

Assessing Congruence Among Ultrametric Distance Matrices

Véronique Campbell

Université de Montréal, Canada

Pierre Legendre

Université de Montréal, Canada

François-Joseph Lapointe

Université de Montréal, Canada

Abstract: Recently, a test of congruence among distance matrices (CADM) has been developed. The null hypothesis is the incongruence among all data matrices. It has been shown that CADM has a correct type I error rate and good power when applied to independently-generated distance matrices. In this study, we investigate the suitability of CADM to compare ultrametric distance matrices. We tested the type I error rate and power of CADM with randomly generated dendrograms and their associated ultrametric distance matrices. We show that the test has correct type I error rates and good power. To obtain the significance level of the statistic, a single (as in the Mantel test) or a double (as in the double permutation test, DPT) permutation procedure was used. The power of CADM remained identical when the two permutation methods were compared. This study clearly demonstrates that CADM can be used to determine whether different dendrograms convey congruent information.

Keywords: Incongruence; Power; Statistical test; Simulation; Type I error rate.

We would like to thank the members of the LEMEE (Laboratoire d'Écologie Moléculaire et d'Évolution) of Université de Montréal for their constructive comments on a preliminary version of this manuscript as well as three anonymous reviewers. This study was supported by NSERC and FQRNT scholarships to VC and by NSERC grants OGP0007738 to PL and OGP0155251 to FJL.

Corresponding Author's Address : Véronique Campbell Département de sciences biologiques, Université de Montréal, C.P. 6128, succ. Centre-ville, Montréal, Québec, H3C 3J7, Canada, email: veronique.campbell@umontreal.ca

1. Introduction

Often, in classification studies, different sets of variables are used to derive dendrograms for the same set of objects. Depending on the set of variables, classifications may differ. Therefore, it is important to know to which extent the information conveyed by each set of variables is congruent to the others. Also, when different classifications of the same objects are available but the information on the variables used is not, assessing the degree of resemblance or congruence of different classifications may be of interest.

Congruence or incongruence tests, depending on how the null hypothesis is postulated, have been extensively studied. Planet (2006) has recently classified congruence tests in two categories: those based on character information and those based on tree shape or topology. Character congruence tests compare the fit of the data on two competing trees (e.g. ILD: Mickevich and Farris 1981; Farris, Källersjö, Kluge, and Bult 1994; Templeton Test: Templeton 1983; and T-PTP: Faith 1991). These tests will not be reviewed further here, given that they present a different approach than the one discussed in this study. In contrast, topological congruence tests compare the branching pattern (topology) of different trees without considering the underlying data. Such tests are based on numerical measurements of topological difference obtained from indices calculated on consensus trees (e.g. Consensus Fork Index: Colless 1980; Rohlf Consensus Index: Rohlf 1982) or from tree distances (e.g. Partition Metric: Robinson and Foulds 1981; Quartets Distance: Estabrook, McMorris and Meacham 1985; and Path Difference Metric: Steel and Penny 1993). Significance testing is generally possible by comparing the statistic to a reference distribution generated by permutations (e.g. Steel and Penny 1993) or by using non-parametric bootstrap of the original data (Page 1996).

Along with different indices that have been proposed to quantify the similarity between dendrograms, a classical approach is to calculate a cophenetic (or matrix) correlation coefficient between two ultrametric matrices representing dendrograms (Sokal and Rohlf 1962). If the dendrograms come from independent data tables, the null hypothesis of a correlation equal to zero can be tested using a Mantel's generalized permutation test strategy, where only the object labels are permuted (Mantel 1967). Amongst other, a double-permutation test (DPT: Lapointe and Legendre 1990), which takes into account the topology, label positions and cluster heights of the dendrograms, has also been proposed (Lapointe and Legendre 1995; see also Podani 2000). It has been shown that only DPT provides correct rates of type I error when a correlation between a pair of dendrograms is used as the test statistic (Lapointe and Legendre 1995).

Although many congruence tests were developed in a phylogenetic context, they are often used in other fields such as ecology, anthropology, archaeology, sociology and classification (Legendre and Lapointe 2004). Unfortunately, the majority of these tests only apply to the comparison of two datasets (or matrices) at a time. Legendre and Lapointe (2004) described a test of congruence among distance matrices (CADM) that is applicable to more than two matrices. Based on Kendall's W concordance statistic, CADM is an extension of the Mantel test that can be used to assess congruence of multiple matrices. CADM presents several advantages with respect to other congruence tests. 1) The statistic is calculated directly from the distance matrices; thus different types of data can be compared if converted to distance matrices using an appropriate function. 2) The matrices can be weighted differentially if needed. 3) *A posteriori* tests can be performed to discriminate incongruent from congruent matrices. Previous simulations have shown that the global and *a posteriori* CADM tests have a correct rate of type I error and good power when applied to independently-generated distance matrices (Legendre and Lapointe 2004). In this study, we have tested the type I error rate and power of the global and *a posteriori* tests of CADM using randomly generated dendrograms and their associated ultrametric distance matrices. To assess the significance of the statistic, we tested two different permutation procedures: a simple Mantel's permutation test (Mantel 1967) and a double-permutation test (DPT: Lapointe and Legendre 1990).

2. CADM Test

The null hypothesis (H_0) for the global CADM test is the incongruence of all distance matrices (Legendre and Lapointe 2004). That is, matrices are statistically independent from each other and convey distinct information about the relationships among the objects under study. Rejecting H_0 indicates that at least two matrices contain congruent information. In those cases, *a posteriori* CADM tests can be performed to determine the contribution of each matrix to the overall congruence. *A posteriori* tests can be used to identify incongruent and congruent matrices in a set, but it does not specify the pairs or groups of congruent matrices. To this end, complementary Mantel tests based upon ranks can be used. Following that, congruent matrices can be combined in a classification analysis. A summary of the computations to perform the CADM test follows:

1. The upper off-diagonal section of each distance matrix is unfolded and written into a vector corresponding to row i in a worktable.
2. The entries of each row are transformed into ranks according to their values.

3. The sum of ranks R_j is calculated for each column j of the table.
4. The mean \bar{R} of all R_j values is calculated.
5. Kendall's coefficient of concordance (W) is computed using the following formula:

$$W = \frac{12S}{p^2(n^3 - n) - pT},$$

where p is the number of matrices, n is the number of objects in each matrix, S is obtained using:

$$S = \sum_{j=1}^n (R_j - \bar{R})^2,$$

and T is a correction factor for tied ranks:

$$T = \sum_{k=1}^m (t_k^3 - t_k),$$

in which t_k is the number of tied ranks for each (k) of m groups of ties.

6. W is transformed into a Friedman's χ^2 , which is a pivotal statistic appropriate for testing, using the following formula:

$$\chi^2 = p(n-1)W.$$

The observed Friedman's χ^2 (χ_{ref}^2) is tested against a distribution of the statistic obtained under permutation (χ^{2*}). For the global CADM test, all matrices are permuted at random, whereas for *a posteriori* tests only the matrix tested is permuted. After p_n permutations, the one-tailed probability of the data under H_0 is computed as the number of χ^{2*} values greater than or equal to χ_{ref}^2 divided by $(p_n - 1)$. In *a posteriori* comparisons, the P-value should be adjusted to maintain an adequate experimentwise error rate using a method designed specifically to correct for multiple testing (e.g. Holm 1979). Two different permutation models were compared in this study (see section 3.1 in the Simulation Procedure). More details about the CADM method can be found in Legendre and Lapointe (2004).

3. Simulation Procedure

Computer simulations were performed to assess the Type I error (α) rate and power of CADM when applied to ultrametric distance matrices. The type I error rate is the probability of incorrectly rejecting a true H_0 and should not be larger than the nominal significance level (α) of the test

(Edgington 1995). The type II error (β) rate refers to the probability of failing to reject a false H_0 . The power of the test is the rate of rejection of a false H_0 (i.e. $1 - \beta$).

3.1 Global CADM Test

We generated, at random, independent ultrametric distance matrices (IM) representing dendrograms, according to the completely random ultrametric matrix algorithm described by Lapointe and Legendre (1991, Sect. 7). To examine a range of different parameter values, the number of objects within each matrix ($n = 5, 10, 20$ and 50) as well as the number of independent ultrametric matrices (IM = 2, 3, 4, 5 and 10) varied. Comparing IM corresponds to a situation where H_0 is “true” by construct (i.e. all the dendrograms are incongruent, see Figure 1A). To estimate the type I error rate, the rejection rate (i.e. the proportion of replicates for which the “true” H_0 was rejected) was calculated at different significance levels ($\alpha = 0.01, 0.05, 0.10, 0.20, 0.30, 0.40, 0.50, 0.60, 0.70, 0.80, 0.90$) along with its 95% confidence interval (CI) computed according to a binomial distribution. A thousand replicate simulations were performed in each case. For each replicate, 999 random permutations of the dendrograms were computed to construct the reference distribution for significance testing. Two different permutation models were compared to obtain the reference distribution: Mantel (χ_M^{2*}) and DPT (χ_{DPT}^{2*}). For this study, we included the DPT procedure as an option in the CADM program (available at www.bio.umontreal.ca/casgrain/en/labo/cadm.html).

To estimate the power of CADM, congruent ultrametric distance matrices (CM) were generated (see Figure 1B). CM are partially similar distance matrices that were generated by permutation (described below) of an original random ultrametric distance matrix. Figure 1 illustrates the steps involved in simulations to test the type I error and power of CADM with a set of five ultrametric distance matrices. Two different permutation procedures were used to generate CM. (1) A fixed number of randomly chosen objects were permuted on the dendrogram, corresponding to the permutation of rows and columns (labels) within the ultrametric matrix. This is similar to the type of permutation that is performed in the Mantel test but restricted to some objects only (CM_M). (2) A fixed number of randomly chosen objects and cluster heights were permuted so that the dendrogram topology and the objects were permuted. This second permutation approach to construct CM is similar to that used in the DPT test but restricted to some objects only (CM_{DPT}). The generated CM were more or less congruent depending on the number of objects that were permuted (see Figure 2). For power simulations, the proportion of permuted objects was identical regardless of the matrix sizes (i.e. 40%

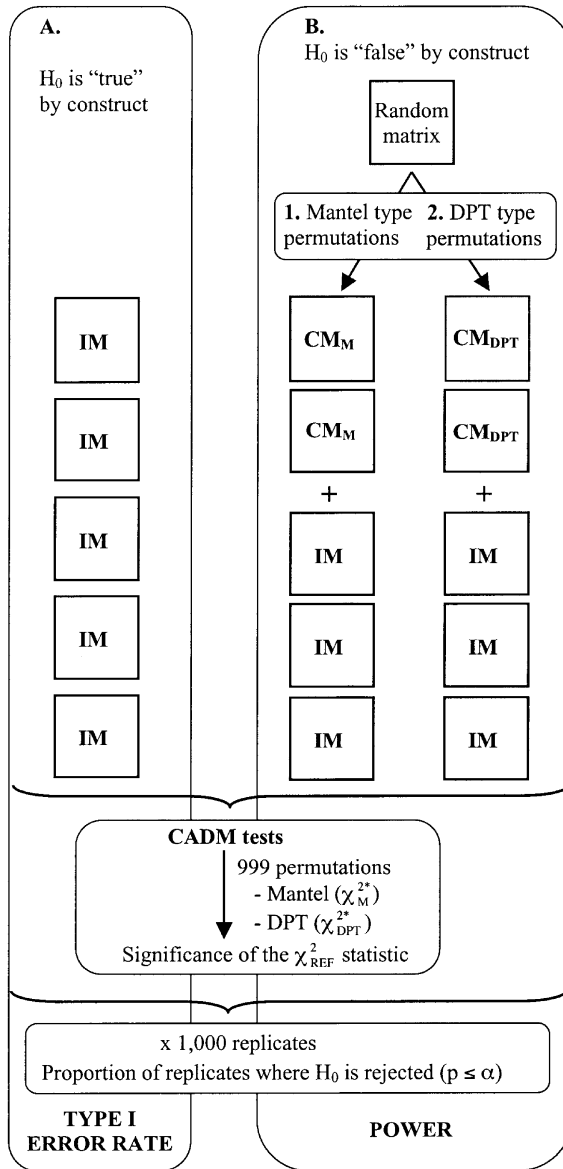


Figure 1. Flowchart of the protocol used to estimate type I error rate and power of CADM, for five ultrametric distance matrices. A. H_0 is “true” by construct, and it includes five incongruent ultrametric distance matrices (IM). B. H_0 is “false” by construct, and it includes two partially similar matrices (CM) and three randomly generated matrices (IM). (1) In Mantel type permutations, CM were generated by permuting rows and columns of an initial matrix (CM_M). (2) For DPT, CM were generated by permuting rows and columns as well as cluster heights of an initial matrix (CM_{DPT}). CADM tests were performed on each set separately to estimate the type I error rate in A and power in B.

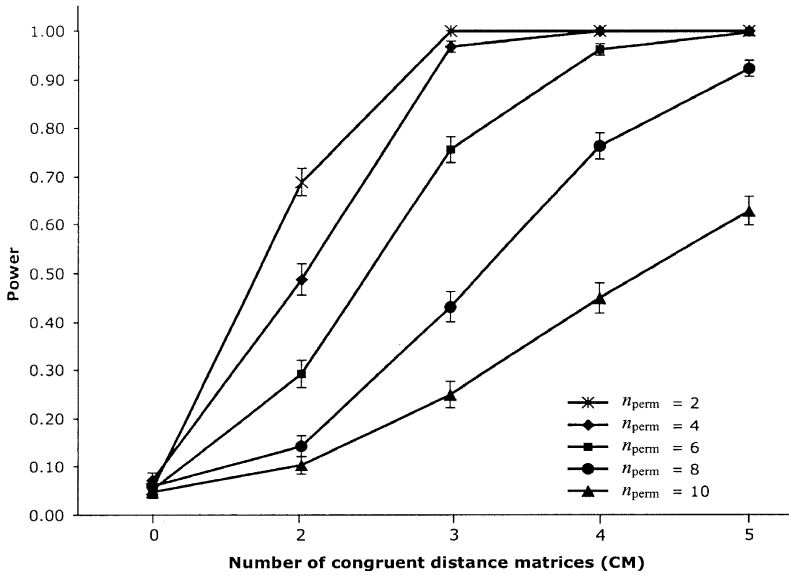


Figure 2. Estimated power (mean and 95% CI) obtained in simulations of global CADM tests using Mantel permutations, for $\alpha = 0.05$, and $n = 20$, with different levels of congruence among CM (represented by different symbols). When CM = 0, all matrices are incongruent and thus H_0 is “true” by construct. For $CM \geq 2$, H_0 is “false”.

with $n_{perm} = 2, 4, 8$ or 20 for matrices of size $n = 5, 10, 20$ and 50 respectively).

In each trial, the total number of distance matrices was fixed to either five or ten, but the number of CM and IM varied. When $CM = 0$ ($IM = 5$ or 10), H_0 is “true”, otherwise H_0 is “false” by construct ($CM \geq 2$). The power of the test, which corresponds to the proportion of replicates where H_0 is rejected when false, was calculated for each combination of parameters. For each replicate of the CADM test, 999 random permutations were computed to estimate the reference distribution using the Mantel (χ_M^{2*}) and DPT (χ_{DPT}^{2*}) randomization procedures. H_0 was rejected when χ_{ref}^2 was greater than or equal to 95% of the χ^{2*} (which corresponds to a one-tailed test with an alpha level of 5%). The rejection rate of H_0 (out of 1000 replicates) was calculated along with its 95% confidence interval (CI).

3.2 A Posteriori CADM Tests

Simulations have also been performed to assess the type I error rate and power of *a posteriori* CADM tests. The H_0 in such cases is the incongruence of the matrix subjected to the test with respect to all other

Table 1. CADM Type I Error Rates Obtained for the Global Tests on Pairs of Ultrametric Distance Matrices (IM = 2), for Different Numbers of Objects (n). The corresponding 95% confidence intervals are in parentheses.

Permutation Models	n	Significance Levels		
		0.01	0.05	0.10
Mantel				
	5	0.0003 (0.0002 – 0.0005)	0.033 (0.032 – 0.035)	0.079 (0.077 – 0.081)
	10	0.001 (0.0005 – 0.002)	0.053 (0.049 – 0.058)	0.100 (0.094 – 0.106)
	20	0.010 (0.008 – 0.013)	0.050 (0.044 – 0.056)	0.100 (0.092 – 0.109)
DPT				
	5	0.008 (0.007 – 0.009)	0.042 (0.040 – 0.043)	0.095 (0.092 – 0.097)
	10	0.009 (0.007 – 0.011)	0.050 (0.045 – 0.054)	0.100 (0.095 – 0.110)
	20	0.013 (0.010 – 0.016)	0.054 (0.048 – 0.060)	0.100 (0.093 – 0.110)

matrices. Therefore, only the matrix subjected to the test is permuted. The sets of five ultrametric distance matrices generated to assess power of the global test were also used for *a posteriori* CADM tests ($n = 5, 10$ and 20). Again, the rejection rate of H_0 , out of 1000 replicates, was calculated along with its 95% confidence interval (CI) for an alpha level of 0.05. For each replicate of the CADM test, 999 random permutations were computed to estimate the reference distribution using Mantel (χ_M^{2*}) and DPT (χ_{DPT}^{2*}) randomization procedures.

4. Simulation Results

The results in Table 1 show that the CADM test underestimated the number of cases where H_0 should have been rejected when $IM = 2$ and $n = 5$; the 95% CI of the type I error rate did not include the nominal significance level (α) when compared to a χ_M^{2*} distribution. Similar results were observed when using a χ_{DPT}^{2*} distribution; however the type I error

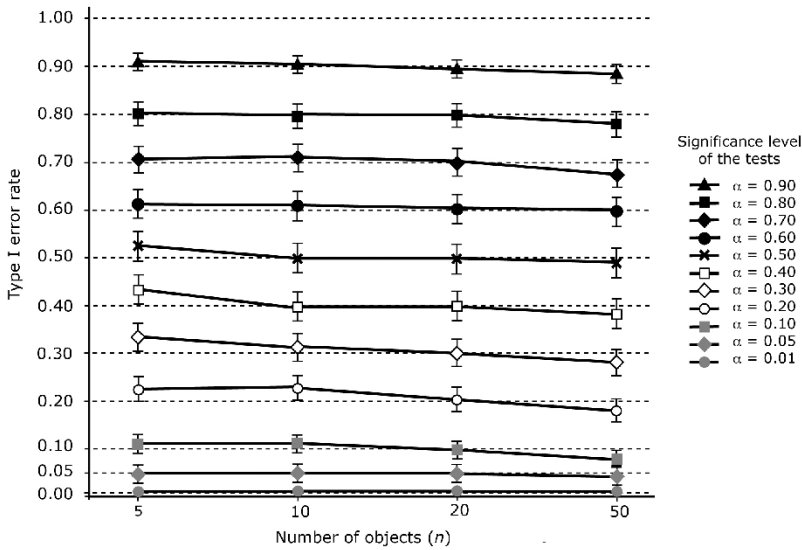


Figure 3. Type I error rates obtained in simulations of global CADM tests using Mantel permutations. Results are shown for tests involving ten IM and varying numbers of objects ($n = 5, 10, 20$ and 50) along with different significance levels (α). The corresponding 95% CI (bars) are provided for each rejection rate (based on 1,000 replicates).

rate and its 95% CI were closer to the nominal values. Nevertheless, a test whose error rate under H_0 is lower than the alpha level remains valid (Edgington 1995). When $IM = 2$ and $n > 5$ or when IM was larger than two (i.e. $IM = 3, 4, 5$ and 10), the global CADM test had an adequate estimated Type I error rate for both types of permutations. Type I error rates obtained with 10 IM ($n = 5, 10, 20$ and 50) are shown in Figure 3. The 95% CI of the rejection rates included the nominal significance level (α) in nearly all cases.

Power was estimated by calculating the proportion of replicates where H_0 was rejected when H_0 was “false” by construct. Simulation results were nearly identical when CADM was tested using CM_M or CM_{DPT} , thus only the CM_M results will be reported here. Power obtained with CADM when different numbers of matrices were included in the analysis and for matrices with varying number of objects is shown in Figure 4. For $IM + CM = 5$ (Figure 4A) and $IM + CM = 10$ (Figure 4B), an increase in power was observed with (1) an increase in the number of objects and (2) an increase in the number of CM relative to the total number of matrices. For sets of five matrices, a power of 1.0 was obtained only when matrices of 50 objects were used, whereas a power of 1.0 was obtained with smaller size matrices (i.e. 20 objects) for sets of ten matrices. Thus, power was higher when the total number of matrices

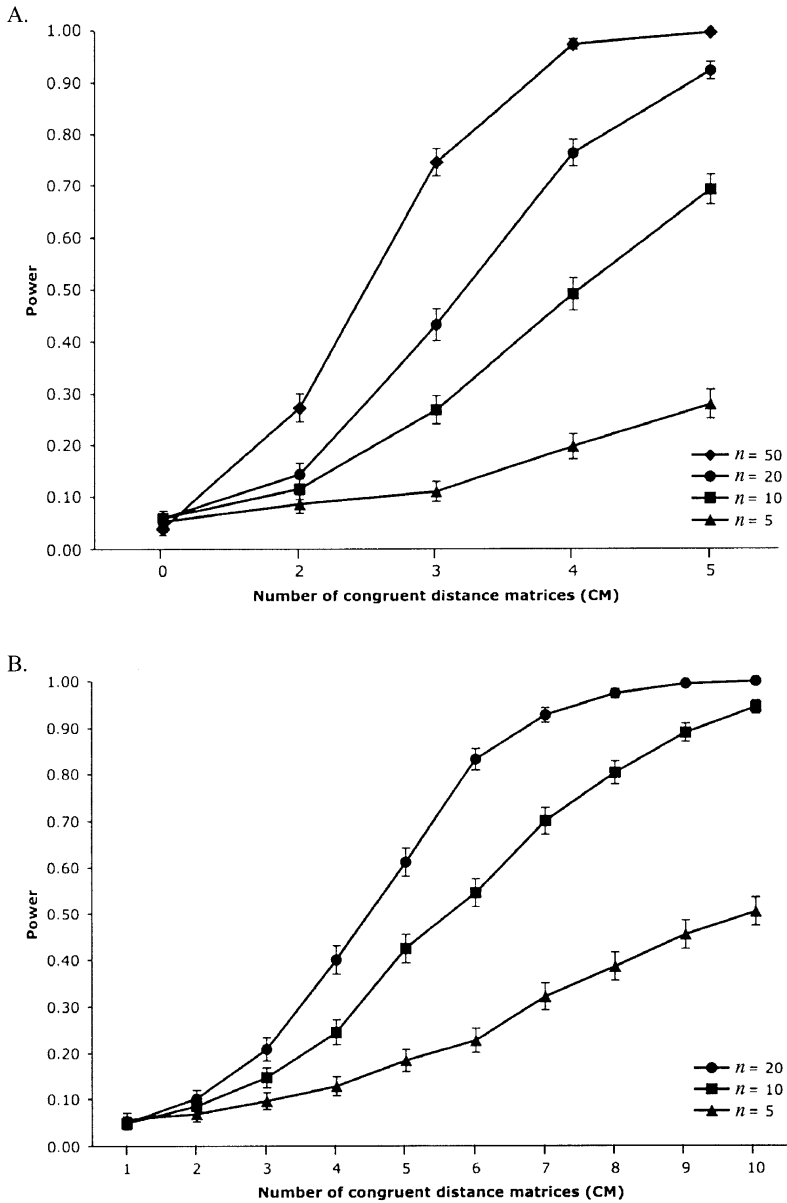


Figure 4. Estimated power (mean and 95% CI) obtained in simulations of global CADM tests using Mantel permutations, for different numbers of objects, and $\alpha = 0.05$. When $CM = 0$, all matrices are incongruent and thus H_0 is “true” by construct. For $CM \geq 2$, H_0 is “false”. A. Five distance matrices ($IM + CM = 5$) in each set with different numbers of objects. B. Ten distance matrices ($IM + CM = 10$) in each set with different numbers of objects.

included in the analysis was larger. Also, a comparison of Figures 4A and 4B reveals that for a given number of CM, power was higher when the number of IM was lower. Hence, it was easier to detect four congruent matrices out of five than four out of ten. This trend was accentuated when the size of the matrices increases.

Simulation results of *a posteriori* CADM tests are presented in Figure 5. In *a posteriori* comparisons, only the matrix that was subjected to the test was permuted. Therefore, rejection rates were obtained for each matrix permuted individually. For simplicity, power is only shown for matrix number 1 (i.e. only the first matrix was permuted, Figure 5A) and matrix number 5 (i.e. only the fifth matrix was permuted, Figure 5B). When H_0 was “true” (CM = 0), the rejection rate for the permuted matrix was close to 0.05, which was expected at the α level of 0.05 used to perform the tests. When $CM \geq 2$, the rejection rate was greater than 0.05 when the permuted matrix was a CM but it was near 0.05 when the permuted matrix was an IM. Consequently, matrix number 1, which was congruent by construct with matrix number 2 in all cases except when $CM = 0$, showed a rejection rate greater than 0.05 when $CM = 2, 3, 4$ or 5 (Figure 5A). In contrast, matrix number 5, which was congruent with the other matrices in the set only when $CM = 5$, showed a rejection rate greater than 0.05 only when $CM = 5$ (Figure 5B). Figure 5 illustrates that *a posteriori* CADM tests have an accurate type I error rate when tested at an α level of 0.05, regardless of the number of CM versus IM, as observed when H_0 was “true” by construct. Similarly to the results obtained for the global CADM test, power was good and increased with the number of objects and with the number of CM.

5. Discussion

CADM represents a powerful approach to test whether some matrices are incongruent to others. As opposed to the Mantel test, which compares matrices in a pairwise fashion, the global CADM test allows for comparisons among multiple matrices in a single analysis, without the need for a multiple testing correction (Legendre and Lapointe 2004). Our results support and generalize those of Legendre and Lapointe (2004), where the type I error and power of CADM was tested using random distance matrices. The results presented in this paper clearly show that CADM can be used to assess congruence among ultrametric distance matrices associated to dendrograms. The simulation results revealed that both the global and *a posteriori* CADM tests have correct type I error rates and good power when tested on ultrametric distance matrices, using either Mantel permutations or DPT. In classification studies, it can be used to determine if dendrograms defined on the same set of objects are congruent

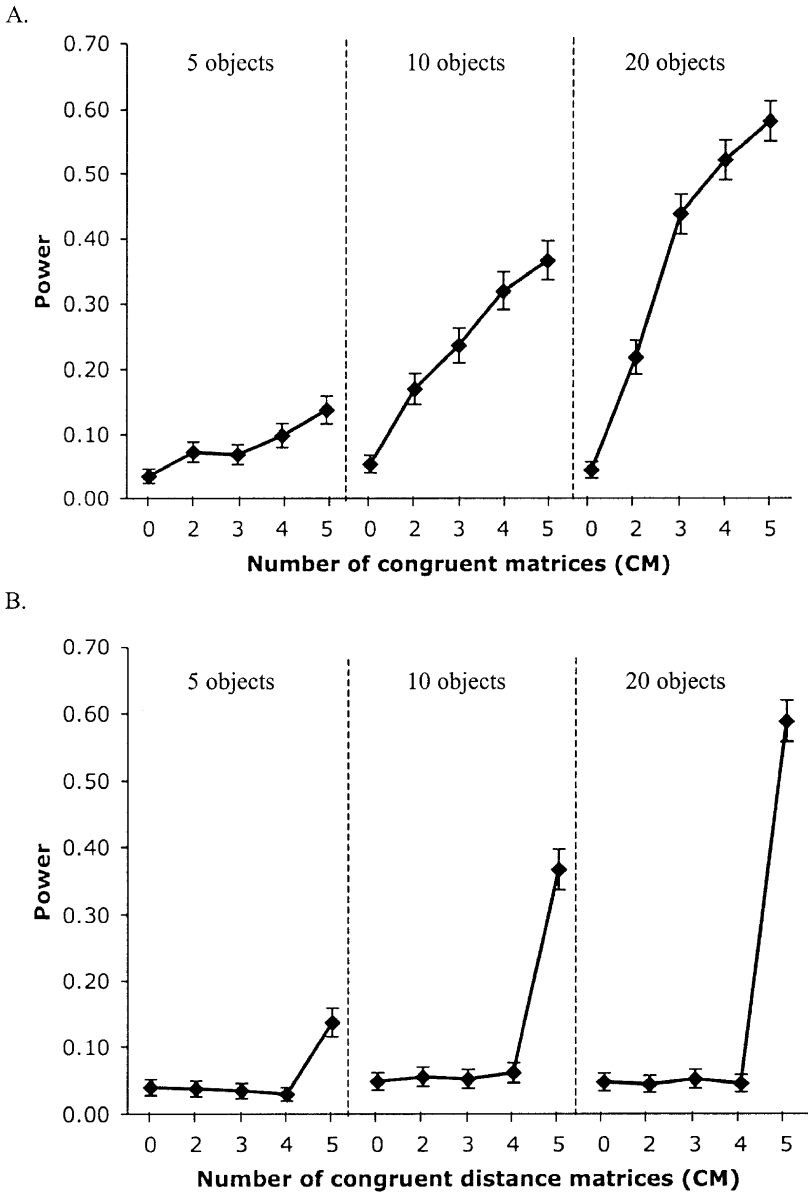


Figure 5. Estimated power (mean and 95% CI) obtained in simulations of *a posteriori* CADM tests, for different numbers of objects, and $\alpha = 0.05$. CM is the number of congruent matrices (in all cases IM + CM = 5). When CM=0, all matrices are incongruent and thus H_0 is “true” by construct. For CM ≥ 2 , H_0 is “false”. CM are numbered so that when CM = 2, matrices 1 and 2 are congruent. The symbols correspond to rejection rates of H_0 obtained when a given matrix was permuted. A. Matrix number 1 was permuted. B. Matrix number 5 was permuted.

and thus support a similar classification. In cases where some dendrograms are incongruent, *a posteriori* tests can be used to determine which ones provide different information about the classification. Incongruent dendrograms can thus be compared to identify conflicting parts of the classification. Otherwise, dendrograms can be combined to derive a global classification using a consensus method.

The difference between the Mantel and the DPT randomization procedures derives from different aspects of the dendrogram being permuted (Lapointe and Legendre 1995). The Mantel procedure permutes the label positions on a fixed topology whereas DPT also permutes cluster heights, thus changing the tree topology. Therefore, the permutation set is larger when using DPT as a randomization procedure. Hence, for five objects, there are 60 different labelings of a topology, while 180 different dendrograms are possible when cluster heights are also randomized. For the global CADM significance test, all the dendrograms are randomized and the number of permutation possibilities increases exponentially with the number of dendrograms compared. Lapointe and Legendre (1995) have suggested that the DPT test might be more appropriate to compare dendrograms since a statistical bias may be introduced when using Mantel randomization, which samples only a subset of the reference distribution. While evaluating the Mantel test and DPT as testing procedures to compare correlations between random pairs of dendrograms, Lapointe and Legendre (1995) have shown that the Mantel test was more conservative than DPT and that only the latter test provided unbiased type I error. They concluded that the Mantel randomization procedure was more likely to overlook congruent matrices. However, their study did not compare matrices that included more than five objects.

In our study, a Friedman's χ^2 statistic was used instead of a correlation coefficient because the test of concordance compares several matrices simultaneously, and matrices of different sizes were tested. Similarly to Lapointe and Legendre (1995), DPT randomization provided an improved type I error rate over the Mantel permutation when matrix pairs of five objects were compared. However, our results suggest that both permutation methods produce unbiased type I error rates when larger matrices ($n > 5$) are compared or when more than two matrices are included in the analysis, probably because a relatively small subset of the tree space is sampled with both types of randomizations when using 1,000 permutations.

No significant difference in power was observed between the Mantel and DPT randomization procedures. Power curves were similar to those obtained by Legendre and Lapointe (2004) where CADM was tested with random distance matrices. That is, an increase in power was observed with (1) an increase in matrix size and (2) an increase in the number of

CM relative to the total number of matrices. Also, for a given number of CM, power was higher when the number of IM was lower. In this study, good power was generally achieved, even though 40% of objects were permuted in each matrix. Power was further increased when comparing more congruent CM, i.e. CM that were generated using fewer permuted objects (as shown on Figure 2). Although DPT is included as an option in the CADM program (see the method section), we recommend to use Mantel permutations even when comparing ultrametric distance matrices since it is less time consuming and performs identically to the DPT randomization, except in the particular case of two IM and five objects.

In an effort to further demonstrate the utility of CADM in different fields, we are currently testing its performance with additive distance matrices corresponding to phylogenetic trees. Further development may also include a generalization to allow comparisons of partially overlapping trees and matrices, which could be used for supertree methods.

References

- COLLESS, D.H. (1980), "Congruence between Morphometric and Allozyme Data for Menidia Species - A Reappraisal," *Systematic Zoology*, 29, 289–299.
- EDGINGTON, E.S. (1995), *Randomization Tests* (3rd ed.), New York: Marcel Dekker.
- ESTABROOK, G.F., MCMORRIS, F.R., and MEACHAM, C.A. (1985), "Comparison of Undirected Phylogenetic Trees Based on Subtrees of Four Evolutionary Units," *Systematic Zoology*, 34, 193–200.
- FAITH, D.P. (1991), "Cladistic Permutation Tests for Monophyly and Nonmonophyly," *Systematic Zoology*, 40, 366–375.
- FARRIS, J.S., KÄLLERSJÖ, M., KLUGE, A.G., and BULT, C. (1994), "Testing Significance of Incongruence," *Cladistics*, 10, 315–319.
- HOLM, S. (1979), "A Simple Sequentially Rejective Multiple Test Procedure," *Scandinavian Journal of Statistics*, 6, 65–70.
- 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. (1994), "A Classification of Pure Malt Scotch Whiskies," *Applied Statistics*, 43, 237–257.
- LAPOINTE, F.J., and LEGENDRE, P. (1995), "Comparison Tests for Dendrograms: A Comparative Evaluation," *Journal of Classification*, 12, 265–282.
- LEGENDRE, P., and LAPOINTE, F.J. (2004), "Assessing Congruence among Distance Matrices: Single-Malt Scotch Whiskies Revisited," *Australian and New Zealand Journal of Statistics*, 46, 615–629.
- MANTEL, N. (1967), "The Detection of Disease Clustering and a Generalized Regression Approach," *Cancer Research*, 27, 209–220.
- MICKEVICH, M.F., and FARRIS, J.S. (1981), "The Implications of Congruence in Menidia," *Systematic Zoology*, 30, 351–370.
- PAGE, R.D.M. (1996), "On Consensus, Confidence, and Total Evidence," *Cladistics*, 12, 83–92.
- PENNY, D., WATSON, E.E., and STEEL, M.A. (1993), "Trees from Languages and Genes Are Very Similar," *Systematic Biology*, 42, 382–384.

- PLANET, P.J. (2006), "Tree Disagreement: Measuring and Testing Incongruence in Phylogenies," *Journal of Biomedical Informatics*, 39, 86–102.
- PODANI, J. (2000), "Simulation of Random Dendrograms and Comparison Tests: Some Comments," *Journal of Classification*, 17, 123–142.
- ROBINSON, D.F., and FOULDS, L.R. (1981), "Comparison of Phylogenetic Trees," *Mathematical Biosciences*, 53, 131–147.
- ROHLF, F.J. (1982), "Consensus Indices for Comparing Classifications," *Mathematical Biosciences*, 59, 131–144.
- SOKAL, R.R., and ROHLF, F.J. (1962), "The Comparison of Dendrograms by Objective Methods," *Taxon*, 11, 33–40.
- STEEL, M.A., and PENNY, D. (1993), "Distributions of Tree Comparison Metrics - Some New Results," *Systematic Biology*, 42, 126–141.
- TEMPLETON, A.R. (1983), "Phylogenetic Inference from Restriction Endonuclease Cleavage Site Maps with Particular Reference to the Evolution of Human and the Apes," *International Journal of Organic Evolution*, 37, 221–244.