

## Society of Systematic Biologists

---

Report on Seventeenth International Numerical Taxonomy Conference

Author(s): Pierre Legendre

Source: *Systematic Zoology*, Vol. 33, No. 1 (Mar., 1984), pp. 117-121

Published by: Taylor & Francis, Ltd. for the Society of Systematic Biologists

Stable URL: <https://www.jstor.org/stable/2413139>

Accessed: 26-07-2018 14:56 UTC

---

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact [support@jstor.org](mailto:support@jstor.org).

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <https://about.jstor.org/terms>



JSTOR

*Society of Systematic Biologists, Taylor & Francis, Ltd.* are collaborating with JSTOR to digitize, preserve and extend access to *Systematic Zoology*

## Report on Seventeenth International Numerical Taxonomy Conference

PIERRE LEGENDRE

Département de Sciences biologiques, Université de Montréal,  
C.P. 6128, Succursale A, Montréal, Québec H3C 3J7, Canada

The Seventeenth International Numerical Taxonomy Conference (NT-17) was held at the University of Ottawa, 7-9 October 1983. The meeting was organized, in both its material and scientific aspects, by a team composed of John McNeill (Dep. Biol., Univ. Ottawa) and Leonard P. Lefkovitch (Agriculture Canada, Ottawa). Twenty-two papers were presented to 53 registered participants.

Following the conference format introduced by Ted Crovello at NT-16 (Univ. Notre Dame), this year's meeting featured two preconference tutorials held on Friday afternoon, before the customary Friday night registration period and social gathering. L. P. Lefkovitch used the first tutorial to present a detailed discussion of the new method of *conditional clustering*, which is primarily aimed at obtaining a partition of objects in a character space (rather than a hierarchy). Given a predefined dissimilarity matrix, the method eliminates unlikely subgroups from consideration by using Maxwell's generalization of an ellipse, and retains only those which are possible. Redundant groupings are then eliminated by a logical minimalization process. Overlapping groups can be fused into *musters*, which form a partition. To choose among possible optimal solutions, the maximum entropy principle is first used, which depends only on group membership and their overlap, and then a maximum joint probability solution is employed to obtain a set covering or partition.

In the second tutorial, W. H. E. Day (Memorial Univ. Newfoundland, St. John's) gave a presentation on the relevance of computational complexity to classification research. He surveyed important concepts

in the evaluation of the complexity of problems and algorithms, such as intractability, NP-completeness, and approximation strategies. These concepts were illustrated by examples drawn from classification problems. Numerous participants showed up at both of these tutorials, but the two groups were mostly composed of different people (Jaccard similarity between attendances at these two events: 0.1). Tutorials seem to be an appropriate format for considering problems that cannot be presented in sufficient depth in regular 30-min talks; participants warmly expressed their satisfaction to the respective speakers.

The meeting speakers can be divided into *people with methods* and *people with data*; further branching of the first group divides the papers presented into *general-purpose methods* and *methods for phylogenetic reconstruction*. The meeting was organized roughly along these lines, although sessions bore other titles: most of the general-purpose methods were presented during sessions I and II, entitled "Clustering Algorithms" and "Consensus Methods" (chaired by L. P. Lefkovitch and W. R. Atchley, respectively). Sessions III and IV, both entitled "Cladistics" and chaired by B. R. Baum and J. Felsenstein considered methods connected with the reconstruction and evaluation of phylogenetic trees, without reference to Hennigian principles, so that these session titles were somewhat misleading. Finally, most of the speakers in sessions V to VIII (entitled "Applications"; chaired by H. Howden, W. W. Moss and J. McNeill) used methods to analyze problems of various flavors. These were intermingled with a few methodological presentations, which will be re-

ported together with the papers of groups 1 or 2.

1. *General-purpose methods.*—Diversity was high among this group of papers, thus showing the vitality of present research on numerical data analysis methodologies. W. H. E. Day looked for efficient agglomerative hierarchical clustering algorithms. He described algorithms requiring  $O(n^2)$  time to cluster  $n$  objects for all reducible combinatorial clustering methods and for many combinatorial methods that permit reversals.

R. Hansell (Univ. Toronto) showed how most SAHN clustering algorithms are related to underlying measures of predictivity. He pointed out that data analysts should ask themselves what questions (predictions) they want their classification to answer, since different questions correspond to different information and prediction measures.

F. J. Rohlf, S. Ferson and R. K. Koehn (SUNY, Stony Brook) discussed a way for summarizing shapes into a small number of descriptors which could then be used for morphometric analysis. Using samples from two electrophoretically distinguishable morphs of the blue mussel (*Mytilus edulis*) they: digitized TV images of each shell; automatically extracted the contour outlines; and computed both moment invariant and elliptic Fourier coefficients from these outlines. The coefficients were then used as characters to describe shell shape in discriminant analyses that were able to distinguish the two morphs.

Regionalization analysis is performed by a class of clustering and ordination methods that use both the information related to the variate or multivariate of interest in a given study, and the geographic information about the spatial relationships of the observation points. D. Wartenberg (SUNY, Stony Brook) first reviewed the four major types of methods: (1) *a posteriori* geographic testing of clustering solutions obtained without reference to geography; (2) clustering or ordination obtained with a spatial contiguity constraint; (3) geographic scaling of nongeographic variables; and (4) probabilistic ordering or

clustering. Then he presented his own method of canonical trend surface analysis, which pertains to class 3, and illustrated its use with allele frequency data of human HLA blood types in Europe. C. H. Smith (Univ. Illinois, Urbana), on the other hand, used nonconstrained clustering and ordination methods to produce a classification of world mammalian faunal regions in which component regions were explicitly equivalent in rank. He then showed how the latter quality could be used to represent interregional relationships; papers explaining these methods in detail are now in press (Smith, 1983a, b, 1984).

Three papers dealt with consensus, describing new consensus measures or comparing existing methods. L. P. Lefkovich first established that resolution and consistency were desirable properties for a consensus method. Then he showed how to obtain a consensus tree with these characteristics, using properties of well-known ordination methods: (1) for each dendrogram, obtain a set of principal coordinates equivalent to it; (2) fuse these into a consensus set of coordinates using generalized Procrustes analysis; (3) reversing step 1, determine the dendrogram equivalent to this set of coordinates. R. Stinebrickner (Berea College, Berea, Kentucky) discussed an approach for extending previously described intersection consensus methods for  $n$ -trees to consensus methods for dendrograms representing numerically stratified hierarchic nested classifications. He also described an algorithm for computing the consensus  $n$ -trees which result from Neumann's generalized intersection methods. This algorithm involves operations on the matrices of cophenetic values representing the original  $n$ -trees rather than intersections of subsets from the original  $n$ -trees. K. Shao (SUNY, Stony Brook) used Monte Carlo simulations to produce tables for testing the significance of 14 types of consensus indices. Input data were randomly-generated bifurcating or multifurcating trees. These tables can be used to test whether a consensus tree obtained from real data differs significantly

from a randomly generated one. Shao also studied the influence of number of OTUs, characters, character states, similarity measures, clustering methods, and consensus tree methods. He found that (1) they all affect the consensus index distributions, and (2) certain consensus indices perform better with particular clustering methods. This leads to an important warning against the usefulness of consensus indices when comparing results obtained from different taxonomic methods.

2. *Methods for phylogenetic reconstruction.*—All the papers under this heading dealt with some aspect of phylogenetic reconstruction, and presented data in support of the new methodologies. J. Felsenstein (Univ. Washington, Seattle) described a solution to the problem of placing confidence limits on the topologies of phylogenies, without assuming arbitrary branch lengths. Assuming instead that the branch lengths are controlled by an evolutionary clock, he showed how to compute confidence intervals for rooted three-species tree topologies, using a test for the difference between two trees of the same three species, measured in numbers of steps. Two different statistics were proposed, leading to essentially similar results.

Three papers discussed the analysis of reticulate evolution patterns, which can be represented by single-rooted acyclic directed graphs and not by trees. Lefkovich proposed that these graphs be called *reticulograms*. P. G. Smith (Univ. Western Ontario, London) showed that the presence of hybrids can lead to serious errors in phylogenetic reconstructions. On the other hand, the identification of hybrids may be achieved in the course of the phylogenetic analysis, especially through character compatibility analysis, since hybrid individuals are expected to generate incompatibilities between characters which might otherwise be perfectly compatible. He also discussed methods of differentiating hybridization from other sources of character incompatibility, such as homoplasy, or coding and polarization mistakes. D. Sankoff and V. Ferretti (Univ. Montréal) discussed the problem of enu-

merating reticulations in a phylogenetic reconstruction. To achieve this, they divided the problem into two separate structures: the phylogenetic history without character reconstruction (Schröder projection) on the one hand, and the character state network (precedence network) on the other. W. M. Fitch and W. R. Atchley (Univ. Wisconsin, Madison) analyzed reticulate evolution among 10 inbred strains of mice, using electrophoretic differences for 61 protein loci. The authors presented a phenetic method for detecting and reconstructing reticulate phylogenies, and tested it with their mice data, for which the true reticulogram is known. The hybrid was only detected once in several combinations involving it and its parental lineage. The authors tested the UPGMA, Fitch and Margoliash (F-M), parsimony, and four-point methods against their data. Only F-M and four-point gave the correct phylogeny (apart from hybridization), but the others were not markedly different.

Three other papers, all from Stony Brook, also dealt with the comparison of methods for reconstructing phylogenies. K. L. Fiala and R. R. Sokal used computer simulation to evaluate factors affecting the accuracy of phylogenetic reconstruction. Trees were simulated with various types and amounts of evolutionary events. The model also produced a data matrix of the character states of the recent species for each tree. The Wagner, compatibility, and UPGMA methods (used to reconstruct phylogenetic history) were about equally accurate as estimates of the true phylogeny; no method was especially good. The single largest determinant of accuracy of cladogram estimation appears to be "stemminess," a measure of the topology of the tree that looks promising as a general method for comparing trees. G. Hart, on the other hand, tested the stability of the same three methods under addition of new species, using a real data set (Leptopodomorpha). Stability was measured by strict consensus. All three methods proved about equally stable for this data set, contrary to the findings of Schuh and Farris (1981). Finally, R. R. Sokal, K. L. Fiala and G. Hart

tested once again the same methods of phylogenetic reconstruction using the Caminalcules data set (Sokal, 1983), whose true phylogeny is known. Methods were compared (1) by the criterion of stability under addition of new species, as described above, and also (2) with reference to the true phylogeny. The cladistic methods did better than UPGMA at reconstructing the true cladogram, regardless of number of OTUs employed. However, when samples are compared with their standards based on the entire tree, established for each reconstruction method (which would be the typical situation in real studies), stability seems to be a function of OTU sample size. For low OTU numbers, it is higher for the cladistic classifications, but for high OTU numbers it is higher for the phenetic classifications. Further work was reported showing the relative advantage of phenetic methods when the ratio of number of characters to number of OTUs is low and, conversely, the advantage of cladistic methods when the ratio is high. This series of three papers leads to the conclusion that none of the tree-reconstruction methods investigated by these workers is definitely better at reconstructing phylogenies. The combined use of several methods, as is common among practitioners, still seems advisable.

D. Sankoff, A. Alary and D. Roby (Univ. Montréal) proposed a dynamic programming method for identifying alleles in isozyme separation gels. The comparison of populations using electrophoretically separated isozymes is best carried out if it is known which bands (alleles) belong to which loci. In the case of a monomeric enzyme, an individual may show one band (homozygous) or two (heterozygous). The dynamic programming method proposed is based in the tendency of competing alleles to have only slightly different mobilities and, hence, to be represented by bands that are adjacent on the electrophoretic record. Discussants indicated that the problem whose solution was presented here is general in isozyme data research.

3. *Using methods to analyze data.*—Application papers covered topics in biogeog-

raphy, flora data base management, microbial taxonomy, and morphometrics. D. E. McAllister and F. W. Schueler (Nat. Mus. Nat. Sci., Ottawa) analyzed the distribution data from Lee's (1980) *Atlas of North American Freshwater Fishes*. They localized major centers of endemism and found them to be located mostly outside areas glaciated during the Wisconsin ice age, such as the Tennessee and Cumberland basins, the Ozarks and Ouchitas rivers, and the Rio Grande. On the other hand, synthetic climatic variables obtained by PCA were related to species diversity (number of species) distribution; aridity had the highest correlation with fish diversity. Neither dispersionalist nor Hennigian theories of biogeography survived unrefuted.

N. P. Moreno, A. Gómez-Pompa, L. Giddings, R. Allkin and L. Gama (Instituto Nacional de Investigaciones sobre Recursos Bióticos, Veracruz) reported on the data base management in the Flora of Veracruz project, initiated some 15 years ago (Gómez-Pompa et al., 1984). Their system, which many a museum could envy, includes: a curatorial data base (over 70,000 entries); a floristic list of all species with synonymies; accessory data bases to localities and authors' names; a data base for taxonomic terms and definitions in six languages, including Latin; as well as other accessory systems that complement the published version of the Flora. These include a descriptive data base for the angiosperm families in Veracruz and other experimental systems for automatic identification and description. These methods are or will be applied to floras of other parts of Mexico.

The University of Maryland (College Park) contributed three papers on microbial taxonomy. M. T. MacDonell and R. R. Colwell discussed the role of ribosomal RNA base sequence comparisons in bacterial systematics. Using this technique, they estimated the relationship of *Vibrio cholerae* to various other bacteria of the same and other genera, by clustering the difference matrix. R. R. Colwell and M. T. MacDonell presented a retrospective of how numerical taxonomy has influenced

the taxonomy of Gram-negative bacteria in the past 25 years. NT results first proved consistent with those obtained by classical methods, and were then used to assess taxa from the natural environment, as well as from food and clinical problems. More recently, NT results for Gram-negative bacteria were shown to be astonishingly well correlated with both nucleic acid hybridization and ribosomal RNA sequences. Finally, A. T. Wortman, A. M. Baya, D. J. Grimes and R. R. Colwell presented a NT study of abyssal marine bacteria, from depths of 4,300 and 4,700 m in the Bay of Biscay. Forty-three strains were examined for 130 characters, and compared by clustering to reference cultures isolated from surface waters. Whereas many phenotypic traits were shared by deep-sea and surface bacteria, clusters formed by the deep-sea organisms were distinct from those comprised of organisms isolated from surface waters, demonstrating the uniqueness of bacteria found in the abyssal environment.

The last talk of the meeting, delivered by R. E. Strauss (Univ. Michigan, Ann Arbor), discussed the problem of "static" allometry, found in many biological variables, in relation to the catfish genus *Corydoras* (about 100 nominal species). Many of these species have been described on the basis of meristic and morphometric characters that are highly correlated with body size among species. Thus, many of the nominal species are probably artifacts of habitat sampling and statistical variation. Those bivariate ratios that can be shown to be uncorrelated with size and correlated with shape factors may be useful for traditional taxonomists.

The banquet address, delivered by Charles Gruchy, Assistant Director of the National Museum of Natural Sciences of Canada, brought the participants back to reality, after the day's incursions into the most sophisticated concepts and methods for systematic investigation. The talk, entitled "The State of Systematics in Canada," showed how little effort is being devoted to inventory and systematic studies in Canada. In a country where much of

the economy depends on accurate knowledge and proper management of natural resources, this is a most astonishing and disheartening situation.

What is the immediate future of Numerical Taxonomy, as can be foreseen from this and the last few NT conferences? From the present reporter's point of view (admittedly biased), three fields of interest seemed to emerge or to grow stronger at the Ottawa conference, and should be heard of again in the next few years. These are: (a) spatial analysis of genetic, taxonomic and ecological phenomena, including biogeographic problems; (b) numerical methods for analyzing reticulate evolution, as can be depicted by *reticulograms*; (c) tree comparisons, including measures (like "stemminess") and tests of significance of the topology, as well as measures of consensus, where a lot of work still has to be done in order to compare the multitude of methods available and to sort out recommendations for practitioners. Future meetings will show how clairvoyant these predictions are. The NT-18 conference is to be held at Cornell University in the fall of 1984.

#### REFERENCES

- GÓMEZ-POMPA, A., N. MORENO, L. GAMA, V. SOSA, AND R. ALLKIN. 1984. Flora of Veracruz: Progress and prospects. Pages 165-174 in *Databases in systematics* (R. Allkin and F. Bisby, eds.). Academic Press, London and New York.
- LEE, D. S., ET AL. 1980. Atlas of North American freshwater fishes. North Carolina State Mus. Nat. Hist., Raleigh.
- SCHUH, R. T., AND J. S. FARRIS. 1981. Methods for investigating taxonomic congruence and their application to the Leptopodomorpha. *Syst. Zool.*, 30: 331-351.
- SMITH, C. H. 1983a. A system of world mammal faunal regions. Part 1. Logical and statistical derivation of the regions. *J. Biogeogr.* (in press).
- SMITH, C. H. 1983b. A system of world mammal faunal regions. Part 2. The distance decay effect upon interregional affinities. *J. Biogeogr.* (in press).
- SMITH, C. H. 1984. Areographic representation of faunal characteristics through a 'second order' relational approach. *Evol. Theory* (in press).
- SOKAL, R. R. 1983. A phylogenetic analysis of the Caminalcules. I. The data base. *Syst. Zool.*, 32:159-184.

Received 8 November 1983; accepted 21 December 1983.