Genet. Res., Camb. (1981), **38**, *pp.* 71–92
With 4 text-figures
Printed in Great Britain

Relationship between allozyme heterozygosity and rates of divergence

By D. O. F. SKIBINSKI

Department of Genetics, University College of Swansea, Singleton Park, Swansea SA2 8PP, U.K.

and R. D. WARD

Department of Human Sciences, University of Technology, Loughborough, Leicestershire, LE11 3TU, U.K.

(*Received* 8 *July* 1980 *and in revised form* 12 *February* 1981)

SUMMARY

The relationship between locus variability and genetic identity is analysed in pooled data from a collection of *Drosophila* and vertebrate allozyme surveys. For vertebrates the analysis provides evidence that loci with high expected heterozygosity diverge at a faster rate than loci with low heterozygosity in comparisons involving pairs of populations and pairs of species. This relationship is not observed in comparisons of *Drosophila* species pairs.

1. INTRODUCTION

The accumulation of numerous electrophoretic surveys of enzyme variation in animal species over the last decade has provided a large body of allozyme data for use in testing evolutionary hypotheses. For example, many attempts have been made to assess the adaptive significance of observed patterns of allele and heterozygosity variation within and between populations (e.g. Johnson, 1972; Kirby & Halliday, 1973; Ayala & Gilpin, 1974; Yamazaki & Maruyama, 1975; Nei, 1975; Chakraborty, Fuerst & Nei, 1978).

This paper considers the relationship between within population and within species enzyme variability and evolutionary rate or divergence using data extracted from a number of allozyme surveys. The specific hypothesis tested is that more variable loci diverge at a faster rate than less variable loci in comparisons involving pairs of populations or species. Corroboration of this hypothesis would be of significance for evolutionary theory, for it would suggest that in evolution the rate of future divergence at protein loci among taxa could be predicted from knowledge of the contemporary level of genetic variation within taxa (Sokal, 1978; Riska, 1979). Determinations of evolutionary rates by comparing amino acid sequences of homologous proteins in diverse organisms have been made for a small number of proteins (see, for example, Wilson, Carlson & White, 1977), but only a few of these are routinely screened for allozyme variation in population surveys. Thus

a direct comparison between evolutionary rate and heterozygosity cannot be made at present. Positive associations have been reported between within and among population or species variation for allozyme data using a number of different measures of variability and genetic divergence (Johnson & Mickevich, 1977; Koehn & Eanes, 1978; Pierce & Mitton, 1979). Associations have also been reported between within and among population morphometric variation (Kluge & Kerfoot, 1973; Sokal, 1976, 1978). Riska (1979) has, however, drawn attention to a number of potential biases in the interpretation of the relationship between variability and divergence. In this paper we present an approach to analysis which considers the covariation of locus heterozygosity, locus genetic identity and overall genetic identity in pooled data consisting of a large number of paired species or population comparisons taken from a collection of allozyme surveys. We believe the method used is free of bias and provides unambiguous evidence that the more variable loci are indeed those that diverge most rapidly.

2. MATERIALS

Estimates of single locus heterozygosity and identity were made for pairs of populations and species for three separate sets of allele frequency data: *Drosophila* inter-species, vertebrate inter-species and vertebrate intra-species. The mean heterozygosity of *Drosophila* is two or three times that of vertebrates (Selander & Kaufman, 1973; Powell, 1975; Selander, 1976). The *Drosophila* and vertebrate data were therefore analysed separately to avoid confounding effects arising from this difference. The vertebrate inter-species and intra-species data were analysed separately to allow the test of a hypothesis relating to gene flow. The vertebrate data were not differentiated into the classes of fishes, amphibians, reptiles, birds and mammals. Members of these classes show similar overall levels of heterozygosity and similar trends following various analyses of allozyme data (Selander, 1976; Fuerst, Chakraborty & Nei, 1977; Chakraborty, Fuerst & Nei, 1978; Ward, 1978). Only studies screening a minimum of 15 individuals per species or population for a minimum of 15 loci were used. Data were abstracted from the following surveys:

Drosophila (25 species plus 4 subspecies). Yang, Wheeler & Bock, 1972 (4+1); Zouros, 1973 (4+1); Ayala *et al.* 1974 (5+2); Ayala, 1975 (7); Prakash, 1977 (3); Sene & Carson, 1977 (2).

Vertebrate inter-species (153 species plus 7 subspecies). *Fish* (28 species plus 2 subspecies): Johnson, 1975 (5); Gorman, Kim & Rubinoff, 1976 (3); Fujio, 1977 (2); Gorman & Kim, 1977 (2); Buth & Burr, 1978 (3+2); Siebenaller, 1978 (2); Ward & Galleguillos, 1978 (3); Kirkpatrick & Selander, 1979 (2); Kornfield *et al.* 1979 (6). *Amphibians* (24 species plus 2 subspecies): Hedgecock & Ayala, 1974 (3+1); Highton & Webster, 1976 (2); Case, 1978 (5); Feder, Wurst & Wake, 1978 (2); Larson & Highton, 1978 (3+1); Tilley *et al.* 1978 (2); Wake, Maxson & Wurst, 1978 (2); Duncan & Highton, 1979 (5). *Reptiles* (40 species plus 1 subspecies): Webster, Selander & Yang, 1972 (4); Hall & Selander, 1973 (2); Yang, Soulé & Gorman, 1974 (9); Gorman *et al.* 1975 (2); Webster, 1975 (2); Gorman & Kim, 1976 (12); Adest, 1977 (5); Bezy *et al.* 1977 (2+1); Gartside, Rogers & Dessauer, 1977 (2). *Birds* (11 species): Smith & Zimmerman, 1976 (7); Barrowclough & Corbin,

1978 (4). *Mammals* (50 species plus 2 subspecies): Selander, Hunt & Yang, 1969 (1+1); Johnson & Selander, 1971 (8); Selander *et al.* 1971, Smith, Selander & Johnson, 1973, Avise, Smith & Selander, 1974 (4); Johnson *et al.* 1972 (2); Patton, Selander & Smith, 1972 (2); Johnson & Packard, 1974 (3); Kilpatrick & Zimmerman, 1975, 1976, Zimmerman, Hart & Kilpatrick, 1975 (6); King & Wilson, 1975 (2); Greenbaum & Baker, 1976 (2); Patton, MacArthur & Yang, 1976 (2); Straney, O'Farrell & Smith, 1976 (2); Cothran, Zimmerman & Nadler, 1977 (3); Nozawa *et al.* 1977 (4+1); Zimmerman & Nejtek, 1977 (3); Britton & Thaler, 1978 (2); Mascarello, 1978 (4).

Interspecies comparisons of those species typed for more than one population were performed using the unweighted population mean allele frequencies.

Vertebrate intra-species. A representative selection of data was made from the vast amount of published work. 205 sets of population data were analysed from 34 species. *Fish* (55 populations, from 9 species): Avise & Selander, 1972 (8); Frydenberg & Simonsen, 1973 (2); Johnson, 1975 (3, 10, 8, 5); Ward & Beardmore, 1977 (2); Buth & Burr, 1978 (5); Merritt, Rogers & Kurz, 1978 (12). *Amphibia* (51 populations, from 5 species): Matthews, 1975 (15); Nevo, Dessauer & Chuang, 1975 (11); Highton & Webster, 1976 (6); Larson & Highton, 1978 (11, 8). *Reptiles* (43 populations, from 8 species): McKinney *et al.* 1972 (13); Webster, Selander & Yang, 1972 (4, 2); Gorman *et al.* 1975 (9, 7); Taylor & Gorman, 1975 (2); Gartside, Rogers & Dessauer, 1977 (3); Gorman, Kim & Taylor, 1977 (3). *Birds* (4 populations, from 1 species): Nottebohm & Selander, 1972 (4). *Mammals* (52 populations, from 11 species): Johnson & Selander, 1971 (4, 6, 5); Nevo & Shaw, 1972 (4); Smith, Selander & Johnson, 1973 (3); Avise, Smith & Selander, 1974 (6); Nevo *et al.* 1974 (10); Patton, Yang & Myers, 1975 (4); Gill, 1976 (3); Browne, 1977 (5); Cothran, Zimmerman & Nadler, 1977 (2).

3. ANALYSIS

Let x_i and y_i be the frequencies of the ith allele at the jth locus in populations or species X and Y respectively. Then H_j is defined as $(2-\Sigma x_i^2-\Sigma y_i^2)/2$ which is the arithmetic mean of the expected heterozygosity, according to Hardy–Weinberg, at the jth locus in samples X and Y. I_j is defined as $\Sigma x_i y_i/(\Sigma x_i^2 \Sigma y_i^2)^{\frac{1}{2}}$ which is the genetic identity value for X and Y at the jth locus. The overall genetic identity between X and Y is defined as $I = J_{xy}/(J_x J_y)^{\frac{1}{2}}$, where J_{xy}, J_x and J_y are the arithmetic means of $\Sigma x_i y_i$, Σx_i^2, and Σy_i^2 over all loci (Nei, 1972).

For each of the three sets of data all possible pairwise comparisons of populations (or species) were used to calculate values of I, and for each individual comparison values of H_j and I_j were calculated for the j loci.

I is regarded as an overall measure of the amount of divergence at allozyme loci between two populations. It is likely that the value of I is related in some way to the time since splitting of two populations from a common ancestor, as a number of studies have found progressively smaller I values at various taxonomic levels from inter-population to inter-genus comparisons (for example, Ayala, 1975: Avise, 1978). Attempts have also been made to use genetic distance values to estimate the actual time since splitting (for example, Yang, Soulé & Gorman, 1974), though

it is likely that the accuracy of such estimates is low. The use of I as an overall measure of divergence in this study does not assume any particular relationship between I and evolutionary time, and population or species pairs showing similar levels of divergence (as measured by I) may have separated from their common ancestors at quite different times in the past. H_j is regarded as a measure of locus variability and I_j as a measure of locus divergence. A problem that arises in the

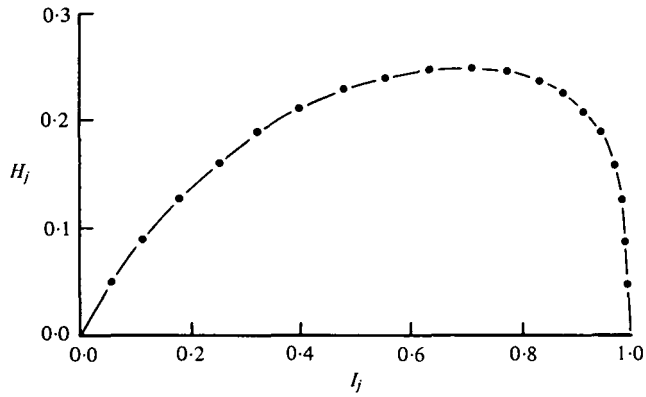

Fig. 1. Theoretical relationship between H_j and I_j for two populations evolving from

Allele	Pop X	Pop Y		Allele	Pop X	Pop Y
1	1·00	1·00		1	1·00	0·00
			to			
2	0·00	0·00		2	0·00	1·00

In allele frequency steps of 0·05. See text.

analysis of such data is that different values of I_j calculated from a single published survey are not all independent in the sense that knowledge of some I_j values might allow deduction of others. For example, with three populations fixed identically at a locus only two of the three comparisons possible will be independent. In general with k populations there will be $k-1$ independent comparisons out of a total of $k!/(k-2)!2! = k(k-1)/2$. This value for the number of independent comparisons is certainly conservative; for example, if the three populations were fixed differently, knowledge of two I_j values would not allow deduction of the third. To take account of any bias that might arise through the use of non-independent comparisons each of the three sets of data has been analysed twice, first using all possible comparisons, and second using independent comparisons only. Independent comparisons were made by pairing population 1 with population 2, then population 2 with population 3 and so on to obtain $k-1$ I values (with their associated I_j and H_j values) for each individual survey. The three original sets of data thus gave the following six sets for analysis: DROS – *Drosophila*, all possible comparisons; DROSI – *Drosophila*, independent comparisons only; VES – vertebrate inter-species, all comparisons; VESI – vertebrate inter-species, independent comparisons; VAS – vertebrate intra-species, all comparisons; VASI – vertebrate intra-species, independent comparisons.

The hypothesis that the more variable loci diverge more rapidly than the invariable loci cannot be tested by examining the correlation between H_j and I_j.

https://doi.org/10.1017/S0016672300020413 Published online by Cambridge University Press

Correlations between H_j and I_j are expected for non-biological reasons (Riska, 1979). The reason for this can be seen by considering Fig. 1 which plots the minimum H_j value possible for a given I_j; points below the curve are mathematically impossible. In comparing two populations with high overall similarity (high I), H_j values of 0 will have I_j values of 1 (very rarely of 0) whereas I_j values between 0 and 1 must necessarily be accompanied by H_j values greater than 0. Thus sampling error alone will generate a negative correlation between I_j and H_j even when samples are taken from two populations which are identical in allele frequencies. Also, as the distribution of I_j can be very different at high and low H_j, comparisons of representative values of the I_j distribution (such as the mean) for different H_j values must have little biological meaning.

An attempt has been made to overcome this problem by considering how the relationship between I_j and H_j changes with decreasing I. Thus instead of a two dimensional analysis (H_j, I_j), a three dimensional analysis (H_j, I_j, I) is used. To facilitate analysis each dimension has been divided into intervals and the data may be visualised as a three way contingency table. Thus in each of the six data sets pairwise population comparisons have been sorted into classes corresponding to the following intervals for I: (*a*) $\geqslant 0{\cdot}9-1{\cdot}0$, (*b*) $\geqslant 0{\cdot}8-0{\cdot}9$, (*c*) $\geqslant 0{\cdot}7-0{\cdot}8$, (*d*) $\geqslant 0{\cdot}6-0{\cdot}7$, (*e*) $\geqslant 0{\cdot}5-0{\cdot}6$, (*f*) $< 0{\cdot}5$. For each of these classes the frequency distribution of H_j has been determined; first for loci having I_j values in the range $0{\cdot}98-1$, and second for loci having I_j values in the range 0–1. Thus the first distribution of H_j is for loci which, it is assumed, have diverged little or not at all. The second distribution of H_j is for loci over the whole range of I_j. This representation of the data permits the following hypothesis to be formulated. As I decreases and divergence increases through classes (*a*)–(*f*), the number of loci with high I_j values (defined as N_A, the first distribution) will decrease relative to the total number of loci (defined as N_T, the second distribution). Thus the ratio N_A/N_T will decrease as I decreases. If, however, the more variable loci diverge more rapidly than the less variable loci, the ratio will show a greater proportional decrease with decreasing I for the higher than the lower end of the H_j distribution. That is as I decreases, loci with high H_j will be 'lost' from the I_j interval 0·98–1 more rapidly than loci with low H_j. To test this hypothesis, H_j has been subdivided into three intervals: (1) $H_j \leqslant 0{\cdot}02$; with the number of loci falling in this interval for the first and second distributions denoted N_{Aw} and N_{Tw} respectively, (2) $0{\cdot}02 < H_j \leqslant 0{\cdot}1$; with numbers of loci in this interval denoted by N_{Am} and N_{Tm}, (3) $H_j > 0{\cdot}1$; with numbers of loci in this interval denoted by N_{Ah} and N_{Th}. The subscripts w, m and h stand for the weakly variable, moderately variable and highly variable heterozygosity classes respectively. The upper limits of the w and m intervals correspond roughly with the 0·99 and 0·95 frequency criteria for polymorphic loci. The prediction now is that as I decreases through classes (*a*)–(*f*) the ratios N_{Ah}/N_{Th} and N_{Am}/N_{Tm} will show a more rapid relative decrease than the ratio N_{Aw}/N_{Tw}.

Table 1 gives values of N_{Tw}, N_{Tm}, N_{Th}, N_A/N_T, N_{Aw}/N_{Tw}, N_{Am}/N_{Tm} and N_{Ah}/N_{Th} for the six data sets. $N_A = N_{Aw}+N_{Am}+N_{Ah}$ and $N_T = N_{Tw}+N_{Tm}+N_{Th}$. For the intra-species sets VAS and VASI I values lower than 0·6 were not observed. For VESI classes (*e*) and (*f*) have been combined.

It can be seen from Table 1 that, in accordance with the hypothesis, with

https://doi.org/10.1017/S0016672300020413 Published online by Cambridge University Press

Table 1. N_{Tw}, N_{Tm}, N_{Th}, N_A/N_T, N_{Aw}/N_{Tw}, N_{Am}/N_{Tm}, N_{Ah}/N_{Th} *and* $N^*_{Ah}/(N^*_{Aw}+N^*_{Am}+N^*_{Ah})$ *for the six data groups VAS, VASI, VES, VESI, DROS and DROSI*

	I	N_{Tw}	N_{Tm}	N_{Th}	N_A/N_T	N_{Aw}/N_{Tw}	N_{Am}/N_{Tm}	N_{Ah}/N_{Th}	$N^*_{Ah}/(N^*_{Aw}+N^*_{Am}+N^*_{Ah})$
VAS	0·9–1·0	8755	1141	2761	0·888	0·995	0·968	0·427	0·103
	0·8–0·9	1842	205	495	0·776	0·920	0·776	0·242	0·067
	0·7–0·8	523	37	95	0·711	0·826	0·568	0·137	0·044
VASI	0·9–1·0	2480	299	668	0·905	0·997	0·977	0·531	0·113
	0·8–0·9	247	28	60	0·797	0·931	0·714	0·283	0·068
	0·7–0·8	59	10	16	0·694	0·831	0·700	0·188	0·051
VES	0·9–1·0	567	122	222	0·856	0·984	0·967	0·469	0·124
	0·8–0·9	767	220	328	0·750	0·936	0·791	0·287	0·086
	0·7–0·8	1026	266	333	0·681	0·822	0·737	0·201	0·069
	0·6–0·7	891	238	408	0·571	0·755	0·605	0·150	0·058
	0·5–0·6	277	61	104	0·477	0·635	0·410	0·096	0·047
	< 0·5	708	131	177	0·324	0·401	0·313	0·023	0·016
VESI	0·9–1·0	392	81	151	0·851	0·985	1·000	0·424	0·118
	0·8–0·9	329	88	153	0·728	0·942	0·784	0·235	0·075
	0·7–0·8	291	81	107	0·656	0·811	0·765	0·150	0·055
	0·6–0·7	239	56	102	0·582	0·732	0·661	0·186	0·077
	< 0·6	373	69	114	0·347	0·448	0·275	0·061	0·043
DROS	0·9–1·0	70	35	80	0·784	0·986	0·971	0·525	0·370
	0·8–0·9	37	19	45	0·663	0·973	0·947	0·289	0·242
	0·7–0·8	87	39	93	0·571	0·828	0·821	0·226	0·230
	0·6–0·7	85	51	128	0·511	0·777	0·667	0·273	0·290
	0·5–0·6	66	43	90	0·457	0·712	0·488	0·256	0·306
	< 0·5	187	159	515	0·179	0·337	0·208	0·113	0·298
DROSI	0·9–1·0	64	28	70	0·778	0·984	0·964	0·515	0·328
	0·8–0·9	16	12	27	0·636	1·000	0·917	0·296	0·229
	0·7–0·8	42	14	33	0·929	0·762	0·929	0·333	0·275
	0·6–0·7	30	11	27	0·588	0·833	0·545	0·333	0·297
	0·5–0·6	22	20	35	0·442	0·636	0·550	0·257	0·294
	< 0·5	37	33	115	0·238	0·378	0·273	0·183	0·333

decreasing I the ratios N_{Ah}/H_{Th} and N_{Am}/N_{Tm} appear to decrease more rapidly than the ratio N_{Aw}/N_{Tw}. This may be seen more clearly by expressing N_{Ah}/N_{Th} and N_{Am}/N_{Tm} as a proportion of N_{Aw}/N_{Tw}. Thus h/w is defined as $(N_{Ah}/N_{Th})/(N_{Aw}/N_{Tw})$ and m/w as $(N_{Am}/N_{Tm})/(N_{Aw}/N_{Tw})$. If, as predicted, loci with high H_j values are lost rapidly from high I_j regions as I decreases, h/w and m/w should decrease with decreasing I. On the other hand, in the absence of an association between heterozygosity and divergence, h/w and m/w should be constants (see appendix for justification of this null hypothesis).

In Fig. 2 m/w and h/w are plotted against N_A/N_T for the six data sets. N_A/N_T is used as a measure of the overall divergence between the pairs of populations contributing to each of classes $(a)-(f)$. Rank and product moment correlation coefficients between m/w or h/w with N_A/N_T are shown in Table 2. Considering Fig. 2 and Table 2 the following observations may be made. First, there is a clear indication overall of both m/w and h/w decreasing with decreasing I as predicted. Secondly, for VAS, VASI, VES and VESI, trends in m/w and h/w are accompanied by high correlation coefficients. There is no obvious indication of large differences between the results for the independent and non-independent data sets. Thirdly, for the *Drosophila* data, a decrease in m/w though not h/w is apparent, and the correlation coefficients for DROS m/w are significant.

Differences among the I classes have been analysed further by means of a G test (Sokal & Rohlf, 1969). The hypothesis that loci with higher H_j values will be 'lost' more rapidly from the 0·98–1 I_j interval could be tested by examining the relative frequencies of N_{Aw}, N_{Am} and N_{Ah} among the six I classes. Unfortunately such variation might simply reflect variation in the frequencies of N_{Tw}, N_{Tm} and N_{Th}. (A three-way G test would not be appropriate to test the null hypothesis of equal probabilities of loss for the three heterozygosity classes (see appendix).) To permit a G test with this confounding effect removed, adjusted values of N_{Aw}, N_{Am}, and N_{Ah} have been calculated using the formula:

$$N^*_{Ax} = N_{Ax} \times (\Sigma N_{Tx}/\Sigma N_T) \times (N_T/N_{Tx}),$$

where x can be w, h, or m, * denotes the adjusted values and Σ represents summation over the different I classes. Under the null hypothesis, the ratio $N^*_{Ax}/(N^*_{Aw}+N^*_{Am}+N^*_{Ah})$ will be the same for all I classes (see appendix). Contingency tables have been constructed to test for variation among the I classes in the relative frequencies of (i) N_{Tw}, N_{Tm} and N_{Th}, (ii) N_{Aw}, N_{Am}, N_{Ah} and (iii) N^*_{Aw}, N^*_{Am} and N^*_{Ah}. For VAS and VASI these are 3(w, h, and m, H_j intervals) × 6 (I classes) tables; for VES and VESI, 3 × 5 tables; for DROS and DROSI 3 × 6 tables. Total and some component G values are given in Table 3 with probability values and also values for the contingency coefficient, a measure of association in contingency tables (Roscoe, 1969). The following observations can be made. First, significant G values are obtained for analysis (i) for all data sets except VASI. This justifies our concern to correct values of N_{Ax} for variation in N_{Tx}. Secondly, for the vertebrate data G values for analysis (ii) are also all significant and have larger associated contingency coefficients. The adjusted values also give significant G values. Thus variation among the I classes in the H_j distribution is greater for the I_j interval 0·98–1 than for the interval 0–1. Thirdly, for the *Drosophila* data the

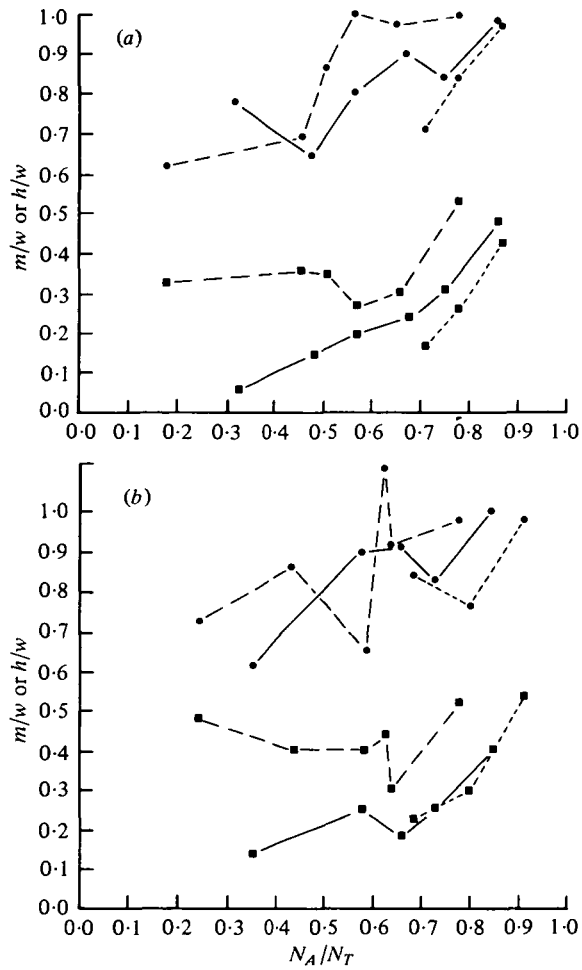

Fig. 2. m/w (●) and h/w (■) plotted against N_A/N_T for the six data groups. (*a*) Data groups VAS (· · · · · · ·), VES (——), and DROS (– – –). (*b*) Data groups VASI (· · · · · · ·), VESI (——) and DROSI (– – –).

Table 2. *Product moment* (r) *and Kendall's rank* (R) *correlation coefficients between* h/w *and* N_A/N_T, *and between* m/w *and* N_A/N_T *for the six data sets*

	$(h/w, N_A/N_T)$		$(m/w, N_A/N_T)$	
	r	R	r	R
VAS	1·00*	1·00	0·99	1·00
VASI	0·97	1·00	0·65	0·33
VES	0·96**	1·00**	0·76	0·73*
VESI	0·83	0·60	0·88*	0·60
DROS	0·43	0·07	0·89*	0·73*
DROSI	−0·06	0·00	0·51	0·26

* $P < 0{\cdot}05$. ** $P < 0{\cdot}01$.

https://doi.org/10.1017/S0016672300020413 Published online by Cambridge University Press

contingency coefficients for analysis (ii) are smaller than for the analysis (i) and analysis (iii) gives non-significant G values. Fourthly, for the vertebrate data partitioning of the G statistic shows that a larger proportion of the total G value can be attributed to the w vs h than the w vs m comparisons. The values of $N^*_{Ah}/(N^*_{Aw}+N^*_{Am}+N^*_{Ah})$ for each I class are given for the six groups in the final

Table 3. *Analysis of G statistics for frequencies of* (i) N_{Tw}, N_{Tm}, N_{Th}; (ii) N_{Aw}, N_{Am}, N_{Ah}; *and* (iii) N^*_{Aw}, N^*_{Am}, *and* N^*_{Ah} *over I classes*

(The dimensions of the contingency tables are 3×3 for VAS and VASI, 3×6 for VES, DROS and DROSI, and 3×5 for VESI. C is the contingency coefficient and equals $G/(G+N)$, where N is total number for the contingency table.)

Source of variation		D.F.	G	P	C
VAS	N_T total	4	44·18	< 0·001	0·053
	N_A total	4	114·32	< 0·001	0·092
	N^*_A total	4	50·48	< 0·001	0·061
	w vs h	2	46·89	< 0·001	—
	w vs m	2	6·39	< 0·05	—
VASI	N_T total	4	1·45	NS	0·019
	N_A total	4	11·77	< 0·05	0·058
	N^*_A total	4	9·30	NS	0·052
	w vs h	2	8·63	< 0·05	—
	w vs m	2	1·20	NS	—
VES	N_T total	10	57·04	< 0·001	0·091
	N_A total	10	84·08	< 0·001	0·139
	N^*_A total	10	59·76	< 0·001	0·117
	w vs h	5	56·65	< 0·001	—
	w vs m	5	5·64	NS	—
VESI	N_T total	8	18·92	< 0·05	0·085
	N_A total	8	30·60	< 0·001	0·134
	N^*_A total	8	20·73	< 0·01	0·110
	w vs h	4	17·74	< 0·01	—
	w vs m	4	4·33	NS	—
DROS	N_T total	10	54·70	< 0·001	0·170
	N_A total	10	18·64	< 0·05	0·159
	N^*_A total	10	10·94	NS	0·124
	w vs h	5	6·17	NS	—
	w vs m	5	4·06	NS	—
DROSI	N_T total	10	35·38	< 0·001	0·230
	N_A total	10	16·78	NS	0·218
	N^*_A total	10	3·61	NS	0·105
	w vs h	5	1·49	NS	—
	w vs m	5	1·87	NS	—

column of Table 1. This is to confirm that the significant G values for the vertebrate data groups are in large part attributable to a consistent decrease in this proportion from I class (a) to I class (f). Taken together with the trends in m/w and h/w these results do therefore provide evidence for the vertebrate data of the existence of the predicted relationship between heterozygosity and rate of divergence for allozyme loci. This relationship does not appear to hold for the *Drosophila* data.

The analysis described above was repeated substituting the wider I_j interval of 0·9–1 for the interval 0·98–1. The results were qualitatively very similar though changes were somewhat less striking. The heterozygosity interval $H_j > 0{\cdot}1$ was also subdivided in an attempt to detect more subtle changes in the H_j distribution at very high heterozygosity. The results were uninteresting because of the small numbers of loci giving high H_j values in the lower I classes.

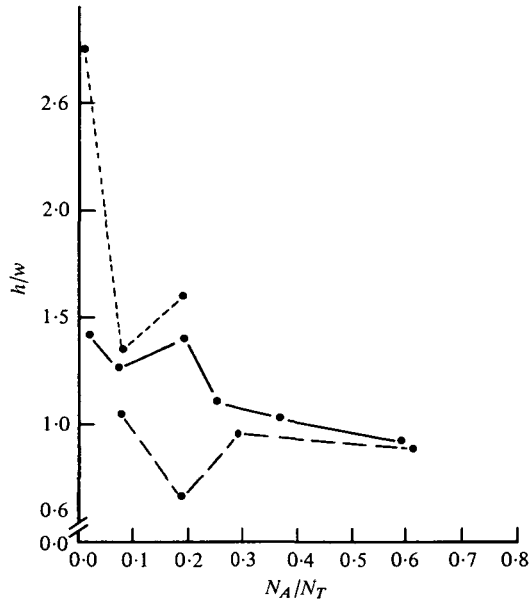

Fig. 3. h/w plotted against N_A/N_T for the 0–0·02 I_j interval for the data groups VAS (· · · · · · ·), VES (——), and DROS (– – –). Note that the ordinate is broken.

An analysis was also made substituting the I_j interval 0–0·02 for the interval 0·98–1. It might be predicted that in the initial stages of divergence the more heterozygous loci might 'arrive' at the low end of the I_j distribution first. At this point the proportion of H_j values in the high heterozygosity interval should be relatively high. Subsequently as divergence continues this high relative frequency should decline. This is a less certain prediction than that for the 0·98–1 I_j interval because it includes the additional assumption that the first loci to begin diverging will be those that are first to diverge completely.

The numbers of loci having H_j values in the I_j interval 0–0·02 are relatively few and only VAS, VES and DROS provided sufficient data for an adequate test. In Fig. 3 h/w ratios for these groups are shown plotted against N_A/N_T. There is some indication of the predicted decrease in h/w with decreasing N_A/N_T for VAS and VES. Values of m/w, which are not plotted, gave no indication of trends. Analyses of G performed in an analogous way to that for the 0·98–1 I_j interval gave non-significant values for all three data sets. There is, therefore, much weaker evidence for trends in h/w in the I_j interval 0–0·02 than in the I_j interval 0·98–1.

The prediction of changes in relative frequencies of the three H_j classes asumes the existence of a correlation in locus heterozygosity between the populations or

species being compared. This assumption can be tested by examining the correlation over loci between the expected heterozygosity values H_{jX} and H_{jY} that are averaged to give H_j for each pairwise population or species comparison. A positive correlation is expected. By dividing the heterozygosity range into two intervals, $\leqslant 0{\cdot}1$ and $> 0{\cdot}1$, 2×2 contingency tables were constructed for each of the six I classes, the four cells being (I) $H_{jx} \leqslant 0{\cdot}1$, $H_{jY} \leqslant 0{\cdot}1$, (II) $H_{jX} \leqslant 0{\cdot}1$, $H_{jY} > 0{\cdot}1$,

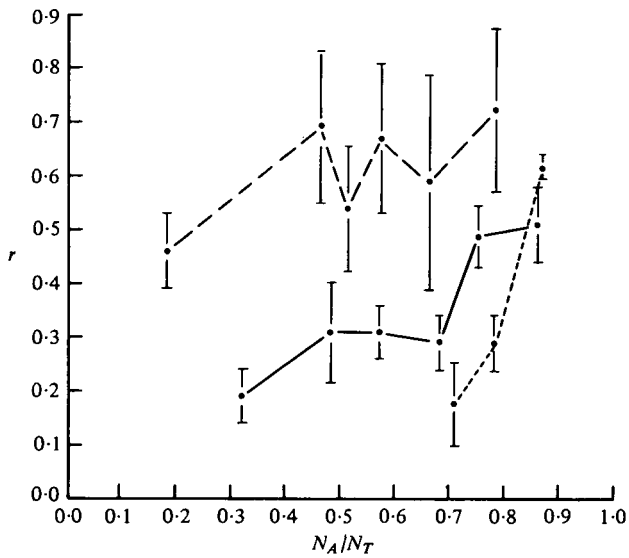

Fig. 4. Four point correlation coefficients (r) for locus heterozygosity plotted against N_A/N_T for VAS (·····), VES (——), and DROS (– – –).

(III) $H_{jX} > 0{\cdot}1$, $H_{jY} \leqslant 0{\cdot}1$, (IV) $H_{jX} > 0{\cdot}1$, $H_{jY} > 0{\cdot}1$. A positive correlation would occur when there are greater than expected numbers in cells (I) and (IV). Four-point correlation coefficients between H_{jX} and H_{jY} with 95% confidence intervals are plotted against N_A/N_T in Fig. 4 for VAS, VES and DROS. The correlation was calculated as the determinant of the 2×2 table divided by the square root of the product of the four column and row totals. Three observations can be made. First, all correlation values are significantly greater than zero. Secondly, for VAS and VES there is a decline in correlation with decreasing N_A/N_T. Thirdly, the *Drosophila* correlation values are consistently greater than the vertebrate values.

Finally an estimate has been made of the relative probabilities for loci in the different heterozygosity intervals of remaining within the I_j interval 0·98–1 in the transition from the high to low I classes. The method of estimation is given in the appendix. The values obtained, which are given in Table 4, show that the values for the h heterozygosity interval are consistently lower than for w. The values for m are intermediate in most cases which also is in line with the hypothesis. For the vertebrate interspecies data probabilities for the transitions $(a)-(c)$ (which allows comparison with the intraspecies data) and (c)–(f) are shown. For (a)–(c) the values for the inter- and intraspecies data are remarkably similar. If the relative

probabilities of remaining for w and h stay the same over two stages of decline in I the ratio of the logs of the probabilities should remain constant. These ratios of logs are shown in the final column of Table 4. For VES and VESI the value of this ratio is greater in the later stages of divergence ((*c*)–(*f*)) than in the initial stages ((*a*)–(*c*)). This suggests that the probability of remaining for h is greater relative to w in the later than in the initial stages of divergence. That is, the difference between w and h is less marked in the later stages.

Table 4. *Probabilities of loci remaining within the I_j interval* 0·98–1 *in the three heterozygosity classes w, m and h for comparisons of high and low I classes*

(The final column shows the ratio of the log of the probabilities for the w and h heterozygosity classes.)

	w	m	h	log w/log h
VAS (*a*)–(*c*)	0·83	0·59	0·35	0·177
VASI (*a*)–(*c*)	0·83	0·72	0·35	0·177
VES (*a*)–(*c*)	0·83	0·76	0·43	0·220
VES (*c*)–(*f*)	0·49	0·42	0·11	0·323
VES (*a*)–(*f*)	0·41	0·32	0·05	0·297
VESI (*a*)–(*c*)	0·82	0·76	0·35	0·189
VESI (*c*)–((*e*)+(*f*))	0·55	0·36	0·41	0·670
VESI (*a*)–((*e*)+(*f*))	0·45	0·27	0·14	0·406
DROS (*a*)–(*f*)	0·34	0·21	0·12	0·508
DROSI (*a*)–(*f*)	0·38	0·28	0·35	0·921

4. DISCUSSION

A relationship between within and among population variability for morphometric characters has been recorded by several investigators (Kluge & Kerfoot, 1973; Sokal, 1976, 1978; Johnson & Mickevich, 1977). Attempts to demonstrate a similar relationship for within and among population allozymic variability have not been so successful. The analytical methods used by Johnson & Mickevich (1977) in their test of this hypothesis have been criticized by Riska (1979) (who also criticized their analysis of morphometric variants) and the analysis of Pierce & Mitton (1979) appears to be subject to several of the same criticisms. In particular, these studies do not give due allowance to the fact that intra- and interpopulation allozymic variation might necessarily be related if these two variables are simply plotted against each other (see text, and Riska, 1979).

We believe the analysis we have undertaken to be substantially free of such bias, and conclude that, for vertebrate species at least, there is a clear relationship between locus variability and locus divergence. This relationship holds for both intra- and interspecies comparisons. These associations were far less clear in the interspecies *Drosophila* data where several tests gave non-significant results. One might have expected relationships between heterozygosity and divergence to be more readily discernible in *Drosophila* than vertebrates, since *Drosophila* are on the whole considerably more heterozygous than vertebrates (Selander, 1976). The solution to this paradox may be found in the relatively small body of *Drosophila* data examined (the total number of I_j comparisons for DROS being 1829 and for

VES 6846). Possibly if further data became available, trends similar to those displayed by vertebrates would appear.

Sokal (1978) and Riska (1979) have pointed out that gene flow between local populations might be a cause of greater within population variation for characters showing greater among population variation. This could be invoked in an interpretation of the vertebrate intraspecies association reported here, but not of the interspecies association where the amount of gene flow can be assumed to be negligible.

Strictly speaking, the observed relationship applies only to the rate at which loci are 'lost' from the I_j interval 0·98–1, the highly heterozygous loci disappearing from this interval more rapidly than the less variable as overall divergence increases, and perhaps also to the rate at which they 'appear' in the I_j interval 0–0·02. A comparative analysis of the more central portion of the I_j range is not informative because of the relatively small numbers of loci in this range. For example, only 13 and 14% respectively of the total vertebrate intra- and interspecies locus comparisons fall in the I_j interval 0·02–0·98. It should also be noted that our conclusion stands only for divergence up to the level of the genus, a lack of data prohibiting comparisons between more distant taxonomic groupings.

Evolution results from both divergent and repetitive change. Thus if the reasonable assumption is made that the total amount of repetitive evolution is not greater for the less variable loci, these results also allow the conclusion that evolutionary rate itself is greater for the more variable loci.

The results may be considered in the light of the neutralist-selectionist controversy. By the neutral hypothesis neutral mutation rate is positively associated with heterozygosity (Kimura & Crow, 1964; Ohta & Kimura, 1973) and with rate of divergence (Nei, 1972). A number of studies have suggested that the assumption of variable neutral mutation rate should be incorporated into neutral theory (for example, Nei, Fuerst & Chakraborty, 1976; Fuerst, Chakraborty & Nei, 1977). Therefore it may be predicted that the more variable loci, those having higher neutral mutation rates, should diverge more rapidly.

Unfortunately, such an observation could also be made consistent with a selectionist viewpoint, by suggesting that in general the more variable loci, by providing more variation on which selection could act, will evolve and diverge more rapidly. The argument is consistent with the observed results even if most polymorphisms are maintained most of the time by balancing or frequency dependent selection when the actual variance in fitness may be low. This is because the change at a locus from a high to a low I_j value may take place in a relatively short time. Such rapid change is evidenced here by the small number of I_j values in the interval 0·02–0·98.

At first sight the evidence for the vertebrate interspecies data showing a higher relative probability of remaining for the highly heterozygous loci in the later than in the initial stages of divergence might be regarded as good evidence against a neutralist interpretation of the results, which would predict the same relative probabilities of remaining (measured as the ratio of log probabilities) at all stages of divergence. Apart from the problem that we have been unable to attach significance levels to these differences this result would also be a predicted effect

of and might therefore be explained by the lower observed heterozygosity correlations at higher levels of divergence.

The extent to which the results can be extrapolated to explain the reported relationships between within and among population variability for morphometric characters (Kluge & Kerfoot, 1973; Sokal, 1976, 1978) depends on the genetic architecture of such characters which is largely unknown. Even if differences in phenotypic variation reflect differences in genetic variation, the genetic variance of quantitative characters may be influenced by the number of segregating loci and the phenotypic effect of substitution of individual alleles as well as by differences in allele frequences and thus heterozygosity at individual loci. The rates of evolution of both structural and regulatory loci for such variation are also unknown and it would certainly be unwise to claim that the allozyme results obtained here can be relevant to the evolution of regulatory loci.

The use of I_j as a measure of locus divergence might be questionable if substitution of amino acids leading to a change in protein charge resulted in a greater probability of loss from the I_j interval 0·98–1·0 for highly than for weakly heterozygous loci. Consider a locus comparison where $I_j = 1$. Substitution of a charged amino acid in one population would convert I_j to zero for a monomorphic locus. Substitution at a variable locus in one population would cause a relative shift in the allele frequency distributions in the two populations but there would still be some overlap so that the new I_j value would be greater than zero. There is in fact some chance that a highly variable locus would not be lost from the I_j interval 0·98–1·0. The higher probability of loss for the highly heterozygous loci observed here would therefore be consistent with a higher rate of charged amino acid substitutions at such loci.

The results are subject to a number of sources of error which would tend to reduce or obscure the relationship between heterozygosity and divergence rate. These include errors in I resulting from the sampling of both alleles and loci (Nei & Roychoudhury, 1974; Nei, 1978), sampling errors in the estimation of allele frequency which would lead to errors in H_j and I_j, and low values for the correlation of locus heterozygosity. A potentially serious source of error might occur if the sizes of the samples used to estimate I_j showed a consistent decrease from class (*a*) to class (*f*). This would result in a greater 'loss' from the I_j interval 0·98–1 for highly heterozygous loci than for, say, monomorphic loci, because the former would show a greater increase in sampling variance of I_j as sample size decreases. To test this possibility average sample sizes have been computed for the populations contributing to each I class for the VESI and VASI data groups. These values are given below:

	I class					
	(*a*)	(*b*)	(*c*)	(*d*)	(*e*)	(*f*)
VASI	43	33	27	—	—	—
VESI	68	58	78	100	54	100

For VESI there is no clear trend, and if anything it is in the opposite direction to that required to produce an artifact. For VASI there is a slight trend showing

a decrease from I class (*a*)–(*c*). By simulating artificial data with different sample sizes we have confirmed that this decrease would be much too slight to explain the results obtained.

It might also be argued that the grouping of data from a large number of surveys (although essential for this type of analysis) has introduced several possible confounding factors. First, there is the uneven distribution of enzymes scored in the different surveys. This could be important because of the evidence of enzyme specific differences in heterozygosity (e.g. Johnson, 1974; Ayala *et al.* 1974; Koehn & Eanes, 1976; Zouros, 1976; Ward, 1977, 1978; Koehn & Eanes, 1978) and enzyme specific variation in contribution to among population variation in I_j (Koehn & Eanes, 1978). Secondly, it is well known that species differ in their levels of overall genetic variability (see, for example, Nevo, 1978), and it may be that species pairs differ in their relationships between heterozygosity and genetic identity. The grouping here of a large number of surveys might have been expected to have randomized or cancelled out such influences, yet limitations on its effectiveness can be gauged by the evidence of highly significant variation among the I classes in the relative frequencies of N_{Tw}, N_{Tm}, and N_{Th} (Table 3).

A further potential source of error or bias occurs where pairs of populations having identical distributions of I_j may give different I values because of differences in H_j or allele frequencies at individual loci. For example, the reduction in I caused by (1) adding a highly heterozygous locus having $I_j = 0$ to a population comparison will be less than (2) adding a monomorphic locus having $I_j = 0$. This is because the numerator of the equation for I will be unaltered while the denominator will show a greater increase with (2) than with (1). The effect of (1) will be to increase N_{Th} but not N_{Ah} and the effect of (2) to increase N_{Tw} but not N_{Aw}. This will result in h/w being higher for (2) than (1). Thus the comparison of (1) and (2) (which have identical I_j distributions) will show a decline in I (from (1) to (2)) accompanied by an increase in h/w. This trend is the opposite to that observed in the analysis of the allozyme data. Similarly the effects of (3) adding a heterozygous locus and (4) adding a monomorphic locus both having $I_j = 1$ can be compared. It is easy to demonstrate algebraically that the increase in I for (3) is always less than for (4). As h/w will be greater for (3) than (4), the decline in I (from (4) to (3)) is once again accompanied by an increase in h/w. The conclusion to be drawn from the above is that a trend in h/w with decreasing I might be observed in allozyme data as a result of differences in the numbers of weakly and highly variable loci among paired comparisons even if such differences have no effect on the distribution of I_j. But the direction of such a trend even if of appreciable magnitude (which is unlikely) would be in the opposite direction to and could not explain the observed trends in the allozyme data.

The prediction of neutral theory of lower similarity between population heterozygosity values for loci with lower I_j values has been corroborated by Chakraborty *et al.* (1978). The evidence here of a decline in correlation of heterozygosity with decreasing N_A/N_T for the vertebrate data also supports this prediction. The relatively high values of the heterozygosity correlations for the *Drosophila* data particularly with low genetic identity were also found and discussed by Chakraborty

et al. (1978). The non-zero correlation coefficients observed for the I classes with low values are consistent with neutral theory if the assumption of variable neutral mutation rate is made.

Finally, while the results do not easily permit discrimination between the selectionist and neutral theory they do fit well with an evolutionary interpretation. A creationist would have to explain why populations or species showing greater overall divergence, as measured by I, tend to be less similar at the more variable loci than at those less variable.

5. APPENDIX: THE NULL HYPOTHESIS

The model considers the transition from a high I class (for example, class (a)) to a low I class (for example, (f)) under the null hypothesis of equal rates or probability of loss of loci from the I_j interval 0·98–1 for the extreme high and extreme low heterozygosity intervals. In class (a) let N_{Aw} and N_{Ah} be represented by p and r respectively and let (N_{Tw}–N_{Aw}) and (N_{Th}–N_{Ah}) be represented by s and u respectively: In the transition from class (a) to (f) the probability of loss of a locus from the I_j interval 0·98 to 1 for the w heterozygosity class is equal to that for the h class. Let this probability be x. We will also consider the probabilities of loci 'returning' to the interval 0·98–1 from the interval 0 to 0·98; let these probabilities for the w and h heterozygosity classes be X and Z respectively. X and Z therefore reflect a component of repetitive evolution. If N_{Tw} and N_{Th} are the total numbers of loci in class (a), let $i \times N_{Tw}$ and $k \times N_{Th}$ be the numbers in class (f). The numbers of loci according to I class, heterozygosity interval, and I_j interval, can be represented by the 3 way table shown below, the numbers in class (f) being given as a function of the numbers in class (a).

	I_j, 0·98–1 (N_{Ax})		I_j, 0–0·98 ($N_{Tx}-N_{Ax}$)	
	Weakly het.	Highly het.	Weakly het.	Highly het.
I class (a)	p	r	s	u
I class (f)	$i(p(1-x)+sX)$	$k(r(1-x)+uZ)$	$i(s(1-X)+px)$	$k(u(1-Z)+rx)$

For class (a),

$$w = p/(p+s), \quad h = r/(r+u), \quad \text{and} \quad h/w = r(p+s)/p(r+u).$$

For class (f),

$$w = i(p(1-x)+sX)/\{i(p(1-x)+sX)+i(s(1-X)+px)\} = (p(1-x)+sX)/(p+s),$$
$$h = k(r(1-x)+uZ)/\{k(r(1-x)+uZ)+k(u(1-Z)+rx)\} = (r(1-x)+uZ)/(r+u),$$

and

$$h/w = (r(1-x)+uZ)(p+s)/(p(1-x)+sX)(r+u).$$

Thus h/w for class (a) equals h/w for class (f) under the null hypothesis when

$$r(p+s)/p(r+u) = (r(1-x)+uZ)(p+s)/(p(1-x)+sX)(r+u),$$
$$r/p = (r(1-x)+uZ)/(p(1-x)+sX),$$
$$pr(1-x)+rsX = pr(1-x)+puZ.$$

https://doi.org/10.1017/S0016672300020413 Published online by Cambridge University Press

If X and Z are negligible compared with x then $LHS = RHS$, and the value of h/w is the same for the two heterozygosity classes under the null hypothesis. Otherwise eliminating $pr(1-x)$,

$$rsX = puZ, \quad \text{or} \quad X/Z = pu/rs.$$

For class (a) for VES, for example, $p = 558$, $u = 118$, $s = 9$, and $r = 104$. Substituting these values gives $X/Z = 70$.

Thus for the value of h/w for class (a) to be greater than for class (f) under the null hypothesis, X must be more than 70 times greater than Z. There is no reason to expect X and Z to differ greatly. If anything we might expect Z to be greater than X. Thus even if X and Z are not negligible compared with x, a decline in h/w as I decreases is extremely unlikely under the null hypothesis.

If it is assumed that $X = 0$, x may be estimated for the highly heterozygous class by equating the observed numbers corresponding to p, s, $ip(1-x)$ and $i(s-px)$. For example, for VES I class (a) $p = 558$ and $s = 9$, and for VES class (f) $ip(1-x) = 284$ and $i(s+px) = 424$. Then,

$$558i(1-x)/i(9+558x) = 284/424,$$

$$x = 234036/395064 = 0{\cdot}59.$$

The probabilities of loss for the other two heterozygosity classes may be calculated in a similar fashion. The probability of a locus remaining within the I_j interval 0·98–1 as I changes is of course simply $1-x$.

The validity of the G tests using the corrected values N^*_{Ax} rests on the assumption that under the null hypothesis the ratio $N^*_{Ax}/(N^*_{Aw}+N^*_{Am}+N^*_{Ah})$ is the same for the different I classes. This assumption can be tested as follows. Replacing the symbol N with n for I class (f), and where Σ represents the summation over classes (a)–(f) the corrected values for I classes (a) and (f) are:

	N^*_{Ax}	
	Weakly het.	Highly het.
I class (a)	$A = N_{Aw} \times (\Sigma N_{Tw}/\Sigma N_T) \times (N_T/N_{Tw})$	$B = N_{Ah} \times (\Sigma N_{Th}/\Sigma N_T) \times (N_T/N_{Th})$
I class (f)	$C = n_{Aw} \times (\Sigma N_{Tw}/\Sigma N_T) \times (n_T/n_{Tw})$	$D = n_{Ah} \times (\Sigma N_{Th}/\Sigma N_T) \times (n_T/n_{Th})$

Then,

$$A/(A+B) = (N_{Aw} \times \Sigma N_{Tw} \times N_{Th})/((N_{Aw} \times \Sigma N_{Tw} \times N_{Th}) + (N_{Ah} \times \Sigma N_{Th} \times N_{Tw})),$$

and

$$C/C+D) = (n_{Aw} \times \Sigma N_{Tw} \times n_{Th})/((N_{Aw} \times \Sigma N_{Tw} \times n_{Th}) + (n_{Ah} \times \Sigma N_{Th} \times n_{Tw})).$$

Substituting with p, u, r, s, x, z, i and k gives

$$A/(A+B) = p(i+1)/(p(i+1)+r(k+1),$$

$$C/(C+D) = (p(1-x)+sX)(i+1)/((p(1-x)+sX)(i+1)+(r(1-x)+uZ)(k+1)).$$

Then it can be shown that $A/(A+B) = C/(C+D)$ under the null hypothesis when

$$pr(1-x)+rsX = pr(1-x)+puZ.$$

https://doi.org/10.1017/S0016672300020413 Published online by Cambridge University Press

This is the same condition discussed above in relation to the ratio h/w. Thus under the null hypothesis a decline in, for example, the ratio $N^*_{Ah}/(N^*_{Aw}+N^*_{Am}+N^*_{Ah})$ from I class (a) to I class (f) is not expected unless X is much greater than Z.

Finally we will consider the validity of computing second order interactions in the above three way table as a method of testing the null hypothesis. Assume first that no second order interactions exist; when this happens it follows that:

$$pk(r(1-x)+uZ)/ri(p(1-x)+sX) = sk(u(1-z)+rx)/ui(s(1-X)+px).$$

With the simplifying assumption $X = Z = 0$, this becomes

$$pr(1-x)/pr(1-x) = s(u+rx)/u(s+px),$$
$$p/s = r/u.$$

Substituting I class (a) values for VES gives $p/s = 558/9$ and $r/u = 104/118$. The condition that $p/s = r/u$ is obviously not met. This means that even if the probabilities of loss for the w and h classes are equal a non zero value for the second order interaction would be obtained and the interaction would in fact only be zero under the restrictive condition $p/s = r/u$. To find an interaction when the probabilities of loss are the same is disconcerting even if the interaction is non-significant. Because of this we chose not to test the hypothesis by computing second order interactions in the three way table.

We wish to thank Dr Bruce Riska and also an anonymous referee for their helpful critical comments of earlier drafts of this paper.

REFERENCES

ADEST, G. A. (1977). Genetic relationships in the genus *Uma* (Iguanidae). *Copeia*, pp. 47–52.

AVISE, J. C. & SMITH, M. H. (1977). Gene frequency comparisons between sunfish (Centrarchidae) populations at various stages of evolutionary divergence. *Systematic Zoology* **26**, 319–335.

AVISE, J. C. & SELANDER, R. K. (1972). Evolutionary genetics of cave-dwelling fishes of the genus *Astyanax*. *Evolution* **26**, 1–19.

AVISE, J. C., SMITH, M. H. & SELANDER, R. K. (1974). Biochemical polymorphism and systematics in the genus *Peromyscus*. VI. The *boylii* species group. *Journal of Mammalogy* **55**, 751–763.

AYALA, F. J. (1975). Genetic differentiation during the speciation process. *Evolutionary Biology* **8**, 1–78.

AYALA, F. J. & GILPIN, M. E. (1974). Gene frequency comparisons between taxa: support for the natural selection of protein polymorphisms. *Proceedings of the National Academy of Science, U.S.A.* **71**, 4847–4849.

AYALA, F. J., TRACEY, M. L., BARR, L. G., MCDONALD, J. F. & PÉREZ-SALAS, S. (1974). Genetic variation in natural populations of five *Drosophila* species and the hypothesis of the selective neutrality of protein polymorphisms. *Genetics* **77**, 343–384.

BARROWCLOUGH, G. F. & CORBIN, K. W. (1978). Genetic variation and differentiation in the *Parulidae*. *Auk* **95**, 691–702.

BEZY, R. L., GORMAN, G. C., KIM, Y. J. & WRIGHT, J. W. (1977). Chromosomal and genetic divergence in the fossorial lizards of the family *Anniellidae*. *Systematic Zoology* **26**, 57–71.

BRITTON, J. & THALER, L. (1978). Evidence for the presence of two sympatric species of mice (genus *Mus* L.) in southern France based on biochemical genetics. *Biochemical Genetics* **16**, 213–225.

BROWNE, R. A. (1977). Genetic variation in island and mainland populations of *Peromyscus leucopus*. *American Midland Naturalist* **97**, 1–9.

BUTH, D. G. & BURR, B. M. (1978). Isozyme variability in the cyprinid genus *Campostoma*. *Copeia*, pp. 298–311.

CASE, S. M. (1978). Biochemical systematics of the genus *Rana* native to Western North America. *Systematic Zoology* **27**, 299–311.

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

COTHRAN, E. G., ZIMMERMAN, E. G. & NADLER, C.F. (1977). Genic differentiation and evolution in the ground squirrel subgenus *Ictidomys* (genus *Spermophilus*). *Journal of Mammalogy* **58**, 610–622.

DUNCAN, R. & HIGHTON, R. (1979). Genetic relationships of the eastern large *Plethodon* of the Ouachita Mountains. *Copeia*, pp. 95–110.

FEDER, J. H., WURST, G. Z. & WAKE, D. B. (1978). Genetic variation in western salamanders of the genus *Plethodon*, and the status of *Plethodon gordoni*. *Herpetologica* **34**, 64–69.

FRYDENBERG, O. & SIMONSEN, V. (1973). Genetics of *Zoarces* populations. V. Amount of protein polymorphism and degree of genic heterozygosity. *Hereditas* **75**, 221–231.

FUERST, P. A., CHAKRABORTY, R. & NEI, M. (1977). Statistical studies on protein polymorphism in natural populations. 1. Distribution of single locus heterozygosity. *Genetics* **86**, 455–483.

FUJIO, Y. (1977). Natural hybridization between *Platichthys stellatus* and *Kareius bicoloratus*. *Japanese Journal of Genetics* 52, 117–124.

GARTSIDE, D. F., ROGERS, J. S. & DESSAUER, H. C. (1977). Speciation with little genic and morphological differentiation in the ribbon snakes *Thamnophis proximus* and *T. sauritus* (Colubridae). *Copeia*, pp. 697–705.

GILL, A. E. (1976). Genetic divergence of insular populations of deer mice. *Biochemical Genetics* **14**, 835–848.

GORMAN, G. C. & KIM, Y. J. (1976). *Anolis* lizards of the eastern Caribbean: a case study in evolution. II. Genetic relationships and genetic variation of the *bimaculatus* group. *Systematic Zoology* **25**, 62–77.

GORMAN, G. C. & KIM, Y. J. (1977). Genotypic evolution in the face of phenotypic conservativeness. *Abudefduf* (Pomacentridae) from the Atlantic and Pacific sides of Panama. *Copeia*, pp. 694–697.

GORMAN, G. C., KIM, Y. J. & RUBINOFF, R. (1976). Genetic relationships of three species of *Bathygobius* from the Atlantic and Pacific sides of Panama. *Copeia*, pp. 361–364.

GORMAN, G. C., KIM, Y. J. & TAYLOR, C. E. (1977). Genetic variation in irradiated and control populations of *Cnemidophorus tigris* (Sauria, Teiidae) from Mercury, Nevada, with a discussion of genetic variability in lizards. *Theoretical and Applied Genetics* **49**, 9–14.

GORMAN, G. C., SOULÉ, M., YANG, S. Y. & NEVO, E. (1975). Evolutionary genetics of insular Adriatic lizards. *Evolution* **29**, 52–71.

GREENBAUM, I. F. & BAKER, R. J. (1976). Evolutionary relationships in *Macrotus* (Chiroptera): biochemical variation and karyology. *Systematic Zoology* **25**, 15–25.

HALL, W. P. & SELANDER, R. K. (1973). Hybridization of karyotypically differentiated populations in the *Sceloporus grammicus* complex (Iguanidae). *Evolution* **27**, 226–242.

HEDGECOCK, D. & AYALA, F. J. (1974). Evolutionary divergence in the genus *Taricha* (Salamandridae). *Copeia*, pp. 738–746.

HIGHTON, R. & WEBSTER, T. P. (1976). Geographic protein variation and divergence in populations of the salamander *Plethodon cinereus*. *Evolution* **30**, 33–45.

JOHNSON, G. B. (1972). Evidence that enzyme polymorphisms are not selectively neutral. *Nature New Biology* **237**, 170–171.

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

JOHNSON, G. L. & PACKARD, R. L. (1974). Electrophoretic analysis of *Peromyscus comanche* Blair, with comments on its systematic status. *Occasional Papers of the Museum. Texas Technical University* **24**, 1–16.

JOHNSON, M. S. (1975). Biochemical systematics of the atherinid genus *Menidia*. *Copeia*, pp. 662–691.

JOHNSON, M. S. & MICKEVICH, M. F. (1977). Variability and evolutionary rates of characters. *Evolution* **31**, 642–648.

JOHNSON, W. E. & SELANDER, R. K. (1971). Protein variation and systematics in kangaroo rats (genus *Dipodomys*). *Systematic Zoology* **20**, 377–405.

Johnson, W. E., Selander, R. K., Smith, M. H. & Kim, Y. J. (1972). Biochemical genetics of sibling species of the cotton rat (*Sigmodon*). *University of Texas Publication*, no. 7213, 297–305.

Kilpatrick, C. W. & Zimmerman, E. G. (1975). Genetic variation and systematics of four species of mice of the *Peromyscus boylii* species group. *Systematic Zoology* **24**, 143–162.

Kilpatrick, C. W. & Zimmerman, E. G. (1976). Biochemical variation and systematics of *Peromyscus pectoralis*. *Journal of Mammalogy* **57**, 506–522.

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

Kimura, M. & Crow, J. F. (1964). The number of alleles that can be maintained in a finite population. *Genetics* **49**, 725–738.

King, M. & Wilson, A. C. (1975). Evolution at two levels in humans and chimpanzees. *Science* **188**, 107–116.

Kirby, G. C. & Halliday, R. B. (1973). Another view of neutral alleles in natural populations. *Nature* **241**, 463–464.

Kirkpatrick, M. & Selander, R. K. (1979). Genetics of speciation in lake whitefishes in the Allegash Basin. *Evolution* **33**, 478–485.

Kluge, A. G. & Kerfoot, W. C. (1973). The predictability and regularity of character divergence. *American Naturalist* **107**, 426–442.

Koehn, R. K. & Eanes, W. F. (1976). The correlation of rare alleles with heterozygosity. In *Population Genetics and Ecology* (ed. S. Karlin and E. Nevo). New York: Academic Press.

Koehn, R. K. & Eanes, W. F. (1978). Molecular structure and protein variation within and among populations. *Evolutionary Biology* **11**, 39–100.

Kornfield, I. L., Ritte, U., Richler, C. & Wahrman, J. (1979). Biochemical and cytological differentiation among cichlid fishes of the Sea of Galilee. *Evolution* **33**, 1–14.

Larson, A. & Highton, R. (1978). Geographic protein variation and divergence in the salamanders of the *Plethodon welleri* group (Amphibia, Plethodontidae). *Systematic Zoology* **27**, 431–448.

Mascarello, J. T. (1978). Chromosomal, biochemical, mensural, penile, and cranial variation in desert woodrats (*Neotoma lepida*). *Journal of Mammalogy* **59**, 477–495.

Matthews, T. C. (1975). Biochemical polymorphism in populations of the Argentine toad, *Bufo arenarum*. *Copeia*, pp. 454–465.

McKinney, C. O., Selander, R. K., Johnson, W. E. & Yang, S. Y. (1972). Genetic variation in the side-blotched lizard (*Uta stansburiana*). *University of Texas Publication*, no. 7213, 307–318.

Merritt, R. B., Rogers, J. F. & Kurz, B. J. (1978). Genic variability in the longnose dace, *Rhinichthys cataractae*. *Evolution* **30**, 116–124.

Nei, M. (1972). Genetic distance between populations. *American Naturalist* **106**, 283–292.

Nei, M. (1975). *Molecular Population Genetics and Evolution*. Amsterdam, Oxford and New York: North-Holland.

Nei, M. (1978). Estimate of average heterozygosity and genetic distance from a small number of individuals. *Genetics* **89**, 583–590.

Nei, M. & Roychoudhury, A. K. (1974). Sampling variances of heterozygosity and genetic distance. *Genetics* **76**, 379–390.

Nei, M., Fuerst, P. A. & Chakraborty, R. (1976). Testing the neutral mutation hypothesis by distribution of single locus heterozygosity. *Nature* **262**, 491–493.

Nevo, E. (1978). Genetic variation in natural populations: Patterns and theory. *Theoretical Population Biology* **13**, 121–177.

Nevo, E. & Shaw, C. R. (1972). Genetic variation in a subterranean mammal, *Spalax ehrenbergi*. *Biochemical Genetics* **7**, 235–241.

Nevo, E., Dessauer, H. C. & Chuang, K. C. (1975). Genetic variation as a test of natural selection. *Proceedings of the National Academy of Science* **72**, 2145–2149.

Nevo, E., Kim, Y. J., Shaw, C. R. & Thaeler, C. S. (1974). Genetic variation, selection and speciation in *Thomomys talpoides* pocket gophers. *Evolution* **28**, 1–23.

Nottebohm, F. & Selander, R. K. (1972). Vocal dialects and gene frequencies in the Chingolo sparrow (*Zonotrichia capensis*). *Condor* **74**, 137–143.

Nozawa, K., Shotake, T., Ohkura, Y. & Tanabe, Y. (1977). Genetic variations within and between species of Asian macaques. *Japanese Journal of Genetics* **52**, 15–30.

Ohta, T. & Kimura, M. (1973). A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. *Genetical Research* **22**, 201–204.

PATTON, J. L., MACARTHUR, H. & YANG, S. Y. (1976). Systematic relationships of the four-toed populations of *Dipodomys heermanni*. *Journal of Mammalogy* **59**, 159–163.

PATTON, J. L., SELANDER, R. K. & SMITH, M. H. (1972). Genic variation in hybridizing populations of gophers (genus *Thomomys*). *Systematic Zoology* **21**, 263–270.

PATTON, J. L., YANG, S. Y. & MYERS, P. (1975). Genetic and morphologic divergence among introduced rat populations (*Rattus rattus*) of the Galapagos archipelago, Ecuador. *Systematic Zoology* **24**, 296–310.

PIERCE, B. A. & MITTON, J. B. (1979). Relationship of genetic variation within and among populations: An extension of the Kluge–Kerfoot phenomenon. *Systematic Zoology* **28**, 63–70.

POWELL, J. R. (1975). Protein variation in natural populations of animals. *Evolutionary Biology* **8**, 79–119.

PRAKASH, S. (1977). Genetic divergence in closely related sibling species *Drosophila pseudoobscura*, *Drosophila persimilis* and *Drosophila miranda*. *Evolution* **33**, 14–23.

RISKA, B. (1979). Character variability and evolutionary rate in *Menidia*. *Evolution* **33**, 1001–1004.

ROSCOE, J. T. (1969). *Fundamental Research Statistics*. Holt, Rinehart and Winston.

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

SELANDER, R. K. & KAUFMAN, D. W. (1973). Genic variability and strategies of adaptation in animals. *Proceedings of the National Academy of Science, U.S.A.* **70**, 1875–1877.

SELANDER, R. K., HUNT, W. G. & YANG, S. Y. (1969). Protein polymorphism and genic heterozygosity in two European subspecies of the house mouse. *Evolution* **23**, 379–390.

SELANDER, R. K., SMITH, M. H., YANG, S. Y., JOHNSON, W. E. & GENTRY, J. B. (1971). Biochemical polymorphism and systematics in the genus *Peromyscus*: I. Variation in the old-field mouse (*Peromyscus poliolotus*). *University of Texas Publication*, no. 7103, 49–90.

SENE, F. M. & CARSON, H. L. (1977). Genetic variation in Hawaiian *Drosophila*. IV. Allozymic similarity between *D. silvestris* and *D. heteroneura* from the island of Hawaii. *Genetics* **86**, 187–198.

SIEBENALLER, J. F. (1978). Genetic variability in deep-sea fishes of the genus *Sebastolobus* (Scorpaenidae). In *Marine Organisms: Genetics, Ecology and Evolution* (ed. B. Battaglia and J. A. Beardmore). New York and London: Plenum.

SMITH, J. K. & ZIMMERMAN, E. G. (1976). Biochemical genetics and evolution of North American blackbirds, family Icteridae. *Comparative Biochemistry and Physiology*. B **53**, 319–324.

SMITH, M. H., SELANDER, R. K. & JOHNSON, W. E. (1973). Biochemical polymorphism and systematics in the genus Peromyscus. III. Variation in the Florida deer mouse (*Peromyscus floridanus*), a Pleistocene relict. *Journal of Mammalogy* **54**, 1–13.

SOKAL, R. R. (1976). The Kluge–Kerfoot phenomenon re-examined. *American Naturalist* **110**, 1077–1091.

SOKAL, R. R. (1978). Population differentiation: Something new or more of the same? In *Ecological Genetics: The Interface* (ed. P. F. Brussard). New York, Heidelberg and Berlin: Springer-Verlag.

SOKAL, R. R. & ROHLF, F. J. (1969). *Biometry*. San Francisco: Freeman.

STRANEY, D. O., O'FARRELL, M. J. & SMITH, M. H. (1976). Biochemical genetics of *Myotis californicus* and *Pipistrellus hesperus* from southern Nevada. *Mammalia* **40**, 344–347.

TAYLOR, C. E. & GORMAN, G. C. (1975). Population genetics of a 'colonising' lizard: natural selection for allozyme morphs in *Anolis grahami*. *Heredity* **35**, 241–247.

TILLEY, S. G., MERRITT, R. B., WU, B. & HIGHTON, R. (1978). Genetic differentiation in salamanders of the *Desmognathus ochrophaeus* complex (Plethodontidae). *Evolution* **32**, 93–115.

WAKE, D. B., MAXSON, L. R. & WURST, G. Z. (1978). Genetic differentiation, albumin evolution and their biogeographic implications in plethodontid salamanders of California and Southern Europe. *Evolution* **32**, 529–539.

WARD, R. D. (1977). Relationship between enzyme heterozygosity and quaternary structure. *Biochemical Genetics* **15**, 123–135.

WARD, R. D. (1978). Subunit size of enzymes and genetic heterozygosity in vertebrates. *Biochemical Genetics* **16**, 799–810.

WARD, R. D. & BEARDMORE, J. A. (1977). Protein variation in the plaice, *Pleuronectes platessa* L. *Genetical Research* **30**, 45–62.

Ward, R. D. & Galleguillos, R. A. (1978). Protein variation in the plaice, dab and flounder, and their genetic relationships. In *Marine Organisms: Genetics, Ecology and Evolution* (ed. B. Battaglia and J. A. Beardmore). New York and London: Plenum.

Webster, T. P. (1975). An electrophoretic comparison of the Hispaniolan lizards *Anolis cybotes* and *A. marcanoi*. *Breviora* **431**, 1–8.

Webster, T. P., Selander. R. K. & Yang, S. Y. (1972). Genetic variability and similarity in the *Anolis* lizards of Bimini. *Evolution* **26**, 523–535.

Wilson, A. C., Carlson, S. S. & White, T. J. (1977). Biochemical evolution. *Annual Reviews of Biochemistry* **46**, 573–639.

Yamazaki, T. & Maruyama, T. (1975). Similarity of gene frequencies in different *Drosophila* species: evidence for the neutrality hypothesis of protein polymorphisms. *National Institute of Genetics (Japan). Annual Report* **25**, 62–65.

Yang, S. Y., Soulé, M. & Gorman, G. C. (1974). *Anolis* lizards of the eastern Carribbean: a case study in evolution. 1. Genetic relationships, phylogeny, and colonization sequence of the *roquet* group. *Systematic Zoology* **23**, 387–399.

Yang, S. Y., Wheeler, L. L. & Bock, I. R. (1972). Isozyme variations and phylogenetic relationships in the *Drosophila bipectinata* species complex. *University of Texas Publication*, no. 7213, 213–227.

Zimmerman, E. G. & Nejtek, M. E. (1977). Genetics and speciation of three semispecies of *Neotoma*. *Journal of Mammalogy* **58**, 391–402.

Zimmerman, E. G., Hart, B. J. & Kilpatrick, C. W. (1975). Biochemical genetics of the *truei* and *boylei* groups of the genus *Peromyscus* (Rodentia). *Comparative Biochemistry and Physiology*, B **52**, 541–545.

Zouros, E. (1973). Genic differentiation associated with early stages of speciation in the *mulleri* subgroup of *Drosophila*. *Evolution* **27**, 601–621.

Zouros, E. (1976). Hybrid molecules and the superiority of the heterozygote. *Nature* **262**, 227–229.