

Comparison Tests for Dendrograms: A Comparative Evaluation

François-Joseph Lapointe

Pierre Legendre

Université de Montréal

Université de Montréal

Abstract: Classifications are generally pictured in the form of hierarchical trees, also called dendrograms. A dendrogram is the graphical representation of an ultrametric (= cophenetic) matrix; so dendrograms can be compared to one another by comparing their cophenetic matrices. Three methods used in testing the correlation between matrices corresponding to dendrograms are evaluated. The three permutational procedures make use of different aspects of the information to compare dendrograms: the Mantel procedure permutes label positions only; the binary tree methods randomize the topology as well; the double-permutation procedure is based on all the information included in a dendrogram, that is: topology, label positions, and cluster heights. Theoretical and empirical investigations of these methods are carried out to evaluate their relative performance. Simulations show that the Mantel test is too conservative when applied to the comparison of dendrograms; the methods of binary tree comparisons do slightly better; only the double-permutation test provides unbiased type I error.

This work was supported by NSERC grant no. A7738 to Pierre Legendre and by a NSERC scholarship to F.-J. Lapointe.

Authors' address: Département de Sciences biologiques, Université de Montréal, C.P. 6128, Succ. Centre-ville, Montréal, Québec, Canada H3C 3J7.

Résumé: Les arbres utilisés pour illustrés les groupements sont généralement représentés sous la forme de classifications hiérarchiques ou dendrogrammes. Un dendrogramme représente graphiquement l'information contenue dans la matrice ultramétrique (= cophénétique) correspondant à la classification. Dès lors, il est possible de comparer des dendrogrammes à partir des matrices ultramétriques correspondantes. Nous comparons trois méthodes permettant d'évaluer la signification statistique du coefficient de corrélation mesuré entre deux matrices ultramétriques. Ces trois tests par permutations tiennent compte d'aspects différents pour comparer des dendrogrammes: le test de Mantel permute les feuilles de l'arbre, les méthodes pour arbres binaires permudent les feuilles et la topologie, alors que la procédure à double permutation permute les feuilles, la topologie et les niveaux de fusion des dendrogrammes comparés. L'efficacité relative des trois méthodes est évaluée empiriquement et théoriquement. Nos résultats suggèrent l'utilisation préférentielle du test à double permutation pour la comparaison de dendrogrammes: le test de Mantel s'avère trop conservateur, tandis que les méthodes pour arbres binaires ne sont pas toujours adéquates.

Keywords: Binary tree; Dendrogram; Classification; Permutation test; Ultrametric tree.

1. Introduction

Modern methods of numerical classification have a multi-national origin. Polish anthropologist Czekanowski (1909) and ecologist Kulczynski (1928) are probably the oldest references to classifications obtained by numerical methods; they were using a technique known nowadays as seriation. Later, a group of Polish mathematicians (Florek, Lukaszewicz, Perkal, Steinhaus and Zubrzycki) led by Lukaszewicz (1951) developed nearest-neighbour analysis, which is the basis of single linkage clustering, shortly after ecologist Sørensen (1948) in Denmark had proposed complete linkage clustering. But it was left to numerical taxonomists, Sneath (1957) in Great Britain and Sokal and Michener (1958) in the USA, to develop the first machine-based methods of numerical classification, which led to the first, well-known synthesis of the field (Sokal and Sneath 1963) and later to the foundation of Classification Societies throughout the world.

Before Classification became a widely accepted field of investigation, numerical taxonomy was the field where mathematical tools burgeoned, to construct trees from empirical data (Sneath and Sokal 1973). Consensus theory subsequently contributed to the development of several comparison indices to quantify the resemblance of such trees (Rohlf 1982). Little effort has been put, however, on testing these resemblance values, and on assessing the distributions and statistical significance levels of particular comparison indices. Two approaches are available to do so: (1) the first one is to develop

a specific statistical test to evaluate the significance of the resemblance values computed among trees, while (2) the second option is to derive a probability distribution for a particular consensus index, either empirically or analytically. Empirical and exact distributions of some consensus indices are already known: Day (1983, 1986) has assessed the distributions of four metrics to compare trees; studies by Shao and Rohlf (1983) and Shao and Sokal (1986) provided tables of significance for ten consensus indices; and, more recently, Steel and Penny (1993), improving on the Hendy et al. (1984) and Steel (1988) studies, derived statistical distributions for three more comparison statistics, based on different underlying tree distributions. Most of these studies only addressed the distribution problem (option 2 above), providing statistical tables to compare unweighted trees, i.e., trees without branch lengths (for weighted tree distributions, see Lapointe and Legendre 1992). The present study will focus on various statistical approaches (option 1) that have been proposed to test for the resemblance of a pair of trees. We will restrict ourselves, however, to the comparison of dendrograms, and will evaluate the relative performance of three specific statistical tests to do so.

2. What is a Dendrogram?

A dendrogram can formally be defined as a rooted terminally-labeled weighted tree in which all terminal nodes are equally distant from the root (Lapointe and Legendre 1991). But, for the purpose of the present paper, a dendrogram can adequately be described by three types of characteristics only: topology, labels, and cluster heights.

1. The topology of a dendrogram represents its furcation structure, or its shape (Harding 1971). The topology actually corresponds to an unweighted and unlabeled tree; that is, a tree without specified branch lengths or cluster heights, and with no labels assigned either to the terminal or the internal nodes (Fig. 1a). Dendrogram topologies are always rooted at a given internal node, and are usually binary, bearing $n - 1$ internal nodes for n terminal nodes; in the remainder of this paper, we will consider binary tree topologies only, since multifurcations can always be decomposed into a series of bifurcations with null internode distances. The number of possible binary topologies is a function of the number n of terminal nodes of the dendrogram (Murtagh 1984).
2. A dendrogram must also bear labels identifying the n objects being classified. These labels are attached to the terminal nodes of the topology (Fig. 1b), and the label positions allow to tell apart otherwise similar trees (Murtagh 1984). The number of terminally-labeled binary trees is a function of the number of objects n , but the number of distin-

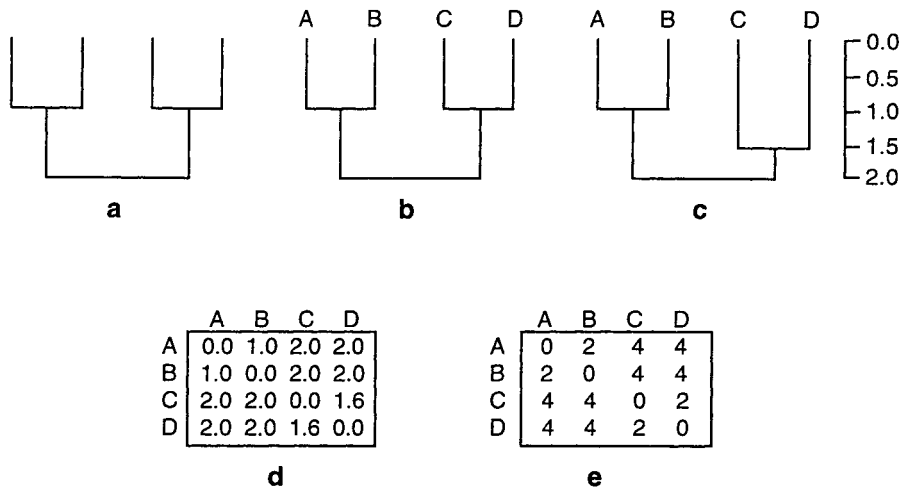


Figure 1. (a) An unlabeled and unweighted tree is a topology. (b) When adding labels, it becomes an unweighted tree. (c) When adding cluster heights, one obtains a dendrogram. In that case, the weighted topology corresponding to this dendrogram can be represented in packed form by {1.0, 2.0, 1.6}. The cophenetic matrix (d) and the cardinality matrix (e) respectively represent the distance and topological relationships among objects.

guishable ways to distribute the labels is constrained by the dendrogram topology (Lapointe and Legendre 1991).

3. Finally, a dendrogram must have cluster heights. Dendrograms actually represent a special form of weighted tree (i.e., a tree with length values attached to the branches) in which all terminal nodes are equidistant from the root (Fig. 1c). This property, which is referred to as the ultrametric condition (Hartigan 1967; Johnson 1967), allows the distances among objects to be described by a height scale associated to the $n - 1$ internal nodes (i.e., clusters) of the dendrogram and going from the root to the level of the leaves. The number of distinguishable dendrograms for a fixed set of cluster heights (i.e., fusion levels) is a function of the number of objects n (Lapointe and Legendre 1991).

The weighted shape of a dendrogram can be uniquely encrypted in a packed representation, formed by listing the cluster heights (i.e., the maximum distance between objects linked by the given fusion level) encountered when traversing the topology from left to right (see Fig. 1). Likewise, any dendrogram can be uniquely represented by an ultrametric matrix (i.e., a cophenetic matrix *sensu* Rohlf and Sokal, 1981) bearing the fusion levels among all pairs of objects i and j (Fig.

1d). For n objects, there are $n(n-1)/2$ such pairs, but at most $n-1$ different cluster heights, each corresponding to an internal node. An unranked binary tree (Fig. 1b) may also be represented by a unique matrix, using cardinalities (i.e., the number of vertical edges along the path between objects) in place of cluster heights (Fig. 1e); it is equivalent to define the cardinality distance between objects i and j to be the size of the smallest unpruned subtree containing both i and j . There is a one-to-one correspondence between dendrograms and cophenetic matrices, and also between unranked binary trees and cardinality matrices. This reduces the comparison of dendrograms to the problem of comparing matrices.

3. How to Compare Dendrograms?

We will consider three different statistical approaches to compare dendrograms. All three are randomization tests, based on the comparison of matrices and using the matrix correlation coefficient as their comparison index (Rohlf 1982). We test for agreement between dendrograms. In all three cases, the null hypothesis states that the correlation observed between two trees is not different from zero, or that the trees are not more similar to one another than random pairs of trees would be. In the particular case of dendrograms, one is evaluating whether the correlation between a given pair of dendrograms is larger than that between random dendrograms. The test proceeds as follows:

- a) Compute the correlation between two dendrograms using the matrix correlation coefficient (i.e., the correlation between corresponding values in two half-matrices, diagonal excluded).
- b) Generate a pair of random dendrograms (the pair of random dendrograms must have the same number of terminal nodes, representing the same set of objects, as in the data being studied).
- c) Compute the correlation between the two random dendrograms.
- d) Repeat steps b and c a large number of times (say, 1000 times).
- e) Construct a null distribution of the matrix correlation coefficient for the values obtained in c.
- f) Reject the null hypothesis (no correlation) when most random pairs of dendrograms (say, more than 95% for a type I error of 0.05) exhibit a smaller correlation than the original pair of dendrograms (one-tailed test).

All three groups of randomization tests compared in this study follow that same statistical approach. The difference lies in step b of the algorithm above: how to generate random dendrograms? The characteristics of a

dendrogram taken into account vary with each test. The first method looks at the label positions only, the second approach considers the labels and the topology, whereas the last one takes all three aspects of a dendrogram (i.e., topology, label positions, and cluster heights) into account in the randomization process.

3.1 The Random Label Model

The easiest way to randomize any matrix is by permutation. Likewise, the simplest model for comparing dendrograms is to randomize the two corresponding cophenetic matrices. The Mantel test (1967) is such a randomization procedure designed specifically to compare dissimilarity matrices. It has been used frequently in numerical taxonomy (Sokal 1979), population genetics (Sokal and Wartenberg 1983), population ecology (Sokal and Unnasch 1987; Hudon and Lamarche 1989), community ecology (Burgman 1987; Legendre and Fortin 1989), biogeography (Schnell et al. 1986; Page 1987; Ardisson et al. 1990), psychometry (Hubert and Levin, 1976), anthropology (Sokal et al. 1986, 1987), social sciences (Krackhardt and Porter 1986; Krackhardt and Kilduff 1990) and ethology (de Wall and Luttrell 1988). In their application of the method, Hubert and Baker (1977) have focussed on the comparison of dendrograms.

The test consists of assessing the significance of the correlation between two actual distance matrices using a distribution of correlation coefficient values between matrices with randomly permuted rows and columns. In the case of cophenetic matrices representing dendrograms, this is like randomizing the label positions, i.e., permuting the object labels on the leaves of a fixed topology. The population of trees accounted for by the Mantel test thus corresponds to the number of different but equiprobable ways of relabeling a given dendrogram:

$$L_{(n,m)} = n! / 2^m \quad (1)$$

where n is the number of objects, m is the number of internal nodes with exactly two terminal leaves on the tree (Lapointe and Legendre 1991), and $L_{(n,m)}$ represents the number of distinguishable labelings.

When applied to cophenetic matrices, the Mantel test evaluates the null hypothesis that a given pair of dendrograms are not more similar than pairs of dendrograms relabeled at random. The topology and cluster heights are kept constant in the permutations; so, this random label model considers only one characteristic of dendrograms. When the topology is also of interest, one has to use a different randomization method, based on more than label positions alone.

3.2 The Random Tree Model

If the topology and labels are both important, random trees have to be generated accordingly, with random label positions and a random topology. If we want to compare unweighted binary trees, this amounts to generating and comparing random cardinality matrices. Therefore, the main difficulty in designing a comparison test that considers the topology lies in the generation of random binary trees, and of their corresponding matrices. Several tree generation algorithms (Furnas 1984; Oden and Shao 1984; Quiroz 1989) are now available to do so, and several tests and statistical distributions (Day 1983; Shao and Rohlf 1983; Day 1986; Shao and Sokal 1986; Simberloff 1987; Page 1988; Steel and Penny 1993) have been proposed for binary tree comparisons. These methods evaluate whether the similarity observed between two binary trees could have occurred by chance alone. All tests are based on the same randomization approach, though they sometimes rely on different distributions of trees.

Three types of assumptions may govern the generation of binary tree distributions (Simberloff et al. 1981): (1) every topology is equiprobable, (2) every tree is equiprobable, and (3) every branching node is equiprobable when growing a tree. The first assumption implies that every topology has the same probability of being generated; in other words, a completely symmetric topology and a fully asymmetric one are equally likely, independently of the label positions. The second assumption requires a topology to be represented in the population in proportion to the number of distinguishable trees possessing that particular topology; when generating such trees, the addition of a new branch is equiprobably distributed among all existing branches (Steel and Penny 1993). The last assumption is related to the second one. The difference is that the location of the next branch, when growing a tree, is equiprobably distributed among terminal branches only, and not among internal branches. The number of trees distinguishable for all three assumption is the same; only the proportion of topologies differs (Fig. 2). The population size corresponding to this model is (Phipps 1975; Felsenstein 1978):

$$B_n = (2n - 3)! / 2^{n-2} (n - 2)! \quad (2)$$

where n is the number of objects, and B_n is the number of distinguishable binary trees.

When using a test designed for the comparison of binary trees to test the correlation of two dendrograms, one has to neglect the distance information associated with cluster heights. In practice, the comparison test involves cardinality matrices and assesses the null hypothesis that two matrices representing dendrograms are not more correlated to one another than cardin-

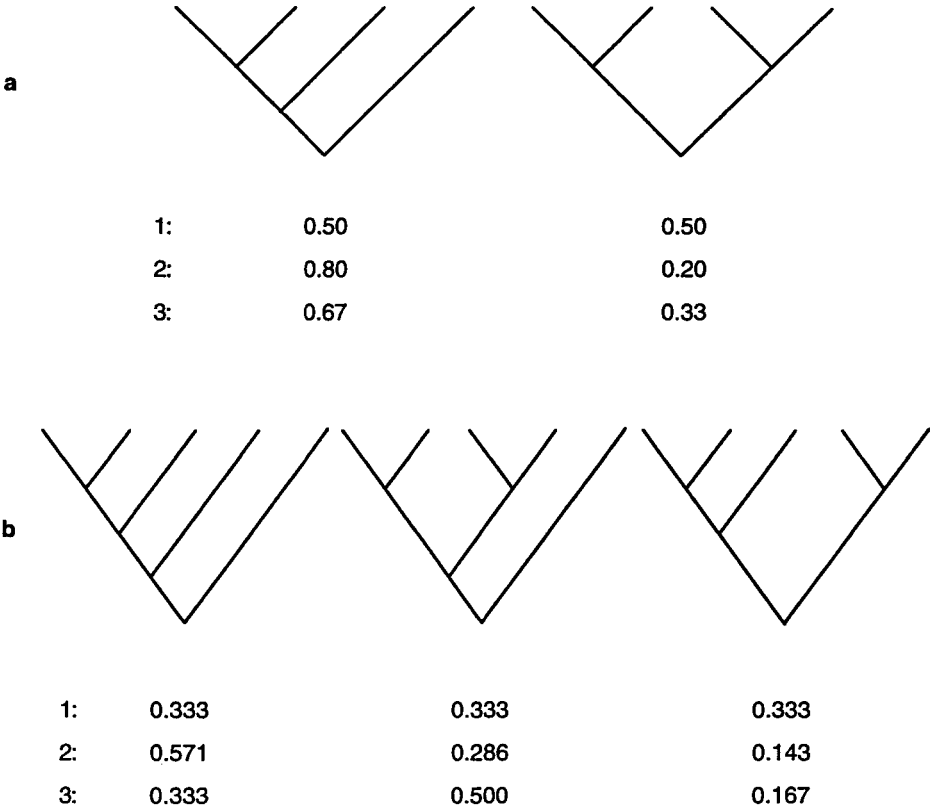


Figure 2. Neglecting labels, (a) two topological forms are distinguishable for 4 objects, while (b) there are three possible topologies for 5 objects. The probability of each topology is presented under different hypotheses: 1) the equiprobable topology distribution, 2) the equiprobable tree distribution, 3) the equiprobable branching nodes distribution. Adapted from Savage (1983).

ality matrices associated to random binary trees would be. The branch length information is left out of the randomization process. When distances are also to be taken into consideration, another test should be used that randomizes all three characteristics of dendrograms.

3.3 The Random Dendrogram Model

The double-permutation test of Lapointe and Legendre (1990) is a method specifically designed to compare cophenetic matrices representing dendrograms. Published applications include comparisons of phylogenetic trees (Lapointe and Legendre 1990; Page 1990; Lapointe 1992), taxonomic classifications (Cuerrier et al. 1992), and phenograms (Lapointe and

Legendre, 1994). This test considers all three characteristics of dendrograms in the generation process. Its specificity lies in the randomization of cluster heights, that the other two families of tests do not consider. Instead of assigning random lengths onto the branches of a tree (Furnas 1984), cophenetic matrices are randomized directly by a double-permutation procedure.

The algorithm for generating a random dendrogram proceeds by a permutation of the packed representation to randomize the topology and cluster heights, followed by a permutation of the labels, as in the Mantel test, to randomize their positions. The significance of the correlation between the original matrices is then assessed by comparing it to a distribution of correlation values between dendrograms with a randomized topology, random label positions and random permutation of the cluster heights (Lapointe and Legendre 1990). It has been shown that this procedure can equiprobably generate all distinguishable dendrograms (Lapointe and Legendre 1991), like if one was drawing at random from a population of size (Frank and Svensson 1981):

$$D_n = n! (n - 1)! / 2^{n-1} \quad (3)$$

where D_n is the number of distinguishable dendrograms relating n objects. Interestingly, the distribution of cardinality matrices corresponding to such equiprobable dendrograms is identical to the distribution of binary trees, when generated with equiprobable branching nodes (Page 1991).

When comparing dendrograms using the double-permutation test, the null hypothesis is the same as for the other models. That is, the observed correlation between dendrograms is not statistically different from zero. However, the testing of the null hypothesis relies here on correlation values obtained for random dendrograms, not simply random label positions, or random topologies. These differences in the definition of randomness and their effects on the statistical comparison of dendrograms are discussed below.

4. Theoretical Comparison of the Three Methods

The tests under investigation all call for a specific randomization model. The main difference among them is the reference population from which the random trees are drawn. Sampling is equiprobable in every case, but the population size varies depending upon the test selected. Differences among the tests are then reflected in the different population sizes. Table 1 illustrates that, for any given number of objects ($n > 3$), the following order always holds:

$$L_{(n,m)} < B_n < D_n \quad (4)$$

Table 1

The number of labelings (L), binary trees (B), and dendrograms (D) distinguishable for n objects, in the case where $m = 1$.

n	L	B	D
1	1	1	1
2	1	1	1
3	3	3	3
4	12	15	18
5	60	105	180
6	360	945	2 700
7	2 520	10 395	56 700
8	20 169	135 135	1 587 600
9	181 440	2 027 025	57 153 600
10	1 814 400	34 459 425	2 571 912 000

Table 2

Results of the simulations involving 1000 pairs of random dendrograms ($n = 5$). The number of cases where the null hypothesis has been rejected at different levels of significance is presented for each test corresponding to the three different random models. Figures in parentheses are empirical probabilities.

Models	Significance levels		
	0.01	0.05	0.10
Random labels	0 (0.000)	16 (0.016)	41 (0.041)
Random binary trees			
Hypothesis 1	4 (0.004)	28 (0.028)	62 (0.062)
Hypothesis 2	4 (0.004)	24 (0.024)	62 (0.062)
Hypothesis 3	4 (0.004)	28 (0.028)	80 (0.080)
Random dendrograms	12 (0.012)	50 (0.050)	100 (0.100)

where m is the number of internal nodes with exactly two terminal leaves on the tree. This relation remains true for any value of m since $L_{(n,m)}$ can only become smaller for larger m when n is fixed. It means that the population of dendrograms is always larger than the population of binary trees and that of the labeling possibilities. Differences increase as the number of objects n gets larger, not only because more topologies are distinguishable among dendrograms, but also because there are more ways of labeling a dendrogram than a binary tree with the same topology (Fig. 3).

Notice that the labeling and binary tree populations are subsets of the dendrogram population. Figure 4 shows the 18 distinguishable dendrograms for 4 objects. Among them, there are 15 distinguishable binary trees (Fig. 4, trees a to o) when ignoring cluster heights; this is the reference population for the random binary tree model. The different labeling possibilities are a function of the topology, however. Two subgroups are distinguishable in the example: topology A ($m = 1$, Fig. 4, trees a to l) allows for 12 different label positions, whereas topology B leads to only 6 different label positions ($m = 2$, Fig. 4, trees m to r). Each topology encompasses a sub-population of the entire set of dendrograms. The random label procedure samples from either sub-population A or sub-population B, not both.

These comparisons suggest that the double-permutation test may be more appropriate than the other two methods for dendrogram comparisons. A statistical bias may be introduced by considering only subsets of the dendrogram population. Monte Carlo simulations were thus carried out to evaluate these theoretical findings.

5. Empirical Comparison of the Three Methods

The purpose of these simulations is to compare the different methods in empirical situations. We want to evaluate the effect of different randomization models for dendrograms when testing the significance of the matrix correlation coefficient. We have used the following approach in our simulations: (1) realizations of the null hypothesis were obtained by generating pairs of random dendrograms; (2) the correlation between the cophenetic and cardinality matrices associated with these dendrograms were computed; and (3) the significance of each correlation value was tested using the different testing procedures. Finally, to evaluate the performance of the different methods under our null hypothesis, the probability of rejecting type I error (i.e., the probability of rejecting the null hypothesis of no correlation when H_0 is in fact true) for fixed H_0 (0.10, 0.05, 0.01) was calculated for each random model (these probabilities are computed as the number of comparison tests rejecting H_0 , divided by the total number of comparisons), and the results were compared.

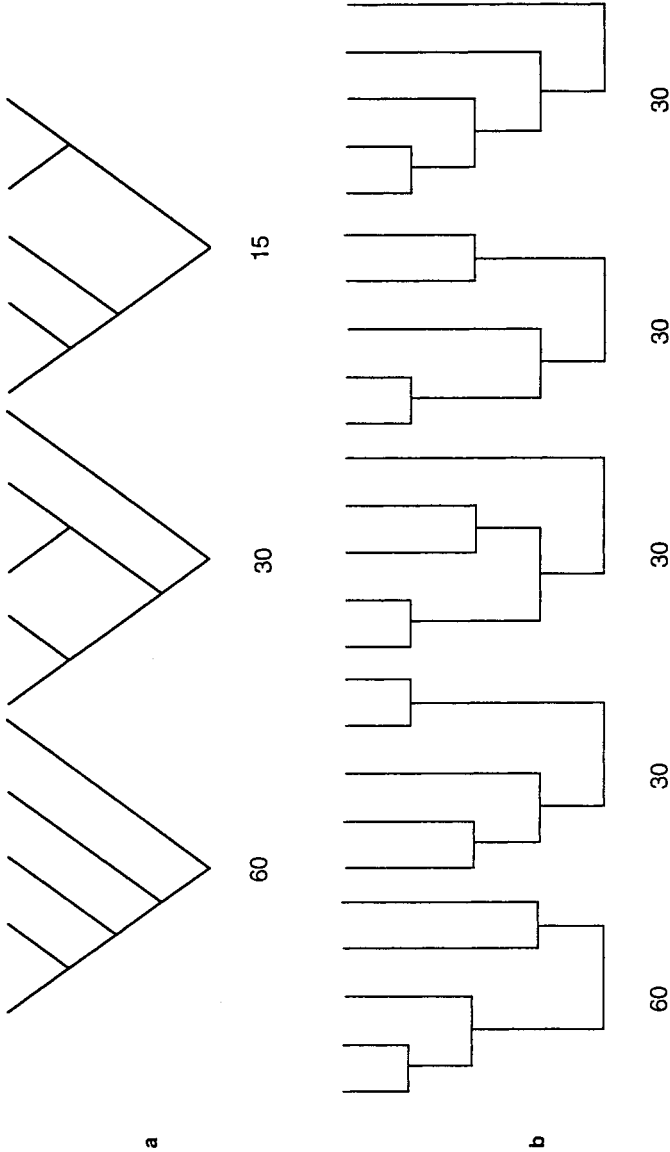


Figure 3. (a) Three binary topologies are distinguishable for 5 objects, while in (b) five dendrogram shapes are possible, notwithstanding labels. The figure under each form indicates the number of different labeling possibilities of each topology.

We used the completely random cophenetic matrix algorithm described in Lapointe and Legendre (1991, section 7) to generate 2000 dendrograms bearing five leaves. One thousand pairs of random dendrograms were thus obtained and compared. Larger dendrograms were not considered in our study since a significant difference among tests for five objects necessarily implies greater discrepancies with larger numbers of objects. Smaller numbers of objects were also ignored, because a comparison involving four objects (or fewer) will always accept the null hypothesis for a type I error of 0.05; the minimal probabilities would respectively be 1/12 for the random labels model, 1/15 for the random binary tree model, and 1/18 for the random dendrogram model (Table 1). In that case, even two identical dendrograms could not be declared more similar than what one would expect from pairs of randomly generated dendrograms.

The 1000 pairs were compared using the three different methods (random labels, random binary trees, and random dendrograms), including three binary tree distributions, for a total of 5000 comparison tests. Matrix correlation coefficients were computed between the cophenetic matrices or the cardinality matrices, depending on the test. The probabilities of the Mantel tests were obtained by complete enumeration of the distinguishable label permutations of the matrices. Exact tests could not be computed for the other methods, however. For five objects, one would need 11,025 binary tree comparisons and 32,400 dendrogram comparisons to compute exact probabilities. So, for each test, a reference distribution was constructed by sampling 10,000 trees at random from the relevant population to form 5,000 random pairs. These pairs consisted of random cophenetic matrices for the double-permutation test, and of random cardinality matrices for the three tests involving binary trees. The null hypothesis was rejected when the correlation for the actual matrix pair was larger than or equal to 95% of the correlations between random matrices (one-tailed test).

The results of the simulations are presented in Table 2. It becomes obvious when looking at this Table that the three methods compared are not equally appropriate for testing the correlation between random dendrograms.

- (1) The Mantel test, as applied to dendrograms, is too conservative in its conclusions. In every situation, it has underestimated the number of cases where the null hypothesis should have been rejected. The implication for comparison tests is that dendrograms are less likely to be declared similar when using the Mantel procedure than the selected type I error level. Furthermore, one should be careful in interpreting the results of this test since a significant Mantel test does not necessarily imply a significant double-permutation test; this results from the fact that a subset only of the set of possible dendrograms is considered by

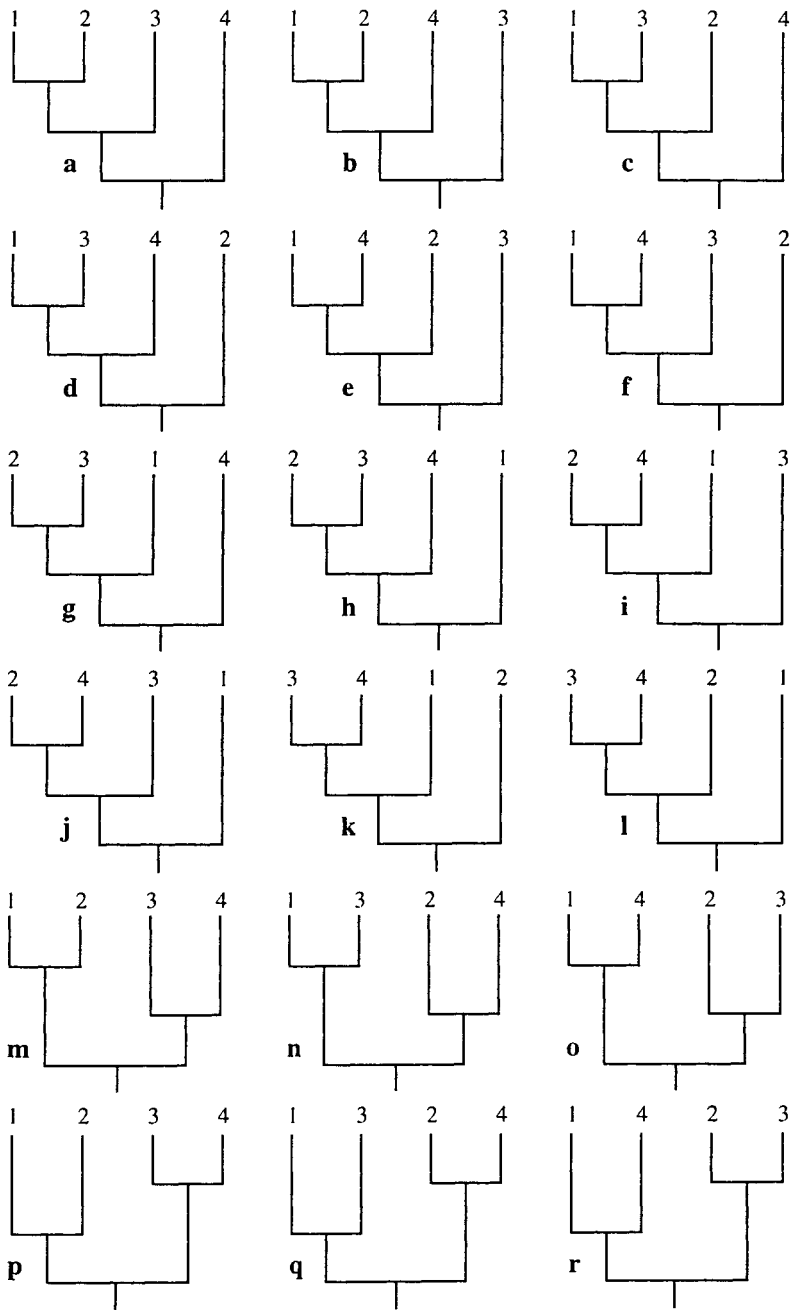


Figure 4. All possible distinguishable dendrograms for 4 objects (from Lapointe and Legendre, 1991).

- the Mantel test.
- (2) The three methods based on binary tree populations do better than the random label test in the simulations. Still, the expected probability of the null hypothesis is underestimated in all situations. The best binary tree model corresponds to the third distribution of Simberloff et al. (1981). On the other hand, the equiprobable topologies and equiprobable tree distributions are very similar. Even in the simple case exemplified in these simulations, all the binary tree methods greatly differ from expectations; it does not seem justified to retain any of these procedures for comparing dendrograms. We can expect larger discrepancies to appear when the population of dendrograms increases as a function of the number of objects.
 - (3) The double-permutation test is behaving well in these empirical simulations. The expected probabilities are well recovered by the method, and the recovered type I error is correct. This not only shows that cluster heights are an important aspect when comparing dendrograms. It also indicates that the constrained generation approach used in the double-permutation test (Lapointe and Legendre 1990) is appropriate to compare dendrograms generated by the completely random algorithm (Lapointe and Legendre 1991); this implies that permuting the actual cluster heights is statistically identical to generating a new set of cluster height values. Therefore, it seems justified to use the double-permutation method in real tests involving dendrograms, as well as the significance tables of the matrix correlation coefficient generated from completely random dendrograms (Lapointe and Legendre 1992).

6. Conclusion

Any comparison test for trees relies on a particular consensus index, a tree population to randomly draw from, and a tree distribution corresponding to the population. In this study, we have evaluated the effect of different tree populations and distributions on the matrix correlation test for dendrograms. Three aspects are important when comparing dendrograms: topology, label positions, and cluster heights. Random dendrograms can be generated using a permutation model considering only the label positions, a random binary tree model considering labels and topology, or a random model that takes into account all three aspects of a dendrogram. On theoretical grounds, it was shown that the dendrogram population includes the binary tree population and the labeling population as subsets. Therefore, the double-permutation test is more appropriate to compare dendrograms: it accounts for cluster heights and draws at random from the entire reference population. Still, we wanted to see whether dendrograms with random cluster heights could be compared

using a permutation approach, and if so, how better the results were, compared to other randomization methods. Our results showed that correlation tables (Lapointe and Legendre, 1992) generated by the "complete" generation algorithm (Lapointe and Legendre, 1991) can be used instead of actually doing the permutations.

References

- ARDISSON, P.-L., BOURGET, E., and LEGENDRE, P. (1990), "Multivariate Approach to Study Species Assemblages at Large Spatiotemporal Scales: The Community Structure of the Epibenthic Fauna of the Estuary and Gulf of St. Lawrence," *Canadian Journal of Fisheries and Aquatic Sciences*, 47, 1364-1377.
- BURGMAN, M.A. (1987), "An Analysis of the Distribution of Plants on Granite Outcrops in Southern Western Australia Using Mantel Tests," *Vegetatio*, 71, 79-86.
- CUERRIER, A., BARABÉ, D., and BROUILLET, L. (1992), "Bessey and Engler: A Numerical Analysis of their Classification of the Flowering Plants," *Taxon*, 41, 667-684.
- CZEKANOWSKI, J. (1909), "Zur Differentialdiagnose der Neandertalgruppe," *Korrespondenz-Blatt der deutschen Gesellschaft für Anthropologie, Ethnologie und Urgeschichte*, 40, 44-47.
- DAY, W. H. E. (1983), "Distribution of Distances Between Pairs of Classifications" in *Numerical Taxonomy*, Ed., J. Felsenstein, NATO Advanced Studies Institute, Ser. G. (Ecological Sciences) 1, Springer Verlag, Berlin, 127-131.
- DAY, W. H. E. (1986), "Analysis of Quartet Dissimilarity Measures Between Undirected Phylogenetic Trees," *Systematic Zoology*, 35, 325-333.
- DE WALL, F. B. M., and LUTTRELL, L. M. (1988), "Mechanisms of Social Reciprocity in Three Primate Species: Symmetrical Relationship Characteristics or Cognition?" *Ethology and Sociobiology*, 9, 101-118.
- FELSENSTEIN, J. (1978), "The Number of Evolutionary Trees," *Systematic Zoology*, 27, 27-33.
- FRANK, O., and SVENSSON, K. (1981), "On Probability Distributions of Single-Linkage Dendrograms," *Journal of Statistics and Computer Simulation*, 12, 121-131.
- FURNAS, G. W. (1984), "The Generation of Random, Binary Unordered Trees," *Journal of Classification*, 1, 187-233.
- HARDING, E. F. (1971), "The Probabilities of Rooted Tree-Shapes Generated by Random Bifurcation," *Advances in Applied Probability*, 3, 44-77.
- HARTIGAN, J. A. (1967), "Representation of Similarity Matrices by Trees," *Journal of the American Statistical Association*, 62, 1140-1158.
- HENDY, M. D., LITTLE, C. H. C., and PENNY, D. (1984), "Comparing Trees with Pendant Vertices Labelled," *SIAM Journal of Applied Mathematics*, 44, 1054-1065.
- HUBERT, L. J., and BAKER, F. B. (1977), "The Comparison and Fitting of Given Classification Schemes," *Journal of Mathematical Psychology*, 16, 233-253.
- HUBERT, L. J., and LEVIN, J. R. (1976), "Evaluating Object Set Partitions: Free-Sort Analysis and Some Generalizations," *Journal of Verbal Learning and Verbal Behavior*, 15, 459-470.
- HUDON, C., and LAMARCHE, G. (1989), "Niche Segregation Between American Lobster *Homarus americanus* and Rock Crab *Cancer irroratus*," *Marine Ecology Progress Series*, 52, 155-168.
- JOHNSON, S. C. (1967), "Hierarchical Clustering Schemes," *Psychometrika*, 32, 241-254.

- KRACKHARDT, D., and KILDUFF, M. (1990), "Friendship Patterns and Culture: The Control of Organizational Diversity," *American Anthropologist*, 92, 142-154.
- KRACKHARDT, D., and PORTER, L. W. (1986), "The Snowball Effect: Turnover Embedded in Communication Networks," *Journal of Applied Psychology*, 71, 50-55.
- KULCZYNSKI, S. (1928), "Die Pflanzenassoziationen der Pieninen," *Bulletin international de l'Académie polonaise des Sciences et des Lettres. Classe des Sciences mathématiques et naturelles, Série B, Supplément II (1927)*, 57-203.
- LAPOINTE, F.-J. (1992), "On the Congruence of Brain Evolution with Taxonomic Distances and Eco-ethological Affinities: A Statistical Evaluation," Unpublished Dissertation Thesis, Université de Montréal.
- LAPOINTE, F.-J., and LEGENDRE, P. (1990), "A Statistical Framework to Test the Consensus of Two Nested Classifications," *Systematic Zoology*, 39, 1-13.
- LAPOINTE, F.-J., and LEGENDRE, P. (1991), "The Generation of Random Ultrametric Matrices Representing Dendrograms," *Journal of Classification*, 8, 177-200.
- LAPOINTE, F.-J., and LEGENDRE, P. (1992), "Statistical Significance of the Matrix Correlation Coefficient for Comparing Independent Phylogenetic Trees," *Systematic Biology*, 41, 378-384.
- LAPOINTE, F.-J., and LEGENDRE, P. (1994), "A Classification of Pure Malt Scotch Whiskies," *Applied Statistics*, 43, 237-257.
- LEGENDRE, P., and FORTIN, M.-J. (1989), "Spatial Pattern and Ecological Analysis," *Vegetatio*, 80, 107-138.
- LUKASZEWICZ, J. (1951), "Sur la liaison et la division des points d'un ensemble fini," *Colloquium mathematicum*, 2, 282-285.
- MANTEL, N. (1967), "The Detection of Disease Clustering and a Generalized Regression Approach," *Cancer Research*, 27, 209-220.
- MURTAGH, F. (1984), "Counting Dendrograms: A Survey," *Discrete Applied Mathematics*, 7, 191-199.
- ODEN, N. L., and SHAO, K. T. (1984), "An Algorithm to Equiprobably Generate All Directed Trees with k Labeled Terminal Nodes and Unlabeled Interior Nodes," *Bulletin of Mathematical Biology*, 46, 379-387.
- PAGE, R. D. M. (1987), "Graphs and Generalized Tracks: Quantifying Croizat's Panbiogeography," *Systematic Zoology*, 36, 1-17.
- PAGE, R. D. M. (1988), "Quantitative Cladistic Biogeography: Constructing and Comparing Area Cladograms," *Systematic Zoology*, 37, 254-270.
- PAGE, R. D. M. (1990), "Temporal Congruence and Cladistic Analysis of Biogeography and Cospeciation," *Systematic Zoology*, 39, 205-226.
- PAGE, R. D. M. (1991), "Random Dendrograms and Null Hypotheses in Cladistic Biogeography," *Systematic Zoology*, 40, 54-62.
- PHIPPS, J. B. (1975), "The Numbers of Classifications," *Canadian Journal of Botany*, 54, 686-688.
- QUIROZ, A. J. (1989), "Fast Random Generation of Binary, t-ary, and Other Types of Trees," *Journal of Classification*, 6, 223-231.
- ROHLF, F. J. (1982), "Consensus Indices for Comparing Classifications," *Mathematical Biosciences*, 59, 131-144.
- ROHLF, F. J., and SOKAL, R. R. (1981), "Comparing Numerical Taxonomic Studies," *Systematic Zoology*, 30, 459-490.
- SAVAGE, H. M. (1983), "The Shape of Evolution: Systematic Tree Topology," *Biological Journal of the Linnean Society*, 20, 225-244.
- SCHNELL, G. D., DOUGLAS, M. E., and HOUGH, D. J. (1986), "Geographic Patterns of Variation in Offshore Spotted Dolphins (*Stenella attenuata*) of the Eastern Tropical Pacific Ocean," *Marine Mammal Science*, 2, 186-213.

- SIMBERLOFF, D. (1987), "Calculating Probabilities that Cladograms Match: A Method of Biogeographical Inference," *Systematic Zoology*, 36, 175-195.
- SIMBERLOFF, D., HECK, K. L., McCOY, E. D., and CONNOR, E. F. (1981), "There Have Been no Statistical Tests of Cladistic Biogeographical Hypotheses," in *Vicariance Biogeography: A Critique*, Eds., G. Nelson and D. Rosen, Columbia University Press, New York, 40-63.
- SHAO, K., and ROHLF, F. J. (1983), "Sampling Distributions of Consensus Indices when all Bifurcating Trees are Equally Likely" in *Numerical Taxonomy*, Ed., J. Felsenstein, NATO Advanced Studies Institute, Ser.G. (Ecological Sciences) 1, Springer Verlag, Berlin, 132-137.
- SHAO, K., and SOKAL, R. R. (1986), "Significance Tests of Consensus Indices," *Systematic Zoology*, 35, 582-590.
- SNEATH, P. H. A. (1957), "The Application of Computers to Taxonomy," *Journal of General Microbiology*, 17, 201-226.
- SNEATH, P. H., and SOKAL, R. R. (1973), *Numerical Taxonomy*, San Francisco: W.H. Freeman and Co.
- SOKAL, R. R. (1979), "Testing Statistical Significance of Geographic Variation Patterns," *Systematic Zoology*, 28, 227-232.
- SOKAL, R. R., and MICHENER, C. D. (1958), "A Statistical Method for Evaluating Systematic Relationships," *University of Kansas Science Bulletin*, 3, 1409-1438.
- SOKAL, R. R., SMOUSE, P. E., and NEEL, J. V. (1986), "The Genetic Structure of a Tribal Population, the Yanomama Indians. XV. Patterns Inferred by Autocorrelation Analysis," *Genetics*, 114, 259-287.
- SOKAL, R. R., and SNEATH, P. H. A. (1963), *Principles of Numerical Taxonomy*, San Francisco: W.H. Freeman and Co.
- SOKAL, R. R., and UNNASCH, R. S. (1988), "Geographic Covariation of Hosts and Parasites: Evidence from *Populus* and *Pemphigus*," *Zeitschrift für zoologische Systematik und Evolutionsforschung*, 26, 73-88.
- SOKAL, R. R., UYTTERSCHAUT, H., RÖSING, F.W., and SCHWIDETZKY, I. (1987), "A Classification of European Skulls from Three Time Periods," *American Journal of Physical Anthropology*, 74, 1-20.
- SOKAL, R. R., and WARTENBERG, D. E. (1983), "A Test of Spatial Autocorrelation Using an Isolation-by-Distance Model," *Genetics*, 105, 219-237.
- SØRENSEN, T. (1948), "A Method of Establishing Groups of Equal Amplitude in Plant Sociology Based on Similarity of Species Content and its Application to Analysis of the Vegetation on Danish Commons," *Biologiske Skrifter*, 5, 1-34.
- STEEL, M. A. (1988), "Distribution of the Symmetric Difference Metric on Phylogenetic Trees," *SIAM Journal of Discrete Mathematics*, 1, 541-551.
- STEEL, M. A. and PENNY, D. (1993), "Distributions of Tree Comparison Metrics - Some New Results," *Systematic Biology*, 42, 126-141.