Phenetic Approaches to Classification

W. WAYNE MOSS

Academy of Natural Sciences, Philadelphia, Pennsylvania 19103

SYNOPSIS. An assessment of similarity among individuals, species, and higher taxa is an essential component of biological classification. Multivariate phenetic techniques can be used to carry out this task, and to provide a logical basis for the recognition of taxa and the construction of phylogenies. Cluster analysis and ordination are current techniques used to group by similarity. These techniques are discussed and some of them are applied to a representative taxon of parasites, the mite family Harpyrhynchidae. Taxonomic structure within the family is described in terms of parasite morphology, host association, and probable coevolution.

INTRODUCTION

A systematist’s driving avocation is to harvest large quantities of novel information from his taxon of interest. New theoretical constructs and new techniques for analysis also appear constantly. Advances in data and technique can provide additional dimensions of knowledge, but they also pose a continual challenge: How does one integrate and reconcile them with existing classifications? Some taxonomists are inclined to believe that only new data sources and new techniques yield crucial insight; others are more inclined to stretch the existing classifications rather than to demolish them.

The background of an expanding data base has been coupled with the appearance of three major and frequently conflicting conceptual schools of taxonomy: the conventional or evolutionary (Mayr, 1969), the phenetic (Sneath and Sokal, 1973; McNeill, 1979), and the cladistic (Hennig, 1966). Papers by Drs. Lichtenfels and Ferris in this volume describe the essence of the conventional and cladistic schools; the present paper describes some phenetic approaches. The various conceptual approaches are rather different, and sometimes competitive in their professed style and methods, yet their application tends to produce roughly equivalent classifications (e.g., Kluge, 1976). Differences in classifications have been reported, but it is still unclear whether these indicate differential efficiency in method or merely a predilection for one taxonomic philosophy over another. Relatively few published studies compare results obtained from the best techniques of all three approaches applied to the same taxon. Despite the scarcity of comparisons, however, the relative merits of these approaches have been debated extensively on theoretical grounds. The continued existence of such diversity in outlook is clearly an indication of health in systematic biology, but the bitterness of the current debate seems to me to be of questionable benefit.

A systematist faced with the task of analyzing a large data set has two basic options. One may wish to assess some aspect of similarity among the species or other taxonomic units; or one may wish to reconstruct some possible phylogenies. In the first task one applies the methods of phenetics, in the second the methods of cladistics. These tasks must be separated logically, but in practice they are not mutually exclusive; they usually involve some interaction and feedback. Similarity among individuals and taxa can be assessed without reference to common ancestry. For example, one could find some measure of similarity among three taxonomic units from three independently-derived solar systems, as long as there were some comparable charac-
teristics. One can use any size data base to obtain an assessment of similarity, although generally one strives to obtain as much data as possible for reduction by defined, operational techniques.

Reconstruction of ancestry cannot take place without an assessment of similarity, for one must first of all choose a taxon with recognizably similar members for analysis, then locate a nearest-relative (i.e., most-similar) taxon to guide decisions on character state polarity. Finally, one places species together on the basis of their similarity in derived character states. Thus phenetics and cladistics are not totally independent of one another. Further discussion of this point is given elsewhere (Moss and Hendrickson, 1973; Moss, 1978).

Similarity analyses, like cladistic analyses, depend on the taxonomist's ability to recognize spatially-corresponding points of reference (homologies). The task of recognizing these points is made much easier by the fact that the organisms being studied have a real evolutionary history. A basic but frequently unspoken tenet of similarity analysis in biology is that organisms are related in some way by ancestry and therefore can be compared logically to determine the extent of their similarity. Cladistic analysis goes a step further by saying that a measure of similarity in turn provides the key to a reconstruction of phylogeny. Both approaches can be used to classify and to recognize taxa. Both approaches seem to me to be valid avenues for dealing with the extensive data bases that challenge all systematists.

This use of the term "phenetic" as essentially equivalent with "similarity" sensu lato differs slightly from some recent usage. The term phenetic as initially introduced by Cain and Harrison (1958) referred to an estimate of similarity assessed over all possible characters, equally weighted. Later workers have recognized the impossibility of obtaining data from all characters, investigated the phenetics of character subsets, and realized that "all" characters does not mean morphological characters only. Further, the organization of similarities need not be restricted to the technique of cluster analysis, nor their visual representation to a phenogram. Such restrictions are straight-jackets on the development of phenetic methods. Different levels of phenetics derive from conscious or unconscious decisions to delete or weight characters differentially and to investigate different character sets (Moss, 1972). Studies on the accuracy of phenograms have shown that these diagrams are extremely poor vehicles by which to represent similarity; in fact, phenograms are generally useful only as rough guides to taxonomic structure. Much more accurate and useful results are obtained from ordination approaches based on multidimensional scaling (Rohlf, 1972; Moss et al., 1977). Dr. Freeman in his introduction to this symposium has noted the tendency for terms, concepts and definitions to evolve with time. A good example is the term phenetic, which has emerged and broadened from the restricted usage originally proposed. A phenetic analysis has come to mean a similarity analysis in the broad sense, i.e., a taxonomic analysis in which taxa are placed together and defined on the basis of some measure of similarity. Numerical and conventional phenetic procedures are basic components of biological classification.

MATERIALS

The nuances of philosophical niceties for dealing with data can be debated endlessly, but they acquire more immediate meaning when one has a specific taxon to analyze. My group of interest is the Harpyrhynchidae, a family of 114 mostly undescribed species of predominantly host-specific, ectoparasitic bird mites. Harpyrhynchids occur on members of 16 different orders of birds. The loss and hypertrophy of setae and other structures on these mites make them convenient material for the study of both similarity and ancestry. Their occurrence on a wide spectrum of hosts provides interesting data on patterns of host association and coevolution. Harpyrhynchids appear to exhibit both host and resource tracking (Kethley and Johnson, 1975; Moss, 1979). The mites are generally assumed monophyletic because of the presence of an enlarged palpal femur-genu that bears three modified setae, and other presumably homologous
correspondences in body structure and segmentation.

Immatures are known for only a few species of harpyrhynchids. The hexapod larvae of known species seem more similar to each other than adults are to each other; the nymphs are intermediate in this respect. The larvae of *H. brevis* Ewing and *H. novoplumaris* Moss et al. are very similar to each other and have the same general body plan as the adult of *H. brevis*, while the proto- and deutonymphal stages of the latter undergo a progressive reduction of the appendages and expansion of the body until the adult condition is reached (Mossedal., 1968). A numerical classification of the larvae would presumably show less distinctive clusters than those described below for the adult, but work to test this hypothesis is still in progress.

Preliminary studies of adult harpyrhynchids have demonstrated the existence of two morphologically distinct subfamilies, the Harpypalpinae and Harpyrhynchinae (Fain, 1972; Moss and Wojcik, 1978). Harpypalpines include only a dozen known species, with generalized morphology (complete leg segments, claws on all legs, and a more complete complement of dorsal and ventral setae). The body plan of these cyst-forming mites appears very much as one would expect in a cheyletid mite that has only recently adopted a parasitic existence. Harpypalpines are restricted to Passeriformes, generally considered to be the most recently-evolved taxon of birds. Harpyrhynchines, in contrast, comprise a much larger group of about 100 species. The harpyrhynchine body plan shows marked evidence of morphological reduction including loss of legs, and leg segments, loss of claws from the posterior legs, and a reduced complement of setae. Harpyrhynchines form cysts, attach to feathers, or move freely on the skin surface. Their overall appearance is that of parasites that have had a relatively long association with their diverse host groups, and in fact these mites occur on 16 orders of birds including passeriforms and those generally considered more ancient in origin. Components of the family have been studied by conventional means (Fain, 1972, and other references cited by Moss and Wojcik, 1978). The family is currently being analyzed for similarity among its component taxa, and will be analyzed cladistically when an adequate data base has been assembled and decisions reached on the validity of the taxa and character states now included.

**METHODS**

The essence of any systematic study is a table of data based on the features of the organisms. In a numerical phenetic analysis one may measure similarity in two ways from such a table. First, one considers each pair of species (or other taxonomic units) for congruence over the available suite of character values. There are many different measures of similarity, including coefficients of match/mismatch, correlation, and distance. There are reasons for preferring one coefficient over the other based on the kind of data available; Sneath and Sokal (1973) provide guidelines. For harpyrhynchid studies I have relied chiefly on the taxonomic distance coefficient. The immediate result of similarity assessment is a table of distance coefficients that indicates how similar (or in this case, how dissimilar) each pair of species is, based on the suite of characters considered. Clustering of these coefficients yields a phenogram that portrays similarities visually.

A second approach is to assess character relationships. In theory, each character should require a different axis to plot the readings for each species; this would require a multidimensional space for adequate representation (Sokal, 1965). In practice, however, one often finds that two or more different characters have very similar readings over the set of species considered; a set of such highly correlated characters can be replaced by a common axis. Other characters that show different patterns of correlation can be grouped into other common axes or factors, the end result being a reduction in the taxonomic space necessary to visualize the similarity between species and taxa. Ordination is the representation of species in a reduced character space defined by the major axes of variation. The usual pathway to ordination is via character correlation followed by
principal component analysis, in which the characters are replaced by major axes of variation, and the axes ranked according to the amount of variation that each explains (Rummel, 1970). Numerous other methods to reduce character dimensionality exist, but empirical studies have shown that principal component analysis seems to work best with biological data.

Phenograms and ordination diagrams have both strengths and weaknesses. A phenogram is constructed from the tips down and thus tends to acquire distortions that increase as clustering progresses. A phenogram shows close-relative similarities well, but not distant-order similarities unless the similarity coefficients are representable accurately in a single dimension. This condition is not often met for morphological data, but seems to hold for some biomolecular data. Thus phenograms as a general rule should be mistrusted as vehicles to represent similarity.

An ordination diagram based on principal component analysis has the arrangement of its axes determined by the relative magnitude of their explanation of character variation. Distances between taxa are greater than distances within taxa; one therefore finds that ordination diagrams tend to show distant-order similarities well, but not close-relative similarities (Rohlf, 1972). The impasse between phenograms and ordinations was broken in psychometric studies with the development of nonmetric multidimensional scaling (MDS) by Kruskal (1964) and others. This useful technique allows one to portray both close- and distant-order similarities well in a single diagram, and tends to achieve an accuracy of visual representation that is much improved over phenogram and principal component ordination. Even with MDS, however, a further weakness of an ordination diagram is that some species may appear closer together than they really are when one views them along a given axis of variation; this distortion effect, a function of the reduced character space, may be corrected by connecting the species with a minimum spanning tree based on their taxonomic distances (Sokal, 1974). One would expect to achieve additional insight by the insertion of hypothetical, ancestral OTUs through application of a Wagner network algorithm (T. Uzzell, personal communication). The capability of carrying out all the above operations, with the exception of the Wagner network, currently exists in the program system NT-SYS (Rohlf et al., 1972). Recent applications of these techniques to parasitic mites include papers by Moss et al. (1977), and by Moss and Wojcik (1978). Earlier papers on parasitic nematodes were provided by Moss and Webster (1969, 1970).

RESULTS

Similarities among taxa in the Harpyrhynchidae are illustrated below (Fig. 1). The mites in this study are representatives of the major taxa recognized by Fain (1972); one hundred characters were used. Figure 1 was constructed purely as an indicator of similarity; for example, the Harpyrhynchinae are separated in the diagram at a considerable distance from other species. Yet this figure lends itself to cladistic interpretation as well. The morphologically generalized harpypalpines are located to the right of Figure 1, the more reduced harpyrhynchines to the left (a detailed explanation of characters associated with the major axes is given by Moss and Wojcik, 1978). The connections of the minimum spanning tree indicate a successive loss of morphological structures as one proceeds from right to left. Members of the Harpyrhynchus agapornis species group are essentially harpypalpines with reduced body setation and shortened legs III and IV. The H. jacana group is made up of H. agapornis-like species that have lost leg IV. Harpyrhynchus nidulans is essentially an H. agapornis-like species with an elongated body. The members of the H. reductus group have lost legs III and IV, the dorsal shield and most of their setation. Connections with the H. squamiferus, H. monstruosus and H. porphyrio groups seem somewhat less desirable. The members of the H. squamiferus group are large compared to other members of the family. They attach to feathers as adult females and seem unlikely to have been derived from a member of the H. jacana group because members of the former have a fourth pair of legs that is
FIG. 1.—Ordination diagram of similarities among species and species groups in the acarine family Harpyrhynchidae. For interpretation of factor axes and further details on computation, see Moss and Wojcik (1978).

reduced and seems unlikely to have been gained after its loss in the H. jacana group. H. porphyrio likewise has four pairs of legs, although all of its legs are reduced and telescoped in their segmentation. Finally, members of the H. monstruosus group seem unlikely to have been derived from H. nidulans on morphological grounds. Thus the ordination diagram is attractive in some respect, but not entirely satisfactory in others. A preliminary Hennigian cladistic analysis based on the possession of derived character states yielded identical groups except for the placement together of the representatives of H. monstruosus, H. reductus and H. porphyrio. This is an unsatisfactory grouping, based essentially on reduction which has taken place in different ways in these cyst-forming mites. A re-assessment of existing characters and a search for additional characters seems in order for these groups.

The ordination diagram does not reflect the neatly nested hierarchy of harpyrhynchines recognized by Fain (1972) who proposed but did not adequately diagnose four genera and four subgenera. There are some correspondences and differences among our findings, but I prefer not to force the harpyrhynchines into a hierarchy at this time. Formal names are premature and would have a transient existence. A clearly-marked hierarchy does exist for the harppypalpines, however, as described below.

An analysis of the best-marked taxon, the genus Harppypalpus, provided some initial difficulties but eventually yielded worthwhile results. J. F. Wojcik and I first attempted to classify the twelve species of Harppypalpus by an intuitive approach. We studied the available material and summarized our impressions of similarity in a phenogram provided with an arbitrary scale of similarity, then compared distances read from this scale with taxonomic distances based on the data used to obtain Figure 1. Generally such sets of distances show a high positive correlation; however, the matrix correlation between these intuitive and numerical distances was essentially zero. A careful analysis showed that we had de-emphasized male: female differences in preparing our intuitive classification, and we had also recognized several species groups that were extremely difficult to
define. On the other hand, our ordination diagram did not indicate groups clearly, and had limited explanatory value. The program results were checked by hand and a variety of data combinations and character analyses tried with no further illumination until I returned to the specimens and eventually discovered several additional characters. Incorporation of these yielded an ordination diagram with two distinct species groups. These proved to be easily definable for both males and females and, in addition, two recently discovered new species fell easily into each group. Further details on this study and a key are provided in a revision of the genus Harpypalpus recently completed.

Thus neither a conventional nor a numerical phenetic analysis was initially successful for Harpypalpus. Uncritical acceptance of either approach would have yielded an unacceptable classification. Dissatisfaction with the results of both approaches and a subsequent re-analysis led eventually to a useful solution that was not anticipated initially. It should be noted that the genus Harpypalpus, unlike many other taxa, has no prior history of taxonomic analysis and opinion. In fact, even the number of included species was unclear at the beginning of the study.

Completion of the harpyrhynchid study will require a detailed re-analysis of most reduced harpyrhynchine groups, a reassessment of similarity over all taxa within the family and, finally, a cladistic analysis of the entire family.

SUMMARY

A systematist seeks estimates of similarity and ancestry among his taxa of interest. There are many ways to obtain such estimates and to represent them visually. Some approaches seem to work better on one taxon than on others; more applications are needed. A preliminary analysis of an entire taxon can be followed by a detailed analysis of constituent groups, and concluded by a re-analysis of the entire taxon using new characters discovered in the course of piece-wise analysis. Conventional studies often yield useful results, but we live in a multivariate world, and numerical techniques are helpful in dealing with complex patterns of variation. New techniques are constantly being developed and improved to meet the challenges of new data.

Similarity and ancestry are two fundamental components of biological relationship and classification. These components must be separated logically in a taxonomic analysis, but in practice they interdigitate and complement each other. Systematics will be the poorer for it if one component is prosecuted slavishly and totally in place of, in conflict with, or in ignorance of the other. Assessment of similarity without thought to the implications of ancestry is a logically defensible but biologically shallow operation. Ancestry inferred from restricted character sets without careful and controlled assessment of similarity based on the whole organism is also shallow and simplistic, because it can ignore or suppress relevant data.

There are no formula answers to the resolution of problems in systematic parasitology. Eclecticism in methods and scepticism toward results remain important elements of systematics research.

REFERENCES

Moss, W. W. 1979. Patterns of host-specificity and co-evolution in the Harpyrhynchidae (Acari). In