Genet. Res., Camb. (1989), **53**, *pp.* 119–128 *With* 1 *text-figure* *Printed in Great Britain*

Inbreeding depression and the maintenance of deleterious genes by mutation: model of a *Drosophila* chromosome

JOHN A. SVED AND ALAN N. WILTON*
School of Biological Sciences A12, Sydney University, NSW 2006, Australia

(*Received 16 August 1988*)

Summary

Inbreeding experiments in *Drosophila*, particularly those carried out using the 'balancer equilibration' technique, have revealed high levels of inbreeding depression. It has been estimated that non-lethal chromosomes have a fitness of 20% or less in homozygous condition compared to chromosome heterozygotes. Deleterious recessive genes are, in principle, capable of explaining such inbreeding depression. In this paper we have asked quantitatively whether the observed high levels are consistent with what is known about numbers of loci and mutation rates. We find that accepted mutation rates are easily high enough, provided that the deleterious genes are fully recessive. Partial dominance, even to the extent of 10% or less, reverses this conclusion. These calculations have been made assuming the multiplicative model. However the arguments are potentially sensitive to certain types of selective interactions, and a model which proposes quadratic gene interaction allows for higher levels of partial dominance. We also test the effect of taking into account a further constraint. Crow and Mukai have argued from estimates of the persistence of new deleterious mutations affecting viability that heterozygotes have a reduction in fitness of around 1–2% per locus, similar to the estimate for lethal genes. Application of this additional constraint would markedly reduce the range of permissible selection coefficients. However we argue that the selective disadvantages in heterozygotes of most mutations affecting fitness are unlikely to be as high as estimated for mutations affecting viability.

1. Introduction

Inbreeding depression is a phenomenon which has been noted in most, if not all, normally outbreeding diploid organisms. The corresponding phenomenon of heterosis, found in outcrossing a normally inbred organism, is also widely found. We use both terms 'inbreeding depression' and 'heterosis' throughout this paper and the accompanying paper (Wilton, Joseph & Sved, 1989), although for *Drosophila* the former term is more appropriate.

Inbreeding depression is an observation which is made at the level of the organism. For some experimental designs it is possible to associate levels of inbreeding depression with particular chromosomal genotypes. However, no direct observation can be made at the level of the gene. Indeed it has been known for a long time that either dominance or overdominance at the level of the individual gene can be consistent with the observation of inbreeding depression or heterosis at the chromosome level.

* Current address: School of Biological Sciences, Macquarie University, NSW, Australia.

A priori, the dominance hypothesis must be regarded as the more likely of the two hypotheses, since it depends on a class of genes, deleterious recessives, whose existence is well established. By contrast, there has been little evidence to suggest the widespread occurrence of overdominance (e.g. Kimura, 1983), although Lewontin (1974) has commented that the potentially small selective values involved make detection of such selection problematical.

The key difference between the two hypotheses is that the dominance hypothesis relies on mutation to replenish the supply of deleterious genes. Although Crow (1948) and Fisher (1949) argued that overall mutation rates would probably be insufficient to account for observed levels of heterosis in maize, more recent estimates of mutation rates, particularly from *Drosophila*, seem to have removed some of the strength of this argument (see e.g. Crow & Simmons, 1983). However, there appears to have been no quantitative evaluation in recent years of the dominance argument. Such an evaluation is needed, particularly because of the finding (see below) that levels of inbreeding depression may be much higher than previously

recognized. The purpose of this paper is therefore to build a model of a *Drosophila* chromosome which incorporates large numbers of loci each mutating to deleterious genes. In particular, we are interested in the prediction of inbreeding depression given by accepting various levels of mutation and selection, and the extent to which the results from inbreeding experiments therefore constrain the possible values of mutation rates and selective intensities for natural populations.

The principal experimental result upon which we rely in the paper comes from the 'balancer equilibration' (BE) technique for measuring the selective value of a chromosome homozygote (Sved & Ayala, 1970). The technique is a simple one in which a population cage is set up with descendants from just two chromosomes, one a balancer chromosome which is lethal in homozygous condition, and the other a chromosome, usually a wild-type chromosome, whose fitness in homozygous condition is to be measured. This technique, with minor modifications, has been carried out on several different species and chromosomes (see Simmons & Crow, 1977; Table 4). In all cases the results have indicated a low fitness of chromosome homozygotes. As a rough median, we have accepted a figure in this paper of 20% as the mean fitness of chromosome homozygotes. It should be mentioned that only non-lethal chromosomes are investigated by this technique, and that chromosomes causing complete sterility make a comparatively small contribution in the experiments.

2. The model

(i) *Single locus selection intensities*

We follow the usual single-locus selection terminology for a deleterious gene (cf. Crow & Kimura, 1970, p. 183), s being the selection coefficient and h the degree of dominance (Table 1, line 3). While the discussion is directed at recessive genes, we cannot assume complete recessiveness, and, indeed, the degree of dominance is shown to be perhaps the most important parameter in the discussion below. There is conflicting evidence on the likely magnitude of h for genes in natural populations (Simmons & Crow, 1977). However, the weight of evidence (e.g. Mitchell & Simmons, 1977) seems to favour the notion that genes of small deleterious effect are incompletely recessive. The bulk of this evidence comes from genes affecting viability, so that its applicability to the description of fitness in populations is still unproven. In this discussion we consider possible values of h in the range $0 \leqslant h < 0{\cdot}5$. Since the discussion is directed at deleterious genes, we will not consider negative values of h (overdominance). Similarly, although s can take any value from 1 to $-\infty$, in this discussion we restrict our attention to small positive values of s.

Table 1. *Frequencies and selective values for a single locus deleterious gene model*

Genotype	$+/+$	$+/d$	d/d
Frequency	p^2	$2pq$	q^2
Selective value	1	$1-hs$	$1-s$

(ii) *The nature of selective interactions*

Nothing can be calculated without some understanding about how deleterious genes might interact. This is especially the case if many genes are involved, as in the present discussion of genes on an entire chromosome. Unfortunately, consideration of general models of selective interactions indicates that virtually *any* expected inbreeding depression can be consistent with a given set of deleterious genes. The following simple model serves to illustrate this.

We are interested in this paper in two rather different types of populations. The first is that of a natural population in which deleterious genes are maintained at a mutation–selection balance. If deleterious genes at k loci are present at frequencies $q_1, q_2, \ldots, q_k$ then in a random-mating population the expected number of deleterious homozygous loci will be $n_A = \Sigma(q_i)^2$. The second type of population is the balancer equilibration population in which chromosomes from the natural population are made homozygous. In such populations, the number of deleterious homozygotes is $n_B = \Sigma q_i$. Clearly for low values of q_i, n_A is expected to be much lower than n_B.

Let us assume that there exists a genotype-fitness function in which the fitness of a genotype depends only on the number of homozygous loci. For example, the multiplicative, or log linear, model leads to such a function, specifically a linear function if fitness is plotted on a log scale (Fig. 1, line 1). Now let us assume instead that the fitness function is one with the same slope in the region of point n_A, i.e. in the range of genotypes such as expected in the natural population, while having a slope of zero outside of this range (Fig. 1, line 2). Clearly in this case, homozygous genotypes will perform better than expected from the single locus selective values applying in the natural population. But from experiments which consider only genotypes in the normal range there is no way of knowing which of the two models will apply.

Even more pathological models are possible, e.g. Fig. 1, line 3. In this case, the fitness of genotypes with n_B deleterious genes is greater than that of genotypes with no deleterious genes. Thus genes which are deleterious in the natural population become advantageous in the inbreeding test, and it becomes difficult to distinguish what is, and what is not, a deleterious gene. Note that this is not intended as a realistic model. However, it is one which illustrates that no unequivocal prediction for the inbreeding test can be

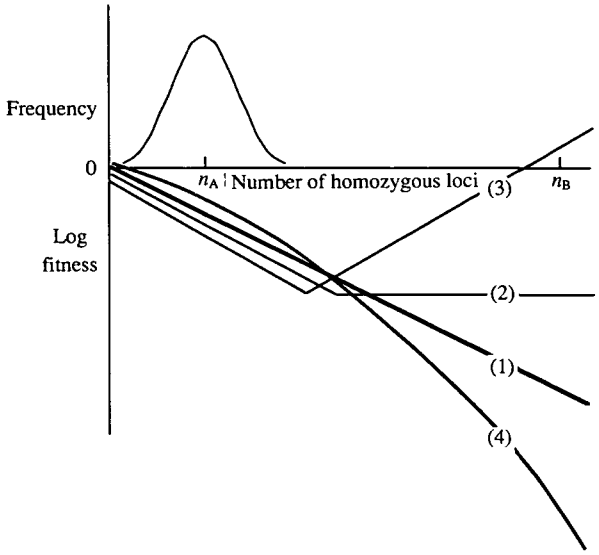

Fig. 1. Four possible relationships between genotype and fitness. The linear (1) and quadratic (4) models are shown in thick lines, and two other possible relationships in thin lines. The position of the natural population distribution is illustrated by the normal distribution around a mean n_A, and the position of a chromosome homozygote is indicated by n_B.

made from the behaviour of genes under mutation-selection balance in a natural population.

From these considerations we conclude that no calculations can be made unless it can be assumed that the interaction of deleterious genes follows some 'reasonable' rules. Unfortunately there are few studies which would allow us to test this assumption directly. We need estimates of the relationship between the fitness of whole- and part-chromosome homozygotes. Wilton *et al.* (1987) have attempted to measure the fitness of part-chromosome homozygotes in balancer equilibration cages, and obtained results consistent with a fitness of such genotypes approximately intermediate between the fitness of chromosome heterozygotes and complete chromosome homozygotes. It follows from this result that if homozygous fitness is attributable to deleterious genes, then there is approximate linearity of the interaction over at least part of the range.

We have used several models in an attempt to determine the likely role of gene interaction. All of these may be considered as falling into the class of 'reasonable' interaction models. The results from two of these models, the multiplicative and quadratic interaction models (Fig. 1, lines 1 and 4), will be presented in some detail below. We have also considered in some detail the 'additive' and 'truncation' models. However for reasons which will be described below, the results from these are not as important for considerations of the applicability of the dominance model.

(iii) *Number of loci*

An important parameter in any attempt to model the selective situation for a whole *Drosophila* chromosome is the total number of loci on the chromosome. There is at present no simple estimate of this number. Based on an analysis of mutation rates to lethal genes on the *X* chromosome, Lefevre & Watkins (1986) have questioned the validity of estimates based on equating band, or chromomere, numbers with numbers of functional units identified by saturation mutagenesis (e.g. Judd *et al.* 1972). Perhaps even more importantly, the evidence from molecular walks has shown that there are more mRNA transcripts than predicted from chromomere numbers (see Fristrom & Clegg, 1988, pp. 553–560 for an instructive summary of the current evidence).

From our point of view, the distinction between the class of genes identified by saturation mutagenesis, and those identified by the existence of transcripts, is an important one. As emphasized by Fristrom & Clegg (see also Young & Judd, 1978) the obvious explanation for the difference in counts is that the majority of genes are not capable of mutating to lethal or visibly distinct alleles. It is precisely these non-lethal genes which are of potential importance in the balancer equilibration experiments. Also, although lethal genes are specifically excluded from these experiments, we do not exclude the possibility that non-lethal mutations are making an important contribution at the loci capable of mutating to lethality. We therefore accept the overall estimate given by transcript counting, rather than that given by chromomere counting, as a relevant estimate of the number of loci at which deleterious recessives can occur.

As a confirmation of the results from counting transcripts, which has necessarily been carried out over a small number of regions, an overall estimate can be obtained of the fraction of DNA giving rise to mRNA (see Fristrom & Clegg, 1987, p. 560). This has given reasonable agreement with the transcript counting experiments, and an overall estimate of gene number of about three or four times the number of chromomeres. Accepting a figure of 2000 chromomeres per major autosome (Lefevre & Watkins, 1986) leads to an estimate of around 6000–8000 genes.

3. Calculations

(i) *Chromosome fitness under the multiplicative model*

Assuming that deleterious genes occur in natural populations at equilibrium given by the balance of mutation and selection, our aim in these calculations is to estimate the expected fitness of chromosome homozygotes. Since our principal interest is to describe the situation in BE populations, the calculation is tailored specifically to the procedure of this type of experiment. Two stages are involved: (1) predicting

https://doi.org/10.1017/S0016672300028007 Published online by Cambridge University Press

the fitness of wild-type chromosome homozygotes relative to heterozygotes containing the balancer chromosome, (2) predicting the fitness of balancer chromosome heterozygotes relative to chromosome heterozygotes in BE populations set up with a mixture of wild-type chromosomes.

Because of the large number of unknown parameters in the formulation, we have tried to simplify as much as possible by reversing the natural direction of the calculation. We start with an assumed value of the homozygous fitness, and work backwards to find what mutation value is consistent with this.

A typical chromosomally homozygous individual will be homozygous for some number of deleterious genes, which we take to be i. Chromosomally homozygous individuals cannot carry any deleterious genes in heterozygous condition. Assuming multiplicativity, the selective value of such an individual, ignoring contributions from genes on other chromosomes, will be

$$w_{++} = (1-s)^i.$$

The selective value of such a genotype is measured against heterozygotes containing the balancer heterozygote. We assume that a balancer heterozygote contains no homozygous deleterious genes. However it will contain i heterozygous genes contributed by the + chromosome, plus some contribution from the balancer chromosome. The selective value of this genotype is conveniently expressed in the form

$$w_{B+} = (1-hs)^i B,$$

where B is taken to be a constant contribution from the balancer chromosome. The balancer chromosome will typically contain gene(s) of major effect which are of a different class than those carried by the + chromosome.

The overall selective value of chromosome homozygotes compared to balancer heterozygotes is

$$w_{\text{hom}} = w_{++}/w_{B+} = \frac{1}{B}\left[\frac{1-s}{1-hs}\right]^i, \quad (1)$$

We now take into account the frequency of deleterious genes in the population from which the chromosomes are drawn. If the frequency of chromosomes having i deleterious genes is $f(i)$, then the mean selective value of chromosome homozygotes will be

$$\bar{w}_{\text{hom}} = \sum_{i=0}^{L} \frac{1}{B}\left[\frac{1-s}{1-hs}\right]^i f(i),$$

where L is the total number of loci. Assuming that the deleterious genes are in linkage equilibrium in the population from which the chromosomes are drawn, $f(i)$ is a binomial distribution with parameters L and q, where q is the population frequency of deleterious genes, assumed to be the same at all loci. Thus the above equation becomes

$$\bar{w}_{\text{hom}} = \sum_{i=0}^{L} \frac{1}{B}\left[\frac{1-s}{1-hs}\right]^i C_i^L q^i (1-q)^{L-i}.$$

Combining the terms $[1-s/1-hs]^i$ and q^i, and simplifying, gives

$$\bar{w}_{\text{hom}} = \frac{1}{B}\left[1 - sq\left(\frac{1-h}{1-hs}\right)\right]^L. \quad (2)$$

To complete the calculation, we need to give expectations for the selective values of genotypes in a BE population containing chromosome heterozygotes rather than homozygotes. The overall selective value of chromosome homozygotes must be judged relative to chromosome heterozygotes, eliminating the contribution from the balancer chromosome.

Two wild-type chromosomes in a heterozygous population will have respectively i and j deleterious genes. Then the selective value of such a genotype will be

$$w_{+/+} = (1-hs)^{i+j}.$$

Since two chromosomes are sampled together in heterozygous cages, the mean selective value of such genotypes will be

$$\bar{w}_{+/+} = \sum_{i=0}^{L}\sum_{j=0}^{L} (1-hs)^{i+j} C_i^L q^i (1-q)^{L-i} C_j^L q^j (1-q)^{L-j},$$

which simplifies to

$$\bar{w}_{+/+} = (1-hsq)^{2L}.$$

Note that the summation is carried out here before comparing against the balancer genotypes, since a single population in this case consists of a mixture of genotypes.

The selective value of balancer genotypes is, as previously,

$$w_{B/+} = (1-hs)^i B.$$

Summing over all possible wild type chromosomes in this case gives

$$\bar{w} = (1-hsq)^L B.$$

The selective value of chromosome heterozygotes compared to balancer heterozygotes is therefore

$$\bar{w}_{\text{het}} = \bar{w}_{+/+}/\bar{w}_{B/+}, = \frac{1}{B}(1-hsq)^L.$$

The overall selective value of chromosome homozygotes compared to heterozygotes is now obtained by dividing $\bar{w}_{\text{hom}}$ by $\bar{w}_{\text{het}}$. The contribution B due to the balancer chromosome cancels out. Note that this division is an artificial one for the balancer equilibration procedure, since the chromosome homozygotes and heterozygotes are never compared directly against each other. However this procedure is what

https://doi.org/10.1017/S0016672300028007 Published online by Cambridge University Press

has been used to obtain the actual estimates of homozygous fitness (e.g. Sved & Ayala, 1970). Denoting this fitness simply by $\bar{w}$, we get

$$\bar{w} = \left[\frac{1-sq\left(\frac{1-h}{1-hs}\right)}{1-hsq}\right]^L. \tag{3}$$

Equation (3) is exact under the assumptions made. The equation may be simplified, if s is sufficiently small that we can ignore terms in s^2, to

$$\bar{w} = [1-sq(1-2h)]^L, \tag{4a}$$

or, alternatively to

$$\bar{w} = [1-s(1-2h)]^{Lq}. \tag{4b}$$

A simple derivation of equation (4*b*) follows from noting that the 'average' chromosome has Lq deleterious genes, and therefore that the average chromosome homozygote has selective value $(1-s)^{Lq}$, while the average chromosome heterozygote has selective value $(1-sh)^{2Lq}$. Dividing the former by the latter, and simplifying by ignoring terms in s^2, gives (4*b*). This shows that, to a first approximation, the sampling processes can be ignored, and the selective value of the average chromosome used.

The assumption of equal selective values at all loci is made for convenience. An explicit solution may be obtained in the case of unequal selective values at individual loci. If the selective disadvantages of homozygote and heterozygote at the ith locus are s_i and $s_i h_i$ respectively, with gene frequency q_i, then it is shown in Appendix I that the equivalent equation to (3) may be derived as

$$\bar{w} = \prod_{i=1}^{L} Z_i \tag{5}$$

where

$$Z_i = \frac{1-s_i q_i\left(\frac{1-h_i}{1-h_i s_i}\right)}{1-h_i s_i q_i}.$$

In practice, conclusions cannot readily be drawn for models in which the selective values are unequal. All numerical calculations given are for the case of equal selective values.

As previously mentioned, we are reversing the natural direction of the calculation to determine what population parameters are consistent with the observed chromosome fitness. Thus we take the value of $\bar{w}$ as fixed, and equal to 0·2, and calculate the gene frequency q consistent with a range of values of s and h. Reversing equation (3) gives

$$q = \frac{(1-W)(1-hs)}{s(1-h-hW+hsW)}, \tag{6}$$

where $W = (\bar{w})^{1/L}$.

Assuming a value for $\bar{w}$ of 0·2, values of q calculated from equation (6) for various values of the selection coefficient s and degree of dominance h are tabulated in Table 2. Rather than tabulating the value of q, it is convenient to tabulate the value of Lq, which is the number of deleterious genes are chromosome. A small approximation is involved here, as in passing from equation (4*a*) to (4*b*). The values in Table 2 have been calculated assuming that $L = 10000$. The values calculated assuming $L = 2000$ differ from these by less than 1%.

The table shows that there is a strong, approximately inverse, dependence of Lq on s, and a weak dependence on h excepting in the higher range of h values. As the value of h approaches 0·5, no inbreeding depression is expected at all, which accounts for the severe rise in values of Lq in the range $h = 0{\cdot}25$–$0{\cdot}40$.

(ii) *Mutation–selection balance*

We now wish to calculate the level of mutation required to maintain deleterious genes in a natural population at the levels given in Table 2. In extending the argument in this way, we are making the important assumption that the selective intensities in natural populations are comparable to those found in the population cage.

Assuming that deleterious genes are at equilibrium as determined by a balance between mutation and selection, the required rate of mutation can be given as

Table 2. *Number of deleterious genes per chromosome (Lq) calculated from the multiplicative interaction model*

	s						
h	1/2	1/4	1/8	1/16	1/32	1/64	1/128
1/2·5	9·2	24·1	55·6	119·5	248·0	505·3	1020·0
1/4	5·3	11·7	24·6	50·3	101·8	204·8	410·7
1/8	4·0	8·3	16·9	34·0	68·3	137·0	274·3
1/16	3·6	7·2	14·6	29·3	58·7	117·6	235·3
1/32	3·4	6·8	13·7	27·4	54·9	109·8	219·7
1/64	3·3	6·6	13·3	26·6	53·1	106·3	212·6
1/128	3·3	6·5	13·1	26·1	52·3	104·6	209·2
0	3·2	6·4	12·9	25·7	51·5	103·0	206·0

https://doi.org/10.1017/S0016672300028007 Published online by Cambridge University Press

Table 3. *Mutation rates ($\times 10^6$) required to maintain deleterious genes in natural populations, assuming the multiplicative interaction model. Two possible estimates of the total number of loci are considered*

	s						
h	1/2	1/4	1/8	1/16	1/32	1/64	1/128
				10000 Loci			
1/2·5	184·0	241·6	278·7	300·6	313·8	323·8	335·0
1/4	66·3	73·3	77·1	79·4	81·1	83·3	86·8
1/8	24·9	26·0	26·6	27·1	27·8	29·0	31·2
1/16	11·2	11·4	11·6	11·9	12·4	13·4	15·3
1/32	5·3	5·4	5·6	5·8	6·2	7·1	8·9
1/64	2·6	2·7	2·8	3·0	3·4	4·3	6·0
1/128	1·3	1·4	1·5	1·7	2·1	3·0	4·6
0	0·1	0·1	0·2	0·4	0·8	1·7	3·3
				2000 Loci			
1/2·5	920·6	1212·4	1406·7	1536·1	1643·3	1775·3	1996·4
1/4	332·9	369·8	392·9	412·4	437·8	481·5	565·4
1/8	125·9	132·4	138·3	146·4	160·8	188·7	244·0
1/16	56·9	59·4	62·8	68·9	80·9	104·6	152·0
1/32	27·7	29·3	32·2	37·8	48·8	70·9	115·1
1/64	14·2	15·6	18·3	23·6	34·3	55·7	98·4
1/128	7·7	9·0	11·6	16·9	27·4	48·4	90·5
0	1·3	2·6	5·2	10·4	20·7	41·4	82·8

in equation (7), which is obtained by slight simplification from equation (6.2.6) of Crow & Kimura (1970).

$$u = q^2 s(1-2h) + qhs. \qquad (7)$$

Substituting for q from (6) gives the mutation rate required to balance particular selective values. Table 3 gives the value of u for a range of values of h and s, for two different values of L, 10000 and 2000, which span the estimated value of 6000–8000 discussed earlier. The results show that there is a comparatively weak dependence of mutation rate on selection intensity s, except for nearly recessive genes. On the other hand, there is a strong dependence on the degree of dominance. Clearly for high values of h, i.e. low dominance, unrealistically high mutation rates, as well as gene numbers (Table 2), must be postulated in order to give a mutation-selection balance which explains high levels of inbreeding depression. In the lower range of h values, there is an approximately inverse relationship between h and u until for very low values of h, selection against the recessive becomes the dominant selective force. The considerable differences between the two sets of values of Table 3, especially for low values of h, highlights the necessity for an accurate estimate of locus number L.

If we accept that mutation rates cannot be much higher than 10^{-5} (Simmons & Crow, 1977), the results from Table 3 rule out the possibility of values of h much higher than 1/16. This is true even if we accept the more conservative results corresponding to a locus number per chromosome of 10000. However the results imply very weak limits on the selection intensity s.

There are two reasons for the strong dependence of inbreeding depression on the value of h. The first, already discussed in connection with Table 2, is that heterosis occurs only if there is some dominance at the loci in question. However, Table 2 shows that this effect is only important at higher values of h. More important in the present calculation is the way in which the equilibrium frequency of the deleterious genes falls as h rises. From equation (7), there is an approximately inverse relationship between q and h.

An additional factor which may come into play at very low values of h is the size of the population. The calculations have all assumed an infinite population size, which is probably not critical for partially dominant genes. However, Robertson (1962) has emphasized that the population size needs to be extremely large before formulae calculated for the balance between mutation and selection for a recessive gene are accurate. Thus the values of u in Table 3 for very low values of h are undoubtedly underestimates for realistic population sizes.

(iii) *The rate of production of new mutations*

Crow (1979) has used estimates of (1) the rate of mutation to new deleterious genes, and (2) the population incidence, to obtain estimates of the selective disadvantage in populations. For example, for lethal genes the estimated rate of production of lethals on chromosome II (U) is 0·005 per generation, while the population incidence ($Q = Lq$) is around 0·25. Assuming that the mutation rate is u at all L loci, we can use equation (7) with $s = 1$ to estimate h. This

https://doi.org/10.1017/S0016672300028007 Published online by Cambridge University Press

gives approximately $h = (U/Q)\cdot(1-Q^2/LU)$. For any value of L greater than a few hundred, the estimate of h becomes very close to U/Q, or 1/50 in this case.

A similar argument can be made for non-lethal deleterious genes to show that the ratio of newly produced loads to equilibrium population loads should again directly estimate the selective disadvantage of heterozygotes, *hs* in this case. Crow (1979) has used estimates of Mukai *et al.* (1972) from newly accumulated viability polygenes to estimate the value of *hs* in this case as 1/30, surprisingly a higher estimate than the estimate for lethal genes. It is perhaps safer to assume an estimate for *hs* in the range 1/50–1/30.

Since similar estimates of *hs* are obtained for two different classes of genes, it seems worthwhile investigating whether a similar estimate would be consistent with the calculations of the present paper. Unfortunately we cannot test this question experimentally, since we have no estimate of the effect of mutation from the BE technique. Clearly if we assume that *hs* is equal to 1/32, the range of permissible values of h and s is greatly reduced. For low values of s, e.g. $s = 1/16$, even if we assume for conservatism that *hs* is only 1/128, i.e. 1/4 as high as the estimated value of 1/32, the lowest mutation rate corresponds to $h = 1/4$ and $L = 10000$, and is equal to $27{\cdot}1 \times 10^{-6}$, which seems unacceptably high.

(iv) *The quadratic interaction model*

The potential importance of the model of gene interactions has already been indicated. The calculations to date have assumed the multiplicative interaction model. We have attempted to give a numerical assessment of the importance of this assumption by considering three other interaction models, which may be described as the 'additive', 'truncation' and 'quadratic' models respectively. We will only be presenting results for the latter model, partly because of the complexity of the calculations with the other models. More importantly, however, for a given inbreeding depression in the BE population, the truncation model predicts a *greater* selection intensity per locus than does the multiplicative model. Thus the restrictions shown above for the multiplicative model are accentuated by the truncation model. The quadratic model predicts the opposite, thereby allowing us to see whether more realistic combinations of parameter values can be achieved with the deleterious gene model. The additive model predicts a result somewhere between the multiplicative and quadratic models, and so does not add significantly to the argument.

The quadratic model, in which log fitness falls in a curvilinear manner (Fig. 1, curve 4) derives support from the data of Mukai (1969). The rate of decrease of mean viability over a long time course followed a convex curve. The model implies a type of synergism such that the joint effect of a number of deleterious genes is greater than the product of the individual genes.

The general form of the quadratic fitness function is taken to be

$$w = e^{-(al+bl^2)}.$$

For a completely recessive model, l is simply the number of deleterious genes in homozygous condition. For a model with partial dominance, measured by the parameter h as in Table 1, l is equal to $i+hj$, where i is the number of loci homozygous for deleterious genes and j the number of loci heterozygous for deleterious genes. For the linear (multiplicative) model we set $b = 0$, and can confirm that this formulation leads to the same calculations as given above for the multiplicative model if the parameter a is equated to the selective value s.

For a pure quadratic model we set $a = 0$. The selective value s cannot be used immediately in this calculation. The quadratic model implies a particular set of fitness values in the population in terms of the parameter b. From these values we can calculate the mean, or marginal, selective value at individual loci. To do this, we focus attention on one particular locus, and calculate the mean fitness of individuals homozygous for the deleterious gene relative to the mean fitness of individuals homozygous for the wild type gene at that locus. This is equal to

$$e^{-b[L(q^2+2pqh)+1]^2}/e^{-b[L(q^2+2pqh)]^2},$$

where L is again the total number of loci on the chromosome, q is the frequency of deleterious genes and $p = 1-q$. The term $L(q^2+2pq)$ represents the average 'homozygosity', while the $+1$ term represents the additional homozygosity at the locus in question. Note that what is being calculated is the fitness of the mean individual in the population rather than the mean fitness in the population, but the calculations with the multiplicative model showed that little error is introduced by this device. The above quantity can be equated with $(1-s)$, where s is the marginal selective disadvantage of deleterious homozygotes. Simplifying the above expression then gives

$$1-s = e^{-b[2L(q^2+2pqh)+1]}. \qquad (8)$$

Equation (8) may be inverted to solve for the parameter b in terms of s, but the formulation leads to a confounding of b with the gene frequency q.

We now calculate the consequences of this model for selection in the BE population. If l is the mean number of deleterious genes in the BE cage, then the average chromosome homozygote will be homozygous for l loci and heterozygous for none, while the average chromosome heterozygote will be heterozygous for $2l$ loci and homozygous for none. As previously, the parameter l is the same as the product Lq. The mean selective value of chromosome homozygotes compared to heterozygotes in BE populations will be

$$\bar{w} = e^{-b[Lq]^2}/e^{-b[2hLq]^2}. \qquad (9)$$

https://doi.org/10.1017/S0016672300028007 Published online by Cambridge University Press

Table 4. *Mutation rates ($\times 10^6$) required to maintain deleterious genes in a population assuming a quadratic gene interaction model. Format of the table is as for Table 3*

	s						
h	1/2	1/4	1/8	1/16	1/32	1/64	1/128
				10000 Loci			
1/2·5	218·4	255·2	272·0	281·0	287·2	294·1	305·5
1/4	48·7	52·3	53·4	54·0	54·7	55·9	58·2
1/8	14·5	13·7	12·6	11·9	11·6	11·6	12·0
1/16	5·8	5·0	4·2	3·6	3·2	3·0	3·1
1/32	2·6	2·2	1·7	1·4	1·1	1·0	0·9
1/64	1·3	1·0	0·8	0·6	0·5	0·4	0·3
1/128	0·6	0·5	0·4	0·3	0·2	0·2	0·1
0	0·01	0·01	0·02	0·02	0·02	0·02	0·02
				2000 Loci			
1/2·5	1095·4	1285·5	1382·3	1452·6	1537·3	1688·4	2023·7
1/4	244·5	264·1	272·5	281·2	296·8	329·9	409·2
1/8	72·8	69·2	64·8	62·7	64·0	70·9	90·8
1/16	29·5	25·6	21·9	19·3	18·2	19·2	24·3
1/32	13·5	11·2	9·0	7·5	6·5	6·3	7·4
1/64	6·6	5·4	4·3	3·4	2·9	2·6	2·9
1/128	3·4	2·8	2·2	1·8	1·5	1·4	1·4
0	0·3	0·4	0·4	0·4	0·4	0·4	0·5

This quantity is equated to 0·2 if we accept the same figure as previously for homozygous fitness.

The combination of equations (8) and (9) now enables the two unknown parameters, b and q, to be calculated in terms of the parameters s, h and L. Then, substituting for q in equation (7), the mutation rates compatible with particular selection coefficients can be calculated (Table 4). This table shows that the conclusions regarding values of h are somewhat alleviated by comparison with the multiplicative model (Table 3). Values of h greater than 10% are not excluded for the higher estimate of L.

4. Discussion and conclusions

The results as summarized in Tables 3 and 4 show that there are combinations of parameters for which mutation rates to deleterious genes of 10^{-5} or less predict selective values as low as 20% for chromosome homozygotes. It is therefore clear that there is a range of parameter values for the dominance model which will satisfactorily explain chromosome heterosis. The magnitude of selective values at individual loci is relatively unimportant in these calculations, a finding which is in agreement with genetic load calculations (Crow & Kimura, 1977, p. 300). However when dominance is incomplete, the inbreeding depression consistent with a given mutation rate falls rapidly for increasing h. Depending on the number of loci assumed in the calculation, and the mode of interaction, the critical value of h could be as low as 1/128 or as high as 1/8.

Even the upper limit of this range, $h = 1/8$ is not high. As mentioned previously, Crow (1979) has compared the rate of production of new deleterious mutations with the incidence in populations, and has shown that the persistence in populations is much less than expected for a fully recessive gene. Crow's study considered only mutations affecting viability, presumably a sub-set of the mutations affecting fitness. Interestingly, although the study provides an estimate of the mean value of the parameter combination sh, which comes to about 1/50–1/30, it does not provide an estimate of either s or h individually. Thus if the values of s in the study were reasonably high, say 1/4–1/8, the corresponding value of h would not be inconsistent with the upper limit $h = 1/8$ mentioned above. We would argue that it is likely that only the more severe mutations are detectable by their effects on viability, and that therefore the estimate for sh, 1/50–1/30, is only applicable to mutations with moderately high values of s. Patently the value of sh cannot be as high as 1/50 for mutations for which $s < 1/50$. Thus if most mutations affecting fitness are of low selective effect, and if the value of h does not generally rise above 1/8, then observed levels of heterosis are consistent with the dominance hypothesis.

As mentioned previously, a key assumption of the argument is that selective intensities at individual loci in the population cage are comparable with those in natural populations. In fact we do not know whether the population cage fitness estimates are likely to be higher or lower than those in natural populations.

https://doi.org/10.1017/S0016672300028007 Published online by Cambridge University Press

Haymer & Hartl (1982) have raised the possibility that the 20% fitness estimate from population cages is too low. From the point of view of crowding it seems possible that the situation in the population cage is more extreme than in the wild, and therefore that the intensity of selection is overestimated in the population cage. However it has been argued by Sved (1976) that many of the selectively important activities in natural populations, e.g. finding food, shelter, mates and egg-laying sites, are eliminated or reduced to chance events in the population cage. On this argument the population cage selective intensities are more likely to be underestimates of the natural population intensities. This would accentuate the problem of accounting for the heterosis in terms of deleterious recessives.

The arguments of the paper implicitly assume that linkage equilibrium occurs in natural populations. Given that we are discussing genes at low frequency spread over the genome, such an assumption seems warranted. No assumption is necessary for the population cage, since the experiment is set up to measure the effect of all genes on a single chromosome, and recombination is deliberately suppressed.

Some mention should also be made of the assumption that all deleterious genes have equal effects on fitness. As previously mentioned, it is possible to relax this assumption, but it is then difficult to establish summary principles. Perhaps more importantly, models in which there is a mixture of gene effects are expected to give results which are intermediate between different models of equal gene effects. For example, we can consider a model in which half the genes have selective disadvantage s and half have selective disadvantage $10s$. Clearly this model will give results which are intermediate between the models in which all genes have selective value s and $10s$ respectively. For order-of-magnitude calculations such as those made in the paper, the intermediate result would add little to the argument.

Finally, it should be mentioned that although the dominance rather than overdominance model has been assumed in all calculations, from one point of view it does not matter which model is invoked. On either model, each chromosome homozygote must contain a deleterious genotype at numerous loci. The calculations of gene numbers and selective effects (Table 2) could equally well be interpreted in terms of the overdominance model. It is only the calculations which invoke mutation to explain the maintenance of these selective values which strictly demand the dominance model.

Numerous suggestions and criticisms on this manuscript from Professor James Crow are gratefully acknowledged.

References

Crow, J. F. (1948). Alternative hypotheses of hybrid vigor. *Genetics* **33**, 477–487.

Crow, J. F. (1979). Minor viability mutants in *Drosophila*. *Genetics* **92**, s165–s172.

Crow, J. F. & Kimura, M. (1970). *An Introduction to Population Genetics Theory*. New York: Harper and Row.

Crow, J. F. & Simmons, M. J. (1983). The mutation load in *Drosophila*. In *The Genetics and Biology of Drosophila*, vol. III (ed. M. Ashburner and J. N. Thompson). New York: Academic Press.

Fisher, R. A. (1949). *The Theory of Inbreeding*. Edinburgh: Oliver and Boyd.

Fristrom, J. W. & Clegg, M. T. (1988). *Principles of Genetics*, 2nd edn. New York: Chiron.

Haymer, D. S. & Hartl, D. L (1982). The experimental assessment of fitness in *Drosophila*. I. Comparative measures of competitive reproductive success. *Genetics* **102**, 455–466.

Judd, B. H., Shen, M. W. & Kaufman, T. C. (1972). The anatomy and function of a segment of the *X* chromosome of *Drosophila melanogaster*. *Genetics* **71**, 139–156.

Kimura, M. (1983). *The Neutral Theory of Molecular Evolution*. Cambridge University Press.

Lefevre, G. & Watkins, W. (1986). The question of the total gene number in *Drosophila melanogaster*. *Genetics* **113**, 869–895.

Lewontin, R. C. (1974). *The Genetic Basis of Evolutionary Change*. New York: Columbia University Press.

Mitchell, J. A. & Simmons, M. J. (1977). Fitness effects of EMS-induced mutations on the *X* chromosome of *Drosophila melanogaster*. II. Hemizygous fitness effects. *Genetics* **87**, 775–783.

Mukai, T. (1969). The genetic structure of natural populations of *Drosophila melanogaster*. VII. Synergistic interaction of spontaneous mutant polygenes controlling viability. *Genetics* **61**, 749–761.

Mukai, T., Chigusa, S. I., Mettler, L. E. & Crow, J. F. (1972). Mutation rate and dominance of genes affecting viability of polygenes in *Drosophila melanogaster*. *Genetics* **72**, 335–355.

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

Simmons, M. J. & Crow, J. F. (1977). Mutations affecting fitness in *Drosophila* populations. *Annual Review of Genetics* **12**, 289–328.

Sved, J. A. (1976). The relationship between genotype and fitness for heterotic models. In *Population Genetics and Ecology* (ed. S. Karlin and E. Nevo). New York: Academic Press.

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

Wilton, A. N., Joseph, M. G. & Sved, J. A. (1989). Can chromosomal heterosis in *Drosophila* be explained by deleterious recessive genes? Negative results from a dichromosomal population test. *Genetical Research* **53**, 129–140.

Wilton, A. N., Sved, J. A., Hu, K. & Ayala, F. J. (1987). Fitness of half chromosome homozygotes of *D. melanogaster* relative to balancer heterozygotes in population cages. *Genetics* **116**, s46.

Young, M. W. & Judd, B. H. (1978). Nonessential sequences, genes, and the polytene chromosome bands of *Drosophila melanogaster*. *Genetics* **88**, 723–742.

Appendix I. Derivation of fitness formulae for unequal selective values

We begin by deriving an equation equivalent to (2) for the fitness of a chromosome homozygote in the balancer equilibration cage. The fitness of a chromo-

some having the deleterious gene at some combination of loci $i_1, i_2, \ldots, i_k$ is equal to

$$w_{i_1, i_2, \ldots, i_k} = \frac{1}{B}\left(\frac{1-s_{i_1}}{1-s_{i_1}h_{i_1}}\right)\left(\frac{1-s_{i_2}}{1-s_{i_2}h_{i_2}}\right)\ldots\left(\frac{1-s_{i_k}}{1-s_{i_k}h_{i_k}}\right).$$

The probability of a genotype such as this is equal to

$$f_{i_1, i_2, \ldots, i_k} = q_{i_1}q_{i_2}\ldots q_{i_k}(1-q_1)(1-q_2)\ldots(1-q_L)$$

where the terms in $(1-q)$ involve all loci other than $i_1, i_2, \ldots, i_k$. This may be rewritten as

$$\left(\frac{q_{i_1}}{1-q_{i_1}}\right)\left(\frac{q_{i_1}}{1-q_{i_1}}\right)\ldots\left(\frac{q_{i_k}}{1-q_{i_k}}\right)$$
$$(1-q_1)(1-q_2)\ldots,(1-q_L),$$

where this time the terms in $(1-q)$ involve a contribution from all loci.

The overall mean fitness is equal to

$$\bar{w}_{\text{hom}} = \sum w_{i_1, i_2, \ldots, i_k} f_{i_1, i_2, \ldots, i_k},$$

where summation is over all combinations of deleterious genes. This expression involves a constant term,

$$\frac{1}{B}(1-q_1)(1-q_2)\ldots(1-q_L).$$

Each term in the summation involves the product of terms of the form

$$\left(\frac{q_i}{1-q_i}\right)\left(\frac{1-s_i}{1-s_ih_i}\right).$$

If we put

$$X_i = 1+\frac{(1-s_i)q_i}{(1-h_is_i)(1-q_i)},$$

then it may be seen by evaluating the product term $\prod_{i=1}^{L}(X_i)$ that

$$\bar{w}_{\text{hom}} = \frac{1}{B}(1-q_1)(1-q_2)\ldots(1-q_L)\prod_{i=1}^{L}(X_i),$$

X_i simplifies to

$$\frac{1}{(1-q_i)}\left[1-s_iq_i\left(\frac{1-h_i}{1-h_is_i}\right)\right].$$

Thus all terms in $(1-q)$ cancel, giving

$$\bar{w}_{\text{hom}} = \frac{1}{B}\prod_{i=1}^{L}Y_i, \tag{10}$$

where

$$Y_i = 1-s_iq_i\left(\frac{1-h_i}{1-h_is_i}\right).$$

Equation (10) is the analogue of equation (2) with variable selective values.

The derivation of equation (5) follows in an exactly analogous manner. The mean heterozygous fitness, $\bar{w}_{\text{het}}$, is equal to

$$\frac{1}{B}\prod_{i=1}^{L}(1-h_is_iq_i).$$

Dividing equation (10) by this product leads to equation (5).

https://doi.org/10.1017/S0016672300028007 Published online by Cambridge University Press