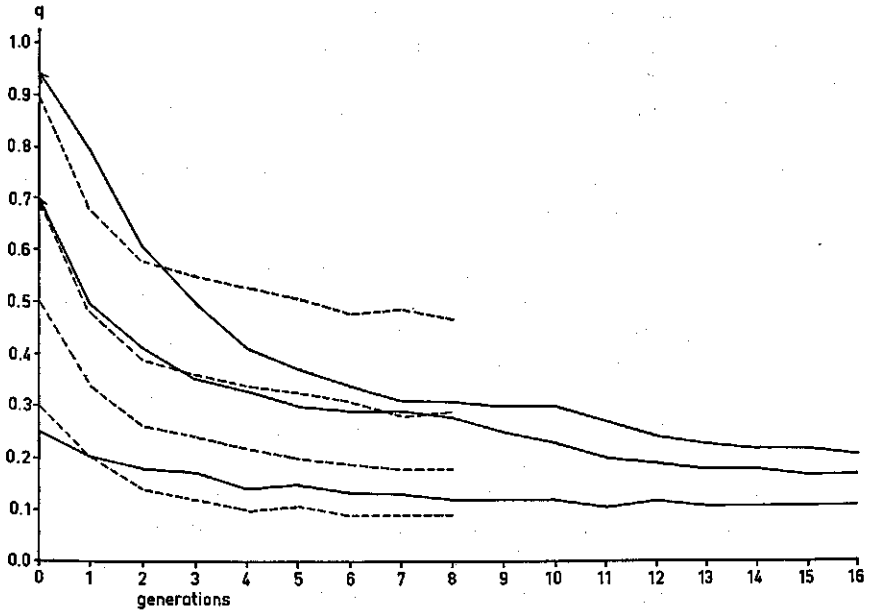


FIG. 4.7. Results of the simulation runs, adjusted for initial gene frequency (–) and of the population experiments (– – – –). q: frequency of b allele.

cients in the simulation and the population experiments are s and $\alpha.s.$, respectively, say ($\alpha<1$ because s changes slower in the simulation than in the population experiment). Starting from a given gene frequency (q), the changes in gene frequency in the simulation ($(\Delta q)_s$) and the population experiment ($(\Delta q)_p$) are

$$(\Delta q)_s = \frac{q(1-sq)}{1-sq^2} - q$$

and

$$(\Delta q)_p = \frac{q(1-\alpha sq)}{1-\alpha sq^2} - q,$$

respectively. The difference between these changes (which is a measure for the discrepancy between simulation and experiment) is

$$(\Delta q)_p - (\Delta q)_s = \frac{sq^2(1-q)\,(1-\alpha)}{(1-sq^2)\,(1-\alpha sq^2)}.$$

Upon inspection it is seen that the above quantity strictly increases when q increases from 0 to 2/3 for any value of $0<s<1$ and $0<\alpha<1$. This means that the difference between gene frequency shifts in the experiment and the simulation always increases when q increases from 0 to 2/3. This trend is indeed observed, as can be seen from Fig. 4.7 (see e.g. the changes in q from generation 3

to generation 4). Thus, the greater discrepancy (measured as the difference in Δq after some generations) between the experiment and the simulation for the populations with the higher initial gene frequencies is satisfactorily explained by the role of the non-marker chromosomes, which has been neglected in the simulation runs.

Returning now to the graphs of Figs 4.2-4.4 one sees that the limiting value of P_{11} (..) for the three marker genotypes is zero because every unit, except the unit containing the marker locus, is treated as a single dominant locus. In other words: the input of 1 units from CAB eventually is completely lost. This would mean that the complete marker chromosomes, except the neutral unit containing the *b* allele, from CAB become eventually lost. In an actual population however, this will not happen, because not every locus is a fitness locus. Nevertheless, when fitness neutral units would have been dispersed between the non-neutral units, this would not have essentially altered the results, because any other neutral locus will behave in a similar way as the *b* locus.

It is clear that, meeting a situation where 'new' genetic material from a (partially) inbred population is introduced into a non inbred breeding stock (as might be the case in a plant breeding programme) one should be aware of the possibility of losing a great deal of the newly introduced genetic material by natural selection. This phenomenon also may be important when natural populations are pooled spontaneously, as is the case with introgression.

5. SUMMARY

In chapter 2 a number of the mechanisms are discussed through which genetic polymorphisms can be maintained in natural populations: overdominance, frequency dependent selection and neutral alleles with associative overdominance. The overdominance model is emphasized because overdominance is also the basic feature of the associative overdominance model. Different theoretical relationships between number of heterozygous loci and fitness are explored, including their implications with regard to mean population fitness and selection coefficients at individual loci in an ideal population. From these, King's threshold model for multiple gene action on fitness proved to be the most satisfactory in all respects: it accomodates fairly high selection coefficients at individual loci without implying too heavy a load; it further explains different inbreeding depressions for different organisms and for different environments, as well as genotype by environment interaction. The model of associative overdominance, with the incorporation of King's threshold model for multiple gene action, has been chosen as an operational hypothesis for explanation of my experimental results (chapter 3) and as a basis for the simulation study (chapter 4).

Chapter 2 further discusses the implications of associative overdominance (which is a result of overall linkage disequilibrium in finite populations) when linkage disequilibrium is generated artificially by using a small sample to found a new population. In this situation pseudo-frequency dependent selection is expected to occur at selectively neutral loci. An experimental design is proposed which distinguishes between apparent and real frequency dependent selection.

Chapter 3 presents the experiments: individuals from two laboratory stocks of *Tribolium castaneum* HERBST, together with their F_1, were used to initiate a set of polymorphic populations (for the black locus) with different frequencies of the marker allele. These experiments, jointly taken, indicate that the black locus itself is selectively neutral under the current experimental conditions and rule out the possibility of real frequency dependent selection. There was however apparent selection against the mutant allele due to initial linkage disequilibrium. This linkage disequilibrium is described in terms of the different genotypic backgrounds of the components (wild type and mutant stock and F_1) of the founder population: in the mutant stock there is an excess of homozygosity which may be randomly distributed over the chromosomes or may be partially or wholly concentrated in a chromosome region near the marker locus. This confirms the expectation formulated in the Introduction (chapter 1). The initial linkage disequilibrium in these experiments is not so much due to small samples from the founder stocks as to the different genotypic backgrounds of the founder stocks, and, with respect to neutral loci, implies associative dominance rather than associative overdominance. The apparent decrease in selection against the *b* allele is a result of the approach to linkage equilibrium.

A comparison of the fitness differences among the original marker genotypes (wild type, mutant black and F_1) on the one hand and the marker genotypes of an F_2 population on the other hand, showed that the fitness loci closely linked to the marker locus and the joint non-linked fitness loci made approximately equal contributions to the fitness contrast between the two founder stocks (i.e. a lower fitness of the mutant stock). It also showed that, under the current experimental conditions, differential viability only played a minor role, if any, in the gene frequency changes of the *b* allele in the pooled populations.

Chapter 4 presents a computer model for (stochastic) simulation of the population experiments. This model is based on the hypothesis of overdominance at the chromosomal level and on the assumption that only the marker chromosome contributes to the fitness difference between the founder stocks. For this purpose FRASER's technique of binary representation of genotypes was adopted.

After correction for some discrepancies between the simulation model and the experiments (in the simulation only the marker chromosome is considered and selection acts through differential viability), the results of simulation proved to be in fairly good agreement with the experimentally obtained results. The simulation model can readily be adapted to other situations, e.g. both founder stocks being 'inbred', tracking the gene frequencies at more than one neutral locus, and any arbitrary function relating the number of heterozygous loci to fitness.

The final conclusion from both the experiments and the simulation study is, that after introducing the relatively 'inbred' mutant stock into the wild type population, a great deal of the genetic material of the mutant stock is lost by natural selection. For practical breeding this implies a risk of losing part of the genetic material, as a result of natural selection, from small samples of (relatively) inbred populations which are introduced into a breeding stock. Of course, the breeder may artificially select in favour of a fitness-neutral character introduced by the 'fresh' genetic material; however, the effect of artificial selection may be greatly reduced when (in the initial generations) natural selection outweighs artificial selection, since natural selection 'acts' against the desired character through linkage disequilibrium with fitness loci.

REFERENCES

DuMouchel, W. H. and Anderson, W. W.: The analysis of selection in experimental populations. Genetics **58** (1968): 435–449.

Fraser, A. S. and Burnell, D.: Computer models in genetics. McGraw Hill, New York, 1970.

Hubby, J. L. and Lewontin, R. C.: A molecular approach to the study of genetic heterozygosity in natural populations. I. The number of alleles at different loci in *Drosophila pseudoobscura*. Genetics **54** (1966): 577–594.

Kimura, M.: Diffusion models in population genetics. Methuen's monographs on applied probability and statistics (Ed. M. S. Bartlett). Methuen and Co, London, 1964.

Kimura, M.: Evolutionary rate at the molecular level. Nature **217** (1968): 624–626.

Kimura, M.: Theoretical foundation of population genetics at the molecular level. Theor. Pop. Biol. **2** (1971): 174–208.

Kimura, M. and Crow, J. F.: The number of alleles that can be maintained in a finite population. Genetics **49** (1964): 725–738.

Kimura, M. and Ohta, T.: Protein polymorphism as a phase of molecular evolution. Nature **229** (1971): 467–469.

King, J. L.: Continuously distributed factors affecting fitness. Genetics **55** (1967): 483–492.

Kojima, K. I.: The distribution and comparison of 'genetic loads' under heterotic selection and simple frequency dependent selection in finite populations. Theor. Pop. Biol. **2** (1971): 159–173.

Kojima, K. I. and Yarbrough, K. M.: Frequency dependent selection at the esterase-6 locus in *Drosophila melanogaster*. Proc. Nat. Acad. Sci. U.S.A. **57** (1967): 645–649.

Kojima, K. I. and Tobari, Y. N.: Selective modes associated with karyotypes in *Drosophila ananassae*. II. Heterosis and frequency dependent selection. Genetics **63** (1969): 639–651.

Lerner, I. M.: The genetic basis of selection. Wiley and Sons, New York, 1958.

Maruyama, T.: Effective number of alleles in a subdivided population. Theor. Pop. Biol. **1** (1970): 273–306.

Milkman, R.: Heterosis as a major cause of heterozygosity in nature. Genetics **55** (1967): 493–495.

Ohta, T. and Kimura, M.: Development of associative overdominance through linkage disequilibrium in finite populations. Genet. Res. **16** (1970): 165–177.

Prout, T.: The estimation of fitness from genotype frequencies. Evolution **19** (1965): 546–551.

Prout, T.: The estimation of fitness from population data. Genetics **63** (1969): 949–967.

Robertson, A.: Selection for heterozygotes in small populations. Genetics **47** (1962): 1291–1300.

Robertson, S.: in Population Biology and evolution (ed. R. C. Lewontin). Syracuse Univ. Press, New York, 1968.

Sokal, R. and Karten, I: Competition among genotypes in *Tribolium castaneum* at varying densities and gene frequencies (the black locus). Genetics **49** (1964): 195–211.

Sokoloff, A.: The genetics of *Tribolium* and related species. Adv. Genet. suppl. **1**. Acad. Press., New York, 1966.

Stam, P.: Estimators for viability in experimental populations; their variances and covariances. Theor. Pop. Biol. **2** (1971): 51–59.

Sved, J. A.: The stability of linked systems of loci with a small population size. Genetics **59** (1968): 543–563.

Sved, J. A., Reed, T. E. and Bodmer, W. F.: The number of balanced polymorphisms that can be maintained in a natural population. Genetics **55** (1967): 469–481.

Sved, J. A. and Ayala, F. J.: A population cage test for heterosis in *Drosophila pseudoobscura*. Genetics **66** (1970): 97–113.

Tobari, Y. N. and Kojima, K. I.: Selective modes associated with karyotypes in *Drosophila ananassae*. I. Frequency dependent selection. Genetics **57** (1967): 179–188.

VANN, E. G.: The fate of X-ray induced chromosomal rearrangements introduced into laboratory populations of *Drosophila melanogaster*. Amer. Natur. **100** (1966): 425–449.
WALLACE, B.: Topics in population genetics. W. W. Norton and Co., New York, 1968.
WILLS, C., CRENSHAW, J. and VITALE, J.: A computer model allowing maintenance of large amounts of genetic variability in Mendelian populations. I. Assumptions and results for large populations. Genetics **64** (1969): 107–123.

De auteur, Peter Stam, werd in 1943 geboren te Ochten. Hij bezocht de Rijks H.B.S. te Tiel en studeerde van 1961 tot 1969 aan de Landbouwhogeschool (studierichting plantenveredeling). Sinds zijn afstuderen is hij werkzaam bij de afdeling Erfelijkheidsleer van de Landbouwhogeschool.