

THE FITNESS CONSEQUENCES OF MULTIPLE-LOCUS HETEROZYGOSITY:  
THE RELATIONSHIP BETWEEN HETEROZYGOSITY AND GROWTH RATE  
IN PITCH PINE (*PINUS RIGIDA* MILL.)

ROBIN M. BUSH

*Department of Biology, University of Michigan, Ann Arbor, MI 48109*

PETER E. SMOUSE

*Department of Human Genetics and Department of Biology, University of Michigan,  
Ann Arbor, MI 48109*

AND

F. THOMAS LEDIG

*Institute of Forest Genetics, Pacific Southwest Forest and Range Experiment Station,  
U.S.D.A. Forest Service, Berkeley, CA 94701*

**Abstract.**—Positive correlations between measures of “fitness” and the number of electrophoretic loci for which an individual is heterozygous have been observed in many species. Two major hypotheses have been proposed to explain this phenomenon: inbreeding depression and overdominance. Until recently, there has been no way to distinguish between these hypotheses. The overdominance model devised by Smouse (1986) is used here in a reanalysis of Ledig et al.’s (1983) study of heterozygosity and growth rate in eight populations of pitch pine and is contrasted with an inbreeding-depression analysis. Ledig et al. (1983) regressed mean growth rate per heterozygosity class on the number of heterozygous loci, a method of analysis which, although it points to general trends in the data, does not differentiate between hypotheses. The correlations they obtained in four populations were significant only because regressing on the means eliminates most of the sum of squares for error and does not weight the unequally sized heterozygosity classes. Reanalysis of Ledig et al.’s data using individuals, not means, showed no significant correlations between heterozygosity and fitness.

A major assumption of Smouse’s overdominance model is that genetic polymorphism is in part a reflection of selection for heterozygotes at genetic equilibrium. The homozygote for the most frequent allele at a locus should be more fit than a homozygote for a less frequent allele, with the heterozygote superior to both homozygotes. Smouse’s model predicts a negative, linear relationship between fitness and “adaptive distance,” a variable that for a heterozygote is zero and for homozygotes is equal to the inverse of the frequency of the corresponding allele. The adaptive-distance model accounted for between 15% and 50% of the variation in growth rate within eight *P. rigida* population samples by accounting for genotypic differences at eight polymorphic loci. This is over twice as much of the variation in growth rate accounted for by Ledig et al.’s (1983) analysis using individuals. Significant correlations were found between adaptive distance and growth rate in four of the eight populations, but in only two of the populations were more of the partial coefficients negative than positive, as would be predicted by the overdominance hypothesis. The remaining two populations in which correlations were significant did not lend themselves to such clear-cut interpretation, as the majority of the partial coefficients were positive. Positive partial coefficients indicate that the growth rate of the heterozygote is inferior to that of at least one of the homozygotes. The adaptive-distance analysis provides evidence that specific genotypes do play a role in determining growth rate in pitch pine. The correlation between growth rate and adaptive distance increased significantly with the age of the population, possibly reflecting competition subsequent to crown closure.

Received July 17, 1986. Accepted January 15, 1987

Many studies have tested for correlations between heterozygosity and fitness (see Mitton and Grant [1984] for a review). Most commonly, groups of individuals are studied by regressing a surrogate measure of fitness on the proportion of surveyed loci for which those individuals are heterozygous. While some studies have found no signifi-

cant associations between fitness and heterozygosity (Mukai et al., 1974), or even negative correlations (Gaines et al., 1978; Mitton et al., 1981), the majority of these studies have yielded positive correlations (Schaal and Levin, 1976; Singh and Zouros, 1978; Bottini et al., 1979; Mitton et al., 1981; Ledig et al., 1983). While the phenomenon

is not universal, so many positive correlations have been found that the problem warrants further investigation.

Two major hypotheses have been proposed to explain these positive correlations, but there has been no way to attribute the observed correlations to one hypothesis or the other. The first of these two hypotheses, hereafter called the "overdominance" hypothesis, proposes that heterozygosity of the loci under examination, or of closely linked loci, leads to increased fitness. Evidence exists to suggest that heterozygotes for some electrophoretically assayed loci have physiological advantages in particular environments (Koehn, 1969; Koehn et al., 1980; DiMichele and Powers, 1982; Watt et al., 1983). In addition, Turelli and Ginzburg (1983) have shown that under any of several broad forms of balancing selection, one should expect to see a general increase in fitness with increasing heterozygosity.

The second hypothesis, hereafter called the "inbreeding" hypothesis, proposes that the loci under examination are themselves selectively neutral and that the observed heterozygosity is merely indicative of the degree of genomic heterozygosity. Under this hypothesis, relatively low heterozygosity of the assayed loci is correlated with homozygosity for deleterious recessive alleles in the genome as a whole. However, as Mitton and Pierce (1980) and Chakraborty (1981) have shown, heterozygosity calculated from a small number of loci is only loosely correlated with genomic heterozygosity. Their work indicates that the estimates of genomic heterozygosity used in most empirical studies may not be very accurate.

One of the more intriguing studies that has explored the relation between heterozygosity and fitness is that of Ledig, Guries, and Bonefeld (1983), henceforth referred to as LGB, who related heterozygosity level, as determined from 21 segregating electrophoretic loci, to growth rate in pitch pine (*Pinus rigida* Mill.). Growth rate is a fairly good surrogate for fitness in pines, and the large number of loci used offers some possibility of large differences in heterozygosity level. Three of their eight populations exhibited large and significant positive correlations between heterozygosity and growth

rate, four had slightly positive or slightly negative correlations that were not significant, and one showed a significant negative correlation. Because they found no evidence of single-locus relationships between heterozygosity and fitness, they judged the surveyed loci to be selectively neutral, thus rejecting the overdominance hypothesis.

The purpose of this paper is to reexamine the relationship between heterozygosity and growth rate in pitch pine, using the "adaptive distance" model introduced by Smouse (1986). Three central questions will be addressed: 1) Is there a relationship between heterozygosity and growth rate that gives consistent results across populations? 2) Which of the two theoretical models provides the better description of the patterns that exist? 3) Do all loci contribute equally to the correlations obtained, or are some loci more telling than others?

#### MATERIALS AND METHODS

Information on the eight study sites, which spanned the range of pitch pine, along with details of the sampling, mensural, and electrophoretic procedures, and allele frequencies may be found in Guries and Ledig (1978, 1982), Guries et al. (1978), and Ledig et al. (1983).

*Measuring Fitness.*—The measurement of "fitness" for an individual is an exceedingly difficult task. Even if we were able to extract a reliable measure of the "relative reproductive contribution to future generations" for a single individual, fitness is usually defined as the expected value of that measure, an average over all individuals of a single genotype. The deviation of each tree's fitness measure from this expected fitness value is included in our model as the error term, a term which we could only minimize by having many replicates of each genotype.

We have chosen the growth rate of the cross-sectional area of pitch pine trunks (basal area growth rate) as our surrogate measure of fitness for a number of reasons. Pitch pine regenerates following fires, frequently in large numbers. Early mortality is heavy, and by the time the trees reach breast height (the usual measuring point), the potential for considerable selection may have

TABLE 1. Means, standard deviations, and sample sizes for log(MAI) values, adjusted to the age of the youngest tree in eight populations of pitch pine. BR = Blue Ridge, NC; SS = Shawnee State Forest, OH; MS = Michaux State Forest, PA; HM = Helmetta, NJ; BF = Bradley Field, CT; MR = Marconi Station, MA; and ST = St. Chrysostôme, Quebec.

Population	Average log(MAI)	SD log(MAI)	N	Adjusted age
BR	2.23	0.48	36	17
SS	0.66	0.70	31	31
MS	1.36	0.44	35	10
LS	0.47	0.48	47	15
HM	0.88	0.47	55	21
BF	1.18	0.57	44	12
MR	-0.26	0.76	58	11
ST	0.18	0.56	52	10

already occurred. We have no real way to assay the selective impact of this early mortality, but there remain differences in the growth rates of the survivors. The rate of basal-area growth after reaching breast height is correlated with the rapidity with which a seedling reaches breast height, and trees that grow rapidly will outcompete their neighbors. In addition, the crown volume of the adult tree is highly correlated with the basal area (Chisman and Schumacher, 1940), so that trees of larger diameter have more sites for cone and seed production, everything else being equal. A variety of microhabitat factors will also influence the growth of any particular tree, of course, as will the idiosyncratic competitive histories of neighboring individuals. Basal-area growth rate is nevertheless a reasonable first approximation to fitness.

The measure of basal-area growth rate chosen by LGB was the mean annual increment [ $MAI = \pi r^2/t$ ], where  $r$  is the radius of the core, and  $t$  is the age of the core (number of years elapsed since reaching breast height). This is a standard measure of basal-area growth, but it has some limitations. The trees LGB used were of a wide range of ages within populations. MAI varies with the age of the individual, generally increasing for the first few decades of life, then leveling off, and finally decreasing as the tree approaches maturity. LGB attempted to compensate for this feature of MAI by linearly adjusting the MAI of each tree to what

it would be if each tree were as old as the mean age of the trees in its population. This introduces error into the MAI measurement, because the estimates are obtained using a linear model to represent a relationship that generally deviates from linearity.

For our surrogate measure of fitness, we have used log(MAI), measured for each tree when it was the same age as the youngest sampled tree in its population. The growth rates of all trees were thus measured directly from the cores, not estimated. A logarithmic transformation tends to homogenize the variances, which otherwise increase with growth rate. In addition, seed output is proportional to crown volume, and the relationship between basal area and crown volume should be roughly a power function. Therefore, the relationship between log(seed output) and log(radial growth) should be roughly linear. We present the means and standard deviations of log(MAI), along with the sampling ages in Table 1. Note that the standard deviations of the transformed variables do not vary systematically with the mean values.

*Genetic Marker Data.*—The most frequent allele at each of the 21 loci was the same for all populations, with the single exception of the *Mdh-2* locus in the St. Chrysostôme population. All alleles except the most frequent were pooled into an "other alleles" class, thus reducing the problem to a 21-locus, two-allele form. As pointed out by Smouse (1986), this decision makes little difference in the outcome of the analysis, and we are forced to it in any case by sample-size constraints. As it is, 21 loci are still too many for the sample sizes at our disposal, and some further reduction in the dimensionality of the analysis is appropriate. Five of the 21 loci (*Pgi-1*, *Ald-1*, *Fum*, *Gdh*, *Adh*) had a common allele with a frequency greater than 0.99 over all populations; we removed this subset from further consideration on two grounds. First, allelic variation at these virtually monomorphic loci is very probably not maintained by balancing selection. Second, we have no real statistical resolution on such loci, because a population would be made up almost entirely of a single homozygous genotype. In addition, two of the remaining loci (*Got-1*

TABLE 2. Frequencies of the most common allele at each of 15 loci and zygotic sample sizes for each of eight populations of pitch pine. The population codes are the same as those in Table 1.

Population	Locus															N
	Mdh-1	Mdh-2	Pgm-1	Pgm-2	6 Pg-1	6 Pg-2	Lap-1	Lap-2	Got-1	Got-2	Alf-2	Idh	G6pd	Acp	Aco	
BR	1.000	0.702	0.988	1.000	0.798	0.738	0.811	0.893	0.940	0.893	0.988	0.917	0.869	0.988	0.488	42
SS	0.976	0.610	0.976	0.951	0.622	0.902	0.920	0.927	0.963	0.839	0.988	0.963	0.927	0.988	0.756	41
MS	0.974	0.731	0.987	0.962	0.808	0.859	0.962	0.910	0.949	0.962	0.974	0.923	0.987	0.987	0.526	39
LS	0.964	0.775	0.978	0.957	0.768	0.703	0.899	0.775	0.899	0.920	0.971	0.913	0.906	0.942	0.688	69
HM	0.959	0.721	0.992	0.934	0.623	0.598	0.951	0.770	0.893	0.951	1.000	0.959	0.918	0.926	0.598	61
BF	0.978	0.717	0.946	0.989	0.707	0.783	0.815	0.826	0.957	0.880	0.967	0.946	0.935	0.978	0.620	46
MR	0.935	0.484	0.968	1.000	0.758	0.726	0.887	0.750	0.879	0.968	1.000	0.879	0.887	0.944	0.484	62
ST	0.982	0.527	0.991	1.000	0.964	0.827	0.955	0.764	1.000	0.945	0.873	0.636	1.000	0.982	0.836	55

and *Pgi-2*) have been shown to be tightly linked (Guries et al., 1978), so we have deleted the less polymorphic of the two (*Pgi-2*) from our analysis. The current theory (Smouse, 1986) requires that loci not be closely linked. We have therefore limited our analysis to the remaining 15 loci, the allele frequencies for which are shown in Table 2.

*Adaptive Distance.*—The first of the two models to be examined here is the adaptive-distance model, which is based on the assumption of multiplicative overdominance (Smouse, 1986). The adaptive-distance model is a method for inferring the relationship between  $Y = \log(\text{fitness})$ , here represented by  $\log(\text{MAI})$  and  $X = \text{adaptive distance}$ , where the three genotypes of a two-allele locus are assigned adaptive-distance values on the basis of the allele frequencies  $P_A = f(A_1)$  and  $Q_A = f(A_2)$ . Smouse (1986) shows that when  $S_A$  and  $T_A$  are the selective differentials of the  $A_1A_1$  and  $A_2A_2$  homozygotes, respectively, the equilibrium allele frequencies are functions of  $S_A$  and  $T_A$ . If we define an “adaptive-distance” measure as follows:

Genotype	$A_1A_1$	$A_1A_2$	$A_2A_2$
Adaptive distance	$P_A^{-1}$	0	$Q_A^{-1}$
Fitness	$e^{-S_A}$	1	$e^{-T_A}$

then the relation between  $X$  (the adaptive-distance value) and  $Y$  (the logarithm of fitness) will be linear at genetic equilibrium (Fig. 1). The regression coefficient for  $X_A$  is the so called “segregational genetic load” for the locus (Morton et al., 1956),  $\alpha = [S_A T_A / (S_A + T_A)]$ , so that we may write the regression model for the  $j$ th individual as:

$$Y_j = 0 - \alpha X_{Aj} + \epsilon_j, \tag{1}$$

where  $\epsilon_j$  is an error term reflecting both estimation errors in  $Y_j$  and model failure. For the multiplicative overdominance model, the adaptive-distance values for different loci are additive (Smouse, 1986), and (1) can be expanded to the form

$$Y_j = 0 - \alpha X_{Aj} - \beta X_{Bj} - \dots - \kappa X_{Kj} + \epsilon_j. \tag{2}$$

The use of a zero intercept implies the use of a  $K$ -locus heterozygote as the reference

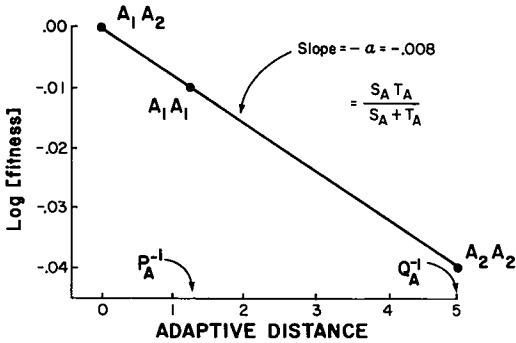


FIG. 1. Example plot of log(fitness) against adaptive distance (AD) for the three possible genotypes at locus *A*. The AD value of the heterozygote is zero. The AD value for each homozygote is equal to the inverse of the equilibrium frequency of the corresponding allele. Here,  $P_A = 0.80$  and  $Q_A = 0.20$ .

genotype. Because we seldom encounter the *K*-locus heterozygote in a finite sample, it is usually necessary to employ another reference genotype, or alternatively to add a nonzero intercept to the model. This nonzero intercept represents a mere shift of scale, however, and does not influence our inference.

For expository purposes, Smouse (1986) assumed that allele frequencies, and therefore the adaptive distances ( $X$ 's), were known without error. In practice, allele frequencies are always estimated with some degree of error; for this study, the imprecision involved in using log(MAI) as a measure of fitness is so much greater than that involved in estimating the allele frequencies that we feel justified in ignoring small imprecision in the estimation of  $X$  values.

The usual binomial estimates of  $P$  and  $Q$  are given by

$$\hat{P} = \frac{Z}{N} \quad \text{and} \quad \hat{Q} = \frac{N - Z}{N} \quad (3a)$$

where  $Z$  is the allelic count for  $A_1$  and  $(N - Z)$  is that for  $A_2$ . Although the estimates  $(1/\hat{P})$  and  $(1/\hat{Q})$  are biased for polymorphic loci, we can largely circumvent this problem if we substitute  $\tilde{P}$  and  $\tilde{Q}$ , defined as follows, for  $\hat{P}$  and  $\hat{Q}$  in (3a):

$$\tilde{P} = \frac{Z + 1}{N + 1}$$

and

$$\tilde{Q} = \frac{N - Z + 1}{N + 1} \quad (3b)$$

The reciprocals of these latter estimates have expectations (Smouse and Chakraborty, 1986):

$$E\left[\frac{1}{\tilde{P}}\right] = \frac{1}{P}[1 - Q^{N+1}]$$

and

$$E\left[\frac{1}{\tilde{Q}}\right] = \frac{1}{Q}[1 - P^{N+1}] \quad (4)$$

Except for very small values of  $P$  or  $Q$  (say  $1/N$ ), the terms in the right-hand brackets are very close to unity. For the allele frequencies we are dealing with in our populations of pitch pine,  $\tilde{Q}$  is quite small in some instances (Table 2), so that  $P^{N+1}$  is not trivial; in these cases our values of  $\tilde{Q}^{-1}$  are underestimates. However, only the rarer homozygote would be assigned this value, and with  $0.01 < \tilde{Q} < 0.02$ , we do not encounter any of these genotypes in the sample.

The regressions of log(MAI) on adaptive distance were done separately for each population, first using all 15 loci together, and then using only the eight loci which contributed most to the regression sum of squares of the 15-locus regressions, when averaged over all populations.

*Inbreeding Method.*—The treatment of the heterozygosity/fitness phenomenon as used by LGB involves a regression of growth rate on the number of heterozygous loci. Two parameters are estimated for this regression: the mean and the slope. The adaptive-distance model requires estimation of one parameter (allele frequency) per locus in addition to the estimation of a grand mean. If the LGB treatment explains the same amount of variability in fitness as does the adaptive-distance model, then the inbreeding hypothesis is the better explanation for the phenomenon, since it requires fewer parameters.

Heterozygote superiority should result in regression lines with negative slopes in the adaptive-distance treatment, so for easier comparison we will use the number of

homozygous loci rather than the number of heterozygous loci for our LGB-type regressions.

### RESULTS

*The Inbreeding Treatment.*—Quite apart from our usage of log(MAI), we have made three additional changes from the analysis of LGB. First, we have pooled all but the most common allele at a locus into a single “other alleles” class. This resulted in moving 12% of the trees to the next highest homozygosity class. Second, we have removed five of the virtually monomorphic loci, and one of a pair of linked loci from consideration, again because of the small sample size. The latter change raised about nine percent of the trees to the next highest homozygosity class. However, these two changes alter the numeric outcome of LGB’s regressions only slightly.

For each of their homozygosity classes, LGB computed an average MAI value. They then conducted an unweighted regression analysis of these averages on homozygosity count, a method we will refer to as an LGB-type regression (Table 3 column d). The residual variation from an analysis of averages, or “lack of fit,” will almost always be smaller than the variation in an analysis of individuals, as a consequence of ignoring the “pure error” component of the variation among trees of the same homozygosity class. In the case of LGB’s data, *F* tests of the lack-of-fit components were not significant, indicating that the linear model itself is adequate; however, the pure-error components amounted to an average of 87% of the residual variation over all populations, indicating that a more conservative procedure would be to include the pure error in the residual variation. In addition, LGB’s regressions were not weighted by sample sizes, which are quite uneven. Consequently, the nominally significant correlations LGB obtained in the Bradley Field and Lebanon populations were mainly due to the analytical effects of single trees that were the sole representatives of their homozygosity classes.

We have reanalyzed LGB’s data using an analysis of individual trees, which we will refer to as “LGB-type regressions using in-

dividuals.” These regressions were done in two different ways: a) by regression on the number of homozygous loci out of all 15 loci listed in Table 2 and b) by regression on the number of homozygous loci out of the eight loci that contributed most to the sums of squares for regression in our adaptive-distance analysis, when totaled over all eight populations (see below). Either way, with the “pure error” restored, the correlations were low (Table 3 column c). Because 15 loci should provide a better estimate of genomic homozygosity than eight loci when testing the inbreeding hypothesis, we shall limit our discussion of the inbreeding method to the 15-locus results. The positive correlations LGB found using means in the Shawnee and Helmetta populations ( $r = +0.90$  and  $r = +0.83$ , respectively; both  $P < 0.05$ ) were reduced to small, positive, nonsignificant correlations when using individuals ( $r = +0.32$  and  $r = +0.21$ ; both ns). As discussed above, the correlations LGB found in the Lebanon and Bradley Field populations ( $r = +0.85$  and  $r = -0.68$ , respectively; both  $P < 0.05$ ) depend heavily on the contributions of a few individuals that represent whole homozygosity classes. It is therefore not surprising that the effects of these few trees were diluted in an analysis of individuals, and LGB’s correlations all but disappear ( $r = +0.12$  and  $r = +0.04$ , respectively; both ns). Over all populations, the LGB-type regressions using individuals explain less than 3% of the variation in basal-area growth, and none of the regressions is significant.

*The Overdominance Treatment.*—Regressions of growth rate on adaptive distance, using all fifteen loci, gave a significant correlation in only one population. The proportional contribution of each of the fifteen loci to the adaptive-distance regression varied among populations, but the greater part of any of the correlations was generally due to only eight of the 15 loci (*Mdh-2*, *Idh*, *Pgm-1*, *Pgm-2*, *G6pd*, *Lap-2*, *Got-1*, and *Ald-2*). We shall focus on the regressions using these eight loci from this point forward (Table 3 column a). Because the subset of eight loci was chosen after looking at the results of the overall 15-locus analysis, the *R* values are slight overestimates and the *P*

TABLE 3. Correlations ( $r$ ) and tail probabilities ( $P$ ) for the regressions of  $\log(\text{MAI})$  values on a) adaptive distance, c) the number of homozygous loci (LGB method), using values for individual trees, and d) the number of homozygous loci (from LGB's 1983 analysis), using mean  $\log(\text{MAI})$  values for each homozygosity class. Column b shows results from an eight-locus LGB-method analysis of individuals in which each locus is analyzed separately. Population codes as in Table 1.

Population	a) Adaptive distance				b) LGB method using separate coefficients		c) LGB method using individuals				d) LGB method using means
	(15 loci)		(8 loci)		(8 loci)		(15 loci)		(8 loci)		(21 loci)
	$R$	$P$	$R$	$P$	$R$	$P$	$r$	$P$	$r$	$P$	$r$
BR	0.76	0.03	0.61	0.04	0.57	0.10	+0.05	0.79	+0.20	0.24	+0.12
SS	0.82	0.06	0.70	0.03	0.68	0.01	-0.32	0.08	-0.50	0.00	-0.90*
MS	0.59	0.77	0.54	0.27	0.54	0.27	+0.07	0.68	+0.16	0.36	+0.23
LS	0.62	0.28	0.47	0.24	0.42	0.44	-0.12	0.42	-0.05	0.71	-0.85*
HM	0.59	0.15	0.53	0.02	0.47	0.09	-0.21	0.13	-0.28	0.04	-0.83*
BF	0.67	0.15	0.58	0.05	0.63	0.01	-0.04	0.79	+0.10	0.54	+0.68*
MR	0.51	0.32	0.43	0.09	0.50	0.02	-0.17	0.20	-0.21	0.12	-0.33
ST	0.53	0.30	0.39	0.16	0.34	0.32	-0.09	0.51	+0.06	0.68	-0.03

\*  $P < 0.05$ .

values are slight underestimates (Cohen and Cohen, 1975). A significant correlation, such as the one found for the Bradley Field population ( $P < 0.05$ ), might therefore be considered to be only marginally significant in this case. The important point, however, is that the adaptive-distance model is considerably more efficacious than is the inbreeding-depression model when the same loci are used. The statistical tests must allow for the fact that more parameters, and hence more degrees of freedom, are used in the adaptive-distance analysis, but even so, treating each locus separately improves the predictive utility of the model considerably.

It is simple to attribute a significant adaptive-distance correlation to either heterozygote or homozygote superiority when using only a single locus; the regression line has a negative slope when the heterozygote is more fit and a positive slope when both (or in some cases only one) of the homozygotes are more fit. However, the response function resulting from a multiple regression using eight independent variables is eight-dimensional, so partial regression coefficients must be examined (Table 4). The response function was judged to have a "positive" or "negative" slope when these partial regression coefficients were predominantly positive or negative for a population. In our adaptive-distance results, the eight partial regression coefficients are never all positive or all negative in any one pop-

ulation. It can be seen that the sign and magnitude of the partial coefficient of any one locus are not constant from population to population.

In the Shawnee and Helmetta populations, the adaptive-distance model explained over twice as much of the variation in  $\log(\text{MAI})$  as did the LGB method using individual trees ( $R^2 = 0.49$  and  $R^2 = 0.28$ , respectively; both  $P < 0.05$ ). The partial coefficients in these two regressions were predominantly negative (Table 4). The adaptive-distance method resulted in a strong correlation ( $R^2 = 0.37$ ,  $P < 0.05$ ) in the Blue Ridge population, with most of the partial correlations being positive. The correlations obtained for the Bradley Field ( $R^2 = 0.34$ ,  $P < 0.05$ ) and Marconi Station ( $R^2 = 0.18$ ,  $P < 0.10$ ) populations were harder to interpret, as the partial regression coefficients were of mixed sign. No significant correlations were found in the regressions for the Michaux, Lebanon, or St. Chrysostôme populations, in keeping with LGB's earlier results.

That significant adaptive-distance correlations ( $P < 0.05$ ) were obtained in four populations leads us to question the assumption of neutrality, because the genotypes at particular loci were correlated with growth rate. Different loci have different predictive values in different populations, which is not surprising, given the considerable range of habitats sampled and the

TABLE 4. Partial correlation coefficients resulting from the multiple regression of log(MAI) on adaptive distance. Population codes as in Table 1.

Population	Genetic locus							
	<i>Mdh-2</i>	<i>Idh</i>	<i>Pgm-1</i>	<i>Pgm-2</i>	<i>G6pd</i>	<i>Lap-2</i>	<i>Got-1</i>	<i>Ald-2</i>
BR	+0.15	+0.05	+0.24	0.00	+0.43	-0.50	+0.22	+0.41
SS	-0.46	+0.07	-0.20	+0.17	0.00	-0.45	+0.22	-0.36
MS	-0.12	-0.14	+0.34	+0.16	+0.31	+0.08	-0.27	+0.23
LS	+0.10	+0.14	-0.33	+0.15	+0.24	-0.03	-0.02	0.00
HM	+0.16	+0.05	0.00	-0.36	-0.31	-0.07	-0.24	0.00
BF	+0.39	+0.09	+0.24	-0.03	-0.29	+0.07	+0.35	-0.02
MR	-0.26	-0.30	0.00	0.00	-0.08	+0.19	+0.02	0.00
ST	+0.28	+0.03	-0.13	0.00	0.00	-0.16	0.00	+0.20

allele frequency variation already documented by LGB. However, in the populations where the partial correlations were positive, the results do not indicate overdominance; for some of the assayed loci, the heterozygote is growing more slowly, not more rapidly, than at least one of the homozygotes.

The extent to which overdominance plays a role in growth rate is somewhat ambiguous from these results. Our sample sizes were so small that we have very few individuals homozygous for the rarer of the two alleles at a locus. For most loci, our adaptive-distance regressions are based only on the contribution of those individuals homozygous for the common allele and a smaller number of heterozygotes. A positive slope would be entirely compatible with the fitness differences at a locus with rare deleterious recessives.

The gain in predictive value hoped for by adding information about differences in relative allele frequencies was not attainable, as we often lacked the rarer homozygote. This lack of resolution is demonstrated by the fact that if the adaptive distance values (0,  $P^{-1}$ ,  $Q^{-1}$ ) are replaced with 1 when an individual is homozygous and 0 if heterozygous, the results (Table 3 column b) are very similar to those of our adaptive-distance analysis. When the rarer homozygote is completely absent from the sample, the results are identical. With either analysis, however, it is clear that there is added predictive power in treating each locus separately, which argues strongly against a generalized inbreeding model.

The predictive value of the adaptive-dis-

tance model greatly exceeds that of an LGB analysis. To show this advantage visually, we have plotted observed log(MAI) against the predicted log(MAI), drawn both from the adaptive-distance regressions (Fig. 2) and from the LGB regressions (Fig. 3). The raw data are plotted for all populations in Figures 2a and 3a, with numbers indicating multiple data points. Population-specific contour lines are presented in Figures 2b and 3b. For clarity, the contour lines delineate the range of variation in only four of the eight populations, but all population contours are similarly shaped. The LGB model using individuals predicts virtually the same log(MAI) value for all individuals within any one population, regardless of genotype. The inbreeding model is virtually useless; recall that on the average, less than 3% of the variation is explained in this fashion. The adaptive-distance model better predicts the observed growth rates, but the major variation in growth rate is that between populations. Even so, there remains a great deal of (presumably environmental) variation in growth rate within the species that we cannot predict. Still, considering the inherent limitations of the data and sampling frame we have used, the fact that we can account for between 15% and 50% of the variation in growth rate within populations by accounting for particular electrophoretic genotypes at eight polymorphic loci is very encouraging.

## DISCUSSION

The real difference between the inbreeding and overdominance models lies in the relevance of the assayed loci. If, as assumed



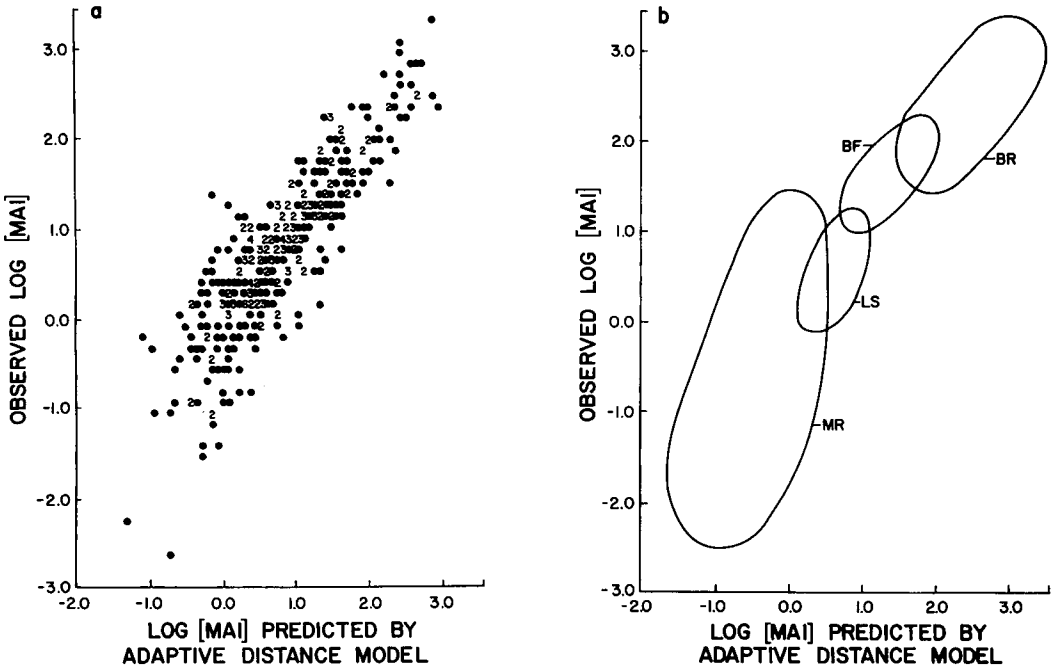


FIG. 2. Observed log(MAI) versus predicted log(MAI) from the adaptive-distance model: a) individual data points for all eight populations, with numbers indicating multiple data points; b) contours delineating the Blue Ridge (BR), Bradley Field (BF), Lebanon (LS), and Marconi Station (MR) populations.

under the inbreeding model, the surveyed loci merely indicate the amount of genomic heterozygosity, applying the overdominance model should give no further information beyond what is expected by chance when including eight irrelevant parameters. The fact that separate treatment of each locus significantly improves the prediction of growth rate is strong evidence that the generalized inbreeding model is not an adequate explanation of the basic phenomenon.

The strength of the correlation between homozygosity and log(MAI) increases significantly with the age at which the population was sampled. If we correlate those correlation coefficients with age, we obtain  $R = +0.77$  ( $P < 0.05$ ) for the adaptive-distance model, and  $r = +0.85$  ( $P < 0.05$ ) for the LGB model. In addition, there was a tendency for the partial correlation coefficients to be more negative in the older populations. There is also some evidence for an increase in heterozygosity with age in populations of conifers (Tigerstedt, 1984; Plessas and Strauss, 1986). Competition for

light, subsequent to crown closure, may be the primary factor responsible for the observed age effect. Said another way, the correlation between heterozygosity and growth rate may increase with age due to the accumulated effects of competition. This is in contrast to a number of studies of marine bivalves, in which the magnitude of the correlation between heterozygosity and growth declines with age (Zouros et al., 1980; Koehn and Gaffney, 1984).

Guries and Ledig (1981), working with the eleven most polymorphic loci from this data set, found significant correlations between six loci and four climatic variables, and argued for the adaptive significance of these loci. Four of these six loci (*Mdh-2*, *Idh*, *G-6-P* and *Got-1*) are among the eight loci that have contributed the most to our adaptive-distance regressions, reinforcing our conclusion that these polymorphic loci are themselves of interest, relative to growth rate.

As pointed out by Smouse (1986), the adaptive-distance analysis does not address the question of whether it is overdominance

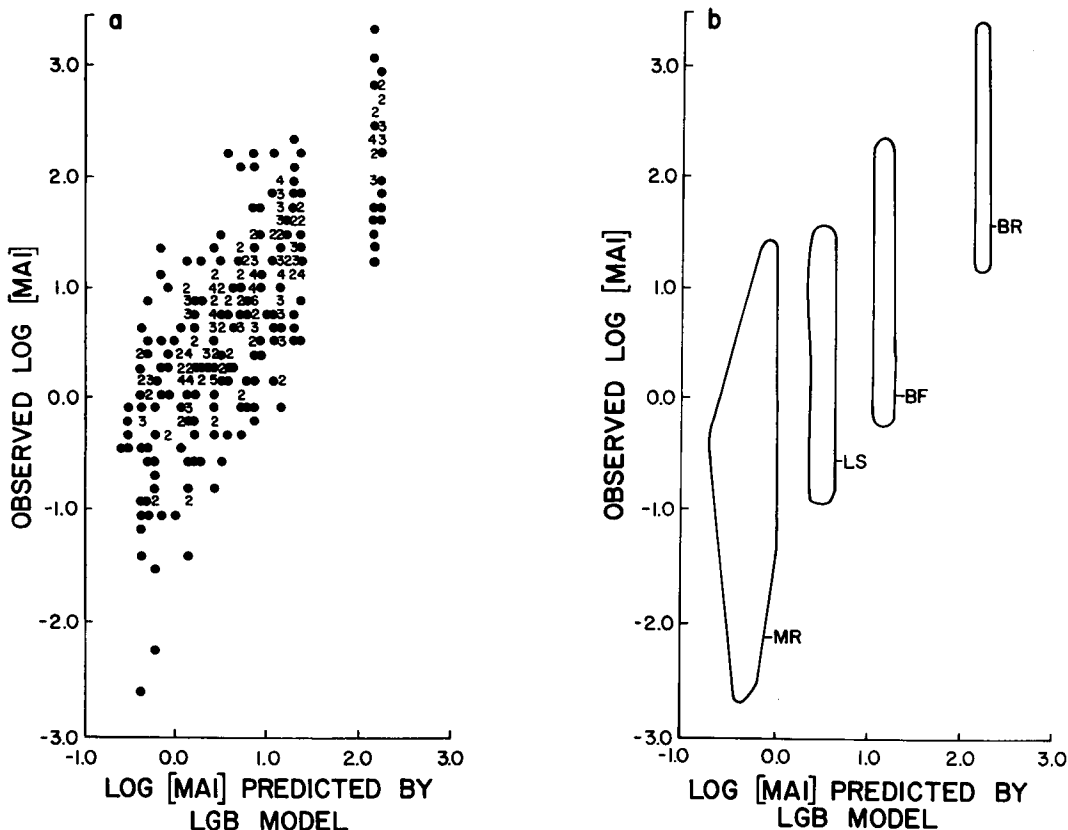


FIG. 3. Observed log(MAI) versus predicted log(MAI) from the LGB model: a) individual data points for all eight populations, with numbers indicating multiple data points; b) contours delineating the Blue Ridge (BR), Bradley Field (BF), Lebanon (LS), and Marconi Station (MR) populations.

of the assayed electrophoretic loci themselves, or associative overdominance of closely linked loci that is responsible for the observed correlations. It is important, however, that the hypothesis of neutrality of the assayed loci (or the segments of chromosome they mark) be rejected before assuming that any kind of locus-specific selection is taking place. We have accomplished that limited objective for some of the loci used in this study.

This study could be improved in several respects. A number of additional measures relating to ultimate reproductive output, such as rate of height growth and rate of cone and seed production, would improve the accuracy of our surrogate fitness measure. While growth rate is a reasonable first approximation to fitness, differences among populations in history and site quality

strongly influence growth rates (Figs. 2, 3), suggesting that there are important environmental influences on growth. Even within populations, growth rate is undoubtedly affected by microhabitat heterogeneity, especially in the vulnerable and unobserved early years of growth. The nature and severity of competition from other plants changes over the lifetime of an individual, affecting growth in ways that are not easy to interpret after the competitors have disappeared. A large population, all planted at the same time and at regular spacing in a fairly uniform habitat, could be expected to yield more accurate fitness measures than those available from natural stands.

A recent analysis by Strauss (1986) of the crossbred and inbred progeny of a group of knobcone pine (*Pinus attenuata*) maternal trees tested the prediction that a correlation

between growth rate and heterozygosity should appear only among inbred progeny if inbreeding depression were the correct explanation of the relationship between growth rate and heterozygosity. This is because the crossbred trees would have few deleterious recessive alleles that were identical by descent, the condition necessary for inbreeding depression. While the correlation between heterozygosity and fitness was stronger among inbred progeny, it was also substantial among the outbred progeny. These results are quite intriguing but do not yield a conclusive resolution of the problem. It would be interesting to subject data such as these to the sorts of analyses we have described above.

Work of this type has been done on primarily outcrossing organisms. Inbreeding organisms supposedly maintain fewer deleterious recessive alleles than outcrossers, yet populations of plants that regularly inbreed are often more heterozygous than would be expected by chance alone (Brown, 1979). Although the theory needed to apply the adaptive-distance model to breeding systems involving other than random mating has yet to be developed, comparison of LGB-type and adaptive-distance models in various organisms having other types of breeding systems should shed some additional light on the relationship between heterozygosity and fitness.

#### ACKNOWLEDGMENTS

The authors express their appreciation to S. Frank, S. Hoffman, S. Strauss, and two anonymous reviewers, all of whom contributed helpful critique on earlier drafts of the manuscript. The responsibility for any remaining errors of omission or commission is our own. During this study, R.M.B. was supported by the University of Michigan Biological Station, P.E.S. was supported by NIH-GM-RO1-32589, and F.T.L. was supported by NSF-DEB-74-11794.

#### LITERATURE CITED

- BOTTINI, E., F. GLORIA-BOTTINI, P. LUCARELLI, A. POLZONETTI, F. SANTORO, AND A. VARVERI. 1979. Genetic polymorphisms and intrauterine development. Evidence of decreased heterozygosity in light-for-dates human newborn babies. *Experientia* 35:1565-1567.
- BROWN, A. D. H., 1979. Enzyme polymorphism in plant populations. *Theoret. Popul. Biol.* 15:1-42.
- CHAKRABORTY, R. 1981. The distribution of the number of heterozygous loci in an individual in natural populations. *Genetics* 98:461-466.
- CHISMAN, H. H., AND F. X. SCHUMACHER. 1940. On the tree-area ratio and certain of its applications. *J. Forest.* 38:311-317.
- COHEN, J., AND P. COHEN. 1975. *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences.* Wiley, N.Y.
- DI MICHELE, L., AND D. A. POWERS. 1982. Physiological basis for swimming endurance differences between LDH-B genotypes of *Fundulus heteroclitus*. *Science* 216:1014-1016.
- GAINES, M. S., L. R. MCCLENNAGHAN, JR., AND R. K. ROSE. 1978. Temporal patterns of allozymic variation in fluctuating populations of *Microtus ochrogaster*. *Evolution* 32:723-739.
- GURIES, R. P., S. T. FRIEDMAN, AND F. T. LEDIG. 1978. A megagametophyte analysis of genetic linkage in pitch pine (*Pinus rigida* Mill.). *Heredity* 40:309-314.
- GURIES, R. P., AND F. T. LEDIG. 1978. Inheritance of some polymorphic isoenzymes in pitch pine (*Pinus rigida* Mill.). *Heredity* 40:27-32.
- . 1981. Genetic structure of populations and differentiation in forest trees, pp. 30-37. *In* M. T. Conkle (tech. coord.), *Proceedings of the Symposium on Isozymes of North American Forest Trees and Forest Insects.* Pacific Southwest Forest and Range Experiment Station, Berkeley, CA.
- . 1982. Genetic diversity and population structure in pitch pine (*Pinus rigida* Mill.). *Evolution* 36:387-402.
- KOEHN, R. K. 1969. Esterase heterogeneity: Dynamics of a polymorphism. *Science* 163:943-944.
- KOEHN, R. K., B. L. BLYNE, M. N. MOORE, AND J. F. SIEBENALLER. 1980. Salinity related physiological and genetic differences between populations of *Mytilus edulis*. *Biol. J. Linn. Soc.* 14:319-334.
- KOEHN, R. K., AND P. M. GAFFNEY. 1984. Genetic heterozygosity and growth in *Mytilus edulis*. *Mar. Biol.* 82:1-7.
- LEDIG, F. T., R. P. GURIES, AND B. A. BONEFELD. 1983. The relation of growth to heterozygosity in pitch pine. *Evolution* 37:1227-1238.
- MITTON, J. B., AND M. C. GRANT. 1984. Associations among protein heterozygosity, growth rate and developmental homeostasis. *Ann. Rev. Ecol. Syst.* 15:479-499.
- MITTON, J. B., P. KNOWLES, K. B. STURGEON, Y. B. LINHART, AND M. DAVIS. 1981. Associations between heterozygosity and growth rate variables in three western forest trees, pp. 27-34. *In* M. T. Conkle (tech. coord.), *Proceedings of the Symposium on Isozymes of North American Forest Trees and Forest Insects.* Pacific Southwest Forest Range Experiment Station, Berkeley, CA.
- MITTON, J. B., AND B. A. PIERCE. 1980. The distribution of individual heterozygosity in natural populations. *Genetics* 95:1043-1054.
- MORTON, N. E., J. F. CROW, AND H. J. MULLER. 1956. An estimate of the mutational damage in man from

- data on consanguineous marriages. *Proc. Nat. Acad. Sci. USA* 42:855-863.
- MUKAI, T., J. K. WATANABE, AND O. YAMAGUCHI. 1974. The genetic structure of natural populations of *Drosophila melanogaster*. XII. Linkage disequilibrium in a large local population. *Genetics* 77:771-793.
- PLESSAS, M. E., AND S. H. STRAUSS. 1986. Allozyme differentiation among population, stands, and cohorts in Monterey pine. *Can. J. For. Res.* 16:1155-1164.
- SCHAAL, B. A., AND D. A. LEVIN. 1976. The demographic genetics of *Liatris cylindracea* Michx. (Compositae). *Amer. Natur.* 110:191-206.
- SINGH, S. M., AND E. ZOUROS. 1978. Genetic variation associated with growth rate in the American oyster (*Crassostrea virginica*). *Evolution* 32:342-353.
- SMOUSE, P. E. 1986. The fitness consequences of multiple-locus heterozygosity under the multiplicative overdominance and inbreeding depression models. *Evolution* 40:946-957.
- SMOUSE, P. E., AND R. CHAKRABORTY. 1986. The use of restriction fragment length polymorphisms in paternity analysis. *Amer. J. Hum. Genet.* 38:918-939.
- STRAUSS, S. H. 1986. Heterosis at allozyme loci under inbreeding and crossbreeding in *Pinus attenuata*. *Genetics* 113:115-134.
- TIGERSTEDT, P. M. A. 1984. Genetic mechanisms for adaptation: The mating system of Scots pine, pp. 317-322. *In* V. L. Chopra, B. C. Joshi, R. P. Sharma, and H. C. Bansai (eds.), *Genetics: New Frontiers. Proceedings of the XV International Congress of Genetics, Vol. IV. Applied Genetics*. Oxford and IBH, New Delhi, India.
- TURELLI, M., AND L. R. GINZBURG. 1983. Should individual fitness increase with heterozygosity? *Genetics* 104:191-209.
- WATT, W. B., R. C. CASSIN, AND M. S. SWAN. 1983. Adaptation at specific loci. III. Field behavior and survivorship differences among *Colias* PGI genotypes are predictable from in vitro biochemistry. *Genetics* 103:725-739.
- ZOUROS, E., S. M. SINGH, AND H. E. MILES. 1980. Growth rate in oysters: An overdominant phenotype and its possible explanations. *Evolution* 34:856-867.

Corresponding Editor: M. K. Uyenoyama