

Are polymorphism and evolutionary rate of allozyme proteins limited by mutation or selection?

DAVID O. F. SKIBINSKI*† & ROBERT D. WARD‡

†School of Biological Sciences, University of Wales, Swansea SA2 8PP, U.K. and ‡CSIRO Division of Marine Research, GPO Box 1538, Hobart, Tasmania, 7001, Australia

It is well established that some allozyme proteins are more polymorphic and evolve at a faster rate than others, and that allozyme protein heterozygosity and genetic distance are correlated. These empirical observations could be explained either by differences in mutation rate or by differences in the type or intensity of selection between allozyme proteins. Computer simulation is used here to study a variety of models of allozyme evolution to gauge the potential impact of mutation and selection. The results suggest that the range of models that can account for variation in allozyme heterozygosity and genetic distance through variation in the type or intensity of adaptive directional or overdominant selection is rather narrow, and that many selection models can be rejected as an explanation. However, the variation is easily explained by variation in mutation rate in a wide range of models. The simulation results thus provide support for the view that evolution of allozyme proteins is mutation-limited.

Keywords: allozymes, evolutionary rate, mutation rate, natural selection, polymorphism.

Introduction

It is now well established that some allozyme proteins are more polymorphic and evolve at a faster rate than others. For example, across species, phosphoglucumutase loci have, on average, higher values of heterozygosity and genetic distance than glutamate dehydrogenase loci (e.g. Ward *et al.*, 1992; Skibinski *et al.*, 1993, and references cited therein). In the past, some authors have sought adaptive explanations of such differences by focusing on physiological properties such as substrate specificity (e.g. Gillespie & Kojima, 1968) or on enzyme regulation (e.g. Johnson, 1974). Others have focused on structural features such as subunit size (e.g. Koehn & Eanes, 1977) or quaternary structure (e.g. Ward, 1977), an approach more in line with neutral theory with its emphasis on the importance of selective constraint. In recent years, much effort has been devoted to the analysis of DNA sequences in natural populations and development of a variety of DNA tests for distinguishing neutrality from selection. A recent review suggests that roughly half of such tests applied to individual protein loci refute neutrality

and half do not, thus implicating selection (Moriyama & Powell, 1996). Most of the DNA data are obtained from *Drosophila* species and are still much too few to throw any clear light on the causes of the global differences in heterozygosity and evolutionary rate between allozyme proteins.

A powerful approach for inferring the action of natural selection is the study of the relationship between intraspecific polymorphism and interspecific divergence (Pierce & Mitton, 1979; Skibinski & Ward, 1982; Kimura, 1983; Hudson *et al.*, 1987; Clark, 1994). Neutral theory predicts a positive correlation between polymorphism and evolutionary rate arising from variation in neutral mutation rate. Thus, the observation of a negative correlation or the observation of significant heterogeneity in evolutionary rate for a given amount of polymorphism could suggest selective influences. This approach is the basis of the HKA test (Hudson *et al.*, 1987) applied frequently to DNA data. For allozymes, global positive correlations between average heterozygosity and genetic distance are in line with neutral theory but residual variation suggests some selection (Skibinski & Ward, 1982; Ward & Skibinski, 1985; Skibinski *et al.*, 1993). Although neutral theory does predict a positive correlation and only this, support

*Correspondence. E-mail: d.o.f.skibinski@swansea.ac.uk

for neutral theory from an empirical positive correlation would be weakened if the same prediction could be derived from selection models.

Gillespie (1994) examined the relationship between heterozygosity and evolutionary rate in five models of molecular evolution. One was the neutral model, the other four had different kinds of selection. Gillespie showed that the selection models resemble the neutral model in producing a positive correlation and concluded that the discovery of a positive correlation is therefore not informative for distinguishing between these models. His conclusions have been restated by several authors (Singh & Zeng, 1994; Kreitman & Akashi, 1995; Watt, 1995). Gillespie remarked further that the discovery of a negative correlation could be used to reject all of the models and referred to a *Drosophila* study (Choudhary *et al.*, 1992), where a negative correlation was observed.

When considering such results, it is important to distinguish between the causes of differences between individual loci and the causes of differences between proteins (where the values for proteins are averages over many loci). Thus, one approach to studying the correlation between heterozygosity and evolutionary rate would be to compute a correlation over a sample of allozyme loci in two related species. In this circumstance, selection acting in varied ways could easily produce any kind of relationship between the two variables across the loci. The theoretical expectation of neutral theory is of a positive correlation, given that the loci differ in mutation rate; however, the stochastic (drift) variance associated with individual locus values might be so large as to obscure completely this expectation. In addition, a recent bottleneck could easily generate a negative correlation across loci by converting heterozygosity to genetic distance as observed in human data (Livshits & Nei, 1990). Thus this approach is unlikely to be sufficiently powerful to identify precisely the evolutionary forces responsible for whatever relationship is observed between the variables across loci.

A second approach, relevant to the simulation results of Gillespie (1994), would be to try to minimize as far as possible the stochastic variance, and other errors, by averaging over many replicate loci. For allozyme loci this can be achieved by averaging over species, and leads to the familiar observation, mentioned above, that proteins differ in levels of polymorphism and evolutionary rate. The explanation of these global differences between proteins necessarily requires factors that are protein specific. There are two main possibilities. The differences

between proteins, and the correlation, could result from differences in mutation rate. Alternatively, they could result from differences between proteins in the nature or intensity of selection. The parameter effective population size that determines the extent of genetic drift can be excluded as a cause. This is because proteins, unlike species and populations, do not differ in the value of this parameter. Whatever the explanation for the global variation between proteins, all evolutionary forces — drift, selection, mutation, and gene flow, as well as historical influences — might contribute to the dynamics at individual loci.

In the selection models simulated by Gillespie (1994), selection is represented by a fixed parameter and variation in mutation rate causes the variation in expected heterozygosity, evolutionary rate and the correlation between these two variables. Mutation rate limits the rate of evolution in a variety of theoretical models of molecular evolution (Gillespie, 1991), including nearly neutral models (Ohta, 1976, 1992; Tachida, 1991) and models with strong positive directional selection. However, little attention has been paid to the question of whether variation in the type or intensity of selection could cause a positive correlation between heterozygosity and evolutionary rate. The aim of the present study is to investigate in greater depth the potential roles of mutation and selection as causes of this correlation across allozyme proteins. Computer simulation is used to investigate a variety of selection models and identify the circumstances in which variation in selection or mutation might give rise to realistic correlations.

An important principle of neutral theory is the emphasis on the predominance of mutation pressure (Kimura, 1989). It is thus useful to draw this distinction between selection and mutation as causal factors of global differences between proteins, even if the role of drift and selection, acting at individual loci, is unresolved. An important role for selection would implicate external environmental factors as the cause of the variation between proteins, an adaptive explanation. An important role for mutation would provide a nonadaptive explanation more in keeping with neutral theory (Li, 1997).

Simulation

Models analysed

Although there is still controversy over an appropriate mutational model for allozyme variation, it

seems likely that it should include a stepwise element. The infinite-allele model (Kimura & Crow, 1964) predicts a linear relationship between heterozygosity and genetic distance, which is inconsistent with allozyme data for which the relationship is curvilinear (e.g. Skibinski *et al.*, 1993). Modelling of allozyme variation with the infinite-allele model is inappropriate for other reasons (Barbadilla *et al.*, 1996). Thus a stepwise mutation model (Ohta & Kimura, 1973) was used in the present study for most of the simulations, although infinite-allele and two-allele models were also analysed for some parameter values. The simulations are concerned primarily with directional selection, overdominance and underdominance in isolated populations, although a wide range of other models are also considered (see below).

The models considered a diploid population with a single segregating locus, random mating, and nonoverlapping generations. In each simulation, a single population, initially monomorphic at the locus, was allowed to evolve for T generations under the influence of mutation, drift and selection to allow the population to approach equilibrium. However, some simulations were also carried out using nonequilibrium conditions. The population was then split into two daughter populations, each of the same size as the ancestral population. These two daughter populations then evolved, usually in isolation, for a further t generations. The mean heterozygosity of, and genetic identity between, the two daughter populations was then computed. For each combination of parameter and variable values an 'experiment' of 1000 replicate simulations was made, each replicate being initialized with a different value for the pseudorandom number generator. Heterozygosity according to Hardy-Weinberg equilibrium and genetic identity (Nei, 1972) were used as the measures of polymorphism and evolutionary rate, respectively. Locus heterozygosity and genetic identity were averaged over the 1000 replicate simulations, and these average values, with the average genetic identity converted to genetic distance (Skibinski *et al.*, 1993), plotted on the graphs.

Parameter space

The parameter values used for most simulations were as follows.

(i) Population size of $N = 50$ with $T = 500$ and $t = 360$ or $t = 40$. This gives a realistic value for genetic distance in interspecific ($t = 360$) or intraspecific ($t = 40$) comparisons, given $4Nu$. The results were extended by scaling to larger N for some simu-

lations for which $N = 250$ was used with $T = 1800$ and $t = 1770$.

(ii) Values of $4Nu = 0.2, 0.1$ and 0.01 . These give heterozygosity values in line with mean values reported for allozyme proteins (Skibinski *et al.*, 1993), but the effect of intermediate or more extreme values can be inferred easily from the graphs.

(iii) Values of $Ns = 45, 25, 10, 5, 1, 0.1$ and 0 . An additional value of $Ns = 225$ is used with $N = 250$. With directional selection, the fitness values for each homozygote either (a) fluctuate between 1 and $1-s$, making the model highly parsimonious with respect to the number of selection parameters needed to describe it, or (b) fluctuate between values drawn at random from a normal distribution. Thus, for example, for (a) the fitness of any homozygote will have a certain probability (determined by parameter A , see below) of changing from 1 to $1-s$ or vice versa every generation. In both models (a) and (b), heterozygotes have intermediate fitness. With overdominance, fitness values of homozygotes and heterozygotes are set at $1-s$ and 1 , respectively, and vice versa for underdominance models.

(iv) Three values of a parameter (A) causing periodic fluctuations in fitness for directional selection models. The probability of a change in fitness for a homozygote is $1/A$ each generation, where A takes a constant value of $0.04N$, $4N$ or $400N$ in each experiment. When $A = 0.04N$ with $N = 50$, fluctuations are relatively rapid, occurring on average every second generation; when $A = 4N$, fluctuations occur on a time-scale roughly equal to that of the simulation; and when $A = 400N$ fitnesses are essentially constant. Increasing A can be interpreted as an increasing level of autocorrelation of the environment.

Comparison with other published studies

Two other simulation studies that produced results bearing on the relationship between polymorphism and evolutionary rate used comparable ranges of parameter values. For example, values used by Tachida (1991) were: (i) $N = 50$; (ii) $4Nu = 0.4, 0.2, 0.1$ and 0.04 ; (iii) $Ns = 5.0, 2.5, 0.5, 0.25, 0.05$ and 0 . Values used by Gillespie (1994) were: (i) $N = 1000$; (ii) $4Nu = 5.0, 1.0, 0.5$ and 0.1 ; (iii) $Ns = 25$.

The study of Tachida (1991) uses the house of cards model (Kingman, 1978) where selection coefficients for new mutations are drawn from a normal distribution and remain fixed thereafter. Gillespie (1994) uses the SAS-CFF model. This has fluctuating selection coefficients drawn from a normal distri-

bution. These models use the infinite-allele model rather than the stepwise mutation model. This aside, in the present study the directional selection model with normally distributed fluctuating selection coefficients and $A = 0.04N$ is similar to the SAS-CFF model with the parameter $c = 1$, whereas when $A = 400N$, the model is similar to the house of cards model. Analytical expressions have been derived in strictly neutral theory for the single stepwise mutation model for heterozygosity (Ohta & Kimura, 1973) and genetic distance (Nei & Chakraborty, 1973; Li, 1976) but not for the selection models investigated in the present study.

Although the range of parameter values is in line with that used in previous studies, an attempt has also been made in the present study to investigate in limited depth a range of diverse models. Different mutational models, gene flow, polygenic variation, gene conversion, stabilizing and frequency-dependent selection, interspersed periods of neutrality with selection, and nonequilibrium conditions were all simulated. The purpose of this was to extend the parameter space in a qualitative sense, and it transpires that the conclusions about the impact of $4Nu$, Ns and A on the relationship between heterozygosity and distance are largely robust to these qualitative variations.

Simulation results

The simulation results shown in Fig. 1(a) are for a single-step mutation model with directional selection. The homozygote fitness values fluctuate between two values, 1 and $1-s$, as described above. The three clusters of points with heterozygosity at 0.01, 0.09 and 0.15, corresponding to values of $4Nu$ of 0.01, 0.1 and 0.2, respectively, describe the expected neutral relationship. For each combination of A and $4Nu$, decreasing Ns results in convergence on these clusters. As A decreases from $400N$ to $0.04N$ at a given mutation rate, the relationship between genetic distance and heterozygosity with

increasing Ns moves clockwise in an arc centred on these clusters. When $A = 400N$, selection is purifying; it removes deleterious mutations. (The smaller cluster of points visible at heterozygosity of 0.05 for $4Nu = 0.1$ is the residual neutral variation that arises in this model because 50% of pairs of adjacent

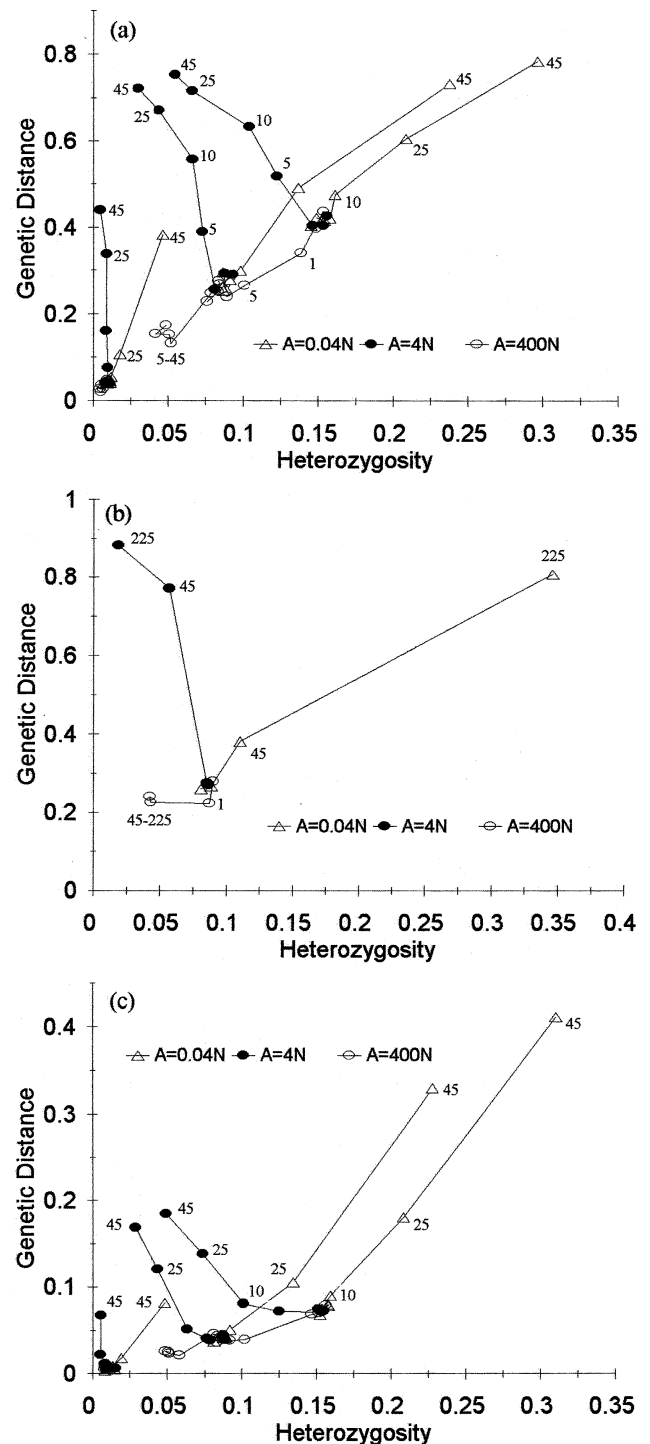


Fig. 1 Relationship between genetic distance and heterozygosity for directional selection in a single-step mutation model with homozygote fitness values fluctuating between 1 and $1-s$. $A = 0.04N$, $4N$, $400N$; $4Nu = 0.01$, 0.1 , 0.2 . Each plotted point is based on 1000 replicate loci (see text). Values of Ns are shown adjacent to many of the plotted points, and can be inferred for other well separated points that are unlabelled. (a) $N = 50$; $t = 360$; $Ns = 45, 25, 10, 5, 1, 0.1, 0$. (b) $N = 250$; $t = 1700$; $Ns = 225, 45, 1, 0$; $4Nu = 0.1$ ($4Nu = 0.01, 0.2$ not simulated). (c) Parameter values as in (a) but with $t = 40$.

alleles in the stepwise series have homozygotes equal in fitness, i.e. $1-s$ or 1 .) When $A = 4N$, the occasional changes in the direction of selection convert accumulated heterozygosity to genetic distance by positive selection, hence the negative correlation. When $A = 0.04N$, increasing Ns increases both heterozygosity and distance by positive selection. With this more rapid fluctuation in the direction of selection the geometric mean fitness of heterozygotes will tend to exceed that of homozygotes, and in this situation the observed maintenance of polymorphism is expected (Haldane & Jayakar, 1963). Qualitatively similar patterns are obtained with a larger population size and stronger selection, up to $Ns = 225$ (Fig. 1b, for $4Nu = 0.1$ alone) and with a shorter divergence time (Fig. 1c). The results suggest that a positive or negative correlation between heterozygosity and genetic distance can be obtained depending on the extent of autocorrelation of the environment. A positive correlation over a realistic range of heterozygosity values for allozyme proteins (0–0.2) can be obtained, with some difficulty, either if both $4Nu$ and A are low or if both $4Nu$ and A are high.

Similar patterns are observed in a diverse variety of other directional selection models. In a fixed selection model, homozygote selection coefficients in the stepwise series are drawn at random from a normal distribution with parameter $2N\sigma$, where σ is the standard deviation of s (Fig. 2). In a polygenic model, the phenotypic value of each allele in the

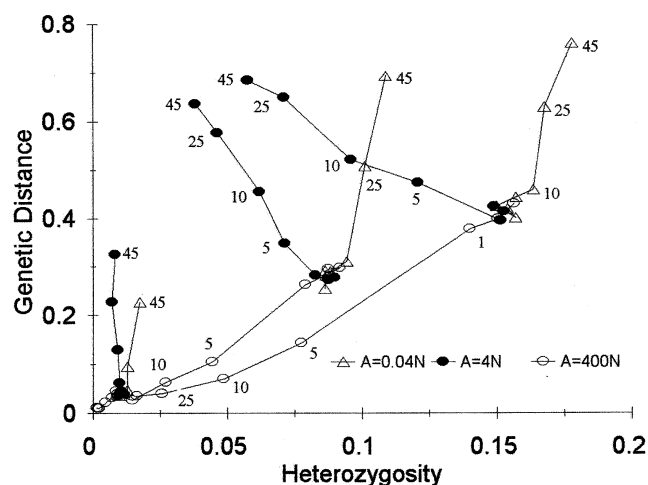


Fig. 2 Relationship between genetic distance and heterozygosity for directional selection in a single-step mutation model with selection coefficients drawn from a fixed normal distribution. $2N\sigma = 45, 25, 10, 5, 1, 0.1, 0$, where σ is standard deviation of s ; $A = 0.04N, 4N, 400N$; $4Nu = 0.01, 0.1, 0.2$; $N = 50$; $t = 360$.

stepwise series is set equal to the additive score for 10 freely recombining diallelic loci each set initially heterozygous $+/-$ with $+$ increasing and $-$ decreasing the score by one unit. Selection acts on the overall scores, which range between -20 and $+20$, as for alleles in a stepwise series. However, heterozygosity and distance are plotted for just one of these diallelic loci (Fig. 3). In a single-step model with gene conversion, heterozygotes chosen as parents have the allele of lower state in the stepwise series converted with the arbitrary probability 0.1 (Fig. 4).

In other simulations (results not shown), the conclusions drawn from consideration of the results of Fig. 1 were robust to variation in population size and divergence time (between sets of parameter values), to fluctuating mutation rate (within experiments), to variation in the mutational model (the infinite-allele model, directionally biased mutation, two-step plus one-step mutation) and to the presence of gene flow ($Nm = 0.5$).

Simulation results for overdominance are shown in Fig. 5. A consistent pattern is evident in all these quite diverse models (although with the two-allele model there is, as expected, evolutionary stasis at heterozygosity of 0.5 with stronger selection). Changing the intensity of selection gives rise to a positive correlation between heterozygosity and genetic distance over a realistic range of heterozygosity values for allozyme proteins only when selection is quite weak ($Ns < 1$). Stronger selection pushes

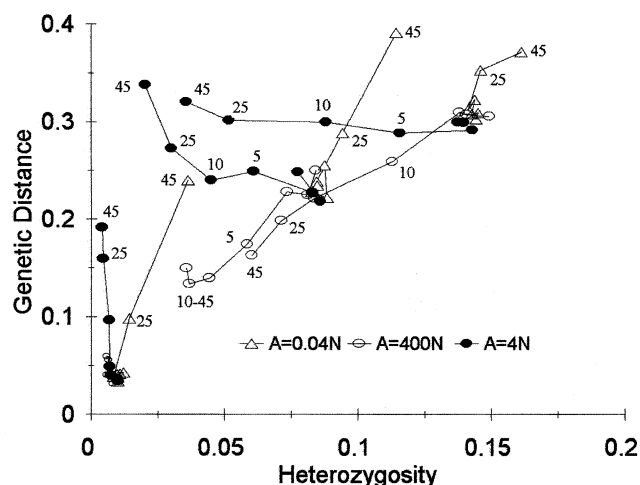


Fig. 3 Relationship between genetic distance and heterozygosity for directional selection in polygenic model with 10 diallelic loci contributing to stepwise series of phenotypic values. $A = 0.04N, 4N, 400N$; $4Nu = 0.01, 0.1, 0.2$; $N = 50$; $t = 360$; $Ns = 45, 25, 10, 5, 1, 0.1, 0$. Values plotted for one of the 10 loci.

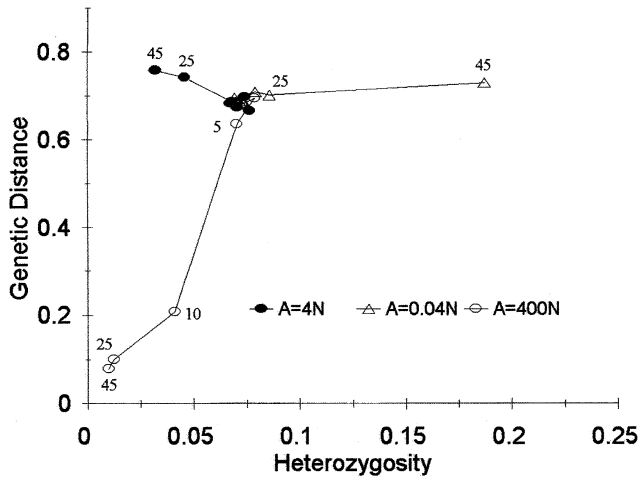


Fig. 4 Relationship between genetic distance and heterozygosity for directional selection in a single-step mutation model with gene conversion. $A = 0.04N$, $4N$, $400N$; $4Nu = 0.01$, 0.1 , 0.2 ; $N = 50$; $t = 360$; $Ns = 45$, 25 , 10 , 5 , 1 , 0.1 , 0 ; conversion probability = 0.1 .

heterozygosity up to unrealistically high levels. This general picture is unaltered with overdominance and gene flow (Fig. 6a) or when overdominance is interspersed with periods of neutrality (Fig. 6b). However, in the latter model, realistic heterozygosity values are obtained with quite strong selection if $4Nu$ is low. In other simulations, the conclusion that selection stronger than about $Ns = 1$ elevates heterozygosity to unrealistically high levels was robust to the variations investigated in the directional selection models (results not shown). In addition, similar conclusions could be drawn for models incorporating differing and fluctuating fitness of homozygotes and models in which periods of overdominance were interspersed with equal periods of directional selection (results not shown).

Simulation results for underdominance and stabilizing selection are shown in Fig. 7. Stabilizing selection in the polygenic model considered above was simulated by giving fitness of one to a zero score ($10+$ and $10-$) and $1-s$ to other scores. In both models an increase in the intensity of selection drives variation from the population. Thus, given that $4Nu$ is sufficiently high initially, these models can result in a positive correlation between heterozygosity and distance over a realistic range of heterozygosity values. The selection is essentially purifying with new mutations being removed from the population. Decreasing Ns is equivalent to an increase in effectively neutral mutation rate. This general pattern obtained with underdominance proved highly robust to the variations considered for

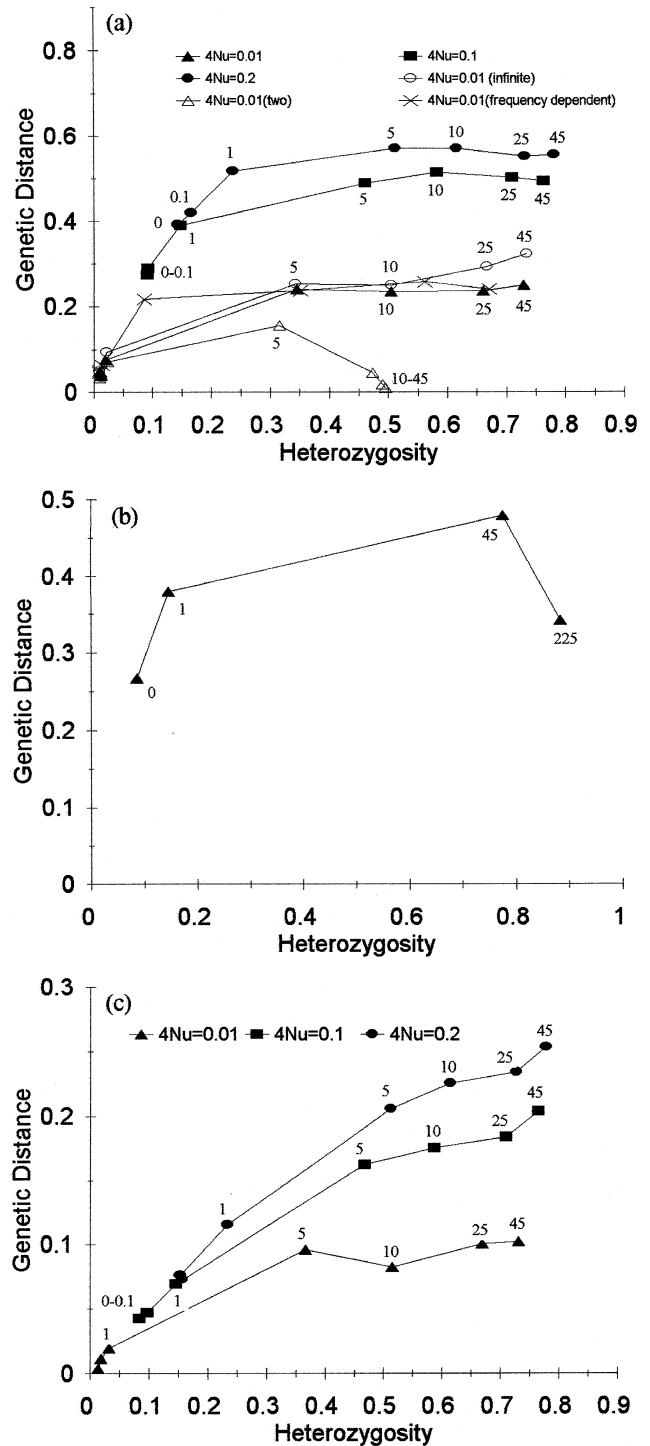


Fig. 5 Relationship between heterozygosity and genetic distance for overdominance. $4Nu = 0.01$, 0.1 , 0.2 . (a) Single-step, infinite-allele, two-allele, and frequency-dependent selection models; $N = 50$; $t = 360$. For frequency dependence, homozygote fitness = $1 - (\text{allele frequency})/w$, where $w = 1.111, 2, 5, 10, 50, 500, 5000$. For other models $Ns = 45, 25, 10, 5, 1, 0.1, 0$. (b) Single-step; $4Nu = 0.1$; $N = 250$; $t = 1700$; $Ns = 225, 45, 1, 0$. (c) Parameter values as in (a) for single-step overdominance but with $t = 40$.

directional selection and overdominance (results not shown).

Simulation results for population contraction or expansion part way through the period of population divergence are shown in Fig. 8. As expected, contraction increases genetic distance at the expense of heterozygosity whereas an expansion has the opposite effect. However, the patterns of points with respect to the relationship between heterozygosity and distance remain essentially the same as in the equilibrium models.

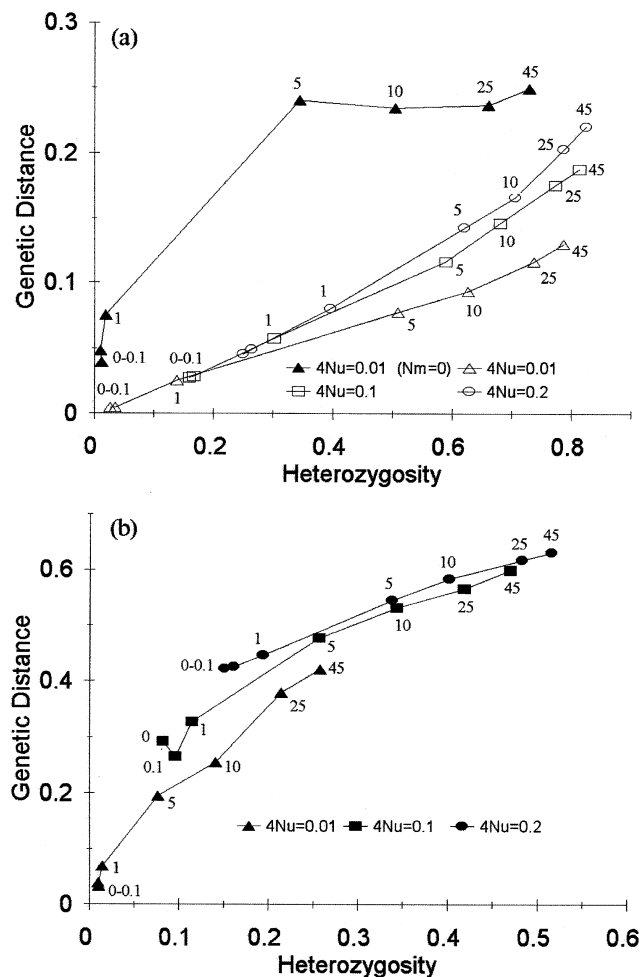


Fig. 6 Relationship between heterozygosity and genetic distance for overdominance in a single-step mutation model. $4Nu = 0.01, 0.1, 0.2$; $N = 50$; $t = 360$; $Ns = 45, 25, 10, 5, 1, 0.1, 0$. (a) Single-step with gene flow ($Nm = 0.5$, where m is migration rate); results for $Nm = 0$ are shown for comparison. (b) Periods of overdominance and neutrality alternate in each replicate (probability of shift between the two modes of selection determined by parameter $A = 4N$).

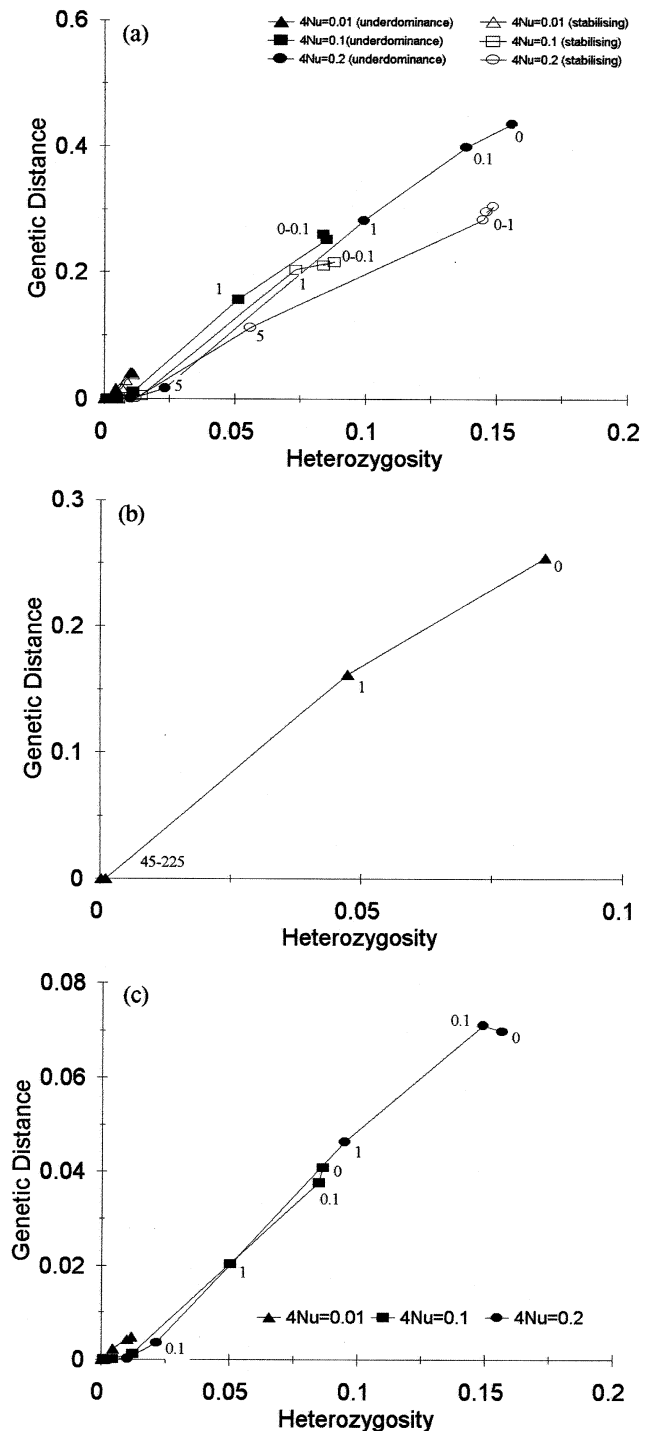


Fig. 7 Relationship between heterozygosity and genetic distance for underdominance and stabilizing selection in a single-step mutation model. $4Nu = 0.01, 0.1, 0.2$. (a) Underdominance and stabilizing selection compared; $N = 50$; $t = 360$; $Ns = 45, 25, 10, 5, 1, 0.1, 0$. (b) Underdominance; $4Nu = 0.1$; $N = 250$; $t = 1700$; $Ns = 225, 45, 1, 0$. (c) Parameter values as in (a) for underdominance but with $t = 40$.

Discussion

This paper is not intended as a contribution to the selectionist–neutralist debate where this concerns an assessment of the relative importance of drift and selection in evolution. Instead, the aim has been to compare the ease with which variation in mutation rate and variation in the type or intensity of selection can explain the observed positive correlation between protein polymorphism and evolutionary rate.

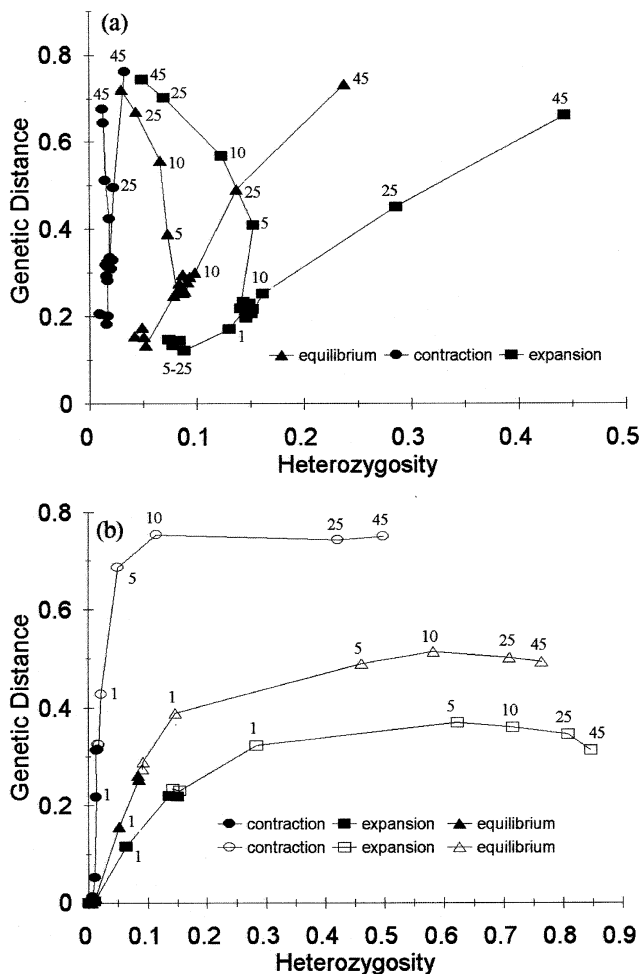


Fig. 8 Nonequilibrium relationship between genetic distance and heterozygosity in a single-step mutation model. N is changed from 50 to 10 (contraction) or from 50 to 250 (expansion) after 260 generations and remains fixed at these new values. $4Nu = 0.1$; $t = 360$; $Ns = 45, 25, 10, 5, 1, 0.1, 0$. (a) Directional selection; $A = 0.04N$, $4N$, $400N$. (b) Overdominance (open symbols) and underdominance (closed symbols). (Note that, unlike Fig. 1 which has different symbols for different values of A , different symbols — squares, triangles and circles — are used here to illustrate the effects of the changing population size.)

The simulation results confirm previous findings that increasing mutation rate causes an increase in heterozygosity and evolutionary rate both in a neutral model and in models with selection (Mukherjee *et al.*, 1987; Tachida, 1991; Gillespie, 1994). To get the same result by varying the intensity of selection is not so straightforward. The simulations suggest two strategies to get a positive correlation. The first is to begin with low mutation rate and heterozygosity and make increasingly intense selection stretch out a positive correlation over a realistic range of heterozygosity values. Figures 1–4 show that this can be achieved with directional selection only with the lowest environmental autocorrelation ($A = 0.04N$). However, with the lowest mutation rate, very strong selection would be needed to produce heterozygosity values as high as 0.2. Figures 5 and 6 show that this strategy might also work with overdominant selection. For example, in Fig. 5(a) with $4Nu = 0.01$, both heterozygosity and genetic distance rise to about 0.2 as the intensity of selection increases from $Ns = 0$ to between $Ns = 1$ and $Ns = 5$. However, if selection is much stronger, heterozygosity is elevated rapidly to unrealistically high values. Thus selection could only be the causal factor within the context of a weak selection ($Ns < 1$) or nearly neutral ($Ns = 1$ approximately) model. In this circumstance, genetic drift would play a major role in the dynamics at individual loci.

The second strategy for selection to give a positive correlation is to begin with high mutation rate and heterozygosity and make increasingly intense selection reduce heterozygosity and genetic distance towards zero. Figure 1 shows that this strategy could be frustrated when some selectively equivalent alleles are present. However, Fig. 2 shows that the strategy can work with directional selection, but only with the highest environmental autocorrelation ($A = 400N$). Figure 7 shows that it can also work with underdominant selection. Both these models are of purifying selection in that selection acts against new mutations and thus depresses variation below the expectation of strict neutrality. Therefore, the models are neutral with the correlation caused in the main by variation in what is essentially effectively neutral mutation rate.

Thus, the possibilities for variation in heterozygosity and genetic distance being caused by adaptive positive directional or balancing selection are limited. Either there must be strong directional selection with a relatively low environmental autocorrelation, or there must be weak overdominant selection. In the latter case drift might dominate selection. There is yet no good empirical evidence

that either of these modes of selection is prevalent, in general, in nature. Directional selection with the moderate level of environmental autocorrelation ($A = 4N$) might seem a plausible candidate *a priori* to explain protein variation. This mode of selection can be generalized as one where periods of evolutionary stasis, in which there is a build of heterozygosity, are interspersed with periods of directional selection when the polymorphism is converted to divergence. Intuitively it might be thought that in this model, more polymorphism should be associated with greater divergence. However, the simulations show that it tends in fact to produce a negative correlation. What is happening is that with strong selection the fate of new mutations is to contribute more to divergence than polymorphism, whereas with weak selection these mutations contribute more to polymorphism. The model is refuted by the observed positive correlation between protein allozyme heterozygosity and genetic distance.

The difficulties with a selective interpretation can be contrasted with the ease with which variation in mutation rate can account for variation between allozyme proteins. The simulation results show that if proteins do differ in mutation rate it is difficult to avoid the conclusion that a large part of protein variation is mutation-limited, whether or not it is neutral.

A positive correlation between protein heterozygosity and genetic distance caused by mutation is consistent with the dynamics at individual loci being determined by selection. For example, in Fig. 1(a), increasing mutation rate with $A = 0.04N$ and $Ns = 45$ gives a positive relationship over the heterozygosity range 0.05–0.3. However, in reality, there would need to be considerable constraint on the Ns values and modes of selection for different proteins to make the values fall on a positive slope with not so much scatter as to obscure the correlation. In general, with strong selection in action, it is not obvious why the selective effects on different proteins should be such as to result fortuitously in a positive correlation. This problem does not arise when $Ns < 1$ for all proteins.

There is now abundant evidence that protein evolution is influenced by selection (e.g. Gillespie, 1991). Analysis of synonymous and nonsynonymous DNA substitutions provides support for the nearly neutral over the strictly neutral theory of molecular evolution (Ohta, 1995). Furthermore, there are now many examples of evidence of selection acting at individual allozyme loci (e.g. Singh & Rhomberg, 1987; Karl & Avise, 1992; Berry & Kreitman, 1993; Johannesson *et al.*, 1995; Pogson *et al.*, 1995; Watt *et*

al., 1996). All this evidence pointing to selection is perfectly compatible with variation in mutation rate being the dominant cause of variation in allozyme protein heterozygosity and genetic distance. The empirical coefficient of determination reported by Skibinski *et al.* (1993) for regression of allozyme protein genetic distance on heterozygosity was about 80% globally. However, significant residual variation in genetic distance was also observed. This cannot be explained by neutral theory but can be explained easily by selection. For example, as shown in the present study, directional selection with occasional changes in direction ($A = 4N$) and overdominance are both effective in causing excess genetic distance over neutral expectation at a given heterozygosity.

In conclusion, the results suggest that mutation is a more promising candidate than adaptive selection for explaining variation in allozyme protein polymorphism and evolutionary rate. In addition, some models of both directional and overdominant selection can definitely be rejected because they lead to predictions inconsistent with empirical data on allozyme variation. Li (1997) in reviewing several lines of evidence from DNA data, argues strongly for the importance of mutation pressure in protein evolution. The results of the present study support the view that allozyme protein evolution should be regarded similarly despite the many case studies suggesting selection acting at individual loci. There are clearly important biological implications if evolution is mutation-limited, irrespective of whether the dynamics at individual loci are dominated by selection or drift. For example, if the relative rates of evolution of quantitative trait loci for different characters were mutation-limited, the relative evolutionary rates of the characters themselves would be constrained within limits. This could represent a major mutational constraint on patterns of phenotypic evolution.

Acknowledgements

We thank Tomoko Ohta for helpful critical comments of an earlier version of this paper.

References

- BARBADILLA, A., KING, L. M. AND LEWONTIN, R. C. 1996. What does electrophoretic variation tell us about protein variation? *Mol. Biol. Evol.*, **3**, 427–432.
- BERRY, A. AND KREITMAN, M. 1993. Molecular analysis of an allozyme cline: alcohol dehydrogenase in *Drosophila melanogaster* on the East Coast of North America. *Genetics*, **134**, 869–893.

- CHOUDHARY, M., COULTHART, M. B. AND SINGH, R. S. 1992. A comprehensive study of genic variation in natural populations of *Drosophila melanogaster*. VI. Patterns and processes of genic divergence between *D. melanogaster* and its sibling species, *Drosophila simulans*. *Genetics*, **130**, 843–853.
- CLARK, A. G. 1994. A neutrality test for continuous characters based on levels of intraspecific variation and interspecific divergence. In: Golding, B. (ed.) *Non-Neutral Evolution*, pp. 101–111. Chapman & Hall, New York.
- GILLESPIE, J. H. 1991. *The Causes of Molecular Evolution*. Oxford University Press, New York.
- GILLESPIE, J. H. 1994. Substitution processes in molecular evolution. II. exchangeable models from population genetics. *Evolution*, **48**, 1101–1113.
- GILLESPIE, J. H. AND KOJIMA, K. 1968. The degree of polymorphism in enzymes involved in energy production compared to that in nonspecific enzymes in two *Drosophila ananassae* populations. *Proc. Natl. Acad. Sci. U.S.A.*, **61**, 582–585.
- HALDANE, J. B. S. AND JAYAKAR, S. D. 1963. Polymorphism due to selection of varying direction. *J. Genet.*, **58**, 237–242.
- HUDSON, R. R., KREITMAN, M. AND AGUADE, M. 1987. A test of neutral molecular evolution based on nucleotide data. *Genetics*, **116**, 153–159.
- JOHANNESSON, K., JOHANNESSON, B. AND LUNDGREN, U. 1995. Strong natural selection causes microscale allozyme variation in a marine snail. *Proc. Natl. Acad. Sci. U.S.A.*, **92**, 2602–2606.
- JOHNSON, G. B. 1974. Enzyme polymorphism and metabolism. *Science*, **184**, 28–37.
- KARL, S. A. AND AVISE, J. C. 1992. Balancing selection at allozyme loci in oysters: implications from nuclear RFLPs. *Science*, **256**, 100–102.
- KIMURA, M. 1983. *The Neutral Theory of Molecular Evolution*. Cambridge University Press, Cambridge.
- KIMURA, M. 1989. The neutral theory of molecular evolution and the world view of neutralists. *Genome*, **31**, 24–31.
- KIMURA, M. AND CROW, J. F. 1964. The number of alleles that can be maintained in a finite population. *Genetics*, **49**, 725–738.
- KINGMAN, J. F. C. 1978. A simple model for the balance between selection and mutation. *J. Appl. Prob.*, **15**, 1–12.
- KOEHN, R. K. AND EANES, W. F. 1977. Subunit size and genetic variation of enzymes in natural populations of *Drosophila*. *Theor. Pop. Biol.*, **11**, 330–341.
- KREITMAN, M. AND AKASHI, H. 1995. Molecular evidence for natural selection. *Ann. Rev. Ecol. Syst.*, **26**, 403–422.
- LI, W. H. 1976. Electrophoretic identity of proteins in a finite population and genetic distance between taxa. *Genet. Res.*, **28**, 119–127.
- LI, W. H. 1997. *Molecular Evolution*. Sinauer Associates, Sunderland, MA.
- LIVSHITS, G. AND NEI, M. 1990. Relationships between intrapopulational and interpopulational genetic diversity in man. *Ann. Hum. Biol.*, **17**, 501–513.
- MORIYAMA, E. N. AND POWELL, J. R. 1996. Intraspecific nuclear DNA variation in *Drosophila*. *Mol. Biol. Evol.*, **13**, 261–277.
- MUKHERJEE, M., SKIBINSKI, D. O. F. AND WARD, R. D. 1987. A simulation study of the neutral evolution of heterozygosity and genetic distance. *Heredity*, **58**, 413–423.
- NEI, M. 1972. Genetic distance between populations. *Am. Nat.*, **106**, 283–292.
- NEI, M. AND CHAKRABORTY, R. 1973. Genetic distance and electrophoretic identity of proteins between taxa. *J. Mol. Evol.*, **2**, 323–328.
- OHTA, T. 1976. Role of slightly deleterious mutations in molecular evolution and polymorphism. *Theor. Pop. Biol.*, **10**, 254–275.
- OHTA, T. 1992. The nearly neutral theory of molecular evolution. *Ann. Rev. Ecol. Syst.*, **23**, 263–286.
- OHTA, T. 1995. Synonymous and nonsynonymous substitutions in mammalian genes and the nearly neutral theory. *J. Mol. Evol.*, **40**, 56–63.
- OHTA, T. AND KIMURA, M. 1973. The model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a genetic population. *Genet. Res.*, **22**, 201–204.
- PIERCE, B. A. AND MITTON, J. B. 1979. Relationship of genetic variation within and among populations: an extension of the Kluge-Kerfoot phenomenon. *Syst. Zool.*, **28**, 63–70.
- POGSON, G. H., MESA, K. A. AND BOUTILIER, R. G. 1995. Genetic population structure and gene flow in the Atlantic cod *Gadus morhua*: a comparison of allozyme and nuclear RFLP loci. *Genetics*, **139**, 375–385.
- SINGH, R. S. AND RHOMBERG, L. R. 1987. A comprehensive study of genic variation in natural populations of *Drosophila melanogaster*. II. Estimates of heterozygosity and patterns of geographic differentiation. *Genetics*, **117**, 255–271.
- SINGH, R. S. AND ZENG, L.-W. 1994. Genetic divergence, reproductive isolation and speciation. In: Golding, B. (ed.) *Non-Neutral Evolution*, pp. 217–232. Chapman & Hall, New York.
- SKIBINSKI, D. O. F. AND WARD, R. D. 1982. Correlations between heterozygosity and evolutionary rate of proteins. *Nature*, **298**, 490–492.
- SKIBINSKI, D. O. F., WOODWARK, M. AND WARD, R. D. 1993. A quantitative test of the neutral theory using pooled allozyme data. *Genetics*, **135**, 233–248.
- TACHIDA, H. 1991. A study on a nearly neutral mutation model in finite populations. *Genetics*, **128**, 183–192.
- WARD, R. D. 1977. Relationship between enzyme heterozygosity and quaternary structure. *Biochem. Genet.*, **15**, 123–135.
- WARD, R. D. AND SKIBINSKI, D. O. F. 1985. Observed relationships between protein heterozygosity and protein genetic distance and comparisons with neutral expectations. *Genet. Res.*, **45**, 315–340.
- WARD, R. D., SKIBINSKI, D. O. F. AND WOODWARK, M. 1992.

- Protein heterozygosity, protein structure, and taxonomic differentiation. *Evol. Biol.*, **26**, 73–159.
- WATT, W. B. 1995. Allozymes in evolutionary genetics: beyond the twin pitfalls of 'Neutralism' and 'Selectionism'. *Rev. Suisse Zool.*, **102**, 869–882.
- WATT, W. B., DONOHUE, K. AND CARTER, P. A. 1996. Adaptation at specific loci. VI. Divergence vs. parallelism of polymorphic allozymes in molecular function and fitness-component effects among *Colias* species (Lepidoptera, Pieridae). *Mol. Biol. Evol.*, **13**, 699–709.