Genet. Res., Camb. (1997), **70**, *pp.* 215–223. *With* 1 *figure.* Printed in the United Kingdom © 1997 Cambridge University Press

Heterozygosity and growth in the marine bivalve *Spisula ovalis*: testing alternative hypotheses

PATRICE DAVID[1]*, BERNARD DELAY AND PHILIPPE JARNE
Génétique et Environnement, Institut des Sciences de l'Evolution – CC065, Université Montpellier II, 34095 Montpellier Cedex 05, France

(*Received 3 February 1997 and in revised form 21 July 1997*)

Summary

Allozyme-associated heterosis has been repeatedly observed in marine bivalves, but its genetic origin remains debatable. A simple explanation is direct overdominance at the enzyme loci scored. An alternative is associative overdominance due to partial inbreeding, affecting the whole genome. The two hypotheses yield different predictions concerning (i) locus-specific effects, (ii) the relationship between heterozygosity and the variance in fitness, and (iii) the expected form of the relationship between the multilocus genotype and mean fitness. The relationship between heterozygosity and growth, a component of fitness, is here analysed in *Spisula ovalis* (1669 individuals, 9 loci), using statistical models designed to test these predictions. In contrast to most other bivalves, *S. ovalis* shells display clear annual growth lines allowing accurate quantification of individual age and growth. Our results show (i) that there is no evidence for locus-specific effects, (ii) that the variance in growth decreases significantly when heterozygosity increases, and (iii) that growth is better predicted by a genetic variable optimized for inbreeding than by a variable optimized for overdominance. In addition, the heterozygosity–growth relationship displays a significant variation among annual cohorts, being more pronounced in young cohorts. Although the need to pool alleles and the occurrence of null alleles may limit the efficiency of some of the models used (especially for result (iii)), our results suggest that the heterozygosity–growth relationship is due to inbreeding effects.

1. Introduction

Positive correlations between allozyme heterozygosity and fitness traits have frequently been reported in natural populations of various organisms, especially marine molluscs (Mitton & Grant, 1984; Zouros, 1987). However, no consensus has emerged about the origin of these correlations (Zouros & Foltz, 1987). Allozymes themselves may determine fitness differences (direct overdominance), or may reflect heterozygosity at other loci (associative overdominance *sensu* Ohta, 1971). Effects of allozyme loci are difficult to distinguish from those of closely linked loci in gametic disequilibrium (Smouse, 1986; Houle, 1994), but they differ markedly from inbreeding or 'general' effects (David *et al.*, 1995). Indeed, when heterozygosity correlations are generated through partial inbreeding, allozyme loci act as mere markers of the inbreeding level, i.e. their heterozygosity reflects inter-individual variations in genomic heterozygosity (Ohta & Cockerham, 1974).

The hypothesis of direct overdominance may be evaluated by comparing observed relationships with theoretical expectations. These were first given by Smouse (1986), who derived the 'adaptive distance' model (hereafter referred to as the *AD* model) explicitly designed to fit overdominance at multiple loci. The good fit of this model was subsequently interpreted as evidence for overdominance in *Pinus* trees (Bush *et al.*, 1987). However, Houle (1994), reanalysing Smouse's model in the one-locus case, showed that such results were actually expected under inbreeding as well. This shows that no conclusion can be drawn by testing the predictions derived from one hypothesis only. Predictions must be derived from both hypotheses (inbreeding and overdominance).

* Corresponding author.
[1] Current address: University College London, Department of Biology, The Galton Laboratory, Wolfson House, 4 Stephenson Way, London NW1 2HE, UK. Tel: +44 (0171) 387 7050. Fax: +44 (0171) 383 2048. e-mail: p.david@ucl.ac.uk.

Table 1. *Predictions for the relationship between heterozygosity and fitness, based on the hypotheses of overdominance and inbreeding*

Hypothesis	Overdominance	Inbreeding
1. Locus-specific effects	Present	Absent
2. Relationship between heterozygosity and variance of the fitness trait	None	Negative
3. Best linear predictor of the fitness trait	Adaptive distances	γ (relative inbreeding level)

Only divergent predictions are useful to distinguish between the alternatives, and should be used to build appropriate statistical tests. David (1997), following this line of reasoning, extended the analysis of Smouse (1986) and Houle (1994), and derived three predictions (summarized in Table 1).

Prediction (*i*). Allozyme loci are expected to affect fitness differentially (Koehn *et al*., 1988) under overdominance, whereas all loci tend to reflect the same underlying parameter (genomic heterozygosity or inbreeding level) under inbreeding. Therefore loci should have significantly different contributions to the heterozygosity–growth relationship under the former, but not the latter, hypothesis. It might be argued that, under the inbreeding hypothesis, variation in the linked load among neutral markers mimics differences in selection intensity. However, this mainly holds in small populations, where linkage disequilibrium occurs as a consequence of sampling effects. In large populations, the correlation between heterozygosities of markers and selected genes (i.e. the identity disequilibrium) is not greatly affected by linkage, at most by a factor 2 with complete linkage (Weir & Cockerham, 1973). Under the overdominance hypothesis, larger differences are expected. For example, some loci may be involved in crucial metabolic pathways that determine the fitness trait (usually growth), whereas other loci are irrelevant. This approach was followed by Koehn *et al*. (1988), who argued for overdominance on the basis of a correspondence between the contribution of specific enzyme loci to the heterozygosity–size relationship and their metabolic importance, in the bivalve *Mulinia lateralis*. However, no statistical test was provided to show that these contributions were significantly unequal and consistent across different samples.

Prediction (*ii*). Under inbreeding, the mean fitness of a given genotype at marker (enzymatic) loci depends on the proportion of inbred individuals among individuals bearing this genotype. As inbred individuals are less fit than outbred ones, homozygous genotypes, among which the proportion of inbred individuals is relatively high, should have both a lower mean fitness and a higher variance in fitness than heterozygous ones, which contain fewer inbred individuals. We therefore expect, under inbreeding, a negative relationship between heterozygosity and the variance of the fitness trait, paralleled by the positive relationship with the trait mean. In contrast, fitness is supposed to be homogeneous within a given genotype under overdominance: the mean fitness depends on marker heterozygosity, but not the variance.

Prediction (*iii*). When inbreeding is low (even if a reliable estimate of inbreeding is not available), it is possible to compute an index (hereafter referred to as γ) representative of the expected proportion of inbred individuals, for each multilocus genotype. Under inbreeding, γ is the best possible linear predictor of fitness, whereas under overdominance the best possible linear predictors of fitness are the adaptive distances (Smouse 1986). In the one-locus case, the adaptive distance and γ are linearly related, and are equally good predictors of fitness (Houle, 1994). Therefore comparison of the models cannot discriminate between local and general hypotheses. However, with several loci, γ is no longer a linear function of the adaptive distances, and the two models are no longer equivalent (David, 1997). Under overdominance, a linear model based on the adaptive distances (the *AD* model) will explain significantly more variance than a model based on γ, while the reverse is expected under inbreeding.

Evaluating the above predictions involves the use of complex linear models, and requires (a) a large sample size and (b) an accurate quantification of fitness-related traits for each individual (Houle, 1994). Growth rate is a suitable fitness index, usually inferred from body size in bivalves (e.g. Gaffney *et al*., 1990). However, as the age of field-grown individuals is generally unknown, this procedure confounds growth and survival effects (Zouros & Foltz, 1987). In the present study, this pitfall is avoided because the bivalve studied, *Spisula ovalis*, displays annual shell lines, which allow age determination and the quantification of individual growth history using a few parameters. A previous study (David *et al*., 1995) showed that the parameter $t_{1/2}$, the age at which an individual reaches half its maximum size, was negatively correlated with heterozygosity at seven allozyme loci in a sample of 239 individuals. We now extend this study to nine loci and 1669 individuals, from different annual cohorts and sites. This large sample is

https://doi.org/10.1017/S0016672397002978 Published online by Cambridge University Press

used to test the three predictions above, in order to evaluate the competing hypotheses of overdominance and inbreeding. In addition, our dataset allows tests of the consistency of the heterozygosity–growth correlation in space (across sites) and time (across cohorts), in contrast to previous studies on marine molluscs.

2. Material and methods

Individuals of *S. ovalis* were dredged in 1993, 1994 and 1995 in three sites (A, B, C) in the Glénans archipelago (France). These sites are very close to each other (distance between two sites about 1 km) and ecologically very similar. They consist of shallow subtidal sandflats 2–10 m deep with no seagrass, and all three sites are protected from oceanic swell by numerous surrounding rocks and islands. All individuals were genetically analysed at nine polymorphic enzyme loci using starch gel electrophoresis. Electrophoretic procedures are described in David *et al.* (1995) for *AAT*, *EST*, *GAL*, *IDH*, *LAP*, *PGI* and *PGM*. *PGD* and *LAP2* were run on Tris-citrate pH 8·0 gels using muscle and digestive gland extracts respectively, and stained as described in Pasteur *et al.* (1988). Gene diversity (H_e, using Levene's correction for finite samples), Weir & Cockerham (1984)'s estimator of F_{is} (f), and exact tests for Hardy–Weinberg equilibrium were computed for each locus using the software Genepop 2.0 (Raymond & Rousset, 1995).

Individuals aged 5 yr or more were used, as a minimum of five shell lines are required to ensure an accurate estimation of growth parameters. A few individuals with unclear growth lines were discarded. Distances from the umbo to all shell lines were measured on right valves using an image analyser, and used to determine for each individual the parameters of the Von Bertalanffy growth model:

$$L(t) = L_{\max} \cdot \exp[-t \cdot \log_e(2)/t_{1/2}],$$

where $L(t)$ is the size at age t, $L_{\max}$ is the maximum size, and $t_{1/2}$ is the age at size $L_{\max}/2$ (for the computation of these parameters from shell line measures see David *et al.*, 1995). We focus here on the parameter $t_{1/2}$, which was previously reported to be negatively correlated with heterozygosity in *S. ovalis*, reflecting a positive relationship between heterozygosity and growth (David *et al.*, 1995).

The basic model used to describe the relationship between heterozygosity and growth is a regression of $t_{1/2}$ on the number of heterozygous loci (multiple-locus heterozygosity or *MLH*):

$$t_{1/2} = \alpha\, MLH + \beta + \epsilon, \quad (1)$$

where α and β are respectively the slope and intercept of the regression, and ϵ the error term. In order to take into account differences in locus-specific effects on growth, *MLH* can be decomposed into single-locus heterozygosities (SLH_i, scoring 1 for a heterozygote, 0 for a homozygote, at locus i), yielding a multiple-regression model with L predictor variables (L being the number of loci):

$$t_{1/2} = \alpha_1 SLH_1 + \alpha_2 SLH_2 + \ldots + \alpha_L SLH_L + \beta + \epsilon. \quad (2)$$

It is also possible to account for fitness differences between different homozygous genotypes at a given locus, as such differences are expected under both overdominance and inbreeding (Houle, 1994). The adaptive distance (*AD*) model has been designed to account for this variation at diallelic loci. Its predictor variables are the single-locus adaptive distances (SAD_i), whereby appropriate values are assigned to the three possible genotypes in order to predict relative fitnesses at equilibrium under overdominant selection ($SAD_i = 1/p_i$, $1/(1-p_i)$ and 0 for genotypes A_iA_i, B_iB_i and A_iB_i respectively, where p_i is the frequency of allele A_i; Smouse, 1986). The equation for the *SAD* model is similar to (2), with SAD_i instead of SLH_i. Note that adaptive distances are by construction expected to be negatively related to the fitness trait, when heterozygosity is positively so (Smouse, 1986). Just as the SLH_i values may be summed to produce *MLH*, SAD_i values sum to a multilocus adaptive distance (*MAD*) that can be used in a simple regression (equation (1) with *MAD* instead of *MLH*). The *SAD* and *MAD* models are designed for diallelic loci and adaptive distances cannot be computed for multiple alleles, as given equilibrium allele frequencies do not correspond to a unique set of genotypic relative fitnesses when more than two alleles are present (Smouse, 1986). However, the model is still valid when composite alleles (several different alleles pooled into a single class) are used. We therefore pooled all alleles but the most frequent (A) into a composite allele (B) for each locus, as recommended by Smouse (1986). The four regression models (*MLH*, *SLH*, *MAD*, *SAD*) were computed for the dependent variable $t_{1/2}$. To assess the effect of allele pooling, we computed the *SLH* and *MLH* models both with and without pooling.

According to prediction (i), locus-specific effects are expected under overdominance but not under inbreeding. This can be tested by comparing the explained variances of univariate versus multivariate models (David, 1997). The multivariate models (*SLH*, *SAD*) allow variation in locus-specific effects whereas the corresponding univariate models (*MLH*, *MAD* respectively) do not. Therefore, if there are locus-specific effects, the variance of the fitness trait explained by multivariate models should be significantly larger than the variance explained by univariate models. As the *MLH* and *MAD* models are indeed simplifications of the *SLH* and *SAD* models respectively, the extra variances explained by the latter two models were tested using *F*-tests (Crawley, 1993). Significant *F*-values would therefore indicate locus-

https://doi.org/10.1017/S0016672397002978 Published online by Cambridge University Press

Table 2. *Locus-specific parameters for the sample studied*

Locus	*Na*	*He*	*f*	*p*(HW)	*b*(*SLH*)	*b*(*SAD*)
PGI	7	0·114	0·015	0·669	−0·032	−0·007
PGM	11	0·222	0·098	**0·000**	−0·042	0·000
LAP	12	0·590	0·093	**0·000**	−0·066	0·013
AAT	9	0·347	0·021	0·500	0·014	−0·005
IDH	8	0·461	−0·008	0·429	−0·123**	0·053*
EST	3	0·469	−0·127	**0·000**	−0·054	0·035
GAL	9	0·696	0·118	**0·000**	−0·047	0·008
PGD	11	0·405	0·003	0·708	−0·110*	0·002
LAP2	7	0·620	0·030	**0·007**	0·012	0·005

Na, number of alleles; *He*, expected heterozygosity after Levene's correction for finite samples; *f*, Weir & Cockerham's (1984) estimator of F_{is}; *p*(HW), the significance of departure from Hardy–Weinberg genotypic proportions (exact test). Bold characters denote significant ($< 0{\cdot}05$) *P* values. *b*(*SLH*) and *b*(*SAD*), the partial regression coefficients from the multiple regression of $t_{1/2}$ on *SLH* and *SAD* respectively.
* $P < 0{\cdot}05$; ** $P < 0{\cdot}01$.

specific effects. These tests were performed for both adaptive distance (*MAD*, *SAD*) and heterozygosity (*MLH*, *SLH*) models, because adaptive distances are theoretically more refined predictors of relative fitness, although pooling alleles entails a loss of information compared with simple heterozygosity counts.

According to prediction (ii), a negative relationship between heterozygosity and variance in the fitness trait is expected only under the inbreeding hypothesis. This would result in a decrease of the squared residuals of the heterozygosity–trait regression when heterozygosity increases. We therefore regressed the log-transformed squared residuals of the $t_{1/2}/MLH$ regression on *MLH* values. We also performed the same test using *MAD* instead of *MLH* values (see previous paragraph). Under inbreeding, we expect a negative and a positive slope for the tests using *MLH* and *MAD*, respectively.

Prediction (iii) states that a variable representing the expected inbreeding level of a genotype should predict fitness better than the adaptive distances when inbreeding effects are involved. The computation of such a variable is derived as follows. Under inbreeding, the mean fitness is a negative linear function of z, the expected proportion of inbred individuals within the multilocus genotype considered (David, 1997). In partly inbred populations,

$$z = \gamma s/(1-s+s\gamma), \tag{3}$$

s being the inbreeding rate and γ the ratio of the expected frequency of the genotype considered among inbred over that among outbred individuals. z is proportional to γ when s is small, and γ is then equivalent to z as a linear predictor of mean fitness. However, γ can be computed without knowledge of s. Under partial selfing, and neglecting recurrent inbreeding, γ can be computed as $[1+1/2(1-p_i)/p_i]$ for genotype A_iA_i, p_i being the estimated frequency of allele A_i, and 1/2 for any heterozygous genotype. Assuming linkage equilibrium among loci, single-locus γ values can be multiplied to give the multilocus γ. As *S. ovalis* is dioecious, selfing does not occur. However, limited inbreeding through other kinds of correlated matings is possible, and could generate inbreeding levels equivalent to that of very small s values. The regression of $t_{1/2}$ on γ was computed with and without pooling alleles, as only the pooled version could be compared with the *AD* model. Moreover, the distribution of γ comprised a small number of unexpectedly high values (see Section 3). As linear regressions are very sensitive to outliers for the independent variable, the computations were performed both with and without these individuals.

To test the consistency of the relationship between heterozygosity and growth, the sample was subdivided into cohorts (i.e. all individuals born the same year), sites, or both. A given cohort in a given site will be referred too here as a 'group'. We tested the linear model $t_{1/2} = \alpha_S + \beta\ MLH + \eta_S\ MLH$, where α_S is the effect of S, the subdivision of the sample considered (site, cohort, or group), β is the linear effect of *MLH*, and the $\eta_S\ MLH$ term represents the variation of the linear effect of *MLH* among subdivisions. The significance of each term of the model was tested using standard model simplification procedures (Crawley, 1993).

3. Results

Genetic parameters are given in Table 2. The nine loci scored showed high levels of genetic polymorphism, as revealed by their numbers of alleles (often 10 or more) and gene diversities (around 0·5). Allele frequencies are given in the Appendix. At four of the nine loci (*PGM*, *LAP*, *GAL*, *LAP2*), significant departures from Hardy–Weinberg expectations were associated with heterozygote deficiencies ($f > 0$), whereas one significant heterozygote excess was detected (*EST*).

https://doi.org/10.1017/S0016672397002978 Published online by Cambridge University Press

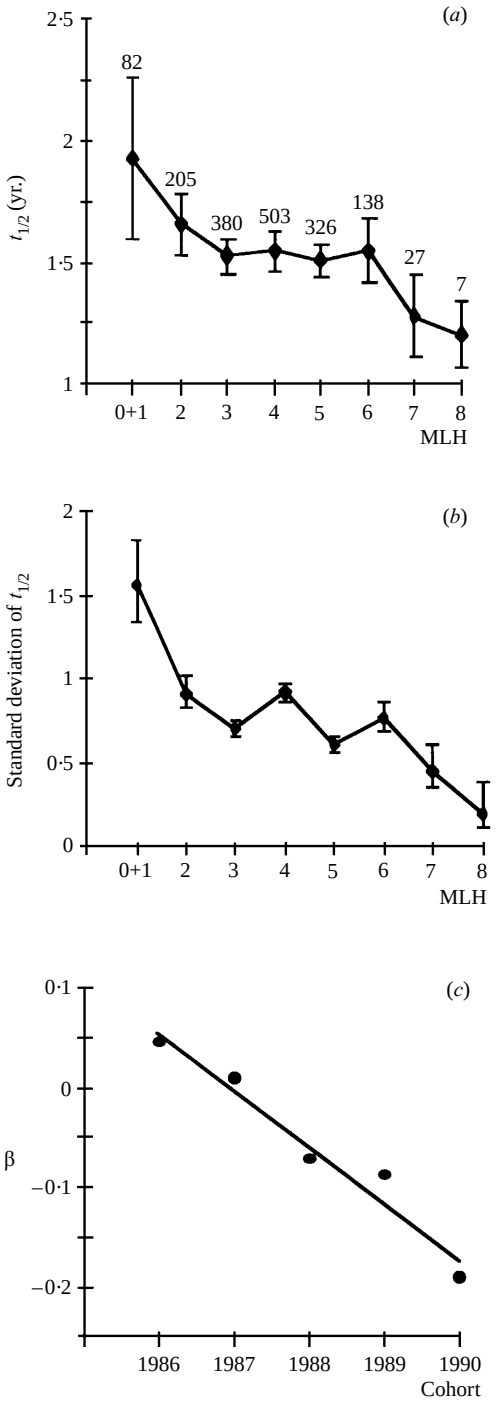

Fig. 1. (*a*) Relationship between multiple-locus heterozygosity (*MLH*) and the growth parameter $t_{1/2}$ in *Spisula ovalis*. Data are presented as means and confidence intervals (95%). The number of individuals in each class is indicated above each point. Classes 0 and 1 were pooled because they displayed very large confidence intervals, due to relatively low sample size and large variance in $t_{1/2}$. (*b*) Relationship between *MLH* and within-class standard deviations of $t_{1/2}$. Same conventions and sample sizes as in (*a*). (*c*) Among-cohort variation of β, the standardized regression coefficient of the *MLH*–$t_{1/2}$ regression line. The correlation between β and birth time (continuous line) is significant ($r^2 = 0{\cdot}9539$, d.f. = 4, $P = 0{\cdot}004$). cohorts 1984 and 1985 were not taken into account as they contained only one and five individuals respectively. Sample sizes were 45, 471, 521, 404 and 222 individuals for cohorts 1986 to 1990 respectively.

Table 3. *Analysis of variance for various regression models on the dependent variable* $t_{1/2}$

Model	*F*	d.f.	*P*
MLH	11·51	1;1667	**0·0007**
MLH_{PO}	7·07	1;1667	**0·0050**
SLH	2·25	9;1659	**0·0170**
SLH_{PO}	1·63	9;1659	0·1004
MAD_{PO}	2·27	1;1667	0·1322
SAD_{PO}	1·00	9;1659	0·4399

The subscript *PO* was added when alleles were pooled before computation (see text). d.f., degrees of freedom. Bold characters denote significant ($< 0{\cdot}05$) *P* values. Slopes of the univariate regressions are $-0{\cdot}053$, $-0{\cdot}044$ and 0·011 for *MLH*, MLH_{PO} and MAD_{PO} models respectively.

$t_{1/2}$ was negatively dependent on *MLH* ($P = 0{\cdot}0007$; Fig. 1*a*), indicating a positive relationship between heterozygosity and growth. The results of the different models are given in Table 3. The *MLH* and *SLH* models were highly significant whereas the *MAD* and *SAD* were not ($P > 0{\cdot}05$). The partial regression coefficients are given in Table 2 for the multivariate models *SAD* and *SLH*. These coefficients represent the contributions of individual loci to the observed relationship. They were not correlated with *f*, the amount of heterozygote deficiency at the same loci ($F_{(1\cdot7)} = 0{\cdot}14$, $P = 0{\cdot}72$). Locus-specific effects were not significant, either using adaptive distances (comparison between *MAD* and *SAD* models: $F_{(8,1659)} = 0{\cdot}84$, $P = 0{\cdot}56$) or using heterozygosities (comparison between *MLH* and *SLH*: $F_{(8,1659)} = 1{\cdot}09$, $P = 0{\cdot}37$, without pooling alleles).

The variance of $t_{1/2}$ decreased with increasing *MLH*, as illustrated in Fig. 1*b*. Indeed, the log-transformed squared residuals or the *MLH* model were strongly negatively dependent on *MLH* (slope $= -0{\cdot}28$, $F_{(1\cdot1667)} = 53{\cdot}72$, $P < 10^{-6}$). When *MAD* was used instead of *MLH*, the dependence was significantly positive (slope $= 0{\cdot}11$, $F_{(1\cdot1667)} = 18{\cdot}59$, $P = 0{\cdot}00002$).

Regression models using γ and *MAD* as predictors of $t_{1/2}$ can be compared when alleles are pooled before computing γ. This comparison is given in Table 4. A few individuals (1·7% of the total sample, or 22 individuals) displayed unexpectedly large γ values ($10 < \gamma \leqslant 181$) and may have disproportionate effects on the regression. When these individuals were removed, the regression of $t_{1/2}$ on γ was highly significant, in contrast to the non-significant regression on adaptive distances (Table 4). The variance explained by γ and *MAD* together was significantly larger than the variance explained by *MAD* alone (Table 4). In other words, the inclusion of γ as a predictor variable significantly increases the goodness of fit of the model, although γ and *MAD* are positively correlated ($r = 0{\cdot}613$, $n = 1669$, $P < 10^{-6}$, alleles pooled for both variables). However, the regressions were not significant when the 1·3% individuals with the largest γ values were included (Table 4).

https://doi.org/10.1017/S0016672397002978 Published online by Cambridge University Press

Table 4. *Comparison of the regressions of* $t_{1/2}$ *on the predictor variables* γ *and MAD*

	γ			*MAD*			$MAD+\gamma^a$	
Predictor	Slope	*F*(d.f.)	*P*	Slope	*F*(d.f.)	*P*	*F*(d.f.)	*P*
All individuals	0·014	3·4 (1, 1667)	0·066	0·011	2·27 (1, 1667)	0·132	1·37 (1, 1666)	0·243
$\gamma < 10$ only	0·063	14·9 (1, 1645)	**0·0001**	0·012	2·59 (1, 1645)	0·107	15·5 (1, 1644)	**0·00008**

Alleles were pooled into two classes prior to analysis, and the models were computed including and excluding individuals with very large γ values (1·3% individuals, cf. Section 3). Significant probabilities are in bold type. d.f., degrees of freedom.
[a] Tests of the significance of the γ term in the bivariate model ($t_{1/2} = \alpha(MAD)+\beta(\gamma)+\delta+\epsilon$) relative to the univariate *MAD* model ($t_{1/2} = \alpha(MAD)+\beta+\epsilon$) using a model simplification procedure (Crawley, 1993).

Table 5. *Linear models of the effect of* MLH *and population subdivision into cohorts, sites or both* (*groups*), *on* $t_{1/2}$

Subdivision	*S* effect			*MLH* effect			Interaction $S \times MLH$		
S	*F*	d.f.	*P*	*F*	d.f.	*P*	*F*	d.f.	*P*
Group	44·85	14; 1653	$< \mathbf{10^{-6}}$	11·30	1; 1652	**0·0008**	5·69	14; 1638	$< \mathbf{10^{-6}}$
Cohort	122·62	5; 1662	$< \mathbf{10^{-6}}$	5·82	1; 1661	**0·0159**	8·59	5; 1656	$< \mathbf{10^{-6}}$
Site	11·07	2; 1665	**0·0047**	11·07	1; 1664	**0·0009**	0·76	2; 1662	0·4659

F-tests were performed using model simplification by subtracting successively the interaction term (η_S *MLH*), the *MLH* main effect (β *MLH* term) and the effect of the 'subdivision' factor (α_S term) from the total model. The associated degrees of freedom (d.f.) and probabilities (*P*) are given. One individual was discarded from the dataset as it was the only representative of its cohort (1984). Bold characters denote significant *P* values.

The regression of $t_{1/2}$ on γ was also performed without pooling alleles, and was significant when outlier ($\gamma > 10$) individuals were excluded (slope = 0·064, $F_{(1\cdot1607)} = 14\cdot71$, $P = 0\cdot0001$) and non-significant when they were included ($F_{(1\cdot1666)} = 0\cdot05$, $P = 0\cdot82$). Another way to reduce the effect of large γ values, instead of removing them, is to use the log-transformation. In this case the regression was significantly positive when all individuals were considered (slope = 0·071, $F_{(1\cdot1666)} = 5\cdot76$, $P = 0\cdot016$).

The sample considered contained six different cohorts (1985–90) collected at three sites. However, the number of groups (a group being a given cohort in a given site) was only 15, as some cohorts were absent from some sites. $t_{1/2}$ displayed some significant variation among groups, cohorts or sites, and the *MLH* main effect was always significant (Table 5). The effect of *MLH* was significantly variable among groups or cohorts, though not among sites, indicating that the heterozygosity–growth correlation is cohort-dependent. Indeed, the *MLH*–$t_{1/2}$ regression was increasingly negative within younger cohorts ($P = 0\cdot004$, Fig. 1 *c*).

4. Discussion

The negative relationship observed between *MLH* and $t_{1/2}$ confirms previous results in *S. ovalis* (David *et al.*, 1995) and is consistent with numerous reports of positive correlations between allozyme heterozygosity and growth or size in bivalves (review in Zouros, 1987). Although significant, the variance explained by *MLH* and other genetic variables is small (of the order of 1%). This order of magnitude is similar to that found in previous bivalve studies (Zouros *et al.*, 1980; Diehl & Koehn, 1985; Gaffney *et al.*, 1990). The statistical power needed to deal with such low signals is provided here by both an age-independent growth index and a large sample size.

$t_{1/2}$ and adaptive distances are expected to be positively correlated under a variety of hypotheses (Houle, 1994). However, neither the multivariate nor the univariate *AD* models were significant. A first explanation could be that growth (as represented by $t_{1/2}$) is not correlated with fitness, violating a basic assumption of the model. However, growth and size have been successfully used as fitness traits for decades in various organisms (see Mitton & Grant, 1984; Bush *et al.*, 1987; Zouros & Foltz, 1987). In bivalves, growth is important both as a way to escape from size-dependent predation (Paine, 1976; Seed & Brown, 1978) and also because reproductive parameters are size-dependent rather than age-dependent (Nakaoka, 1994). Allele pooling might reduce the power of *AD* models. The *SLH* model is indeed significant only without pooling (Table 2). Unfortunately, a multi-allelic adaptive distance is not feasible (Smouse, 1986). We now examine the predictions mentioned in the Section 1.

According to prediction (i), locus-specific effects are expected under overdominance only. They were not observed as the *SAD* model does not explain

https://doi.org/10.1017/S0016672397002978 Published online by Cambridge University Press

significantly more variance than does the *MAD* model. Although the usefulness of this test is limited by allele pooling, this restriction does not apply to the *MLH* and *SLH* models, for which locus-specific effects were not detected either, although both models were individually significant. Although the power may be low, as the explained variance is small, our data therefore provide no support for overdominance. This contrasts with previous results on *Mulinia lateralis* (Koehn *et al*., 1988), in which metabolically important loci (glycolytic or protein catabolic enzymes) had the largest effects on size, although the significance of these differences was not tested.

According to prediction (ii), a positive (or negative) relationship between the variance of the fitness trait and heterozygosity (or adaptive distance) is expected under inbreeding effects. Such a relationship has already been detected in *Crassostrea virginica* (Zouros *et al*., 1980). In *S. ovalis*, it is attested by highly significant regressions with *MLH* and *MAD* residuals (Fig. 1*b*). This reflects the occurrence of more individuals with high $t_{1/2}$ in homozygous classes, rather than a shift of the whole distribution of $t_{1/2}$. This is predicted under inbreeding, as individuals with high $t_{1/2}$ may be unfit inbred individuals. Alternatively, the increased variance may reflect developmental instability resulting directly from enzyme homozygosity. This view has been supported in several species by negative relationships between heterozygosity and fluctuating asymmetry or morphological variance (review in Mitton & Grant, 1984; Palmer & Strobeck, 1986). However, numerous null results have been reported (e.g. Handford, 1980), and additive genetic effects and/or inbreeding may explain the results as well as developmental stability (Chakraborty, 1987). If we assume that enzymic loci directly influence growth, two independent hypotheses are required to explain the effects of heterozygosity on mean and variance of $t_{1/2}$, whereas both are expected consequences of partial inbreeding.

Prediction (iii) states that, under inbreeding, the expected inbreeding level of a genotype (estimated by γ) should be a better predictor of $t_{1/2}$ than is the adaptive distance (*MAD*). Our evidence is only suggestive on this point. γ and *MAD* seem to be equally poor predictors of $t_{1/2}$, as both regressions are non-significant. However, a few outliers with very large γ values bias the whole regression. When they are excluded, the regression on γ, but not on *MAD*, is significantly positive, as expected under inbreeding. The extra variance explained by the model with both γ and *MAD* compared with *MAD* only is significant, indicating that γ indeed improves the predictive accuracy. This is not due to allele pooling, as this result was obtained whether alleles were pooled to compute γ or not. The very large γ values observed are unexpected, indicating highly improbable genotypes (unless extensive inbreeding is assumed), i.e. genotypes homozygous at most loci and/or for rare alleles. Some of these apparent homozygotes could be heterozygotes with a null allele. Indeed, important frequencies of null alleles have been detected in bivalves and may generate heterozygote deficiencies (Foltz, 1986; Gaffney, 1994). Such deficiencies are present in our sample (Table 2), though quite inconsistent across loci, as in previous bivalve studies (see Gaffney *et al*., 1990). They cannot be generated by inbreeding (David *et al*., 1997), and null alleles are a plausible origin. In contrast with the results of Gaffney *et al*. (1990), heterozygote deficiencies are not correlated with the contributions of loci to the heterozygosity–growth relationship. Overall, although our analysis of the relationship between $t_{1/2}$ and γ does not yield a clear-cut result, it is suggestive of inbreeding effects, partially obscured by null alleles.

The consistency of heterozygosity–growth correlations across various conspecific samples has never been tested, although Gaffney (1990) pointed to the lack of repeatability of heterozygosity–viability correlations in *Mytilus edulis*. The heterozygosity–growth correlation in *S. ovalis* varies across cohorts, though not across sites (Table 5). This may originate in environmental and/or genetical variation among cohorts. We now discuss these two hypotheses.

Regardless of their genetic origin, heterozygosity–growth relationships have indeed been suggested to vary with laboratory-induced (Audo & Diehl, 1995) or natural (Borsa *et al*., 1992) levels of environmental stress. However, additional assumptions are required for this hypothesis to account for our data. First, the main temporal pattern is a significant increase in the heterozygosity–growth correlation in young cohorts (Fig. 1*c*). This is not consistent with small, random environmental variation, and can be explained only if we assume long-term, directional change in environmental quality. Second, as there is no site × *MLH* interaction, environmental variation among sites has to be negligible compared with temporal variation. This seems unlikely as (a) the significant main effect of site on $t_{1/2}$ suggests that detectable spatial variations indeed occur, and (b) cohorts n and $n+1$ at a given site have shared a common environmental history for all their life (5 yr or more) except the first year (when cohort n settled and cohort $n+1$ was absent), leaving little room for differential effects on successive cohorts.

Alternatively, the observed variations in *MLH* effects may originate in differences in the genetic background. Heterozygosity–growth correlations generated by associative overdominance depend on the number of segregating deleterious alleles, which may vary among cohorts for two non-exclusive reasons: (a) cohorts come from different gene pools, and (b) the genetic load decreases with age, being purged by natural selection. The former is supported by significant among-cohort population structure in *S. ovalis* (David *et al*., 1997). The latter is supported by more pronounced effects of *MLH* on $t_{1/2}$ in younger cohorts (Fig. 1*c*). On the whole, the observed

pattern allows us to reject the simplest hypothesis of overdominance with a consistent fitness associated with each allozyme genotype. A more sophisticated version of the overdominance hypothesis may include context-dependent selection coefficients (Gillespie & Turelli, 1989), and cannot be excluded. However, this hypothesis is difficult – or impossible – to falsify in natural populations, and we believe that our results are more parsimoniously explained by associative overdominance.

The comparison of observed heterozygosity–growth relationships with theoretical expectations is a valuable tool for inferring the underlying mechanisms. Two restrictions limit the use of such models. First, pooling alleles entails a loss of information. This is of special concern as (a) allozymes in bivalves are very polymorphic (e.g. Gaffney *et al*., 1990; David *et al*., 1995), and (b) a promising approach involves the use of DNA markers (Pogson & Zouros, 1994), such as microsatellites, with potentially numerous alleles. The model using γ is a first step to avoid this problem as multiallelic loci can be dealt with. Second, all models are sensitive to null alleles, observed at both allozymes (Gaffney, 1994) and DNA markers (Hare *et al*., 1996) in bivalves. The *AD* or γ models, penalizing homozygosity for rare alleles, are especially affected because apparent rare homozygotes may often be null heterozygotes. These restrictions mainly apply to prediction (iii).

For tests (i) and (ii), our data fit the predictions of the inbreeding hypothesis better than those of overdominance. The results of test (iii) are, with the abovementioned restrictions, also suggestive of inbreeding effects. The patterns of variation across sites and cohorts are consistent with associative overdominance, although environment-dependent direct effects of enzymatic loci cannot be excluded. The inbreeding hypothesis, predicting most observed patterns, is the most parsimonious. Partial inbreeding has never been documented in marine bivalves, and cannot account for the observed heterozygote deficiencies (cf. Gaffney *et al*., 1990; this study). These deficiencies, whatever their cause, are large enough to mask the potential effects of low inbreeding, which therefore cannot be excluded. Selfing could occur sporadically in hermaphroditic bivalves such as scallops, but biparental inbreeding due to local settlement of related larvae would be a better candidate for most bivalves. The tenuous growth–heterozygosity relationship (about 1% explained variance) could rely on a rather low proportion of inbred individuals (say, of the order of 0·01), provided that they actually suffer from inbreeding depression. Controlled crosses generally revealed severe inbreeding depression in marine bivalves (e.g. Ibarra *et al*., 1995), as expected for any species with naturally low inbreeding levels (Charlesworth & Charlesworth, 1987). More research is needed to identify how low inbreeding could appear in natural populations of bivalves, and to what extent our results are representative of allozyme-associated heterosis in these organisms.

Appendix. Allele frequencies in the sample studied

Allele frequencies are given in order of increasing electrophoretic mobilities; alleles are separated by slashes (/).

PGI.
0·0030/0·0420/0·0009/0·9400/0·0009/0·0126/0·0006.

PGM.
0·0009/0·0003/0·0003/0·0399/0·0027/0·8789/0·0051/0·0618/0·0087/0·0012/0·0003.

LAP.
0·0006/0·0027/0·0111/0·0252/0·0564/0·0045/0·2743/0·0021/0·5740/0·0024/0·0405/0·0063.

AAT.
0·0003/0·0123/0·0009/0·7836/0·0027/0·0006/0·1981/0·0003/0·0012.

IDH.
0·0006/0·0015/0·0003/0·3309/0·0003/0·6553/0·0027/0·0084.

EST. 0·3741/0·0003/0·6256.

GAL.
0·0033/0·0150/0·0426/0·2488/0·0531/0·4739/0·0707/0·0878/0·0048.

PGD.
0·0045/0·0003/0·0384/0·0045/0·7581/0·0003/0·0698/0·0003/0·1199/0·0036/0·0003.

LAP2.
0·0081/0·0917/0·0003/0·3546/0·4937/0·0501/0·0015.

We thank P. Berthou, A. Le Ménac'h and P. Noël for help in collecting clams, M. A. Perdieu, A. F. Pernot, G. Pistre and M. Tomacelli for technical assistance, and two anonymous reviewers for comments on the manuscript. This work was supported by CNRS (UMR 5554) grants. This is contribution no. 97-095 of the Institut des Sciences de l'Evolution.

References

Audo, M. C. & Diehl, W. J. (1995). Effect of quantity and quality of environmental stress on multilocus heterozygosity–growth relationships in *Eisenia fetida* (Annelida: Oligochaeta). *Heredity* **75**, 98–105.

Borsa, P., Jousselin, Y. & Delay, B. (1992). Relationships between allozymic heterozygosity, body size and survival to natural anoxic stress in the palourde, *Ruditapes decussatus* L. (Bivalvia: Veneridae). *Journal of Experimental Marine biology and Ecology* **155**, 169–181.

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.

Chakraborty, R. (1987). Biochemical heterozygosity and phenotypic variability of polygenic traits. *Heredity* **59**, 19–28.

Charlesworth, D. & Charlesworth, B. (1987). Inbreeding depression and its evolutionary consequences. *Annual Review of Ecology and Systematics* **18**, 237–268.
Crawley, M. J. (1993). *GLIM for Ecologists*. Methods in Ecology. Oxford: Blackwell.
David, P. (1997). Modelling the genetic basis of heterosis: tests of alternative hypotheses. *Evolution* **51**, 1049–1057.
David, P., Delay, B., Berthou, P. & Jarne, P. (1995). Alternative models for allozyme-associated heterosis in the marine bivalve *Spisula ovalis*. *Genetics* **139**, 1719–1726.
David, P., Perdieu, M.-A., Pernot, A.-F. & Jarne, P. (1997). Fine-grained spatial and temporal population genetic structure in the marine bivalve *Spisula ovalis*. *Evolution*, **51**, 1318–1322.
Diehl, W. J. & Koehn, R. K. (1985). Multiple-locus heterozygosity, mortality and growth in a cohort of *Mytilus edulis*. *Marine Biology* **88**, 2565–271.
Foltz, D. W. (1986). Null alleles as a possible cause of heterozygote deficiency in the oyster *Crassostrea virginica* and other bivalves. *Evolution* **40**, 869–870.
Gaffney, P. M. (1990). Enzyme heterozygosity, growth rate, and viability in *Mytilus edulis*: another look. *Evolution* **44**, 204–210.
Gaffney, P. M. (1994). Heterosis and heterozygote deficiencies in marine bivalves: more light? In *Genetics and Evolution of Aquatic Organisms* (ed. A. R. Beaumont), pp. 146–153. London: Chapman & Hall.
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.
Gillespie, J. H. & Turelli, M. (1989). Genotype–environment interaction and the maintenance of polygenic variation. *Genetics* **121**, 129–138.
Handford, P. (1980). Heterozygosity at enzyme loci and morphological variation. *Nature* **286**, 261–262.
Hare, M. P., Karl, S. A. & Avise, J. C. (1996). Anonymous nuclear DNA markers in the American oyster and their implications for the heterozygote deficiency phenomenon in marine bivalves. *Molecular Biology and Evolution* **13**, 334–345.
Houle, D. (1994). Adaptive distance and the genetic basis of heterosis. *Evolution* **48**, 1410–1417.
Ibarra, A. M., Cruz, P. & Romero, B. A. (1995). Effects of inbreeding on growth and survival of self-fertilized scallop larvae, *Argopecten circularis*. *Aquaculture* **134**, 37–47.
Koehn, R. K., Diehl, W. J. & Scott, T. M. (1988). The differential contribution by individual enzymes of glycolysis and protein catabolism to the relationship between heterozygosity and growth rate in the coot clam, *Mulinia lateralis*. *Genetics* **118**, 121–130.
Mitton, J. B. & Grant, M. C. (1984). Associations among protein heterozygosity, growth rate, and developmental homeostasis. *Annual Review of Ecology and Systematics* **15**, 479–499.
Nakaoka, M. (1994). Size-dependent reproductive traits of *Yoldia notablis* (Bivalvia: Protobranchia). *Marine Ecology Progress Series* **114**, 129–137.
Ohta, T. (1971). Associative overdominance caused by linked detrimental mutations. *Genetical Research* **18**, 277–286.
Ohta, T. & Cockerham, C. C. (1974). Detrimental genes with partial selfing and effects on a neutral locus. *Genetical Research* **23**, 191–200.
Paine, R. T. (1976). Size-limited predation: an observational and experimental approach with the *Mytilus–Pisaster* interaction. *Ecology* **57**, 858–874.
Palmer, A. R. & Strobeck, C. (1986). Fluctuating asymmetry: measurement, analysis, patterns. *Annual Review of Ecology and Systematics* **17**, 391–421.
Pasteur, N., Pasteur, G., Bonhomme, F., Catalan, J. & Britton-Davidian, J. (1988). *Practical Isozyme Genetics*. Ellis Horwood Ltd, Chichester.
Pogson, G. H. & Zouros, E. (1994). Allozyme and RFLP heterozygosities as correlates of growth rate in the scallop *Placopecten magellanicus*: test of the associative overdominance hypothesis. *Genetics* **137**, 221–231.
Raymond, M. & Rousset, F. (1995). GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity* **86**, 248–249.
Seed, R. & Brown, R A. (1978). Growth as a strategy for survival in two marine bivalves, *Cerastoderma edule* (L.) and *Modiolus modiolus* (L.). *Journal of Animal Ecology* **47**, 283–292.
Smouse, P. E. (1986). The fitness consequences of multiple-locus heterozygosity under the multiplicative overdominance and inbreeding depression models. *Evolution* **40**, 946–957.
Weir, B. S. & Cockerham, C. C. (1973). Mixed self and random mating at two loci. *Genetical Research* **21**, 247–262.
Weir, B. S. & Cockerham, C. C. (1984). Estimating *F*-statistics for analysis of population structure. *Evolution* **38**, 1358–1370.
Zouros, E. (1987). One the relation between heterozygosity and heterosis: an evaluation of the evidence from marine mollusks. *Isozymes* **15**, 255–270.
Zouros, E. & Foltz, D. W. (1987). The use of allelic isozyme variation for the study of heterosis. *Isozymes* **13**, 1–59.
Zouros, E., Singh, S. M. & Miles, H. E. (1980). Growth rate in oysters: an overdominant phenotype and its possible explanations. *Evolution* **34**, 856–867.