Interspecific isozymic substitution is not random

D. J. COLGAN

Ken and Yasuko Myer Molecular Evolutionary Biology Unit, The Australian Museum, P.O. Box A285 Sydney South, Australia 2000

The validity of the neutral theory of molecular evolution, which holds that almost all allelic substitution during evolution has been due to random genetic drift, is a pivotal question in evolutionary biology. Here it is shown that the theory predicts a 1:1 ratio among isozymes in a species between those which are faster (i.e. have greater anodal mobilities) and those which are slower (i.e. have lesser anodal mobilities) than their homologues in a related species. In fact, observed patterns of isozymic variation display deviations from such a ratio. Significant departures from expected values were found for reptiles, fish and insects. A near significant result was seen for amphibia. The overall deviations are highly significant. The results strongly suggest that a substantial fraction of allozymic substitution has been due to positive Darwinian selection.

Keywords: isozyme, natural selection, neutral theory, non-random substitution.

Introduction

The proposal that the majority of allelic substitution during evolution has been due to random genetic drift (Kimura, 1968; King & Jukes, 1969; Kimura & Ohta, 1971) aroused a spirited debate for many years after it was first suggested. Latterly, the controversy regarding this neutral hypothesis has been somewhat subdued. This is not because it has been decided one way or the other. Nor is it because the question of the comparative contributions of random, and what may be called adaptive processes to evolutionary change are no longer significant. It remains one of the most central questions in evolutionary biology. Rather, the debate is subdued because of the difficulty of posing tests which have sufficient statistical power to discriminate between the alternatives and which are satisfactorily robust against the variations which can be assumed in the underlying mathematical models. An example of these difficulties, which refers rather more particularly to questions of the average selective advantage of alleles in extant genetic polymorphism than to that of successful substitutions, is the use of allozymic frequency distributions. A number of tests have been proposed to ascertain if observed allelic frequency distributions (Ewens, 1972, 1979; Gilpin *et al.*, 1976; Watterson, 1977, 1989; Schaffer *et al.*, 1977) or genotypic segregation ratios in the offspring of heterozygous parents

(Colgan, 1981a, 1984) are in accord with those expected under drift processes. None of these has been generally successful in, or indeed even widely applied to, resolution of the claims of competing hypotheses. This is partly due to low statistical power and partly due to the dependence of expected frequencies on the mode of selection envisaged in the test.

There have been a number of studies of particular groups of animals or specific loci which have strongly supported the operation of differential selection between electrophoretically inferred genotypes. These include the demonstration by Oakeshott and his colleagues (Oakeshott *et al.*, 1981, 1982) of parallel clines at several loci in Northern and Southern hemisphere populations of *Drosophila melanogaster*. The alcohol dehydrogenase locus of *D. melanogaster* has been the subject of considerable investigation following the pioneering studies of Clarke and his group (Clarke, 1975). The trend in these investigations has been the finding of some degree of selective differential between genotypes. Various aspects of selection at the locus have been explored by McKenzie & McKechnie (1978), van Delden (1982) and Sullivan *et al.* (1990). Note, however, that some large-scale studies have failed to find evidence of selection (Yoshimaru & Mukai, 1979). There are large differences in observed phenotypic frequencies at the leucineaminopeptidase (LAP) loci in mussels from those which are expected

under the Hardy–Weinberg equilibrium (Koehn *et al.*, 1976; Colgan, 1981b, 1987; Zouros & Foltz, 1984). Part of the explanation for these deficiencies has been shown to be differential salinity tolerances of the LAP genotypes (Hilbish, 1985). Clarke (1979) pointed out that if enzymes were required to serve multiple functions (a possible basis of heterozygous advantage), there would tend to be a fixation of gene duplications and the loss of allelic variation at each member of the resulting gene family. This argument was applied by Colgan to the isozymic changes in glycolytic enzymes which occur during the development of grasshoppers (Colgan, 1986). Genes whose expression is developmentally variable are presumably required to serve multiple functions. Those where there are no developmental changes are either not required to serve multiple functions or have not had the fortune to be duplicated. Then, under Clarke's (1979) argument, developmentally variable genes should have lower allozyme variability than developmentally invariant genes. This was indeed found to be the case in grasshoppers (Colgan, 1989). It could then be argued that a fraction of the allozymic variation in developmentally invariant genes was selectively maintained.

A number of other examples of enzyme loci where selection is probably operative could have been chosen but although such specific studies are a *sine qua non* for detailed understanding of the selective factors actually experienced by a population, they can be regarded, as tests of the overall validity of the neutral hypothesis, as anecdotal. The operation of significant selective positive differentials in a relatively few cases has never been denied by the neutral hypothesis. A new test is developed in this paper. Under the neutral hypothesis, assuming functional constraints, there should be no tendency for a given species to have isozymes with either a greater of lesser mobility than those of a related species with which it is compared. That is, where mobilities differ, half of a species' isozymes should be faster than the corresponding forms in the second taxon and half should be slower. This is intuitively appealing because of its symmetry. How is it developed from the neutral hypothesis? For instance, it is shown in the next paragraph that the expectation requires justification because of the probability that, during the evolutionary processes leading to the separation of two taxa, substitution has occurred in populations of different sizes.

The rate of substitution of neutral alleles has often been treated in mathematical population genetics (Crow & Kimura, 1970; Kimura, 1983). For a diploid population in which the process has been occurring for a long time and it is assumed that each mutant represents a new site at which alternative forms are not already segregating (the 'infinite sites' model), the rate is equal to $2Nvu$, where N is the population size, v is the mutation rate per gamete per generation and u is the probability of a single mutant ultimately reaching fixation. It can be shown that u is equal to $1/2N$ so that the rate of substitution is equal, after substituting for u to v. That is, under the neutral hypothesis, the rate of substitution of mutant alleles is independent of population size. The same argument applies where actual size is replaced by effective population numbers. A deterministic treatment would suggest for neutral alleles that the rates of substitution should be the same in both branches of a dichotomy, regardless of relative population sizes. In reality, however, there would be stochastic variances in the numbers of substitutions along various lineages. If the newly arising mutants have a mobility frequency bias, then departures from the 1:1 expectation of mobility classes would be anticipated in the populations at the end of the evolutionary lineages.

Expected frequencies of electrophoretic mobility classes under the neutral hypothesis

Point mutations in coding regions may have one of a number of results affecting electrophoretic mobility. Many changes may result in protein conformational differences which are distinguishable using the gel sieving properties of some media. Better quantified results involve one of five types of substitution producing amino acids with different charges.

1 They may change a codon coding for a residue which is positively charged at physiological (and normal electrophoretic) pH to one coding for a neutral (many ways) or negatively charged residue (His → Asp; Lys → Glut). This leads to an increase in relative mobility towards the anode.

2 A codon for a negatively charged residue may change to one coding for a positively charged or a neutral residue, leading to a decrease in anodal mobility.

3 A codon for a neutral amino acid may change to one for a positively charged one, with electrophoretic consequences as in (2).

4 A codon for a neutral residue may change to one for a negative residue with consequences as in (1).

5 Histidine codons may change to those for lysine or arginine giving an increase in net positive charge (at a pH up to 9).

The frequencies of all of these effects must be considered in determining the overall frequency of mobility classes amongst newly arising mutants. Clearly, the impact of the effects will depend on the relative abundance of the various amino acid codons, which are known to exhibit usage biases (Grantham *et*

al., 1980; Sharp & Li, 1987) and on the relative frequency of mutations from one codon to another. The biases in the utilization of alternative codons for the same amino acid are now understood to be due to the preferential incorporation of codons for which the corresponding tRNA is abundant in proteins which are strongly expressed (Ikemura, 1981, 1985). This understanding does not immediately provide a way to average the effects of the bias over the entire genome but it does suggest an approach which can, and which has, in fact, been developed by proponents of the neutral theory to counter the argument that the biases are themselves evidence against neutrality (Kimura, 1986, 1990). This counter is to invoke 'functional constraints' and to emphasize that in questions of substitution we are concerned with successful mutants rather than those which are eliminated from the population. It is clear that proteins are subject to functional constraints which are such that many, if not most, mutants are unable to allow normal viability. Such mutants are relatively rapidly eliminated.

The neutral hypothesis regarding substitutions at any given site in a protein-encoding sequence proceeds as follows. There are a range of bases which are acceptable at the site in that they form part of codons which satisfy the functional constraints on this section of the DNA sequence and the constraints on the amino acid residues in the corresponding position of the encoded protein. The range of acceptable alternative bases at most sites is probably restricted to the incumbent specified by the currently utilized codon. At sites where there are multiple acceptable alternatives, the fate of mutations from the incumbent base will be determined by genetic drift. Other changes will generally be rapidly eliminated from the population, although there is a very small fraction of selectively advantageous mutations. It is considered in this paper that the neutral hypothesis must predict that it is essentially a matter of chance which, of the alternative acceptable bases, occupies a given site at a given time, if the process has continued long enough to establish a steady-state. If there is a charge difference between two residues specified by members of an acceptable range of bases then a newly arising neutral mutation from one member of the pair to the other has an equal probability of increasing or decreasing net negative charge (because the mutation has an equal probability of occurring in either direction). This scenario also applies to changes between residues which do not cause charge differences but which nevertheless result in conformational differences detectable by electrophoresis. Stochastic deviations from equality will tend to be balanced out when a range of proteins from a variety of phylogenetically diverse taxa are considered.

That is, under the neutral hypothesis, the overall frequency amongst successful new mutants of those which are faster than their antecedent is expected to be the same as the frequency of those which are slower.

The expected frequencies of mobility classes of successful new mutants cannot be estimated from observed patterns of codon usage for at least two reasons. Firstly, because genetic manipulation is easier for highly expressed genes (as the cDNA for the mRNA can be obtained more readily), the genes which have been sequenced will present a distorted picture of the overall degree of codon usage bias, if Ikemura's (1985) explanation of the bias is correct. Secondly, irrespective of the first difficulty, there is no available satisfactory method by which *a priori* one can deduce the acceptable range of neutral substitutions for any given site. Hence, any arguments which purport to show differences from an expected 1:1 ratio in the relative mobility classes of new mutations, and which are based on codon usage biases or amino acid abundance, must be considered erroneous.

Patterns of apparent substitution which are generated by the fixation of one allele (of two or more in a pre-existing polymorphism) along each branch of an evolutionary dichotomy would also be expected to show a 1:1 ratio in expected frequencies of the relative mobility classes under the neutral hypothesis. Such polymorphisms originate with the appearance of a new allele in a population. The relative mobility of this allele with respect to its antecedent would be expected to follow the 1:1 ratio explained above. There is also an additional level of randomness, however, under the neutral hypothesis, in this scenario. Chance effects, as well as population size and allelic frequency would also play a role in determining what representative of a locus predominates in what lineage.
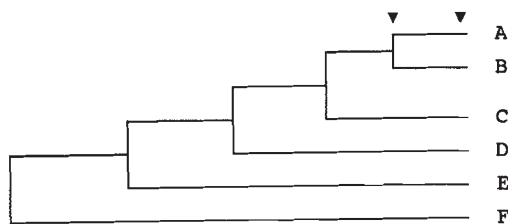
Methods

The data considered in this paper were taken from a survey of the literature on animal electrophoresis. A survey of the plant literature could provide a useful additional test of the results of the analyses of the animal data treated here. In the present instance, the survey was conducted by searching tables of contents and indices of major taxonomic and evolutionary journals. The journal coverage will be obvious from perusal of the list of comparisons between taxa used in the analyses, which is available from the author and which has also been lodged with the editorial office of *Heredity*.

The mechanics of making the comparisons between taxa are straightforward. Publications that present details of allelic frequencies and relative mobility in

more than one taxon could be used, providing that, where there were three or more taxa, a dendrogram of relatedness was included. Many publications could not be used because allelic frequency data were not presented. Many comparisons between taxa were not relevant to the present purpose because only one locus differed between the samples. One study could not be used because only intraspecies relative mobilities were detailed. Otherwise, there was never any apparent confusion as to the relative mobilities of isozymes.

The procedure for selecting species for comparison can be seen in Fig. 1. The routine method was to compare those pairs identified as sisters (by phenetic or cladistic means) in the dendrogram regarded as most reliable by the authors of the relevant publication. Comparisons were also routinely made between such pairs as (C, D) in Fig. 1. Comparisons were not routinely made for pairs such as (E, F). Where such a comparison was made it was identified as being between 'distantly related' taxa. Only one such comparison was made for each study. This was chosen to compare the most distantly related taxa in the study. Where more than six taxa were investigated this would not be the pair (E, F). In such cases these taxa were not further treated. Where there were less than five taxa, or where all (or all but one) terminal taxa were included in sister-pair comparisons, the 'distant' comparison was of a pair such as (A, E). In this case there is an overlap between distant and sister comparisons of an evolutionary branch (indicated by arrows in the figure). With this exception, the above procedure was adopted to minimize the double counting of evolutionary lineages.

There is some question as to what should be considered as a substitution event for the purposes of the

comparisons used here. In most of the studies, the sample numbers are too small to be certain that all of the rarer alleles at a locus have been detected in a study. Consequently, even though a particular pair of taxa may appear to show a fixed difference, one or other allele may be present at a low frequency in the member of the pair in which it is not the most abundant. Hence, it was decided to record, as substitution events, those cases where there is a great difference between taxa in the frequencies of alleles. As a convention, an event was supposed to have occurred if the common allele in each taxon comprised less than 0.1 of the sample of alleles from the other taxon. This will probably lead to the registration of some cases as substitutions of a particular allozyme which subsequent evolution will show to have been an unwarranted assumption. Nevertheless, where frequencies are essentially driven by drift, the vast majority of alleles present at frequencies less than 0.1 would go to extinction.

A further complication arises in the small number of cases (<0.5% of loci) when there are multiple alleles in a taxon which are all present in substantial frequencies (>0.1) and which are all faster or all slower than the set of alleles in the second taxon. This is again regarded as a single substitution. A locus was ignored if it was polymorphic (i.e. had two allozymes with frequencies >0.1) and the mobilities of its allozymes straddled the mobility of the corresponding isozyme(s) in the second species of a comparison.

Comparisons were registered by scoring the number of loci where the allozymes of one taxon were faster than those of the second member of the pair. They could equally well have been registered as cases where slower forms were counted. Loci which were invariant or where, according to the conventions above, a substitution had not occurred were ignored in further analysis.

Results

Summaries of the results are presented in Tables 1 and 2. More detailed breakdowns of the types of comparisons within each taxonomic group can be obtained from the author. Both tables show the distribution of the numbers of comparisons (i.e. species pairs) with the specified absolute difference in the numbers of loci for which each member of the pair has the faster allozyme. The rows indicate the number of loci on which the comparisons are based. Statistical testing of these distributions proceeds as follows. Given that a 1:1 ratio of relative mobility classes is expected under the neutral hypothesis, comparing the mobilities of the allozymes in a taxon with those at corresponding loci in the

Fig. 1 Dendrogram explaining the choice of species pairs for comparison. Pairs such as (A, B) and (C, D) in the dendrogram regarded as most reliable by the original authors were chosen for comparison. Comparisons were not routinely made for pairs such as (E, F). One comparison was made to compare the most distantly related taxa in each study. With six taxa, this would be the pair (E, F), with more taxa it would not. Where all (or all but one) terminal taxa were included in other comparisons, the 'distant' comparison was of a pair such as (A, E). Here, there is an overlap of an evolutionary branch between comparisons (indicated by arrowheads).

Table 1 The distribution of isozymic relative mobilities in species-pairs sorted by taxonomic group. The rows indicate the number of substitutions between the members of a given pair. The columns show the absolute difference between the numbers of loci for which the members of the pair have the isozyme(s) with greater anodal mobility. The figures in the cells are the numbers of species-pairs with the specified combination of row and column characteristics. All species-pairs with absolute differences greater than 10 are indicated in the final column

Locus level	0	1	2	3	4	5	6	7	8	9	10	11+
Mammals												
2	5	8										
3		7		3								
4	4		6		1							
5			2		3							
6	2		1		3							
7		6		1		1						
8	2		1		2							
9		1		2								
10			2		1							
11				2								
12	2		3		1		1					
13				1		2		1				
14	1		1								1	
15		1				1				1		
Birds												
2	4		1									
3		1		1								
4	1		1									
5		1										
7		1				1						
8			2									
9		1				1						
10			1									
14			1								1	
15						1						
17		1										
Reptiles												
2	2		2									
3		3		7								
4	1		6		1							
5		1		3								
6	1		1		1							
7				2								
8	1		1		2							
9		2				1						
10			1		2							
11				1		1			1			
12			1		1							
13		2										
15		1										
16												1 × 14
17						1		1				1 × 15
												1 × 17

Table 1 *Continued*

Locus level	0	1	2	3	4	5	6	7	8	9	10	11+
Amphibians												
2	3		4									
3		7										
4	2		4		1							
5		2										
6	1		4		1							
7		2										
8			1		1							
10		1										
12										1		
13				1								
14	1				1							1 × 14
15						1						
16	1									1		
Fish												
2	1		5									
3			2									
4	5		3									
5		1		3								
6	1		2		1							
7		1										
8	1		1									
9		2				2						
10	2		1				1			1		
11						1		1				
12					1					1		1 × 12
14			1									
15		2										
17												1 × 17
Insects												
2	4		7									
3		5		3								
4	2				1							
5		6		4								
6	2		3		1		1					
7		2				3						
8			1		3		1					
9				2		1						
11				3		1						
12			1		1							
13		1										
Other invertebrates												
2	1		1									
3		2										
4	4		2									
5		4		1								
6	2		2									
7		1		2								
9				1		1						
10	1											
13		1			1							

Table 2 The overall distribution of isozymic relative mobilities in species-pairs. The rows indicate the number of loci where substitutions have occurred between the members of a given pair. The columns headed by numerals show the absolute difference between the numbers of loci for which each member of the pair has the isozyme(s) with greater anodal mobility. The figures in the cells are the numbers of species-pairs with the specified combination of row and column characteristics. The second last column is the expected mean (see text for method of calculation) and the last the observed mean for each row. The last two rows give the total expected and observed numbers of each column

Locus level	Difference												Expected mean	Observed mean
	0	1	2	3	4	5	6	7	8	9	10	11+		
2	20		28										1.00	1.17
3		25		16									1.50	1.78
4	19		22		4								1.50	1.29
5		17		14									1.87	1.90
6	8		13		7	1							1.87	2.07
7		13		5		5							2.19	2.30
8	4		7		8		1						2.19	2.60
9		6		5		5		1					2.46	3.12
10	3		6		3		1				1		2.46	2.86
11			6		3		1		1				2.71	4.46
12	2		5		4		1				2	1 × 12	2.71	4.27
13		4		2		3		1					2.93	3.20
14	2		3		1		1				2	1 × 14	2.94	5.00
15		5				3				1			3.14	3.22
16	1										1	1 × 14	3.16	8.00
17		1				1		1				1 × 15 2 × 17	3.38	10.3
Total observed	59	71	84	48	27	20	5	4	—	2	6	6		
Total expected	64.9	86.0	85.1	44.3	25.5	13.2	6.6	3.6	1.5	0.8	0.3	0.2		

other member of a pair can be considered as a series of Bernoulli trails (with 'success' being a faster mobility). The number of successes follows a binomial distribution with P, the probability of a success in any one trial being $1/2$. Clearly there is no *a priori* means to pick one member of a pair of taxa for counting of successes so we work with the absolute difference between the number of successes and failures in any given comparison. The expected frequencies of the various numbers of absolute differences can be determined by reference to a table of binomial probabilities. For example, with five loci in the comparison, a difference of three eventuates whether there are one or four 'successes'. Summing the probability of these two cases gives an expected frequency of 0.312. The first test that can be made with these frequencies is to compare the means of the observed and hypothesized distributions for each 'locus level' (this term will be used to designate the number of loci on which a comparison between two species is based). A sign test can be used to investigate whether there are consistent deviations in either direction.

A second test is based on the following rationale. The hypothesized distribution of outcomes for each locus level is a random variable. An individual comparison can be considered as a sample from this variable and the n_i observations at this locus level as a sample of this size. The entire set of observations can be treated as a sample from the random variable with a density given by the following relation:

$$f(x) = \sum_{i=1}^{k} n_i f_i(x)/N$$

where N is the total number of comparisons, $f_i(x)$ is the density of the random variable generated by the neutral hypothesis at locus level i and k is the largest locus level. Note that the overall sample is not strictly random as its variation is somewhat reduced by fixing the size of the sample at each locus level. This means that whatever null hypothesis is being tested would be marginally favoured by the bias. Various tests are possible to compare observed patterns of absolute differences to those expected from the distribution of this random variable. These include chi-squared goodness-of-fit tests and tests based on the convergence, under the Central Limit Theorem, of the deviation of the sample mean from its true value to a normal distribution.

These tests can be applied to the overall dataset. Comparing observed and hypothesized locus level means, only one of the 16 trials gives lower values than expected (Table 2, last two columns). The probability of this occurring by chance is less than 0.005, which indicates that there is a tendency for the absolute value of the difference between taxa in the number of loci at which they have faster allozymes to be greater than predicted by the neutral hypothesis. This is also observed for cases where a smaller number of substitutions have occurred. One only of the first nine locus levels has an observed mean less than expected. This is significant at the 0.05 level. This remark remains true if the values of the singleton outliers at the 9 and 10 locus levels are rounded down to the next highest observed value (i.e. the data are 'winsorized' — Sachs, 1982).

To test whether observed means of the absolute difference differ from expected, a standardized normal statistic was constructed according to the expression

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt[3]{N}}$$

where the mean and variance are calculated from the hypothesized distribution of $f(x)$. For the overall dataset, the probability of obtaining the observed result by chance is negligible. We may be interested, however, in ignoring the contribution of outliers. If all the observed absolute differences greater than 10 are assumed equal to 10 then the probability of observing the result by chance is < 0.001. If all observed absolute differences greater than 10 are completely ignored, $z = 3.039$ ($P < 0.002$). If all above 7 are rounded down to 7, $z = 3.284$ ($P < 0.002$). Much the same picture emerges for the chi-square tests of goodness-of-fit, using $f(x)$ to obtain expected values. Pooling all cells with expected absolute differences of seven or more gives a statistic (with 7 d.f.) with $P < 0.005$. Pooling all differences of six or more gives a statistic of 14.756 ($P < 0.025$). For locus levels up to 10 the observed mean absolute difference is significantly greater than expected ($P < 0.05$).

Clearly the overall dataset is significantly different from expectations under the neutral hypothesis. But, it is also of interest to know whether this is true for subsets of the data. Firstly, does double counting of phylogenetic branches in sister-pair and distant comparisons affect the conclusions? The same battery of tests was applied to the two subsets where (i) distant pair comparisons involving overlaps were omitted and (ii) where sister-pair comparisons with overlaps were ignored. The sign tests had the same level of significance for both subsets as for the overall data. Omitting sister-pairs with overlaps, the total observed mean

absolute difference had a negligible P value. Ignoring all comparisons with absolute differences greater than 10, $P < 0.002$. Whilst this is changed to non-significance if all differences above nine are ignored, rounding all observed extreme values down to nine produces a z of 4.25. The chi-squared test of goodness-of-fit with all expected differences of more than seven pooled has a probability of less than 0.005 of being returned under the null hypothesis. When distant-pair comparisons with overlaps are ignored, the total subset mean absolute difference is highly significantly greater than expected ($P < 0.001$); assuming all observations above nine are equal to nine gives $P < 0.001$. The chi-square test of goodness-of-fit, where all values above seven are pooled, is significant at the 0.005 level.

The pattern of the parametric test results for locus levels up to 10 is much the same in the two subsets we are now considering, as in the overall dataset. The patterns for the subset omitting sister-pair comparisons will be treated as an example. The observed mean absolute difference is significantly greater than expected ($P < 0.05$). However, if the outlying observation at the 10 locus level is assumed to be the same as the next most extreme observation (8), the calculated z is only marginally larger than the 0.05 significance level.

Turning to consideration of individual taxonomic groups, sufficient data for testing were available for mammals, birds, reptiles, amphibia, fish and insects (Table 1). A miscellaneous grouping comprising other invertebrate taxa was also examined. Sign tests were performed for the comparison of sample locus level means with expected values for all of these groups. None of the tests of an individual group was significant although combining all gave a highly significant excess of trials where the observed mean was greater than the expected. The finding of 11 (of 14) means larger than expected for reptiles borders on significance ($P \approx 0.052$). In other groups, nine of 14 locus levels for mammals had observed means larger than expected, as did five of 11 in birds, 10 of 13 in amphibia, nine of 14 in fish, eight of 11 in insects and three of seven in other invertebrates. Tests (two-sided) of the standardized overall taxon group mean against an $N(0, 1)$ distribution gave significant z values for reptiles ($P < 0.001$), fish ($P < 0.01$) and insects ($P < 0.02$) and non-significant values for other groups. For mammals, $0.20 > P > 0.15$; for amphibians, $0.15 > P > 0.10$; for birds $0.30 > P > 0.20$ and for the miscellaneous invertebrates $0.40 > P > 0.30$. In all of these groups except the last, the observed mean was higher than the expected.

The analyses to this point have shown that the departures from random expectations are not solely

due to a small number of comparisons with large absolute differences in the two relative mobility classes. Inspection of the last two columns of Table 2 suggests, however, that there may be a tendency to an increase in the excess of the observed mean over the expected mean at any given locus level as the number of loci in the comparisons increases. This tendency was tested by examining the significance of the linear regression coefficient of $(\bar{x}_i - \mu_i)/\mu_i$ against the number of loci. For all locus levels, the t_{14} statistic is 3.59 ($P < 0.005$). If the last three locus levels are excluded as may be suggested by the unusually low (level 15) or unusually high (levels 16 and 17) observed means, the t_{11} value is 2.886 $P < 0.01$). So, as the number of observed differences (locus levels) between species increases, the observations become more different from those which are expected under the neutral hypothesis.

Discussion

The test developed here provides a more powerful general evaluation of trends in substitution than is presently available. In the test, a significant tendency was found for the mobilities of a species' isozymes to deviate from a 1:1 ratio between those which are faster and those which are slower relative to homologues in a second taxon. It is difficult to see how this could be explained without invoking positive Darwinian selection. The test is not based on a particular mathematical model of mutation, such as the infinite alleles or infinite sites (Kimura, 1969) models. And even if there were deviations from the steady-state patterns of substitution envisaged by the neutral theory with functional constraints, it is not clear how under the hypothesis this could induce a mobility bias.

The dataset and taxonomic range on which the conclusions in this paper are based is large, although it could be increased by the consideration of the plant literature. The dataset could also be increased by provision of frequency data by the authors of papers which do not present such data and which hence could not be included in this study.

Some elaborations of the overall conclusion that the neutral hypothesis cannot explain the observed distribution of relative electrophoretic mobilities should be noted.

1 Departures from expectations are found in a wide range of organisms: significant deviations being found in three large taxonomic groups, the insects, reptiles and fish.

2 There is an excess of observed over expected mean absolute differences at almost all locus levels, suggesting that the departures from expectations are not solely due to outlier effects of pairs of genetically distant species. However, there is also a significant regression of the deviation of observed means (from expected means) on the number of loci on which a comparison is based. This will be discussed further below.

3 The present tests compare species. It could be argued that intra-specific substitution is somehow different.

4 The present tests are based on electrophoretically observable changes which might be supposed to be qualitatively evolutionarily different to other types of substitution. This is possible as charge or conformational differences themselves may be of selective significance. However at least 1/3 (and frequently a much greater fraction) of single-site mutations are detectable by electrophoresis (Ramshaw et al., 1979; McLellan & Inouye, 1986). The failure of the neutral hypothesis in such a substantial subset of substitution would prohibit its general validity.

5 The present results do not impinge directly on the question of whether there is a substantial amount of genetic variability actively maintained in populations by factors such as heterozygous advantage or frequency-dependent selection.

It would be interesting to obtain an estimate of the fraction of allozymic substitutions which have been selectively driven. Calculation of $f(x)$ with varying P produces concordance of expected and observed means with a value between 0.60 and 0.63. The actual fraction of selectively driven substitutions cannot be directly determined from these figures. Some mobility differences in the present dataset may actually comprise multiple substitutions. Under the neutral hypothesis this does not alter the expected 1:1 ratio of mobility classes but if a bias between relative mobility classes of new successful mutants arises from other causes then successive substitutions would inflate observed deviations from expectation. In the unlikely event that there were as many as five substitutions at each locus in the data included in the present study, an apparent P value of 0.63 could be obtained from a true P of 0.52. This is, however, still far too high for the neutral theory to be considered accurate. The occurrence of multiple substitutions may, however, be the explanation of the significant regression of the excess of observed over expected means (normalized by the expected mean) on the locus level. Selectively advantageous mutations would tend to be accumulated over evolutionary time if the basis of their advantage remains unchanged. Codon usage biases (Grantham et al., 1980; Sharp & Li, 1987) have presumably arisen in this fashion.

The question arises as to what might be the selective basis of this deviation of electrophoretic mobility class frequencies from neutral theory expectations. Are they

due to selective pressures favouring one type of charge (or conformational) change *per se*? Is it due to a property of proteins other than charge which is better promoted by amino acids of a given type? Or is it due to selection favouring substitution of a specific amino acid? Perhaps an environmental change favours a change in flux through a biochemical pathway which is most easily achieved by a number of substitutions of one sort of amino acid which is particularly suitable because of relative tRNA abundances, present codon usage bias, etc. Probably all three of these causes, and possibly others, have played a role in the determination of the observed patterns of substitution. It would be premature to speculate overly regarding this, but some comments may be made about the pre-requisities for operation of the first factor. It may be anticipated that ion concentration, and indeed other properties relevant to optimizing protein charge, would be subject to strong homoeostatic mechanisms (Masoro & Siegel, 1977; McGilvery, 1979). However, species-specific steady-state profiles may have developed in the course of evolution. More likely, perhaps, is that there may be differences in the response to perturbations from the steady-state (caused by exertion, starvation of particular nutrients, etc.) which reflect the selective pressures of the different environments in which species are required to survive.

Acknowledgements

References

CLARKE, B. 1975. The causes of biological diversity. *Sci. Am.*, **233**, 50–60.

CLARKE, B. 1979. The evolution of genetic diversity. *Proc. Roy. Soc. Lond. B*, **205**, 453–474.

COLGAN, D. J. 1981a. The relative viabilities of alternative human genotypes. *Hum. Hered.* **31**, 172–179.

COLGAN, D. J. 1981b. Spatial and temporal variation in the genotypic frequencies of the mussel *Brachidontes rostratus*. *Heredity*, **46**, 197–208.

COLGAN, D. J. 1984. The effects of linked loci on estimates of relative viability derived from pedigree analysis. *Heredity*, **53**, 69–77.

COLGAN, D. J. 1986. Developmental changes in the isoenzymes controlling glycolysis in the acridine grasshopper, *Caledia captiva*. *Roux's Arch Dev. Biol.*, **195**, 197–201.

COLGAN, D. J. 1987. Deficits of heterozygotes in relation to selective interaction between loci in the mussels *Brachidontes rostratus* and *Xenostrobus pulex*. *Heredity*, **59**, 355–362.

COLGAN, D. J. 1989. Developmental variation and allozymic polymorphism in the enzymes of glycolysis in grasshoppers. *J. Evol. Biol.*, **2**, 141–151.

CROW, J. F. AND KIMURA, M. 1970. *An Introduction to Population Genetics Theory*. Harper and Row, New York.

VAN DELDEN, W. 1982. The alcohol dehydrogenase polymorphism in *Drosophila melanogaster*: selection at an enzyme locus. *Evol. Biol.*, **15**, 187–222.

EWENS, W. J. 1972. The sampling theory of selectively-neutral alleles. *Theor. Pop. Biol.*, **3**, 87–112.

EWENS, W. J. 1979. *Mathematical Population Genetics*. Springer-Verlag, New York.

GILPIN, M., SOULÉ, M. ONDRICEK, A. AND GILPIN, E. E. 1976. Overdominance and U-shaped gene frequency distributions. *Nature*, **263**, 497–499.

GRANTHAM, R., GAUTIER, C. AND GOUY, M. 1980. Codon frequencies in 119 individual genes confirm consistent choices of degenerate bases according to genome type. *Nucl. Acids Res.*, **8**, 1893–1912.

HILBISH, T. J. 1985. Demographic and temporal structure of an allele frequency cline in the mussel *Mytilus edulis*. *Mar. Biol.*, **86**, 163–171.

IKEMURA, T. 1981. Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes. *J. Mol. Biol.*, **146**, 1–21.

IKEMURA, T. 1985. Codon usage and tRNA content in unicellular and multicellular organisms. *Mol. Biol. Evol.*, **2**, 13–34.

KIMURA, M. 1968. Evolutionary rate at the molecular level. *Nature*, **217**, 624–626.

KIMURA, M. 1969. The number of heterozygous nucleotide sites maintained in a finite population due to a steady flux of mutations. *Genetics*, **61**, 893–903.

KIMURA, M. 1983. *The Neutral Theory of Molecular Evolution*. Cambridge University Press, Cambridge.

KIMURA, M. 1986. DNA and the neutral theory. *Phil. Trans. R. Soc. Lond. B*, **312**, 344–354.

KIMURA, M. 1990. Some models of neutral evolution, compensatory evolution and the shifting balance process. *Theoret. Pop. Biol.*, **37**, 150–158.

KIMURA, M. AND OHTA, T. 1971. Protein polymorphism as a phase of molecular evolution. *Nature*, **229**, 467–469.

KING, J. L. AND JUKES, T. H. 1969. Non-Darwinian evolution: random fixation of selectively neutral alleles. *Science*, **164**, 788–798.

KOEHN, R. K., MILKMAN, R. AND MITTON, J. B. 1976. Population genetics of marine pelecypods. IV. Selection, migration and genetic differentiation in the blue mussel *Mytilus edulis*. *Evolution*, **30**, 2–32.

MCGILVERY, R. W. 1979. *Biochemistry: a Functional Approach*. W. B. Saunders, Philadelphia.

MCKENZIE, J. A. AND MCKECHNIE, S. W. 1978. Ethanol tolerance and the ADH polymorphism in a natural population of *Drosophila melanogaster*. *Nature*, **272**, 75–76.

MCLELLAN, T. AND INOUYE, S. 1986. The sensitivity of isoelectric focusing and electrophoresis in the detection of sequence differences in proteins. *Biochem. Genet.*, **24**, 571–577.

MASORO, E. J. AND SIEGEL, P. D. 1977. *Acid-Base Regulation*. W. B. Saunders, Philadelphia.

OAKESHOTT, J. G., CHAMBERS. G. K., GIBSON, J. B. AND WILLCOCKS, D. A. 1981. Latitudinal relationships of esterase-6 and

phosphoglucomutase gene frequencies in *Drosophila melanogaster. Heredity*, **47**, 385–396.

OAKESHOTT, J. G., GIBSON, J. B., ANDERSON, P. R., KNIBB, W. R., ANDERSON, D. G. AND CHAMBERS, G. K. 1982. Alcohol dehydrogenase and glycerol-3-phosphate dehydrogenase clines in *Drosophila melanogaster* on different continents. *Evolution*, **36**, 86–96.

RAMSHAW, J. A. M., COYNE, J. A. AND LEWONTIN, R. C. 1979. The sensitivity of gel electrophoresis as a detector of genetic variation. *Genetics*, **93**, 1019–1037.

SACHS, L. 1982. *Applied Statistics*. (transl. Z. Reynarowych). Springer-Verlag, New York.

SHAFFER, H. E., YARDLEY, D. G. AND ANDERSON, W. W. 1977. Drift or selection: a statistical test of gene frequency variation over generations. *Genetics*, **87**, 371–379.

SHARP, P. M. AND LI, W.-H. 1987. The rate of synonymous substitution in Enterobacterial genes in inversely related to codon usage bias. *Mol. Biol. Evol.*, **4**, 222–230.

SULLIVAN, D. T., ATKINSON, P. W. AND STARMER, W. T. 1990. Molecular evolution of the alcohol dehydrogenase genes in the genus *Drosophila. Evol. Biol.*, **24**, 107–147.

WATTERSON, G. A. 1977. Heterosis of neutrality? *Genetics*, **85**, 789–814.

WATTERSON, G. A. 1989. The neutral alleles model with bottlenecks. In: Feldman, M. W. (ed.), *Mathematical Evolutionary Theory*, pp. 26–40. Princeton University Press, Princeton.

YOSHIMARU, H. AND MUKAI, T. 1979. Lack of experimental evidence for frequency dependent selection at the alcohol dehydrogenase locus in *Drosophila melanogaster. Proc. Nat. Acad. Sci., U.S.A.*, **76**, 876–878.

ZOUROS, E. AND FOLTZ, D. W. 1984. Possible explanations of heterozygote deficiency in bivalve molluscs. *Malacologia*, **25**, 583–591.