Comparison of dendrograms: a multivariate approach[1]

J. PODANI[2] AND T. A. DICKINSON[3]

Department of Plant Sciences, University of Western Ontario, London, Ont., Canada N6A 2R1

Received April 17, 1984

PODANI, J., and T. A. DICKINSON. 1984. Comparison of dendrograms: a multivariate approach. Can. J. Bot. **62**: 2765–2778.

Three new descriptors of dendrogram structure (cluster membership divergence, subtree membership divergence, and partition membership divergence) have been proposed which supplement existing ones (cophenetic difference and topological difference). These enable multivariate comparisons among dendrograms in a manner that minimizes the drawbacks of individual descriptors and emphasizes their agreement. The new descriptors are illustrated and compared with the existing ones by reference to an artificial data set. The multivariate comparison of dendrograms is further illustrated using two real data sets, one phenetic and one cladistic. These applications demonstrate the ability of this approach to reveal features of dendrograms that may be data dependent, method dependent, or due to the interaction of data and method, and that may not be readily apparent otherwise.

PODANI, J., et T. A. DICKINSON. 1984. Comparison of dendrograms: a multivariate approach. Can. J. Bot. **62**: 2765–2778.

Les auteurs présentent trois nouveaux descripteurs de la structure d'un dendrogramme: divergence dans l'appartenance à un groupe ("cluster membership divergence"), divergence dans l'appartenance à un sous-arbre ("subtree membership divergence") et divergence dans l'appartenance à une division ("partition membership divergence"). Ces descripteurs complètent ceux qui existent déjà (différence cophénétique, différence topologique). Ils permettent d'effectuer des comparaisons entre dendrogrammes de manière à minimiser les lacunes des descripteurs individuels et à accentuer leur concordance. Les nouveaux descripteurs sont illustrés et comparés aux anciens à l'aide d'un ensemble de données artificielles. La comparaison multivariée des dendrogrammes est aussi illustrée par deux ensembles de données réelles, l'un phénétique et l'autre cladistique. Ces applications montrent que notre approche est capable de révéler les caractéristiques des dendrogrammes qui peuvent découler des données, celles qui peuvent découler de la méthode ou encore qui peuvent être dues à l'interaction des données et de la méthode; ces caractéristiques ne seraient pas faciles à percevoir autrement.

[Traduit par le journal]

Introduction

In numerical taxonomic studies it is customary to represent relationships among the objects under study in the form of trees or dendrograms regardless of whether the approach taken is cladistic (relationships among objects are described in terms of putative ancestry) or phenetic (relationships among objects are described in terms of overall similarity). It is commonly observed that different methods may produce quite dissimilar results when applied to the same set of data (e.g., Sokal and Rohlf 1962; Hartigan 1975; Jardine and Sibson 1971). This statement is relevant not only to biological taxonomy but also to any field of science where numerical classification is practiced. However, in biological applications of numerical taxonomy many decisions concerning methods are critical without there being hard and fast guidelines to follow. Not only must resemblance (similarity and dissimilarity) coefficients and classification algorithms be chosen, but also other decisions must be made, such as: (*i*) selection of characters; (*ii*) selection of character weights and data transformations, if any; and (*iii*) selection of a method of character coding. Moreover, in cladistic analyses the results may be affected by the sequence in which the objects are analyzed (Colless 1983). The complexity of this situation may be appreciated by examining the left portion of Fig. 1. For example, dendrograms may be obtained along the path $a-b-c$. If we consider that each arrow in the scheme is associated with a multitude of choices to be made, the number of possible dendrograms for the same set of objects is too large to comprehend. Consequently, it is very important to examine how results are influenced by these choices and which stages of the analysis are most critical for the consistency and interpretability of the classifications obtained. Studies of dendrogram congruence of this kind would supplement the optimality tests recently reviewed by Rohlf and Sokal (1981).

The effect upon dendrogram structure of any change in either algorithm or data may be evaluated by the comparison of results. The comparison of dendrograms (D–D comparisons, in terms of Fig. 1) has been employed in numerical taxonomy (Rohlf 1974; Rohlf and Sokal 1981; Lachance and Starmer 1982; Hopper and Burgman 1983) but not routinely in botanical studies, according to Duncan and Baum (1981). In the most widely adopted approach each dendrogram is described in terms of an $n \times n$ matrix summarizing the pairwise relationships among the n objects studied in terms of some descriptor of dendrogram structure. Two dendrograms are then compared by various matrix correlation methods (Duncan *et al.* 1980; Rohlf and Sokal 1981). Note that in this paper attributes of objects will be designated by reference to characters, whereas the attributes of dendrograms that describe relationships among objects will be designated by reference to descriptors. The distinction between the terms character and descriptor is made here solely for the sake of clarity.

Dendrogram similarity may also be expressed without comparing descriptor matrices. Examples of methods of direct dendrogram comparison are the ultrametric dissimilarity coefficient (Dobson 1975; the terminology is ours), the cluster distortion technique (Farris 1973), versions of an edge-

[1]This paper is based on portions of theses by the authors, submitted in partial fulfillment of the requirements for the Ph.D. degree at the University of Western Ontario. The underlying principles of the method introduced and the *Crataegus* example were presented at the 16th International Numerical Taxonomy Conference at the University of Notre Dame, Notre Dame, IN, U.S.A., 22–24 October 1982 (see Jensen 1983).

[2]Permanent address: Research Institute for Botany, Hungarian Academy of Sciences, Vácrátót, Hungary H-2163.

[3]Present address: School of Forestry, Lakehead University, Thunder Bay, Ont., Canada P7B 5E1.

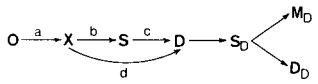


FIG. 1. Scheme illustrating the pathways of numerical taxonomic studies, supplemented by the multivariate comparison of dendrograms (expanded from schemata described by Rohlf and Sokal 1981 and Duncan and Baum 1981). The set of objects (O) is described by a raw data matrix (**X**) from which the resemblance matrix for objects (**S**) is obtained. D denotes dendrograms derived either from the resemblance matrix or directly from the data. Comparison of several dendrograms leads to their resemblance matrix (S_D) and then to an ordination (M_D) or clustering (D_D). Lower case letters indicate pathways along which the investigator is faced with multiple choices in a conventional numerical taxonomic study.

matching coefficient (Robinson and Foulds 1979; Podani 1982), and the number of nearest-neighbor interchanges (Robinson 1971; Waterman and Smith 1978; Jarvis et al. 1983).

Finally, dendrogram comparisons are implicit when consensus trees (Adams 1972) are used to summarize the information contained in alternative dendrograms (Seaman and Funk 1983; Smith and Phipps 1984). Related approaches for explicit comparisons may employ consensus indices (Mickevich 1978; Rohlf 1982). Such methods are obviously inappropriate for study of the problems addressed above, since they are based on determining the departure from a standard of the consensus tree for a set of dendrograms obtained using a particular method or set of data.

It must be noted that the comparison of the resemblance matrices from which the dendrograms were derived (**S**−**S** comparisons, Fig. 1) may appear to represent a simpler method of analysis, e.g., of the effect of character selection alternatives. However, D−D comparisons are more widely applicable, for several reasons. (*i*) Except for special purposes (e.g., Gilmartin 1974, 1980), taxonomists are in fact interested in the classifications derived from dendrograms more than they are interested in resemblance structures. (*ii*) Dendrograms may be generated without calculating resemblance matrices (**X** → D, Fig. 1; character compatibility methods (Duncan *et al.* 1980) as well as monothetic divisive clustering methods (Orlóci 1978)). (*iii*) The correlation between resemblance matrices is not always meaningful, as in the case of that between a correlation matrix and a distance matrix. (*iv*) The effect of sorting algorithm can only be analyzed by means of D−D comparisons. (*v*) It is not known at the outset whether minor changes in the resemblance structure can cause drastic changes in hierarchical relationships, or whether substantial perturbations of the resemblance structure may nevertheless lead to similar hierarchies, if the group structure is sharp. Consequently, **S**−**S** comparisons may produce misleading results with respect to the relationships among dendrograms and hence, among classifications.

Nevertheless, as shown by Douglas and Endler (1982) and by Dietz (1983), **S**−**S** comparisons may still prove useful in taxonomic and evolutionary studies, but further discussion of these methods is beyond the scope of this paper. So too is a discussion of methods concerned with comparisons of taxonomic classifications, as distinct from the dendrograms from which they may be derived, with respect to their predictiveness (Hawksworth *et al.* 1968; Duncan and Baum 1981).

All methods of D−D comparisons currently available suffer from the drawback that only one aspect of dendrogram resemblance is taken into account (this insufficiency is discussed at length below). This fact is all the more surprising because all other stages of a numerical taxonomic study are multivariate in nature. The present paper is an attempt, therefore, to meet a long-standing need by providing a means for dendrogram comparisons such that several features of dendrogram structure are incorporated into a single resemblance function.

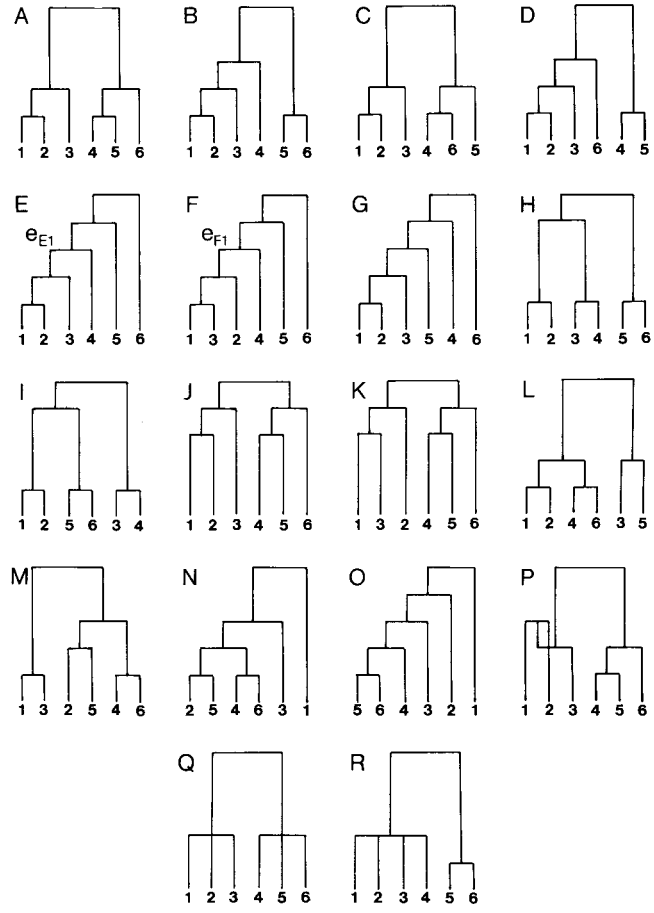


FIG. 2. Artificial dendrograms illustrating different classifications of six objects.

Materials

In addition to analyses of an artificial data set we illustrate the method for multivariate comparison of dendrograms described below by means of analyses of two real data sets. The latter analyses employed a number of computer programs beside the one for calculating dendrogram descriptors described in the Appendix. These included programs EUCD and PCAR, given by Orlóci (1978) and the first author's programs for cluster analysis and principal coordinates analysis, NCLAS and PRINCOOR, respectively (Podani 1980, 1984). All of these ran on the DEC system 1090 installation of the University of Western Ontario Computing Centre.

Artificial dendrograms

Eighteen different dendrograms for six objects were constructed (Fig. 2). For the sake of simplicity only six hierarchical levels were distinguished. Dendrograms A−P are binary trees, while Q and R are nonbinary ones. Dendrogram P differs from all of the others in that it contains a reversal. Only dendrograms A−O are included in the analysis below; the other three merely illustrate certain aspects of the terminology used in presenting the method.

Most of the artificial classifications were generated by elementary transformations from arbitrarily constructed ones. Dendrogram A was transformed into B, C, D, J, and K by interchanges of objects or by shifting hierarchical levels. Dendrograms A and J have similar topo-

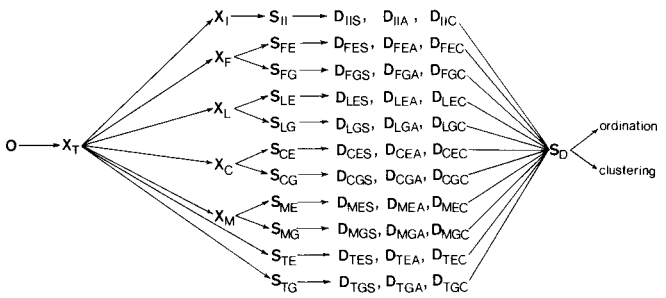


FIG. 3. Flowchart illustrating the experimental design used in the analysis of *Crataegus* phenograms. The meaning of symbols O, **X**, **S**, and D is the same as in Fig. 1. The first subscript indicates the type of data, as described in Table 1. The second subscript stands for the type of resemblance function used: I, information radius; E, Euclidean distance; G, generalized distance. The third subscript indicates the type of classification strategy: S, single linkage; A, average linkage; and C, complete linkage. $\mathbf{S}_D$ represents the matrix of Euclidean distances among phenograms calculated using the dendrogram descriptors described in the text.

logies but different hierarchical levels. Dendrogram C was obtained from A by nearest-neighbor interchange at the lowest level. Dendrogram A was transformed to D by the relocation of object 6. Dendrogram B was obtained by removing object 1 from the "seed" of a subtree in A. Another group of dendrograms (E, F, G, and O) demonstrates a typical series of fusions called chaining. Dendrogram F differs from E in a nearest-neighbor interchange at a low level, while G was obtained from E by interchanging two objects at a higher level. The fusion sequence of objects in O is just the opposite of that in dendrogram E (chain inversion). Dendrograms H and I imply the same three-cluster classification, while M is as different from H and I at this level as possible. Dendrogram N was obtained from M by destroying one subtree in M. Dendrogram L represents a unique classification, apparently dissimilar to all others.

Six matrices of Euclidean distances were calculated among dendrograms A–O. The first five are based each on one of the five dendrogram descriptors described below, while in the sixth all five are incorporated simultaneously, using function 9 given below. Each distance matrix was subjected to principal coordinates analysis (PCoA) and minimum variance clustering (Ward 1963). Other ordination and clustering methods might equally be used; those employed in the examples here were selected because they are commonly used and their properties are well-known.

Dendrograms of phenetic relationships

The first of the two real data sets comes from a study of variation in *Crataegus crus-galli* L. *sensu lato* in Ontario (Dickinson 1983; Dickinson and Phipps 1984, 1985). The objective of the portion of the study described here was to determine the range of taxonomic structures (classifications) supported by the data available. It was of interest to see to what extent particular classifications were the outcome of particular combinations of method and character suite. Sixty hawthorn trees sampled randomly at five sites in southern Ontario were used as operational taxonomic units (OTUs). This sample comprised four taxa: three making up *Crataegus crus-galli* L. *sensu lato* (*Crataegus* Sect. *Crus-galli* : *C. crus-galli s. str.*, $N = 32$; *C. fontanesiana* (Spach) Steud., $N = 18$; *C. ?grandis* Ashe, $N = 2$) plus *C. punctata* Jacq. (*Crataegus* Sect. *Punctatae*; $N = 8$). Further details concerning sampling sites and methods, and the composition of the sample studied here, are given in Dickinson (1983) and Dickinson and Phipps (1984, 1985).

Each *Crataegus* OTU was scored for a total of 17 characters (Dickinson 1983; Table 2 in Dickinson and Phipps 1985). For each OTU, reproductive characters were scored on 10–20 flowers and fruits, while leaf characters were scored on six to eight short shoot terminal leaves. Using the sample maxima and minima for each character, OTU means were rescaled to a {0,1} interval to ensure commensurability (condensation, Crovello 1968; ranging, Gower 1971*a*). These means made up the basic data matrix used in subsequent analyses.

TABLE 1. Key to the data sets ($\mathbf{X}_\cdot$) used in creating the 33 phenograms of 60 *Crataegus* OTUs represented in Fig. 7

	Multistate	Continuous	
Flowers, fruit	PROJ	WFL	
	ANTH	LCAL	
	TCAL	LFR	$\mathbf{X}_F$
	PUB1	WFR	
	PUB2		
Leaves		*X*	
	NUMSEC	*Y*	
	ANGSEC	*Z*	$\mathbf{X}_L$
		TEETH	
	$\mathbf{X}_M$	$\mathbf{X}_C$	All[a] X_T

NOTE: Entries are characters, sorted by organ system and scale of measurement. See Dickinson (1983) and Dickinson and Phipps (1985) for complete explanation of the abbreviations and corresponding characters. $\mathbf{X}_I$ = observed character–state frequency distributions for the characters in $\mathbf{X}_F$, plus STAM and STYL.

[a] Including descriptors STAM and STYL.

The experimental design of the study is shown in Fig. 3. In addition to the complete basic data matrix, $\mathbf{X}_T$, four other data sets were assembled, each representing a subset of $\mathbf{X}_T$. The characters were divided according to whether they related to flowers and fruits or to leaves (Table 1). This division yielded two data matrices denoted $\mathbf{X}_F$ (flower and fruit characters) and $\mathbf{X}_L$ (leaf characters). Another classification was made based on whether characters were scored as continuous or meristic variables, or as ordered multistate ones (Table 1). The corresponding data matrices are denoted $\mathbf{X}_C$ and $\mathbf{X}_M$, respectively. Note that two flower characters, style number per flower (STYL) and stamen number per flower (STAM) were omitted from $\mathbf{X}_F$ and $\mathbf{X}_C$ because of the way in which by themselves these two characters sufficed to assign all but the two OTUs of *C. ?grandis* to the correct taxon. The sixth data matrix, $\mathbf{X}_I$, differs from the preceding ones as it contains observed character-state frequency distributions for all 11 flower and fruit characters (see Dickinson (1983) and Dickinson and Phipps (1985) for details).

For each of $\mathbf{X}_T$, $\mathbf{X}_F$, $\mathbf{X}_L$, $\mathbf{X}_C$, and $\mathbf{X}_M$ two distance matrices were calculated: a matrix of Euclidean distances ($\mathbf{S}_{\cdot E}$) and one of Mahalanobis' generalized distances ($\mathbf{S}_{\cdot G}$). Both matrices were produced using program EUCD. However, in the latter case a R algorithm, principal components analysis was performed first, using the character covariance matrix. Between-OTU generalized distances were calculated from principal component scores for each OTU instead of from the OTU character means. In doing so, the differences between OTUs on each principal component were divided by the corresponding eigenvalue of the covariance matrix (Rohlf 1970; Orlóci 1978). An eleventh resemblance matrix, of information radii among OTUs ($\mathbf{S}_{II}$), was calculated using $\mathbf{X}_I$ and the formulation given by Prentice (1979).

A total of 33 phenograms were obtained from these 11 resemblance matrices using three sorting algorithms with well-known characteristics (Sneath and Sokal 1973): single linkage (space-contracting), average linkage (space-neutral), and complete linkage (space-dilating). Values of the five dendrogram descriptors were calculated for each phenogram as described below, by program DENDAT (see Appendix). These values were used to calculate a matrix of Euclidean distances among the 33 phenograms (expression 9, below). Relationships among the phenograms were then examined by means of minimum variance clustering and PCoA.

Dendrograms of phylogenetic relationships

Recently, Penny *et al.* (1982) published 39 dendrograms of

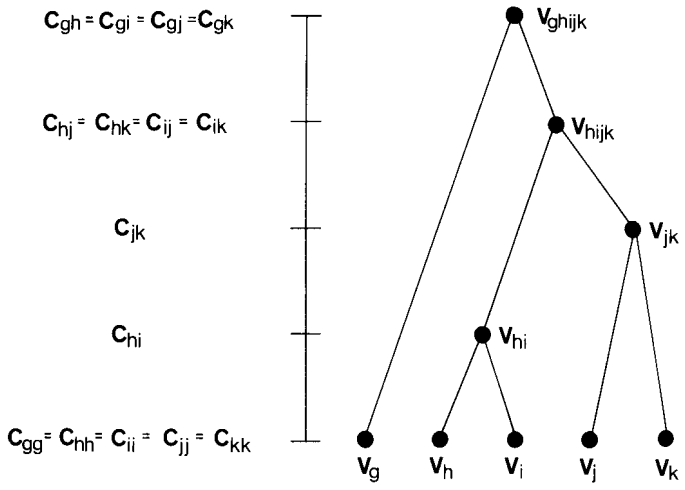


FIG. 4. Cophenetic levels (C) and vertices (V) in a dendrogram.

11 mammalian taxa, based on evolutionary relationships deduced from the amino acid sequences of five proteins considered singly and all together. We have used their cladograms to illustrate the usefulness of the method described below in detecting relationships between taxonomic structure and data source in a sample of phylogenetic trees.

Since the cladograms given by Penny *et al.* (1982) illustrate only topological relationships among the 11 taxa, cophenetic differences among them could not be calculated. However, values of the remaining four descriptors were obtained for each cladogram (program DENDAT) and used to calculate a matrix of Euclidean distances (expression 9, below) that was the basis for a minimum variance clustering and PCoA ordination of the cladograms. Although Penny *et al.* (1982) published a total of 39 cladograms, two of them (numbers 8 and 38) are identical. The analyses described here were carried out using only the 38 different cladograms.

Method

Dendrograms can be treated as individual objects. Consequently, dissimilarities among them can be defined and a set of dendrograms may be subjected to ordination and clustering in much the same way as any other group of entities (see e.g., Schnell 1970; Phipps 1971). Dendrograms, however, are objects whose adequate description requires knowledge of basic graph theoretical terms. Accordingly a brief summary of the relevant terminology is given prior to discussion of the methodological details.

Definitions

Let $S = \{s_i\}$ be a set of $j = 1 \ldots n$ objects to be subjected to hierarchical classification. The resulting dendrogram is a connected, undirected graph in which each terminal vertex v_i represents object s_i (Fig. 4). The terminal vertices are connected by edges through interior vertices. The degree of a vertex is the number of edges belonging to it, so that terminal vertices are of degree one.

A dendrogram is an unrooted binary tree if all interior vertices have degree three. Examples are some evolutionary trees (e.g., Waterman and Smith 1978). Numerical taxonomic studies are frequently concerned with the construction of rooted trees, however. The root is usually a degree two vertex corresponding to the last fusion (or first division) point in the clustering process. A rooted binary tree or dendrogram has $n - 1$ interior vertices and $2n - 2$ edges. The number of interior edges is $n - 2$. In many cases the degree of interior vertices and the root is higher than three or two, respectively. Such nonbinary dendrograms are obtained if the fusion of more than two clusters at one time is allowed, or in the case of divisive algorithms, if clusters are allowed to split simultaneously into more than two groups. Additionally, a nonbinary tree will result if a divisive clustering process is arrested at a specified termination point. Examples of nonbinary trees are dendrograms Q and R in Fig. 2.

There is a characteristic number associated with each interior vertex identifying the hierarchical level at which the corresponding groups were fused or divided. The lowest hierarchical level, that of the terminal vertices, is the similarity of objects to themselves. Based on these hierarchical levels the tree structure of a dendrogram may be completely described by an $n \times n$ matrix (cf. Hartigan 1967) here designated **C**. An element of this matrix, c_{jk}, represents the hierarchical level assigned to the interior vertex (v_{jk} in Fig. 2) closest to the root along the path from v_j to v_k. This value measures the cophenetic difference (Sokal and Rohlf 1962), designated CD. In most cases c_{jk} satisfies the criteria for being an ultrametric (Jardine and Sibson 1968). That is, for any i, j, and k ($i \neq j \neq k$, $i = 1 \ldots n$; $j = 1 \ldots n$; $k = 1 \ldots n$) the following relation holds:

[1] $c_{jk} \leq \max\{c_{ij}, c_{ik}\}$

Using other symbols, any triplet of vertices can be labelled p, q, and r so that

[2] $c_{pq} \leq c_{pr} = c_{qr}$

These relations may not hold true for all triplets if, for instance, centroid sorting algorithms are used and reversals occur in the dendrogram, as illustrated by dendrogram P in Fig. 2.

The removal of an edge e from a dendrogram G_1 creates two subtrees, G_1' and G_1''. This operation divides the set of objects into two subsets and represents the partition $P_1 = \{S_1', S_1''\}$ of S_1. Let G_1 and G_2 denote two dendrograms to be compared, and further, let e_{11} and e_{21} be, respectively, edges of G_1 and G_2. e_{11} and e_{21} are said to be matched if the partitions P_1 and P_2 generated by their deletion are equivalent. An example is shown in Fig. 2, where edges e_{E_1} and e_{F_1} are matched. In this respect interior edges have specific importance since the terminal edges are necessarily matched for any G on S.

Finally, the minimum difference between two dendrograms is defined in terms of tree topology. If G_1 is a binary tree, the minimum step necessary to obtain a topologically different tree G_2 is the change of the positions of two vertices that are two interior vertices apart. For instance, the change of v_5 and v_6 in dendrogram A will result in dendrogram C (Fig. 2). This transformation is termed nearest-neighbor interchange (Waterman and Smith 1978) or crossover (Robinson 1971).

Descriptor variables

The most common approach to dendrogram comparisons is to define a descriptor variable expressing the relative position of two objects in the hierarchy. These values are written in a $n \times n$ symmetric matrix, hereafter called a descriptor matrix. Two dendrograms are then compared using any meaningful function based on the respective elements of the corresponding matrices.

The first definition for descriptor variables was the concept of the cophenetic difference (CD) described above. The product–moment correlation between $\mathbf{C}_1$ and $\mathbf{C}_2$ was then calculated to measure linear relationship between G_1 and G_2. Other coefficients such as rank correlations and stress functions are also used (Jackson 1969; Cunningham and Ogilvie 1972). One advantage of this approach is that the dendrograms can be compared with the original resemblance matrices, and the distortion implied by the dendrogram can be measured. However, the method has limited applicability since comparison of the **C** matrices to the original resemblance matrices is often meaningless, if not impossible (e.g., in the case of minimum variance clustering). Moreover, since only hierarchical levels are considered, dendrograms with different topologies may prove as similar or more so than two which imply the same hierarchy. For instance, consider dendrograms A, C, and J (Fig. 2): $r_{AC} = 0.976$ but $r_{AJ} = 0.925$, although dendrograms A and J differ only in hierarchical levels. If a distance coefficient is used the difference may be even greater.

Another descriptor was suggested independently by several authors (Farris 1969; Phipps 1971; Williams and Clifford 1971). Instead of considering levels, the vertices (or edges, whose number will be one greater) are counted along the paths between all possible pairs of terminal vertices. This number is called the cladistic difference (Farris

1969) or topological distance (Phipps 1971) and is denoted by t_{jk} for v_j and v_k. Since the use of the term cladistic is restricted to cladograms and the use of the term distance may cause confusion with intervertex and interdendrogram distances, t_{jk} will be called the topological difference (TD). The descriptor matrix **T** for the dendrogram of Fig. 4 is given by

$$[3]\quad \mathbf{T} = \begin{array}{c|ccccc} & g & h & i & j & k \\ g & 0 & 3 & 3 & 3 & 3 \\ h & & 0 & 1 & 3 & 3 \\ i & & & 0 & 3 & 3 \\ j & & & & 0 & 1 \\ k & & & & & 0 \end{array}$$

The topological difference is not an ultrametric, and there is not a one-to-one relationship between TD and the hierarchic classification (Rohlf and Sokal 1981). Two matrices may be compared element by element using an appropriate resemblance function. The Manhattan metric, Euclidean distance, or the correlation coefficient are most often used. It is clear that a low topological difference does not necessarily imply high similarity, and vice-versa. Another potentially undesirable property of topological difference is that a nearest-neighbor interchange affects $2n - 4$ values in the descriptor matrix. Comparisons employing TD cannot be made when the number of interior vertices in G_1 and G_2 is not the same (i.e., when one or both are nonbinary).

Three new, alternative descriptor variables will now be introduced. Rohlf (1974) pointed out that one drawback of TD is that the location of the root has little effect on dendrogram resemblance. A more fundamental problem with this descriptor is that it gives equal weight to interior vertices. Clearly, the root should be the most weighted vertex since it represents the highest hierarchical level. Vertices close to the root should be given less weight, and so on. Direct weighting using CDs would result in an improved descriptor, but one which would largely ignore topological relationships. However, the tree structure will be exactly preserved if each vertex is weighted according to the number of objects that are fused into the same cluster at the corresponding level and instead of counting all vertices along the path between v_j and v_k, only the vertex with the largest weight is considered. Therefore, if **W** denotes the descriptor matrix to be defined, w_{jk} will simply equal the number of objects in the smallest cluster which contains both s_j and s_k. For the dendrogram in Fig. 4 matrix **W** will be given by

$$[4]\quad \mathbf{W} = \begin{array}{c|ccccc} & g & h & i & j & k \\ g & 1 & 5 & 5 & 5 & 5 \\ h & & 1 & 2 & 4 & 4 \\ i & & & 1 & 4 & 4 \\ j & & & & 1 & 2 \\ k & & & & & 1 \end{array}$$

The advantage of this strategy over TD is clear from the following example. If we consider objects s_g, s_i and s_k in the dendrogram in Fig. 4 (vertices v_g, v_i, and v_k), the corresponding measures are $t_{gi} = 3$, $t_{gk} = 3$, $t_{ik} = 3$, $w_{gi} = 5$, $w_{gk} = 5$, and $w_{ik} = 4$. That is, this new descriptor takes into account the fact that $c_{ik} < c_{gi}$ and $c_{ik} < c_{gk}$, whereas TD does not. The "difference" between s_i and s_k is less than w_{gi} and w_{gk}, as required. The hierarchical levels are not preserved because $w_{hi} = w_{jk}$ but $c_{hi} < c_{jk}$. This example suggests that the new descriptor is intermediate in its behavior between CD and TD. This is not to say that w_{jk} is a better descriptor than the others in all circumstances, but it does have some properties that may prove advantageous. These include the following: (*a*) $w_{jk} - 1$ is ultrametric under all conditions even if reversals are present and **W** has an exact tree structure; (*b*) a nearest-neighbor interchange affects only $2m - 4$ values in **W**, where m is the weight of the more weighted vertex of the interchange in question; and (*c*) it follows from (*b*) that a nearest-neighbor interchange near the terminal vertices has less effect on dendrogram similarity than does a change close to the root. The descriptor w_{jk} will be called the *cluster membership divergence* (CMD).

The next descriptor suggested utilizes the property that any dendrogram implies a series of partitions. For example, the dendrogram in Fig. 4 represents the following sequence of nonhierarchical classifications:

$$\begin{array}{c} \{v_g\},\{v_h\},\{v_i\},\{v_j\},\{v_k\} \\ \{v_g\},\{v_h,v_i\},\{v_j\},\{v_k\} \\ \{v_g\},\{v_h,v_i\},\{v_j,v_k\} \\ \{v_g\},\{v_h,v_i,v_j,v_k\} \\ \{v_g,v_h,v_i,v_j,v_k\} \end{array}$$

Excluding the trivial case of one-group classification, but keeping the division into single objects, it is seen that the number of different partitions is $n - 1$. However, this value is only a theoretical maximum not reached by trees in which two or more hierarchical levels are equivalent (as in dendrograms A,B,C,D,H,I,J,K,L,M, and N in Fig. 2). The relative position of two objects in the hierarchy can be expressed in terms of the number of partitions in which they are not assigned together to a group. This value will be termed *partition membership divergence* (PMD) and denoted by b_{jk} for v_j and v_k. The descriptor matrix **B** for the dendrogram in Fig. 4 will be

$$[5]\quad \mathbf{B} = \begin{array}{c|ccccc} & g & h & i & j & k \\ g & 0 & 4 & 4 & 4 & 4 \\ h & & 0 & 1 & 3 & 3 \\ i & & & 0 & 3 & 3 \\ j & & & & 0 & 2 \\ k & & & & & 0 \end{array}$$

Comparison with matrix **W** [4] immediately shows that in this example there is only one substantial difference between the two descriptors, namely $b_{hi} < b_{jk}$, whereas $w_{hi} = w_{jk}$. That is, PMD preserves the ordering of hierarchical levels although it is a completely topological descriptor. PMD and CMD are similar in that a nearest-neighbor interchange affects only $2m - 4$ values in the matrix (see above). PMD has the further advantage that the magnitude of the change of these values depends on the number of hierarchical levels falling between the two levels on which the interchange took place. In this way the hierarchical structure is more explicit in PMD than in CMD. However, b_{jk} is ultrametric only if there are no reversals in the dendrogram. A more serious difficulty is that the comparison of G_1 and G_2 is meaningless if the number of hierarchical levels, p_1 and p_2, is different in G_1 and G_2. A possible remedy is to consider only the first p' most important levels or $\min\{p_1,p_2\}$ upper levels in both dendrograms being compared, and ignore the others. Fortunately, for quantitative data the chance for two levels to be equal is negligible, although in the case of binary data that chance is rather higher.

The third descriptor to be introduced characterizes dendrograms on the basis of their interior graph structure. In a binary tree there are $n - 1$ subtrees (including the whole dendrogram) which correspond to the $n - 1$ interior vertices. Each of these subtrees represents a hierarchical classification of a subset of objects. The position of two objects in the hierarchy relative to the others can be described as the number of these subclassifications in which they do not occur together. This quantity will be called *subtree membership divergence* (SMD) and denoted by l_{jk} for v_j and v_k. Matrix **L** for the dendrogram in Fig. 4 clarifies the meaning of this descriptor.

$$[6]\quad \mathbf{L} = \begin{array}{c|ccccc} & g & h & i & j & k \\ g & 3 & 3 & 3 & 3 & 3 \\ h & & 1 & 1 & 2 & 2 \\ i & & & 1 & 2 & 2 \\ j & & & & 1 & 1 \\ k & & & & & 1 \end{array}$$

As seen, SMD differs fundamentally from the other four descriptors in that the values of the main diagonal in **L** are not necessarily equal. This is due to the fact that the terminal vertices are not equidistant from the root. If t^*_{jr} denotes the number of edges between v_j and the root, l_{jj} will be given by

$$[7]\quad l_{jj} = n - t^*_{jr} - 1$$

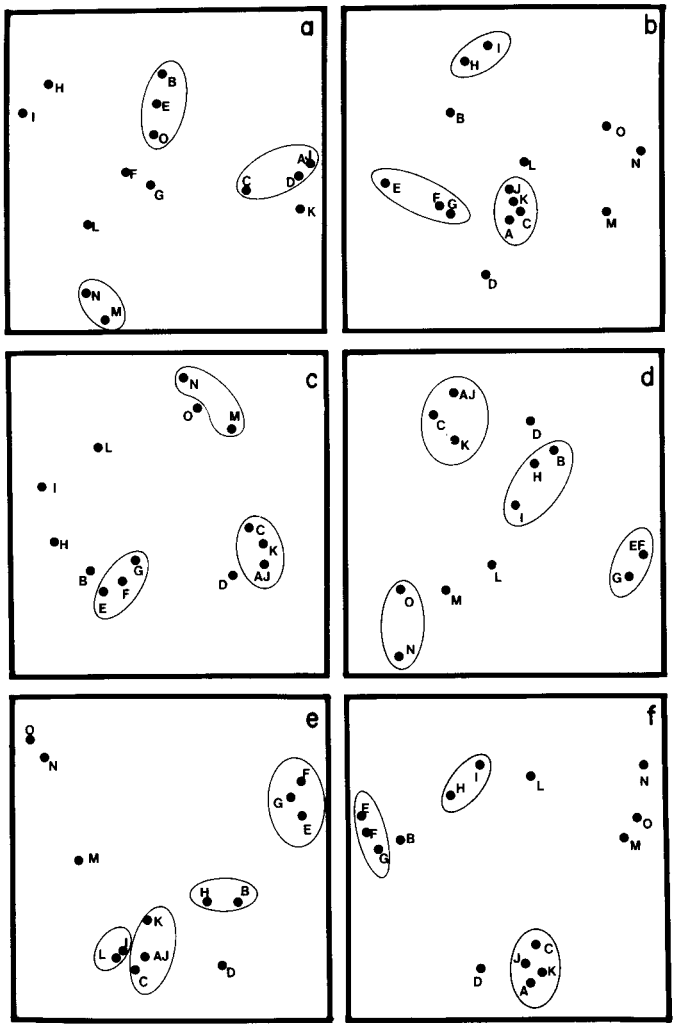


FIG. 5. Artificial dendrograms A–O of Fig. 2 in the space defined by the first two principal coordinate axes derived from matrices of dendrogram descriptors: (*a*) topological difference; (*b*) cophenetic difference; (*c*) cluster membership divergence; (*d*) subtree membership divergence; (*e*) partition membership divergence; and (*f*) Euclidean distances calculated from all five descriptors. Clusters existing in the first three dimensions are outlined. In Figs. 5*a*–5*f*, the eigenvalues corresponding to the first three axes account for at least 67% of the sum of all 15 eigenvalues.

That is, t^*_{jjr} influences the actual value of the diagonal element. This property may or may not be considered in defining a resemblance between two trees. It may also be seen that the subtree membership divergence between v_j and v_k is

$$[8] \quad l_{jk} = n - t^*_{jkr} - 2$$

where t^*_{jkr} is the number of edges between the interior vertex closest to the root along the path from v_j and v_k and the root itself. Owing to the inequality of the diagonal elements of **L**, SMD is not ultrametric, although all nondiagonal elements would satisfy the requirements even if reversals occur. For this reason, the complete tree structure is easily recovered from the **L** matrix. The use of Euclidean distance is suggested to compare G_1 and G_2, provided that both dendrograms are binary. For nonbinary trees this descriptor becomes less efficient as the number of high-degree (>3) interior vertices increases.

Multiple comparisons based on the combined distance between dendrograms

The comparison of two dendrograms using the corresponding descriptor matrices is univariate even though $n^2 - n$ paired elements are involved in the calculation of dendrogram resemblance. A cophenetic correlation coefficient reflects similarity only in terms of CD, a distance calculated from TDs is affected only by specific topological relationships, and so forth. Consequently, the analysis (scaling or clustering) of a $k \times k$ matrix of resemblance coefficients among k dendrograms will also be univariate in this sense. An analytical method may be sought which incorporates as many aspects of dendrogram description as possible, so as to make the comparison really multivariate, and hence more reliable. This goal may be accomplished by involving all five descriptor variables in the definition of a distance among dendrograms.

To avoid excessive differences in the importance of descriptor variables, some manipulations of the data are necessary. CDs are rescaled for each dendrogram to fall within the interval {0,1}. The CMD values are readily standardized if divided by n. The minimum number of partitions, p_{min}, in the set of dendrograms being analyzed is found. Then, only the upper p_{min} partitions are considered for each dendrogram in calculating the PMD. Division by p_{min} results in a {0,1} range for this descriptor. SMD and TD are standardized by the global maxima. The maximum for the SMD is determined by counting all terminal vertex–root topological differences, finding the minimum and subtracting it from n. It may be shown that the maximum of the topological differences pertains to the terminal vertex whose SMD is the minimum. This fact considerably simplifies the computations.

Having standardized the descriptor variables, the combined Euclidean distance between dendrograms G_1 and G_2 is computed according to

$$[9] \quad d^2_{12} = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \sum_{a=1}^{5} (x(a)_{1ij} - x(a)_{2ij})^2$$

where each state of $x(a)$ represents a descriptor. The distance matrix **D** of k dendrograms ($\mathbf{S}_D$, in terms of Fig. 1) is subjected to ordination or clustering. The results of these analyses may yield information useful in interpreting the dendrograms, whether or not discontinuities exist among alternative dendrograms.

Results and discussion

Artificial dendrograms

The corresponding ordinations (Fig. 5) and classifications (Fig. 6) of the 15 artificial dendrograms are in good agreement with each other, although in a few cases ordination groups outlined in the scattergrams have been changed in the classification. The most striking example of this is group {A,C,D,J} of Fig. 5*a* (TD), from which object C has been removed and placed in a distinct cluster in Fig. 6*a*. In the remaining comparisons such drastic differences are not observed.

As expected, there are both similarities and substantial differences among the performances of the five dendrogram descriptors. These may be summarized as follows. (*i*) The most consistently recognized group is {A,C,J,K}, with the exception of TD (Figs. 5*a*, and 6*a*) as mentioned above. This suggests that nearest-neighbor interchanges affect all five descriptors considered simultaneously and all but one (TD) of the individual descriptors in a similar way. (*ii*) The uniqueness of TD is also obvious if another aspect of dendrogram similarity is considered. The relative closeness of dendrograms E and O in ordination (Fig. 5*a*) and classification (Fig. 6*a*) demonstrates that TD is insensitive to chain inversions, unlike the four other descriptors. (*iii*) SMD also produced some obviously unacceptable groups. Dendrograms I and L, although quite dissimilar, have been assigned together in Figs. 5*e* and 6*e*. Also, H and I, which differ at the two-group level but imply similar three-cluster classifications, fall far apart (Figs. 5*e* and 6*e*). (*iv*) The position of A and J in Figs. 5*b* and 6*b* demonstrates the problem of hierarchic levels in cophenetic comparisons. Although A and J imply the same hierarchy, these are not the most similar

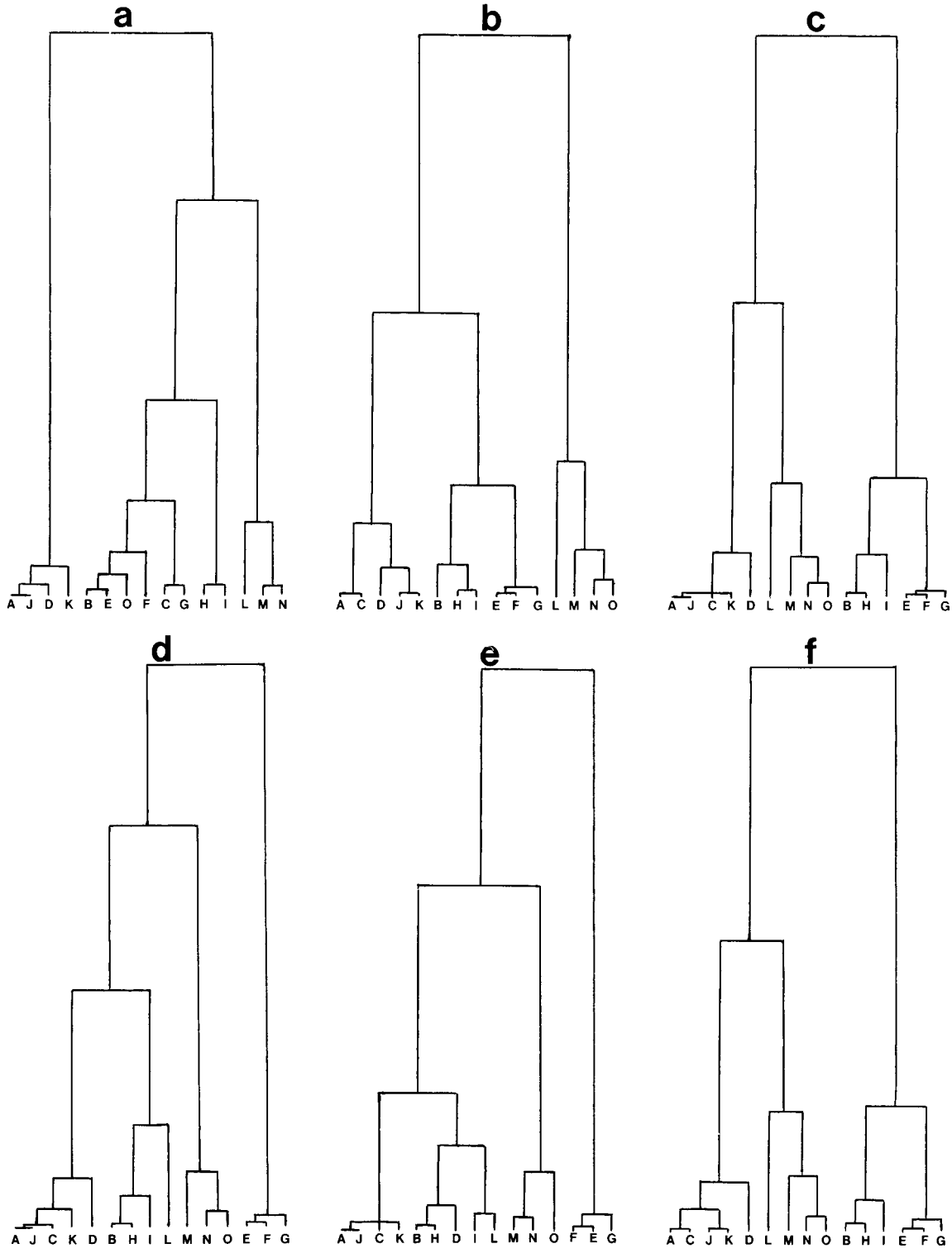
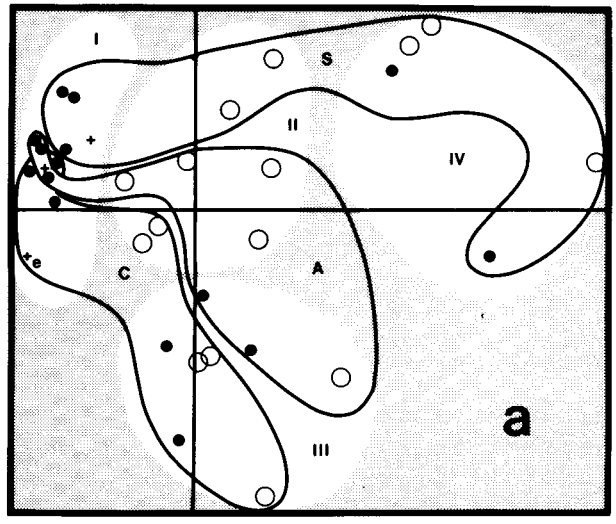
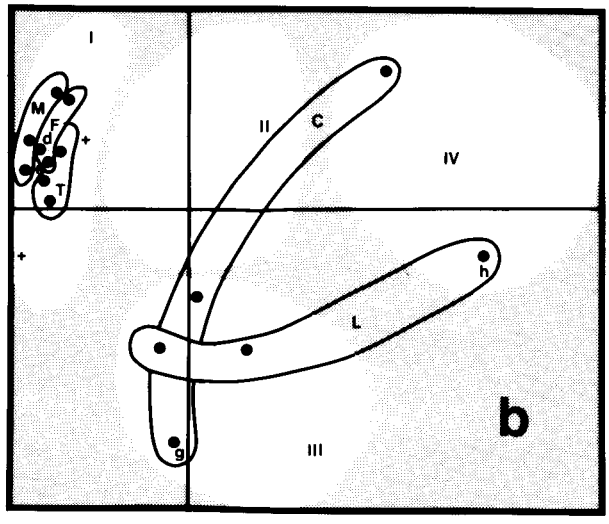
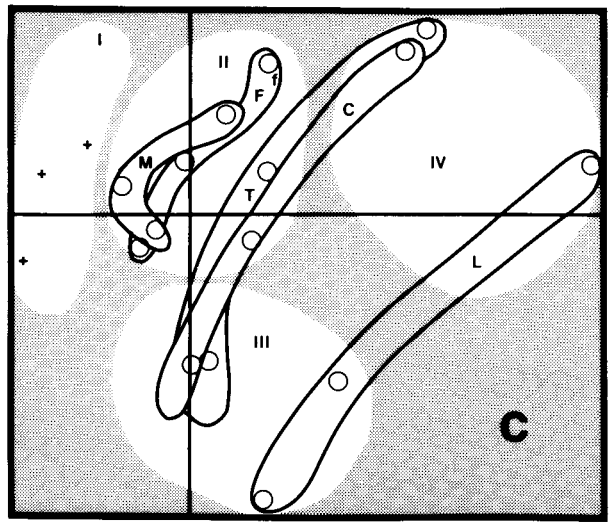


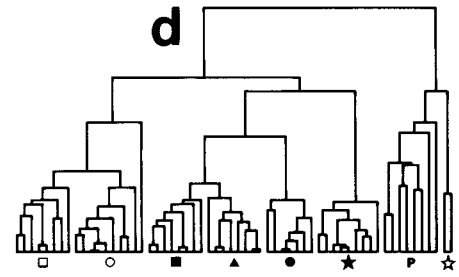
FIG. 6. Minimum variance cluster analysis of artificial dendrograms A–O of Fig. 2 based on: (*a*) topological difference; (*b*) cophenetic difference; (*c*) cluster membership divergence; (*d*) subtree membership divergence; (*e*) partition membership divergence; and (*f*) Euclidean distances calculated from all five descriptors.

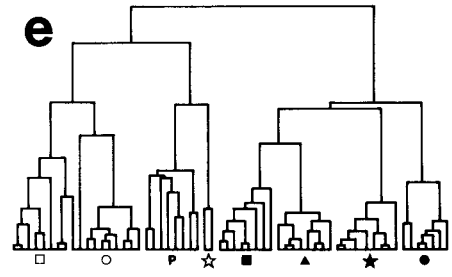
dendrograms according to CD. (*v*) Three descriptors, CD, CMD, and PMD yielded fairly consistent classifications at the four-group level: {A,C,D,J,K}, {M,N,O}, {B,H,I}, and {E,F,G} (Figs. 6*b*–6*d*). The affiliation of L in these classifications was ambiguous. Not surprisingly, this grouping also arose when all descriptors were considered simultaneously (Fig. 6*f*). (*vi*) Similarly, ordinations based on CD, CMD, and PMD are in best agreement with one another and with the combined ordination (Figs. 5*b*–5*d*, 5*f*). Despite the difficulty with hierarchic levels, in this data set the CD-based ordination suggests exactly the same group structure as the combined ordination (Figs. 5*b*, 5*f*). (*vii*) It may be concluded that the drawbacks of individual descriptors are diminished, whereas agreement among them is emphasized, if the various

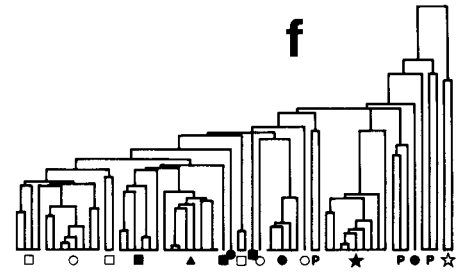

I
S
II
IV
C
A
III
a

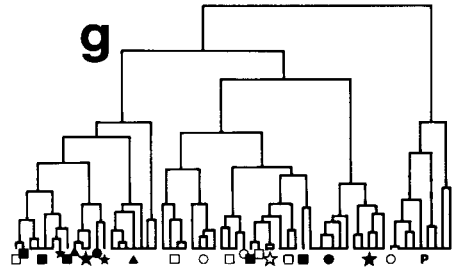

I
M
F
d
T
II
C
IV
h
L
g
III
b

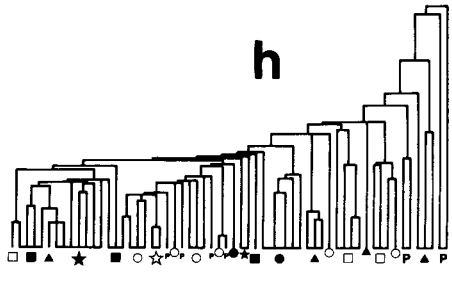

I
II
F
f
C
M
T
IV
L
III
c


d


e


f


g


h

aspects of dendrogram comparisons are considered simultaneously. The relative arrangement and grouping of dendrograms A–O in Figs. 5*f* and 6*f* agree with intuitive ideas about their interrelationships.

Phenograms

The alternative *Crataegus* data sets used here (Table 1) varied in the extent to which they produced phenograms congruent with an *a priori* classification of the sample according to site and taxon (Figs. 7*d*–7*h*). In general, the flower and fruit ($\mathbf{X}_F$) and the multistate ($\mathbf{X}_M$) characters tended to vary considerably between both sites and taxa while varying much less within these groups (Dickinson 1983; Dickinson and Phipps 1985). Conversely, the continuous ($\mathbf{X}_C$) and leaf ($\mathbf{X}_L$) characters were much more variable within groups, and so differentiated those groups much less sharply (Dickinson 1983; Dickinson and Phipps 1984, 1985). Multivariate comparisons of phenograms derived from these data sets enabled visualization of the way in which differences among resemblance coefficients and sorting algorithms were superimposed on these contrasts among data sets (Figs. 7*a*–7*c*).

Calculation of Euclidean distances among the 33 *Crataegus* phenograms based on five dendrogram descriptors ($\mathbf{S}_D$, Fig. 3), and minimum variance clustering, led to recognition of four clusters of phenograms (I–IV). In the phenograms of the first cluster all or virtually all OTUs were sorted according to site and taxon (Figs. 7*d* and 7*e*). In contrast, for the phenograms in the other three clusters, especially those of cluster IV, the sorting of OTUs was much less precise (compare Figs. 7*f*–7*h* with Fig. 7*d* and 7*e*). Sorting of OTUs according to species and especially section (*Crus-galli* versus *Punctatae*) was generally more successful (Figs. 7*d*–7*h*).

The first PCoA axis (associated with an eigenvalue representing 26% of the sum of all 33 eigenvalues of the distance matrix $\mathbf{S}_D$) corresponds to a trend in the sample from phenograms showing very similar, sharply defined groups (cluster I, Fig. 7*a*) to those in which groups are less well defined and of more variable composition, as a result either of chaining (single linkage phenograms, cluster IV; Fig. 7*a*) or of the difference between Euclidean and generalized distances (Figs. 7*b* and 7*c*). With Euclidean distances the sharpness of group structure may be enhanced by the pattern of correlations among the characters used. Generalized distances, however, are calculated in a manner independent of character correlations (here, from scores on mutually uncorrelated principal component axes). Phenograms produced from such distances thus may show correspondingly weaker group differentiation, as in the present case.

The second PCoA axis (associated with an eigenvalue representing 15% of the sum of all 33 eigenvalues) relates to the effect of sorting algorithm on the hierarchy of groups in a phenogram. Thus the phenograms produced by single linkage clustering have high scores on axis II (Fig. 7*a*) and tend to distinguish *C. crus-galli s. str.* at one site in particular (solid stars, Figs. 7*d*–7*h*) from the remaining samples of this taxon and *C. fontanesiana*. The corresponding average linkage phenograms present a simpler characterization of the sample of OTUs in which the samples from each site are grouped uniformly by taxon (Fig. 7*d*). Complete linkage clustering frequently imposed a hierarchy corresponding to differences in style and stamen numbers, at the highest level (Fig. 7*e*).

The third PCoA axis (not shown; 9% of the sum of all 33 eigenvalues) corresponds to a contrast between the three phenograms based on generalized distances calculated for all 17 characters (D_{TGS}, D_{TGA}, and D_{TGC}) and the six based on both Euclidean and generalized distances calculated for the six leaf characters (D_{LES}, D_{LEA}, D_{LEC}, D_{LGS}, D_{LGA}, and D_{LGC}). In the space defined by the second and third PCoA axes cluster I forms the most dense group of phenograms, close to the origins of the two axes. In the same space clusters II, III, and IV are increasingly spread out. This spreading, or lack of congruence, corresponds to increased variation in the composition and distinctness of groups, and in their hierarchical relationships. It is indicative of the decreasing stability of classifications as character set, resemblance function and sorting algorithm contribute to weakening group structure. In the space defined by the first two PCoA axes (Figs. 7*a*–7*c*) this weakening can be seen to be a function of not only continuous and leaf characters, generalized distances and the space-contracting effect of single linkage clustering but also the interaction among these factors.

The way in which the flower and fruit data gave rise to phenograms that reproduced the *a priori* classification of the sample is not surprising, in view of the heavy use made of flower and fruit characters in *Crataegus* taxonomy (Phipps and Muniyamma 1980; Dickinson and Phipps 1985). Conversely, the variability of leaf characters, or of exclusively continuous ones, makes group recognition based on these data sets more dependent on choices of method.

Comparison of the results of using different sorting algorithms here and elsewhere (Dickinson 1983; Dickinson and Phipps 1985) also indicates some possible drawbacks associated with space-dilating methods such as complete linkage and minimum variance. These methods tend to form a relatively small number of groups the distinctness of which may be exaggerated, if comparison is made with ordinations or other information (Sinnott and Phipps 1983). Moreover, to the extent that OTUs are forced into a small number of perhaps artificially distinct groups, these groups may be more heterogeneous

FIG. 7. Multivariate comparison of *Crataegus* phenograms. (*a*–*c*) Thirty-three phenograms of 60 *Crataegus* OTUs in the space defined by the first two principal coordinate axes for the matrix of Euclidean distances calculated from five dendrogram descriptors. Shading indicates clusters of dendrograms (I–IV) produced by minimum variance clustering using the same distances. Resemblance functions are indicated by the symbols used: information radius, +; Euclidean distance, ●; and generalized distance, ○. In (*a*), solid lines enclose dendrograms produced by single linkage (S), average linkage (A), and complete linkage (C) clustering. In (*b*), solid lines enclose dendrograms based on Euclidean distances calculated from the data sets described in Fig. 3 and Table 1: all 17 characters (T); nine flower and fruit characters (F); six leaf characters (L); seven multistate characters (M); and eight continuous characters (C). In (*c*), solid lines enclose dendrograms based on generalized distances calculated from the data sets described in Fig. 3 and Table 1; abbreviations as in (*b*). In Figs. 5*d*–5*h*, selected dendrograms corresponding to the points in Figs. 5*a*–5*c* labelled *d*, *e*, *f*, *g*, and *h*: (*d*) flower and fruit characters, Euclidean distance, average linkage (D_{FEA}); (*e*) information radii, complete linkage (D_{IIC}); (*f*) flower and fruit characters, generalized distance, single linkage (D_{FGS}); (*g*) continuous characters, Euclidean distance, complete linkage (D_{CEC}); and (*h*) leaf characters, Euclidean distance, single linkage (D_{LES}). In Figs. 7*d*–7*h* terminal vertices represent OTUs whose taxonomic affiliation is indicated by the associated symbols: *C. crus-galli s. str.*, solid symbols; *C. fontanesiana*, open circles and squares; *C. ?grandis*, open stars; and *C. punctata*, the letter P (one site only). Symbol shape (◨, ◑, ◭, ★) identifies four collecting sites. See Dickinson (1983) and Dickinson and Phipps (1985) for complete details concerning sampling and character sets used.

than is desirable. This is illustrated by the way in which 10-stamen *Crataegus* OTUs (*C. crus-galli s. str.*) were distinguished from 20-stamen ones (*C. fontanesiana, C. ?grandis* and *C. punctata*) in some complete linkage phenograms (Fig. 7*e*).

Finally, the degree of congruence among the phenograms based on flower and fruit characters, regardless of whether based on Euclidean (cluster I; Figs. 7*b* and 7*d*) or generalized distances (cluster II; Figs. 7*c* nd 7*f*), suggests that these characteristics are unlikely to be controlled by only a few pleiotropic genes (Atchley *et al.* 1982). Instead, there could be an appreciable genetic component to the phenotypic differentiation observed among sites in taxa belonging to *C. crus-galli s. l.* (Dickinson 1983; Dickinson and Phipps 1985).

Cladograms

Penny *et al.* (1982) argued that the congruence of evolutionary trees (cladograms) constructed from the amino acid sequences of different proteins by maximum parsimony methods (implying minimum evolution) for a common set of taxa or evolutionary units (EUs) should be a criterion by which the theory of evolution could be tested. They proposed that the degree of congruence of cladograms so constructed should be compared with that expected if the cladograms were selected at random. For nine different values of the measure of congruence that they used (Robinson and Foulds 1979), Penny *et al.* (1982) were able to calculate expected frequencies, and to compare these with the observed frequencies in a sample of 39 minimal and near minimal (maximally parsimonious) cladograms for 11 mammals. Penny *et al.* (1982) found that the degree of congruence among the 38 different cladograms was considerably greater than expected for randomly generated cladograms. They concluded that "the different protein sequences give trees that are markedly similar, showing a relationship among them that is consistent with the theory of evolution."

In the present study single, average and complete linkage cluster analyses grouped all or most of the 38 different cladograms produced by Penny *et al.* into three to five principal groups in the same way as did minimum variance clustering (clusters I–V, Fig. 8), in each case based on Euclidean distances calculated from four dendrogram descriptors. The five clusters formed by both minimum variance and average linkage sorting each consisted principally of the cladograms obtained from one class of protein sequences (Fig. 8). This corresponds to the finding by Penny *et al.* (1982, Table 4) that the average distance between cladograms derived from the same kind of sequence was smaller than that between those derived from different sequences.

Subjecting the matrix of Euclidean distances among the 38 cladograms to PCoA (Fig. 8*a*) demonstrated a contrast along the first axis (not shown; associated with an eigenvalue representing 27.4% of the sum of the 38 eigenvalues of the distance matrix) between cladograms derived from cytochrome *c* sequences in which a few clades, each consisting of several EUs, are present (cladogram 17, Fig. 8*e*) and ones derived from fibrinopeptide sequences in which chaining is conspicuous and all or most EUs form a clade by themselves (cladogram 24, Fig. 8*c*). This pattern resembles that seen along the first PCoA axis in comparisons of the 33 *Crataegus* phenograms (Fig. 7).

Contrasts along PCoA axes two through four are associated more with alternative phylogenies than with the presence or absence of chaining. The relative positions in the cladograms of dog (D) and horse (E) appear to account for the separation of clusters III and IV, consisting of all but two of the cladograms derived from hemoglobin A and B sequences, from the rest of the sample (Fig. 8*a*; compare Figs. 8*b* and 8*d* with Figs. 8*c*, 8*e*–8*g*; Penny *et al.* 1982). It is of interest that an examination of variation in the relative position of dog between the cladograms based on hemoglobin sequences (cladograms 27–39) and the remainder of the sample has led D. Penny and M. D. Hendy (unpublished material) to suggest checking dog hemoglobin sequences for possible errors. The method employed here thus has the additional advantage that it too may indicate which data in a study require verification, if they result in unusually incongruent dendrograms. Separation of cluster III from the other clusters along the fourth PCoA axis is associated with the relative position of mouse (M) in the cladograms (compare 36, Fig. 8*d*, with Figs. 8*b*, 8*c*, 8*e*–8*g*).

Our results demonstrate how the multivariate comparison of cladograms, in this case, can greatly facilitate summarization and interpretation of the contrasts among a large number of complex alternatives. Moreover, these results make it clear that while the 38 different minimal and near-minimal cladograms obtained by Penny *et al.* (1982) are quite similar, it can also be argued that particular proteins support rather different phylogenies than do others. Whether these differences are sufficient to constitute a falsification of the prediction made by Penny *et al.* (1982; that different proteins in related organisms should support similar phylogenetic schemes) is unclear. We suggest, however, that in view of the very large numbers of alternative phylogenetic hypotheses (cladograms) generated by different automatic reconstruction methods, data sets and coding schemes (Colless 1983), comparisons using a method like the one described here are needed to detect in the results patterns such as we have demonstrated (Fig. 8*a*) in those of Penny *et al.* (1982).

Methodological remarks

The multivariate technique suggested for the analysis of dendrograms utilizes five descriptor variables. Other possibilities for D–D comparisons, such as those offered by ultrametric relationships, nearest-neighbor interchanges, edge removals, etc. were deliberately neglected in constructing the distance coefficient [9]. It is of course conceivable that as many aspects of dendrogram structure as possible could be considered by incorporating all direct and indirect techniques into a single method. An important point, however, is that the calculation and standardization of dendrogram descriptors is a fast procedure, whereas most direct techniques (e.g., ultrametric dissimilarity) are much more time-consuming. A further difficulty is that no efficient algorithm is available for calculating nearest-neighbor interchanges, and the use of existing ones is restricted to relatively small dendrograms (Day 1983; Jarvis *et al.* 1983). Conversely, for the dendrogram descriptors the upper limit of dendrogram size and number is determined only by computer memory. Despite the problems it would be worthwhile in the future to examine whether or not features of dendrogram structure not incorporated in the approach given here in fact give substantial additional information about the relationships among alternative dendrograms.

A serious objection to the use of dendrogram descriptors might be that the cells of a descriptor matrix are not independent (Farris 1969, 1973). For this reason the descriptor matrix cannot be used to assess the statistical significance of the comparison. (Gower (1971*b*) suggested a technique which avoids

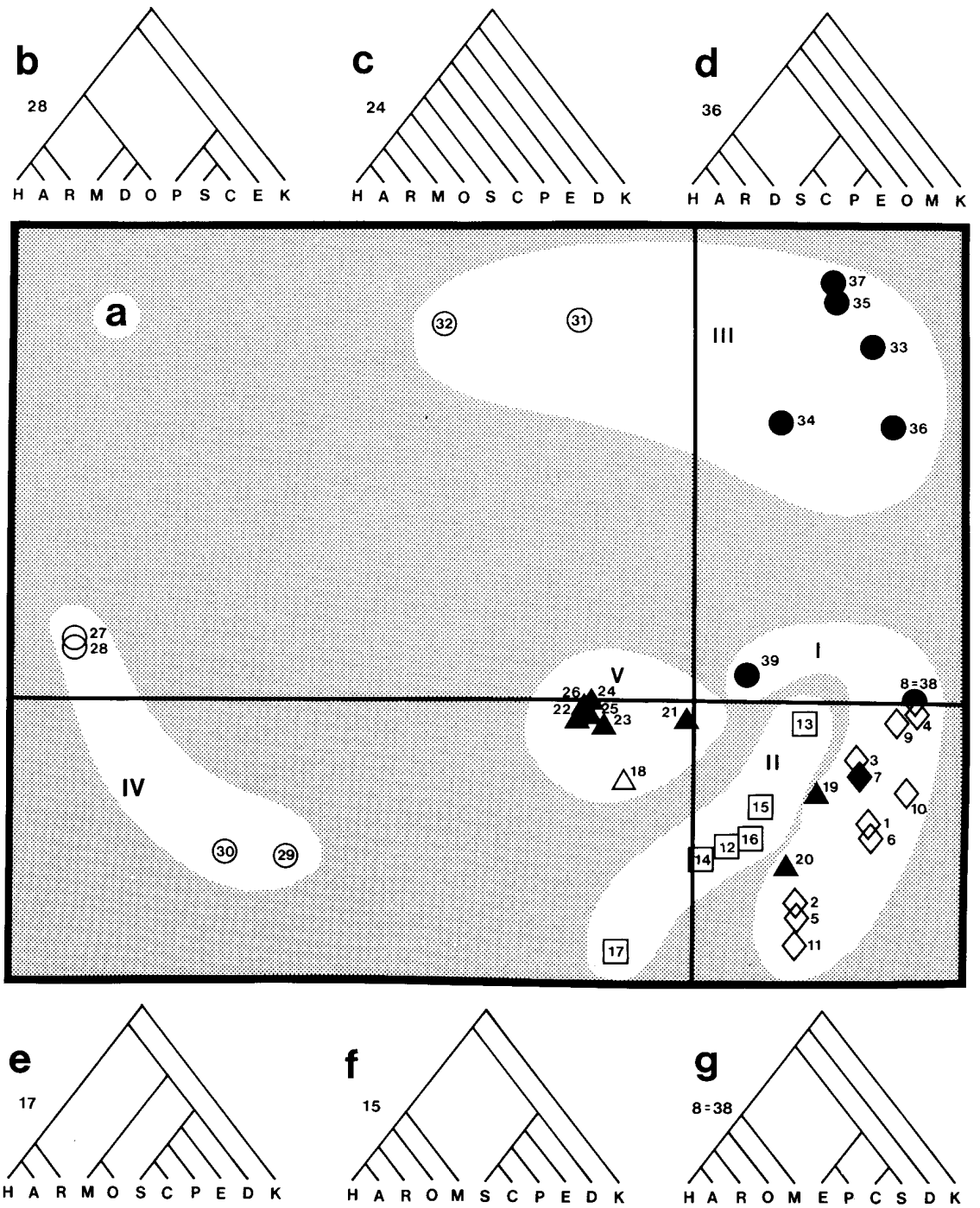


FIG. 8. Multivariate comparison of 38 different cladograms for 11 mammals published by Penny *et al.* (1982). (*a*) Cladograms in the space defined by the second and third principal coordinate axes for the matrix of Euclidean distances calculated from four dendrogram descriptors (TD, CMD, PMD, and SMD). Shading indicates clusters of cladograms (I−V) produced by minimum variance clustering using the same distances. Cladograms are coded according to the protein sequences on which they are based: □, cytochrome *c* (cladograms 12−17); ▲, fibrinopeptide (cladograms 18−26); ◑, hemoglobin (cladograms 27−39); and ◇, data for all proteins combined (cladograms 1−11). Open symbols indicate the A form, solid symbols the B form of the corresponding proteins. The combination symbol (solid circle, open lozenge) represents cladograms 8 and 38, which are identical. The solid lozenge represents cladogram 7 which is identical to the consensus cladogram obtained by Penny *et al.* (1982). Figs. 8*b*−8*g*. Selected cladograms for 11 mammals published by Penny *et al.* (1982) which illustrate extrema in the principal coordinates analysis shown in (*a*). Key to taxa: H, human; A, ape; R, Rhesus monkey; O, rabbit; M, mouse; D, dog; P, pig; C, cow; S. sheep; E, horse; K, kangaroo. Cladograms are based on data for (*b*) hemoglobin A, (*c*) fibrinopeptide B, (*d*) hemoglobin B, (*e*) cytochrome *c*, (*f*) cytochrome *c*, and (*g*) all proteins.

this problem, but the distributional properties of his R^2 statistic are unknown.) However, any statistic based on descriptors can still be used as an indicator of dendrogram congruence by keeping in mind that the differences in magnitude observed are interpretable only in the framework of the given study. Another potential criticism is that the five descriptors are correlated, and that this interdependence remains uncontrolled during the analysis. Correlation is present in practically every multivariate situation, and guarantees that agreements among different descriptors will tend to be emphasized, rather than

disagreements. Thus, for the descriptive purposes illustrated here, correlation of descriptors is in fact an advantage.

Because of the nature of the descriptors, the method presented here is applicable only to strictly binary dendrograms. For this reason, dendrograms differing in the number of interior vertices and thus implying unequal numbers of partitions and subtrees should not be compared in this way, since the distortion occurring in the results would be unpredictable. Nevertheless, experience with many data sets suggests that if the number of objects is much higher than the number of dendrograms, a few tri- or poly-furcations at low hierarchic levels may be allowed without jeopardizing the interpretability of the final results (e.g., J. Podani, unpublished study of 80 phytosociological units). We admit that the condition of binarity represents a serious limitation since in many practical situations hierarchies are neither detailed nor of great importance at low levels. Undoubtedly our technique could be modified so as to obtain a more widely applicable tool for dendrogram comparisons.

Conclusions

The three new descriptors of dendrogram structure that have been introduced here (CMD, SMD, and PMD) together with the two existing ones (CD and TD) make possible multivariate comparisons of dendrograms. Analysis of simple artificial dendrograms revealed striking differences among these five descriptors. TD, a descriptor currently used in many comparative studies (e.g., Duncan et al. 1980; Lachance and Starmer 1982; Hopper and Burgman 1983; cf. also the relationship between TD and some consensus indices), provided the most unique characterization of dendrogram resemblance. Comparisons based on SMD also differed from those based on the other descriptors. The remaining three descriptors tended to be in greater agreement with one another, and this agreement turned out to have a decisive influence on the results based on all five descriptors. Of course, far-reaching conclusions should not be drawn from this observation since dendrograms A–O were arbitrarily selected from the set of all possible ones. Nevertheless, we feel that this example demonstrates the greater objectivity of simultaneous comparisons of dendrograms based on all five descriptor variables.

The relevance of multivariate dendrogram analysis to the study of current numerical taxonomic problems is clear from the results of its application to two real data sets. The *Crataegus* study exemplified how illuminating our approach may be when examining the relative importance of choices associated with paths *a*, *b*, and *c* (Fig. 1). The experimental design of the study (Fig. 3) was deliberately made rather complex to demonstrate the possibilities present in a single analysis. Thus, the effects on the results of cluster analyses of character suite, scale of measurement, dissimilarity coefficient, and sorting algorithm were examined simultaneously. Character suite, as it related to the partitioning of variation between and within taxonomic groups in the sample, turned out to be the most important single factor (Fig. 7). In the same way, scale type and sorting algorithm were also found to be important. The choice of resemblance function in this case appeared to be much less critical.

The analysis of cladograms showed that even if no significant differences exist among the trees obtained by Penny *et al.* (1982), trends determined by the nature of the original input data and not otherwise revealed are nevertheless present, and of some interest (Fig. 8).

We conclude that an optimal survey of any kind of trees should include both significance tests and multivariate comparisons. Such an approach may prove extremely rewarding, regardless of whether or not there is reason to hypothesize the existence of a single, true dendrogram representation of the sample.

Acknowledgements

For supervision of the theses on which the work presented here is based and for financial support from their NSERC operating grants, the authors are deeply indebted to L. Orlóci (J.P.) and J. B. Phipps (T.A.D.). Each of us also gratefully acknowledges financial support from an Ontario Graduate Scholarship. We thank P. G. Smith, J. McNeill, and an anonymous reviewer for their comments and suggestions made on drafts of this paper. We thank D. Penny and M. D. Hendy of Massey University, Palmerston North, New Zealand, and L. R. Foulds of the University of Florida at Gainesville for allowing us to use some of the illustrations from their paper as our Figs. 8*b*–8*g*. These are reproduced by permission from Nature Vol. 297, No. 5863, page 198. Copyright © 1982 Macmillan Journals Limited.

Appendix

The Fortran program DENDAT has been written to prepare data for the multivariate analysis of alternative dendrograms. It is assumed that all dendrograms being analyzed are binary and that there is no reversal at the highest hierarchical level in any dendrogram.

The input data file for this program consists of a series of merge matrices. Each merge matrix describes the fusion sequence in a dendrogram in terms of m − 1 rows, where m is the number of objects. In each row there are five values: the first two refer to the identification numbers of the clusters being amalgamated, the second two indicate the number of objects in these clusters, while the fifth value is the hierarchical level at which the two clusters are fused. For example, the merge matrix for dendrogram A in Fig. 2 will be

```
1  2  1  1  1.0
4  5  1  1  1.0
1  3  2  1  2.0
4  6  2  1  2.0
1  4  3  3  5.0
```

Note that the identification number of a cluster always corresponds to the smallest of the identification numbers of its component objects. The recent versions of the classification programs NCLAS, HMCL, and INFCL of the SYN-TAX package (Podani 1984) include a subroutine which automatically creates a data file with the merge matrix as soon as the classification process is terminated. Alternatively, the merge matrix could be prepared by hand.

The user of DENDAT can choose which of the five descriptors will be calculated by specifying values for variables I1–I5. Hierarchical levels must be included in the merge matrix even if cophenetic differences are not requested. As in the case of the cladograms analyzed in this paper, dummy values reflecting the relative magnitude of the hierarchical levels are used when precise values are unavailable.

Other input parameters to be provided by the user are the number of objects in each dendrogram (M), the number of dendrograms (NC), the number of high-level partitions to be retained for calculation of PMD (KPART; default = M − 1), and the lowest value in the dendrogram (BAS; the diagonal value of the resemblance matrix among objects, usually either zero or unity).

The output is either a distance matrix for dendrograms (MOPT = 0) or a data matrix from which the distance matrix will be calculated by another program (MOPT = 1). In the first case the upper half of the NC × NC distance matrix, including the principal diagonal, is output by columns. In the second, output consists of an $I \times (M \times (M-1)/2)$ (rows) × NC(columns) matrix, output by rows, where I is the number of descriptors calculated. In both cases the descriptor variables are standardized as described earlier. Printed output from DENDAT lists the values of all input parameters plus those of the constants used to standardize the descriptors.

The listing of program DENDAT and some input–output examples are available from the authors on request.

ADAMS, E. N. 1972. Consensus techniques and the comparison of taxonomic trees. Syst. Zool. **21**: 390–397.

ATCHLEY, W. R., E. V. NORDHEIM, F. C. GUNSETT, and P. L. CRUMP. 1982. Geometric and probabilistic aspects of statistical distance functions. Syst. Zool. **31**: 445–460.

COLLESS, D. H. 1983. Wagner trees in theory and practice. *In* Numerical taxonomy. *Edited by* J. Felsenstein. Springer-Verlag, Berlin and Heidelberg. pp. 259–278.

CROVELLO, T. J. 1968. The effect of alteration of technique at two stages in a numerical taxonomic study. Univ. Kans. Sci. Bull. **47**: 761–786.

CUNNINGHAM, K. M., and J. C. OGILVIE. 1972. Evaluation of hierarchical grouping techniques. A preliminary study. Comput. J. **15**: 209–213.

DAY, W. H. E. 1983. Properties of the nearest-neighbor interchange metric for trees of small size. J. Theor. Biol. **101**: 275–288.

DICKINSON, T. A. 1983. *Crataegus crus-galli* L. *sensu lato* in southern Ontario: phenotypic variation and variability in relation to reproductive behavior. Ph.D. thesis, University of Western Ontario, London. (National Library of Canada Microfiche TC 58716.)

DICKINSON, T. A., and J. B. PHIPPS. 1984. Studies in *Crataegus* (Rosaceae: Maloideae). IX. Short-shoot leaf heteroblasty in *Crataegus crus-galli sensu lato*. Can. J. Bot. **62**: 1775–1780.

——— 1985. Studies in *Crataegus* L. (Rosaceae: Maloideae). XIII. Degree and pattern of phenotypic variation in *Crataegus* section *Crus-galli* in Ontario. Syst. Bot. **10**. In press.

DIETZ, E. J. 1983. Permutation tests for association between two distance matrices. Syst. Zool. **32**: 21–26.

DOBSON, A. J. 1975. Comparing the shape of trees.Lect. Notes Math. **452**: 95–100.

DOUGLAS, M. E., and J. A. ENDLER. 1982. Quantitative matrix comparisons in ecological and evolutionary investigations. J. Theor. Biol. **99**: 777–795.

DUNCAN, T., and B. R. BAUM. 1981. Numerical phenetics: its uses in botanical systematics. Annu. Rev. Ecol. Syst. **12**: 387–404.

DUNCAN, T., R. B. PHILLIPS, and W. H. WAGNER, JR. 1980. A comparison of branching diagrams derived by various phenetic and cladistic methods. Syst. Bot. **5**: 264–293.

FARRIS, J. S. 1969. On the cophenetic correlation coefficient. Syst. Zool. **18**: 279–285.

——— 1973. On comparing the shapes of taxonomic trees. Syst. Zool. **22**: 50–54.

GILMARTIN, A. J. 1974. Variation within populations and classification. I. Taxon, **23**: 523–536.

——— 1980. Variation within populations and classification. II. Patterns of variation within Asclepiadaceae and Umbelliferae. Taxon, **29**: 199–210.

GOWER, J. C. 1971*a*. A general coefficient of similarity and some of its properties. Biometrics, **27**: 857–871.

——— 1971*b*. Statistical methods of comparing different multivariate analyses of the same data. *In* Mathematics in the archaeological and historical sciences. *Edited by* F. R. Hodson, D. G. Kendall, and P. Tăutu. University of Edinburgh Press, Edinburgh. pp. 138–149.

HARTIGAN, J. A. 1967. Representation of similarity matrices by trees. J. Am. Stat. Assoc. **62**: 1140–1158.

——— 1975. Clustering algorithms. John Wiley & Sons, New York.

HAWKSWORTH, F. G., G. F. ESTABROOK, and D. J. ROGERS. 1968. Application of an information theory model for character analysis in the genus *Arceuthobium* (Viscaceae). Taxon, **17**: 605–619.

HOPPER, S. D., and M. A. BURGMAN. 1983. Cladistic and phenetic analyses of phylogenetic relationships among populations of *Eucalyptus caesia*. Aust. J. Bot. **31**: 35–49.

JACKSON, D. N. 1969. Comparison of classifications. *In* Numerical taxonomy. *Edited by* A. J. Cole. Academic Press, New York. pp. 91–113.

JARDINE, N., and R. SIBSON. 1968. The construction of hierarchic and nonhierarchic classifications. Comput. J. **11**: 177–184.

——— 1971. Mathematical taxonomy. John Wiley & Sons, New York.

JARVIS, J. P., J. K. LUEDMAN, and D. R. SHIER. 1983. Comments on computing the similarity of binary trees. J. Theor. Biol. **100**: 427–433.

JENSEN, R. J. 1983. Report on the Sixteenth International Numerical Taxonomy Conference. Syst. Zool. **32**: 83–89.

LACHANCE, M.-A., and W. T. STARMER. 1982. Evolutionary significance of physiological relationships among yeast communities associated with trees. Can. J. Bot. **60**: 285–293.

MICKEVICH, M. F. 1978. Taxonomic congruence. Syst. Zool. **27**: 143–158.

ORLÓCI, L. 1978. Multivariate analysis in vegetation research. Dr. W. Junk bv Publishers, The Hague.

PENNY, D., L. R. FOULDS, and M. D. HENDY. 1982. Testing the theory of evolution by comparing phylogenetic trees constructed from five different protein sequences. Nature (London), **297**: 197–200.

PHIPPS, J. B. 1971. Dendrogram topology. Syst. Zool. **20**: 306–308.

PHIPPS, J. B., and M. MUNIYAMMA. 1980. A taxonomic review of *Crataegus* (Rosaceae) in Ontario. Can. J. Bot. **58**: 1621–1699.

PODANI, J. 1980. SYN-TAX: computer program package for cluster analysis in ecology, phytosociology and taxonomy. Abstr. Bot. (Budapest), **6**: 1–158.

——— 1982. Spatial processes in the analysis of vegetation. Ph.D. thesis, University of Western Ontario, London. (National Library of Canada Microfiche TC 56124.)

——— 1984. SYN-TAX II. Computer programs for data analysis in ecology and systematics. Abstr. Bot. (Budapest). In press.

PRENTICE, H. C. 1979. Numerical analysis of infraspecific variation in European *Silene alba* and *S. dioica* (Caryophyllaceae). Bot. J. Linn. Soc. **78**: 181–212.

ROBINSON, D. F. 1971. Comparison of labelled trees with valency three. J. Comb. Theory, **11**: 105–119.

ROBINSON, D. F., and L. R. FOULDS. 1979. Comparisons of weighted labelled trees. Lect. Notes Math. **748**: 119–126.

ROHLF, F. J. 1970. Adaptive hierarchical clustering schemes. Syst. Zool. **19**: 58–82.

——— 1974. Methods of comparing classifications. Annu. Rev. Ecol. Syst. **5**: 101–113.

——— 1982. Consensus indices for comparing classifications. Math.

Biosci. **59**: 131–144.

ROHLF, F. J., and R. R. SOKAL. 1981. Comparing numerical taxonomic studies. Syst. Zool. **30**: 459–490.

SCHNELL, G. D. 1970. A phenetic study of the suborder Lari (Aves). II. Phenograms, discussion and conclusions. Syst. Zool. **19**: 264–302.

SEAMAN, F. C., and V. C. FUNK. 1983. Cladistic analysis of complex natural products: developing transformation series from sesquiterpene lactone data. Taxon, **32**: 1–27.

SINNOTT, Q. P., and J. B. PHIPPS. 1983. Variation patterns in *Crataegus* series *Pruinosae* (Rosaceae) in southern Ontairo. Syst. Bot. **8**: 59–70.

SMITH, P. G., and J. B. PHIPPS. 1984. Consensus trees in phenetic analyses. Taxon, **33**(4). In press.

SNEATH, P. H. A., and R. R. SOKAL. 1973. Numerical taxonomy. W. H. Freeman & Co., San Francisco.

SOKAL, R. R., and F. J. ROHLF. 1962. The comparison of dendrograms by objective methods. Taxon, **11**: 33–40.

WARD, W. H., JR. 1963. Hierarchical grouping to optimize an objective function. J. Am. Stat. Assoc. **58**: 236–244.

WATERMAN, M. S., and T. F. SMITH. 1978. On the similarity of dendrograms. J. Theor. Biol. **73**: 789–800.

WILLIAMS, W. T., and H. T. CLIFFORD. 1971. On the comparison of two classifications of the same set of elements. Taxon, **20**: 519–522.