Associative overdominance, heterozygosity and fitness

PEKKA PAMILO* & SNAEBJÖRN PÁLSSON

Program of Conservation Biology, Department of Genetics, University of Uppsala, Box 7003, 750 07 Uppsala, Sweden

Our aim in this study was to examine the power of associative overdominance in creating a correlation between individual heterozygosity and fitness and in maintaining genetic polymorphism at neutral loci. This was undertaken by simulating a diploid model with five chromosomes, each with 1000 loci that could have deleterious mutations linked to neutral markers located on the same chromosomes. The simulations were carried out with various combinations of the following parameters: population size (N), number of crossovers (c), selection coefficient (s) and dominance (h). All combinations of parameter values resulted in a positive regression of the fitness on the individual marker heterozygosity, although the fitness differences were not very large when $c = 2$. The association between individual heterozygosity and fitness was clearest for recessive mutations (small h) with intermediate selection coefficients in small populations. The level of marker heterozygosity in the population was higher than the neutral expectation for many parameter values, even though regression of fitness on heterozygosity was not always steep at the individual level. These results agreed with the conclusion that associative overdominance could help to maintain polymorphisms in small populations. Strong selection occasionally yielded gene diversities lower than the neutral expectation, reflecting the prediction that strong selection reduces the effective population size.

Keywords: associative overdominance, background selection, deleterious mutations, heterozygosity, linkage.

Introduction

Selection on harmful mutations affects the genetic variation at linked loci. Two alternative hypotheses with opposite effects have been put forward, those of background selection (Charlesworth *et al.*, 1993) and associative overdominance (Frydenberg, 1963; Ohta, 1971). Directional selection against harmful mutations results in fitness variation among individuals, lowering the effective population size and the level of neutral variation below the value predicted by a single-locus neutral model (Charlesworth *et al.*, 1993; Nordborg *et al.*, 1996). The general effect of such background selection can be described by considering the fraction (f_o) of genomes with the smallest number of harmful mutations. In the absence of recombination, the future gene pool is expected to originate from these genomes, and the effective population size is reduced to the fraction f_o of the total population size (Charlesworth *et al.*,

1993; Hudson & Kaplan, 1995). Background selection is based on directional selection against harmful mutations and the theory applies to large populations, although simulations show that heterozygosity is reduced even in small populations ($N = 100$) as long as the product $Nhs > 1$, where h is the dominance coefficient and s the selection coefficient against recessive homozygotes (Charlesworth *et al.*, 1993; Nordborg *et al.*, 1996). The alleles are expected to behave as if neutral when $Nhs < 1$ (Charlesworth *et al.*, 1993).

When homologous chromosomes carry harmful mutations at different loci, the individuals that are chromosomal heterozygotes are at the same time heterozygous at these selected loci and, at least with an elevated probability, at other loci in the same linkage group. If the harmful mutations are completely or partially recessive, selection on them leads to associative overdominance which is 'due to nonrandom linkage between the observed pair of allelic units and the entire remainder of genetic alternatives present in the same chromosome pair'

*Correspondence. E-mail: pekka.pamilo@genetik.uu.se

(Frydenberg, 1963). Associative overdominance depends on linkage disequilibrium between selected and neutral loci, and the effect is expected to be strongest in small populations and in genomes with restricted recombination. We can distinguish two different features of associative overdominance, the association of individual heterozygosity and fitness and the maintenance of polymorphism at linked loci.

Ohta (1971) developed an analytical model of associative overdominance which included the effects of selection, inbreeding and linkage. The model is based on a one-locus mutation-selection balance that is extended to multiple loci by assuming additivity. By application of this model to data from *Drosophila*, Ohta concluded that associative overdominance caused by recessive deleterious alleles 'is probably responsible for most of the observed superiority of heterozygotes'. Ohta (1973) also concluded that associative overdominance may be partly responsible for the maintenance of genetic variability, although she first considered it ineffective (Ohta, 1971). Similar conclusions were reached by Sved (1972) on the basis of multilocus simulations. A two-locus hitchhiking model showed that significant associative effects required that the product between the effective population size and the recombination fraction between the two loci must be less than 5 (Ewens, 1979, p. 207), and it seems that tight linkage and small *N_h*s are required for associative overdominance.

There is a large literature on positive correlation between individual heterozygosity at putatively neutral allozyme loci and some fitness components (Mitton, 1993; David, 1997). It has been questioned how well the small number of loci used in the empirical studies reflects genomic heterozygosity, as the correlation of heterozygosities at markers and in the total genome is expected to be $\sqrt{p}$, where *p* is the proportion of loci studied (Chakraborty, 1981). Zouros & Mallet (1989) listed six specific hypotheses to explain the observed correlation between individual heterozygosity and performance in natural populations: functional overdominance, multiple-locus dominance, balanced enzyme pathways, null alleles, chromosomal loss and associative overdominance. David (1997) showed how it could be empirically possible to distinguish whether such heterotic effects were caused by overdominance or inbreeding depression. Experimental work has shown diminishing evidence for overdominance at the level of single loci (Houle, 1989), and Turelli & Ginzburg (1983) demonstrated theoretically that if polymorphisms are maintained (at equilibrium) by any of several forms of balancing selection, that should result in a

positive correlation between heterozygosity and fitness.

Our aim is to test the predictions from Ohta's (1971) model by using multilocus simulations. We use a simulation model to examine the power of associative overdominance in creating a correlation between individual heterozygosity and fitness in small populations, and in maintaining genetic polymorphism at neutral loci. We also aim to assess the conditions where the two opposing effects, heterozygosity-reducing background selection and heterozygosity-increasing associative overdominance, are important; in other words, when can associative overdominance override the effects of background selection. The simulations are also used to examine the correlation of heterozygosities at neutral markers and at the linked background loci. The basic scheme of our approach is as follows. New mutations, which are commonly slightly deleterious and recessive, occur at linkage disequilibrium with the alleles at the linked loci. This leads to associative overdominance at the linked loci, because heterozygous individuals are also heterozygous for the deleterious alleles, whereas homozygotes at a neutral locus are likely to be homozygous for some deleterious alleles and to have reduced fitness. It should be noted that associative overdominance at a neutral locus can be caused by functional overdominance at a linked locus. However, throughout this paper, we use the concept of associative overdominance solely in reference to the effects of deleterious alleles.

Models

Simulations

The model had five pairs of chromosomes, each one having 1001 loci. The locus in the middle of each chromosome was a selectively neutral marker and was surrounded by 1000 loci (500 on each side) that could have deleterious mutations. We classified individuals into heterozygosity classes (*HET*, ranging from 0 to 5) depending on the number of heterozygous marker loci they had.

The selected background loci determined the individual fitnesses. They could have two types of alleles, + (wild type) and - (a deleterious mutant). The fitnesses of the three genotypes (+ +, + -, - -) at a single locus were 1, 1-*h_s* and 1-*s*, respectively, and the multilocus fitness determined by all of the 5000 background loci was given multiplicatively as

$$w = (1-s)^m(1-hs)^n,$$

where m and n were the numbers of loci homozygous or heterozygous for deleterious alleles, respectively.

Mutations at the neutral loci followed the infinite allele model and the haploid mutation rate was $v = 2 \times 10^{-4}$ per locus and generation. The rate of deleterious mutations was $u = 2 \times 10^{-5}$ per locus, which gave the total mutation rate of $U = 0.2$ per genome and generation.

All the loci were evenly spaced in the chromosomes and the crossing-over frequency was the same for each chromosome. Crossovers occurred in tetrads and the expected number of crossovers per chromosome pair was c ; the value used in the simulations was either $c = 2$ (i.e. 0.5 crossovers between any pair of chromatids) or $c = 0$. The numbers of new mutations and crossovers were considered to be Poisson-distributed random variables when forming new gametes. Their locations were then selected using uniformly distributed random numbers.

The parameter c is related to the recombination distance r between the marker and a specific background locus i according to Haldane's mapping function

$$r = (1 - e^{-ic/1000})/2,$$

where i is the ranking number of the background locus when counting from the neutral marker towards the tips of the chromosome ($i = 1, \dots, 500$).

The population was considered to be completely panmictic and hermaphroditic, but self-fertilization was not allowed. (Self-fertilization would have produced highly homozygous offspring.) The population size N was kept constant in any one simulation. Each new offspring was formed by selecting randomly (with replacement) two parents and forming the gametes in the same way as in the study of Pamilo *et al.* (1987). The probability of an individual i being selected as a parent was directly proportional to its fitness w_i as given above.

The model was simulated using fixed values of the population size ($N = 100$ or $N = 500$), of selection coefficient ($s = 0.02, 0.1$ or 0.5), and of dominance ($h = 0.1$ or 0.01). The simulations were first run for 200 generations with tenfold mutation rates in order to allow harmful mutations to accumulate. The following 200 generations were used for drift and selection to stabilize the population, and the results were recorded from generations 401–1000. Simulations were replicated 3–47 times for each set of parameter values (Table 1; the number varied because of long computing times).

As any given generation would have given erratic results depending on the specific associations of the

alleles at the marker loci and background loci and on stochastic events, the results were averaged over the last 600 generations in each replicated run. The successive generations were not independent of each other, but this should not bias the average association between heterozygosity and fitness. The standard errors were calculated over replicated runs. The results were calculated for the mean heterozygosity at the neutral marker loci and for the association between individual heterozygosity class and fitness.

Fitnesses

The absolute fitnesses, w , could not easily be compared over generations and replicates, as the mean fitnesses were likely to fluctuate and change with the accumulation of harmful mutations. We used two methods to standardize the fitnesses in each generation. This was achieved by setting either

Table 1 The coefficients of associative overdominance: s' calculated from Ohta's (1971) model and that observed in the simulations, $b/(1+b)$, with its standard error. NA (not available) indicates simulations that were not performed, and n is the number of replicate simulations

N	c	h	s	Associative overdominance				n
				s' (Ohta)	$b/(1+b)$	SE		
100	0	0.1	0.02	0.108	0.030	0.002	17	
			0.1	0.052	0.053	0.009	20	
			0.5	0.014	0.003	0.001	20	
		0.01	0.02	0.848	0.067	0.003	7	
			0.1	0.805	0.161	0.013	18	
			0.5	0.625	0.112	0.014	47	
	2	0.1	0.02	0.004	0.002	0.000	16	
			0.1	0.005	0.002	0.000	20	
			0.5	0.004	0.001	0.000	27	
		0.01	0.02	0.010	0.002	0.000	5	
			0.1	0.033	0.005	0.001	10	
			0.5	0.060	0.009	0.000	20	
500	0	0.1	0.02	0.052	0.018	0.002	7	
			0.1	0.014	0.010	0.003	15	
			0.5	0.003	0.001	0.000	16	
		0.01	0.02	0.805	0.032	0.001	13	
			0.1	0.625	0.076	0.004	13	
			0.5	0.279	0.022	0.006	10	
	2	0.1	0.02	NA				
			0.1	0.001	0.001	0.000	3	
			0.5	0.001	0.001	0.000	6	
		0.01	0.02	0.008	0.001	0.000	5	
			0.1	0.016	0.003	0.001	4	
			0.5	0.017	0.005	0.001	4	

the mean fitness of the heterozygosity class 0 (W_0) or the mean fitness of the whole population (W) as equal to one and dividing all the individual fitnesses with the chosen reference ($w'_i = w_i/W_0$, or $w'_i = w_i/W$). Such relative fitnesses were calculated in each generation, and the means were calculated over generations by weighting with the observed numbers of individuals in each heterozygosity class. The two ways to standardize the fitnesses gave very similar results and we present only the results obtained when standardizing with the population mean fitness.

The results from each replicate were summarized by calculating the linear regression coefficient (b) of the standardized mean fitness on the heterozygosity (HET) class. The regressions were calculated by weighting each heterozygosity class with the number of generations for which it occurred in the simulations. The slopes b of the regression lines measure how much individual fitness increases per one heterozygous marker. The effects of different parameter values on b were studied by ANOVA, following a stepwise procedure recommended for unbalanced designs (Venables & Ripley, 1994).

The single-locus model of associative overdominance gave the selection coefficient per selected background locus at a recombination distance r from the neutral marker as

$$s' = \frac{u(1/h - 2)}{2N(r + hs) + 1 + r/(2hs)}$$

when $hs \gg u$ (Ohta, 1971). The coefficient of associative overdominance in our simulations is given as $s' = b$ when the fitness of a heterozygote is 1 and that of a homozygote is $1 - b$, or as $s' = b/(1 + b)$ when the fitnesses are $1 + b$ and 1, respectively. With small b these two are almost identical, and we report the latter values.

Heterozygosity

We used two methods for calculating the association between the genotypes of the markers and of the background loci. First, we calculated the mean number of heterozygous background loci within each heterozygosity (HET) class. This number was standardized within each generation by dividing by the mean number over all individuals. The difference between two successive heterozygosity classes thus indicates how much the background heterozygosity increases per one heterozygous marker.

Secondly, we made calculations on an individual chromosomal basis and calculated the following two

ratios: $BACKHET$ = the number of heterozygous background loci in a pair of chromosomes in individuals heterozygous for the neutral marker, divided by the respective number in individuals homozygous for that marker, and $BACKHOM$ = the same ratio for numbers of background loci that are homozygous for the deleterious mutants. The values were then averaged over chromosomes and generations.

Variation at the population level was measured as the mean gene diversity (or expected heterozygosity, H) over the five marker loci. The gene diversity at a single locus was given by $1 - \sum x_i^2$, where x_i refers to the frequency of the i th allele. The mean over the 600 generations gave the data point for each simulation repeat.

Results

Individual heterozygosity and fitness

All combinations of parameter values resulted in associative overdominance and a positive regression of the fitness on the individual heterozygosity (Table 1), and the strength of overdominance depended strongly on the parameter values used. The effect was strongest when there were no cross-overs ($c = 0$) within the chromosomal pairs (Table 1). In such cases the mean selection coefficients could approach 20%, and even exceed that in individual runs. When crossing-over was allowed, the selection coefficients were at most a few percentage (Table 1), and ANOVA showed that the difference between the regression slopes for the cases $c = 0$ and $c = 2$ was significant (Table 2). In general, the observed associative overdominance was weaker than predicted by Ohta's (1971) model (Table 1), particularly in the absence of crossing over and with strong dominance ($h = 0.01$).

Table 2 Analysis of variance for the increase in individual fitness per one heterozygous marker (b). The values result from fitting a parsimonious model in a stepwise procedure (Venables & Ripley, 1994)

Factor*	d.f.	MS	F	P
c	1	0.291	100.9	$< 10^{-8}$
h in c	2	0.204	70.7	$< 10^{-8}$
N in c	2	0.075	25.9	$< 10^{-8}$
s in c	4	0.018	6.4	6×10^{-5}
$h \times N$ in c	2	0.034	11.7	1×10^{-5}
$h \times s$ in c	4	0.008	2.7	0.03
Residuals	307	0.003		

*'a in b' refers to the effect of the factor a within b , and 'a $\times$ b' refers to the interaction of the two factors.

The effects of the parameters used in the simulations were summarized by calculating in which direction, and how much, the regression slopes (regression of fitness on individual heterozygosity) deviated from the overall mean. This was carried out separately for $c = 0$ and $c = 2$. The population size had a clear effect on the regression slopes, and the association between fitness and heterozygosity (*HET*) was stronger in small ($N = 100$) than in large ($N = 500$) populations, although the effect was not significant when crossing-over was allowed.

The dominance coefficient (h) also had a clear effect (Table 2). Associative overdominance was particularly strong when h was small ($h = 0.01$), i.e. when the deleterious mutations were recessive and well masked in heterozygous conditions. The interaction among N and h (Table 2) could be detected in the absence of crossovers (Fig. 1a,b), and increased dominance had larger effects when $N = 100$ than when $N = 500$. This interaction was, however, not significant when $c = 2$.

The strength of selection had significant effects on the regression coefficient (Table 2), but it also showed significant interactions with dominance (Fig. 1). Associative fitness effects were largest with intermediate selection ($s = 0.1$) when $c = 0$, particularly when $h = 0.01$. With strong selection ($s = 0.5$), very few mutations segregated in the populations, and some clear fitness consequences were seen only when crossing-over was forbidden and h was small (0.01) (Table 1). When crossing-over was allowed ($c = 2$), fitness increased with selection intensity when $h = 0.01$ but not when $h = 0.1$, but the differences were small (Fig. 1c). With weak selection ($s = 0.02$), mutations did accumulate in the chromosomes but the number segregating as polymorphic at any given time was on one hand too small to cause

severe associative effects (without crossovers), and on the other hand large enough to prevent any single locus from determining associative effects when crossovers were allowed (Table 1).

Correlation between markers and background loci

The association between the genotypes at the marker loci and background loci were recorded both for each chromosome separately and for different heterozygosity (*HET*) classes. In both cases, the results indicated to what extent an increase of marker heterozygosity predicted the genotypes at linked loci. The regressions of the number of heterozygous background loci on the heterozygosity class were linear and the slopes corresponded well with the values calculated on a chromosomal basis (*BACKHET*).

The ratios of background heterozygosity showed that an individual heterozygous for a marker had a clearly elevated probability of carrying heterozygous background loci (Table 3, *BACKHET*). It was also clear that individuals that had heterozygous markers had a reduced probability of being homozygous for the harmful background mutants (Table 3, *BACKHOM*).

Population variability

The expected heterozygosity under the infinite allele, neutral model is $4N_e\nu/(4N_e\nu + 1)$. With the values of N and ν used in the simulations, the neutral expectations were 0.074 for $N = 100$, and 0.286 for $N = 500$. The simulations without any background selection gave heterozygosities agreeing well with these expectations, although the means were somewhat higher (Fig. 2).

Fig. 1 Interaction effects of h and s in determining the regression of fitness on heterozygosity. The curves show the mean regression coefficients ($100 \times b$) for different selection coefficients and for the dominance values $h = 0.01$ (dashed lines) and $h = 0.1$ (solid lines). The other parameter values are: (a) $c = 0$, $N = 100$; (b) $c = 0$, $N = 500$; and (c) $c = 2$, $N = 100$.

In the presence of background selection, the simulated mean heterozygosities at the neutral marker loci (H) exceeded the neutral values (simulated or expected) with most combinations of parameter values (Fig. 2), particularly when $Nhs < 1$. When $Nhs > 1$, the observed H was sometimes, but not always, below the neutral expectation. The heterozygosity was particularly high when the mutant alleles were only slightly deleterious, i.e. when selection was weak. The general effects of h and c were less evident, but in a small population ($N = 100$), heterozygosities were clearly largest with $h = 0.01$ and no crossing-over. Strong selection ($s = 0.5$) decreased the mean heterozygosity below the neutral expectation when no crossovers were allowed. The parameters affected H partly in the same way as they affected the regression of fitness on heterozygosity. The parametric correlation coefficient between H and the regression slope b was 0.53 when $N = 100$

Table 3 The relative frequency of heterozygous (*BACKHET*) and homozygous (*BACKHOM*) background loci in a chromosome which is heterozygous for the marker locus. The standard errors have been calculated over simulation replicates. NA (not available) indicates simulations that were not performed

N	c	h	s	<i>BACKHET</i>		<i>BACKHOM</i>	
				Mean	SE	Mean	SE
100	0	0.1	0.02	1.49	0.00	0.37	0.00
			0.1	1.35	0.00	0.42	0.01
			0.5	1.27	0.01	0.46	0.01
		0.01	0.02	1.56	0.00	0.23	0.01
			0.1	1.44	0.00	0.23	0.00
			0.5	1.21	0.00	0.46	0.01
	2	0.1	0.02	1.04	0.00	0.92	0.01
			0.1	1.04	0.00	0.67	0.01
			0.5	1.26	0.01	0.45	0.01
		0.01	0.02	1.03	0.00	0.95	0.01
			0.1	1.03	0.00	0.84	0.02
			0.5	1.07	0.00	0.73	0.01
500	0	0.1	0.02	1.31	0.00	0.28	0.01
			0.1	1.12	0.00	0.36	0.01
			0.5	1.18	0.00	0.46	0.01
		0.01	0.02	1.25	0.00	0.19	0.00
			0.1	1.21	0.00	0.22	0.00
			0.5	1.13	0.00	0.42	0.01
	2	0.1	0.02	NA			
			0.1	1.02	0.00	0.63	0.02
			0.5	1.16	0.01	0.55	0.01
		0.01	0.02	1.01	0.00	0.92	0.01
			0.1	1.02	0.00	0.81	0.01
			0.5	1.02	0.00	0.59	0.02

($t_{10} = 1.95$, $P = 0.08$) and 0.43 when $N = 500$ ($t_9 = 1.43$, $P = 0.19$). This correlation was significant when we pooled the results from the two population sizes by dividing the observed gene diversities by the neutral expectations ($r = 0.56$, $t_{21} = 3.12$, $P = 0.005$).

Discussion

Theoretical results

The mean number of deleterious mutations per gamete is expected to be $U/(2hs)$ in infinite populations assuming multiplicative fitness interactions between loci (Kimura & Maruyama, 1966). However, harmful mutations can keep accumulating beyond this in small populations. The evolutionary consequences of deleterious mutations include: (i) deterioration of the genome (Pamilo *et al.*, 1987; Lynch *et al.*, 1995); (ii) inbreeding depression (Charlesworth & Charlesworth, 1987); (iii) background selection and fitness variation among individuals within the population (Charlesworth *et al.*, 1993); (iv) associative overdominance (Ohta, 1973 and this study); and (v) long-term evolutionary fate at linked loci (Stephan *et al.*, 1992; Charlesworth, 1994).

Our results give some support to the hypothesis that associative overdominance in a small population might contribute to the correlation between individual heterozygosity and fitness, even though the phenomenon may, in natural populations, more often result from inbreeding. The correlation depends on the linkage disequilibrium between the markers and background loci, and the strength of this disequilibrium depends on the mutation rates at the two categories of loci, on the recombination rate and on the population size. The association can diminish quickly with an increasing rate of crossing-over and with population size. Furthermore, the correlation is strongest for intermediate selection and high dominance (small h). The extension of Ohta's (1971) model multiplicatively (or additively) to multiple loci assumes that the background loci behave independently of each other. This is not the case in small populations, for which reason the observed effects in our simulations were generally smaller than predicted by the model.

The genomic mutation rate used at the background loci, $U = 0.2$, was somewhat larger than in other similar studies (e.g. Charlesworth *et al.*, 1995) but lower than estimated, e.g. in *Drosophila*, where the estimated haploid rate of harmful mutations per one of the 5000 cytological chromosome bands is $u = 10^{-4}$ (Hudson & Kaplan, 1995). The mutation

rate at the markers (2×10^{-4} per locus per generation) corresponds best with that found at microsatellite loci, and is clearly higher than that found at allozyme loci.

The associative fitness effects depend on the linkage disequilibrium between the loci. The expected length of the homozygous chromosomal segment surrounding a locus at which the genes are identical by descent is approximately $(\log N - 1)/2N$ (Sved, 1971). So, the effect is strongest in small populations. The coefficient of associative overdominance has been predicted by using a model that assumes one neutral and one linked locus with a flux of deleterious mutations (Ohta, 1971). With only one such background locus, there will be no associative overdominance in a specific generation, but taken over generations the fitnesses of the homozygotes will average less than those of heterozygotes. A similar phenomenon arises when there are many background loci. The magnitude of associative overdominance should be positively related to the muta-

tion rate and inversely related to dominance, population size and recombination distance (r). This agrees with our simulation results (we did not vary the mutation rate). Assuming a constant h , s' in Ohta's (1971) model has a maximum with respect to hs at $hs = \sqrt{r/(2N)}$. This means that intermediate selection is expected to lead to strongest effects, also in agreement with our results. This result also agrees with earlier findings that fitness consequences caused by accumulation of harmful mutations are most severe with intermediate selection (Pamilo *et al.*, 1987). Ohta's (1971) model also predicts that the maximal effect is reached by weaker selection when the population size increases and the recombination distance decreases.

At an individual level, associative overdominance in our simulations led to a positive regression of fitness on heterozygosity. The association largely disappeared when crossing-over was allowed within the chromosomes. It should, however, be noted that occasionally linkage disequilibrium between a

Fig. 2 The observed gene diversities (H) at the marker loci for: (a) $N = 100$; and (b) $N = 500$. The broken horizontal lines show the upper and lower 95% confidence limits of the marker gene diversity without background selection, and the boxplots give the observed values with background selection. Each boxplot shows the median (white bar), the middle half of the data points (the box), and the main range of observations (whiskers). The outliers are shown separately.

marker and a nearby background locus could remain for a long period even with recombination, creating a temporary association and increasing the variance among the replicates. This could be seen as a positive skewness of the regression coefficients. The linkage also created clear intrachromosomal correlations between the markers and the background loci (*BACKHET* and *BACKHOM*). Individuals heterozygous at marker loci had clearly reduced frequencies of homozygous background mutations. This shows that the markers reflect genomic heterozygosity better than suggested for unlinked loci, for which the expected correlation would have been $\sqrt{p} = 0.03$ in our simulations (Chakraborty, 1981).

The simulations also gave interesting results concerning population level phenomena in the absence of crossing over. The marker heterozygosity (H) was higher than the neutral expectation particularly when $Nhs < 1$, even when there was no steep regression of fitness on heterozygosity at the individual level. These results agree with the conclusions of Ohta (1973) that associative overdominance could be responsible for maintaining polymorphisms. Strong selection ($Nhs > 1$) in the absence of crossing-over occasionally yielded gene diversities lower than the neutral expectation, reflecting the prediction of Charlesworth *et al.* (1993) that selection reduces the effective population size. However, when crossing-over was allowed, the observed H agreed fairly well with neutral expectations with a slight tendency for elevated values (Fig. 2). This agrees largely with the conclusions of Charlesworth *et al.* (1993) that a recombination distance of 10^{-3} between adjacent loci makes background selection ineffective.

Tests of associative overdominance

Empirical studies have had difficulties in distinguishing among the explanations of functional and associative overdominance (e.g. Pogson & Zouros, 1994). Genotypic correlations resulting in associative overdominance can result from a variety of reasons which have been classified into two categories: correlations caused by gametic correlation or inbreeding and those caused by linkage disequilibrium (Zouros, 1993).

If the observed correlations between individual heterozygosity and fitness traits were directly caused by the small number of loci studied, this would extrapolate to a level of genomic selection too high to be compatible with the evidence from molecular data. Bush *et al.* (1987) showed that the loci and alleles used to measure the correlation do not contribute equally to its strength. Pogson & Zouros

(1994) further found that allozyme loci correlated more than noncoding DNA markers with growth rate in the scallop *Placopecten magellanicus*. This observation does not support the associative overdominance hypothesis, as coding and noncoding markers are expected to be similarly affected by linkage effects.

Our results complement these earlier studies in several ways. First, they indicate that variation at neutral loci can predict the genomic heterozygosity, and that the interlocus correlation of heterozygosities increases when there is a strong linkage disequilibrium between loci. As mentioned above, it has been doubted whether heterozygosity at the studied loci would accurately predict genomic heterozygosity because of the low expected correlation among a set of unlinked loci (Chakraborty, 1981). Secondly, strong linkage can result in associative overdominance. Empirical allozyme studies have, however, generally shown little evidence for such linkage disequilibria (Pogson & Zouros, 1994). Thirdly, our results show that large associative overdominance is expected in small populations. Most examples of the positive correlation between heterozygosity and fitness tend to come from species that either have a heterozygote deficiency (a putative sign of inbreeding) or that live in small and structured populations (Houle, 1989; Zouros & Pogson, 1994). In a simulation study of linkage in partially selfing populations, Charlesworth (1991) came to a similar conclusion, the model of associative overdominance appearing capable of explaining the relationship between heterozygosity and fitness in such populations. A recent study of Beaumont *et al.* (1995) on diploid and triploid mussels found, however, support for the linkage hypothesis of associative overdominance, as heterozygote deficiency was not detected and inbreeding could not explain the observed fitness relationships.

Our simulations showed that associative overdominance caused by harmful recessive mutations affects both the individual level (heterozygote superiority) and the population level (maintenance of genetic variation) features. These results agree qualitatively with those of Ohta (1971, 1973). It is likely that the effects observed here disappear in large populations and with increasing rate of crossing-over. It is clear that background selection reduces the effective population size and, as a consequence, neutral variation at linked loci (Charlesworth *et al.*, 1993; Nordborg *et al.*, 1996). In small populations this effect can be opposed by associative overdominance that increases the marginal fitnesses of marker heterozygotes. As small populations may

not persist very long, it will be important to examine the role of associative overdominance in a set of populations with restricted migration, where isolation takes the role of reducing recombination while the total population size can be large.

Acknowledgements

Our work has been supported by the Swedish Environmental Protection Agency and Carl Trygger's Foundation. We thank the population genetics discussion group in the Uppsala Genetic Centre and the referees for valuable comments, and Dr M. Nei for initially suggesting that it might be worthwhile to model associative overdominance.

References

- BEAUMONT, A. R., FAIRBROTHER, J. E. AND HOARE, K. 1995. Multilocus heterozygosity and size: a test of hypotheses using triploid *Mytilus edulis*. *Heredity*, **75**, 256–266.
- BUSH, R. M., SMOUSE, P. E. AND 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.
- CHAKRABORTY, R. 1981. The distribution of the number of heterozygous loci in an individual in natural populations. *Genetics*, **98**, 461–466.
- CHARLESWORTH, B. 1994. The effects of background selection against deleterious mutations on weakly selected, linked variants. *Genet. Res.*, **63**, 213–227.
- CHARLESWORTH, B., MORGAN, M. T. AND CHARLESWORTH, D. 1993. The effect of deleterious mutations on neutral molecular variation. *Genetics*, **134**, 1289–1303.
- CHARLESWORTH, D. 1991. The apparent selection on neutral marker loci in partially inbreeding populations. *Genet. Res.*, **57**, 159–175.
- CHARLESWORTH, D. AND CHARLESWORTH, B. 1987. Inbreeding depression. *Ann. Rev. Ecol. Syst.*, **18**, 237–268.
- CHARLESWORTH, D., CHARLESWORTH, B. AND MORGAN, M. T. 1995. The pattern of neutral molecular variation under the background selection model. *Genetics*, **141**, 1619–1632.
- DAVID, P. 1997. Modeling the genetic basis of heterosis: tests of alternative hypothesis. *Evolution*, **51**, 1049–1057.
- EWENS, W. J. 1979. *Mathematical Population Genetics*. Springer-Verlag, Berlin.
- FRYDENBERG, O. 1963. Population studies of a lethal mutant in *Drosophila melanogaster*. I. Behaviour in populations with discrete generations. *Heredity*, **48**, 89–116.
- HOULE, D. 1989. Allozyme-associated heterosis in *Drosophila melanogaster*. *Genetics*, **123**, 789–801.
- HUDSON, R. R. AND KAPLAN, N. L. 1995. Deleterious background selection with recombination. *Genetics*, **141**, 1605–1617.
- KIMURA, M. AND MARUYAMA, T. 1966. The mutational load with epistatic gene interactions in fitness. *Genetics*, **54**, 1337–1351.
- LYNCH, M., CONERY, J. AND BÜRGER, R. 1995. Mutational meltdown in sexual populations. *Evolution*, **49**, 1067–1080.
- MITTON, J. B. 1993. Theory and data pertinent to the relationship between heterozygosity and fitness. In: Thornhill, N. C. (ed.) *The Natural History of Inbreeding and Outbreeding*, pp. 17–41. University of Chicago Press, Chicago.
- NORDBORG, M., CHARLESWORTH, B. AND CHARLESWORTH, D. 1996. The effect of recombination on background selection. *Genet. Res.*, **67**, 159–174.
- OHTA, T. 1971. Associative overdominance caused by linked detrimental mutations. *Genet. Res.*, **18**, 277–286.
- OHTA, T. 1973. Effect of linkage on behavior of mutant genes in finite populations. *Theor. Pop. Biol.*, **4**, 145–162.
- PAMILO, P., NEI, M. AND LI, W.-H. 1987. Accumulation of mutations in sexual and asexual populations. *Genet. Res.*, **49**, 135–146.
- POGSON, G. H. AND ZOUROS, E. 1994. Allozyme and RFLP heterozygosities as correlates of growth rate in the scallop, *Placopecten magellanicus*: a test of the associative overdominance hypothesis. *Genetics*, **137**, 221–231.
- STEPHAN, W., WIEHE, T. H. E. AND LENZ, M. W. 1992. The effect of strongly selected substitutions on neutral polymorphism: analytical results based on diffusion theory. *Theor. Pop. Biol.*, **41**, 237–254.
- SVED, J. A. 1971. Linkage disequilibrium and homozygosity of chromosome segments in finite populations. *Theor. Pop. Biol.*, **2**, 125–141.
- SVED, J. A. 1972. Heterosis at the level of the chromosome and at the level of the gene. *Theor. Pop. Biol.*, **3**, 491–506.
- TURELLI, M. AND GINZBURG, L. R. 1983. Should individual fitness increase with heterozygosity? *Genetics*, **104**, 191–209.
- VENABLES, W. N. AND RIPLEY, B. D. 1994. *Modern Applied Statistics with S-Plus*. Springer, New York.
- ZOUROS, E. 1993. Associative overdominance: evaluating the effects of inbreeding and linkage disequilibrium. *Genetica*, **89**, 35–46.
- ZOUROS, E. AND MALLET, A. L. 1989. Genetic explanations of the growth/heterozygosity correlation in marine mollusks. In: Ryland, J. S. and Tyler, P. A. (eds) *Reproduction, Genetics and Distributions of Marine Organisms*, pp. 317–324. 23rd Eur. Mar. Biol. Symp., Olsen and Olsen, Fredensborg, Denmark.
- ZOUROS, E. AND POGSON, F. H. 1994. Heterozygosity, heterosis and adaptation. In: Beaumont, A. R. (ed.) *Genetics and Evolution of Aquatic Organisms*, pp. 135–146. Chapman and Hall, London.