(glyoxysomes) in castor bean endosperm. It has been established that these organelles, containing the enzymes of the glyoxylate cycle, are involved in the mobilization of lipid reserves and their subsequent conversion to carbohydrate. The intimate association between lipid bodies and peroxisomes in maize endosperm, observed during the present study, would support the contention that maize peroxisomes function in a similar fashion.

It may be concluded then, that m-MDH and p-MDH are distinct molecular species with catalytic differences and unique metabolic functions.

D. B. Hayden

2. Data analysis and maize cytogenetics.

The application of standard statistical tests is generally well understood and the techniques have been described in many elementary texts. Less well known are a number of techniques which in the biological field are more commonly used by numerical taxonomists and ecologists. In general, these are techniques which are concerned not so much with validating hypotheses as with the manipulation of data to bring to light information not apparent in the original data. Information in this sense may well include new problems or the rephrasing of old questions in a way which allows of more productive experimental design in subsequent investigations.

Perhaps the greatest problem is the difficulty which is faced in evaluating a large body of diverse data which contains many simultaneous variables. Presented with such a mass of data, one's tendency is to examine it in parts, looking, the while, to see if the rest of the data supports impressions gained from the part. This is a notoriously unreliable procedure, and consequently the whole basis of the investigative method rests on the sequential examination of limited hypotheses. While reliable, the method suffers from the fact that very frequently one is examining the wrong hypothesis. Clearly, any technique which will extract more information from a mass of multivariate data will enable specific hypotheses to be formulated much more efficiently. It is possible, indeed, that a meaningful answer can only be obtained by taking the mass of data as a whole.

A generalised, non-mathematical description of the problem makes the available techniques more easily understood. Consider a population of objects, characterised by a number of descriptors, each of which may assume two or more values. It may be desired to ask questions concerning the objects or the descriptors, or to ask that the relationships be simplified within certain constraints. In many cases, which elements are taken to represent descriptors and which to represent objects is optional, depending on the questions which are being addressed. An example, using Principal Components Factor Analysis, will make this point clear. A detailed account of this technique will be found in Seal (1964). Very briefly, Factor Analysis of this type is a technique whereby the descriptors of a population are replaced by linear combinations of the descriptors, called factors. The technique maximizes the variance removed from the system by the linear combination of descriptors contributing to the first factor. The process is repeated for the second and subsequent factors. Complete description of the object relationships will require as many factors as there were original descriptors, even in intrinsically simple systems, due to error variance. The distribution of information, however, has been changed. It is now concentrated in the first few factors: hence the description of the objects has been simplified.

The data obtained by Douglas and Filion (see contributions below) illustrate this interchangeability of objects and description.

A number of clustering techniques are available (see Orloci, 1968) where the object is the detection of groups of similar objects, rather than the simplification of relationships. As in the case of factor analysis, a degree of flexibility exists in designating objects and descriptors. Applicability of clustering techniques varies considerably and must be considered individually. The choice of clustering technique may be influenced by prior logical deductions concerning the groups of similar objects considered in spatial terms. If, by the nature of the data, the groups must necessarily be spherical or hyperspherical, then multiple linkage (not to be confused with genetic linkage) techniques, possibly Multidimensional Group Analysis (Jancey, 1966), are most appropriate. Failing such information, a cluster technique more

appropriate to the demonstration of elongate clinal variation should be used. A single linkage approach, such as the Graph Theory Model, (Wirth, Estabrook and Rogers, 1966) would be suitable. For a fuller discussion of clustering techniques, see Sokal and Sneath (1963).

Work in the field of Information Theory has led to the development of a number of powerful analytical tools well suited to cytological data. It is possible to measure the total information content of various arrangements of data, thus providing an objective basis for discussion and an effective means of selecting the most meaningful hypothesis. Another valuable approach is the measurement of information contained in one set of data, conditional on knowing a second set. Such an approach is directional and applicable to any data that can be expressed in the form of frequencies. It is generally more appropriate than either correlation or regression techniques. A discussion of the application of some Information Theory models to biological problems can be found in Estabrook (1967) and in Orloci (1971).

Finally, a technique should be mentioned which, while it involves conventional tests of significance, is extremely useful in the development of fresh or difficult to evaluate hypotheses. It involves the combining of probabilities from a number of independent statistical tests. The probabilities may be based on different sets of data and may refer to different statistical hypotheses. They should, however, all test the same scientific hypothesis, and the probabilities should be known exactly. A full account of the method, including a worked example, is given in Sokal and Rohlf (1969).

In conclusion, it may be said that a valuable collection of analytical tools are available to geneticists. Though they are as yet little known to biologists in general, they are for the most part well proven in other fields, and exist as fully developed computer programs.

References

Estabrook, George F. 1967. An information theory model for character analysis. Taxon 16:86-97.

Jancey, R. C. 1966. Multidimensional group analysis. Aust. J. Bot. 14:127-30.

Orloci, L. 1968. A model for the analysis of structure in taxonomic collections. Can. J. Bot. 46:1093-1097.

Orloci, L. 1971. An information theory model for pattern analysis. J. Ecol. (in press).

Seal, H. 1964. Multivariate statistical analysis for biologists. Wiley. New York.

Sokal, R. R. and Rohlf, F. J. 1969. Biometry. W. H. Freeman and Co., San Francisco.

Sokal, R. R. and Sneath, P. H. A. 1963. Principles of numerical taxonomy. W. H. Freeman and Co., San Francisco.

Wirth, M., Estabrook, George F. and Rogers, David J. 1966. A graph theory model for systematic biology, with an example from the Oncidiinae (Orchidaceae), Systematic Zoology 15:59-69.

<div align="right">R. C. Jancey</div>

3. Reanalysis of Longley's data: translocations in maize.

A. E. Longley (1958, 1961) has provided the most extensive cytological data on translocations available from any organism (Lindsley and Grell (1968) list approximately 800 translocations in Drosophila). Whereas the Drosophila data are based entirely on salivary analysis, Longley's data are determined entirely from pachytene analysis. Longley's data provide assignment of a 'breakage point' to a specific unit within a chromosome arm. He has discussed the limits of his data, e.g. sampling error, and some corrections have been listed in previous Newsletters (e.g. Burnham, 1969).

We have become interested in the distribution of genetic events and undertook to reanalyze Longley's data. Initially, we have attempted to describe the distribution, and in particular to establish any areas within chromosomes in which there can be detected a higher frequency of breakpoints than might be expected on a hypothesis of 'random distribution'. Our present report, which should be considered preliminary, consists largely of a summary of these distribution analyses. Next, however, we are reexamining the distribution in terms of probability statements for particular types of rearrangements. While not yet finished, it is clear that centromere-translocation probabilities far exceed that expected by chance alone. Lastly, designed but not yet processed, is a production of likelihood statements for different length-interstial and interchange-