MULTIVARIATE ANALYSIS OF GENETIC VARIATION

CHARLES E. TAYLOR[1] AND JEFFRY B. MITTON[1]

Dept. of Ecology and Evolution, State University of New York, Stony Brook, N. Y. 11790

Manuscript received May 2, 1973
Revised copy received September 5, 1973

ABSTRACT

Multiple factor analysis was used to interpret geographical variation of gene frequencies. Allelic frequencies at three loci (two esterase loci, *Esr* and *Esh*, and a malic dehydrogenase locus) from ants (*Pogonomyrmex barbatus*) collected throughout Texas and reported by JOHNSON *et al.* (1969) were re-examined for patterns of correlation with five environmental measurements: mean January temperature, mean July temperature, mean annual precipitation, elevation, and annual growing season. These measurements and the associated gene frequencies at each sampling location were subjected to factor analysis. Variables highly correlated with the same factor were hypothesized to be causally related.

Both orthogonal and oblique rotations of the factor solution provided four factors with essentially the same interpretation. Variation at the *Esh* locus was associated with a continuum from cold-wet to hot-dry. Variation at the *Mdh* locus and some of that at the *Esr* locus was related to the severity of winter months. Other allelic frequencies at the *Esr* locus had high correlations with a third factor which seemed to be independent of the environmental measurements. None of the allelic frequencies had high correlations with the fourth factor which was interpretable as an altitudinal gradient.

MOST contemporary evolutionists subscribe to the view that natural selection maintains the majority of genetic variation observed in natural populations. Some of the selective forces impinging upon the genetic structure of natural populations will originate from the physical environment. Therefore, it is likely that some of the genetic variation observed in populations should be correlated with abiotic variables of the environment. Although a direct correspondence may sometimes be found between genetic and environmental variation (KOEHN and RASMUSSEN 1967; MERRITT 1972), such heuristically pleasing relationships will not always be the case (see, for example, SELANDER, HUNT and YANG 1969). More complex relationships would be expected in many circumstances: if there is a non-linear genetic response to an environmental variable, if the response of some loci to the environment is modified by genetic background, or if loci respond to an ensemble of environmental variables. We are concerned here with the elucidation and interpretation of complex relationships of genetic variation and physical components of the environment in natural populations.

[1] Present address: Dept. of Genetics, University of California, Davis, California 95616.

Genetics **76**: 575–585 March, 1974.

MATERIALS AND METHODS

Our data come from JOHNSON *et al.* (1969), who collected harvester ants (*Pogonomyrmex barbatus*) from thirty-one localities throughout Texas. They estimated gene frequencies of alleles at a malic dehydrogenase locus (*Mdh*) and two esterase loci (*Esr* and *Esh*). Each sampling location was also characterized by its elevation, rainfall, mean January temperature, mean July temperature, and length of growing season.

This analysis used only those localities having complete data reported, reducing the number of localities from thirty-one to twenty-five. Allele frequencies were transformed using an arcsin square root transformation (SOKAL and ROHLF 1969) which tends to equalize variances and to give those variables a normal distribution. The twenty-five observations of the thirteen variables were then used to construct a correlation matrix (Table 1).

The first problem is to determine if gene frequencies are in fact correlated with the environmental variables which have been measured. This problem may be rephrased in the following way. Suppose that allele frequencies have been measured for alleles at a single locus and that environmental parameters have been measured at each sample location. Denote the correlation between frequencies of alleles i and j by $r(A_i,A_j)$, between environmental variables E_i and E_j by $r(E_i,E_j)$, and between the frequency of allele i and environmental variable j by $r(A\ ,E_j)$. The $(p+q)\times(p+q)$ symmetric matrix of correlations, $\underset{\sim}{R}$, may then be formed and partitioned into submatrices of intercorrelations amongst the p independent allele frequencies $(\underset{\sim}{R}_{11})$, intercorrelations amongst the q environmental measures $(\underset{\sim}{R}_{22})$, and intercorrelations between allele frequencies and environmental measures $(\underset{\sim}{R}_{12})$. Because the allele frequencies must be independent, p will usually be one less than the total number of alleles. The following matrices result:

$$\left[\begin{array}{ccc|ccc}
r(A_1,\ A_1) & \ldots\ldots & r(A_1,\ A_p) & r(A_1,\ E_1) & \ldots\ldots & r(A_1,\ E_q) \\
\cdot & & \cdot & \cdot & & \cdot \\
\cdot & & \cdot & \cdot & & \cdot \\
\cdot & & \cdot & \cdot & & \cdot \\
r(A_p,\ A_1) & \ldots\ldots & r(A_p,\ A_p) & r(A_p,\ E_1) & \ldots\ldots & r(A_p,\ E_q) \\
\hdashline
r(E_1,\ A_1) & \ldots\ldots & r(E_1,\ A_p) & r(E_1,\ E_1) & \ldots\ldots & r(E_1,\ E_q) \\
\cdot & & \cdot & \cdot & & \cdot \\
\cdot & & \cdot & \cdot & & \cdot \\
\cdot & & \cdot & \cdot & & \cdot \\
r(E_q,\ A_1) & \ldots\ldots & r(E_q,\ A_p) & r(E_q,\ E_1) & \ldots\ldots & r(E_q,\ E_q)
\end{array}\right]$$

$$= \left[\begin{array}{c|c} \underset{\sim}{R}_{11} & \underset{\sim}{R}_{12} \\ \hline \underset{\sim}{R}_{21} & \underset{\sim}{R}_{22} \end{array}\right] = \underset{\sim}{R}.$$

Under the null hypothesis of no relationship between environmental variables and gene frequencies, the parametric values of the correlations in $(\underset{\sim}{R}_{12})$ are zero. Statistical tests have been developed to test this hypothesis and are described in MORRISON (1967). This hypothesis may be tested by observing the ratio of determinants,

$$V = \frac{|\underset{\sim}{R}|}{|\underset{\sim}{R}_{11}|\ |\underset{\sim}{R}_{22}|}.$$

Under the null hypothesis the quantity

$$X^2_f = -(N-1)C^{-1}lnV$$

where

$$C^{-1} = 1 - \frac{1}{12f\,(N-1)}\,(2\Sigma_3 + 3\Sigma_2)$$

$$f = \Sigma_2/2$$

$$\Sigma_s = (p+q)^s - (p^s + q^s) \qquad s=2,3$$

is distributed approximately as χ^2 with f degrees of freedom.

TABLE 1

Correlation matrix of observations on 13 genetic and ecological variables from 25 localities

Esh-4	1.0												
Esh-6	.329	1.0											
Esh-8	—.661	—.889	1.0										
Esr-2	.267	—.286	.100	1.0									
Esr-4	—.066	—.526	.411	.324	1.0								
Esr-6	.004	.536	—.396	—.448	—.986	1.0							
Mdh-2	.235	—.192	.058	.565	.250	—.318	1.0						
Mdh-3	—.256	.197	—.055	—.583	—.232	.301	—.989	1.0					
Elevation	.521	.013	—.171	.200	.176	—.191	.412	—.462	1.0				
Rainfall	—.577	—.718	.757	—.012	.260	—.226	—.228	.244	—.495	1.0			
Mean January temperature	—.336	.247	—.122	—.455	—.234	.287	—.598	.632	—.758	.343	1.0		
Mean July temperature	.514	.462	—.575	.295	—.278	.172	.283	—.278	.157	—.609	—.304	1.0	
Growing season	—.140	.645	—.506	—.366	—.332	.347	—.346	.391	—.634	—.218	.688	.230	1.0

Once it has been determined that the genetic variation is not independent of the physical measures of the environment, the problem becomes one of interpreting these correlations. Factor analysis is designed for this sort of problem.

The idea behind this method is to construct common factor variables, $F_1 \ldots F_m$, such that each observed variable can be represented by a linear combination of these factors plus a term which is unique to that variable. The model is therefore:

$$\begin{aligned} x_1 &= a_{11}F_1 + a_{12}F_2 + \ldots\ldots\ldots\ldots a_{1m}F_m + e_1 \\ x_2 &= a_{21}F_1 + a_{22}F_2 + \ldots\ldots\ldots\ldots a_{2m}F_m + e_2 \\ &\cdot \\ &\cdot \\ &\cdot \\ &\cdot \\ x_n &= a_{n1}F_1 + a_{n2}F_2 + \ldots\ldots\ldots\ldots a_{nm}F_m + e_n \end{aligned}$$

where x_1, $x_2, \ldots x_n$ are observed gene frequencies or environmental measurements, a_{ij}'s are parameters reflecting the weighting of the jth factor on the ith variable, and $e_1 \ldots e_n$ are specific factor variates. Simplification can be achieved by using fewer common factors than variables if a small number of factors can adequately account for the correlations between variables. The process of finding the factors and their loadings is called factor analysis. Accounts of this technique may be found in VAN DE GEER (1971); SEALE (1968); and HARMON (1960).

For concise descriptions of the complex relations between the variables in this study, the final result of the factor analysis, simple structure, is presented. Simple structure may be achieved under either or both of two assumptions: If the model assumes that the factors are not correlated, then the solution is called "orthgonal". If the factors are allowed to be correlated the rotation is termed "oblique". The precise methods for computing these loadings were minimum residual (MINRES) solutions followed by VARIMAX and COVARMIN (=OBLIMIN, $\gamma = .5$) rotations respectively.

RESULTS

Association between allozyme frequencies and weather was tested by the determinantal ratio test described above. This gave for *Esh*, $\chi^2_{15} = 54.73$ ($p<.001$); for *Esr* $\chi^2_{15} = 23.86$ ($.05<p<.1$); and for *Mdh* $\chi^2_{10} = 13.679$ ($.1<p<.2$). The relationship between gene frequency and weather is statistically significant for the *Esh* locus and is on the borderline of statistical significance for the *Esr* and *Mdh* loci, in essential agreement with the results obtained by JOHNSON *et al.* (1969), who analyzed the data in a somewhat different manner.

Given that there is a significant correlation between the two classes of variables, one ideally would want the most parsimonious description of that relationship in order to understand which variables are involved in the correlations, and to what degree. Simple structures for orthogonal and oblique rotations of the factor matrix are presented in Table 2.

The presentation of results may be considerably simplified by constructing "path diagrams" as shown in Figures 1 and 2 where the lines indicate standardized partial regressions of variables on factors which are, in the orthogonal case, equal to correlations. Only regressions larger than 0.5 are shown. Interpretation of these diagrams may be facilitated by noting that temperatures, rainfall, etc. are not of interest in themselves, but merely indicate what sort of environment the population lived in. Although the particular environmental variables in this study could conceivably be the direct causes of genetic variation, it is more

TABLE 2

Factor pattern matrices rotated to simple structure

Variable	Factor I	II	III	IV	Communality
2a. Orthogonal rotation					
Esh-4	—.64	.34	.05	.15	.562
Esh-6	—.81	—.25	—.31	—.21	.872
Esh-8	.96	.11	.16	.11	.973
Esr-2	—.01	.16	.28	.58	.447
Esr-4	.25	.13	.86	.14	.846
Esr-6	—.18	—.10	—1.03	—.22	1.162
Mdh-2	—.07	.20	.11	.92	.912
Mdh-3	.07	—.25	—.08	—.93	.946
Elevation	—.32	.85	.11	.18	.887
Rainfall	.86	—.20	.13	—.13	.806
Mean January temperature	.10	—.71	—.11	—.50	.783
Mean July temperature	—.63	—.01	—.13	.33	.530
Growing season	—.40	—.85	—.13	—.2	.952

2b. Oblique rotation

Variable	Factor I	II	III	IV	Communality
Esh-4	—.60	.33	.10	—.00	.566
Esh-6	—.83	—.18	—.32	—.11	.845
Esh-8	.96	.08	.15	.06	.943
Esr-2	.17	—.15	—.16	.84	.936
Esr-4	.10	—.00	.91	—.24	.946
Esr-6	—.11	.05	—1.01	.04	1.004
Mdh-2	.01	.09	.06	.84	.723
Mdh-3	—.08	.06	.14	—.96	1.048
Elevation	—.24	.81	.06	.16	.809
Rainfall	.86	—.27	.02	.03	.887
Mean January temperature	.16	—.53	—.35	.44	.857
Mean July temperature	—.62	.08	.09	—.14	.510
Growing season	—.49	—.37	.22	—.62	.970

Factor correlation matrix

	1	2	3	4
1	1.0			
2	—.12	1.0		
3	—.08	.34	1.0	
4	.28	.05	—.17	1.0

Entries under the heading "Factor" are the loadings of each factors onto the indicated variable. Entries under "Communality" indicate the proportion of variance of each variable accounted for by the common factors. A few communalities are greater than one because the linear model sometimes predicts more variation than is actually observed.

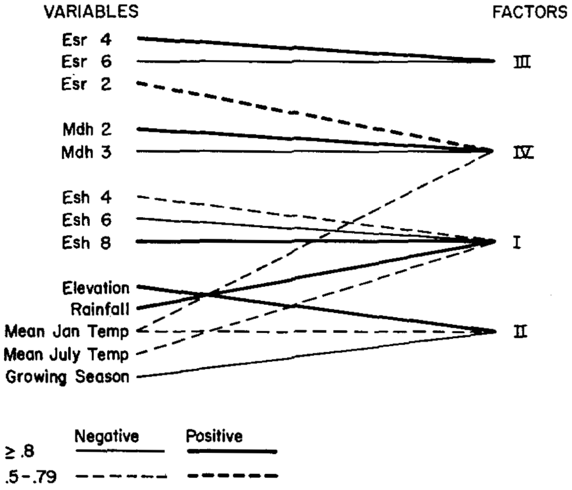

FIGURE 1.—Path diagram of factor pattern after orthogonal rotation.

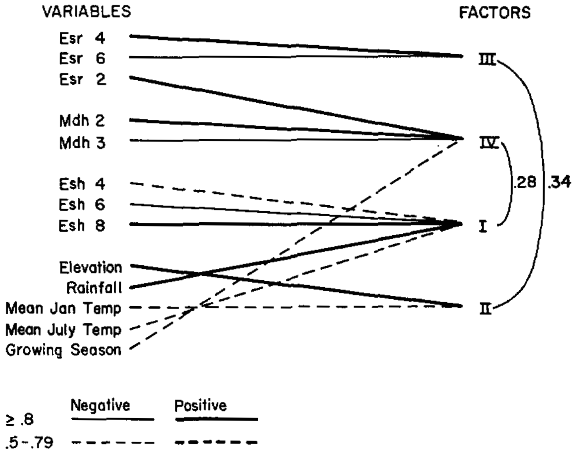

FIGURE 2.—Path diagram of factor pattern after oblique rotation.

probable that they are simply correlated with the real causes. Thus, in Figure 1 the effects of factor IV should, strictly speaking, be interpreted as, "the factor which caused variation at *Esr-2, Mdh-2,* and *Mdh-3* was correlated with mean January temperatures." We interpret this as "Lower winter temperatures caused increased frequencies of *Esr-2* and *Mdh-2* and a lower frequency of *Mdh-3.*" Similar reasoning would apply to the other factors.

While the results must be interpreted with caution, the following patterns seem to emerge:

1) Factor I represents a gradient of environments ranging from cold and wet to hot and dry. Thus *Esh-8* is replaced by *Esh-4* or *Esh-6* in hot, dry environments.

2) Factor II seems to represent an "altitude" gradient, so that as one comes down from the mountains, winters become warmer and the growing season becomes progressively longer. It is interesting that this pattern of environmental variation has no discernible influence on gene frequencies.

3) Factor III seems to be uncorrelated with any of the environmental variables which were measured. Thus, it seems that observed variation in none of the measured variables much affects replacement of *Esr-4* by *Esr-6* or *vice-versa.*

4) Factor IV is greater where January temperatures are low and July temperatures are high. Thus, it is a measure of temperature extremes, and when a population must adapt to such conditions, *Esr-2* and *Mdh-2* increase in frequency.

DISCUSSION

A problem central to population genetics and evolutionary theory today is whether or not the large amount of genetic variability apparent at the allozyme level is acted upon by natural selection. Many studies of variation in natural populations are designed to answer the question, "Does genetic variation respond to the environment, and if so, how?" It is generally difficult to perform controlled experiments on natural populations. One must usually be satisfied with systematic correlation between genetic and environmental variation. Genetic variation that may be interpreted as an adaptive response has been found in simple correlation studies (O'Gower and Nicoll 1968; Frydenberg *et al.* 1965; Hamrick and Allard 1972; Merritt 1972), but the application of this design is clearly limited.

The adaptive behavior of a locus may well be too complicated to be found by a simple one-to-one correlation with a biotic or physical variable. If environments are patchy, or the behavior of a locus is modified by the allelic complement at another locus, or if the locus responds to a composite of environmental variables, the data collected will be a series of observations on environmental variables that are correlated with one another. Univariate analysis of such data is simply not reliable, and multivariate analysis may be preferred.

There are many types of multivariate analysis which may be performed upon a matrix of observations of several variables. Generally speaking, one wishes to account for the variation of one or several variables by correlation with possible sources of variation. The type of analysis chosen depends upon the goal of the experiment or the question being asked.

Multiple regression has been used (Soulé 1972) to find the best predictor, in an assortment of ecological and geographical variables, of the phenetic variation of populations of the side-blotched lizard, *Uta stansburiana.* Similar methods have been used for analyzing allozyme variation by Smouse and Kojima (1972) and by Kojima *et al.* (1972). Multiple regression is a useful technique for this type of question because it provides a predictive model that may be tested further.

This technique must be used with caution, however, for multiple regression provides a test of the influence of individual independent variables upon one or more dependent variables with the effect of all other independent variables held constant. Clearly, tests of independence and interpretations of relationships between variables depend on which independent variables are included in the data. Consequently, multiple regression is not suited for reducing the number of independent variables needed to explain the variation among the dependent variables. For this task principal components analysis is commonly used.

Principal components analysis is used to reduce the dimensionality of a data matrix, so that the variation may more easily be described. The principal components are linear combinations of the original variables describing orthogonal axes fitted to the greatest amount of scatter in the multivariate space. This method was used by JOHNSON *et al.* (1969) in their original analysis of the data we examined here. They reduced the dimensionality of their data by considering just the first few principal axes of variation of the climatic variables. The correlations between the principal components of the variation of the two esterase polymorphisms and one or the other of the principal components of the weather variables were found to be significant. The principal components of the three allozyme polymorphisms from separate principal components analyses were also found to be correlated with one another. They concluded that the three protein polymorphisms were adaptive, and that the agents of natural selection were the ecological variables, either singly or in combination. Testing the correlation of principal axes from separate principal components analyses, however, is not a direct test of the covariation of climatic variables and genetic variation. If climatic variables with large amounts of variation are unrelated to genetic variation, then no relationships between the sets of variables would be found, even though genetic variation may be intimately associated with climatic variables with low amounts of variation—unless, that is, correlation coefficients were calculated for all or nearly all of the principal components of climatic variation.

A further difficulty with principal components analysis is shared with multiple regression discussed above. Correlations among climatic variables alone would tend to obscure simple relationships that may also exist between a subset of the weather variables and gene frequencies. One would like a means to separate those parts of the climatic variation which are correlated only with other climatic variation and those parts which may be correlation with allelic frequencies. This separation cannot usually be achieved with principal components analysis.

Canonical correlation, another multivariate statistical technique, provides a statistical test for association which is very similar to the ratio of determinants employed above. Canonical correlation finds those independent linear combinations of weather variables which are maximally correlated with other independent linear combinations of gene frequencies. These transformed variables are termed "canonical variates" and the correlations between them are the canonical correlation coefficients. Since there are several weather variables and gene frequencies there will be several canonical correlation coefficients. It is possible to compare the largest of these to the largest coefficient expected under the null hypothesis of no association (MORRISON 1967). The method used here is formally identical to testing the product of all the canonical correlations to that product expected under the null hypothesis. This seems to be a bit more thorough and, since the statistic is distributed as χ^2, complete tables are more readily available.

Beyond providing a test for significance, canonical correlation might be used in another way: the gene frequencies, when expressed as canonical variates, may be regressed on the canonical variates of climate. If only the few variates with the largest correlations are used, then simplification may be achieved. While this method would reduce the number of variates with an eye to maximizing correlations between environment and gene frequencies, it would ignore correlations among only the climatic variables and among only the gene frequencies. It would consequently fail to achieve the *clarification* described. It is desired to reduce the number of variables but, at the same time to retain, to the greatest possible extent, correlations amongst gene frequencies and among weather variables as well as between these sets of variables. One would like to identify all environmental trends which are present, and then identify those which cause variation in gene frequencies and those which do not. Patterns of correlation between gene frequencies and weather are interesting, but one would also like to identify patterns which are independent. It is interesting, for example, that the elevation factor (II) exists and is uncorrelated with gene frequencies, also that *Esr-4* varies with *Esr-6* but independently of any weather variables examined. This desired type of simplification can be achieved only by factor analysis.

Factor analysis is a form of multivariate analysis particularly useful for understanding patterns of covariation and for postulating causality from a data matrix (SOKAL, DALY and ROHLF 1961). The variation is broken down into two major components, the variation specific to a variable (uniqueness) and the variation associated with other variables in the study (communality). The technique then summarizes the variation common to two or more variables as stemming from a limited number of factors. This information may then be presented in the form of a path diagram, which shows which variables are influenced by the different factors, and to what degree. The greater resolution of the correlation matrix and simple representation of the interdependence of variables makes factor analysis a technique particularly useful for understanding complex data sets and for generating hypotheses for further studies or tests. An example of the utility of factor analysis may be taken from the present treatment, by factor analysis, of the Pogonomyrmex data. Although the methodology of JOHNSON *et al.* (1969) exposed correlations of similar magnitude to the ones found in this analysis, no simple interpretations of the relationships between climatic variables and genetic variation could be found in the correlations of principal axes. Elevation had the heaviest loadings on the first principal axis of the principal components analysis of the weather variables, which suggested a direct response of *Esh* and *Mdh* to elevation. The results from the factor analysis, however, indicate no major role in the determination of gene frequencies by elevation *per se*. Rather, elevation is seen to influence gene frequencies only secondarily, by influencing other variables such as summer temperatures and growing season. In addition, the path diagrams allow prediction of gene frequencies for other localities that fit into the described range of weather variables, providing further tests for the hypothesis of selective maintenance of the protein polymorphisms.

It should be stressed that factor analysis cannot *prove* that any particular interpretation of causation is correct. Only future testing can do this. The connection between correlation and causality is always intricate, and is especially so where factor analysis is concerned. An infinite number of path diagrams might be drawn which explain as much of the covariation equally as well as those shown in Figures 1 and 2. Factor analysis merely picks out that possible explanation which results in the simplest path diagrams. While there is no guarantee that simplicity is synonomous with truth, it would seem that the simplest hypothesis is a good one to examine first.

The experimental design of JOHNSON *et al.* (1969) and the methodology presented here may be applied to another related problem. The determination of niche space axes responsible for variation in gene frequencies is one of the most interesting and challenging problems in population genetics. If the environment can be ordinated to reproduce the niche space (in the sense of HUTCHINSON 1957) in which the population is found, then equilibrium gene frequencies would be expected to vary along one or more axes of the niche space. Multiple factor analysis may be of some use in determining the axes of the niche space (for example, the cold-wet, hot-dry axis in this study) and identifying variables which have little or no role in defining the niche (for example, the elevation factor in this study).

It is hoped that the examples and discussion presented here will promote further use of multivariate techniques in the analysis of complex ecological and genetic data.

We gratefully acknowledge guidance and assistance from DR. F. JAMES ROHLF and support from National Science Foundation Grant GB-25343 to DR. RICHARD K. KOEHN. Partial support was also received from National Science Foundation Grant GB-13213 and NIH Grant GM-10476. DRS. F. JAMES ROHLF, RICHARD K. KOEHN and ROBERT W. ALLARD made comments on the manuscript. We would also like to thank HENRY SCHAFFER for his constructive criticism of an earlier draft of this paper. Both authors are presently trainees on NIH training grant 5T01GM00701.

LITERATURE CITED

FRYDENBERG, OVE, D. MOLLER, G. NAEVDAL and K. SICK, 1965 Haemoglobin polymorphism in Norwegian Cod populations. Hereditas **53**: 257–271.

HAMRICK, JAMES and R. W. ALLARD, 1972 Microgeographical variation in allozyme frequencies in *Avena barbata.* Proc. Natl. Acad. Sci. U.S. **69**: 2100–2104.

HARMAN, H. H., 1960 *Modern Factor Analysis.* University of Chicago Press, Chicago.

HUTCHINSON, G. E., 1957 Concluding remarks. Cold Spring Harbor Symposium Quant. Biol. **22**: 415.

JOHNSON, F. M., H. E. SCHAFFER, J. E. GILLASPY and E. S. ROCKWOOD, 1969 Isozyme genotype-environment relationships in natural populations of the harvester ant, *Pogonomyrmex barbatus,* from Texas. Biochemical Genetics **3**: 429.

KOEHN, RICHARD K. and D. I. RASMUSSEN, 1967 Polymorphic and monomorphic serum esterase heterogeneity in Castostomid fish populations. Biochemical Genetics **1**: 131.

KOJIMA, KEN-ICHI, P. SMOUSE, S. YANG, P. S. NAIR and D. BRNCIC, 1972 Isozyme frequency patterns in *Drosophila pavani* associated with geographical and seasonal variables. Genetics **72**: 721–731.

MERRITT, ROBERT R., 1972 Geographic distribution and enzymatic properties of lactate dehydrogenase allozymes in the fathead minnow, *Pimephales promelas.* Amer. Nat. **106**: 173.

MORRISON, DONALD F., 1967 *Multivariate Statistical Methods.* McGraw-Hill, Inc., New York.

O'GOWER, A. K. and P. NICOL, 1968. A latitudinal cline in hemoglobins in a bivalve mollusc. Heredity **23**: 485.

SEALE, HILARY, 1968 *Multivariate Statistical Analysis for Biologists.* Methuen and Co., Lond.

SELANDER, ROBERT K., W. G. HUNT and S. Y. YANG, 1969 Protein polymorphism and genetic heterozygosity in two European subspecies of the house mouse. Evolution **23**: 379.

SMOUSE, PETER E. and K. KOJIMA, 1972 Maximum likelihood analysis of population differences in allelic frequencies. Genetics **72**: 709–719.

SOKAL, ROBERT R., H. V. DALY and F. J. ROHLF, 1961 Factor analysis procedures in a biological model. The University of Kansas Science Bulletin **42**: 1099.

SOKAL, ROBERT R. and F. J. ROHLF, 1969 *Biometry.* W. H. Freeman and Company. San Francisco.

SOULÉ, MICHAEL, 1972 Phenetics of natural populations. III. Variation in insular populations of a lizard. Amer. Nat. **106**: 429.

VAN DE GEER, JOHN P., 1971 *Introduction to Multivariate Analysis for the Social Sciences.* W. H. Freeman and Co., San Francisco.

Corresponding editor: R. ALLARD