Genet Res., Camb. (1991), **57**, *pp.* 159–175 *With 6 text figures Printed in Great Britain*

The apparent selection on neutral marker loci in partially inbreeding populations

D. CHARLESWORTH
Department of Ecology and Evolution, University of Chicago, Barnes Laboratory, 5630 S. Ingleside Ave., Chicago, IL 60637

(Received 11 April 1990 and in revised form 16 July 1990)

Summary

Deterministic computer calculations were used to investigate the effects on the fitnesses of genotypes at neutral loci that are caused by associations with several linked or unlinked selected loci, in partially self fertilizing populations. Both mutation to partially recessive alleles and heterozygote advantage at the selected loci were studied. In the heterozygote advantage models, either arbitrary linkage between all loci was modelled, with a single neutral locus, or many unlinked selected and neutral loci were modelled. Large apparent overdominance could be generated in all types of model studied. As has previously been suggested, these types of effect can explain the observed associations between fitness and heterozygosity in partially inbreeding populations. There were also apparent fitness differences between the genotypes at the neutral locus among the progeny produced by selfing, especially with linkage between the neutral and selected loci. There is thus no genotype-independent fitness value for these progeny. Marker based methods for estimating the relative fitness of selfed and outcrossed progeny assume equality of these fitnesses, and will therefore be inaccurate (with in most cases a bias towards overestimating the degree of inbreeding depression) when there is linkage between the neutral marker loci and loci determining fitness.

1. Introduction

There have been many studies of allozyme variation in natural populations, and attempts have often been made to estimate the selective significance of this genetic variation (Mitton & Grant, 1984; Strauss, 1986; Houle, 1989). It has become clear that significant effects of genotypes at allozyme loci on fitness are found only in populations that have deficiencies of heterozygotes compared with the Hardy–Weinberg expectations (Zouros & Foltz, 1984; Strauss, 1986; Gaffney, 1990). One possible cause of the fitness differences at the allozyme loci is that these populations are partially inbreeding and that this produces associations between the heterozygosity at these loci and that of other loci in the genome which are the true determinants of variation in fitness. It is important to try to decide whether this explanation can account for the findings, or whether the allozyme loci themselves affect fitness.

In partially inbreeding populations there will be associations between the homozygosity of different loci, even between unlinked loci. There is because when a population includes individuals with different inbreeding coefficients, homozygosity at different loci will be correlated (Haldane, 1950; Kimura, 1958; Bennet & Binet, 1956). This phenomenon is usually referred to as 'genotypic association' or 'identity disequilibrium' defined by Weir & Cockerham (1973). Linked loci can show such associations even when all individuals have the same expected inbreeding coefficient (Weir & Cockerham, 1973), and the associations are stronger than for unlinked loci. The existence of a correlation of homozygosity between loci means that the fitnesses of genotypes at a given locus cannot be treated in isolation from the rest of the genome, but will be affected by genetic variation in fitness in the population caused by allelic variation at other loci with effects on fitness. Therefore a locus whose alleles are neutral in their effects on fitness will show apparent selection (Cockerham & Rawlings, 1967; Ohta, 1971; Ohta & Cockerham, 1974; Strobeck, 1980). In particular, homozygosity for a neutral locus will tend to be correlated with (and thus indicate) homozygosity for the selected loci, and will therefore be associated with lowered fitness. This is true if the selected locus or loci have heterozygote advantage, or if these loci have low frequency deleterious alleles maintained in the population by mutation. Each of these possibilities has been studied for the case of a

single selected locus and a linked or unlinked neutral locus. The case of heterozygote advantage was studied by Strobeck (1980), and that of mutation–selection balance by Ohta & Cockerham (1974). Evidently a single selected locus will not generate strong selection at the neutral locus, but a large effect could be expected if there are many loci that can mutate to deleterious alleles.

In the present paper I study the effects on neutral marker loci of genetic variation at several selected loci. The genetic variation in fitness studied is assumed to be due either to deleterious alleles, or to heterozygote advantage at the selected loci. These studies enable me to investigate apparent fitnesses at a neutral locus in a model that is more biologically realistic than the previously studied models in which a single locus determined fitness. I also assess the performance of the marker based procedure of Ritland (1990) for estimating inbreeding depression in natural populations. All the studies reported here are deterministic models. I am thus not concerned with linkage disequilibrium between the marker loci and the selected loci, which can be another cause of associations between genotypes at marker loci and linked selected loci (Ohta, 1971; Thomson, 1977).

2. Computer programs

(i) *Models of mutational load with unlinked loci, and of heterozygote advantage*

All the results were generated using Fortran programs written for a Macintosh computer. The programs used have been described previously for the case of heterozygote advantage at the selected loci (Charlesworth & Charlesworth, 1990), and mutation–selection balance, respectively (Charlesworth, Morgan & Charlesworth, 1990). The calculations for the mutational model are based on the method of Kondrashov (1985). Those for the case of heterozygote advantage were either done by a program that calculates the deterministic transition equations for a small number of loci with any selection coefficients and any linkage between loci, in a partially selfing population, or else assumed a set of unlinked loci with symmetrical overdominance, using the method of Ziehe & Roberds (1989) adapted by Charlesworth & Charlesworth (1990).

The only modifications needed to the programs for the purpose of the work described here were the addition of statements to calculate the apparent fitnesses of the genotypes at the neutral locus. Each neutral locus was assumed to have two alleles. For each genotype of progeny produced, the genotype at the neutral locus was found. The total production of each genotype at the neutral locus in the new zygotes was recorded, and also the frequencies after selection had occurred. The fitness of each genotype, for the whole population of zygotes produced in part by selfing and in part by outcrossing, was calculated as the ratio of the frequency after selection, but before normalisation, to the frequency among the zygotes. These calculations were also done separately for the zygotes produced by selfing and by outcrossing. I will use the notation w_s and w_x for the mean fitness of progeny produced by selfing and by outcrossing, respectively. The inbreeding depression is therefore defined by

$$\delta = 1 - \frac{w_s}{w_x}. \tag{1}$$

The fixation indices (defined as 1 − the ratio of the frequency of the neutral locus heterozygotes to their expected frequency under the Hardy–Weinberg assumptions) will be denoted by F among the adult population after selection, and F' among zygotes.

In what follows, I shall use the symbol S to denote the selfing rate of the population as a whole. The selection coefficients were assumed to be the same for all selected loci, in all the models studied. For the case of deleterious mutation, the symbol s will be used for the selection coefficient against mutant alleles in homozygotes, and the dominance coefficient will be denoted by h. The mutation rate per diploid genome is U. For this model, all the loci are assumed to be unlinked.

The numbers of loci in this model were not specified, as these depend on the strength of selection and the values of the other parameters, and were computed when the population reached equilibrium, as described previously (Charlesworth, Morgan & Charlesworth, 1990). For moderate selection ($s = 0{\cdot}2$) with the values of the parameters as in Table 1, the mean number of mutant alleles heterozygous per individual is 39 for a population with a selfing rate of 0·2, with lower values when there is more inbreeding. The numbers for the case of very strong selection were much smaller, especially when, as here, a lower mutation rate for this selection coefficient was assumed for the weaker selection. Weak selection was not studied, because in that case the numbers of mutant alleles became too high for the program to run in a reasonable time.

For the overdominance model with L linked loci, the recombination fraction between adjacent selected loci will be denoted by r. The neutral locus could be linked to the selected loci (either between them or at a location outside the group of selected loci). To study several neutral loci, I made a simple modification of Charlesworth & Charlesworth's (1990) adaptation of Ziehe & Roberds (1989) model for unlinked overdominant loci with symmetrical overdominance, to include a number of neutral loci. All loci in this model were unlinked. The selection coefficients for the case of overdominance will be denoted by t_1 and t_2.

The runs were done by allowing the population to reach equilibrium for the selected loci and storing these equilibrium genotype frequencies. Variation at

Table 1. *Results of calculations with the mutational load model with unlinked loci, showing the apparent selection coefficients at the neutral marker locus, in the population as a whole. The dominance coefficient,* h, *was equal to* 0·1 *for all the runs. The selection coefficient against homozygotes at the selected loci,* s, *and the mutation rate to deleterious alleles for the diploid genome as a whole,* U, *are shown in the table*

S	w_s	w_x	w_s/w_x	Frequency of neutral allele	Apparent selection coefficients		Apparent average relative fitness of homozygotes
					s_1	s_2	
Moderate selection, $s = 0{\cdot}2$, high mutation rate, $U = 1{\cdot}0$							
0·2	0·083	0·440	0·189	0·005	0·081	0·761	0·918
0·2	0·083	0·440	0·189	0·05	0·085	0·552	0·911
0·2	0·083	0·440	0·189	0·25	0·105	0·249	0·878
0·2	0·083	0·440	0·189	0·5	0·147	0·147	0·853
0·5	0·271	0·668	0·406	0·005	0·152	0·515	0·847
0·5	0·271	0·668	0·406	0·05	0·157	0·464	0·838
0·5	0·271	0·668	0·406	0·25	0·184	0·324	0·812
0·5	0·271	0·668	0·406	0·5	0·235	0·235	0·765
0·9	0·561	0·866	0·648	0·005	0·168	0·214	0.832
0·9	0·561	0·866	0·648	0·05	0·170	0·212	0·829
0·9	0·561	0·866	0·648	0·25	0·177	0·201	0·817
0·9	0·561	0·866	0·648	0·5	0·188	0·188	0·812
Strong selection, $s = 0{\cdot}9$, low mutation rate, $U = 0{\cdot}1$							
0·2	0·850	0·947	0·897	0·005	0·010	0·085	0·990
0·2	0·850	0·947	0·897	0·05	0·010	0·063	0·989
0·2	0·850	0·947	0·897	0·25	0·013	0·029	0·985
0·2	0·850	0·947	0·897	0·5	0·018	0·018	0·982
0·5	0·911	0·970	0·939	0·005	0·013	0·039	0·987
0·5	0·911	0·970	0·939	0·05	0·013	0·035	0·986
0·5	0·911	0·970	0·939	0·25	0·015	0·026	0·983
0·5	0·911	0·970	0·939	0·5	0·019	0·019	0·981
0·9	0·943	0·981	0·961	0·005	0·012	0·014	0·988
0·9	0·943	0·981	0·961	0·05	0·012	0·014	0·988
0·9	0·943	0·981	0·961	0·25	0·012	0·014	0·987
0·9	0·943	0·981	0·961	0·5	0·013	0·013	0·987

the neutral locus was then introduced by multiplying each adult genotype frequency by the same quantity between 0 and 1, and changing the frequency of the genotype heterozygous for the neutral allele from zero to this new value. It is also necessary to reduce by the same amount the frequency of each corresponding genotype with the original allele at the neutral locus. The fitness values were recorded at equilibrium.

The equilibrium F values were used to estimate the value of the ratio of the fitnesses of selfed and outcrossed progeny (w_s/w_x), by the method proposed by Ritland [1990, eqn. (8)]. Since the populations studied had reached equilibrium, the two methods given by Ritland (1990) are equivalent. These estimates were compared with the actual values observed in the runs. For the model with several neutral loci, the heterozygosity at the neutral loci can be found from the mean number of loci heterozygous per individual, divided by the number of neutral loci assumed in the model. This can be done for both the zygotes and the adults, to yield the F values.

(ii) *Model of mutation occurring at completely linked loci*

To provide an extreme contrast with the case where the loci are assumed to be unlinked, deterministic calculations were also done for mutations occurring at a set of completely linked loci. To deal with a set of completely linked loci on a single chromosome, one must keep records of the frequencies of gametes with different numbers of mutant loci. Let the frequency of gametes with mutations at i loci be f_i. One can calculate these frequencies, given the set of frequencies of genotypic classes characterized by three numbers, the numbers (j and k) of mutant loci on each of the two chromosomes, and the number (m) of loci that are homozygous for mutant alleles. I shall call this frequency q_{jkm}, using a similar notation to that used by Kondrashov (1985) for the case of unlinked mutations. The calculation begins with the mutation process, and then finds the values of q'_{jkm}, the frequencies of the genotypic classes after mutation, assuming as for the unlinked case that mutation increases the number of heterozygous loci (Kondrashov, 1985). To produce the progeny of selfing is simple for the case of complete linkage; $\frac{1}{4}$ of the

https://doi.org/10.1017/S0016672300029244 Published online by Cambridge University Press

progeny fall into the class j,j,j, $\frac{1}{4}$ will be k,k,k, and $\frac{1}{2}$ will be of the same genotypic class as the parent, j,k,m. As in Kondrashov's (1985) original model, I assume that homozygosity for mutations arises solely as a result of selfing. For outcrossing, I therefore first find the frequencies of the gamete classes f_i; the frequencies of the genotypic classes j,k,O can then be found by multiplying the gamete class frequencies together:

$$q''_{xjkO} = (1-S)f_j f_k.$$

As in the unlinked model, the zygotes produced by selfing or outcrossing can be subjected to selection, and the average fitnesses of the two classes of zygotes can be calculated, to yield the inbreeding depression. The apparent fitnesses were calculated as described above for the other models.

3. Apparent overdominance due to deleterious mutations

(i) *Unlinked selected loci and unlinked neutral locus*

Given some level of inbreeding depression due to the presence of mutant alleles, homozygosity for a very rare allele at the neutral locus strongly indicates inbreeding. Such homozygotes would therefore be expected to have the same fitness as that of inbred progeny, whereas the more frequent homozygote would have a fitness close to the mean fitness of the population as a whole, and the heterozygotes would have a higher fitness, between that of outcrossed progeny and the population mean. There would thus be the appearance of asymmetrical overdominance, with the heterozygotes having higher fitness than the mean of the two homozygotes (Ohta, 1971). The heterozygotes will have the same apparent fitness for any allele frequency of the neutral locus, because their frequencies both before and after selection are functions of the product of the allele frequencies, p and $1-p$. With increasing frequencies of the neutral mutation, the asymmetry in apparent fitnesses of the homozygotes would diminish, until at an allele frequency of 0·5 there would be no asymmetry. As can be seen from Table 1, these expectations are confirmed by the calculations.

With multiple selected loci, strong apparent overdominance at the neutral locus can be generated, especially when the selection coefficient at the selected locus is small. The apparent selection coefficients are often as high as 0·1 or 0·2 (Table 1). However the apparent fitness of homozygotes, averaged over both homozygous genotypes, is not as severely depressed as is that of the rare homozygote, because the rare genotype contributes only slightly. The difference associated with the strength of selection at the selected loci is not due to the assumption of a lower mutation rate for the case of strong selection, but was found even with the unrealistic assumption of the same

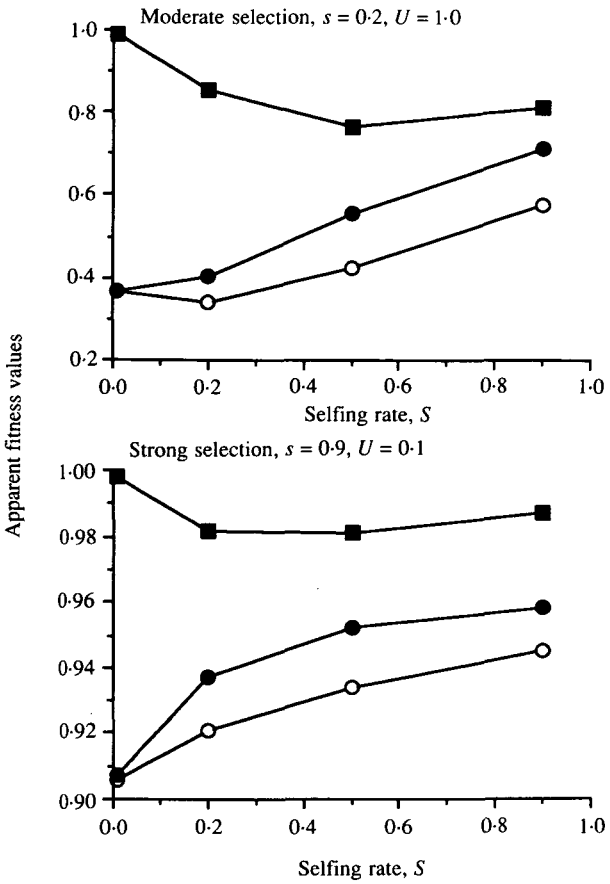

Fig. 1. Apparent fitnesses of heterozygotes and homozygotes at the neutral locus, and apparent relative fitness of homozygotes, for the mutational model when both alleles at the neutral locus are at a frequency of 0·5. ○, w_{11}; ●, w_{12}; ■, relative fitness of homozygotes.

mutation rate for that case as for the runs shown with moderate selection. The larger effect of mutation to alleles with weaker effects on fitness is expected since the apparent selection is caused by the difference in the average fitnesses of inbred and outcrossed progeny, and it has previously been found (Charlesworth *et al.* 1990) that this fitness difference is greatest for weak selection on the mutant alleles.

Fig. 1 shows the effect of variation in the selfing rate on the apparent fitnesses, for the case when the two alleles at the neutral locus are at the same frequencies. This case yields the largest apparent selection coefficient at the neutral locus (Table 1). The apparent overdominance was greatest for intermediate selfing rates. This is expected because there should be no effect due to associations of homozygosity at different loci when all progeny have the same inbreeding coefficient (Haldane, 1950), as would be the case for wholly outbreeding or inbreeding populations.

Fig. 2 shows the effect of changing the dominance coefficient. The effect is shown in terms of the apparent fitness of homozygotes at the neutral locus divided by that of heterozygotes, i.e. 1−the apparent overdominance. For simplicity, the frequencies of the two alleles were assumed to be equal, for these calculations.

https://doi.org/10.1017/S0016672300029244 Published online by Cambridge University Press

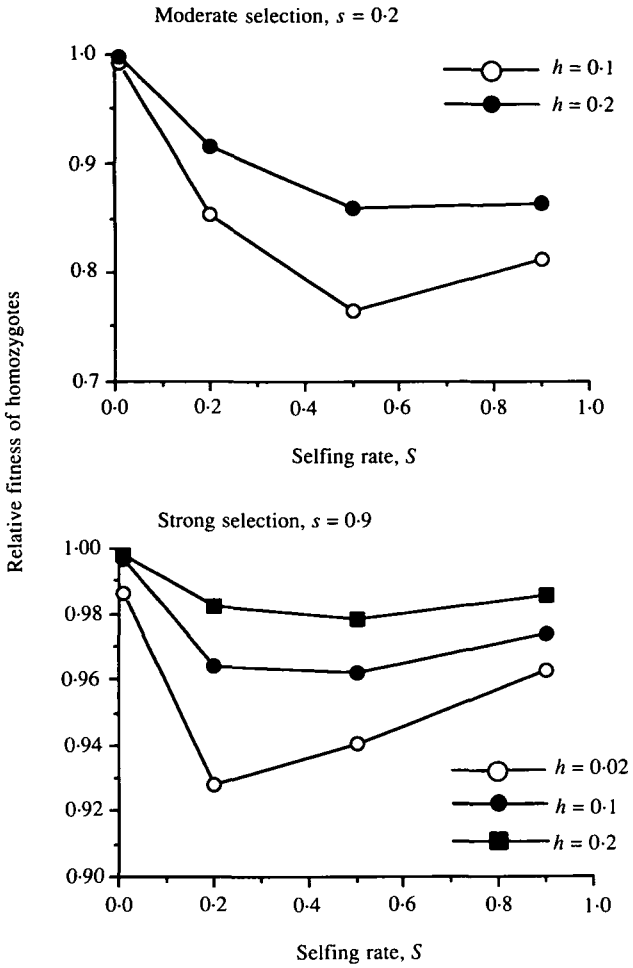

Fig. 2. Effect of the dominance coefficient for deleterious alleles in the mutational model on the apparent average relative fitness of homozygotes (= 1 − apparent selection coefficient against mutant alleles in the homozygotus state).

As expected from the well-known lowering of inbreeding depression when the dominance coefficient increases, the apparent overdominance was lower when the dominance coefficient was lower.

The apparent fitnesses among the progeny derived from outcrossing are, as expected, the same for all genotypes among the outcrossed progeny (because these progeny are produced by random mating and should show no associations between the heterozygosity of the marker locus and the selected loci, so that all genotypes at the neutral locus have the same apparent fitness as the mean fitness of outcrossed progeny). Among the progeny produced by selfing, however, there are small differences in the apparent fitness of the different genotypes. Table 2 shows some of the apparent selection coefficients within the progeny of selfing, for two selection coefficients at the selected loci, with other parameter values the same as for the runs shown in Table 1. The fitness values for the progeny produced by selfing (for all genotypes) were similar to the average fitness values for selfed progeny (see Table 1). With moderate selection (s = 0·2), the heterozygotes had the highest apparent fitness. With intense selection (s = 0·9), they had the lowest fitness value.

These differences in the apparent fitness of the marker genotypes among the selfed progeny can be understood in terms of the fact that these progeny include individuals with various different inbreeding coefficients, due to their past genealogical history of inbreeding. For example, some individuals are products of selfing of maternal plants that were produced by selfing in the previous generation, while others come from maternal plants that themselves were products of outcrossing. Thus a slight association of

Table 2. *Results of calculations with the mutational load model with unlinked loci. The table shows the apparent selection coefficients at the neutral marker locus in the progeny produced by selfing, when the two alleles are at equal frequencies, the fixation indices, and the value of* w_s/w_x *estimated by the marker-based method from the fixation indices. The dominance coefficient,* h, *was equal to 0·1 for all the runs. The selection coefficient against homozygotes at the selected loci,* s, *and the mutation rate to deleterious alleles for the diploid genome as a whole,* U, *are shown in the table*

S	Apparent selection coefficient within selfed progeny	F'	F	Estimated w_s/w_x
	Moderate selection, s = 0·2, high mutation rate, U = 1·0			
0·2	0·0190	0·1026	0·0229	0·1876
0·5	0·0394	0·2912	0·1641	0·3953
0·9	0·0340	0·7823	0·7384	0·6272
	Strong selection, s = 0·9, low mutation rate, U = 0·1			
0·2	−0·0079	0·1101	0·1024	0·9025
0·5	−0·0173	0·3303	0·3218	0·9488
0·9	−0·0151	0·1864	0·8142	0·9732

https://doi.org/10.1017/S0016672300029244 Published online by Cambridge University Press

Table 3. *Results of calculations with the mutational load model with all selected loci completely linked, showing the apparent relative fitness of homozygotes at the neutral marker locus when the allele frequencies are equal. The dominance coefficient,* h, *was equal to 0·1 for all the runs. The selection coefficient against homozygotes at the selected loci,* s, *and the mutation rate to deleterious alleles for the diploid genome as a whole,* U, *are shown in the table*

S	w_s	w_x	w_s/w_x observed	Neutral locus recombination fraction	Apparent selection coefficient against homozygotes	w_s/w_x (estimated)
Moderate selection, $s = 0{\cdot}2$, high mutation rate, $U = 1{\cdot}0$						
0·2	0·289	0·542	0·5325	0·5	0·085	0·5309
0·2	0·289	0·542	0·5325	0·3	0·100	0·4488
0·2	0·289	0·542	0·5325	0·05	0·155	0·1466
0·5	0·393	0·679	0·5780	0·5	0·165	0·5681
0·5	0·393	0·679	0·5780	0·3	0·190	0·5077
0·5	0·393	0·679	0·5780	0·05	0·296	0·2242
0·9	0·565	0·857	0·6593	0·5	0·188	0·6267
0·9	0·565	0·857	0·6593	0·3	0·208	0·5902
0·9	0·565	0·857	0·6593	0·05	0·301	0·4165
Strong selection, $s = 0{\cdot}9$, moderate mutation rate, $U = 0{\cdot}1$						
0·2	0·854	0·947	0·9025	0·5	0·017	0·9075
0·2	0·854	0·947	0·9025	0·3	0·019	0·8923
0·2	0·854	0·947	0·9025	0·05	0·097	0·8284
0·5	0·912	0·969	0·9406	0·5	0·019	0·9498
0·5	0·912	0·969	0·9406	0·3	0·022	0·9410
0·5	0·912	0·969	0·9406	0·05	0·037	0·9017
0·9	0·943	0·981	0·9612	0·5	0·013	0·9741
0·9	0·943	0·981	0·9612	0·3	0·014	0·9691
0·9	0·943	0·981	0·9612	0·05	0·028	0·9439

heterozygosity between the neutral locus and the selected loci is expected among the selfed progeny, even when the loci are unlinked as assumed in the mutational model under discussion here.

Table 2 also shows the values of the fixation indices for the neutral locus at the zygote stage before selection (F') and for the adults after selection (F). These are independent of the allele frequencies of the neutral locus. As expected, they are positive and lower after selection than before, due to the differential loss of homozygous genotypes as is consistent with their lower apparent fitness. The table shows the estimated ratios of survival of self progeny to that of outcrossed progeny, calculated from the fixation indices (Ritland, 1990). The agreement was good. For moderate or weak selection, the estimates were slightly lower than the values calculated by the computer program as described above, but for strongly selected mutations the opposite was true, as would be expected if the discrepancies are due to the inequalities in the apparent genotypic fitnesses among the progeny produced by selfing. Thus, unless the genetic load is due to very highly deleterious mutations, the inbreeding depression values are usually slightly overestimated.

(ii) *Linked selected loci, with linked or unlinked neutral locus*

With linkage, inbreeding depression is expected to be smaller than with unlinked loci, and with complete linkage it cannot exceed one-half. This is because the progeny produced by selfing behave at worst as if there was a balanced lethal system, if identity by descent for any particular chromosome is lethal. With such a system, half of the progeny of selfing would die, but assuming no identity between different gametes that unite to form outcrossed progeny, these progeny would have a viability of one.

It would be valuable to study the effect of any degree of linkage between the selected loci. This is not possible with a deterministic model, but the case of complete linkage provides an extreme contrast to the case of unlinked selected loci. Comparison of Table 3 with Table 1 shows that linkage reduces the inbreeding depression (increases the w_s/w_x ratios), as expected. Table 3 also shows the effects on the apparent selection coefficients against homozygotes at the marker locus, when the two alleles at the neutral locus were at the same frequency, i.e. for the case yielding the largest average apparent overdominance. As expected from the reduction in inbreeding depression, the apparent selection coefficients at unlinked neutral marker loci are slightly smaller than when the selected loci are unlinked. With linked markers, however, much larger apparent selection coefficients are found (Table 3).

The marker based method of Ritland (1990) assumes that the average fitness of selfed or outcrossed progeny determine the survival to the adult stage, regardless of the marker genotype whereas, as has been shown, the probability of survival from zygote to adult is not

Table 4. *Apparent fitness values of genotypes at a neutral marker locus in the presence of two loci segregating for alleles with heterozygote advantage. The population selfing rate was assumed to be 0·3. The recombination fraction between the selected loci is denoted by* r, *and when the selected loci were unlinked the neutral locus was also assumed unlinked to the selected loci. When the selected loci were assumed to be linked, the neutral locus was either linked and assumed to be located between them, or unlinked*

r	Linked to selected locus	w_s	w_x	w_s/w_x (observed)	Frequency of neutral allele	Apparent selection coefficients		Apparent relative fitness of homozygotes
						s_1	s_2	
				Strong, symmetrical selection, $t_1 = t_s = 0{\cdot}5$				
0·5	No	0·429	0·563	0·763	0·01	0·038	0·211	0·962
0·5	No	0·429	0·563	0·763	0·5	0·064	0·064	0·936
0·5	No	0·429	0·563	0·763	0·5	0·083	0·483	0·916
0·1	Yes	0·453	0·530	0·805	0·5	0·142	0·142	0·858
0·1	No	0·453	0·530	0.805	0·01	0·031	0·170	0·969
0·1	No	0·453	0·530	0·805	0·5	0·052	0·052	0·948
				Strong, asymmetrical selection, $t_1 = 0{\cdot}5$, $t_2 = 0{\cdot}7$				
0·5	No	0·375	0·501	0·748	0·01	0·040	0·222	0·960
0·5	No	0·375	0·501	0·748	0·5	0·068	0·068	0·923
0·1	Yes	0·413	0·503	0·820	0·01	0·096	0·563	0·903
0·1	Yes	0·413	0·503	0·820	0·5	0·164	0·164	0·836
0·1	No	0·413	0·503	0·820	0·01	0·028	0·153	0·972
0·1	No	0·413	0·503	0·820	0·5	0·047	0·047	0·953
				Weak, symmetrical selection, $t_1 = t_2 = 0{\cdot}2$				
0·5	No	0·719	0·810	0·888	0·01	0·018	0·100	0·982
0·5	No	0·719	0·810	0·888	0·5	0·031	0·031	0·969
0·1	Yes	0·722	0·810	0·891	0·01	0·033	0·183	0·967
0·1	Yes	0·722	0·810	0·891	0·5	0·056	0·056	0·944
0·1	No	0·722	0·810	0·891	0·01	0·018	0·097	0·982
0.1	No	0·722	0·810	0·891	0·5	0·030	0·030	0·970
				Weak, asymmetrical selection, $t_1 = 0{\cdot}2$, $t_2 = 0{\cdot}1$				
0·5	No	0·810	0·869	0·932	0·01	0·011	0·061	0·989
0·5	No	0·810	0·869	0·932	0·5	0·019	0·019	0·981
0·1	Yes	0·811	0·869	0·933	0·01	0·019	0·107	0·981
0·5	Yes	0·811	0·869	0·933	0·5	0·033	0·033	0·967
0·5	No	0·811	0·869	0·933	0·01	0·011	0·060	0·989
0·5	No	0·811	0·869	0·933	0·5	0·018	0·018	0·982

independent of the genotype at the marker genotype, among the progeny produced by selfing. The associations causing these differences are strongest with linkage between the marker locus and the selected loci, and therefore the differences in the apparent fitnesses of the genotypes at the neutral locus, among the progeny of selfing, should be greater with linkage. One might thus expect inbreeding depression estimates from the fixation indices of marker loci to be seriously biased by linkage between the neutral and selected loci. Table 3 shows that linkage between the neutral and selected loci indeed increases the downward bias in the estimates of the w_x/w_x ratio, and upward bias in inbreeding depression.

4. Apparent overdominance due to loci with heterozygote advantage

There is some possibility that loci with heterozygote advantage contribute to genetic variance in fitness in natural populations (reviewed by Bijlsma-Meeles & Bijlsma, 1988; Houle, 1990) so that it is important to consider the behaviour of such systems. Furthermore, even when there are no loci with heterozygote advantage involved in determining fitness, blocks of chromosome containing loci subject to deleterious mutations will tend to carry different mutations in different individuals and will therefore show heterozygote advantage. To study the effect on a neutral locus of linked selected loci, one can therefore use a model of heterozygote advantage.

(i) *Selected loci with linked or unlinked neutral locus*

With a single selected locus, the results were identical to those given by the formulae of Strobeck (1980). Table 4 shows some values of the apparent fitnesses of genotypes at a neutral marker locus in the presence of several other loci segregating for alleles with heterozygote advantage, with selection coefficients t_1 and t_2. The runs shown assumed a selfing rate of 0·3. As with the mutational case of the previous section, the

https://doi.org/10.1017/S0016672300029244 Published online by Cambridge University Press

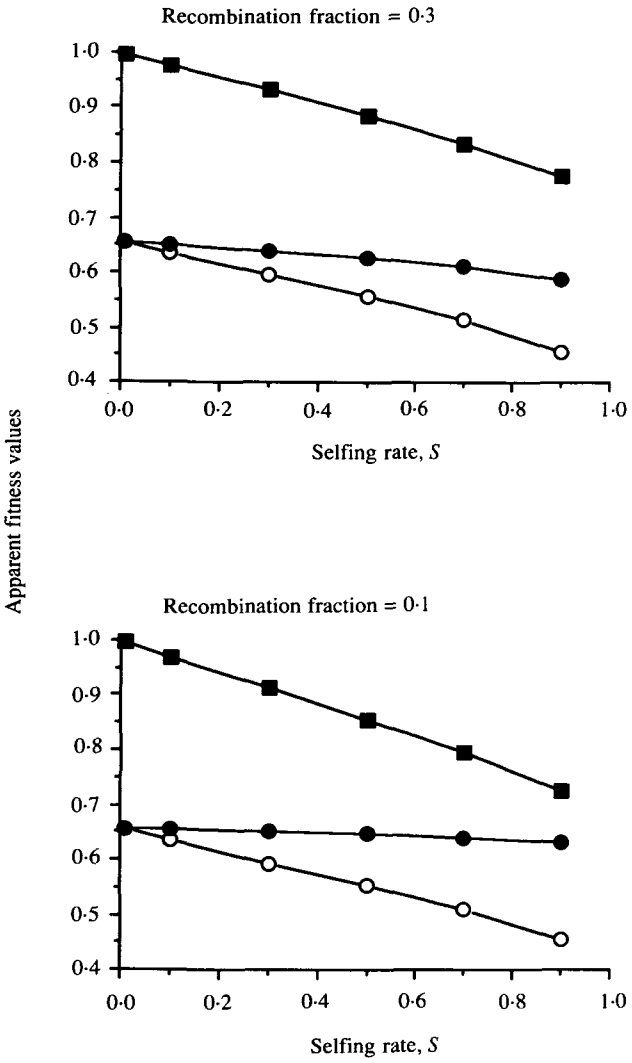

Fig. 3. Apparent fitnesses of heterozygotes and homozygotes, and apparent relative fitness of homozygotes, for the overdominance model. Both alleles at the neutral locus are at a frequency of 0·5. Symmetrical overdominance was assumed at each of four selected loci, and the selection coefficient for each locus was 0·2. The recombination fractions between the selected loci are shown in the figure, and the neutral locus was assumed to lie between the first and second selected loci. ○, w_{11}; ●, w_{12}; ■, relative fitness of homozygotes.

inbreeding depression was greatest when the selected loci were unlinked. When the neutral locus was unlinked to the selected loci, there was slight apparent overdominance at the neutral locus, similar to that observed in the previous case when the selection was against deleterious mutant alleles. The main difference in the present case is the smaller difference between the fitnesses of the selfed and the outcrossed progeny. This is expected because when variation at the selected loci is maintained by heterozygote advantage all genotypes are produced with high frequency by both types of mating. With linkage between the neutral locus and the selected loci, the apparent overdominance was stronger, and large apparent selection coefficients could be observed, in some cases approaching the selection coefficients at the selected loci, at least for the rare allele at the neutral locus. The apparent overdominance was greatest, as for the mutational model, when the alleles at the neutral locus were equally frequent. In contrast to that case, however, it was greatest for the highest selfing rates when symmetrical selection coefficients were assumed (Fig. 3). The results from runs with asymmetrical overdominance were similar to those with symmetrical selection coefficients, with the apparent overdominance at the neutral locus increasing with increasing selfing, until the selfing rate was reached at which the polymorphism could no longer be maintained.

In all cases studied, there was no linkage disequilibrium between the neutral locus and the selected loci, whatever the locations of the loci, or degree of linkage between them. Again, this is expected, since the calculations done here are deterministic, and assume infinite population sizes, and since there is no selection at the neutral locus there is no reason to expect any linkage disequilibrium to develop, even for very tight linkage (Thomson, 1977; Hedrick, Jain & Holden, 1978; Hastings, 1990).

Table 5 shows the apparent selection coefficients among the progeny of selfing, and the fixation indices, for the same parameter values as the results in Table 4. As in the case of mutational load studied above, there could be marked decreases in the fixation indices between the zygote and adult stages. This was especially pronounced with strong heterozygote advantage at the selected loci, and when the marker locus was linked to the selected loci, as expected since this is the situation when apparent selection at the neutral locus is strongest (Table 5). Table 5 also shows that there were differences in the apparent fitnesses of genotypes at the marker locus among the progeny of selfing, as in the case of mutational load discussed above. These differences were greatest when the neutral locus was linked to the selected loci. In the present case, as one might expect, the heterozygotes always had the highest apparent fitness. The fitness differences could be large, especially when the neutral locus was linked to the loci determining fitness.

Table 5 also shows the value of the ratio of the fitnesses of selfed and outcrossed progeny estimated by the method of Ritland (1990). Comparison of the estimated values with the actual w_s/w_x values in Table 4 shows that the inbreeding depression values are overestimated. As can be seen from Fig. 4, the problem is greatest when the selfing rate of the population is high, but is not negligible even for lower S values. Similar results were obtained for just one selected locus, from the analytical questions of Strobeck (1980).

(ii) *Unlinked selected and neutral loci*

To study the relationship between the apparent fitness and number of heterozygous marker loci, I used a

https://doi.org/10.1017/S0016672300029244 Published online by Cambridge University Press

Table 5. *Apparent fitness values of genotypes at a neutral marker locus in the presence of two loci segregating for alleles with heterozygote advantage. The recombination fraction between the selected loci is denoted by* r, *and when the selected loci were unlinked the neutral locus was also assumed unlinked to the selected loci. When the selection loci were assumed to be linked, the neutral locus was located between them*

r	Linked to selected locus	Frequency of neutral allele	Apparent selection coefficient within selfed progeny		Observed F'	Observed F	Estimated w_s/w_x
			Strong, symmetrical selection, $t_1 = t_2 = 0{\cdot}5$				
0·5	No	0·01	0·0182	0·0318	0·1707	0·1383	0·7490
0·5	No	0·5	0·0231	0·0231	0·1707	0·1383	0·7490
0·1	Yes	0·01	0·3077	0·5591	0·1632	0·0879	0·4494
0·1	Yes	0·5	0·3970	0·3970	0·1632	0·0879	0·4494
0·1	No	0·01	0·0104	0·0181	0·1719	0·1458	0·7968
0·1	No	0·5	0·0133	0·0133	0·1719	0·1458	0·7968
			Strong, asymmetrical selection, $t_1 = 0{\cdot}5$, $t_2 = 0{\cdot}7$				
0·5	No	0·01	0·0163	0·0283	0·1704	0·1362	0·7359
0·5	No	0·5	0·0207	0·0207	0·1704	0·1362	0·7359
0·1	Yes	0·01	0·3578	0·6586	0·1609	0·0728	0·3364
0·1	Yes	0·5	0·4637	0·4637	0·1609	0·0728	0·3364
0·1	No	0·01	0·0051	0·0084	0·1723	0·1489	0·8161
0·1	No	0·5	0·0063	0·0063	0·1723	0·1489	0·8161
			Weak, symmetrical selection, $t_1 = t_2 = 0{\cdot}2$				
0·5	No	0·01	0·0090	0·0158	0·1738	0·1586	0·8800
0·5	No	0·5	0·0092	0·0092	0·1738	0·1586	0·8800
0·1	Yes	0·01	0·1092	0·1890	0·1715	0·1436	0·7824
0·1	Yes	0·5	0·1400	0·1400	0·1715	0·1436	0·7824
0·1	No	0·01	0·0089	0·0152	0·1739	0·1591	0·8832
0·1	No	0·5	0·0113	0·0113	0·1739	0·1591	0·8832
			Weak, asymmetrical selection, $t_1 = 0{\cdot}2$, $t_2 = 0{\cdot}1$				
0·5	No	0·01	0·0052	0·0087	0·1749	0·1658	0·1658
0·5	No	0·5	0·0065	0·0065	0·1749	0·1658	0·1658
0·1	Yes	0·01	0·0618	0·1059	0·1736	0·1573	0·1573
0·1	Yes	0·5	0·0780	0·0780	0·1736	0·1573	0·1573
0·1	No	0·01	0·0050	0·0086	0·1749	0·1659	0·1659
0·1	No	0·5	0·0064	0·0064	0·1749	0·1659	0·1659

modification of the method of Ziehe & Roberds (1989) which assumes unlinked loci with symmetrical overdominance (see methods above) and is restricted to the case of unlinked neutral loci. As in the previous cases, there was an increase in fitness with heterozygosity. One would expect that this would level off as fitness approaches the fitness of progeny produced by outcrossing (see discussion below) and this is seen in Fig. 5. The figure shows the apparent fitness values for genotypes with different numbers of loci heterozygous, when each locus has a selection coefficient of 0·1, and for ten, five, or two neutral loci. The slope of the curve is steeper the more neutral loci are included, as one would expect, because when an individual is homozygous for several loci this is a stronger indication of selfing than is homozygosity for fewer loci. The slope of the regression of fitness on number of heterozygous loci can be quite large, even though the neutral loci are unlinked to the selected ones. It is of interest to examine the means and variances of the numbers of heterozygous loci in these runs, to see whether the mean values in equilibrium populations lie in the range in which a high slope of these curves will be seen. For the case of ten neutral loci, with $S = 0{\cdot}9$, 0·16 of the loci were heterozygous on average, with a standard deviation of 0·20. In this case, therefore, most of the populations would lie in the region where the curves are steepest. For lower selfing, the standard deviations were slightly larger, and the means were 0·37 and 0·46 for $S = 0{\cdot}5$ and 0·2, respectively. For these selfing rates, the population would include substantial fractions of individuals with sufficient numbers of heterozygous loci that a levelling-off in apparent fitness should be seen. The numerical values were similar for the case of 5 neutral loci.

Fig. 6 shows that there was again a similar, but weaker, relationship between heterozygosity and apparent fitness among the progeny produced by selfing. Again, this leads to a difference between the observed values of the w_s/w_x ratios and the values estimated according to the method of Ritland (1990). For the parameter values used in the runs with ten selected loci shown in Figs. 5 and 6, the observed value of the fitness ratio was 0·760 for a selfing rate of 0·2, and the estimated value was 0·752. The agreement was thus very good. The agreement worsened as the selfing

https://doi.org/10.1017/S0016672300029244 Published online by Cambridge University Press

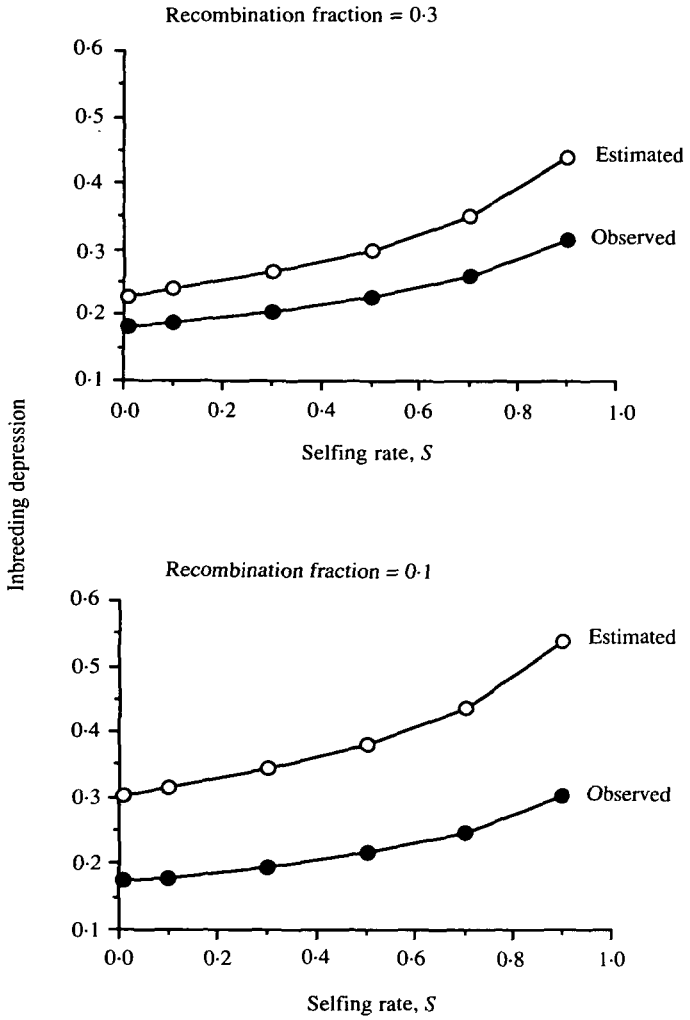

Fig. 4. Comparisons of the observed inbreeding depression with the values estimated from the change in fixation indices. The parameter values were the same as for Fig. 3.

rate increased. For $S = 0{\cdot}9$, the observed value was 0·64, compared with the estimate of 0·56, a 22% overestimate of the inbreeding depression.

5. Discussion

(i) *Apparent fitnesses of genotypes at a neutral locus*

The results described here show that the appearance of strong selection can be produced for a neutral locus in partially selfing populations in which there is inbreeding depression. Although it has seemed intuitively likely that large effects could be produced, no study has previously been done of a multi-locus system, nor have the detailed effects that occur been examined. It is important to realise that this apparent selection does not affect the allele frequencies at the neutral locus. It therefore has no evolutionary importance in determining the genotype frequencies at the neutral locus, and is truly described by the term 'apparent selection'. A point which has not previously been emphasised is that on the apparent selection

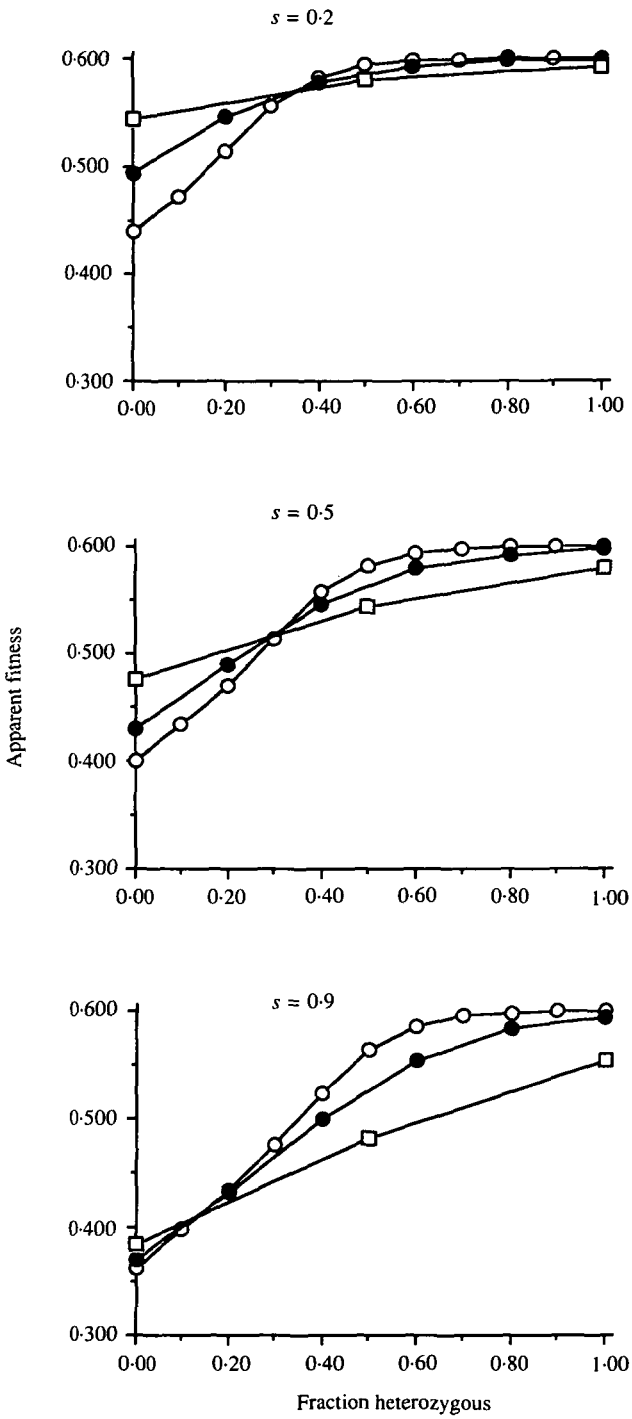

Fig. 5. Apparent fitnesses of genotypes with different numbers of neutral loci heterozygous in a model with ten unlinked loci having symmetrical overdominance with selection coefficient 0·1 for each locus. The neutral loci were unlinked to the selected loci, and to one another, and different numbers of neutral loci were modelled. The results from runs with two neutral loci are shown as open squares, those with five neutral loci as filled circles, and those with ten neutral loci as open circles.

hypothesis the largest selection coefficient will always be against the homozygote for the rarer allele. This is consistently found in the plant data, whenever significant apparent selection at particular loci has been detected (Mitton, 1989), and in one case the

https://doi.org/10.1017/S0016672300029244 Published online by Cambridge University Press

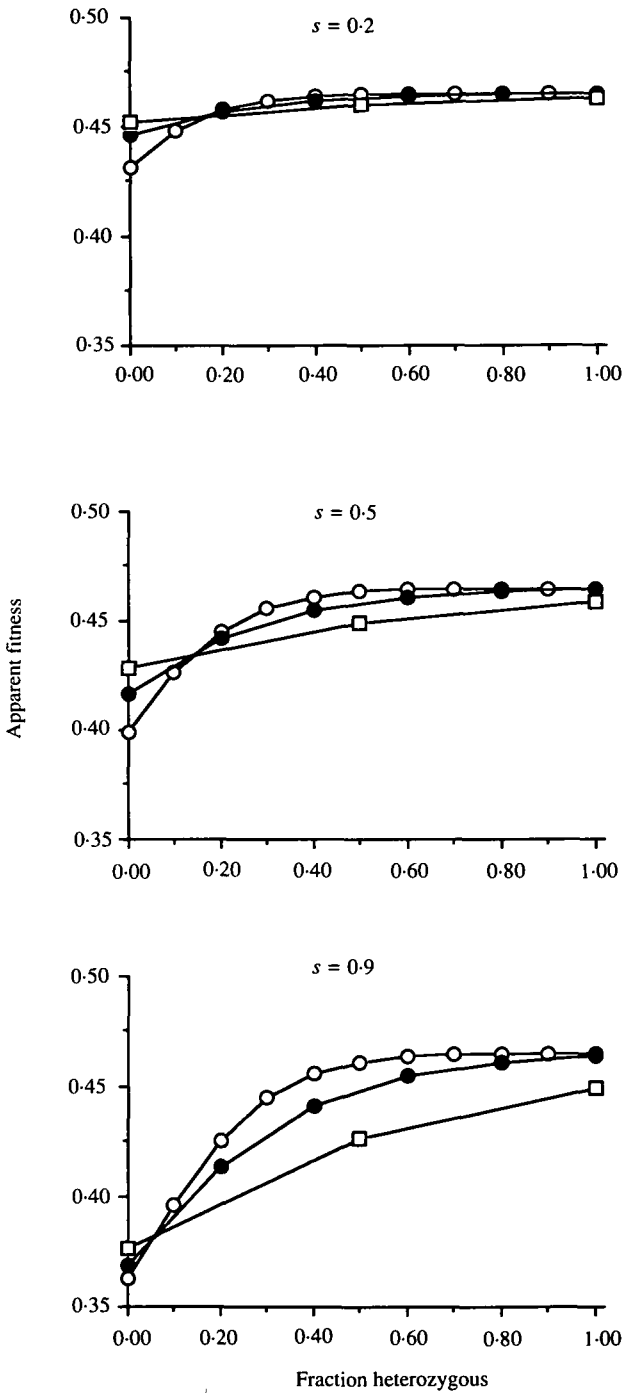

Fig. 6. Apparent fitnesses among the progeny produced by selfing, in the same model as shown in Fig. 5. The results from runs with two neutral loci are shown as open squares, those with five neutral loci as filled circles, and those with ten neutral loci as open circles.

direction of selection was apparently different between populations in which an allele was common, and one in which it was rare (Ganders *et al.* 1977). If the large selection coefficients were due to selection at the marker loci, some loci should have low fitness of both homozygotes, given the frequency of very low homozygote fitnesses in the data as a whole.

Heterozygosity at marker loci need not necessarily be strongly correlated with measures of fitness, because the effect on the average fitness of homozygotes depends on the allele frequencies at the neutral locus, and is strongest when they are in equal frequencies. The fitness of a rare homozygote at the marker locus will be as low as the mean fitness of progeny of selfing, but if this allele is rare, this genotype has little effect on the average fitness of homozygotes. Nevertheless, apparent selection of sufficient magnitude to be detectable in real organisms should be generated. If it were possible to study the genotypes separately at the marker loci, clearer evidence might be obtained. However, the numbers of the genotype homozygous for the rare allele are usually small (Bush, Smouse & Ledig, 1987).

The model of mutational load is probably more biologically realistic than the heterozygote advantage model for genetic variation in fitness in natural populations. There is much evidence for mutational load, and a reasonable degree of quantitative agreement between the values estimated from real populations and those predicted from known values of mutation rates and selection parameters for mutant alleles (Crow, 1970; Charlesworth & Charlesworth, 1987). The results presented here show that this model can generate detectable levels of apparent selection at a neutral locus, especially when there are many loci subject to mutation (at a high rate for the genome as a whole) to mildly deleterious alleles. The mutation rate of 1·0 per diploid genome per generation appears to be a plausible value, base on the information available at present (Kondrashov, 1988; Charlesworth *et al.* 1990). The assumption that all the loci are unlinked is evidently unrealistic, so to investigate the effects of linkage I also studied a model in which there is complete linkage of the selected loci. With linkage, the total amount of inbreeding depression is reduced, because loci segregate together and it is the fitness effects of these blocks that should be considered. However, the associations between selected and neutral loci will be stronger, so that greater apparent selection coefficients could arise. These effects are evident in the comparison between the linked and unlinked models studied here (Table 3, and Tables 1 and 2, respectively).

The inclusion of the heterozygote advantage model in the present paper is an attempt to model partial linkage, at least approximately. Although overdominance at the single locus level is unlikely to be a major contributor to the total inbreeding depression found in populations, it should represent to a good approximation the essential features of the behaviour of chromosomes or chromosome segments that are subject to mutation. Just as for the genome as a whole in the model of mutation studied here, a chromosome will accumulate mutations so that homozygotes produced by selfing will suffer a reduction in fitness if the mutations are partially or wholly recessive. In *Drosophila*, for example, homozygosity for whole chromosomes has severe effects on fitness (reviewed in

Crow,1970; Charlesworth & Charlesworth, 1987). One can therefore consider the genome as a whole to be likely to behave as a set of units each showing heterozygote advantage, with some recombination between them. The results presented here for the case of heterozygote advantage can therefore give an idea of how a system consisting of linked loci will behave. A more accurate way to obtain these results would be to model a multi-locus system with linked loci subject to mutation, but this cannot be done deterministically. Stochastic stimulations to study such a model will be done in the future, but the present results show the general similarity with the mutational model. The apparent fitnesses of genotypes at a neutral marker locus behave similarly, and in both models linkage reduces the inbreeding depression, but can increase the associations between these loci, and thus the apparent fitness differences.

In all the runs studied, there was no linkage disequilibrium between the neutral and selected loci, no matter how close the linkage between the neutral and selected loci. This is expected, as there is no selective or other force present to produce linkage disequilibrium (Thomson, 1977). Some authors (e.g. Gaffney, 1990) have suggested that linkage disequilibrium may be important in producing apparent fitness differences at marker loci, but there is clearly no need to invoke any such effect. The present results using completely deterministic calculations, without even any randomly generated linkage disequilibrium, show that the effects on apparent fitnesses generated in these model inbred populations can be due solely to the associations between homozygosity of different loci. This disagrees with the statement of Mitton & Grant (1984, p. 490) that: 'Associations will only be evident if the genotypes at the linked genes are not independently distributed, that is, if there is linkage disequilibrium between loci', and thus invalidates their conclusion that, since linkage disequilibrium is rarely found: 'The distribution of a genotype at one locus will occasionally reflect selection at other loci, but this is not likely to be the primary explanation of so general a phenomenon.' Such statements are valid only for completely outbred populations.

(ii) *Relationship between heterozygosity at marker loci and fitness*

The effect of correlations in homozygosity between the marker locus and the selected loci produces not only apparent selection at the neutral locus, but should also result in a relationship between fitness measures and heterozygosity at allozyme loci. Writing w_{het} and w_{hom} for the fitnesses of heterozygotes and homozygotes, respectively, then if one could assume independence between the genotypes of different loci, it is evident from considering a single locus that the slope of the regression of fitness on number of heterozygous loci would be

$$\text{slope} = \frac{w_{het} - w_{hom}}{w_{het}}, \qquad (2)$$

which is equal to 1 − the average relative fitness of homozygotes. With more than one neutral locus, one would expect the slope to be greater with more loci scored, as there is more certainty about whether an individual is a product of selfing or of outcrossing [see also Mitton & Jeffers (1987)]. Equation (2) is therefore a lower bound. I have only been able to do calculations for unlinked neutral loci, but the results confirm these expectations (Fig. 5). The slopes of such plots would presumably be greater in the case of linkage to the selected loci.

As can be seen from Figs. 1, 2 and 5 and Tables 1, 3 and 4, quite large slopes could be generated in these models, of the order of 0·1 or 0·2 for moderately deleterious mutations or for heterozygote advantage. There is a clear tendency for the slope to level off with increasing numbers of neutral marker loci heterozygous, as the fitness converges on the mean fitness of outcrossed progeny. This happens because the difference in the probability that two genotypes are

Table 6. *Regression values of fitness components on numbers of marker loci heterozygous, from the published literature*

Species	Fitness component	Number of marker loci	Slope of regression	References
Marine molluscs				
Mytilus edulis	Shell length	5	0·076	Koehn & Gaffney (1984)
Oysters (*Crassostrea virginica*)	Weight	6	0·023	Zouros, Singh & Miles (1980)
Conifers				
Pitch pine (*Pinus rigida*) (several populations)	Annual basal area increment	6–8	0.17–0·16	Ledig *et al.* (1983)
Loblolly pine (*Pinus taeda*)	Height	3	0·036	Bush (1988)
	Diameter	3	0·086	Bush (1988)
Flowering plants				
Maize (F2 between inbred strains)	Grain yield	13	0·029	Stuber (1989)

https://doi.org/10.1017/S0016672300029244 Published online by Cambridge University Press

products of outcrossing, when their heterozygosity differs by one locus, becomes small for genotypes with a moderate number of heterozygous loci. It is interesting to compare these curves with those obtained from real populations (reviewed by Mitton & Grant, 1984 and by Strauss, 1986). In the data, there is no sign of the curvature just described. This is not because I have modelled much larger number of marker loci than are monitored in the studies, as these typically include 5–7 loci (see Table 6), but is probably because of linkage between the neutral and selected loci, which was not taken into account in my model. If each neutral locus is linked to different selected loci, the homozygosity of each marker is informative the homozygosity of its linked selected loci, so the regression should be more nearly linear than those with unlinked loci. It should also be steeper.

It is also interesting to compare the results derived here with regressions of fitness estimates on numbers of heterozygous loci in real populations, even though obviously the values of the parameters used in my calculations are chosen without knowledge of their true values in any real populations. Unfortunately, the data are often expressed as correlations between fitness estimates and the number of heterozygous loci, or the deficiency of heterozygotes (equivalent to $1-F$) averaged across the loci studied (Koehn & Gaffney, 1984). The regressions cannot be extracted from such data. Where suitable data have been published for such characters as size or growth rate, I calculated regressions, using the measurements (from tables or estimated figures) scaled by their mean values. Table 6 shows the results, which show that the model of associative overdominance appears capable of accounting for the magnitudes of the relationships found in natural populations.

In several studies, it has been found that the selection apparently acting on the marker loci is strongest for loci that show the greatest heterozygote deficiency (e.g. Gaffney, 1990). There is no obvious explanation for this if the marker loci are themselves under selection, but on the hypothesis that the loci show apparent selective effects due to association with selected loci, it is expected. Suppose that the inbreeding coefficient differs between loci, due to differences in the history of inbreeding for different regions of the genome. Then a region which is more homozygous in an individual will tend to show a degree of association characteristic of that level of inbreeding. Figs. 1 and 2 above show that, for low levels of inbreeding at least, the apparent overdominance will increase as the amount of inbreeding increases. This might account for the observations just mentioned. It is also important to remember that the equality of selective effects for all selected loci assumed in the models studied here is a simplification. In real populations, homozygosity for different marker loci will be correlated with homozygosity for different regions of the genome, with different effects on fitness.

The result that the apparent fitnesses of the three genotypes at the neutral locus are not the same among the progeny produced by selfing, even when the loci are unlinked, has not previously been pointed out. Ritland (1990) implicity assumes that they are all equal, when he uses the notation w_s for the fitness of progeny of selfing in deriving his method for estimating the ratio w_s/w_x, and thus the inbreeding depression value in natural populations. For example, in his eqn. (3) the viabilities of heterozygotes and homozygotes are given in terms of viabilities for the products of selfing and outcrossing, without distinguishing between the genotypes among the progeny of selfing, but assuming that the viability is the same for all genotypes. However, this assumption of equal fitnesses is not justified. The problem is slight if all the marker loci are unlinked to selected loci, in which case the estimates of w_s/w_x are only slightly higher than the correct values, but when there is linkage the error can be large. Because the heterozygotes at the marker locus usually have higher viability than the homozygotes (except for the case of selection against very severely deleterious mutations, which are unlikely to be of major importance in inbreeding populations), the difference in fitness of selfed and outcrossed progeny estimated by Ritland's method is greater than the true value. In other words, there is a bias in such cases towards underestimation of the w_s/w_x ratio, and thus overestimation of the degree of inbreeding depression. This bias can be quite large when the marker locus is linked to selected loci.

Linkage is of course highly likely in real organisms, with many chromosomes carrying loci that affect fitness (Schuler, 1954; Schuler & Sprague, 1955; Kahler & Wehrhahn, 1986; Stuber, 1990). Moreover, Strauss (1986) showed in a study of a population of *Pinus attenuata* that there was a strong relationship between fitness measures and heterozygosity at allozyme loci among a set of progeny produced by selfing, but a much weaker one for the outcrossed progeny of the same maternal parents. This not only demonstrates that one cannot assume a constant fitness of selfed progeny, independent of genotype at a marker locus, but also suggests that the effect is not a direct one of the loci themselves. If the effect were caused by fitness differences due to the marker loci, the outcrossed progeny should show the effect also. If, however, it is caused by associations between these loci and linked selected loci, the expected result is equality of fitness for the marker locus genotypes among the outcrossed progeny, as Strauss (1986) states.

(iii) *Relative self-fertility of different genotypes*

Mitchell-Olds & Guries (1986) suggested that in populations carrying deleterious alleles at many loci spread throughout the genome, heterozygosity at a marker locus might indicate heterozygosity for recessive or partially recessive deleterious alleles, and

https://doi.org/10.1017/S0016672300029244 Published online by Cambridge University Press

thus heterozygotes would be expected to have lower progeny fitness after selfing than homozygotes. With genetic load due to a few lethal genes, these authors predicted that there should be no such association. They therefore suggested that the nature of the genetic load of a species could be tested by measuring the relative self-fertility of different genotypes at a marker locus, as estimates of their individual genetic loads. The relative self fertility of an individual is the output of viable seeds produced by selfing, relative to that produced by an unrelated donor. This suggestion is based on the reasonable assumption that the self fertility of an individual depends on the number of deleterious alleles for which it is heterozygous. The results obtained here are relevant to this suggestion.

The relative self-fertility of the genotype with alleles i and j at the marker locus is given by

$$\mathrm{RSF}_{ij} = \frac{w_{\text{self progeny of genotype } ij}}{w_{\text{outcrossed progeny of genotype } ij}}. \quad (3)$$

As explained above, assuming that there is no linkage disequilibrium, the fitness of outcrossed progeny is the same regardless of their genotype at the marker locus, or of their parents' genotypes, just as for a neutral locus in a random mating population. The fitness in the denominator of eqn. (3) is therefore a constant. The selfed progeny, however, generally show apparent heterozygote advantage when the genetic load is due to mutation to deleterious alleles with moderate or small effect, or to loci with heterozygote advantage (see Tables 2 and 5 above). Since heterozygotes are the only genotype that can produce heterozygous progeny by selfing, the mean fitness of their progeny produced by selfing will usually be higher than that of the progeny of homozygous genotypes. The opposite would hold when the neutral locus shows apparent underdominance. This was found only for mutation to very strongly deleterious alleles. These conclusions are the opposite of those of Mitchell-Olds & Guries (1986). Furthermore, the effects are likely to be small because they depend on the differences in fitness between the genotypes among the progeny of selfing. In the mutational model, the differences among the genotypes of selfed progeny are much smaller than the difference between the mean of the selfed and outcrossed progeny (see Table 2), though this ignores linkage between the neutral locus and the selected loci, which probably increases the differences [see section 3(ii) above]. It does not seem, therefore, that this method can be used to distinguish between the major gene and multi-genic causes of genetic load, especially since the expectation on either the mutational or the overdominance hypothesis is that heterozygotes at the marker locus should have a somewhat higher self fertility than homozygotes.

(iv) *Heterozygosity in excess of Hardy–Weinberg proportions*

In the results presented above, the inbreeding coefficients at the adult stage are lower than in the zygotes, due to loss of homozygous genotypes that were present among the zygotes because of inbreeding, but which had low survival rates. However strong the selection and the association between the selected loci and the marker locus, this effect can evidently never produce a lower frequency of heterozygotes than the frequency expected under random mating. This is because the genotype frequencies among the progeny of outcrossing will be in Hardy–Weinberg proportions. Thus, even with the greatest possible loss of inbred progeny (when all these offspring die), the adult population would have random mating proportions. Because the progeny of outcrossing show no association between the heterozygosity at different loci (see above), excess heterozygosity cannot be produced because of differential mortality of the least heterozygous individuals, as Mitton & Jeffers (1987) claim. With less differential mortality of inbred zygotes, some residual inbreeding would remain. In contrast to these expectations, studies of natural plant populations have sometimes found heterozygosity in excess of random mating proportions (negative F values), for one or more marker loci (reviewed by Mitton & Jeffers, 1987; Ennos, 1990). Such findings are not common among populations studied (Brown, 1979) but when they are obtained they suggest that the marker loci are themselves subject to selection, and that heterozygotes at these loci have the highest fitness. In that case, excess heterozygosity after selection can be generated.

Brown (1979) reviewed possible causes of F values differing from the values expected on the basis of the observed breeding system of the population. One important point is that there is a bias towards heterozygosity exceeding the Hardy–Weinberg frequency when small samples are used (Levene, 1949; Cannings & Edwards, 1969). The bias is greatest when the allele frequency is close to 0·5. Cannings & Edwards (1969) point out that this could lead to a mistaken claim of overdominance. It is therefore important that the data should be corrected for this problem, but unfortunately most of the published observations do not state whether or not this was done, so that it is difficult to evaluate many of the claims. Certainly, some studies that found heterozygote excess involved small samples [e.g. Yeh *et al.* (1986) sampled 7–12 trees per population, and Levin (1973) sampled 5–28 *Lycopodium* individuals], but this is not true of all studies. Furthermore, the numbers are usually much smaller for adult plants, which show excess heterozygosity, than for seeds or young plants, which usually have deficiencies of heterozygotes (e.g. Cheliak *et al.* 1985; see also Plessas & Strauss, 1986 and Shaw & Allard, 1982 though a correction for bias

was made in these studies). The fact that in some cases the loci with the largest heterozygote frequencies also had the largest excess over Hardy–Weinberg proportions (eg. Cheliak *et al.* 1985; a study in which the number of adults sampled was 30), is in agreement with what is expected if this bias is occurring, but there could, of course, be other explanations (Brown, 1978; Shaw & Allard, 1982).

If selection at the loci scored is the true explanation for the excess heterozygosity, the strength of selection necessary is very large, given that there is homozygote excess at the zygote stage due to inbreeding (Ohta, 1971). It is therefore of interest to examine the fixation indices that result from selection against homozygotes at a locus in partially selfing populations, to see what selection coefficients can produce the observed magnitudes of heterozygote excess. The allele and genotype frequencies expected after selection can easily be calculated for various selection coefficients and selfing rates, using the formulae of Kimura & Ohta (1971, pp. 190–192) for loci with heterozygote advantage, and these can be used to calculate inbreeding coefficients. As one would expect, negative F values can occur in populations with low selfing rates, and are largest in magnitude when the selection is strong. With higher levels of inbreeding, stronger selection is needed to maintain the alleles polymorphic (Kimura & Ohta, 1971), unless selection is symmetrical or nearly so, and negative F values after selection are not found unless selection is very strong. For example, with $S = 0{\cdot}3$, a value of F less than $-0{\cdot}1$ requires symmetrical selection with a selection coefficient of at least 0·4. With $S = 0{\cdot}5$, selection coefficients of 0·5 or more are required. With asymmetrical selection coefficients, stronger selection is required.

The magnitudes of the negative F values observed in real populations are often large, frequently on the order of 0·1–0·2 (reviewed by Mitton & Jeffers, 1987; see also Muona, Yazdani & Rudin, 1987) in populations with selfing rates of 0·1–0·2, or even higher. For example, Ganders, Carey & Griffiths (1977) found values greater than 0·34 (average 0·52) for a locus controlling fruit type in *Plectritis congesta*, in populations estimated to have a selfing rate of 0·33. This requires very strong heterotic selection, as mentioned above. This may not be implausible for a fruit character (Ganders *et al.* 1977). Such strong selection should be measurable directly among outcrossed progeny. As far as I am aware, a direct test of selection has not been done for this character. For the case of allozyme loci, a few experiment tests using progenies of uniform breeding coefficient have been done. Strauss (1986) found little or no relationship between fitness and heterozygosity at allozyme loci among a set of progeny produced by outcrossing between trees from different populations. Similar results were obtained by Beaumont, Beveridge & Budd (1983) and Leary *et al.* (1987).

(v) *Conclusions*

The general conclusion that can be drawn from the results presented here is that many of the observations on the fitnesses of genotypes at allozyme loci are consistent with the view that these loci do not themselves influence fitness, but are associated in their homozygosity with other loci that do so. The most likely reason for such an association is inbreeding. In the present paper, I have studied only self-fertilisation. The properties of other types of inbreeding should be similar, provided that they result in individuals of various inbreeding coefficients in the populations. The finding that the relationship between heterozygosity of the allozyme loci and fitness appears to vary in different populations (Ledig, Guries & Bonefeld, 1983), and from year to year in the same population (Gaffney, 1990), is readily explained if there is occasionally a high degree of outcrossing, since a single generation of outcrossing destroys the associations found under partial inbreeding. This could be tested by testing for differences in the breeding system concurrently with differences in the relationship between marker genotypes and fitness. The findings of excess frequencies of both multiply heterozygous and homozygous genotypes, compared with what would be expected if the heterozygosity of different loci were independent (Gaffney *et al.* 1990; Alvarez, 1989), also suggest with partial inbreeding in the population, and are hard to explain in any other way. With such a breeding system, some individuals are products of inbreeding, and thus tend to be more homozygous than average at all loci, while others are progeny of outcrossing events and therefore more heterozygous than average at all loci. Indeed, a test for such a correlation of heterozygosity between loci would be a very worthwhile part of any study of the relationship between heterozygosity and fitness.

Probably the best test that can be done to distinguish whether this explanation can account for the observed fitness differences between genotypes is to test the genotypes in a set of individuals known to have been produced by outcrossing (or in a naturally outcrossing population). As Houle (1990) also argues, if one studies partially inbreeding populations the associations between the loci studied and the loci under selection will obscure the true relationship between the genotypes scored and their effects on fitness, and it will be hard to detect selection, even if it is happening in such populations. As mentioned above, the few tests that have been done using outcrossed progeny or populations known to be non-inbred have failed to detect selection at the marker loci, and have found very little relationship between heterozygosity and fitness (see review by Houle, 1990; Leberg *et al.* 1990). If, however, heterozygosity values in excess of Hardy–Weinberg expectations (negative F values) are found for marker loci in adult individuals, this does suggest that selection of these loci (or at linked loci in

linkage disequilibrium with the markers) has occurred. Alternative explanations are, however, possible, such as differences in allele frequency between the male and female gamete pools, and these must be ruled out before one could be convinced that selection is really acting on the loci under study (Brown, 1979; Plessas & Strauss, 1986). Since very strong selection would be necessary to produce such results, in inbred populations, direct tests of the selection in outcrossed progeny should readily show whether the loci are themselves involved.

I thank B. Charlesworth, K. Ritland and M. T. Morgan for comments on this manuscript. This work was supported by NSF grants BSR 8516629 and BSR 8817976.

References

Alvarez, G., Zapata, C., Amaro, R. & Guerra, A. (1989). Multilocus heterozygosity and fitness in the European oyster, *Ostrea edulis* L. *Heredity, London* **63**, 359–372.

Beaumont, A. R., Beveridge, C. M. & Budd, M. D. (1983). Selection and heterozygosity within single families of the mussel *Mytilus edulis* (L.) *Marine Biology Letters* **4**, 151–161.

Bennet, J. H. & Binet, F. E. (1956). Association between Mendelian factors with mixed selfing and random mating. *Heredity, London* **10**, 51–55.

Bijlsma-Meeles, E. & Bijlsma, R. (1988). The alcohol dehydrogenase polymorphism in *Drosophila melanogaster*: fitness measurements and predictions under conditions with no alcohol stress. *Genetics* **120**, 743–753.

Brown, A. H. D. (1979). Enzyme polymorphism in plant populations. *Theoretical and Applied Genetics* **15**, 1–42.

Bush, R. M. (1988). The fitness consequences of multiple-locus heterozygosity: The relationship between heterozygosity and growth rate in pitch pine (*Pinus rigida* Mill.) and loblolly pine (*Pinus taeda* L.). Ph.D. thesis, University of Michigan.

Bush, R. M., Smouse, P. E. & Ledig, F. T. (1987). The fitness consequences of multiple-locus heterozygosity: the relationship between heterozygosity and growth rate in pitch pine (*Pinus rigida* Mill.). *Evolution* **41**, 787–798.

Cannings, C. A. & Edwards, A. F. W. (1969). Expected genotype frequencies in a small sample: deviations from Hardy–Weinberg equilibrium. *American Journal of Human Genetics* **21**, 245–247.

Charlesworth, D. & Charlesworth, B. (1990). Inbreeding depression with heterozygote advantage and its effect on selection for modifiers changing the outcrossing rate. *Evolution* **44**, 870–888.

Charlesworth, D., Morgan, M. T. & Charlesworth, B. (1990). Inbreeding depression, genetic load and the evolution of outcrossing rates in a multi-locus system with no linkage. *Evolution* **44**, 1469–1489.

Cheliak, W. M., Dancik, B. P., Morgan, K., Yeh, F. C. H. & Strobeck, C. (1985). Temporal variation of the mating system in a natural population of jack pine. *Genetics* **109**, 569–584.

Cockerham, C. C. & Rawlings, J. O. (1967). Apparent heterosis of a neutral gene with inbreeding. *Ciência et Cultura, São Paolo* **19**, 88–94.

Ennos, R. A. (1990). Detection and measurement of selection: genetic and ecological approaches. In *Plant Population Genetics, Breeding, and Genetic Resources* (ed. A. H. D. Brown, M. T. Clegg, A. L. Kahler and B. S. Weir), pp. 181–199. Sunderland, Massachusetts: Sinauer.

Gaffney, P. M. (1990). Enzyme heterozygosity, growth rate, and viability in *Mytilus edulis*: another look. *Evolution* **44**, 204–210.

Gaffney, P. M., Scott, T. M., Koehn, R. K. & Diehl, W. J. (1990). Interrelationships of heterozygosity, growth rate and heterozygote deficiencies in the coot clam, *Mulinia lateralis*. *Genetics* **124**, 687–699.

Ganders, F. R., Carey, K. & Griffiths, A. J. F. (1977). Natural selection for a fruit dimorphism in *Plectritis congesta* (Valerianaceae). *Evolution* **31**, 873–881.

Haldane, J. B. S. (1950). The association of characters as a result of inbreeding and linkage. *Annals of Eugenics* **15**, 15–23.

Hastings, A. (1990). The interaction between selection and linkage in plant populations. In *Plant Population Genetics, Breeding, and Genetic Resources* (ed. A. H. D. Brown, M. T. Clegg, A. L. Kahler and B. S. Weir), pp. 163–180. Sunderland, Massachusetts: Sinauer.

Hedrick, P. W., Jain, S. K. & Holden, L. (1978). Multilocus systems in evolution. *Evolutionary Biology* **2**, 104–184.

Houle, D. (1989). Allozyme-associated heterosis in *Drosophila melanogaster*. *Genetics* **123**, 789–801.

Kahler, A. L. & Wehrhahn, C. F. (1986). Association between quantitative traits and enzyme loci in the F2 population of a maize hybrid. *Theoretical and Applied Genetics* **72**, 15–26.

Kimura, M. (1958). Zygotic frequencies of a partially self-fertilizing population. *Annual Report of the National Institute of Genetics, Japan* **8**, 104–105.

Kimura, M. & Ohta, T. (1971). *Theoretical Topics in Population Genetics*. Princeton, NJ: Princeton University Press.

Koehn, R. K. & Gaffney, P. M. (1984). Genetic heterozygosity and growth rate in *Mytilus edulis*. *Marine Biology* **82**, 1–7.

Kondrashov, A. S. (1985). Deleterious mutation as an evolutionary factor. II. Facultative apomixis and selfing. *Genetics* **111**, 635–653.

Kondrashov, A. S. (1988). Deleterious mutations and the evolution of sexual reproduction. *Nature* **336**, 435–440.

Leary, R. F., Allendorf, F. W. & Knudsen, K. L. (1987). Differences in inbreeding coefficient do not explain the association between heterozygosity at allozyme loci and developmental stability in rainbow trout. *Evolution* **41**, 1413–1415.

Leberg, P. L., Smith, M. H. & Rhodes, O. E. (1990). The association between heterozygosity and growth of deer fetuses is not explained by the effects of the loci examined. *Evolution* **44**, 454–458.

Ledig, F. T., Guries, R. P. & Bonefeld, B. A. (1983). The relation of growth to heterozygosity in pitch pine. *Evolution* **37**, 1227–1238.

Levene, H. (1949). On a matching problem arising in genetics. *Annals of Mathematical Statistics* **20**, 91–94.

Mitchell-Olds, T. & Guries, R. P. (1986). Genetic load and heterozygosity in the Pinaceae. *Canadian Journal of Genetics and Cytology* **28**, 942–946.

Mitton, J. B. (1989). Physiological and demographic variation associated with allozyme variation. In *Isozymes in Plant Biology* (ed. D. E. Soltis and P. S. Soltis), pp. 127–145. Portland, Oregon: Dioscorides Press.

Mitton, J. B. & Grant, M. C. (1984). Associations among protein heterozygosity, growth rate, and developmental homeostasis. *Annual Reviews of Ecology and Systematics* **15**, 479–499.

Mitton, J. B. & Jeffers, R. M. (1987). The genetic consequences of mass selection for growth rate in Engelmann spruce. *Silvae Genetica* **38**, 6–12.

Muona, O., Yazdani, R. & Rudin, D. (1987). Genetic change between life stages in *Pinus sylvestris*: allozyme

variation in seeds and planted seedlings. *Silvae Genetica* **16**, 39–42.

Ohta, T. (1971). Associative overdominance caused by linked detrimental mutations. *Genetical Research* (*Cambridge*) **18**, 277–286.

Ohta, T. & Cockerham, C. C. (1974). Detrimental genes with partial selfing and effects on a neutral locus. *Genetical Research* (*Cambridge*) **23**, 191–200.

Plessas, M. E. & Strauss, S. H. (1986). Allozyme differentiation among populations, stands, and cohorts in Monterey pine. *Canadian Journal of Forest Research* **16**, 1155–1164.

Ritland, K. (1990). Gene identity and the genetic demography of plant populations. In *Plant Population Genetics, Breeding, and Genetic Resources* (ed. A. H. D. Brown, M. T. Clegg, A. L. Kahler and B. S. Weir), pp. 181–199. Sunderland, Massachusetts: Sinauer.

Schuler, J. F. (1954). Natural mutations in inbred lines of maize and their heterotic effect. I. Comparison of parent, mutant and their F1 hybrid in a highly inbred background. *Genetics* **39**, 908–922.

Schuler, J. F. & Sprague, G. F. (1955). Natural mutations in inbred lines of maize and their heterotic effect. II. Comparison of mother line versus mutant when outcrossed to related inbreds. *Genetics* **41**, 281–291.

Shaw, D. V. & Allard, R. W. (1982). Isozyme heterozygosities in adult and open-pollinated embryo samples of Douglas fir. *Silva Fennica* **16**, 115–121.

Strauss, S. H. (1986). Heterosis at allozyme loci under inbreeding and crossbreeding in *Pinus attenuata*. *Genetics* **113**, 115–134.

Strobeck, C. (1980). Partial selfing and linkage: the effect of a heterotic locus on a neutral locus. *Genetics* **92**, 305–315.

Stuber, C. W. (1989). Isozymes as markers for studying and manipulating quantitative traits. In *Isozymes in Plant Biology* (ed. D. E. Soltis & P. S. Soltis), pp. 206–220. Portland, Oregon: Dioscorides Press.

Stuber, C. W. (1990). Molecular markers in the manipulation of quantitative characters. In *Plant Population Genetics, Breeding, and Genetic Resources* (ed. A. H. D. Brown, M. T. Clegg, A. L. Kahler and B. S. Weir), pp. 181–199. Sunderland, Massachusetts: Sinauer.

Thomson, G. (1977). The effect of a selected locus on a linked neutral locus. *Genetics* **85**, 752–788.

Weir, B. S. & Cockerham, C. C. (1973). Mixed selfing and random mating at two loci. *Genetical Research* (*Cambridge*) **21**, 247–262.

Yeh, F. C., Khalil, M. A. K., El-Kassaby, Y. A. & Trust, D. C. (1986). Allozyme variation in *Picea mariana* from Newfoundland: genetic diversity, population structure, and analysis of differentiation. *Canadian Journal of Forest Research* **16**, 713–720.

Ziehe, M. & Roberds, J. H. (1989). Inbreeding depression due to overdominance in partially self-fertilizing plant populations. *Genetics* **121**, 861–868.

Zouros, E. & Foltz, D. W. (1984). Possible explanations of heterozygote deficiency in bivalve molluscs. *Malacologia* **25**, 583–591.

Zouros, E., Singh, S. M. & Miles, H. E. (1980). Growth rate in oysters: an overdominant phenotype and its possible explanations. *Evolution* **34**, 856–867.