*Heredity* **68** (1992) 241–252

*Received 27 March 1991*

# Sources of variance in protein heterozygosity: the importance of the species–protein interaction

JACK DA SILVA, GRAHAM BELL & AUSTIN BURT*

*Department of Biology, McGill University, 1205 Avenue Docteur Penfield, Montreal, Quebec, Canada H3A 1B1*

We report on a detailed survey of protein heterozygosity in Canadian freshwater fish and mammals. A simple one-way analysis showed substantial variance among species. However, a two-way analysis of species and proteins showed that there was little if any variance among species or among proteins, but a very large species–protein interaction. We could not remove this interaction by analysing taxa separately, by constructing completely balanced datasets, by eliminating study bias or by excluding monomorphic proteins, and we could not decompose the interaction by classifying enzymes according to their form and function. We conclude that most of the variance in protein heterozygosity is attributable to species–protein interaction. This casts some doubt on the interpretation of comparative studies of mean heterozygosity among species or among proteins. Our result seems inconsistent with the neutral theory of protein variation but not with the differential action of natural selection.

**Keywords:** allozyme variation, gene diversity, genetic variation, heterozygosity, natural selection, neutral theory.

## Introduction

The application of gel electrophoresis to population genetics from the mid-1960s onwards has shown that protein structure varies substantially within natural populations. The renewed debate about variation, which was fuelled by this discovery, was graphically described by Lewontin (1974). At first, the new variation that was uncovered seemed to confirm the importance of balancing selection, but it was soon realized that hard selection, acting simultaneously at so many loci, depresses mean fitness so much that populations are unable to survive. This led to the theory that most protein variation is selectively neutral, and represents only a phase in the substitution of alleles by sampling error (Kimura, 1983). The controversy has continued, and has been reviewed recently by Nevo *et al.* (1984), Nei & Graur (1984) and Nei (1987).

During the last quarter-century, variation has been measured for about 100 proteins distributed over about 1000 species of plants and animals. There have been two approaches to testing selectionist and neutralist theories: detailed studies of proposed enzyme function in particular situations, and broad surveys of variation. This paper is concerned only with the second approach. The comparative analysis of protein variation has uncovered many patterns of great interest but has not been successful in distinguishing unequivocally between neutralist and selectionist interpretations of these patterns. The reason for this uncertainty is that variation among neutral alleles is governed by the product of population number and mutation rate. As it is rarely possible to estimate either of these parameters with precision, predicting the variation expected over a wide range of species is impracticable. Consequently, the neutral theory can generate comparative predictions by using covariates of population size and mutation rate; but these covariates can seldom if ever be shown to be devoid of ecological significance, and so an alternative selectionist interpretation is always available. For example, the neutral theory predicts that enzymes of greater molecular weight will be more variable because their total mutation rate will be higher. This seems to be generally true (Koehn & Eanes, 1977; Ward, 1978; Nei *et al.*, 1978), although there are some exceptions, especially in humans (Harris *et al.*, 1977; Eanes & Koehn, 1978); but the same facts can be given a selectionist interpretation (Leigh, Brown & Langley, 1979).

*Present address: Department of Biology, University of California, Santa Cruz, CA 95064, USA.

Enzymes with more subunits should be less variable because complex quarternary structures may constrain the number of possible neutral changes. This also appears to be true (Harris *et al.*, 1977; Ward 1977), but quarternary structure may reflect enzyme function (Zouros, 1975; Ward, 1977). Selectionists have replied by arguing that variability is correlated with enzyme function (Gillespie & Langley, 1974; Johnson, 1974), but this is by no means always the case (Selander, 1976). Comparisons among species seem even less decisive. There is no shortage of surveys which demonstrate correlations between protein variation and ecological, demographic and life-history variables (e.g. Hamrick *et al.*, 1979; Nevo *et al.*, 1984; Wooten & Smith, 1985; Mitton & Lewis, 1989). However, Nei & Graur (1984) claim that variation is correlated with population size, as expected under the neutral theory. Population size is now very generally correlated with individual body size; and body size in turn is correlated with almost every aspect of physiology (Peters, 1983). Hence, it is difficult to use weak correlations with ecological variables to support the selectionist theory as these may readily arise through covariation with body size.

Our initial object in conducting the survey reported here was to test a particular selectionist hypothesis, that protein variation among species of hosts is correlated with the species diversity of their parasites, through a detailed quantitative analysis of a restricted fauna. We have not yet proceeded with this exercise because, to our surprise, we were unable to detect any substantial variation in heterozygosity either among species or among proteins in our material. Instead, the bulk of the variation present is attributable to the interaction between species and proteins. This unexpected discovery has led us to explore in detail the structure of variation in our data and to comment on its interpretation in terms of selection, mutation and genetic drift.

## Materials and methods

### *Range of species surveyed*

We collated data from electrophoretic surveys of within-population variation for North American freshwater fish and land mammals whose geographical range extends into Canada. Only native, non-stocked species, and only exclusively freshwater fish, were used. Samples of the same species of fish were classified by drainage basin, with populations within drainage basins forming the lowest level of analysis and thus contributing the residual variance. No analogue of drainage basin exists for mammals, where populations were nested directly under species. In some cases more than one sample from a population is reported; then the mean of these samples is used as the population value. The primary literature was searched through 1989 for fish and through 1984 for mammals. The 22 species of fish and 17 species of mammals used are listed, with sample sizes and species means, in Table 1. The original references are listed in the Appendix.

### *Characterization of proteins*

We recorded the per-locus mean observed heterozygosity $H_{obs}$ and Hardy–Weinberg expected heterozygosity $H_{exp}$ for each population, when these were reported. Most of our analysis, however, is based on protein heterozygosity, $H_p$, the per-locus mean expected heterozygosity over all loci (including monomorphic loci) resolved for a given protein in a given population in a given study. Single-locus heterozygosity $h$ was calculated in the usual way as $h = 1 - \Sigma x_i^2$, where $x_i$ is the frequency of the $i$th of $n$ alleles at a locus and the sum is taken over all $n$ alleles. $H_p$ was used because the homology of loci among distantly related species is often unclear, whereas the identity of proteins is unequivocal. We calculated $H_p$ for 54 proteins: 45 enzymes and three non-enzyme proteins in fish, and 29 enzymes and five non-enzyme proteins in mammals. The number of loci surveyed per population and the mean values of $H_{obs}$, $H_{exp}$ and $H_p$ for each species are given in Table 1.

Enzymes were classified by function in two ways. Gillespie & Langley (1974) distinguished Group I enzymes, with single substrates, from Group II enzymes, which have several substrates. Johnson (1974) further divided Group I into regulatory and non-regulatory enzymes. Discrepancies between these two papers were scored in favour of Johnson. Quaternary structure was taken from Hopkinson *et al.* (1976), and Ward (1977); if enzymes coded by different loci had different quaternary structures, the value was recorded as missing. Subunit molecular weights are human data from Hopkinson *et al.* (1976), who provide the justification for neglecting variation in molecular weight among vertebrate species. For the purpose of analysis, enzymes were classified by the quartiles of the frequency distribution of subunit molecular weight (17,000–35,250; 35,251–42,000; 42,001–54,750; 54,751–112,000). Characteristics of the proteins and the number of species in which each was scored are listed in Table 2.

### *Statistical analysis*

The arcsine square-root transform of $H_p$ is approximately normally distributed and is used as our

**Table 1** Summary of data by species. References are given in the Appendix

| Number | Species | Populations | Individuals/ population | Loci/ population | Species mean $H_{obs}$ | Species mean $H_{exp}$ | Species mean $H_p^*$ | Reference |
|---|---|---|---|---|---|---|---|---|
| Fish | | | | | | | | |
| 1 | *Salvelinus namaycush* | 13 | 112 | 33 | 0.01067 | 0.04308 | 0.06185 | 12,22 |
| 2 | *Coregonus clupeaformis* | 8 | 80 | 25 | 0.06800 | 0.06925 | 0.03108 | 10,25 |
| 3 | *Esox lucius* | 2 | 43 | 25 | 0.00200 | 0.00200 | 0.00121 | 21 |
| 4 | *Nocomis micropogon* | 2 | 15 | 43 | — | 0.03200 | 0.03073 | 19 |
| 5 | *Notemigonus chrysoleucus* | 1 | 15 | 24 | 0.06800 | 0.06700 | 0.07346 | 2,4 |
| 6 | *Notropis cornutus* | 5 | 9 | 26 | — | 0.04680 | 0.04104 | 7,13,14 |
| 7 | *Ptychocheilus oregonensis* | 1 | 28 | 24 | 0.01100 | 0.01200 | 0.02277 | 2,4 |
| 8 | *Rhinichthys cataractae* | 15 | 24 | 24 | 0.05385 | 0.04673 | 0.06152 | 31,19 |
| 9 | *Carpiodes cyprinus* | 1 | 15 | 31 | 0.08300 | — | 0 | 15,16 |
| 10 | *Catastomus catastomus* | 1 | 15 | 29 | 0.03800 | — | 0 | 15,16 |
| 11 | *Catastomus commersoni* | 1 | 15 | 30 | 0.03500 | — | 0 | 15,16 |
| 12 | *Erimyzon sucetta* | 1 | 15 | 29 | 0.05800 | — | 0 | 15,16 |
| 13 | *Hypentelium nigricans* | 4 | 13 | 35 | 0.00900 | 0.01067 | 0.00994 | 8,15,16 |
| 14 | *Moxostoma erythrurum* | 1 | 15 | 27 | 0.03400 | — | 0 | 15,16 |
| 15 | *Ictalurus punctatus* | 4 | 15 | 23 | — | 0.01725 | 0.01927 | 24,40 |
| 16 | *Ambloplites rupestris* | 1 | 12 | 11 | — | 0.13000 | 0.15729 | 5 |
| 17 | *Lepomis gibbosus* | 1 | 29 | 14 | 0.06700 | — | 0 | 3 |
| 18 | *Lepomis macrochirus* | 1 | 30 | 14 | 0.11400 | — | 0 | 3 |
| 19 | *Micropterus salmoides* | 24 | 20 | 28 | 0.03021 | 0.03196 | 0.03244 | 33,34,35 |
| 20 | *Pomoxis nigromaculatus* | 1 | 18 | 11 | — | 0.01000 | 0.00560 | 5 |
| 21 | *Perca flavescens* | 11 | 157 | 44 | 0.00950 | 0.01773 | 0.00672 | 28,39 |
| 22 | *Cottus cognatus* | 4 | 11 | 33 | — | 0.00287 | 0.00427 | 42 |
| Mammals | | | | | | | | |
| 23 | *Didelphis virginiana* | 6 | 14 | 31 | 0.11567 | 0.11117 | 0.07931 | 27 |
| 24 | *Myotis californicus* | 1 | 32 | 21 | 0.12600 | 0.11600 | 0.14333 | 38 |
| 25 | *Cervus elephas* | 1 | — | 24 | — | 0.01200 | 0.01647 | 9 |
| 26 | *Odocoileus virginianus* | 2 | 108 | 22 | 0.12700 | 0.12050 | 0.11513 | 36 |
| 27 | *Marmota flaviventris* | 1 | — | 20 | 0.08000 | 0.07500 | 0.05308 | 37 |
| 28 | *Spermophilus tridecemlineatus* | 10 | 9 | 28 | — | 0.07440 | 0.03785 | 11 |
| 29 | *Thomomys talpoides* | 10 | 28 | 31 | 0.04680 | 0.05570 | 0.04709 | 32 |
| 30 | *Dipodomys deserticola* | 1 | 13 | 17 | — | 0.01300 | 0.01817 | 23 |
| 31 | *Dipodomys merriami* | 7 | 35 | 17 | 0.04857 | 0.05029 | 0.04898 | 23 |
| 32 | *Dipodomys ordii* | 9 | 44 | 17 | 0.00989 | 0.01022 | 0.01246 | 23 |
| 33 | *Neotoma floridana* | 5 | 7 | 20 | 0.06300 | 0.10700 | 0.11078 | 41 |
| 34 | *Peromyscus leucopus* | 6 | 18 | 27 | 0.07400 | 0.07150 | 0.07726 | 6 |
| 35 | *Peromyscus maniculatus* | 21 | 21 | 26 | 0.06400 | 0.11114 | 0.08093 | 17,29 |
| 36 | *Sigmodon hispidus* | 1 | 30 | 23 | 0.02633 | 0.02467 | 0.28383 | 30 |
| 37 | *Microtus pennsylvanicus* | 1 | 79 | — | — | — | 0.06358 | 26 |
| 38 | *Zapus hudsonius* | 9 | 11 | 21 | 0.01556 | 0.01078 | 0.00739 | 20 |
| 39 | *Ochotoma princeps* | 5 | 39 | 26 | 0.01300 | 0.04100 | 0.01696 | 18 |

*Means were calculated across proteins and then across populations.

response variable. We estimated variance components in the two-way taxon–protein analysis by least-squares, using the Type 1 method in the SAS procedure VARCOMP (SAS Institute, 1985). Negative estimates have been set to zero in the tables. The residual mean square in all analyses confounds true error and the interaction of population with protein.

## Results

### *One-way analysis of species means*

We compared the species mean heterozygosities in our material with those in the much larger dataset collated by Nevo *et al.* (1984). For 16 species of freshwater fish,

**Table 2** Summary of data by protein. See text for definitions of protein classifications Function: V = variable; R = regulatory; N = non-regulatory

| Code | Protein | EC No. | Group | Function | Number of subunits | MW of subunit × $10^3$ | Number of fish species | Fish mean $H_p$ | Number of mammal species | Mammal mean $H_p$ |
|---|---|---|---|---|---|---|---|---|---|---|
| AAT | Aspartate aminotransferase | 2.6.1.1 | 1 | N | 2 | 46 | 16 | 0.054 | 14 | 0.035 |
| ACON | Aconitase | 4.2.1.3 | — | — | 1 | 74 | 2 | 0.000 | 0 | — |
| ACP | Acid phosphatase | 3.1.3.2 | 2 | V | — | 41 | 8 | 0.029 | 5 | 0.031 |
| ADH | Alcohol dehydrogenase | 1.1.1.1 | 2 | R | 2 | 40 | 12 | 0.020 | 5 | 0.012 |
| ADK | Adenylate kinase | 2.7.4.3 | 1 | R | 1 | 31 | 11 | 0.0002 | 1 | 0.025 |
| AGP | α-glycerophosphate dehydrogenase | 1.1.1.8 | 1 | — | — | — | 6 | 0.015 | 0 | — |
| ALB | Albumin | — | — | — | — | — | 0 | — | 16 | 0.054 |
| ALD | Aldolase | 4.1.2.13 | 1 | N | 4 | 38 | 12 | 0.039 | 1 | 0.000 |
| ALDH | Alanine dehydrogenase | 1.4.1.1 | — | — | — | — | 2 | 0.000 | 0 | — |
| ALKP | Alkaline phosphatase | 3.1.3.1 | 2 | V | 2 | 69 | 3 | 0.006 | 1 | 0.000 |
| AO | Aldehyde oxidase | 1.2.3.1 | — | — | 2 | — | 1 | 0.000 | 0 | — |
| AP | Amino peptidase | 3.4.11.1 | — | — | — | — | 2 | 0.000 | 0 | — |
| CAT | Catalase | 1.11.1.6 | — | — | 4 | 60 | 0 | — | 1 | 0.000 |
| CK | Creatine kinase | 2.7.3.2 | — | — | 2 | 41 | 11 | 0.001 | 0 | — |
| CMP | Cathodal muscle protein | — | — | — | — | — | 1 | 0.000 | 0 | — |
| DIA | Diaphorase | 1.6.4.3 | — | — | — | — | 6 | 0.000 | 0 | — |
| EST | Esterase | 3.1.1.1 | 2 | V | — | 35 | 12 | 0.085 | 11 | 0.149 |
| FDP | Fructose diphosphatase | — | — | — | — | — | 2 | 0.000 | 0 | — |
| FUM | Fumerase | 4.2.1.2 | 1 | — | 4 | 49 | 3 | 0.000 | 0 | — |
| G6PD | Glucose-6-phosphate dehydrogenase | 1.1.1.49 | 1 | R | 2 | 53 | 7 | 0.022 | 5 | 0.028 |
| GA3P | Glyceraldehyde-3-phosphate dehydrogenase | 1.2.1.12 | 1 | R | 4 | 36 | 5 | 0.000 | 1 | 0.000 |
| GDH | Glutamate dehydrogenase | 1.4.1.3 | — | — | — | — | 5 | 0.000 | 3 | 0.000 |
| GK | Glucokinase | 2.7.1.2 | — | — | — | — | 1 | 0.000 | 0 | — |
| GO3P | Glycerol-3-phosphate dehydrogenase | 1.1.1.8 | 1 | R | 2 | 36 | 6 | 0.031 | 13 | 0.058 |
| GUS | β-gluconronidase | — | — | — | — | — | 1 | 0.000 | 0 | — |
| H6PD | Hexose-6-phosphate dehydrogenase | — | 1 | — | — | — | 2 | 0.000 | 1 | 0.311 |
| HB | Haemoglobin | — | — | — | — | — | 1 | 0.000 | 13 | 0.103 |
| HBDH | 3-hydroxybutyrate dehydrogenase | 1.1.1.30 | — | — | — | — | 1 | 0.000 | 0 | — |
| HK | Hexokinase | 2.7.1.1 | 1 | R | 1 | 112 | 5 | 0.000 | 0 | — |
| HP | Haptoglobin | — | — | — | — | — | 0 | — | 2 | 0.005 |
| HSD | Homoserine dehydrogenase | 1.1.1.3 | — | — | — | — | 1 | 0.000 | 0 | — |
| IDH | Isocitrate dehydrogenase | 1.1.1.42 | 1 | — | 2 | 48 | 14 | 0.040 | 14 | 0.039 |
| IPO | Indophenol oxidase | 1.9.3.1 | 2 | V | 2 | — | 8 | 0.040 | 13 | 0.016 |
| LAP | Leucine aminopeptidase | 3.4.1.1 | — | — | 1 | — | 4 | 0.000 | 6 | 0.165 |
| LDH | Lactate dehydrogenase | 1.1.1.27 | 1 | N | 4 | 35 | 22 | 0.002 | 16 | 0.033 |
| LEDH | L-leucine dehydrogenase | 1.4.1.9 | — | — | — | — | 2 | 0.000 | 0 | — |

**Table 2** *Continued*

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| MDH | Malate dehydrogenase | 1.1.1.37 | 1 | N | 2 | 35 | 19 | 0.011 | 16 | 0.006 |
| ME | Malic enzyme | 1.1.1.40 | 1 | R | 4 | 60 | 9 | 0.031 | 6 | 0.000 |
| MPD | Mannose phosphate dehydrogenase | — | — | — | — | — | 1 | 0.000 | 0 | — |
| MPI | Mannose phosphate isomerase | 5.3.1.8 | — | — | 1 | 43 | 5 | 0.000 | 0 | — |
| ODH | Octopine dehydrogenase | 1.5.1.11 | — | — | — | — | 1 | 0.000 | 0 | — |
| PALB | Prealbumin | — | — | — | — | — | 0 | — | 2 | 0.038 |
| PEP | Peptidase | 3.4.13.11 | 2 | V | — | 54 | 8 | 0.070 | 3 | 0.028 |
| PGDH | Phosphogluconate dehydrogenase | 1.1.1.44 | 1 | N | 2 | 52 | 15 | 0.070 | 15 | 0.109 |
| PGI | Phosphoglucose isomerase | 5.3.1.9 | 1 | R | 2 | 62 | 16 | 0.037 | 11 | 0.009 |
| PGM | Phosphoglucomutase | 2.7.5.1 | 1 | R | 1 | 55 | 18 | 0.035 | 14 | 0.065 |
| SDH | Sorbitol dehydrogenase | 1.1.1.14 | 1 | N | 4 | 38 | 10 | 0.002 | 8 | 0.141 |
| SHDH | Shikimic acid dehydrogenase | 1.1.1.25 | — | — | — | — | 1 | 0.000 | 0 | — |
| SOD | Superoxide dismutase | 1.15.1.1 | — | — | — | 17 | 11 | 0.026 | 3 | 0.000 |
| SUDH | Succinate dehydrogenase | — | — | — | — | — | 0 | — | 1 | 0.000 |
| TF | Transferrin | — | — | — | — | — | 0 | — | 14 | 0.130 |
| TP | Nonspecific protein | — | — | — | — | — | 6 | 0.027 | 10 | 0.061 |
| TPI | Triosephosphate isomerase | 5.3.1.1 | 1 | N | 2 | 26 | 2 | 0.000 | 1 | 0.066 |
| XDH | Xanthine dehydrogenase | 1.1.1.37 | 1 | — | 2 | — | 10 | 0.000 | 2 | 0.000 |

we have mean ± s.d. of $H_{obs} = 0.043 \pm 0.032$, while for 183 species of marine and freshwater fish Nevo *et al.* report $H_{obs} = 0.051 \pm 0.035$. The means of these two samples are not significantly different ($t_{197} = 0.86$, $P > 0.10$); neither are the variances ($F_{182,15} = 1.21$, $P > 0.25$). For 13 species of mammal, we have $H_{obs} = 0.062 \pm 0.041$, while for 184 species Nevo *et al.* report $H_{obs} = 0.041 \pm 0.035$. Our material appears to be somewhat more heterozygous ($t_{195} = 2.12$, $0.02 < P < 0.05$) but the variances do not differ ($F_{12,184} = 1.42$, $P > 0.10$). On the whole, therefore, our material seems reasonably representative of fish and mammals generally, and encompasses a similar range of variation among species.

One-way analysis of variance shows that heterozygosity varies widely among species. For 95 populations of fish distributed over 15 species, analysis of $H_{exp}$ shows that 48 per cent of the total variance is due to species differences ($F_{14,80} = 6.42$, $P < 0.0001$). For 95 populations of mammals distributed over 16 species, 70 per cent of the variance is due to species ($F_{15,79} = 14.2$, $P < 0.0001$). It is variance of this sort that has fuelled previous comparative studies of heterozygosity. This variance disappears almost completely when a two-way analysis is attempted.

### *Two-way analysis of species and protein*

The complete two-way analysis of taxon and protein is shown in Table 3. Less than 5 per cent of the variance of $H_p$ is either by taxon, at all levels, or by protein. More than half the variance is attributable to taxon–protein interaction, the great bulk of this, amounting to 44 per cent of the overall variance in $H_p$, being represented by the species–protein interaction. Figure 1 depicts this interaction. Taken at face value, this result invalidates attempts to interpret variation either among species or among enzymes, and indicates that variation should be analysed only for a given enzyme among species, or only for a given species among enzymes. Because this is a much more onerous task, we attempted to show that the very large interaction variance is either artefactual or misleading.

First, we re-analysed fish and mammals separately. For fish, this enabled us to use a drainage basin to classify samples within species, and reduce the residual mean square. The taxonomic component remained about the same (6 per cent among families, zero at other levels), while the small protein component in the main analysis disappeared. About three-quarters of overall variance was attributable to interaction: 15 per cent to order–protein, 28 per cent to species–protein, and 31 per cent to drainage–protein. The large drainage–protein variance is especially noteworthy

**Table 3** Sources of variation in protein heterozygosity. The residual variance comprises errors of estimation and sample–protein interaction. Variance components are calculated by least-squares, taking protein and taxon as random effects, successive taxonomic levels being nested in the order given. Because of the unbalanced design, significance could be determined only for the species–protein interaction effect: $F_{203,2174} = 4.79$, $P < 0.0001$

| Source of variance | d.f. | Mean square | Variance component (%) |
|---|---|---|---|
| Protein | | | |
| Protein | 53 | 0.3177 | 4.4 |
| Taxon | | | |
| Class | 1 | 0.2570 | 0 |
| Order | 6 | 0.2980 | 0 |
| Family | 10 | 0.3259 | 3.0 |
| Species | 21 | 0.0872 | 0 |
| Population | 160 | 0.0325 | 1.6 |
| Taxon × protein interaction | | | |
| Class × protein | 27 | 0.1846 | 0 |
| Order × protein | 116 | 0.1414 | 12.8 |
| Family × protein | 137 | 0.0869 | 0 |
| Species × protein | 203 | 0.0862 | 44.1 |
| Residual | 2174 | 0.0180 | 34.2 |

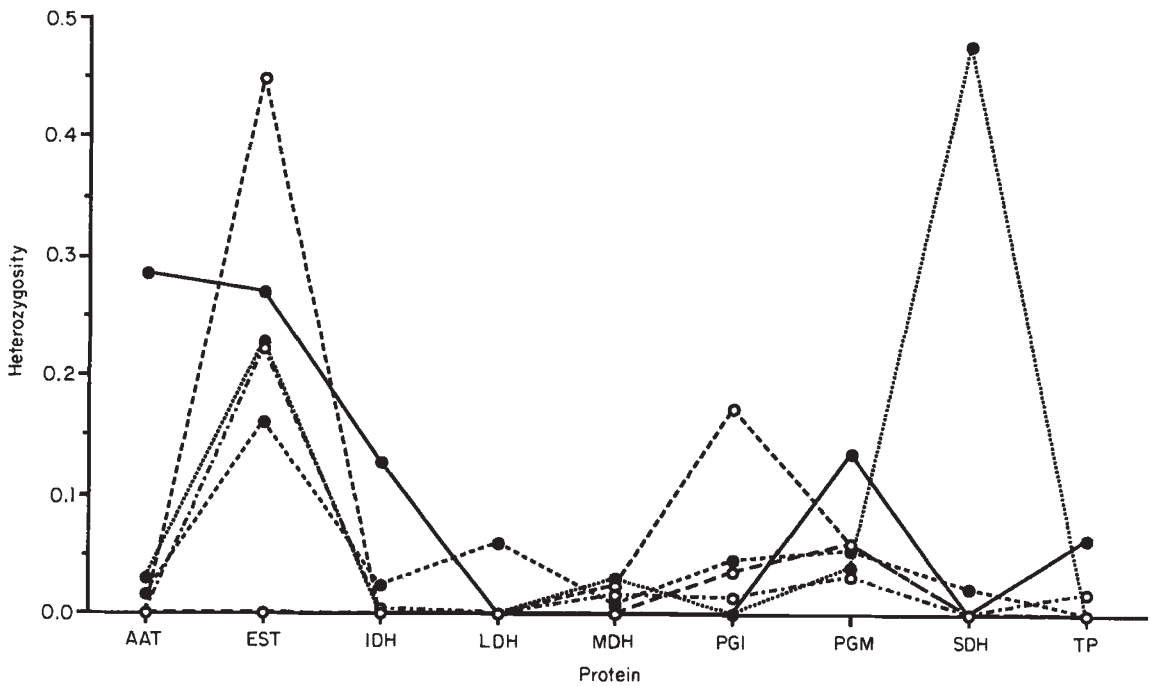


**Fig. 1** The species–protein interaction in explaining variation in protein heterozygosity, $H_p$. Proteins are ordered alphabetically. Species and protein codes are defined in Tables 1 and 2, respectively. Lines connect points for each species to highlight the interaction. The proteins are those reported in the highest numbers of species, and the species are those with the highest numbers of proteins studied, within mammals: (—●—) 35, (···●···) 24, (- -●- -) 29; and fish: (----o----) 6, (-·-o-·-) 8, (- -o- -) 21.

because it implies that the distribution of heterozygosity varies not only between species, but also between isolates of the same species. The result for mammals is virtually the same as in the main analysis: the taxonomic component remains at about 5 per cent, the protein component again disappears, and somewhat more than half the variance is contributed by taxon–protein interaction. Thus, performing the analysis within fish or mammals leaves the result of the main analysis unchanged, as would be expected from the absence of variance at the class–protein level.

Secondly, we argued that the result might be an artefact introduced by the unbalanced nature of the dataset. To test this, we constructed completely balanced subsets of the data. For fish, we used the seven proteins most often reported (AAT, IDH, LDH, MDH, PGDH, PGI, PGM; see Table 2), for two populations each of four species (1, 3, 19, 21; see Table 1). The main effects of protein (4 per cent of total variance) and species (9 per cent) remained small and non-significant, while the species–protein interaction contributed 80 per cent of the variance and was significant at $P<0.0005$. For mammals, we used the eight proteins most often reported (AAT, ALB, IDH, LDH, MDH, PGDH, PGM, TF; see Table 2), for five populations of each of six species (28, 29, 31, 32, 34, 39; see Table 1). The main effects of species and protein were negligible (1 per cent each) and non-significant, while the interaction remained large (39 per cent) and highly significant ($P<0.0005$). Thus, the use of completely balanced datasets confirmed the result of the main analysis.

To illustrate why the species variance component is so markedly reduced in the two-way analysis relative to the one-way analysis, we compared the compositions of the expected mean squares between the two analyses using the balanced mammal dataset. Table 4 shows that in the two-way analysis the expected mean square for species contains both species–protein interaction and population-within-species variance, whereas in the one-way analysis it contains only population-within-species variance. If the species variance in the two-way analysis is estimated while omitting the species–protein variance from the expected mean square, then the species variance (0.0020) is the same in the two analyses. Thus, the species variance estimate from the one-way analysis is partitioned into species and species–protein interaction components in the two-way analysis. This partition, together with the larger total variance in the two-way analysis, results in a smaller percentage of the overall variance being attributable to species.

Thirdly, it has been suggested that a major determinant of levels of heterozygosity is the laboratory in which they are measured (Selander, 1976; Simon & Archie, 1985). This bias might contribute to the species–protein and drainage–protein interactions because the data for a given species or drainage is often derived from a single study. To investigate the effect of study bias on the species–protein interaction among fish, we chose five studies (4, 3, 5, 16, 19; see Appendix), each of which scored more than one species, and none of which had any species in common. Thus, species were nested within study. The main effect of study was very small (3 per cent), but the suspicion of study bias is probably well-founded as 19 per cent of the variance was attributable to study–protein interaction. Nevertheless, the species–protein interaction remained overwhelming (67 per cent of total variance). To investigate the effect of study bias on the drainage–protein interaction, we identified seven studies (39, 8, 13, 31, 22, 24, 33; see Appendix) in each of which a single species was scored from several drainage basins. Only two variance components were substantial. The species–protein interaction is large (32 per cent) but cannot be interpreted because it is confounded with any study–protein interaction. The drainage–protein interaction remains large (36 per cent), confirming that it is independent of study. For mammals, the only comparable analysis that we could attempt was to analyse separately the only study (23; see Appendix) dealing with more than one species. This yielded a large main effect for protein (29 per cent), but only 12 proteins were scored; there was no significant variance among species (7 per cent). The species–protein interaction remained large (28 per cent) and highly significant ($F_{22,154}=4.65$, $P<0.0005$). Thus, we were unable to show that study bias contributes substantially to the taxon–protein interaction. These results also discount the possibility that the interaction is due to non-homologous loci coding for the same protein being resolved in different studies.

Fourthly, we recognized that our estimates of $H_p$ comprised monomorphic and polymorphic proteins, which might be interpreted in different ways. We therefore reanalysed our data by excluding all monomorphic proteins. The results were unchanged. Among fish, species–protein (50 per cent) and drainage–protein (25 per cent) interactions remained large, and the same was true for the species–protein interaction (37 per cent) among mammals, all other sources of variation being zero or small. Thus, the results of the main analysis apply both to all proteins and to polymorphic proteins only.

Finally, we investigated the possibility that classifying enzymes by their form or function would account for taxon–protein interaction in terms of the taxon–enzyme type interaction. The classification of enzymes

**Table 4** Sources of variation in protein heterozygosity. Results are for a balanced subset of the mammal data (see text). Details of analysis are given in Table 3

| Sources of variance | d.f. | Mean square | Expected mean square |
|---|---|---|---|
| One-way analysis | | | |
| Species | 5 | 0.0124 | Var(population) + 5Var(species) |
| Population | 24 | 0.0025 | Var(population) |
| Two-way analysis | | | |
| Protein | 7 | 0.0986 | Var(population × protein) + 5Var(species × protein) + 30Var(protein) |
| Species | 5 | 0.0995 | Var(population × protein) + 5Var(species × protein) + 8Var(population) + 40Var(species) |
| Population | 24 | 0.0201 | Var(population × protein) + 8Var(population) |
| Species × protein | 35 | 0.0863 | Var(population × protein) + 5Var(species × protein) |
| Population × protein | 168 | 0.0202 | Var(population × protein) |

| Variance component | Variance | Total variance (%) | $F$(d.f.) | $P$ |
|---|---|---|---|---|
| One-way analysis | | | | |
| Var(species) | 0.0020 | 44.2 | 4.96(5,24) | $<0.005$ |
| Var(population) | 0.0025 | 55.8 | | |
| Total | 0.0045 | | | |
| Two-way analysis | | | | |
| Var(protein) | 0.0004 | 1.2 | 1.14(7,35) | $>0.25$ |
| Var(species) | 0.0003 | 1.0 | 1.13(7,49)* | $>0.25$ |
| Var(population) | 0.000 | 0.0 | $<1$ | |
| Var(species × protein) | 0.0132 | 38.7 | 4.27(35,168) | $<0.0005$ |
| Var(population × protein) | 0.0202 | 59.2 | | |
| Total | 0.0341 | | | |

*Approximate $F$ calculated according to Damon & Harvey (1987, p. 69).

by function according to Gillespie & Langley (1974), by function according to Johnson (1974), by quaternary structure and by subunit size is described in the Materials and Methods section. Enzyme classification was considered a fixed effect in analyses of variance. None of these four classifications was of any value in decomposing the species–protein or drainage–protein interaction. Variance attributable to species and to enzyme (nested within enzyme type) remained small, as in previous analyses, both for fish and for mammals, except that there was a moderately large contribution of enzyme within all classifications (11–18 per cent) except quaternary structure for mammals. The interaction of species with enzyme type was very small (0–2 per cent) except for a moderately large value (12 per cent) for species-quaternary structure in mammals. The leading result of this analysis is that the interaction of species with enzyme (within type) remains very large in mammals (41–45 per cent), while in fish both the species–enzyme (27–38 per cent) and drainage–enzyme (19–29 per cent) interactions remain large. Thus, these interactions are not removed, or even reduced, by taking into account enzyme form or function.

In short, we have investigated the possibilities that the very large species–protein interaction discovered in the main analysis can be attributed to pooling fish and mammals; to the use of unbalanced datasets; to study bias; to including monomorphic proteins; or to failing to take into account enzyme form and function. None of these possibilities suggest any substantial change to the original conclusion. We feel justified in concluding

that variance among taxa and variance among proteins is very small, and that the great bulk of variation in protein heterozygosity is attributable to the taxon–protein interaction.

## Discussion

### *Lack of variation among species*

The heterozygosity ($H_p$) of different proteins is somewhat correlated among samples of the same species, but poorly correlated among species. The poor correlation among species is equivalent to the substantial variance contributed by species–protein interaction. When the mean expected heterozygosity for a species is calculated, this interaction variance is suppressed. The mean heterozygosities will nevertheless vary to some extent, and this variance will be substantial relative to the residual variance of populations within species. This is why the one-way analysis yields highly significant results, with classification into species contributing a large proportion of the overall variance. The two-way analysis reveals the true nature of this variance to be species–protein interaction. The same conclusion holds below the species level for fish populations in different drainage basins, and perhaps also for taxa above the species level.

It seems difficult to justify the use of species mean heterozygosity when the main effect is clearly so much smaller than the interaction. Nelson & Hedgecock (1980) make a similar point, based on their finding in decapod crustacea, that the correlations between the heterozygosity of an enzyme and ecological variables may depend on the enzyme's function as defined by Gillespie & Langley (1974). Contrary to this finding, however, our analyses showed that neither enzyme function nor form helped explain the species–protein or drainage–protein interaction. It is equally difficult to justify the use of protein heterozygosity averaged across species, for similar reasons. The strong interaction effect may be a serious impediment to comparative studies of the mean heterozygosity of species or of proteins because the correlations that are discovered will depend on the particular set of species and proteins which are used. It seems necessary to compare species for a single protein only, or to compare proteins within a single species only. With hindsight, this conclusion might not be very surprising. Species vary with respect to size, shape, coloration and other aspects of external morphology, but we do not know of any general rule that species in which one aspect of morphology is exceptionally variable tend also to be highly variable with respect to other, independent aspects of morphology. Some species of the gastropod *Cepaea* are extremely variable with respect to shell coloration, while they do not appear to be equally variable with respect to shell size or shape, but this has never caused any surprise. The averaging of variation across proteins may simply reflect our ignorance of their function in natural populations.

### *Interpretation of the species–protein interaction*

Our results suggest that attention should shift from attempting to interpret variation in mean heterozygosity to the interpretation of the much greater variation represented by species–protein and drainage–protein interactions. According to the neutral theory, when genetic differentiation of populations occurs by mutation and genetic drift, the correlation $r$ of single-locus expected heterozygosities between two populations decreases over time $t$ as:

$$r = \exp[-(4v + 1/N)t],$$

where $v$ is the mutation rate and $N$, the effective population size (Li & Nei, 1975). It is expected to take a long time for $r$ to become nearly zero when $N$ is large. For example, if $v$ for electrophoretically detectable alleles is assumed to be about $10^{-7}$ (Nei 1987) and $N$ is $10^6$, then it will take $4.9 \times 10^6$ years for $r$ to become 0.001. As populations diverge, interaction variance is generated by the decaying correlation. No main effect variance need be generated if the population mean heterozygosities do not diverge significantly. The correlation will continue to decay exponentially to zero, after which no additional interaction variance will be generated. The observed taxonomic distribution of interaction variance components (Table 3) suggests that populations diverged rapidly initially, so that the largest interaction component is at the species level, and then the rate of divergence slowed so that a small proportion of the total interaction variance remains detectable at the order level. The reason for a lack of interaction variance at the family level is unclear. Unfortunately, this pattern is also expected under natural selection and there is no simple way to distinguish between the two processes using these observations. The neutral theory, however, makes firm and explicit predictions relating mean heterozygosity to population size (and thus body size) among species and to subunit size among proteins. It does not seem to anticipate the absence or near absence of taxon and protein main effects: small proteins in small populations should display the least variation, while large proteins in large populations should display the most variation. Theories of variation, based on natural selection, do not encounter the same difficulty because it is easy to imagine that the way in which selection acts on

different proteins may depend on the species in which they are expressed. In this sense, our result runs counter to the neutral theory, and appears to support a selectionist interpretation of protein variation. This is a long way from directly demonstrating the importance of selection in maintaining variation; the whole trend of our argument is to suggest that such direct evidence can come only from detailed studies of particular species or particular proteins, rather than from broad surveys of mean heterozygosity. Nevertheless, we hope to have directed attention away from mean heterozygosity and towards the species–protein interaction as a major element in the interpretation of genetic variation.

## Acknowledgements

We would like to thank Dan Schoen and Kurt Sittmann for helpful discussions. This work was supported by a postdoctoral fellowship from the Natural Sciences and Engineering Research Council of Canada to JdaS; by an NSERC Operating Grant to GB; and by an NSERC scholarship to AB.

## References

DAMON, R. A. JR. AND HARVEY, W. R. 1987. *Experimental Design, ANOVA, and Regression.* Harper & Row, New York.

EANES, W. F. AND KOEHN, R. K. 1978. Relationship between subunit size and number of rare electrophoretic alleles in human enzymes. *Biochem. Genet.*, **16**, 971–985.

GILLESPIE, J. H. AND LANGLEY, C. H. 1974. A general model to account for enzyme variation in natural populations. *Genetics*, **76**, 837–848.

HAMRICK, J. L., LINHART, Y. B. AND MITTON, J. B. 1979. Relationships between life-history characters and electrophoretically detectable genetic variation in plants. *Ann. Rev. Ecol. Syst.*, **10**, 173–200.

HARRIS, H., HOPKINSON, D. A. AND EDWARDS, Y. H. 1977. Polymorphism and the subunit structure of enzymes: a contribution to the neutralist-selectionist controversy. *Proc. Nat. Acad. Sci., U.S.A.*, **74**, 698–701.

HOPKINSON, D. A., EDWARDS, Y. H. AND HARRIS, H. 1976. The distribution of subunit numbers and subunit sizes of enzymes: a study of the products of 100 human gene loci. *Ann. Hum. Genet.*, **39**, 383–411.

JOHNSON, G. B. 1974. Enzyme polymorphism and metabolism. *Science*, **184**, 28–37.

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

KOEHN, R. K. AND EANES, W. F. 1977. Subunit size and genetic variation of enzymes in natural populations of *Drosophila*. *Theoret. Pop. Biol.*, **11**, 330–341.

LEIGH BROWN, A. J. AND LANGLEY, C. H. 1979. Reevaluation of genic heterozygosity in natural populations of *Drosophila melanogaster* by two-dimensional electrophoresis. *Proc. Nat. Acad. Sci., U.S.A.*, **76**, 2381–2384.

LEWONTIN, R. C. 1974. *The Genetic Basis of Evolutionary Change.* Columbia University Press, New York.

LI, W.-H. AND NEI, M. 1975. Drift variances of heterozygosity and genetic distance in transient states. *Genet. Res.*, **25**, 229–248.

MITTON, J. B. AND LEWIS, W. M. 1989. Relationships between genetic variability and life-history features of bony fishes. *Evolution*, **43**, 1712–1723.

NEI, M. 1987. *Molecular Evolutionary Genetics.* Columbia University Press, New York.

NEI, M., FUERST, P. A. AND CHAKRABORTY, R. 1978. Subunit molecular weight and genetic variability of proteins in natural populations. *Proc. Nat. Acad. Sci., U.S.A.*, **75**, 3359–3362.

NEI, M. AND GRAUR, D. 1984. Extent of protein polymorphism and the neutral mutation theory. *Evol. Biol.*, **17**, 73–118.

NELSON, K. AND HEDGECOCK, D. 1980. Enzyme polymorphism and adaptive strategy in the decapod Crustacea. *Am. Nat.*, **116**, 238–280.

NEVO, E., BEILES, A. AND BEN-SHLOMO, R. 1984. The evolutionary significance of genetic diversity: ecological, demographic and life-history correlates. In: Mani, G. S. (ed.) *Evolutionary Dynamics of Genetic Diversity*, Springer-Verlag, Berlin, pp. 13–213.

PETERS, R. H. 1983. *The Ecological Implications of Body Size.* Cambridge University Press, Cambridge.

SAS INSTITUTE. 1985. *SAS User's Guide: Statistics, Version 5 Edition.* SAS Institute, Inc., Cary, NC.

SELANDER, R. K. 1976. Genic variation in natural populations. In: Ayala, F. J. (ed.) *Molecular Evolution*, Sinauer Associates, Sunderland, MA, pp. 21–45.

SIMON, C. AND ARCHIE, J. 1985. An empirical demonstration of the lability of heterozygosity estimates. *Evolution*, **39**, 463–467.

WARD, R. D. 1977. Relationships between enzyme heterozygosity and quaternary structure. *Biochem. Genet.*, **15**, 123–135.

WARD, R. D. 1978. Subunit size of enzymes and genetic heterozygosity in vertebrates. *Biochem. Genet.*, **16**, 799–810.

WOOTEN, M. C. AND SMITH, M. H. 1985. Large mammals are genetically less variable? *Evolution*, **39**, 212–215.

ZOUROS, E. 1975. Electrophoretic variation in allelozymes related to function or structure? *Nature*, **254**, 446–448.

## Appendix

1. AQUADRO, C. F. AND KILPATRICK, C. W. 1981. Morphological and biochemical variation and differentiation in insular and mainland deer mice (*Peromyscus maniculatus*). In: Smith, M. H. and Joule, J. (eds), *Mammalian Population Genetics*, University of Georgia Press, Athens, pp. 214–230.
2. AVISE, J. C. AND AYALA, F. J. 1976. Genetic differentiation in speciose versus depauperate phylads: evidence from the California minnows. *Evolution*, **30**, 46–58.
3. AVISE, J. C. AND SMITH, M. H. 1974. Biochemical genetics of sunfish II. Genic similarity between hybridizing species. *Am. Nat.*, **108**, 458–472.

4. AVISE, J. C., SMITH, J. J. AND AYALA, F. J. 1975. Adaptive differentiation with little genic change between two native California minnows. *Evolution*, **29**, 411–426.
5. AVISE, J. C., STRANEY, D. O. AND SMITH, M. H. 1977. Biochemical genetics of sunfish IV. Relationships of centrarchid genera. *Copeia*, **1977**, 250–258.
6. BROWNE, R. A. 1977. Genetic variation in island and mainland populations of *Peromyscus leucopus*. *Am. Midl. Nat.*, **97**, 1–9.
7. BUTH, D. G. 1979. Biochemical systematics of the cyprinid genus *Notropis* — 1. The subgenus *Luxilus*. *Biochem. Syst. Ecol.*, **7**, 69–79.
8. BUTH, D. G. 1980. Evolutionary genetics and systematic relationships in the catostomid genus *Hypentelium*. *Copeia*, **1980**, 280–290.
9. CAMERON, D. G. AND VYSE, E. R. 1978. Heterozygosity in Yellowstone Park elk, *Cervus canadensis*. *Biochem. Genet.*, **16**, 651–657.
10. CASSELMAN, J. M., COLLINS, J. J., CROSSMAN, E. J., IHSSEN, P. E. AND SPANGLER, G. R. 1981. Lake whitefish (*Coregonus clupeaformis*) stocks of the Ontario waters of Lake Huron. *Can. J. Fish. Aquat. Sci.*, **38**, 1772–1789.
11. COTHRAN, E. G., ZIMMERMAN, E. G. AND NADLER, C. F. 1977. Genic differentiation and evolution in the ground squirrel subgenus *Ictidomys* (genus *Spermophilus*). *J. Mammal.*, **58**, 610–622.
12. DEHRING, T. R., BROWN, A. F., DAUGHERTY, C. H. AND PHELPS, S. R. 1981. Survey of the genetic variation among eastern Lake Superior lake trout (*Salvelinus namaycush*). *Can. J. Fish. Aquat. Sci.*, **38**, 1738–1746.
13. DOWLING, T. E. AND BROWN, W. M. 1989. Allozymes, mitochondrial DNA and levels of phylogenetic resolution among four minnow species (*Notropis*: Cyprinidae). *Syst. Zool.*, **38**, 126–143.
14. DOWLING, T. E. AND MOORE, W. S. 1985. Genetic variation and divergence of the sibling pair of cyprinid fishes, *Notropis cornutus* and *N. chrysocephalus*. *Biochem. Syst. Ecol.*, **13**, 471–476.
15. FERRIS, S. D. AND WHITT, G. S. 1977. Duplicate gene expression in diploid and tetraploid loaches (Cypriniformes, Cobitidae). *Biochem. Genet.*, **15**, 1097–1112.
16. FERRIS, S. D. AND WHITT, G. S. 1980. Genetic variability in species with extensive gene duplication: the tetraploid catostomid fishes. *Am. Nat.*, **115**, 650–666.
17. GILL, A. E. 1976. Genetic divergence of insular populations of deer mice. *Biochem. Genet.*, **14**, 835–848.
18. GLOVER, D. G., SMITH, M. H., AMES, L., JOULE, J. AND DUBACH, J. M. 1977. Genetic variation in pika populations. *Can. J. Zool.*, **55**, 1841–1845.
19. GOODFELLOW, W. L. JR., HOCUTT, C. H., MORGAN, II, R. P. AND STAUFFER, J. R. JR. 1984. The biochemical assessment of the taxonomic status of '*Rhinichthys bowersi*' (Pisces: Cyprinidae). *Copeia*, **1984**, 652–659.
20. HAFNER, D. J., PETERSON, K. E. AND YATES, T. L. 1981. Evolutionary relationships of jumping mice (genus *Zapus*) of the southwestern United States. *J. Mammal.*, **62**, 501–512.
21. HEALY, J. A. AND MULCAHY, M. F. 1980. A biochemical genetic analysis of populations of the northern pike, *Esox lucius* L., from Europe and North America. *J. Fish Biol.*, **17**, 317–324.
22. IHSSEN, P. E., CASSELMAN, J. M. AND MARTIN, G. W. 1988. Biochemical genetic differentiation of lake trout (*Salvelinus namaycush*) stocks of the Great Lakes Region. *Can. J. Fish. Aquat. Sci.*, **45**, 1018–1029.
23. JOHNSON, W. E. AND SELANDER, R. K. 1971. Protein variation and systematics in kangaroo rats (genus *Dipodomys*). *Syst. Zool.*, **20**, 377–405.
24. KELSCH, S. W. AND HENDRICKS, F. S. 1986. An electrophoretic and multivariate morphometric comparison of the American catfishes *Ictalurus lupus* and *I. punctatus*. *Copeia*, **1986**, 646–652.
25. KIRKPATRICK, M. AND SELANDER, R. K. 1979. Genetics of speciation in lake whitefishes in the Allegash basin. *Evolution*, **33**, 478–485.
26. KOHN, P. H. AND TAMARIN, R. H. 1978. Selection at electrophoretic loci for reproductive parameters in island and mainland voles. *Evolution*, **32**, 15–28.
27. KOVACIC, D. A. AND GUTTMAN, S. I. 1979. An electrophoretic comparison of genetic variability between eastern and western populations of the opossum (*Didelphis virginiana*). *Am. Midl. Nat.*, **101**, 269–277.
28. LEARY, R. AND BOOKE, H. E. 1982. Genetic stock analysis of yellow perch from Green Bay and Lake Michigan. *Trans. Am. Fish. Soc.*, **111**, 52–57.
29. LOUDENSLAGER, E. J. 1978. Variation in the genetic structure of *Peromyscus* populations. I. Genetic heterozygosity — its relationship to adaptive divergence. *Biochem. Genet.*, **16**, 1165–1179.
30. McCLENAGHAN, L. R. JR. 1980. The genetic structure of an isolated population of *Sigmodon hispidus* from the lower Colorado River Valley. *J. Mammal.*, **61**, 304–307.
31. MERRITT, R. B., ROGERS, J. F. AND KURZ, B. J. 1978. Genic variability in the longnose dace, *Rhinichthys cataractae*. *Evolution*, **32**, 116–124.
32. NEVO, E., KIM, Y. J., SHAW, C. R. AND THAELER, C. S. JR. 1974. Genetic variation, selection and speciation in *Thomomys talpoides* pocket gophers. *Evolution*, **28**, 1–23.
33. PHILIPP, D. P., CHILDERS, W. F. AND WHITT, G. S. 1982. *Biochemical Genetics of Largemouth Bass*. Electric Power Research Institute, Palo, Alto, California.
34. PHILIPP, D. P., CHILDERS, W. F. AND WHITT, G. S. 1983. A biochemical genetic evaluation of the northern and Florida subspecies of largemouth bass. *Trans. Am. Fish. Soc.*, **112**, 1–20.
35. PHILIPP, D. P., CHILDERS, W. F. AND WHITT, G. S. 1985. Correlation of allele frequencies with physical and environmental variables for populations of largemouth bass, *Micropterus salmoides* (Facepede). *J. Fish. Biol.*, **27**, 347–365.
36. RAMSEY, P. R., AVISE, J. C., SMITH, M. H. AND URBSTON, D. F. 1979. Biochemical variation and genetic heterozygosity in South Carolina deer populations. *J. Wildl. Manage.*, **43**, 136–142.
37. SCHWARTZ, O. A. AND ARMITAGE, K. B. 1981. Social substructure and dispersion of genetic variation in the yellow-

bellied marmot (*Marmota flaviventris*). In: Smith, M. H. and Joule, J. (eds) *Mammalian Population Genetics*, University of Georgia Press, Athens, pp. 139–159.

38. STRANEY, D. O., O'FARRELL, M. J. AND SMITH, H. 1976. Biochemical genetics of *Myotis californicus* and *Pipistrellus hesperus* from southern Nevada. *Mammalia*, **40**, 344–347.
39. STRITTHOLT, J. R., GUTTMAN, S. I. AND WISSING, T. E. 1988. Low levels of genetic variability of yellow perch (*Perca flavescens*) in Lake Erie and selected impoundments. In: Downhower, J. F. (ed.), *The Biogeography of the Island Region of Western Lake Erie*, Ohio State University Press, Columbus.
40. YATES, T. L., LEWIS, M. A. AND HATCH, M. D. 1984. Biochemical systematics of three species of catfish (Genus *Ictalurus*) in New Mexico. *Copeia*, **1984**, 97–101.
41. ZIMMERMAN, E. G. AND NEJTEK, M. E. 1977. Genetics and speciation of three semispecies of *Neotoma*. *J. Mammal.*, **58**, 391–402.
42. ZIMMERMAN, E. G. AND WOOTEN, M. C. 1981. Allozymic variation in sculpins, *Cottus confuses* and *Cottus cognatus*. *Biochem. Syst. Evol.*, **9**, 341–346.