An Appropriate Space for Clustering Selected Groups of Western North American Salmo

Pierre Legendre

Such is the true explanation for the spelling of taxonomy, a vocable perfectly proper on account of its long, universal usage in the English language, but certainly not be- cause of its etymology, since "taxonomy" should mean "treatment of the yew"—yew is τάξος in Greek. De Candolle should have written "taxinomie." (In present-day Greek, the reader might be interested to know, a "taxinomos" is a sorter.) Along with "taxidermy," a later term, "taximeter," is correctly constructed. "Taxometrics" is also correct as derived from "taxon," an offshoot of "tax-

onomy." On the other hand, "taxionomy," a word criticized by Mayr and others, becomes acceptable if made with ὄνομα (= name) instead of νόμος (= law, rule, way, manner); it would then restrictively designate the action of assigning names to hierarchical units, or the practice of classifying as opposed to the theory (implied in -nomy)—an occasionally useful shade of meaning.

GEORGES PASTEUR

Ecole Pratique des Hautes Etudes
34060 Montpellier, France

An Appropriate Space for Clustering Selected Groups of Western North American *Salmo*

A previous study (Legendre et al., 1972) looked at the phenetic relationship of nine populations of Western North American Salmonid fishes, many of which had no status in nomenclature. Representatives of the two presumed parental taxa, the rainbow trout *Salmo gairdneri* and the cutthroat trout *Salmo clarki*, were included. Single-linkage clustering, which was the main technique used in that paper, gives an exact representation of the pairwise relation between close neighbors, but because of its notorious tendency towards chaining, which results in a contraction of the space of descriptors in the vicinity of the clusters of objects (fishes), a rather tedious analysis of the connectedness pattern was used to identify the real moats of inter-population isolation. We present here another method of analyzing the same clustering diagram (ibidem, Fig. 1), a method perhaps more objective and certainly much easier to use and explain.

In a numerical-taxonomy study, it is recommended (Sneath and Sokal, 1973:303) to look at both a phenogram, obtained by some clustering technique, and at an ordination. Indeed, a component space of reduced dimensionality gives a faithful representation of the distances between the major groups of objects, and even though it, alone,

may falsify the distance between close neighbors, its use in conjunction with single-linkage clustering leads to a more balanced representation of the similarity relationships among objects.

Principal component analysis cannot be used on nonmetric descriptors, however, while clustering can be performed with any kind of descriptors, through the use of a proper similarity (or distance) coefficient. In the study considered here (Legendre et al., 1972), three of the eight descriptors were qualitative, and the others were ordered but not metric. To obviate this problem, we proceeded as follows: an appropriate similarity index was selected for the data at hand, and the similarity matrix was calculated between objects. In this case, the similarity index was that of Estabrook and Rogers (1966). This matrix was used as a basis for single-linkage clustering, as in the above-mentioned study (Legendre et al., 1972). On the other hand, the same similarity matrix was used as input for Gower's (1966) principal coordinate analysis, which describes in a space of reduced dimensionality the distance relationships between objects, based on any given similarity matrix.

The plot representing the objects in this reduced space was then used as a template
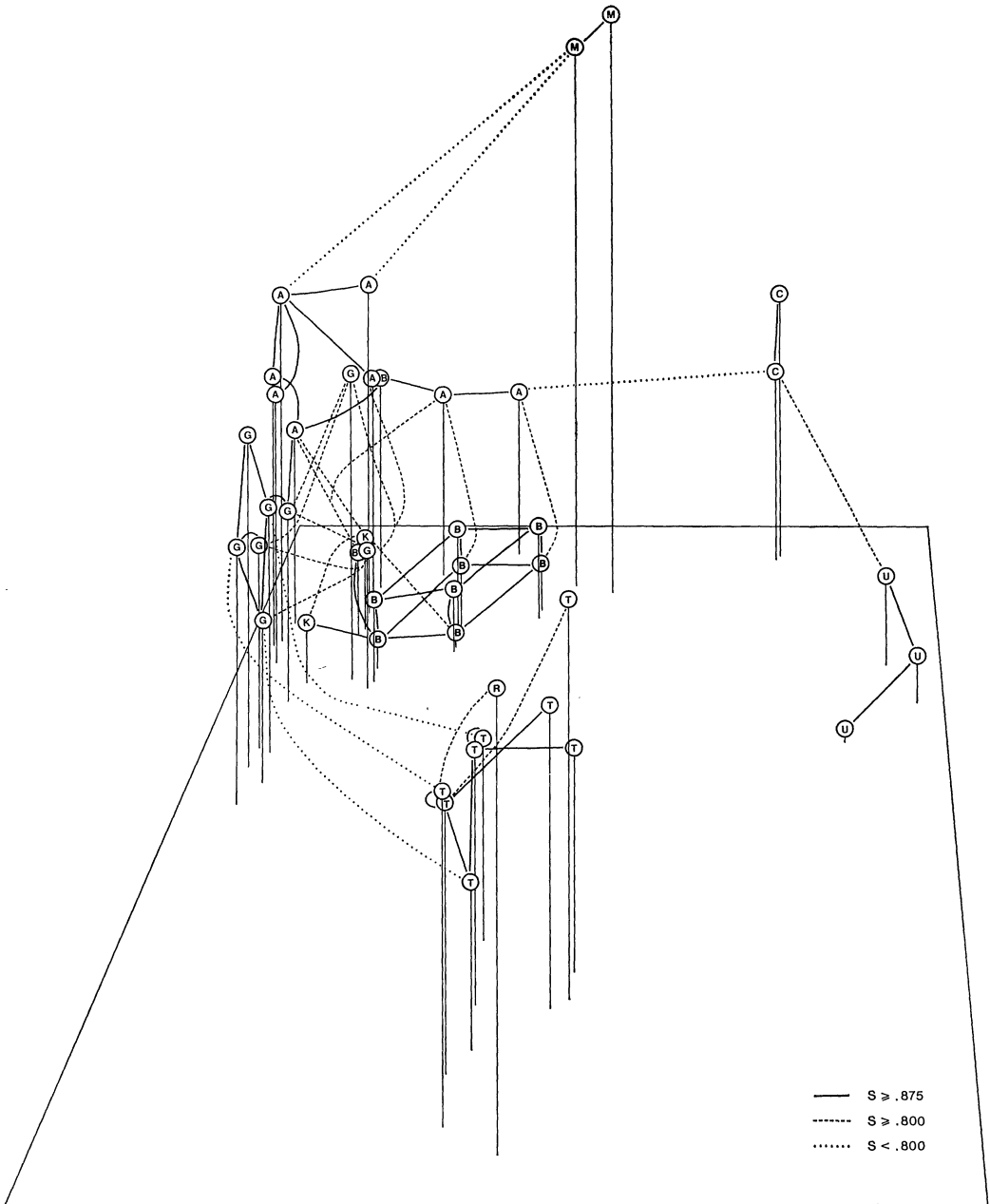
Fig. 1.—The chain of primary connections is drawn onto a plot of the objects (fishes) in a three-dimensional principal coordinates space. See text. The scale on the three axes is the same. The first principal axis (depth) accounts for 35% of the variation, the second (width) for 20% and the third for 16%. The nine populations are symbolized as follows: A = Apache trout (unnamed); B = Red-banded trout (unnamed); C = California golden trout, *Salmo a. aguabonita*; G = Gila trout, *S. gilae*; K = Kern River trout, *S. aguabonita gilberti*; M = Mexican golden trout, *S. chrysogaster*; R = rainbow trout, *S. gairdneri*; T = Rio Truchas trout (unnamed); U = cutthroat trout, *S. clarki*. Several fish may correspond to the same point. The nine fish forms were represented by 104 objects (specimens or collection averages), consisting of a total of 849 specimens.

on which the chain of primary connections (also called dendrites by Lukaszewicz, 1951, or minimum length tree in Sneath and Sokal, 1973) given by the single-linkage analysis was drawn, in the same way as Schnell (1970) did with principal components. The result is presented in Figure 1, and since it presents the advantages of both component analysis and single-linkage clustering, it is thus an appropriate model of the computed similarity matrix. And indeed, the conclusions laboriously reached in 1972 (Legendre et al., 1972) are readily confirmed by this ordinated phenogram: the Apache (A), Red-banded (B) and Kern River (K) trouts form a closely-linked group slightly distinct from the Gila trout (G), the four of which are known as the "golden trout complex." The California golden trout (C), although distinct, joins the cutthroat trout (U) before any other group. The trouts of the "golden trout complex" (A, B, G, and K) are more similar on the first principal axis (depth) to cutthroat trout (U) than to rainbow trout (R). The rainbow-like trout from the Rio Truchas (T) is closely associated to the 12 populations of rainbow trout (R), all of which fall on the same point. The Mexican golden trout (M) is unique in that it dissociates from the eight other trouts on the third principal axis. Had it not been for this form, a plot of the two first principal coordinates would have well rendered 62% of the phenetic variation.

It should be noted that the method presented here of associating single-linkage clustering and principal coordinates analysis is not limited to Q studies: since it is based on any previously computed association (or distance) matrix, it is possible to use any similarity coefficient of one's choice for the computation of this matrix, which makes the method applicable to any kind of data, including the study of the relation between descriptors.

REFERENCES

BLACKITH, R. E., AND R. A. REYMENT. 1971. Multivariate morphometrics. Academic Press, London and New York.

ESTABROOK, G. F., AND D. J. ROGERS. 1966. A general method of taxonomic description for a computed similarity measure. BioScience 16: 789–793.

ESTABROOK, G. F., H. S. FLEMING, P. LEGENDRE, D. J. ROGERS AND M. WIRTH. 1972. Cluster analysis program using graph theory. User's manual, assembled by H. S. Fleming and S. G. Appan, Gulf Universities Research Consortium, NASA/Mississippi Test Facility, Bay St. Louis. Reprinted by: Centre de Recherche écologique, Université du Québec à Montréal.

GOWER, J. C. 1966. Some distance properties of latent root and vector methods used in multivariate analysis. Biometrika 53:325–338.

LEGENDRE, P., C. B. SCHRECK AND R. J. BEHNKE. 1972. Taximetric analysis of selected groups of Western North American Salmo with respect to phylogenetic divergences. Syst. Zool. 21:292–307.

LUKASZEWICZ, J. 1951. Sur la liaison et la division des points d'un ensemble fini. Colloquium Mathematicum 2:282–285.

SCHNELL, G. D. 1970. A phenetic study of the suborder Lari (Aves). I. Methods and results of principal components analyses. Syst. Zool. 19: 35–57.

SNEATH, P. H. A., AND R. R. SOKAL. 1973. Numerical taxonomy—The principles and practice of numerical classification. W. H. Freeman and Co., San Francisco.

PIERRE LEGENDRE

Centre de Recherche en Sciences de l'Environnement
Université du Québec à Montréal
C.P. 8888, Montréal, Québec