A study of interlocus allozyme heterozygosity correlations: implications for neutral theory

M. WOODWARK, D. O. F. SKIBINSKI & R. D. WARD

School of Biological Sciences, University College of Swansea, Singleton Park, Swansea SA2 8PP, U.K. CSIRO, Division of Fisheries, G.P.O. Box 1538, Hobart, Tasmania 7001, Australia

Using a database of allozyme studies, correlations in heterozygosity between selected enzyme loci (MDH, αGPDH, IDH, 6PGDH, LDH, SOD, AAT, PGM, EST, PGI) were calculated across vertebrate species. Large and positive correlations were observed with untransformed heterozygosity values. However, after transformation to correct for mean species heterozygosity, correlations were substantially reduced and median values were closer to zero. Some enzymes were more often involved in significant correlations than others, and correlations calculated across species within vertebrate classes were significant for different enzyme pairs in different classes. There was no evidence that significant correlations occurred primarily between functionally related enzymes. It is suggested that the observed correlations are best explained by variation between enzyme loci in functional constraint and effective neutral mutation rate.

Keywords: allozymes, correlations, database, heterozygosity, neutral theory.

Introduction

Under neutral theory genetic drift does not occur independently at linked loci. Instead it leads to heterozygosity correlations across replicate populations and to linkage disequilibrium (Ohta & Kimura, 1969). The magnitude of the effect increases with decreasing recombination freqency (r) and decreasing population size (N). For example, neutral disequilibrium will only be favoured when $4Nr < 10$, that is in very small populations or when there is tight linkage (Hudson, 1985). Epistatic selection can, however, give rise to disequilibrium between loci which are not closely linked and might also give rise to heterozygosity correlations. Thus the search for disequilibrium and heterozygosity correlations provides a means to test the neutral theory.

Numerous studies have been made of linkage disequilibrium in natural populations. Disequilibrium (D) values for allozyme loci are generally close to zero, unless the loci are closely linked, or inversions are involved, or the studies are of selfing plants (Clegg *et al.*, 1972; Mukai *et al.*, 1974; Weir *et al.*, 1974; Brown *et al.*, 1977; Langley *et al.*, 1977, 1978). Linkage disequilibrium has been observed for closely linked blood group loci such as the MNSs loci (Race & Sanger, 1975), the HLA system (Hedrick *et al.*, 1986) and certain genetic diseases (Thomson & Bodmer, 1977).

Thus experimental results are largely in accord with the expectations of neutral theory.

Little work has been carried out to test the neutral theory by examining heterozygosity correlations. This may be due to problems associated with choosing data suitable for a test of the neutral theory null hypothesis of zero correlation. The loci must not be very closely linked, population samples must be available in order to estimate heterozygosity, populations contributing to a given correlation should not be closely related by descent from a common ancestor, and it must be possible to establish homologies of loci in distantly related populations. In addition, many independent estimates of heterozygosity must be obtainable especially if the correlations are small. DNA sequence data does not at present satisfy these criteria. The present study of heterozygosity correlations is carried out using allozyme data of which there is an abundance in the literature, which do satisfy the above criteria.

A precedent for the approach used in this paper is provided by the work of Gauldie (1984) who examined allozyme data for five enzymes in a range of vertebrate and invertebrate species and found evidence of a negative heterozygosity correlation between PGM and PGI. Selection was invoked to explain this result. In the present study, heterozygosity correlations among 10 enzymes were computed and analysed using a database for vertebrate species including representatives from

all five vertebrate classes. Many significant correlations were observed although these were generally small in magnitude. The occurrence of significant correlations does not appear to be taxon-dependent, however some enzymes are more frequently involved in significant correlations than others.

Materials and methods

Analyses were carried out using a database of allozyme studies of natural outbreeding animal populations. Most of the data has been obtained from published sources but unpublished data is also included (see Acknowledgements). References for much of the data are given in Skibinski & Ward (1981) and Ward & Skibinski (1985), but a complete list is available on request from the authors. For allozyme results to be incorporated into the database, certain selection criteria have to be satisfied. Species must be scored for 15 or more allozyme loci, with 15 or more individuals sampled per locus.

Heterozygosity according to Hardy–Weinberg expectation was computed for each species, for each locus from the observed allele frequencies. Where several loci were scored in a species for a given enzyme one of the loci was chosen at random for analysis and data for the other loci discarded. Sets of populations within a single species were not used to provide several estimates of heterozygosity per locus because their heterozygosity values are expected to be correlated through descent from a common ancestor, particularly in closely related populations (Chakraborty et al., 1978). In this situation, allele frequencies were averaged over populations and expected heterozygosity calculated for each locus from these species averages.

Following computation of the expected heterozygosity for each locus for each species, a two-dimensional species by enzyme matrix of heterozygosity values was drawn up. Rows and columns were then eliminated from the matrix to generate a new species by enzyme matrix without holes in which every remaining species had a heterozygosity value for every remaining enzyme. It was found that the total amount of data remaining was close to its maximum when all but the 10 most frequently scored enzymes in the database were eliminated. Thus the analysis was carried out with these enzymes. The chosen enzymes were: MDH (E.C. 1.1.1.37), αGPDH (E.C. 1.1.1.8), IDH (E.C. 1.1.1.42), 6PGDH (E.C. 1.1.1.44), LDH (E.C. 1.1.1.27), SOD (E.C. 1.15.1.1), AAT (E.C. 2.6.1.1), PGM (E.C. 5.4.2.2), EST (E.C. 3.1.1.-), PGI (E.C. 5.3.1.9). One hundred and twenty-three species were scored for these 10 enzymes.

It is known that species and higher order taxonomic groups differ in mean heterozygosity (Nevo et al. 1984). This might reflect differences in effective population size or selective influences. Thus correlations of heterozygosity between loci across species could be augmented by a factor relating to differences in the overall level of heterozygosity between species. If, for example, species differ in effective population size, then according to the neutral theory, mean heterozygosity will vary across species and heterozygosity correlations will reflect variation in this parameter. The advantage of using a dataset without holes is that it allows correction for this effect. This is achieved through transformations of the heterozygosity values.

The first transformation, giving h_{ij} and carried out by division by the species average, corrects for differences in the overall heterozygosity as a scale transformation. The second and third transformations directly address the problem of variation in effective population size by using equations derived for expected heterozygosity for the infinite allele model (Wright, 1949) and the more realistic stepwise mutation model (Ohta & Kimura, 1973). These are:

$H = 1 - 1/(M + 1)$ (Kimura & Crow, 1964) for the infinite allele model,

and $H = 1 - 1/(2M + 1)^{1/2}$ (Ohta & Kimura, 1973) for the stepwise mutation model,

where $M = 4Nu$, N being the effective population size and u the neutral mutation rate. Rearrangement of these equations gives

$M = H/(1 - H)$ for the infinite allele model

and $M = H(2 - H)/(1 - H)^2$ for the stepwise mutation model.

Using these formulae, an estimate of M from the H values for each enzyme locus can be calculated for both models. If the estimated value of M for each enzyme is divided by the average M value for the species, the term for effective population size, N, cancels out. Thus for enzyme j of species i

$$\bar{M}_i = \frac{1}{10} \sum_{j=1}^{10} M_{ij}$$

$$= \frac{1}{10} \sum_{j=1}^{10} 4N_i u_{ij} = 4N_i \bar{u}_i$$

and

$$m_{ij} = M_{ij}/\bar{M}_i = 4N_i u_{ij}/4N_i \bar{u}_i = u_{ij}/\bar{u}_i.$$

The analytical approach is similar to that devised by Zouros (1978) for estimation of the relative neutral mutation rate. Values of m_{ij} were calculated for each enzyme for each species using the equations for M for the infinite allele model (transformation 2) and the

stepwise mutation model (transformation 3). The structure of the data matrix and transformations are represented algebraically in Table 1.

Non-parametric correlations in heterozygosity values between each of the 45 possible pairs of enzymes across the 123 species were calculated. Correlations were similarly computed for h (transformation 1) and m (transformations 2 and 3). The correlations of m will not be confounded with variation in N between species but will measure the correlation of the neutral mutation rate between enzymes, although the correlations will be confounded by variation in the average species neutral mutation rate ($\bar{u}$).

Other tests were performed on the dataset of 123 species. The dataset was divided up into subsets (the five vertebrate classes) and the correlation analyses described were repeated for each subset. These tests were performed to give indications of any taxon-specific contribution to correlations in heterozygosity between enzyme loci and to test the robustness of correlations that had repeatedly been found to be significant when vertebrate data were considered as a whole.

Results

Spearman, Rank, correlation coefficients among the 10 chosen enzymes are shown in Fig. 1. Each correlation is represented twice, appearing in the columns for both enzymes in the correlation. Nine out of 45 correlations are significant for both the Kendall and Spearman tests (see Sokal & Rohlf, 1981, p. 601 and 607, respectively). Both correlations were performed on the untransformed data and on the data subject to all three transformations but as the results of the two tests are similar,

Table 1 Structure of data matrix. Heterozygosity as tabulated for $i = 123$, $j = 10$. There are no holes in this dataset. Transformation formulae are given below

Species	Enzymes						
	MDH	α – GPDH	IDH	6PGDH	LDH	PGI	Mean
1	$H_{1,1}$	$H_{1,2}$	$H_{1,3}$	$H_{1,4}$	$H_{1,5...}$	$H_{1,10}$	$\bar{H}_1$
2	$H_{2,1}$	$H_{2,2}$	$H_{2,3}$	$H_{2,4}$	$H_{2,5...}$	$H_{2,10}$	$\bar{H}_2$
3	$H_{3,1}$	$H_{3,2}$	$H_{3,3}$	$H_{3,4}$	$H_{3,5...}$	$H_{3,10}$	$\bar{H}_3$
4	$H_{4,1}$	$H_{4,2}$	$H_{4,3}$	$H_{4,4}$	$H_{4,5}$	$H_{4,10}$	$\bar{H}_4$
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
123	$H_{123,1}$	$H_{123,2}$	$H_{123,3}$	$H_{123,4}$	$H_{123,5}$	$H_{123,10}$	H_{123}

Transformation 1

$$\bar{H}_i = \frac{1}{10} \sum_{j=i}^{10} H_{ij}$$

$$h_{ij} = H_{ij}/\bar{H}_i$$

Transformation 2 (infinite allele model)

$$M_{ij} = H_{ij}/(1 - H_{ij})$$

$$\bar{M}_i = \frac{1}{10} \sum_{j=i}^{10} M_{ij}$$

$$m_{ij} = M_{ij}/\bar{M}_i$$

Transformation 3 (stepwise mutation model)

$$M_{ij} = H_{ij}(2 - H_{ij})/(1 - H_{ij})^2$$

$$\bar{M}_i = \frac{1}{10} \sum_{j=i}^{10} M_{ij}$$
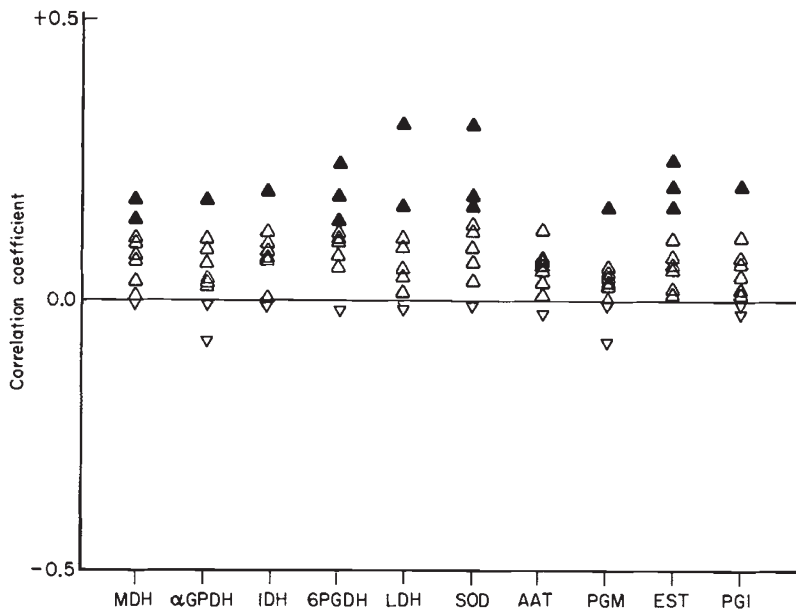
$$m_{ij} = M_{ij}/\bar{M}_i$$

Fig. 1 Graphs of raw heterozygosity correlation distribution. Spearman Rank correlation coefficients). (▲) correlation coefficient significant at 5 per cent level, (△)non-significant.
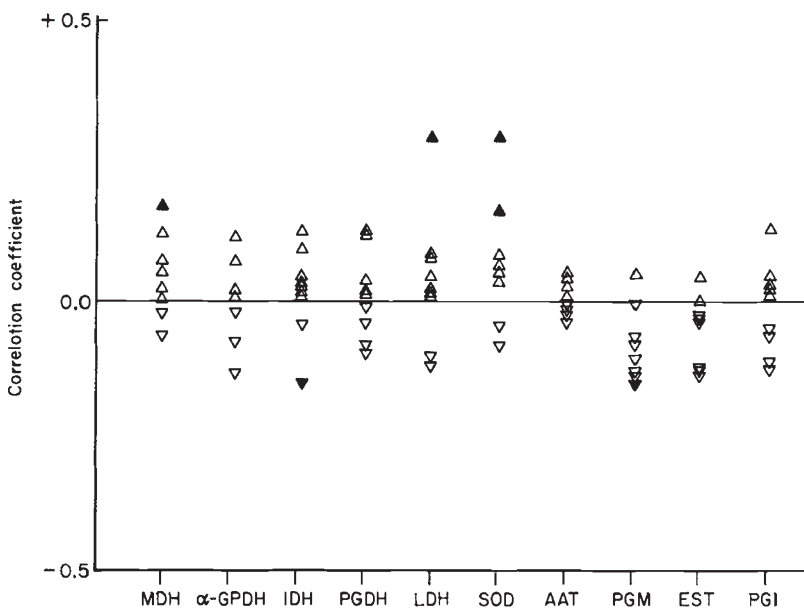

Fig. 2 Graph of effect on distribution of Spearman Rank correlation coefficients of dividing raw data by mean heterozygosity over 10 loci. Symbols as in Fig. 1.

only the results of the more conservative Spearman Rank test will be discussed. The correlations are predominantly positive for the raw data (Fig. 1).

Figure 2 shows the effect on the correlation of dividing each locus heterozygosity by the mean heterozygosity of the 10 enzymes (transformation 1). The effect of species level heterozygosity is confirmed because the number of significant correlations is reduced to three out of 45, and now there are similar numbers of positive and negative values. Figure 3 shows the correlation coefficient values between the estimated m values obtained after transformation 2 (infinite allele model). The results from transformation 3 (stepwise mutation

model) are almost identical to those obtained by using the infinite allele model and thus have not been represented graphically. The results are very similar to those produced by transformation 1 (Fig. 2), showing that these three transformations are roughly equivalent in effect. For all three transformations, the same three or four pairs of loci give significant correlations, with LDH-SOD having the most consistently high values in this dataset. The correlations are generally small, seldom exceeding 0.2 even in Fig. 1, where the number of significant values is greatest.

A high correlation of heterozygosity between two enzymes in the whole dataset can be attributed either

to correlations within each of the subsets (birds, fish, mammals, reptiles, amphibia) or to correlated differences in mean heterozygosity between subsets. In order to test for these effects, each taxonomic group was analysed separately. The number of significant correlations and the median and inter-quartile range of correlation coefficients for each group are given in Table 2 for untransformed data and for the three transformations. Significant correlations were obtained for all the subsets, and for birds and fish the numbers of significant values exceed those obtained for the pooled data for all but the raw heterozygosity correlations. Thus the correlations for the pooled data cannot simply be attributed to correlated differences in mean between subsets.

The inter-quartile ranges of the correlation values given in Table 2 are small, particularly for the transformed data although some of the individual correlations are quite large and significant. For the untransformed data, the median and range suggests an overall tendency towards positive correlations. Following transformation, this tendency is reduced, if not eliminated.

In order to ascertain whether the significant correlations in each subset could have arisen by chance, two tests were performed on the set of correlation coefficients for the whole dataset and for each taxonomic subset separately for each transformation (totalling 24 sets of 45 correlations). First, an *a posteriori* test of significance was carried out. The sequential Bonferroni test of table wide significance (Rice, 1989) was used.

The results of this procedure indicate that, at the 5 per cent significance level, the enzyme pair LDH–SOD was significant for the pooled dataset with each transformation. In birds, 6PGDH–PGM also proved to be significant at the 5 per cent significance level across all four tranformations. The only other significant result was obtained for LDH–PGI in fish for the untransformed dataset. The majority of correlation coefficients significant in the original *a priori* tests are no longer significant following the Bonferroni procedure. There can be some confidence, however, that the few remaining significant correlations reflect real associations between the enzymes.

Secondly, Bartlett's test of sphericity (SPSS-X User's Guide, 1988, p. 485) was carried out to ascertain whether the correlation matrices were singular. The null hypothesis of this test is that all the diagonal values on the matrix have correlation coefficients of 1, and that all the non-diagonal values have correlation coefficients of zero. Product moment correlation coefficients were substituted for the non-parametric values for this test. The results for the raw heterozygosity data indicate that the null hypothesis could be rejected for reptiles ($P < 0.0001$) and the pooled data ($P < 0.05$), but not for any of the other subsets. Because all three transformations were based on a division of the raw heterozygosity data by a mean calculated from all 10 enzymes, Bartlett's test could not be used directly, due to a loss of independence. This problem was circumvented by reducing the number of enzymes to be correlated by one to nine. This had the effect of removing nine corre-

Table 2 Median and inter-quartile range of Spearman Rank correlation coefficients (untransformed and transformations 1–3) for the whole dataset and taxonomic subsets

Class	Untransformed		Transformation 1: Divided by mean 10 loci		Transformation 2: Infinite allele model		Transformation 3: Stepwise mutation model	
	Median	Inter-quartile range	Median	Inter-quartile range	Median	Inter-quartile range	Median	Inter-quartile range
Whole dataset	0.072	0.024–0.127	0.015	−0.053–0.054	0.015	−0.055–0.063	0.018	−0.052–0.062
N	9		3		4		4	
Birds	0.047	−0.198–0.203	0.022	−0.207–0.219	0.049	−0.194–0.214	0.054	−0.194–0.215
N	7		5		6		7	
Fish	0.078	−0.065–0.262	0.012	−0.159–0.206	0.017	−0.153–0.194	0.017	−0.154–0.194
N	7		7		8		9	
Mammals	0.037	−0.048–0.107	−0.015	−0.136–0.079	−0.010	−0.125–0.088	−0.010	−0.127–0.097
N	4		3		2		2	
Reptiles	0.077	−0.015–0.173	−0.022	−0.183–0.107	−0.020	−0.174–0.113	−0.033	−0.163–0.120
N	5		3		3		3	
Amphibia	0.105	−0.027–0.300	0.003	−0.155–0.126	−0.012	−0.177–0.145	−0.019	−0.174–0.142
N	4		2		2		2	

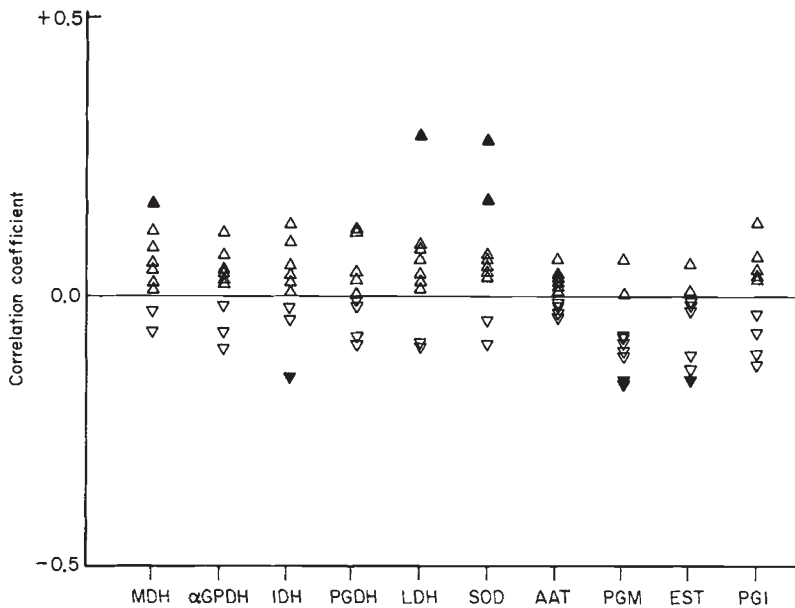N = number of significant ($P < 0.05$) correlations out of 45.

Fig. 3 Graph of effect of infinite allele model transformation on raw data. Symbols as in Fig. 1.

lation coefficients associated with the excised enzyme, leaving a matrix of 36 correlations. The test was thus performed 10 times for each set of correlation coefficients, removing a different enzyme each time. The results for transformation 2 (infinite allele model) alone are reported for this test as it has been shown previously that all three transformations are very similar in effect. Significant results $(P < 0.05)$ were obtained in 75 per cent of cases on removal of the enzyme, which indicates that the correlation matrices were not singular. With regard to the pooled data, all but one of the subsets of nine enzymes proved to be significant at the 0.001 per cent level.

An additional analysis has been carried out to assess the variation in transformed correlation values using transformation 3 (stepwise mutation model). As there are five subsets and 45 enzyme pairs the total number of correlation values is $5 \times 45 = 225$. There are an additional 45 values for the pooled dataset making a grand total of 270. These 270 values have been ranked from high to low and the non-parametric Kruskal–Wallis test (Sokal & Rohlf, 1981, p. 430) used to test for significant differences in mean rank between different groups of transformed correlation values. Four tests were carried out.

1 The mean ranks of the 45 values for each of the five animal group subsets were compared in a single test.
2 The mean rank of the 225 subset values was compared with the mean rank of the 45 pooled dataset values.
3 The mean ranks of the six values for each of the 45 enzyme pairs were compared in a single test.
4 Of the 270 correlation values in the five subsets and

pooled dataset a given enzyme will be involved in $6 \times 9 = 54$ of these. For each enzyme, the mean rank of these 54 values was compared with the mean rank of the remaining $270 - 54 = 216$ correlation values not involving the enzyme.

Table 3 shows the mean ranks for each of the 45 enzyme pairs that were compared in test **3**. The correlation coefficient values are also shown for each enzyme pair for the pooled dataset and which of the individual taxonomic groups exhibit significant correlation values.

Tests **1** and **2** were not significant, indicating no overall difference in the strength of correlations among the subsets and total dataset. This confirmed that the generally positive correlations for the total dataset were unlikely to be due to correlated differences in heterozygosity between the animal groups, which suggested that the correlations occurred within the animal groups, not between them. Test **3** was significant $(\chi^2_{44} = 61.432, P = 0.042)$, which suggested that across all subsets there were differences between enzyme pairs.

In test **4**, an extension of test **3**, enzyme pairs involving PGM and enzyme pairs involving EST were shown to have significantly lower mean rankings than the remaining enzyme pairs $(P < 0.05)$ and enzyme pairs involving αGPDH had significantly higher rankings $(P < 0.05)$. Because correlations are calculated for pairs of enzymes the results for different enzymes are not independent of one another. For example, it can be seen from Table 3 that although αGPDH usually occurs high up in the ranking it also occurs at the bottom when it is paired with PGM, an enzyme with low

Table 3 Results of Kruskal-Wallis test for association (test 3 in text)

Locus pair	Kruskal–Wallis rank	All group r	Locus pair significant in these groups
LDH–SOD	202.33	0.2980**	2, 5, 6
αGPDH–SOD	199.50	0.0842	5
αGPDH–AAT	195.67	0.0345	—
αGPDH–IDH	186.00	0.0285	1
IDH–PGI	183.17	0.1425	1
MDH–SOD	181.67	0.1782*	2, 6
MDH–PGI	173.25	0.0710	—
MDH–αGPDH	172.58	0.0262	—
LDH–PGM	172.25	0.0679	—
MDH–6PGDH	168.00	0.1270	—
MDH–LDH	164.17	0.0932	—
6PGDH–PGM	160.33	−0.0709	1, 2
αGPDH–6PGDH	158.83	0.1230	1
IDH–LDH	155.42	0.0959	3
αGPDH–LDH	153.50	0.0369	—
AAT–PGM	152.33	0.0125	2
MDH–IDH	152.08	0.0448	—
EST–SOD	151.33	0.0643	4
6PGDH–SOD	149.83	0.0449	—
αGPDH–PGI	149.67	−0.0635	—
IDH–AAT	144.33	0.0473	2
LDH–PGI	139.50	0.0377	2, 4
EST–MDH	138.67	0.0181	—
6PGDH–PGI	137.67	0.0416	3
EST–αGPDH	137.58	−0.0271	—
IDH–6PGDH	132.17	0.0032	—
LDH–AAT	126.83	0.0105	—
EST–AAT	126.67	−0.0120	—
SOD–PGI	125.50	0.0458	2
SOD–AAT	121.25	0.0605	—
AAT–PGI	117.50	−0.0272	1
MDH–AAT	114.50	−0.0253	—
PGM–PGI	113.83	−0.1007	—
EST–6PGDH	111.67	0.0320	—
IDH–SOD	109.17	−0.0398	—
6PGDH–AAT	108.67	−0.0094	—
SOD–PGM	99.83	−0.0853	—
EST–LDH	94.17	−0.1028	2
EST–IDH	92.83	−0.0081	—
LDH–6PGDH	79.83	−0.0878	—
MDH–PGM	77.50	−0.0655	—
PGM–αGPDH	70.50	−0.0917	1
EST–PGI	69.67	−0.1398	—
EST–PGM	63.00	−0.1623*	6
IDH–PGM	62.75	−0.1541*	6

r = Spearman Rank correlation coefficient of heterozygosity. Data transformed using the stepwise mutation model (Ohta & Kimura, 1973).
Key to taxonomic groups: 1 = Birds, 2 = Fish, 3 = Mammals, 4 = Reptiles, 5 = Amphibia, 6 = All groups pooled before analysis.
Level of significance: ** = $P < 0.010$; *$P < 0.050$.

mean ranking. The results of tests **1-4** suggest that taxon-specific effects have a smaller influence than enzyme-specific effects on heterozygosity correlations.

Discussion

The results of this study provide evidence for an overall tendency towards significant heterozygosity correlations between enzymes across species. A few individual enzyme pairs were identified as being significant in an *a posteriori* test. In addition, evidence was obtained to indicate that some enzymes are more often involved in significant correlations than others. Transformations which correct for the effect of differences in overall heterozygosity between species, markedly reduce the general tendency towards positive correlations.

Significant correlations in heterozygosity can be explained in a number of ways. Variation in functional constraint between enzyme loci will lead to variation in effective neutral mutation rate and thence to variation in heterozygosity values, with higher functional constraint associated with lower heterozygosity. Both the positive and negative correlations could be explained under neutral theory by positive or negative correlations between enzyme loci in neutral mutation rate across taxa. Thus if two enzymes are both subject to high functional constraint in some species and low constraint in others, a correlation in effective neutral mutation rate will occur and lead at equilibrium to a positive heterozygosity correlation. Functional constraint can arise from a number of causes. Enzymes on a main metabolic pathway carry out an essential role in metabolism, and thus may be subject to more constraint than other enzymes (Gojobori, 1982). Enzymes that are substrate-specific are thought to be more constrained than those that can accept a wider range of substrates (Gojobori, 1982). Multimeric enzymes and those with smaller subunit sizes are also thought to be under higher constraint (Koehn & Eanes, 1978; Nei *et al.*, 1978; Ward, 1978). Other factors that might be involved in constraint include whether an enzyme metabolises glucose (group I) or not (group II) (Kojima *et al.*, 1970) and whether or not it is involved in regulation (Johnson, 1974). All these properties might act in concert to provide an overall functional constraint for an enzyme which will affect its heterozygosity. If these properties vary in a taxon-specific way then heterozygosity correlations might result.

Metabolic and regulatory properties of enzymes might also be involved in a selective interpretation of heterozygosity correlations. For two enzymes it is necessary to postulate that balancing selection favours situations where both enzymes are polymorphic or both have low levels of variation. Koehn *et al.* (1988) suggested that selection might work on biochemical pathways rather than single enzymes to minimize pathway maintenance costs. The costs would be reduced if the activity of successive steps was adjusted to ensure that build up of intermediate substrates did not occur. Because the activity of allozyme heterozygotes tends to be intermediate between the activity of corresponding homozygotes (Falconer, 1989), such a situation might occur when individuals were either multiply heterozygous or multiply homozygous. At the population level, this situation might be achieved when pairs of loci have either high or low levels of polymorphism.

One feature of selective explanations of heterozygosity correlations is that the enzymes involved should be functionally related in some way. The observation of a negative heterozygosity correlation between PGM and PGI was discussed by Gauldie (1984) in these terms. The present study, using a different dataset, does not replicate this result; moreover, significant correlations are observed, after the application of the Bonferroni procedure, between what appear to be functionally related enzymes (6PGDH and PGM are negatively correlated in birds) and functionally unrelated enzymes (LDH and SOD are positively correlated in the pooled dataset). A functional relationship is not required by the constraint argument. Thus while the results for 6PGDH-PGM are consistent with the expectations of selection theory, significant correlations for both 6PGDH-PGM and LDH-SOD are in accord with expectations of neutral theory.

Acknowledgements

References

BROWN, A. H. D., NEVO, E AND ZOHARY, D. 1977. Association of alleles at esterase loci in wild barley, *Hordeum spontaneum* L. *Nature*, **268**, 430–431.

CHAKRABORTY, R., FUERST, P. A. AND NEI, M. 1978. Statistical studies on protein polymorphisms in natural populations. II. Gene differentiation between populations. *Genetics*, **88**, 367–390.

CLEGG, M. T., ALLARD, R. W. AND KAHLER, A. L. 1972. Is the gene the unit of selection? Evidence from two experimental plant populations. *Proc. Natl. Acad. Sci. U.S.A.*, **69**, 2474–2478.

FALCONER, D. S. 1989. *Introduction to Quantitative Genetics.* Longman Scientific and Technical, Harlow, Essex, U.K.

GAULDIE, R. W. 1984. A reciprocal relationship between heterozygosities of the phosphoglucomutase and glucose phosphate isomerase loci. *Genetica*, **63**, 93–104.

GOJOBORI, T. 1982. Means and variances of heterozygosity and protein function. In: *Molecular Evolution, Protein Polymorphism and the Neutral Theory*, Kimura M. (ed.) Japanese Scientific Press, Tokyo.

HENDRICK, P.W., THOMSON, G. AND KLITZ, W. 1986. Evolutionary genetics: HLA as an exemplary system. In: *Evolutionary Processes and Theory* Karlin, S. and Nevo, E. (eds) Academic Press, New York.

HUDSON, R. R. 1985. The sampling distribution of linkage disequilibrium under an infinite allele model without selection. *Genetics*, **109**, 611–631.

JOHNSON, G. B. 1974. Enzyme polymorphism and metabolism. Polymorphism among enzyme loci is related to enzyme function. *Science*, **184**, 28–37.

KIMURA, M. AND CROW, J. F. 1964. The number of alleles that can be maintained in a finite population. *Genetics*, **49**, 725–738.

KOEHN, R. K., DIEHL, W. J. AND SCOTT, T. M. 1988. The differential contribution by individual enzymes of glycolysis and protein catabolism to the relationship between heterozygosity and growth rate in the coot clam, *Mulinia lateralis*. *Genetics*, **118**, 121–130.

KOEHN, R. K. AND EANES, W. F. 1978. Molecular structure and protein variation within and among populations. *Evol. Biol.*, **11**, 39–100.

KOJIMA, K., GILLESPIE, J. AND TOBARI, Y. N. 1970. A profile of *Drosophila* species' enzymes assayed by electrophoresis.I. Number of alleles, heterozygosities, and linkage disequilibrium in glucose-metabolizing systems and some otherenzymes. *Biochem. Genet.*, **4**, 627–637.

LANGLEY, C. H., ITO, K. AND VOELKER, R. A. 1977. Linkage disequilibrium in natural populations of *Drosophila melanogaster*. Seasonal variation. *Genetics*, **86**, 447–454.

LANGLEY, C. H., SMITH, D. B. AND JOHNSON, F. M. 1978. Analysis of linkage disequilibria between allozyme loci in natural populations of *Drosophila melanogaster*. *Genet. Res.*, **32**, 215–230.

MUKAI, T., WATANABE, T. K. AND YAMAGUCHI, O. 1974. The genetic structure of natural populations of *Drosophila melanogaster*. XII. Linkage disequilibrium in a large local population. *Genetics*, **77**, 771–793.

NASS, C. A. G. 1959. The χ^2 test for small expectations in contingency tables with special reference to accidents and absenteeism. *Biometrika*, **46**, 366–385.

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.

NEVO, E., BEILES, A. AND BEN-SHLOMO, R. 1984. The evolutionary significance of genetic diversity: ecological, demographic and life history correlates. *Lect. Not. Biomath.*, **53**, 13–213.

OHTA, T. AND KIMURA, M. 1969. Linkage disequilibrium due to random genetic drift. *Genet. Res.*, **13**, 47–55.

OHTA, T. AND KIMURA, M. 1973. A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. *Genet. Res.*, **22**, 201–204.

RACE, R. R. AND SANGER, R. 1975. *Blood Groups in Man*, 6th edn, J. B. Lippincott, Philadelphia.

RICE, W. R. 1989. Analyzing tables of statistical tests. *Evolution*, **43**, 223–225.

SKIBINSKI, D. O. F. AND WARD, R. D. 1981. Relationship between allozyme heterozygosity and rates of divergence. *Genet. Res.*, **38**, 71–92.

SOKAL, R. R. AND ROHLF, F. J. 1981. *Biometry*. Freeman, San Francisco.

SPSS-X User's Guide, 3rd edn 1988. SPSS-X Inc., Chicago.

THOMSON, G. AND BODMER, W. F. 1977. The genetics of HLA and disease associations. In: Christiansen, F. B. and Fenchel, T. M. (eds) *Measuring Selection in Natural Populations*. Springer-Verlag, New York. pp. 545–564.

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

WARD, R. D. AND SKIBINSKI, D. O. F. 1985. Observed relationships between protein heterozygosity and protein genetic distance and comparisons with neutral expectations. *Genet. Res.*, **45**, 315–340.

WEIR, B. S., ALLARD, R. W. AND KAHLER, A. L. 1974. Further analysis of complex allozyme polymorphisms in a barley population. *Genetics*, **78**, 911–919.

WRIGHT, S. 1949. Genetics of populations. In: *Encyclopedia Britannica* 14th ed. **10**. 111, 111A-D, 112, London.

ZOUROS, E. 1978. Mutation rates, population sizes and amounts of electrophoretic variation of enzyme loci in natural populations. *Genetics*, **92**, 623–646.