Genet. Res. Camb. (1985), **45**, *pp.* 315–340
With 6 *text-figures*
Printed in Great Britain

Observed relationships between protein heterozygosity and protein genetic distance and comparisons with neutral expectations

By R. D. WARD

Department of Human Sciences, Environmental Biology Unit, Loughborough University, Loughborough, Leicestershire, LE11 3TU, U.K.

AND D. O. F. SKIBINSKI

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

(*Received 16 January 1984 and in revised form 21 January 1985*)

SUMMARY

Relationships between protein genetic distance (D) and protein heterozygosity (H) were studied using allele frequency data for 42 proteins derived from multilocus electrophoretic surveys of genetic variation in over 200 invertebrate and over 300 vertebrate species. $\overline{D}$ and $\overline{H}$ values for the different proteins (mostly enzymes) were calculated, and large and significant correlations between $\overline{D}$ and $\overline{H}$ were found in comparisons of both intraspecific and interspecific populations. Empirical relationships between $\overline{D}$ and $\overline{H}$ were compared with neutral expectations under the stepwise model of neutral mutation with the assumption that populations are in equilibrium with respect to the effects of mutation and genetic drift.

At low divergence levels, a linear relationship of $\overline{D}$ on $\overline{H}$ was observed, but at high levels of divergence D tended towards an asymptote at high $\overline{H}$. The results at high divergence cannot be explained using the approximate relationship $D = 2ut$ (where u = mutation rate, t = time). However, computer simulations of neutral models showed that changes of this nature in the relationship between $\overline{D}$ and $\overline{H}$ were to be expected as divergence increases, the equation $D = 2ut$ being a poor approximation at high D. We therefore conclude that the observed relationships between $\overline{D}$ and $\overline{H}$ are, in fact, compatible with equilibrium neutral theory.

1. INTRODUCTION

The presence of substantial amounts of allozymic variability revealed by gel electrophoresis in natural populations of most outcrossing animal and plant species has prompted much speculation concerning the nature of this variation. It is still uncertain whether the bulk of it is maintained by balanced selective forces or whether it arises from mutation and genetic drift acting upon selectively neutral alleles (Lewontin, 1974). Proponents of the latter school have constructed a theoretical framework interrelating such variables as population size, mutation rates and heterozygosity, and in general neutral theory appears adequate to

https://doi.org/10.1017/S0016672300022291 Published online by Cambridge University Press

account for observed levels of variation (e.g. Fuerst, Chakraborty & Nei, 1977; Chakraborty, Fuerst & Nei, 1980; Nei, 1983). Interpopulational parameters such as genetic distance have also been tested for goodness of fit to neutral theory, and the fit found to be acceptable (Chakraborty, Fuerst & Nei, 1978). Some criticisms have been raised of these tests (Phillips & Mayo, 1981), although Nei and his collaborators believe that such comments are largely based on misunderstandings of the papers concerned (Nei, Chakraborty & Fuerst, 1981).

We have devised a further test of neutral theory which examines whether protein genetic distance and protein heterozygosity are interrelated in the manner predicted by neutral theory. An earlier examination of pooled data from many vertebrate interspecific comparisons (Skibinski & Ward, 1982) revealed that protein evolutionary rate estimates obtained from the two techniques of protein electrophoresis and aminoacid sequencing are highly correlated. Furthermore, a high positive correlation was observed between the average genetic distances of different proteins and their average heterozygosities. This is expected under neutral theory if it is assumed that proteins vary in neutral mutation rate: proteins with high mutation rates would accumulate more heterozygosity and would diverge more rapidly than those with lower mutation rates. The observed relationship between protein genetic distance and protein heterozygosity was approximately linear, as expected under both the infinite allele and stepwise mutation neutral models (Kimura & Crow, 1964; Ohta & Kimura, 1973; Nei, 1972), but a linear regression of distance on heterozygosity intercepted the distance axis at a point significantly greater than zero. This non-zero intercept we believed to be inconsistent with neutral theory. Chakraborty & Hedrick (1983) disputed this conclusion on the grounds that sampling errors could cause heterozygosity to be underestimated in cases where the true value of heterozygosity is low, and that furthermore bottleneck effects could distort the relationship between distance and heterozygosity. We questioned whether these factors could in fact account for the observed relationship (Skibinski & Ward, 1983).

It is the purpose of the present paper to expand upon these earlier findings and discussions, firstly by considering not only vertebrate species but also invertebrates, secondly by considering both interspecific and intraspecific comparisons and by describing the changes in relationships between distance and heterozygosity as species diverge, and thirdly by considering in more detail the predictions and expectations of neutral models of evolutionary change. We also consider whether the results support the proposal of Kluge & Kerfoot (1973) that characters that are more variable within populations are also those that diverge most among populations.

2. SOURCES OF DATA

Allele frequency data from a large number of surveys of natural populations were used in the analysis, although only studies screening a minimum of 15 individuals (30 genomes) per species or population for a minimum of 15 loci were used. Surveys of laboratory populations, or of parthenogenetic or haplodiploid species, were not included. Sources of data already given in Skibinski & Ward (1981) are not

https://doi.org/10.1017/S0016672300022291 Published online by Cambridge University Press

repeated here. Each of the references is followed by a figure (or figures) in parentheses. This refers to the number of species or populations that can be compared pairwise, except for the figures following the Skibinski & Ward (1981) reference which gives the total number of species or populations listed in that paper.

Drosophila (27 species plus 4 subspecies). Skibinski & Ward, 1981 (25+4); Gonzalez *et al.* 1982 (2).

Invertebrate interspecies (130 species plus 1 sub-species).

Arthropoda (93 species): Krepp & Smith, 1974 (2); Tracy *et al.* 1975, Hedgecock *et al*, 1977 (2); Brittnacher, Sims & Ayala, 1978 (3); Sluss *et al.* 1978 (2); Harrison, 1979 (4); Hedgecock, 1979 and per. comm. (3, 2, 2, 2, 2, 2, 2, 2, 2); Nemeth & Tracey, 1979 (3, 2); Mulley & Latter, 1980 (6, 6); Brown, 1981 (5); Daly, Wilkinson & Shaw, 1981 (4); Fevolden & Ayala, 1981 (2); Guttman, Wood & Karlin, 1981 (3); Sbordoni *et al.* 1981 (2); Stock & Castrovillo, 1981 (5); Anderson, per comm. (4); Geiger, per. comm. (12). *Mollusca* (35 species plus 1 subspecies): Gould & Woodruff, 1978 (2); Buroker, Hershberger & Chew, 1979*a* (3), 1979*b* (5+1, 3); Chambers, 1980 (6); Dillon & Davis, 1980 (3); Hornbach *et al.* 1980 (4); Ward & Warwick, 1980 (2); Nevo *et al.* 1981 (2); Nevo, per. comm. (8). *Annelida* (2 species): Nicklas & Hoffmann, 1979 (2).

Invertebrate intraspecies (229 *populations from* 52 *species*).

Arthropoda (116 populations, 26 species): Selander *et al.* 1970 (2); Saura, Halkka & Lokki, 1973 (4); Tracey *et al.* 1975 (5); Cockley, Gooch & Weston, 1977 (8); Hedgecock *et al.* 1977 (4); Harrison, 1979 (6, 6, 2); Nemeth & Tracey, 1979 (6, 3); Cianchi, Maini & Bullini, 1980 (2); Matthews & Craig, 1980 (10); Moran, Wilkinson & Shaw, 1980 (5); Smith, McKoy & Machin, 1980 (4); Steiner, Kitzmiller & Osterbur, 1980 (2); Beck & Price, 1981 (9); Guttman *et al.* 1981 (4, 4, 3); Sbordoni *et al.* 1981 (2, 2); Zera, 1981 (3, 3); Anderson, per. comm. (3); Hedgecock, per. comm. (2); Hilburn, per. comm. (14). *Mollusca* (110 populations, 25 species): Gould & Woodruff, 1978 (22, 4); Morgan *et al.* 1978 (2); Buroker *et al.* 1979*a* (3); Chambers, 1980 (3, 8, 3); Dillon & Davis, 1980 (6, 4, 2); Hornbach, McLeod and Guttman, 1980 (2); Hornbach *et al.* 1980 (13); Ward & Warwick, 1980 (6, 3); Daly *et al.* 1981 (2, 4, 3); Ritte & Pashtan, 1982 (3, 2); Nevo, per. comm. (2, 3, 3, 2, 2). *Phoronida* (3 populations, 1 species): Ayala *et al.* 1984 (3).

Vertebrate interspecies (256 species plus 15 subspecies). *Fish* (76 species plus 7 subspecies): Skibinski & Ward, 1981 (28+2); Buth, 1979 (8+1); Smith, Wood & Benson, 1979 (2); Buth, 1980 (3); Buth, Burr & Schenck, 1980 (3); Loudenslager & Gall, 1980 (1+4); Ferguson, Noakes & Danzmann, 1981 (2); Ferris, Buth & Whitt, 1982 (2); McKaye *et al.* 1982 (3); Child, per. comm. (14); Galleguillos, per. comm. (3); Wallis, per comm. (7). *Amphibia* (37 species plus 2 subspecies): Skibinski & Ward, 1981 (24+2); Feder, 1979 (2); Ashton, Braswell & Guttman, 1980 (3); Sattler, 1980 and per. comm. (3); Tilley & Schwerdtfeger, 1981 (2); Yanev & Wake, 1981 (2). *Reptiles* (59 species plus 4 subspecies): Skibinski & Ward, 1981 (40+1); Kim *et al.* 1976 (2); Lawson and Dessauer, 1979 (3+3); Blanc & Cariou, 1980 (4); Gorman *et al.* (4); Mayer, 1981 (2); Mayer & Tiedemann, 1981 (2); Seidel, Reynolds & Lucchino, 1981 (2). *Birds* (18 species): Skibinski & Ward, 1981 (11);

https://doi.org/10.1017/S0016672300022291 Published online by Cambridge University Press

Yang & Patton, 1981 (7). *Mammals* (66 species plus 2 subspecies): Skibinski & Ward, 1981 (50+2); Baker *et al.* 1981 (2); Hafner, Petersen & Yates, 1981 (2); Patton, Sherwood & Yang, 1981 & Patton, per. comm. (7); Simonsen, 1982 (5).

Vertebrate intraspecies (452 populations, from 74 species). *Fish* (140 populations from 24 species): Skibinski & Ward, 1981 (55 populations, 9 species); Sage & Selander, 1975 (3); Smith, Francis & Paul, 1978 (2); Grant & Utter, 1980 (14); McLeod, Wynes & Guttman (2); Winans, 1980 (14); Andersson *et al.* 1981 (3); Davis, Demartini & McGee, 1981 (2); Fairbairn, 1981*a* (5); Fairbairn, 1981*b* (6); Kornfield *et al.* 1981 (5); Stoneking, Wagner & Hildebrand, 1981 (8); Vuorinen, Himberg & Lankinen, 1981 (12); Carlson *et al.* 1982 (3); Galleguillos & Ward, 1982 (4); Wallis, per. comm. (2). *Amphibia* (81 populations, from 9 species): Skibinski & Ward, 1981 (51 populations, 5 species); Case, 1978 (3); Nevo & Yang, 1979 (7); Larson, 1980 (11); Tilley & Schwerdtfeger, 1981 (9). *Reptiles* (81 populations, from 13 species): Skibinski & Ward, 1981 (43 populations, 8 species); Adams, Smith & Baccus, 1980 (3); Gorman *et al.* 1980 (14); Nevo, 1981 (8); Hertz & Zouros, 1981 (7, 6). *Birds* (8 populations, from 2 species): Skibinski & Ward, 1981 (4 populations, from one species); Barrowclough, 1980 (4). *Mammals* (142 populations, from 26 species): Skibinski & Ward, 1981 (52 populations, from 11 species); Glover *et al.* 1977 (5); Schmitt, 1978 (13); Kovacic & Guttman, 1979 (3); Ramsey *et al.* 1979 (2); Smith, 1979 (6); Dew & Kennedy, 1980 (7); Rice, Gardner & O'Brien, 1980 (2); Ryman *et al.* 1980 (14); Smith & Patton, 1980 (15); Hafner *et al.* 1981 (2); Kawamoto & Ischak, 1981 (3); Patton, per. comm. (6, 4, 6, 2).

3. ANALYSIS

The form of analysis described in Skibinski & Ward (1982) is that used here with the minor modification that in each of the categories to be described later, not only the mean genetic distance ($\overline{D}_j$) but also the mean heterozygosity ($\overline{H}_j$) of each protein was adjusted to allow for the possibility that the species or population pairs contributing data for that protein had, for all proteins, atypically high or low distance or heterozygosity values. Thus, in the present paper, $\overline{D}_j$ is equivalent to $\overline{D}_{j^*}$ of Skibinski & Ward (1982), and $\overline{H}_j$ is an equivalent modification of the earlier, unadjusted, $\overline{H}_j$ derivation.

The relationship between $\overline{D}_j$ and $\overline{H}_j$ was studied (as in Skibinski & Ward, 1982) by linear regression and correlation analyses. This is because the neutral equations used in deriving expected relationships (see Results section) predict a straight line relationship between $\overline{D}$ and $\overline{H}$ over the heterozygosity range studied, with zero intercept, and thus such analysis provides a test of neutral theory. As will be seen, the linear approximation is in fact poor as divergence increases, but has been retained throughout as it indicates how with increasing distance the intercept increases as predicted from our simulations of the neutral model. Our conclusions are unaffected if (i) sampling variation in $\overline{H}_j$ (estimated from the standard error of heterozygosity for each protein) is taken into account or (ii) principal axes are used instead of regression analysis.

We have used throughout this paper Nei's standard measures of identity and distance (Nei, 1972), rather than his 'unbiased' estimates which allow for the

https://doi.org/10.1017/S0016672300022291 Published online by Cambridge University Press

effects of sample size (Nei, 1978*a*). Use of the unbiased estimates would have meant excluding an appreciable body of potentially useful data, since authors are not always precise about the exact numbers of individuals used per locus in their electrophoretic surveys. The bias inherent in using the standard estimates is expected to be small, especially as most sample sizes are appreciably greater than our permitted minimum of 15 individuals, and thus would not affect our conclusions in any significant way.

Vertebrate, invertebrate (excluding *Drosophila*) and *Drosophila* data sets were analysed separately, and interspecific comparisons were separated from intraspecific comparisons.

(i) *Interspecific comparisons*

The two data sets consisting of the vertebrate and invertebrate interspecies information were each divided into four files of species pairs showing little, moderate (two levels) or high levels of divergence, corresponding to genetic identity ranges of 0·9–1·0, 0·7–0·9, 0·5–0·7 and 0–0·5 respectively. In addition, the two total data sets [I range 0–1·0) were analysed. The interspecies *Drosophila* file was too small to warrant subdivision, and was analysed in a single file with I range 0–1·0. Thus a total of 11 files were constructed. Each file was analysed twice, firstly using all possible pairwise comparisons of species within each survey (as in Skibinski & Ward, 1982), and secondly using the independent comparisons only.

In the former mode of analysis, only proteins screened from a minimum of four surveys and from a minimum of 30 pairwise species comparisons per protein per file were considered. Here, if a single survey screened k related species within one of the I ranges detailed above, this would give $k(k-1)/2$ species pairs each of which would provide a single estimate of heterozygosity and genetic identity for each protein assayed, estimates which would contribute to the final values of $\overline{H}_j$ and $\overline{D}_j$.

The second analytical method, using independent comparisons, was more conservative in considering comparisons to be included in the analysis. In a survey with k species, species 1 was compared with species 2, species 2 with species 3, and so on to provide $k-1$ estimates of heterozygosity and identity for each protein in that survey. These values would then contribute to the final estimates of $\overline{H}_j$ and $\overline{D}_j$. Note that we are using the term independent in a mathematical rather than biological sense (Skibinski & Ward, 1981), since the estimates are not independent of the constraints that the true phylogeny of a species assemblage places on particular interspecific distance values, given others. Only proteins screened from a minimum of four surveys and from a minimum of 15 independent pairwise species comparisons per file were considered.

(ii) *Intraspecific comparisons*

There were only two data sets under consideration, vertebrates and invertebrates. The invertebrate set excluded *Drosophila*, but there was no separate *Drosophila* file. Each data set was divided into two files of population pairs showing low or high levels of genetic divergence, corresponding to I ranges of 0·95–1·0 and 0·7–0·9 respectively. Again, the two total data sets with I ranges of 0–1·0 were also

https://doi.org/10.1017/S0016672300022291 Published online by Cambridge University Press

analysed, giving a total of six intraspecies files. Only proteins screened from a minimum of four surveys per file, equivalent to a minimum of four species, were considered. Populations of each species were compared pairwise for each protein and these estimates of heterozygosity and identity averaged so that each species contributed a single heterozygosity and a single identity value per protein to the final estimate of $\overline{H}_j$ and $\overline{D}_j$ for that protein. This procedure was followed to avoid swamping the final $\overline{H}_j$ and $\overline{D}_j$ estimates with numerous heterozygosity and identity values derived from those surveys screening large number of populations. In such surveys, the interpopulation correlations of heterozygosity and the interpopulation correlations of identity are expected to be high because of gene flow between populations, and could thus bias the estimates of $\overline{H}_j$ and $\overline{D}_j$. In our analysis, species screened for two populations were accorded equal weight with species screened for twenty.

4. RESULTS

(i) *Overall relationships between $\overline{H}_j$ and $\overline{D}_j$*

Table 1 gives the results of a linear regression and correlation analysis of protein genetic distance and protein heterozygosity. The correlation coefficients are all highly significant, and coefficients of determination range from 0·25–0·50. Thus, in general, proteins with high heterozygosity have high genetic distance, and a substantial proportion of the observed variance in protein distance can be explained by variation in protein heterozygosity. Correlation coefficients for vertebrates are somewhat greater than those for invertebrates: this may reflect the larger database of vertebrates, but in fact these differences are not statistically significant.

Regression coefficients are positive and all are significantly greater than zero. With respect to the interspecies comparisons, regressing $\overline{D}_j$ on $\overline{H}_j$ gives a positive intercept on the $\overline{D}_j$ axis when $\overline{H}_j = 0$, and this intercept is significantly different from zero. Intercepts produced by the intraspecies comparisons, although positive, are very close to, and not significantly different from, zero. The final column of Table 1 expresses the intercept as a proportion of the overall mean genetic distance (D_j) of the proteins of that file, and it can be seen that this proportion increases from around 0·20 for intraspecific analyses to around 0·60–0·70 for interspecific data sets. Thus the intercepts are greater in both absolute and relative terms in the interspecies comparisons.

Overall mean heterozygosity per protein (H_j) increases from vertebrates through invertebrates to *Drosophila*, a trend which has been noted in many other surveys (e.g. Selander, 1976; Nevo, 1978).

Fig. 1 plots mean protein distances against mean protein heterozygosities, using the independent comparisons only for the interspecies plots. The sizes of the graphs shown in Fig. 1 are the same even though the dimensions marked on the axes are different (the maximum values marked on the axes are approximately equal to the observed maximum values). The substantial relative increases in the intercepts of the interspecies comparisons are therefore readily apparent.

The legend of Fig. 1 gives details on sample sizes of the different proteins. Some

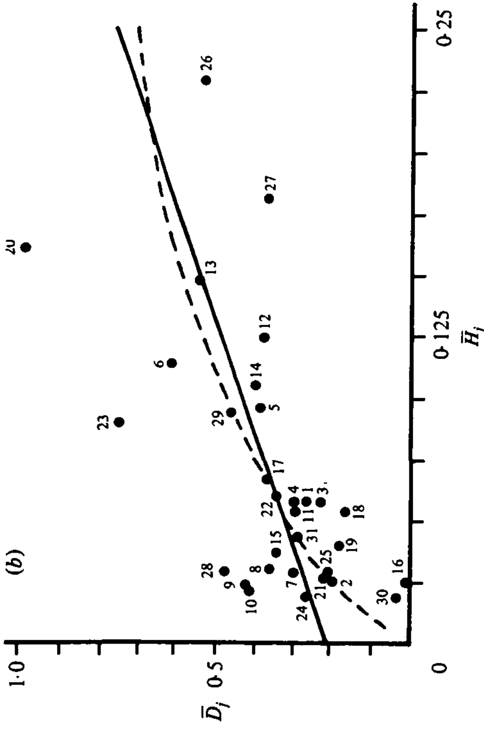
https://doi.org/10.1017/S0016672300022291 Published online by Cambridge University Press

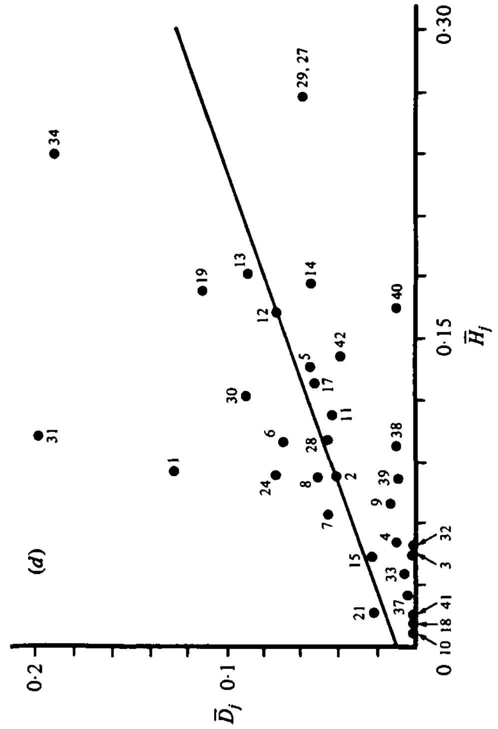
Table 1. *Regression analysis of protein distance* ($\overline{\overline{D}}_j$) *on protein heterozygosity* ($\overline{\overline{H}}_j$), *I ranges 0–1·0.*

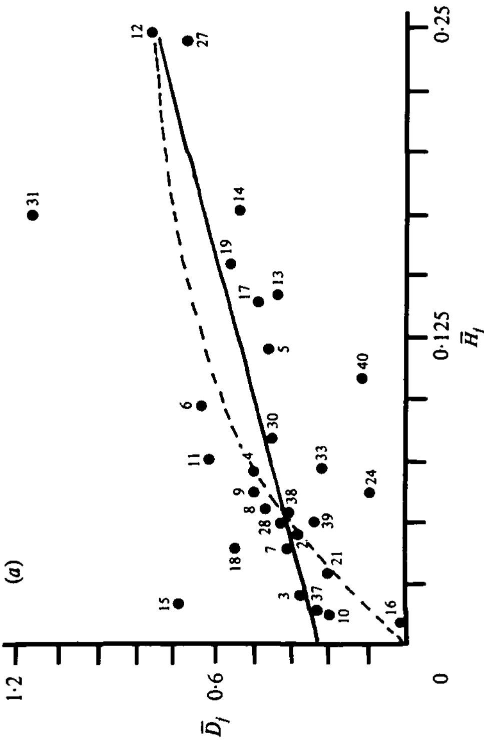
	Number of Proteins	Number of Comparisons	Intercept ± S.E.	Regression coefficient ± S.E.	Correlation coefficient (r)	P	r^2	$\overline{\overline{H}}_j$	$\overline{\overline{D}}_j$	Intercept/ $\overline{\overline{D}}_j$
				Drosophila interspecies						
(A)	20	563	0·310 ± 0·066	1·104 ± 0·352	0·594	0·003	0·353	0·158	0·484	0·641
(B)	22	1682	0·573 ± 0·090	1·171 ± 0·460	0·495	0·010	0·245	0·163	0·763	0·750
				Invertebrate interspecies						
(A)	29	1924	0·270 ± 0·057	2·055 ± 0·538	0·592	< 0·001	0·350	0·085	0·444	0·609
(B)	29	6046	0·511 ± 0·089	2·725 ± 0·817	0·540	< 0·001	0·292	0·076	0·719	0·710
				Vertebrate interspecies						
(A)	31	4397	0·215 ± 0·046	2·214 ± 0·531	0·612	< 0·001	0·375	0·069	0·368	0·584
(B)	31	12245	0·242 ± 0·053	3·164 ± 0·615	0·692	< 0·001	0·479	0·066	0·452	0·537
				Invertebrate intraspecies						
	34	734	0·010 ± 0·013	0·387 ± 0·108	0·543	< 0·001	0·295	0·104	0·051	0·204
				Vertebrate intraspecies						
	33	1124	0·008 ± 0·008	0·527 ± 0·092	0·717	< 0·001	0·514	0·066	0·043	0·189

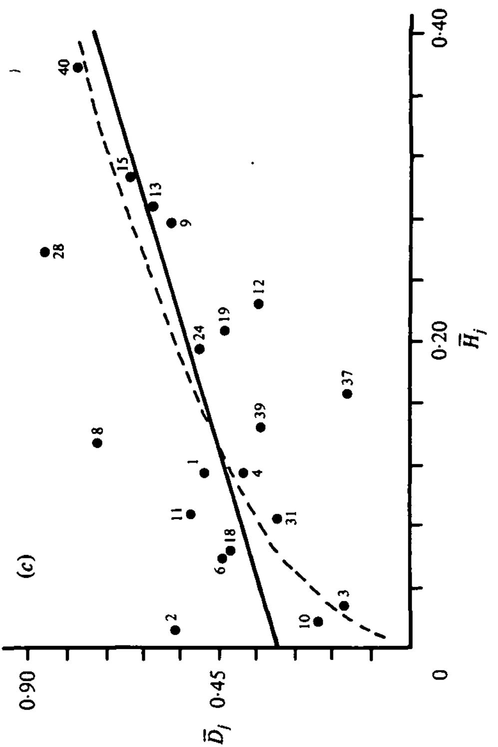
(A) and (B) use independent and non-independent comparisons respectively. P is the probability that $r = 0$ (number of degrees of freedom = number of proteins −2). r^2 is the coefficient of determination. $\overline{\overline{H}}_j = \Sigma\overline{H}_j/n$ and $\overline{\overline{D}}_j = \Sigma\overline{D}_j/n$ where n is the number of proteins essayed.

https://doi.org/10.1017/S0016672300022291 Published online by Cambridge University Press









https://doi.org/10.1017/S0016672300022291 Published online by Cambridge University Press

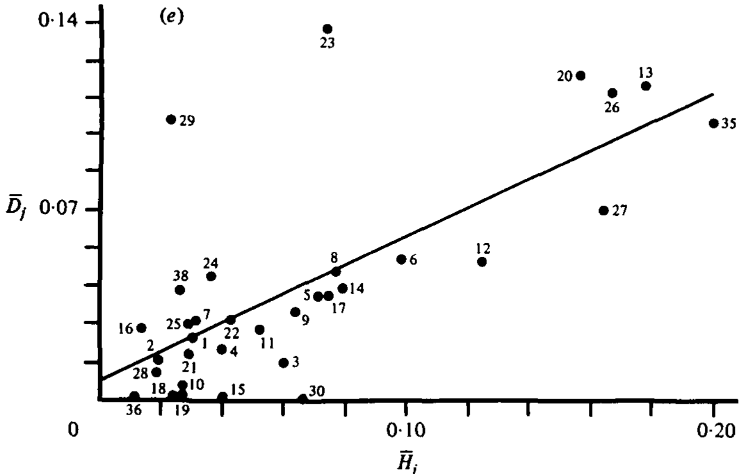


Fig. 1. Genetic distance ($\bar{D}_j$) plotted against heterozygosity ($\bar{H}_j$) for five sets of protein data. (*a*) Invertebrate interspecies comparisons; (*b*) vertebrate interspecies comparisons; (*c*) *Drosophila* interspecies comparisons; (*d*) invertebrate intraspecies comparisons; (*e*) vertebrate intraspecies comparisons

(——), Regression lines of $\bar{D}_J$ on $\bar{H}_J$. (----), Simulated neutral expectations for the stepwise mutation model and are presented for the interspecies plots only. The interspecies plots utilize independent comparisons. Proteins are identified below.

The five sets of numbers in parentheses after each protein are the number of comparisons contributing data for that protein in data sets (*a*) to (*e*) respectively. 1, alcohol dehydrogenase (–, 74, 26, 5, 26); 2, malate dehydrogenase (152, 394, 25, 46, 74); 3, α-glycerophosphate dehydrogenase (52, 178, 24, 29, 51); 4, isocitrate dehydrogenase (111, 227, 20, 41, 69); 5, sorbitol dehydrogenase (37, 75, –, 25, 32); 6, 6-phosphogluconate dehydrogenase (62, 155, 16, 36, 58); 7, lactate dehydrogenase (29, 436, –, 10, 73); 8, malic enzyme (73, 76, 25, 31, 33); 9, glucose-6-phosphate dehydrogenase (16, 28, 19, 17, 23); 10, superoxide dismutase (82, 175, 19, 31, 66); 11, aspartate aminotransferase (144, 291, 20, 42, 66); 12, phosphoglucomutase (100, 269, 22, 51, 73); 13, 'esterases', excluding isozymes specific for umbelliferyl esters (168, 383, 74, 42, 59); 14, glucosephosphate isomerase (96, 222, –, 47, 64); 15, xanthine dehydrogenase (44, 50, 20, 25, 14); 16, glutamate dehydrogenase (15, 46, –, –, 20); 17, peptidase (excluding leucine aminopeptidase) (52, 219, –, 19, 41); 18, fumarase (36, 46, 24, 13, 13); 19, leucine aminopeptidase (75, 76, 25, 36, 18); transferrin (–, 65, –, –, 29); 21, unspecified non-enzymatic proteins (152, 451, –, 22, 54); 22, haemoglobin (–, 97, –, –, 27); 23, albumin (–, 83, –, –, 35); 24, adenylate kinase (58, 42, 28, 14, 14); 25, creatine kinase (–, 69, –, –, 15); 26, adenosine deaminase (–, 22, –, –, 12); 27, mannose phosphate isomerase (38, 44, –, 10, 19); 28, acid phosphatase (63, 25, 30, 28, 16); 29, 'nothing dehydrogenase' (–, 36, –, 6, 6); 30, alkaline phosphatase (36, 21, –, 12, 6); 31, octanol dehydrogenase (22, 22, 26, 12, –); 32, triosephosphate isomerase (–, –, –, 5, –); 33, amylase (20, –, –, 7, –); 34, catalase (–, –, –, 6, –); 35, glutamate pyruvate transaminase (–, –, –, –, 4); 36, umbelliferyl ester specific esterases (–, –, –, –, 6); 37, aldolase (42, –, 35, 13, –); 38, glyceraldehyde-3-phosphate dehydrogenase (50, –, –, 10, 8); 39, hexokinase (62, –, 53, 21, –); 40, aldehyde oxidase (37, –, 32, 12, –); 41, hexanol dehydrogenase (–, –, –, 6, –); 42, phosphoglycerate kinase (–, –, –, 4, –). Average sample sizes per protein are 66·3, 141·8, 28·2, 21·6 and 34·1 for data sets (a)–(e) respectively.

https://doi.org/10.1017/S0016672300022291 Published online by Cambridge University Press

Table 2. *Regression analysis of protein distance* ($\overline{\overline{D}}_j$) *on protein heterozygosity* ($\overline{\overline{H}}_j$) *as species or populations diverge*

I range	Number of Proteins	Number of Comparisons	Intercept ± S.E.	Regression coefficient ± S.E.	Correlation analysis r	Correlation analysis P	r^2	$\overline{\overline{H}}_j$	$\overline{\overline{D}}_j$	Intercept/ $\overline{\overline{D}}_j$
				Invertebrates: (a) intraspecies						
0·95–1·0	33	591	−0·003 ± 0·046	0·194 ± 0·034	0·718	< 0·001	0·515	0·096	0·016	−0·164
0·7–0·9	21	165	0·145 ± 0·301	0·624 ± 1·150	0·123	n.s.	0·015	0·108	0·212	0·684
				Invertebrates: (b) interspecies						
0·9–1·0	4	93	insufficient data							
0·7–0·9	21	679	0·128 ± 0·050	1·160 ± 0·392	0·562	0·004	0·316	0·103	0·247	0·519
0·5–0·7	18	472	0·377 ± 0·118	3·054 ± 1·383	0·484	0·021	0·234	0·066	0·579	0·650
0–0·5	10	214	0·890 ± 0·210	1·142 ± 1·340	0·350	n.s.	0·123	0·123	1·064	0·837
				Vertebrates: (a) intraspecies						
0·95–1·0	33	1091	−0·002 ± 0·002	0·240 ± 0·027	0·847	< 0·001	0·717	0·067	0·015	−0·124
0·7–0·9	25	321	0·037 ± 0·043	2·547 ± 0·645	0·634	< 0·001	0·402	0·052	0·170	0·218
				Invertebrates: (b) interspecies						
0·9–1·0	21	964	0·011 ± 0·013	0·716 ± 0·138	0·766	< 0·001	0·587	0·073	0·063	0·175
0·7–0·9	23	1523	−0·003 ± 0·069	4·356 ± 0·862	0·742	< 0·001	0·551	0·068	0·294	−0·009
0·5–0·7	20	930	0·113 ± 0·110	5·762 ± 1·370	0·665	< 0·001	0·442	0·072	0·533	0·211
0–0·5	14	621	1·356 ± 0·313	−1·094 ± 3·664	−0·086	n.s.	0·007	0·067	1·283	1·058

Interspecies analyses utilize independent comparisons only.

https://doi.org/10.1017/S0016672300022291 Published online by Cambridge University Press

of these proteins in fact represent a mixture of proteins. For example, no attempt has been made here to separate supernatant and mitochondrial isozymes, and category 21, unspecified non-enzymatic proteins, must inevitably be a heterogeneous assemblage. Certain proteins in the analysis are thus far unique to the vertebrate data sets (e.g. 20, transferrin; 22, haemoglobin; 23, albumin; 25, creatine kinase), others unique to the invertebrate sets (e.g. 37, aldolase; 39 hexokinase; 40, aldehyde oxidase). Although we have not here attempted to cross-correlate $\overline{D}_j$ or $\overline{H}_j$ values for those proteins scored in two or more data sets, some proteins are characterized by high heterozygosity and high divergence rates (e.g. 13, esterase; 27, mannose phosphate isomerase) and others by low heterozygosity and low divergence rates (e.g. 3, α-glycerophosphate dehydrogenase; 16, glutamate dehydrogenase; 21, unspecified non-enzymic proteins). It is also noteworthy that many of the outliers from the regression line are those with below average sample sizes.

(ii) *The changing relationship between* $\overline{D}_j$ *and* $\overline{H}_j$ *as species diverge*

A summary of results is given in Table 2. These analyses omit the *Drosophila* data set because of insufficient information. With respect to the interspecies data, analyses were performed using both independent pairwise comparisons and all pairwise comparisons, but since these two modes of analysis gave very similar results, Table 2 presents the results of the former analysis only.

Inspection of Table 2 reveals a number of interesting features. The highest correlation between $\overline{D}_j$ and $\overline{H}_j$, 0·847, is found in the vertebrate intraspecies identity range of 0·95–1·0. Then, for both intraspecies and interspecies comparisons, as divergence increases so the correlation coefficients and coefficients of determination decrease. For the more distantly related pairs of species, those with I values in the range of 0–0·5, the correlation coefficients are not significantly different from zero. However, it should be pointed out that these categories have reduced sample sizes.

The regression coefficient or slope increases in the initial stages of species divergence, but in the later stages ($I = 0$–$0·5$) decreases and becomes not significantly different from zero. As populations diverge, so the regression intercept with the $\overline{D}_j$ axis tends to increase (although note the exception in the vertebrate category $I = 0·7$–$0·9$ and note also that the only vertebrate data sets where the intercept is significantly different from zero are in the I ranges 0–0·5 and the pooled set $I = 0$–$1·0$). This increase can be measured in both absolute terms or, in relative terms, by reference to the ratio intercept/D_j.

(iii) *Expectations of neutral theory*

In neutral theory, genetic distance (D) can be approximated by a function of mutation rate (u) and the period of time (t) that has elapsed since a pair of populations became isolated (Nei, 1972):

$$D = 2ut \quad (1)$$

Heterozygosity can be approximated by a function of u and effective population size (N_e): infinite allele model (Kimura & Crow, 1964):

$$H = 1 - 1/(1 + 4N_e u), \quad (2)$$

https://doi.org/10.1017/S0016672300022291 Published online by Cambridge University Press

or stepwise mutation model (Ohta & Kimura, 1973):

$$H = 1 - 1/(1 + 8\,N_e u)^{\frac{1}{2}} \tag{3}$$

By substituting for u, D can be expressed in both models as a function of H, t and N_e: infinite allele model:

$$D = (t/2N_e)(H/(1-H)), \tag{4}$$

stepwise mutation model:

$$D = (t/4N_e)(1/(1-H)^2 - 1). \tag{5}$$

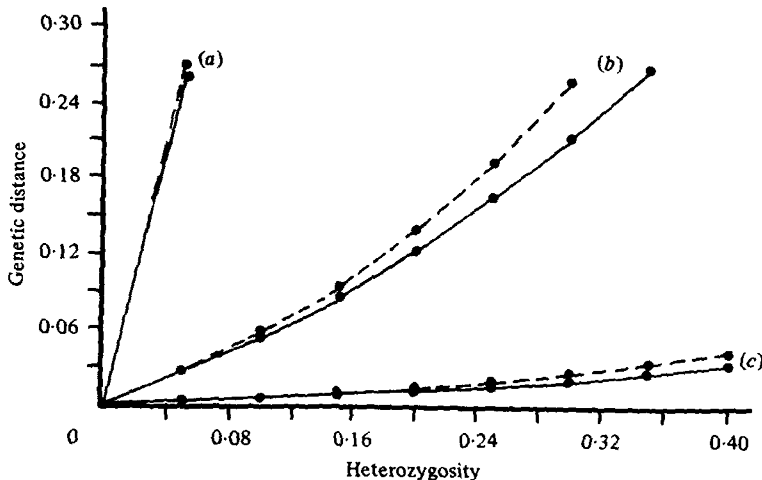


Fig. 2. Expected relationships between D and H derived from equations for both infinite allele (——), and stepwise mutation (----), neutral models. The three pairs of lines are for: (a) $t = 10N_e$, (b) $t = N_e$ and (c) $t = 0{\cdot}1\,N_e$.

Thus for both models the expected relationships between D and H are dependent upon the relative values of t and N_e. Fig. 2 plots these relationships over the heterozygosity range 0·0–0·40 for $t = 10N_e$, $t = N_e$ and $t = 0{\cdot}1\,N_e$. The lines are approximately linear over this H range and, as expected, give $D = 0$ when $H = 0$. As t increases, that is as populations or species diverge, the slopes of the lines increase but still intercept the origin. Linear regressions provided a good fit to these lines, particularly in the H range 0·0–0·20 where most observations fall, but the slight curve means that such regressions in fact intercept the H axis at a point very close to $H \approx 0$. It should also be pointed out that the infinite allele and stepwise mutation models give very similar lines for given values of t and N_e, although for specified values of H, the expected value for D is slightly higher in the stepwise model. At low H, equations (4) and (5) both simplify to

$$D = (t/2N_e)\,H,$$

giving a linear relationship between D and H.

How well do the observed relationships between D and H correspond to these neutral expectations? Briefly, populations with little overall divergence fit expectations very well, but as divergence increase so the fit becomes poorer.

https://doi.org/10.1017/S0016672300022291 Published online by Cambridge University Press

The vertebrate intraspecies file consisting of populations with I values between 0·95 and 1·0 may be considered as an example of low divergence. The observed distribution of proteins, together with the expected stepwise mutation line, is given in Fig. 3. The stepwise mutation model of Kimura & Ohta (1973) was used as a more realistic representation of electrophoretic variation than the infinite allele

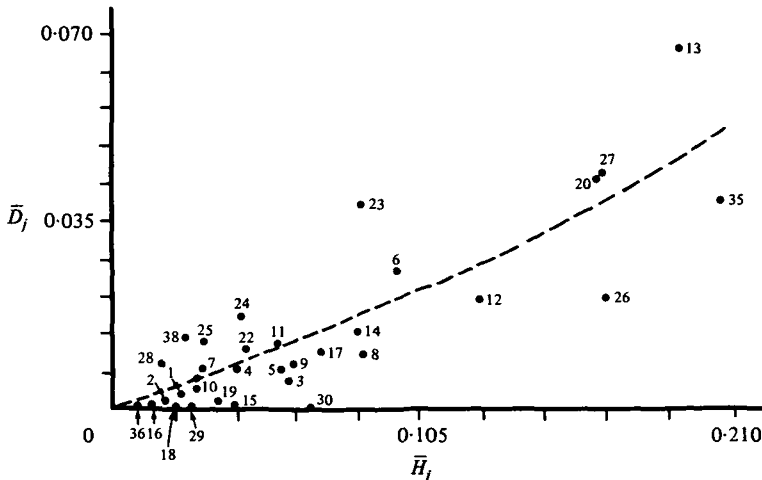


Fig. 3. The relationship between D and H in the intraspecies vertebrate data file $I = 0·95–1·0$. Protein identities are as in Fig. 1. The dashed line is the line expected of the stepwise mutation neutral model, where $t = 0·346\,N_e$.

model. In order to derive this line, expected values of D for the observed values of H were calculated for various values of the parameter t/N_e using equation (5), and then that value of t/N_e chosen which, following a linear regression of expected D on observed H, gave a slope identical to that observed. Thus for the relationship $t = 0·346\,N_e$, expected values of D for the observed values of H should fall on the plotted line. Observed points are in fact scattered around this line, and the observed correlation coefficient, although high ($r = 0·847$, Table 2) is lower than that expected of a perfect fit to neutral theory ($r = 0·996$, less than 1 since the expected relationship is slightly curved). Sampling error may be a major contributory factor here, many of the sampled proteins with D values substantially different from expected having low sample sizes.

In this intraspecies data file, the overall average genetic distance, $\bar{\bar{D}}$, is equal to 0·015 (Table 2). Using equation (1) and neutral mutation rates of between 10^{-6} and 10^{-7} per year, the average divergence time is estimated at between 75000 and 7500 years. Inserting these values into the equation $t = 0·346\,N_e$ gives N_e values of between 200000 and 20000. These estimates of t and N_e must be crude approximations, but they do not seem unreasonable. The general conclusion must be that where divergence is low, observed relationships between D and H correspond satisfactorily with those predicted by neutral theory.

One statistical point must be made here. When considering populations with little divergence, it is necessary to take into account sampling error in allele frequency determination. This alone will generate a positive relationship between D and H even subpopulation allele frequencies are identical and genotype

https://doi.org/10.1017/S0016672300022291 Published online by Cambridge University Press

frequencies accord with Hardy–Weinberg expectations (Nei & Roychoudhury, 1974). The magnitude of this relationship depends on sample size and allele frequencies. Nei (1978*a*) shows that when populations are genetically identical, an approximate value for this 'spurious distance' can be estimated from the formula $\mathrm{E}(D) = (1-G)/(2nG)$ where $G = \Sigma x_i^2 = \Sigma y_i^2$ (x_i and y_i being the sample allele frequencies in populations X and Y respectively) and n is the number of (diploid) individuals sampled.

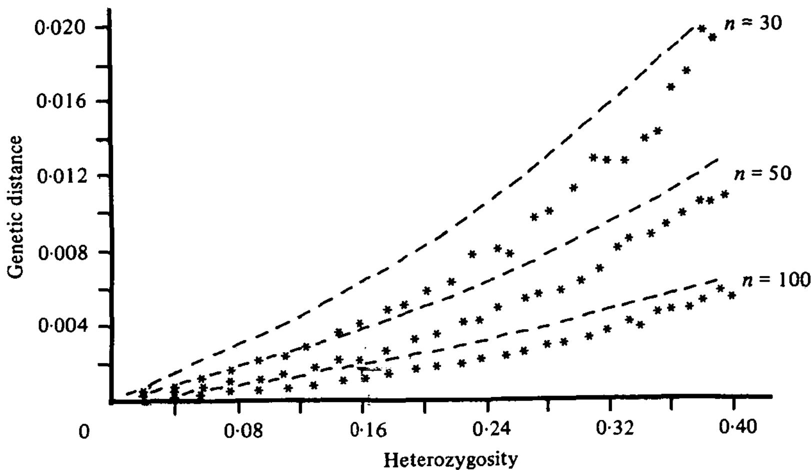


Fig. 4. The relationship between D and H generated by sampling error alone for sample sizes of 30, 50 and 100 genes per subpopulation. Simulation results are presented as asterisks, analytical results as dashed lines. See text for further details.

Fig. 4 gives the relationship between 'spurious distance' and sample heterozygosity using this formula for $n = 15$, 25, and 50, and compares these results with results from Monte Carlo simulations on diallelic polymorphisms. In these simulations, pairs of subpopulations with identical allele frequencies were sampled for 30, 50, or 100 genes (corresponding to 15, 25, or 50 individuals) per subpopulation and the genetic identity and mean sample heterozygosity calculated. At each set of allele frequencies, 1000 such samples were taken and the mean identity (and heterozygosity) calculated. This identity value was then converted to a distance value. Subpopulation allele frequencies ranged from allele $a = 0$ to allele $a = 0{\cdot}28$ in steps of 0·01, corresponding to a heterozygosity range of 0 to about 0·4.

Fig. 4 shows that agreement between the (approximate) formula predictions and the simulation results was reasonable although the former was always somewhat greater than the latter. For example, with a sample heterozygosity of around 0·20 and $n = 15$, the formula generates a 'spurious distance' value of about 0·008 compared with the simulation value of about 0·006. This compares with observed distance values of about 0·05 for heterozygosities of about 0·20 in the vertebrate intraspecies data set of Fig. 3. Most of our sample sizes are substantially greater than 15 and thus sampling errors will generate $\overline{D}_j$ values substantially less than 0·006 at $\overline{H}_j = 0{\cdot}20$. It can be concluded that this source of sampling error can only

https://doi.org/10.1017/S0016672300022291 Published online by Cambridge University Press

account for a small proportion of the observed relationship between $\overline{D}$ and $\overline{H}$, a proportion that decreases as divergence increases.

In general, where divergence is low, the observed relationships between $\overline{D}$ and $\overline{H}$ correspond satisfactorily with those predicted by neutral theory. As divergence increases, however, so goodness of fit to neutral expectations appears to decrease. This is exemplified by intercepts on the $\overline{D}$ axis at points significantly greater than zero and, in the I ranges 0–0·5 (high divergence), a reduction in the slope of the regression. However, it should be noted that the approximation $D = 2ut$ is not applicable in the presence of bottlenecks (Chakraborty & Nei, 1977). Furthermore, the approximation becomes less valid under the stepwise model as t increases, because of the possibility of back mutation (Nei & Chakraborty, 1973; Li, 1976; Nei, 1978*b*). Thus the rate of accumulation of genetic distance slows down as time increases. Li (1976) shows that this decreasing rate of accumulation of genetic distance is dependent on heterozygosity levels.

It should be pointed out that the average identity $\overline{I}_j$ used in our analysis of allozyme data is not precisely equivalent to the genetic identity I used in theoretical analyses. The former is the average of locus identity values while the latter is based on the summation over loci of squared allele frequencies and their cross-products. Where data are pooled from different pairs of species from different studies, I can be calculated but its value can change depending on the arbitrary assignment as species X or species Y of the constituent members of the species pairs. We chose to work with I_j as it is not affected in this way, but in practise the difference in magnitude between I and $\overline{I}_j$ is likely to be small. $\overline{I}_j$ is less easily interpretable than I in evolutionary terms, although in certain circumstances, such as systematic studies, it may be the more useful statistic (Hillis, 1984).

The relationship between $\overline{H}_j$ and $\overline{D}_j$ (i.e. $-\log \overline{I}_j$) under the stepwise mutation model of neutral theory has been determined here by computer simulation. Equations for Nei's I which take account of back mutation have been obtained analytically for conditions of neutral equilibrium (Chakraborty & Nei, 1973; Li, 1976) and the genetic distance values obtained from these equations agree fairly closely with those obtained for the statistic $\overline{D}_j$ from our computer simulations (Skibinski & Ward, unpublished results).

(iv) *Simulations of stepwise neutral theory*

Two populations of equal size were allowed to diverge under the influence of mutation and genetic drift. Mating was at random with non-overlapping generations. The average heterozygosity of the two populations and Nei's genetic identity between them was monitored at intervals. Small population sizes (15 diploid individuals) with high mutation rates were used which permitted substantial genetic distance to accumulate in a relatively small number of generations. Some results are shown in Fig. 5. Each plotted point is based on 1000 replications. The average identity ($\overline{I}_j$) for these replications was computed and converted to genetic distance (D_j). D_j is plotted against average heterozygosity over the 1000 pairs of populations. It is immediately obvious from Fig. 5 that the relationship between H_j and D_j is approximately linear for low amounts of divergence (equivalent here to a low number of generations), as expected from equation (3) and Fig. 3.

https://doi.org/10.1017/S0016672300022291 Published online by Cambridge University Press

However, with high overall levels of divergence (i.e. a high number of generations), the relationship reaches an asymptote with lower heterozygosity proteins accumulating relatively more genetic distance than those of high heterozygosity. This general observation has been found to hold with higher population sizes and lower mutation rates (Skibinski & Ward, unpublished results). The asymptotic nature of the relationship provides a possible explanation for the non-zero intercepts obtained when linear regressions of observed $\overline{D}_j$ on $\overline{H}_j$ are computed at higher levels

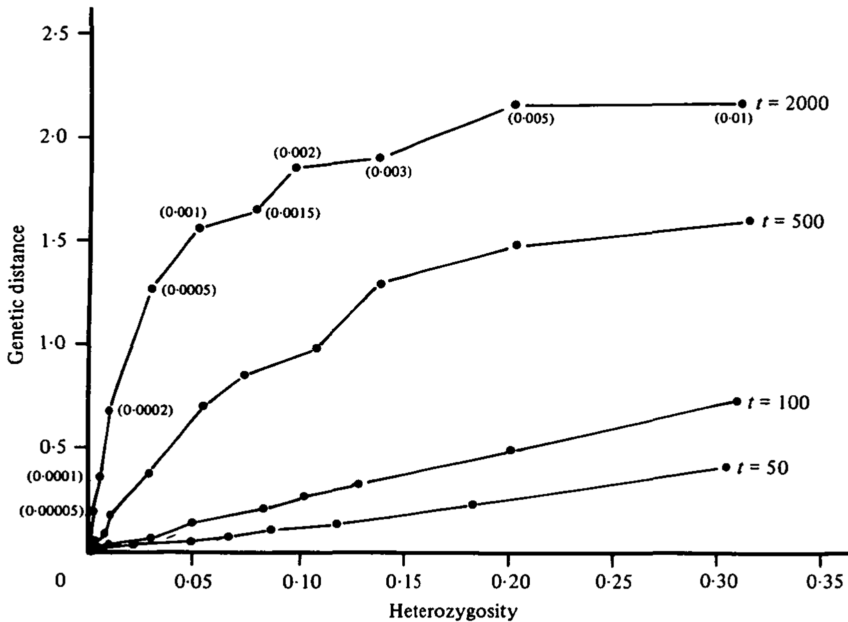


Fig. 5. Simulations of the expected relationships between D and H using the stepwise mutation neutral model. Figures in parenthesis are mutation rates, and t is number of generations. See text for further details.

of divergence. It is in fact not necessary that the average distance accumulated be great, because pooling a few studies having high genetic distance with others of low genetic distance will tend to flatten out the relationship at high heterozygosity more than would be predicted had all studies the same average level of divergence.

The computer simulation results were used to generate expected values of genetic distance to compare with those observed for each protein in a manner which would take into account variation in overall genetic distance between the population or species comparisons contributing data for each protein. The procedure adopted was as follows:

(1) By extrapolation and further simulation, the number of curves of Fig. 5 was increased to give coverage of the graph up to 8000 generations. Thirty extra curves were added in this way.

(2) For each population or species comparison, the average heterozygosity values for the proteins scored in that comparison were used to obtain expected distance values for each protein for each curve in the graph.

(3) The expected protein distance values were then averaged and the curve

https://doi.org/10.1017/S0016672300022291 Published online by Cambridge University Press

Table 3. *Comparisons of linear regressions of observed and expected* D *on* H *using independent interspecies data sets of given* I *ranges. Expected values are those derived by computer simulations of the steady state stepwise allele neutral model*

	Class	Identity range	Number of proteins	Regression coefficient	Intercept ±S.E.	Correlation coefficient	$\overline{\overline{D}}_j$
Observed	*Drosophila*	0–1·0	20	1·104 ± 0·352	0·310 ± 0·066	0·594	0·484
Expected				1·747 ± 0·098	0·176 ± 0·019	0·973	0·452
Observed	Invertebrates	0–1·0	29	2·055 ± 0·538	0·270 ± 0·057	0·592	0·444
Expected				2·913 ± 0·214	0·172 ± 0·023	0·934	0·418
Observed	Vertebrates	0–1·0	31	2·214 ± 0·531	0·215 ± 0·046	0·612	0·368
Expected				2·635 ± 0·198	0·159 ± 0·017	0·927	0·341
Observed	Invertebrates	0–0·5	10	1·142 ± 1·340	0·890 ± 0·210	0·350	1·064
Expected				3·424 ± 0·544	0·912 ± 0·085	0·912	1·332
Observed	Vertebrates	0–0·5	14	−1·094 ± 3·664	1·356 ± 0·313	−0·086	1·283
Expected				9·070 ± 1·574	0·694 ± 0·134	0·857	1·305

https://doi.org/10.1017/S0016672300022291 Published online by Cambridge University Press

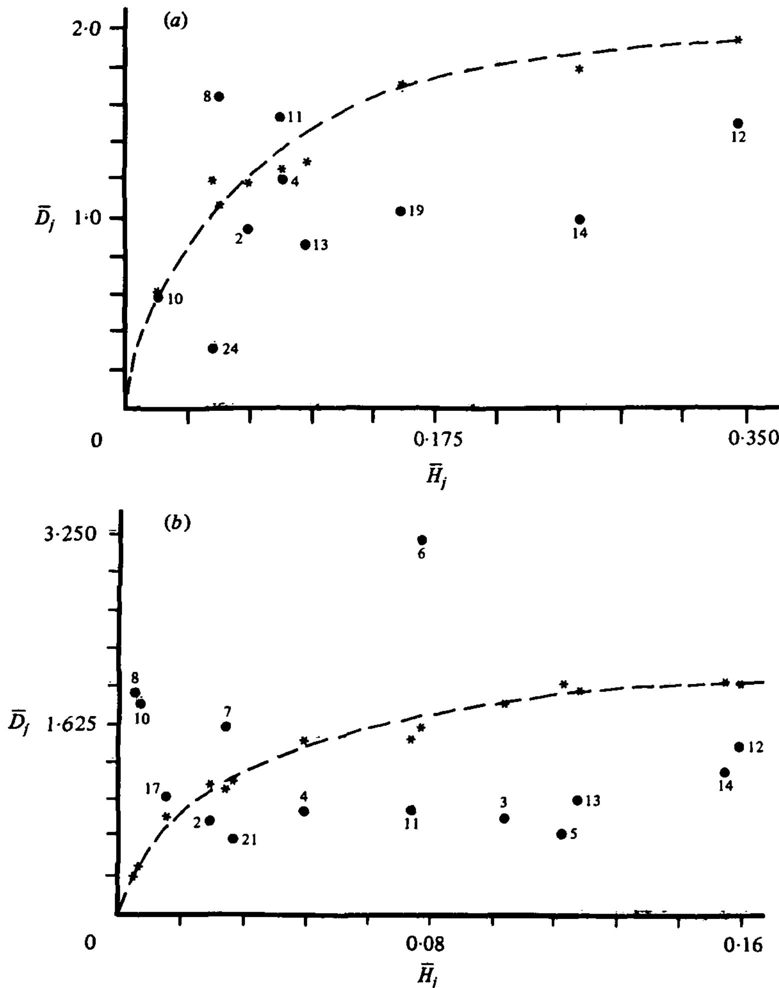


Fig. 6. Observed (●) and expected (*) distributions of D on H for proteins in species pairs showing high divergence ($I = 0·0–0·5$). The expected points are derived using the computer simulations, and the dashed line is a line fitted by eye to these points. Independent comparisons only. (*a*) Invertebrate analysis; (*b*) vertebrate analysis.

which gave the overall average genetic distance closest to that actually observed for that population or species comparison was chosen.

(4) The individual expected protein distance values for this curve were then transformed to identity values and added to running totals for each protein.

(5) Steps 2, 3 and 4 were repeated for each species or population comparison in the overall identity range file.

(6) The summed identities for each of the proteins over all the species or population comparisons were then averaged to give expected $\bar{I}_j$ values for each protein.

(7) The $\bar{I}_j$ values thus obtained were then transformed back to expected distance values.

The dashed lines in the interspecific comparisons of Fig. 1 are lines fitted by eye

https://doi.org/10.1017/S0016672300022291 Published online by Cambridge University Press

to these expected points (which are not themselves shown on the graphs). It can be seen that these fitted lines are not straight but curved and asymptotic, and thus linear regressions of these expected distance values on heterozygosity produce positive intercepts on the genetic distance axes. In fact, in the overall identity range of 0–1·0, Table 3 shows that expected slopes and intercepts are similar to those observed. Fits to expectations are rather less satisfactory for vertebrate and invertebrate species pairs of high divergence (identity range 0–0·5; see Fig. 6 and Table 3). This may be a consequence of the fairly small data sets in these files, and additional data are required for more rigorous comparisons.

5. DISCUSSION

In an earlier paper concerned with vertebrate interspecies comparisons (Skibinski & Ward, 1982), it was shown that there was a significant correlation between the mean genetic distance and mean heterozygosity of proteins. This finding can now be extended to include comparisons between invertebrate species and comparisons between populations of single species of invertebrates and vertebrates.

Similar findings have been reported earlier by Pierce & Mitton (1979). Data sets from three invertebrate and three vertebrate species showed strong correlations between heterozygosity (H) and Nei's genetic distance (D), ranging from 0·48 to 0·87 and all were significantly greater than zero. However, it is clear from the computer simulations detailed in the present paper that sampling error alone can generate a significant correlation between H and D, even when all populations have identical allele frequencies. Thus, it is conceivable that Pierce & Mitton's results could have arisen, at least in part, from this effect. But although such sampling artifacts can generate significant correlations, the regression coefficients produced are rather small (and inversely proportional to sample size). In our data sets, the regression coefficients of D on H are substantially larger than those anticipated through sampling error alone, and thus we believe we have demonstrated a real increase in evolutionary rate in proteins of higher heterozygosity.

The conclusion that a significant correlation exists between mean genetic distance ($\overline{D}_j$) and mean heterozygosity ($\overline{H}_j$) over a sample of different proteins does not necessarily carry with it the conclusion that individual loci that are highly heterozygous will evolve faster than those that are weakly heterozygous. Each protein will be monomorphic in some species and polymorphic in others, although proteins with lower $\overline{D}_j$ tend to have a higher frequency of monomorphic loci (Skibinski & Ward, 1982). One could therefore conceive that the rate of evolution of a protein in a species might be independent of the heterozygosity in that species but that the evolutionary rate differs between proteins. In fact, our observation that the correlation remains when heterozygous loci are excluded in the computation of protein genetic distance (but not in the computation of heterozygosity) is consistent with this (Skibinski & Ward, 1982). However, we favour the interpretation that the relationship arises because the probability of future divergence at a locus is related to its present heterozygosity. This interpretation is supported by an earlier analysis (Skibinski & Ward, 1981), where we demonstrated a correlation between locus heterozygosity and the probability of locus divergence.

https://doi.org/10.1017/S0016672300022291 Published online by Cambridge University Press

Our results can also be considered as a possible example of the somewhat controversial 'Kluge–Kerfoot phenomenon'. This phenomenon, first described by Kluge & Kerfoot (1973) and later supported by many other authors (see Rohlf, Gilmartin & Hart, 1983), relates to the increased diversity between populations of character traits that are highly variable within populations, and was first identified from the analysis of a number of morphological traits in several vertebrate species. Sokal (1976, 1978) reanalysed some of the data sets of Kluge and Kerfoot together with some new (morphological) data sets, and found generally similar results. He pointed out that such effects might arise from environmental effects on phenotypically plastic characters, or, if the trait has a heritable component, from either natural selection or random genetic drift. However, the existence of the Kluge–Kerfoot phenomenon has recently been questioned by Rohlf *et al.* (1983), who conclude that the patterns of morphological variation which appear to support the Kluge–Kerfoot phenomenon are simply statistical artifacts, caused by a general empirical correlation between the mean of a character and both within and among population coefficients of variation. They write that 'whilst there may actually be a correlation between the levels of within- and among-population variability, previous studies do not provide adequate data to allow investigation of such a relationship'.

An earlier study (Skibinski & Ward, 1981) showed that locus heterozygosity and divergence rates were correlated in comparisons of both intraspecific and interspecific populations. We suggest that those results, together with the present data, do provide evidence for the existence of the Kluge–Kerfoot phenomenon, at least at the allozyme level. Whether the patterns we observe arise from selective forces or from genetic drift is a much harder question to resolve. Note that while Sokal (1978) and Riska (1979) both point out that gene flow between local populations might lead to greater within population variation for characters showing greater among population variation, this cannot account for the relationship between heterozygosity and divergence in the interspecies comparisons where gene flow can be assumed to be negligible.

Neutral theory predicts that proteins that are highly variable within populations should also be more diverse between populations. Manipulation of the steady-state or equilibrium neutral equations shows that when distance is plotted against heterozygosity, the expected neutral relationship is approximately linear over the heterozygosity range observed in nature, and gives a distance equal to zero when heterozygosity equals zero. These expectations accord well with the intraspecies observations, particularly in cases when divergence is very low (identity range of 0·95–1). However, as divergence increases, so linear regression intercepts with the distance axis increase and become significantly greater than zero. We originally suggested that therefore not all allozyme polymorphisms were neutral (Skibinski & Ward, 1982), although Chakraborty & Hedrick (1983) did not agree with this conclusion. They proposed that the observed discrepancies from steady-state neutral predictions may have been caused either by sampling errors or by populations not being in neutral equilibrium.

It now appears that the results are consistent with steady-state neutral theory under the stepwise mutation model, the observed differences in distance and

https://doi.org/10.1017/S0016672300022291 Published online by Cambridge University Press

heterozygosity being largely the result of differences in neutral mutation rate. It does not seem necessary to invoke non-equilibrium models nor sampling errors to explain the positive intercepts of the interspecies comparisons. This reinterpretation arises from the observation that while the expected neutral relationship between distance and heterozygosity is approximately linear over the observed heterozygosity range when divergence is low, at high levels of divergence the relationship reaches an asymptote. The observed relationships between distance and heterozygosity agree closely with those predicted by the simulations.

Can the results also be explained by selection theory? Selection theory holds that allozyme polymorphisms are maintained by balanced selective forces such as heterosis or frequency-dependent selection. It is likely that our results could be consistent with selection theory if, for example, the more heterozygous loci are more often able than the less heterozygous loci to respond to changed environmental or ecological conditions by a change in genotype frequency equilibrium values, or by a change to directional selection in favour of a particular allele. Unlike neutral theory, selection theory does not at present make any clear and testable quantitative predictions. To be consistent with our results, such predictions would have to be similar to those of neutral theory in explaining both the initial high correlation between genetic distance and heterozygosity and its decline as divergence increases.

Thus to summarize, our results are consistent with neutral theory and may be consistent with selection theory. Indeed, it is interesting that the relationship which we originally thought not to be completely explicable by neutral theory (Skibinski & Ward, 1982), and which Chakraborty & Hedrick (1983) sought to explain by sampling errors or by non-equilibrium models, is precisely that expected under equilibrium neutral theory. However, certain predictions of neutral theory have not yet been tested in our analyses. We have not tested for the possibility that the scatter of point about the expected lines is greater than could be attributed to sampling error, in particular, it is expected that proteins of similar heterozygosities should have similar genetic distances. More data are being collected to enable us to examine this prediction.

We are indebted to the following who kindly provided us with unpublished information for incorporation into our data files: C. Anderson, A. R. Child, R. Galleguillos, H. J. Geiger, D. Hedgecock, L. R. Hilburn, E. Nevo, J. L. Patton, P. Sattler and G. Wallis. We wish to thank one of the referees for helpful comments on our earler draft of this manuscript.

REFERENCES

Adams, S. E., Smith, M. H. & Baccus, R. (1980). Biochemical variation in the American alligator. *Herpetologica* **36**, 289–296.

Andersson, L., Ryman, N., Rosenberg, R. & Stahl, G. (1981). Genetic variability in Atlantic herring (*Clupea harengus harengus*): description of protein loci and population data. *Hereditas* **95**, 69–78.

Ashton, R. E., Braswell, A. L. & Guttman, S. I. (1980). Electrophoretic analysis of three species of *Necturus* (Amphibia: Proteidae), and the taxonomic status of *Necturus lewisi* (Brimley). *Brimleyana* **4**, 43–46.

Ayala, F. J., Valentine, J. W., Barr, L. G. & Zumwalt, G. S. (1974). Genetic variability in a temperate intertidal Phoronid, *Phoronopsis viridis*. *Biochemical Genetics* **1**, 413–427.

https://doi.org/10.1017/S0016672300022291 Published online by Cambridge University Press

BAKER, R. J., HONEYCUTT, R. L., ARNOLD, M. L., SARICH, V. M. & GENOWAYS, H. H. (1981). Electrophoretic and immunological studies on the relationship of the Brachyphyllinae and the Glossophaginae. *Journal of Mammalogy* **62**, 665–672.

BARROWCLOUGH, G. F. (1980). Genetic and phenotypic differentiation in a wood warbler (Genus *Dendroica*) hybrid zone. *Auk* **97**, 655–658.

BECK, M. L. & PRICE, J. O. (1981). Genetic variation in the terrestrial isopod, *Armadillidium vulgare*. *Journal of Heredity* **72**, 15–18.

BLANC, F. & CARIOU, M.-L. (1980). High genetic variability of lizards of the sand-dwelling Lacertid genus *Acanthodactylus*. *Genetica* **54**, 141–147.

BRITTNACHER, J. G., SIMS, S. R. & AYALA, F. J. (1978). Genetic differentiation between species of the genus *Speyeria* (Lepidoptera: Nymphalidae). *Evolution* **32**, 199–210.

BROWN, K. (1981). Low genetic variability and high similarities in the crayfish genera *Cambarus* and *Procambarus*. *American Midlands Naturalist* **105**, 225–232.

BUROKER, N. E., HERSHBERGER, W. K. & CHEW, K. K. (1979*a*). Population genetics of the family Ostreidae. I. Intraspecific studies of *Crassostrea gigas* and *Saccostrea commercialis*. *Marine Biology* **54**, 157–169.

BUROKER, N. E., HERSHBERGER, W. K. & CHEW, K. K. (1979*b*). Population genetics of the family Ostreidae. II. Interspecific studies of the genera *Crassostrea* and *Saccostrea*. *Marine Biology* **54**, 171–184.

BUTH, D. G. (1979). Biochemical systematics of the Cyprinid genus *Notropis*. 1. The subgenus *Luxilus*. *Biochemical Systematics and Ecology* **7**, 69–79.

BUTH, D. G. (1980). Evolutionary genetics and systematic relationships in the catostomid genus *Hypentelium*. *Copeia* (1980), 280–290.

BUTH, D. G., BURR, B. M. & SCHENCK, J. R. (1980). Electrophoretic evidence for relationships and differentiation among members of the percid subgenus *Microperca*. *Biochemical Systematics and Ecology* **8**, 297–304.

CARLSON, D. M., KETTLER, M. K., FISHER, S. E. & WHITT, G. S. (1982). Low genetic variability in paddlefish populations. *Copeia* (1982), 721–725.

CASE, S. M. (1978). Electrophoretic variation in two species of Ranid frogs, *Rana boylei* and *R. muscosa*. *Copeia* (1978), 311–320.

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.

CHAKRABORTY, R., FUERST, P. A. & NEI, M. (1980). Statistical studies on protein polymorphism in natural populations: III. Distribution of allele frequencies and the number of alleles per locus. *Genetics* **94**, 1039–1063.

CHAKRABORTY, R. & HEDRICK, P. W. (1983). Heterozygosity and genetic distance of proteins. *Nature* **304**, 755.

CHAKRABORTY, R. & NEI, M. (1977). Bottleneck effects on average heterozygosity and genetic distance with the stepwise mutation model. *Evolution* **31**, 347–356.

CHAMBERS, S. M. (1980). Genetic divergence between populations of *Goniobasis* (Pleuroceridae) occupying different drainage systems. *Malacologia* **20**, 63–82.

CIANCHI, R., MAINI, S. & BULLINI, L. (1980). Genetic distance between pheromone strains of the European corn borer, *Ostrinia nubilalis*: different contribution of variable substrate, regulatory and non-regulatory enzymes. *Heredity* **45**, 383–388.

CIANCHI, R., URBANELLI, S., COLUZZI, M. & BULLINI, L. (1978). Genetic distance between 2 sibling species of the *Aedes mariae* complex (Diptera, Culicidae). *Parassitologia* (*Rome*) **20**, 39–46.

COCKLEY, D. E., GOOCH, J. L. & WESTON, D. P. (1977). Genic diversity in cave-dwelling crickets (*Ceuthophilus gracilipes*). *Evolution* **31**, 313–318.

DALY, J. C., WILKINSON, P. & SHAW, D. D. (1981). Reproductive isolation in relation to allozymic and chromosomal differentiation in the grasshopper *Caledia captiva*. *Evolution* **35**, 1164–1179.

DAVIS, B. J., DEMARTINI, E. E. & MCGEE, K. (1981). Gene flow among populations of a teleost (painted greenling, *Oxylebius pictus*) from Puget Sound to southern California, USA. *Marine Biology* **65**, 17–24.

DEW, R. D. & KENNEDY, M. L. (1980). Genic variation in raccoons, *Procyon lotor*. *Journal of Mammalogy* **61**, 697–702.

https://doi.org/10.1017/S0016672300022291 Published online by Cambridge University Press

DILLON, R. T. & DAVIS, G. M. (1980). The *Goniobasis* of southern Virgina and northwestern North Carolina, USA: genetic and shell morphometric relationships. *Malacologia* **20**, 83–98.
FAIRBAIRN, D. J. (1981*a*). Biochemical genetic analysis of population differentiation in Greenland halibut (*Reinhardtius hippoglossoides*) from the Northwest Atlantic, Gulf of St Lawrence, and Bering Sea. *Canadian Journal of Fisheries and Aquatic Science* **38**, 669–677.
FAIRBAIRN, D. J. (1981*b*). Which witch is which? A study of the stock structure of witch flounder (*Glyptocephalus cynoglossus*) in the Newfoundland region. *Canadian Journal of Fisheries and Aquatic Science* **38**, 782–794.
FEDER, J. H. (1979). Natural hybridization and genetic divergence between the toads *Bufo boreas* and *Bufo punctatus*. *Evolution* **33**, 1089–1097.
FERGUSON, M. M., NOAKES, D. L. G. & DANZMANN, R. G. (1981). Morphological and biochemical systematics of chubs, *Nocomis biguttatus* and *N. micropogon* (Pisces: Cyprinidae), in southern Ontario. *Canadian Journal of Zoology* **59**, 771–775.
FERRIS, S. D., BUTH, D. G. & WHITT, G. S. (1982). Substantial genetic differentiation among populations of *Catostomus plebeius*. *Copeia* (1982), 444–449.
FEVOLDEN, S. E. & AYALA, F. J. (1981). Enzyme polymorphism in Antarctic krill (Euphausiacea); Genetic variation between populations and species. *Sarsia* **66**, 167–182.
FUERST, P. A., CHAKRABORTY, R. & NEI, M. (1977). Statistical studies on protein polymorphism in natural populations. I. Distribution of single locus heterozygosity. *Genetics* **86**, 455–483.
GALLEGUILLOS, R. A. & WARD, R. D. (1982). Genetic and morphological divergence between populations of the flatfish *Platichthys flesus* (L.) (Pleuronectidae). *Biological Journal of the Linnean Society*, **17**, 395–408.
GLOVER, D. G., SMITH, M. H., AMES, L., JULE, J. & DUBACH, J. M. (1977). Genetic variation in Pika populations. *Canadian Journal of Zoology* **55**, 1841–1845.
GONZALEZ, A. M., CABRERA, V. M., LARRUGA, J. M. & GULLON, A. (1982). Genetic distance in the sibling species *Drosophila melanogaster*, *Drosophila simulans* and *Drosophila mauritiana*. *Evolution* **36**, 517–522.
GORMAN, G. C., BUTH, D. G., SOULE, M. & YANG, S. Y. (1980). The relationships of the *Anolis cristatellus* species group: Electrophoretic analysis. *Journal of Herpetology* **14**, 269–278.
GOULD, S. J. & WOODRUFF, D. S. (1978). Natural history of *Cerion* VIII. Little Bahama Bank – a revision based on genetics, morphometrics and geographic distribution. *Bulletin of the Museum of Comparative Zoology, Harvard University, USA* **148**, 371–415.
GRANT, W. S. & UTTER, F. M. (1980). Biochemical genetic variation in walleye pollock *Theragra chalcogramma*: population structure in the south eastern Bering Sea and the Gulf of Alaska. *Canadian Journal of Fisheries and Aquatic Science* **37**, 1093–1100.
GUTTMAN, S. I., WOOD, T. K. & KARLIN, A. A. (1981). Genetic differentiation along host plant lines in the sympatric *Enchenopa binotata* complex (Homoptera, Membracidae). *Evolution* **35**, 205–217.
HAFNER, D. J., PETERSEN, K. E. & YATES, T. L. (1981). Evolutionary relationships of jumping mice (genus *Zapus*) of the south western USA. *Journal of Mammology* **62**, 501–512.
HARRISON, R. G. (1979). Speciation in North American field crickets: evidence from electrophoretic comparisons. *Evolution* **33**, 1009–1023.
HEDGECOCK, D. (1979). Biochemical genetic variation and evidence of speciation in *Chthamalus* barnacles of the tropical eastern Pacific Ocean. *Marine Biology* **54**, 207–214.
HEDGECOCK, D., NELSON, K., SIMONS, J. & SHLESER, R. (1977). Genic similarity of American and European species of the lobster *Homarus*. *Biological Bulletin* **152**, 41–50.
HERTZ, P. E. & ZOUROS, E. (1982). Genetic variability in two West Indian anoles (Reptilia, Iguanidae): relation to field thermal biology. *Journal of Zoology* **196**, 499–518.
HILLIS, D. M. (1984). Misuse and modification of Nei's genetic distance. *Systematic Zoology* **33**, 238–240.
HORNBACH, D. J., MCLEOD, M. J. & GUTTMAN, S. I. (1980). On the validity of the genus *Musculium* (Bivalvia: Sphaeriidae): Electrophoretic evidence. *Canadian Journal of Zoology* **58**, 1703–1707.
HORNBACH, D. J. MCLEOD, M. J., GUTTMAN, S. I. & SEILKOP, S. K. (1980). Genetic and morphological variation in the freshwater clam, *Sphaerium* (Bivalvia: Sphaeriidae). *Journal of Molluscan Studies* **46**, 158–170.
KAWAMOTO, Y. & ISCHAK, T. M. (1981). Genetic differentiation of the Indonesian crab-eating

https://doi.org/10.1017/S0016672300022291 Published online by Cambridge University Press

macaque (*Macaca fascicularius*). 1. Preliminary report on blood protein polymorphism. *Primates* **22**, 237–252.

KIM, Y. J., GORMAN, G. C. PAPENFUSS, T. & ROYCHOUDHURY, A. K. (1976). Genetic relationships and genetic variation in the Amphisbaenian genus *Bipes*. *Copeia* (1976), 120–124.

KIMURA, M. & CROW, J. R. (1964). The number of alleles that can be maintained in a finite population. *Genetics* **61**, 763–771.

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

KORNFIELD, I., BELAND, K. F., MORING, J. R. & KIRCHEIS, F. W. (1981). Genetic similarity among endemic Arctic char (*Salvelinus alpinus*) and implications for their management. *Canadian Journal of Fisheries and Aquatic Science* **38**, 32–39.

KOVACIC, D. A. & GUTTMAN, S. I. (1979). Electrophoretic comparison of genetic variability between eastern and western populations of the opossum (*Didelphis virginiana*). *American Midlands Naturalist* **101**, 269–277.

KREPP, S. R. & SMITH, M. H. (1974). Genic heterozygosity in the 13-year cicada, *Magicicada*. *Evolution* **28**, 396–401.

LARSON, A. (1980). Paedomorphosis in relation to rates of morphological and molecular evolution in the salamander *Aneides flavipunctatus* (Amphibia, Plethodontidae). *Evolution* **34**, 1–17.

LAWSON, R. & DESSAUER, H. C. (1979). Biochemical genetics and systematics of garter snakes of the *Thamnophis elegans*, *Thamnophis couchii* and *Thamnophis ordinoides* complex. *Occasional Papers of the Museum of Zoology, Los Angeles State University, USA* No. 56, 1–24.

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

LI, W. H. (1976). Electrophoretic identity of proteins in a finite population and genetic distance between taxa. *Genetical Research* **28**, 119–127.

LOUDENSLAGER, E. J. & GALL, G. A. E. (1980). Geographic patterns of protein variation and subspeciation in cutthroat trout, *Salmo clarki*. *Systematic Zoology* **29**, 27–42.

MCKAYE, K. R., KOCHER, T., REINTHAL, P. & KORNFIELD, I. (1982). A sympatric sibling species complex of *Petrotilapia* Trewavas from Lake Malawi analysed by enzyme electrophoresis (Pisces: Cichlidae). *Zoological Journal of the Linnean Society* **76**, 91–96.

MCLEOD, M. J., WYNES, D. L. & GUTTMAN, S. I. (1980). Lack of biochemical evidence for hybridization between two species of darters. *Comparative Biochemistry and Physiology*, **67** B, 323–325.

MATTHEWS, T. C. & CRAIG, G. B. (1980). Genetic heterozygosity in natural populations of the tree-hole mosquito *Aedes triseriatus*. *Annals of the Entomological Society of America* **73**, 739–743.

MAYER, W. (1981). Electrophoretic investigations on European species of the genera *Lacerta* and *Podarcis*. III. *Podarcis tiliguerta*, species or subspecies? *Zoologischer Anzeiger* **207**, 151–157.

MAYER, W. & TIEDEMANN, F. (1981). Electrophoretic investigations on European species of the genus *Lacerta* and *Podarcis*. II. To the systematic status of the lizards of the island *Piperi* (Northern Sporades, Greece). *Zoologischer Anzeiger* **207**, 143–150.

MORAN, C., WILKINSON, P. & SHAW, D. D. (1980). Allozyme variation across a narrow hybrid zone in the grasshopper, *Caledia captiva*. *Heredity* **44**, 69–81.

MORGAN, R. P., BLOCK, S. B., ULANOWICZ, N. I. & BUYS, C. (1978). Genetic variation in the soft-shelled clam, *Mya arenaria*. *Estuaries* **1**, 255–258.

MULLEY, J. C. & LATTER, B. D. H. (1980). Genetic variation and evolutionary relationships within a group of 13 species of Penaeid prawns. *Evolution* **34**, 904–916.

NEMETH, S. T. & TRACEY, M. L. (1979). Allozyme variability and relatedness in six crayfish species. *Journal of Heredity* **70**, 37–43.

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*a*). Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* **89**, 583–590.

NEI, M. (1978*b*). The theory of genetic distance and evolution of human races. *Japanese Journal of Human Genetics* **23**, 341–369.

https://doi.org/10.1017/S0016672300022291 Published online by Cambridge University Press

NEI, M. (1983). Genetic polymorphism and the role of mutation in evolution. In *Evolution of Genes and Proteins* (ed. M. Nei and R. K. Koehn). Sunderland, Massachusetts: Sinauer.

NEI, M. & CHAKRABORTY, R. (1973). Genetic distance and electrophoretic identity of proteins between taxa. *Journal of Molecular Evolution* **2**, 323–328.

NEI, M., CHAKRABORTY, R. & FUERST, P. A. (1981). Statistical studies of protein polymorphism – a reply. *Genetics* **97**, 495.

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

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

NEVO, E. (1981). Genetic variation and climatic selection in the lizard *Agama stellio* in Israel and Sinai. *Theoretical and Applied Genetics* **60**, 369–380.

NEVO, E., BAR-EL, C., BAR, Z. & BEILES, A. (1981). Genetic structure and climatic correlates of desert landsnails. *Oecologia* **48**, 199–208.

NEVO, E. & YANG, S. Y. (1979). Genetic diversity and climatic determinants of tree frogs in Israel. *Oecologia* **41**, 47–63.

NICKLAS, N. L. & HOFFMANN, R. J. (1979). Genetic similarity between two morphologically similar species of polychaetes. *Marine Biology* **52**, 53–59.

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., SHERWOOD, S. W. & YANG, S. Y. (1981). Biochemical systematics of chaetodipine pocket mice, genus *Perognathus*. *Journal of Mammalogy* **62**, 477–492.

PHILLIPS, P. R. & MAYO, O. (1981). Statistical studies of protein polymorphism in natural populations. *Genetics* **97**, 495.

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

RAMSEY, P. R., AVISE, J. C., SMITH, M. H. & URBSTON, D. F. (1979). Biochemical variation and genetic heterogeneity in South Carolina deer populations. *Journal of Wildlife Management* **43**, 136–142.

RICE, M. C., GARDNER, M. B. & O'BRIEN, S. J. (1980). Genetic diversity in leukemia-prone feral house mice infected with murine leukemia virus. *Biochemical Genetics* **18**, 915–928.

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

RITTE, U. & PASHTAN, A. (1982). Extreme levels of genetic variability in two Red Sea *Cerithium* species (Gastropoda: Cerithidae). *Evolution* **36**, 403–407.

ROHLF, F. J., GILMARTIN, A. J. & HART, G. (1983). The Kluge–Kerfoot phenomenon – a statistical artifact. *Evolution* **37**, 180–202.

RYMAN, N., REUTERWALL, C., NYGREN, K. & NYGREN, T. (1980). Genetic variation and differentiation in Scandinavian moose (*Alces alces*): are large mamals monomorphic? *Evolution* **34**, 1037–1049.

SAGE, R. D. & SELANDER, R. K. (1975). Trophic radiation through polymorphism in cichlid fishes. *Proceedings of the National Academy of Sciences, USA* **72**, 4669–4673.

SATTLER, P. W. (1980). Genetic relationships among selected species of North American *Scaphiopus*. *Copeia* (1980), 605–610.

SAURA, A., HALKKA, O. & LOKKI, J. (1973). Enzyme gene heterozygosity in small island populations of *Philaenus spumarius* (L.) (Homoptera). *Genetica* **44**, 459–473.

SBORDONI, V., ALLEGRUCCI, G., CACCONE, A., CESARONI, D., SBORDONI, M. C. & MATTHAEIS, E. DE (1981). Genetic variability and divergence in cave populations of *Troglophilus cavicola* and *Troglophilus andreinii* (Orethoptera, Rhaphidophoridae). *Evolution* **35**, 226–233.

SCHMITT, L. H. (1978). Genetic variation in isolated populations of the Australian bush rat, *Rattus fuscipes*. *Evolution* **32**, 1–14.

SEIDEL, M. E., REYNOLDS, S. L. & LUCCHINO, R. V. (1981). Phylogenetic relationships among musk turtles (genus *Sternotherus*) and genic variation in *Sternotherus odoratus*. *Herpetologica* **37**, 161–165.

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

https://doi.org/10.1017/S0016672300022291 Published online by Cambridge University Press

Selander, R. K., Yang, S. Y., Lewontin, R. C. & Johnson, W. E. (1970). Genetic variation in the horseshoe crab (*Limulus polyphemus*), a phylogenetic 'relic'. *Evolution* **24**, 402–414.

Simonsen, V. (1982). Electrophoretic variation in large mammals. II. The red fox, *Vulpes vulpes*, the stoat, *Mustela erminea*, the weasel, *Mustela putorius*, the pine marten, *Martes martes*, the beech marten, *Martes foina*, and the badger, *Meles meles*. *Hereditas* **96**, 299–305.

Skibinski, D. O. F. & Ward, R. D. (1981). Relationship between allozyme heterozygosity and rates of divergence. *Genetical Research* **38**, 71–92.

Skibinski, D. O. F. & Ward, R. D. (1982). Correlations between heterozygosity and evolutionary rate of proteins. *Nature* **298**, 490–492.

Skibinski, D. O. F. & Ward, R. D. (1983). Heterozygosity and genetic distance of proteins – reply to Chakraborty & Hedrick. *Nature* **304**, 755–756.

Sluss, T. P., Sluss, E. S., Graham, H. M. & Dubois, M. (1978). Allozyme differences between *Helothis virescens* and *H. zea*. *Annals of the Entomological Society of America* **71**, 191–195.

Smith, M. F. (1979). Geographic variation in genic and morphological characters in *Peromyscus californicus*. *Journal of Mammalogy* **60**, 705–722.

Smith, M. F. & Patton, J. L. (1980). Relationships of pocket gophers (*Thomomys bottae*) populations of the lower Colorado River. *Journal of Mammalogy* **61**, 681–696.

Smith, P. J., Francis, R. I. C. C. & Paul, L. J. (1978). Genetic variation and population structure in the New Zealand snapper. *New Zealand Journal of Marine and Freshwater Research* **12**, 343–350.

Smith, P. J., McKoy, J. L. & Machin, P. J. (1980). Genetic variation in the rock lobsters *Jasus edwardsii* and *Jasus novaehollandiae*. *New Zealand Journal of Marine and Freshwater Research* **14**, 55–63.

Smith, P. J., Wood, B. A. & Benson, P. G. (1979). Electrophoretic and meristic separation of blue maomao and sweep. *New Zealand Journal of Marine and Freshwater Research* **13**, 549–551.

Sokal, R. R. (1976). The Kluge–Kerfoot phenomenon reexamined. *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: Springer-Verlag.

Steiner, W. W. M., Kitzmiller, J. B. & Osterbur, D. L. (1980). Gene differentiation in chromosome races of *Anopheles nuneztovari*. *Mosquito Systematics* **12**, 306–319.

Stock, M. W. & Castrovillo, P. J. (1981). Genetic relationships among representative populations of five *Choristoneura* species: *C. occidentalis*, *C. retiniana*, *C. biennis*, *C. lambertiana*, and *C. fumiferana* (Lepidoptera: Tortricidae). *Canadian Entomologist* **113**, 857–865.

Stoneking, M., Wagner, D. J. & Hildebrand, A. C. (1981). Genetic evidence suggesting subspecific differences between northern and southern populations of brook trout (*Salvelinus fontinalis*). *Copeia* (1981), 810–819.

Tilley, S. G. & Schwerdtfeger, P. M. (1981). Electrophoretic variation in Appalachian populations of the *Desmognathus fuscus* complex (Amphibia: Plethodontidae). *Copeia* (1981), 109–119.

Tracey, M. L., Nelson, K., Hedgecock, D., Shleser, R. A. & Pressick, M. L. (1975). Biochemical genetics of lobsters: genetic variation and the structure of American lobster (*Homarus americanus*) populations. *Journal of the Fisheries Research Board of Canada* **32**, 2091–2101.

Vuorinen, J., Himberg, M. K-J. & Lankinen, P. (1981). Genetic differentiation in *Coregonus albula* (L.) (Salmonidae) populations in Finland. *Hereditas* **94**, 113–122.

Ward, R. D. & Warwick, T. (1980). Genetic differentiation in the molluscan species *Littorina rudis* and *Littorina arcana* (Prosobranchia: Littorinidae). *Biological Journal of the Linnean Society* **14**, 417–428.

Winans, G. A. (1980). Geographic variation in the milkfish, *Chanos chanos*. 1. Biochemical evidence. *Evolution* **34**, 558–574.

Yanev, K. P. & Wake, D. B. (1981). Genic differentiation in a relict desert salamander, *Batrachoseps campi*. *Herpetologica* **37**, 16–28.

Yang, S. Y. & Patton, J. L. (1981). Genic variability and differentiation in the Galapagos finches. *Auk* **98**, 230–242.

Zera, A. J. (1981). Genetic structure of 2 species of water striders (Gerridae, Hemiptera) with differing degrees of winglessness. *Evolution* **35**, 218–225.

https://doi.org/10.1017/S0016672300022291 Published online by Cambridge University Press