

Pedro R. Peres-Neto · Donald A. Jackson

## How well do multivariate data sets match? The advantages of a Procrustean superimposition approach over the Mantel test

Received: 7 September 2000 / Accepted: 11 April 2001 / Published online: 8 June 2001  
© Springer-Verlag 2001

**Abstract** The Mantel test provides a means to test the association between distance matrices and has been widely used in ecological and evolutionary studies. Recently, another permutation test based on a Procrustes statistic (PROTEST) was developed to compare multivariate data sets. Our study contrasts the effectiveness, in terms of power and type I error rates, of the Mantel test and PROTEST. We illustrate the application of Procrustes superimposition to visually examine the concordance of observations for each dimension separately and how to conduct hypothesis testing in which the association between two data sets is tested while controlling for the variation related to other sources of data. Our simulation results show that PROTEST is as powerful or more powerful than the Mantel test for detecting matrix association under a variety of possible scenarios. As a result of the increased power of PROTEST and the ability to assess the match for individual observations (not available with the Mantel test), biologists now have an additional and powerful analytical tool to study ecological and evolutionary relationships.

**Keywords** Procrustes analysis · Mantel test · Randomization test · Multivariate analysis · Matrix association

---

### Introduction

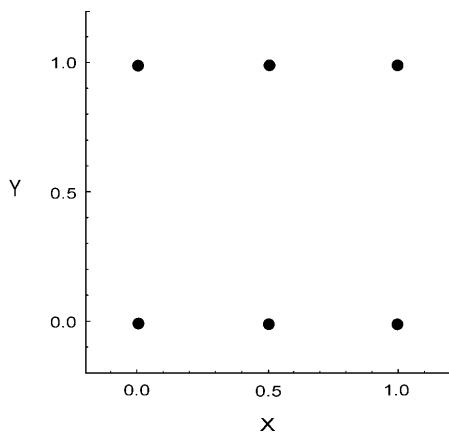
The comparison of multivariate data sets and the testing of their association is a long-standing approach in identifying important processes governing evolutionary and ecological patterns. For instance, ecomorphologists are often interested in testing if size and shape variation are correlated to ecological differences among closely related species (e.g., Douglas and Matthews 1992; Losos and

de Queiroz 1997; Van Damme et al. 1998), whereas comparative biologists are lured by questions of how much of this correlation is due to common ancestry (e.g., Losos 1990; Malhotra et al. 1996). Community ecologists seek to establish relationships between environmental characteristics and species distribution (e.g., Jackson and Harvey 1993; Rodríguez and Lewis 1997; Jenkins and Buikema 1998). Among systematists a common goal is to determine whether or not spatial distribution is related to phenotypic or genetic differentiation among populations or species (e.g., Douglas and Endler 1982; Douglas et al. 1999). These research programs embrace rather different questions and types of multivariate data, but they all involve comparisons between two or more data sets in order to measure their degree of concordance. If statistically significant, the match between data sets contributes to evidence about the processes determining the association.

Due to its simplicity and flexibility, the most common approach used by biologists for assessing the congruence between two multivariate data sets is the test for matrix correlation developed by Mantel (1967; e.g., Douglas and Endler 1982; Manly 1986; Jackson and Harvey 1989; Oden and Sokal 1992; Rosa et al. 1995; Thorpe et al. 1996; Manly 1997; Rodríguez and Lewis 1997; Douglas et al. 1999; Dutilleul et al. 2000 and references in). First each multivariate data set is translated into a pairwise distance matrix that expresses the difference between each pair of objects (e.g., species, sites, populations) in a multivariate space. The Mantel approach consists of calculating the correlation between the two distance matrices; and then a randomization procedure or a parametric approximation is used to evaluate whether the observed correlation is different from random (Jackson and Somers 1989; Manly 1997). The more important features of the Mantel approach are: (1) the possibility of using the large number of possible distance measures available, both in Euclidean and non-Euclidean spaces (e.g., genetic distances, Bray-Curtis); and (2) any type of data can be used to construct association matrices (e.g., continuous, ordinal or binary data). However, raw data

---

P.R. Peres-Neto (✉) · D.A. Jackson  
Department of Zoology, University of Toronto, Toronto, Ontario,  
M5S 3G5, Canada  
e-mail: pperes@zoo.utoronto.ca



**Fig. 1** Simple Pearson product moment correlation between  $X$  and  $Y=0$ , whereas the correlation between distance matrix  $X$  and matrix  $Y=-0.32$ . Matrix correlation was based on separate Euclidean distance matrices, one expressing the pairwise distances between the six observations based on the abscissa values ( $X$ ) and the other on the ordinate values ( $Y$ )

sets must always be transformed into distance matrices. One drawback in performing such transformations is that interpretations in the space of distances are not necessarily equivalent to the ones taken from the space of the original data (Dutilleul et al. 2000). Consider the bivariate orthogonal configuration presented in Fig. 1. The simple Pearson correlation between  $X$  and  $Y$  gives an  $r=0.0$ . Based on two separate Euclidean distance matrices, one expressing the pairwise distances between the 6 observations on the abscissa values ( $X$ ) and the other on the ordinate values ( $Y$ ), the between-matrix correlation is  $r=-0.32$ . Thus, the correlation between distance matrices is not necessarily zero for configurations which are orthogonal to one another (Heo and Gabriel 1998; Dutilleul et al. 2000).

An alternative approach is to assess the overall degree of association between two or more matrices applying a Procrustean superimposition approach (Gower 1971). Rather than transforming data sets into distance matrices, the raw data matrices, or their ordination solutions, are scaled and rotated in order to find an optimal superimposition that maximizes their fit. The sum of the squared residuals between configurations in their optimal superimposition can be then used as a metric of association (Gower 1971). A permutation procedure (PROTEST) implemented by Jackson (1995) can be then used to assess the statistical significance of the Procrustean fit. The Procrustean superimposition approach provides a number of additional features, while keeping Mantel's flexibility (i.e., choice of distance measure and types of data). A particular advantage of the Procrustean approach is that when displaying multivariate data sets in their optimal superimposition, one can visually examine the concordance of observations for each dimension separately, aiding the interpretation of data structure. For the Mantel approach, since each difference between observations for all dimensions is compressed into a single distance, the importance

of each variable (or dimension) cannot be evaluated directly by a simple scatterplot of one distance matrix against the other.

Although fairly common in the realm of morphometric analysis (see Rohlf and Slice 1990; Bookstein 1996 for reviews), the use of Procrustes methods in the interpretation of data sets having different characteristics (e.g., morphological vs ecological features) is restricted mainly because the possibilities of application have not been yet detailed and explored (but see Olden et al. 2001). Furthermore, the permutation procedure (PROTEST) for the Procrustean fit was only recently developed (Jackson 1995) and its statistical performance was never assessed. In order to show ecologists and evolutionary researchers the statistical and analytical advantages of the Procrustean superimposition approach, this article aims to: (1) compare the performance in terms of type I error rates and power of the Procrustean permutation test proposed by Jackson (1995) and the Mantel test; and (2) suggest and illustrate strategies for using Procrustes methods as an analytical tool to describe and test patterns of association among multivariate data sets.

## Materials and methods

### Mantel and Procrustes statistics

The congruence or correlation between distance matrices (usually symmetric, but see Mantel 1967) can be calculated as:

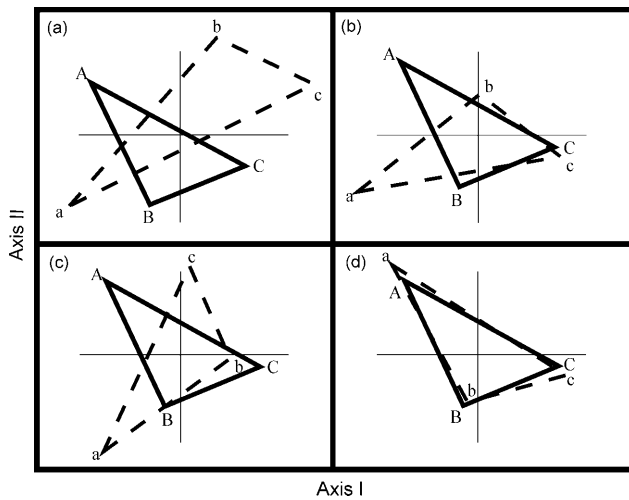
$$Z = \sum_{i,j} X_{ij} Y_{ij}$$

where  $X_{ij}$  and  $Y_{ij}$  are the  $i^{\text{th}}$  and  $j^{\text{th}}$  off-diagonal elements of symmetric matrices (see Douglas and Endler 1982 for a computational example). The diagonal elements of the matrices are not included because the probability of rejection is unaffected. Also, the above equation is computationally simpler but equivalent in terms of probability to its standardized version based on the Pearson product-moment correlation (Manly 1997, p 174).

Procrustes analysis is a procedure that minimizes the sum-of-squared differences between two or more configurations (i.e., data matrices) in a multivariate Euclidean space. There are several strategies for Procrustes analysis (Rohlf and Slice 1990), but the simplest approach is the least-squares superimposition of one configuration  $Y$  ( $n$  rows or observations by  $k$  columns or variables) to a reference configuration  $X$  ( $n \times k$ ). To illustrate the basic approach of a Procrustes analysis, two triangles ( $X$ : A-B-C and  $Y$ : a-b-c) differing in location, size (i.e., sum of their distances) and orientation are presented (Fig. 2a). First, the configurations are re-scaled to a common size and jointly centered (Fig. 2b), and, if necessary, mirror reflected (Fig. 2c) so that their orientation is coincident. In order to find the optimal superimposition, one configuration is kept fixed as reference ( $X$ : A-B-C), while the other ( $Y$ : a-b-c) is rotated successively until the sum-of-the-squared residuals between corresponding coordinates in both configurations ( $\Delta_{12}^2$ ) is minimized (Fig. 2d). Thus, the greater the concordance between data sets, the lower the  $\Delta_{12}^2$  value. In addition, individual residuals between homologous observations can be interpreted separately, indicating their match. Computationally, the Procrustean fit can be achieved in two steps: (1) centering and scaling:

$$X_{\text{sc1}} = (I - P)X / \sqrt{\text{tr}[(I - P)X'(I - P)]}$$

where  $I$  is an  $(n \times n)$  identity matrix,  $P$  is a  $(n \times n)$  matrix with all elements  $=1/n$ . Repeat this step for configuration  $Y$  leading to  $Y_{\text{sc1}}$ . Note that the scaling process does not alter the proportional differ-



**Fig. 2a–d** Representation of steps involved in a Procrustes analysis: **a** original configurations; **b** configurations after standardization (i.e., similar size and common center); **c** configurations after mirror reflection, if necessary; **d** configuration after rotation and dilation so that the sum of the squared differences between homologous observations (A/a, B/b, C/c) is a minimum. Triangle ABC was used as reference configuration

ences in terms of variances among variables so that variables with high variances originally will have more importance in the overall fit. On the other hand, if each variable is standardized to mean 0 and variance 1 prior to the scaling process, all variables will have the same weight in the fit process; and (2) reflection, rotation and the residual sum-of-squares statistics:  $\Delta_{12}^2 = 2(1 - \text{trace } W)$ .  $W$  is obtained by the singular value decomposition ( $X'_{\text{sc1}} Y_{\text{sc1}} = VWU'$ ). In order to make  $\Delta_{12}^2$  vary between the range of 0 and 1, the following transformation is used:  $(\text{trace } W)^2$ . Since both matrices were standardized in step 1,  $m_{12}$  is the same regardless of which matrix is used as the reference matrix. The optimal rotation matrix (i.e., that providing the best fit) can be calculated as  $H = UV'$ . Note that  $\Delta_{12}^2$  can also be calculated by first rotating the matrix  $Y$  (i.e.,  $Y_{\text{rot}} = Y_{\text{sc1}} H$ ) and then finding the sum of the squared distances between  $X_{\text{sc1}}$  and  $Y_{\text{rot}}$ . It is interesting to note that where  $X$  and  $Y$  have only one variable each (e.g., Fig. 1),  $W$  is equal to the Pearson correlation index and therefore  $m_{12} = 1 - r^2$ . Because the superimposition process requires that matrices have the same dimensionality, where this is not the case, the matrix with the smaller number of variables can be filled with columns of zeros until it matches the dimensionality of the larger matrix (Krzanowski and Marriot 1994). Another possibility, which we use in this paper, is to reduce both sets to a common dimensionality by applying an ordination technique (e.g., principal component analysis) to each data matrix separately.

#### Randomization protocol for the Mantel test and PROTEST

In order to evaluate the significance of  $Z$  (Mantel test) and  $m_{12}$  (PROTEST), a randomization test was conducted as follows: (1) calculate the observed statistic of association  $Z_{\text{obs}}$  and  $m_{12-\text{obs}}$ ; (2) randomly permute entire rows in relation to each other of one data matrix; and (3) recalculate the values for the permuted association  $Z_{\text{rnd}}$  and  $m_{12-\text{rnd}}$ ; (4) repeat steps 2 and 3 a large number of times. For the Mantel test, randomly permute rows of one matrix and apply the same randomized order to the columns (see Manly 1997, p 175 for an example). The same permutation of rows was applied for both tests because we wanted to compare them also on the basis of the same permuted sets. Both tests are based on one-tailed probabilities. In the Mantel test large  $Z_{\text{obs}}$  indicates greater correlation between data sets and thus the probability is calculated as:

(number of  $Z_{\text{rnd}}$  equal to or larger than  $Z_{\text{obs}} + 1$ ) / (number of randomizations + 1). In contrast, smaller values of  $m_{12}$  indicate higher concordance between data sets and the probability of rejection is assessed as (number of  $m_{12-\text{rnd}}$  equal to or smaller than  $m_{12-\text{obs}} + 1$ ) / (number of randomizations + 1). The 1 in the numerator and the denominator represents the observed value for the statistic being evaluated, which is considered as a possible value of the randomized distribution.

#### Simulation experiment: type I error rates and power

In this study we follow standard Monte Carlo protocols for estimating probabilities of type I error and power for all methods described above (e.g., Manly 1997; Peres-Neto and Marques 2000; Peres-Neto and Olden 2001). In this case, one simulates population correlation matrices and manipulates them in order to introduce a desirable effect size (i.e., a particular correlation between matrices). Following this, a large number of samples are taken and the test statistic is calculated each time. If the effect size is manipulated to be zero (i.e., the null hypothesis is true), the probability of committing a type I error is estimated as the fraction of tests that erroneously rejected the null hypothesis. If the effect size is set different from zero, the proportion of cases in which the null hypothesis was correctly rejected is used as an estimate of statistical power.

The first step was to design correlation matrices having specific Pearson correlation structure within and between data sets. Each matrix contains 12 variables in total where the first 6 variables were assigned to data matrix  $X$  and the last 6 were assigned to data matrix  $Y$ . Two types were considered:

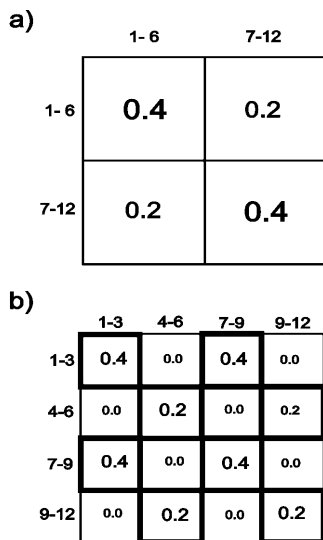
1. Homogeneous matrices have all variables within and between data matrices uniformly correlated (Table 1 and Fig. 3a). For example, matrix 8 (Fig. 3a) has all variables within each data matrix correlated with each other at  $r=0.4$ , and all variables between (i.e., across) data matrices correlated with each other at  $r=0.2$ .
2. Heterogeneous matrices have two levels of correlations within and between data matrices (Table 1 and Fig. 3b). In these cases, the first three variables within each data matrix were correlated at a specified level within and between matrices (e.g.,  $r=0.4$  for matrix 18, Fig. 3b), whereas the other three variables were correlated at a different level ( $r=0.2$  for matrix 18, Fig. 3b). Correlations between corresponding blocks of variables were set at  $r=0.0$  (Fig. 3b).

Generating the samples following designed correlation structures (Table 1) was achieved by applying a modified procedure used by Ganesanandam and Krzanowski (1990) as follows. The steps consists of: (1) generate 12 random normally distributed variables  $N(0, 1)$  with the appropriate number of observations according to sample size; (2) decompose the correlation matrix by a Cholesky decomposition; (3) post-multiply the upper triangular matrix resulting from the matrix decomposition of step 2 by the random matrix of step 1; (4) separate the matrix in step 3 so that the first 6 variables are designated to data matrix  $X$  and the other 6 to data matrix  $Y$ . Note that the resultant data matrices are a sample from a multivariate normal distribution with the specified correlations as in Table 1.

Type I error rates were measured using samples from correlation structures with no association between matrices (Table 1, matrices 1–5). All other matrices present some degree of association between their data matrices and thus are suitable for measuring power. From each correlation matrix, 1,000 random samples of size 10, 20 and 30 observations were extracted and the Mantel test and PROTEST were conducted for each sample matrix using 999 permutations. As mentioned previously, we adopted a paired-test protocol where both statistics were calculated on the same samples and the same randomized data set for all correlation matrices. Each sample matrix had its variables standardized to unit variance and zero mean prior to the tests. For the Mantel test, pairwise (between observations) Euclidean distance matrices (Legendre and

**Table 1** Between- and within-matrix correlation structure of populations used in the simulation study. All populations contain six variables in each of two data matrices. Homogeneous matrices present uniform correlations between and within data matrices (see Fig. 3a for a pictorial example), whereas heterogeneous matrices are divided into blocks of three variables each with non-uniform correlations (see Fig. 3b)

Matrix	Correlation between data matrices	Correlation within data matrices
Homogeneous		
1	0.0	0.8
2		0.6
3		0.4
4		0.2
5		0.0
6	0.2	0.8
7		0.6
8		0.4
9		0.2
10		0.0
11	0.4	0.8
12		0.6
13		0.4
14		0.2
15		0.0
16	0.6	0.8
17		0.6
18		0.4
19		0.2
20		0.0
21	0.8	0.8
22		0.6
23		0.4
24		0.2
25		0.0
Heterogeneous		
16	0.2/0.0	0.2/0.0
17	0.4/0.0	0.4/0.0
18	0.4/0.2	0.4/0.2
19	0.6/0.0	0.6/0.0
20	0.6/0.2	0.6/0.2
21	0.6/0.4	0.6/0.4
22	0.8/0.0	0.8/0.0
23	0.8/0.2	0.8/0.2
24	0.8/0.4	0.8/0.4
25	0.8/0.6	0.8/0.6



**Fig. 3** Matrix representation of the correlation structure for **a** homogeneous correlation matrices with all variables within and between data sets uniformly correlated (the example represents matrix 8 in Table 1) and **b** heterogeneous correlation matrices with two levels of correlations within and between data matrices (the example represents matrix 18 in Table 1). See text for more details. The first six variables (1–6) are used to create data matrix *X* whereas the last six variables (7–12) are used to compose data matrix *Y*

Legendre 1998) were calculated. Principal component analysis (PCA) was applied to each sample matrix in order to evaluate the effects of reduction of data dimensionality on both tests. PCA was conducted on each sample matrix separately, and the scores scaled relative to their eigenvalues for the two first PCA axes were used as input data for both tests. For the Procrustes fit the PCA scores were used as the input data matrix, whereas for the Mantel test Euclidean distance matrices were calculated from the PCA scores.

A significance level of  $\alpha=0.05$  was used throughout and the results for  $\alpha=0.01$  were similar regarding the differences in performance between methods and therefore are not reported. Type I error rates were estimated as the proportion of sample tests out of 1,000 that rejected the null hypothesis when the null hypothesis is true (Table 1, matrices 1–5). On the other hand, power was calculated as the proportion of sample tests that correctly identified true associations (Table 1, all remaining data sets). To compare estimates, we calculated 95% confidence intervals for both power and type I error rate estimates (Manly 1997) as:

$$p \pm 1.96 \sqrt{p(1-p)/N},$$

where  $p$  is the proportion of rejections and  $N$  is the number of sample test trials (i.e., 1,000).

#### Procrustes superimposition as an analytical tool

The Procrustes method is flexible enough that a variety of strategies can be implemented or adapted in order to provide an array of analytical procedures. We present two examples that are both useful and broad enough to serve as a general guide for future applications. The data sets used to demonstrate the analytical approaches were chosen such that the raw data was available in the original publication.

#### Example 1 – Procrustes superimposition plot

The first procedure relates to a plot of the superimposed data matrices in their optimal fit. We refer to this graphical output as a Procrustean superimposition plot. This plot provides a template for multivariate data visualization, where differences between homologous observations in the two matrices can be displayed for each dimension (i.e., variable or multivariate axes) separately. This aids in identifying individual discrepancies and the relative contribution of each variable to the match.

As an example we used the ecomorphological data of Losos (1990) to address the question of whether there is an association between morphology (snout-vent length, mass, foreleg, hindleg, tail lengths and lamella number) and performance capability (sprint, jump and cling) of West Indian *Anolis* lizards. The raw data for 15 species are presented in Table 1 of Losos (1990). Because of missing data in the performance measurements, *Anolis occultus* was excluded from the analysis. We used PCA to summarize the main patterns of variation of morphological and performance data sets. Variables were log-transformed prior to analysis, and in both cases PCA was performed on a correlation matrix. Since we wanted morphological components (i.e., size and shape axes) having the same weight in the Procrustean fit, we scaled the variance of scores on each axis to be 1. The scores on the first two components from each data set were used as input matrices for Procrustes superimposition. Morphology was used as the reference configuration for the rotational process. After centering and scaling the morphological ( $M_{scl}$ ) and performance ( $P_{scl}$ ) data sets, the optimal displacement that minimizes the residual sum-of-squares of  $P_{scl}$  onto  $M_{scl}$  can be achieved as  $P_{rot} = P_{scl} H$ , where  $H$  is the optimal rotational matrix between both data sets (see Procrustes statistics). Displaying  $M_{scl}$  and  $P_{rot}$  jointly will generate the Procrustean superimposition plot. In order to relate the species differences in terms of morphological and performance variables, the eigenvectors of the first two morphological and performance components, which contain the correlation (i.e., loadings) between variables and components (i.e., scores), were also analyzed. To pro-



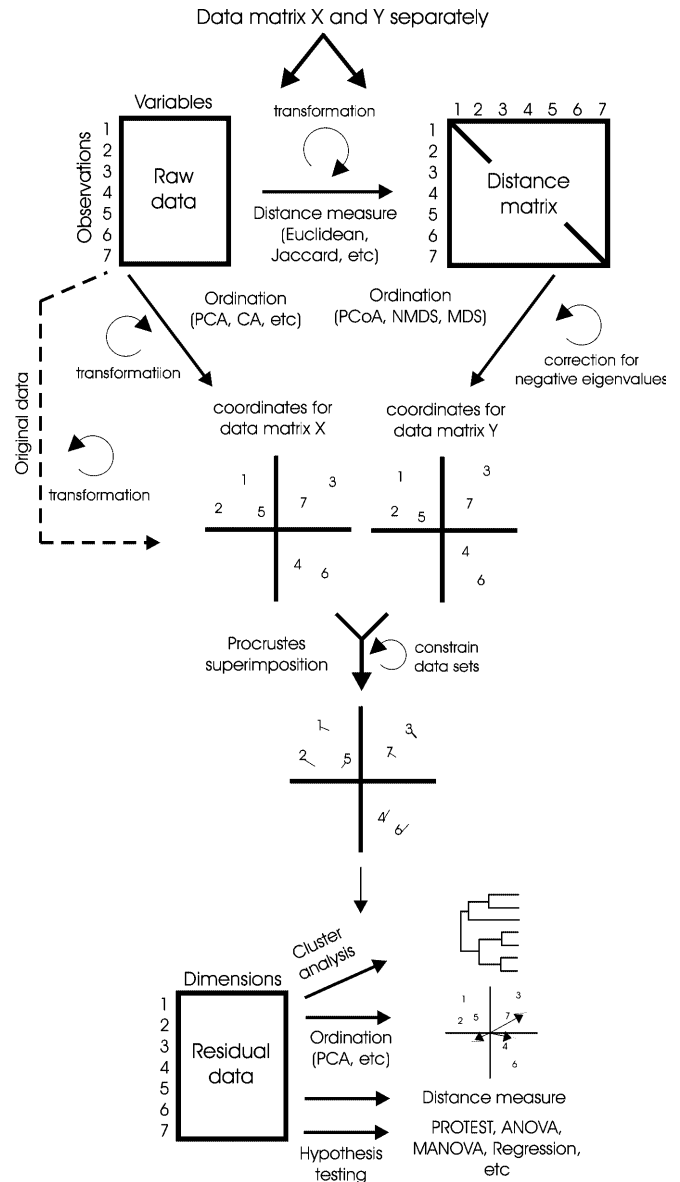
vide this contrast, since the performance scores are now rotated, performance eigenvectors also need to be rotated by matrix  $H$  (i.e., post multiply the eigenvector matrix by  $H$ ).

### Example 2 – partial PROTEST

The second analytical tool that we suggest is a partial approach where the variation in the two target matrices related to other sources of data can be partitioned out and then a PROTEST is performed. We develop this partial PROTEST method as an analogous procedure to a multiple Mantel test or partial Mantel test (Smouse et al. 1986; Thorpe et al. 1996; Douglas et al. 1999), which is an extension of the original test for testing the correlation between two distance matrices where the variation due to one or more matrices is kept fixed (i.e., partial correlation). This approach can be essential in the process of testing the hypothesis of association/causation because it minimizes problems related to spurious correlation, i.e., that the association between two sets of data is largely due to their mutual correlation with other sources of variation rather than to their own.

The partial PROTEST can be best described as a simple extension of partial linear multiple regression (Legendre and Legendre 1998). For simplicity, consider two data matrices  $X$  and  $Y$ , where the variation related to a third data matrix  $Z$  will be partitioned out. In short, partial PROTEST consists of two steps: (1) perform a separate multiple regression for each variable in  $X$  on all variables in  $Z$  and compute the residual, repeating the same process for each variable in  $Y$  against  $Z$ ; and (2) carry out a PROTEST using the two matrices of residuals values as input data sets. Note that the residual matrices have the same dimensionality as the original data, but the variation related to  $Z$  no longer remains. As an example of the procedure, we use the data of Spielman (1973) which contrasts matrices of genetic (gene frequencies), anthropomorphic (Mahalanobis) and geographic (Euclidean) distances between 19 Yanomama villages from northern Brazil and southern Venezuela. We chose this particular study also to show how data that are only available in the form of distance matrices can be analyzed using the Procrustes method. In this case, the first step is to obtain a Euclidean representation in a Cartesian coordinate system of the distance matrices using principal coordinate analysis (PCoA; see Legendre and Legendre 1998 for a review). This ordination method “unfolds” the distance matrix into a coordinate matrix that preserves the relationship between objects. A perfect representation is guaranteed only if the original distance measure is Euclidean and thus when conducting a PCoA on non-Euclidean distances, negative eigenvalues are produced meaning that some of the dimensions are not in real space (i.e., imaginary space), and the relationship among objects are distorted (Gower and Legendre 1986). There are at least three solutions for this problem: (1) correct for the negative eigenvalues directly in the distance matrix and recalculate the PCoA as in Legendre and Legendre (1998, p 434); (2) if the absolute sum of positive eigenvalues is much larger than the sum of negative eigenvalues, use only the PCoA dimensions that present positive eigenvalues (Krzyszowski and Marriott 1994); or (3) use non-metric multidimensional scaling (Legendre and Legendre 1998). McArdle and Anderson (2001) showed that the first two methods inflate the total sum of the squares in the analysis, which may cause a decrease in the nominal type I error. They showed that negative eigenvalues are a negative sum of the squares and should be subtracted from the positive eigenvalues in order to provide the correct total sum of the squares. One solution may be to re-scale the multivariate coordinates such that correct total sum of squares is used. However, because the negative eigenvalues are relatively small for Spielman’s (1973) genetic distance matrix, which is the only non-Euclidean matrix in the study, we chose the second solution.

Our intent is to use the data sets of Losos (1990) and Spielman (1973) as examples of the application of the analytical tools based on Procrustes methods described here to real situations. Therefore comparisons with previous results and discussion of the data will not be performed. Although we demonstrate the flexibility of the

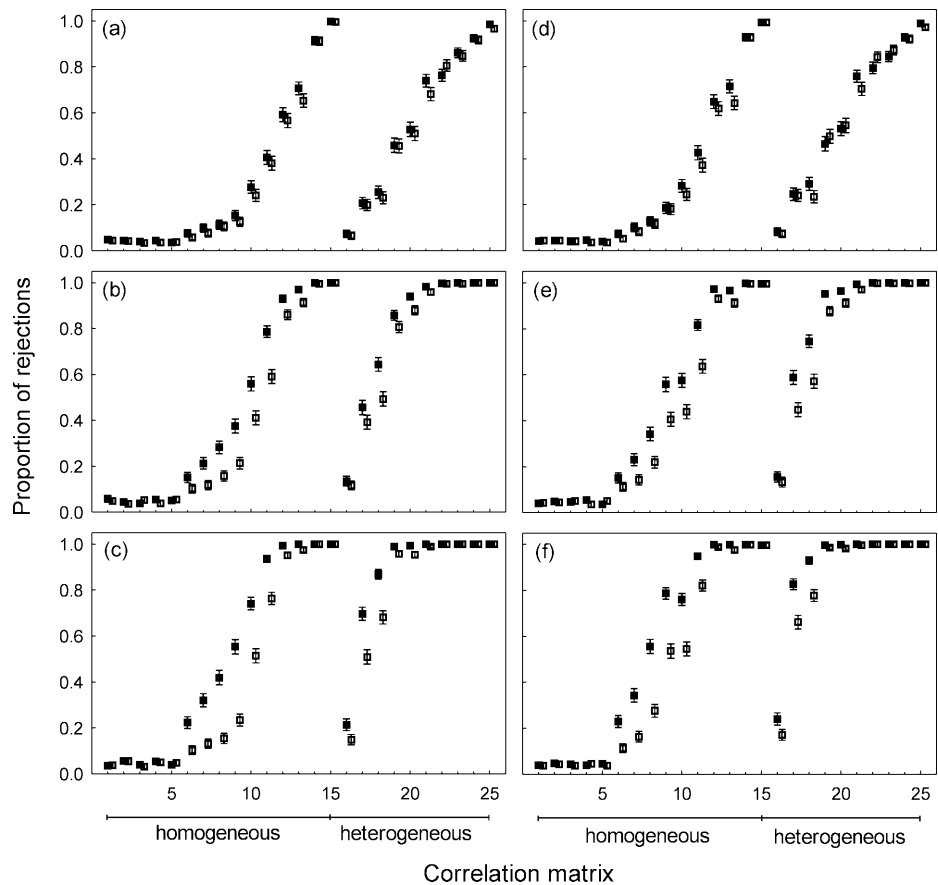


**Fig. 4** Diagrammatic summary of possibilities when applying Procrustes rotations as an analytical tool. Input multivariate data are represented by either raw or distance matrices, or both. For simplicity only two data sets are considered. Data sets are superimposed and residuals can be used as input to other analytical tools (see text for more details)

Procrustes superimposition approach using two examples of analytical strategies, they differ sufficiently to show an avenue of possibilities, which are summarized in Fig. 4. Three main stages are identified as follows:

1. Multivariate data sets are represented in the form of raw data, distance matrices, or both. Where necessary, appropriate transformations should be carried out (e.g., log-transformation).
2. Raw data or ordination solutions for each data set are superimposed. Where necessary, conduct corrections for negative eigenvalues.
3. A residual data matrix is generated and serves as input for other analytical tools (e.g., Procrustean superimposition plot, partial PROTEST, cluster analysis, ANOVA). Note that the sum-of-the-squared residuals also can be used as a measure of over-

**Fig. 5a–f** Type I error and power, measured as the proportion of rejections ( $\alpha=0.05$ ) per 1,000 tests, of PROTEST (solid squares) and Mantel test (open squares). **a** Raw data, sample size  $n=10$ ; **b** raw data,  $n=20$ ; **c** raw data,  $n=30$ ; **d** PCA data,  $n=10$ ; **e** PCA data,  $n=20$ ; **f** PCA data  $n=30$ . Vertical bars indicate confidence intervals for the estimates



all distance between two multivariate data sets, and therefore can be used as an input to pairwise distance matrices contrasting several multivariate data sets (e.g., Jackson 1993).

## Results

### Type I error rates and power

The results for both type I error and power estimates are presented in terms of the proportion of rejections out of 1,000 sample tests for each correlation matrix. Matrices 1–5 (Table 1) were used to estimate type I error rates. The number of rejections was usually equal to the nominal value of 5% for both tests regardless of the sample size, within-matrix structure and whether either raw data (Fig. 5a–c) or data summarized by PCA (Fig. 5d–f) were used. These results indicate suitable efficacy of both tests in controlling type I errors (i.e., rejecting the null hypothesis when it is true), and therefore power comparisons between both tests are straightforward.

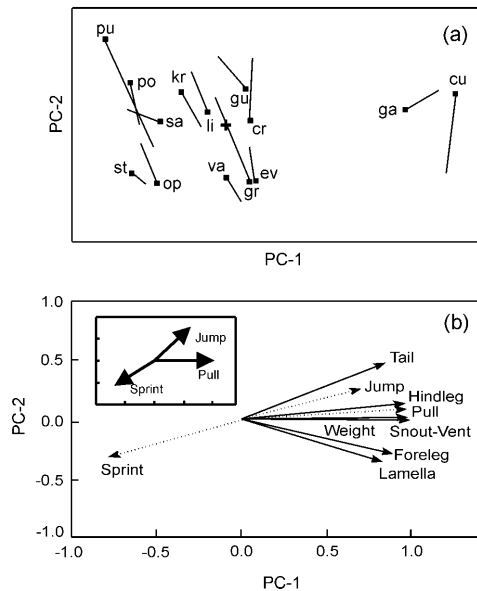
Based on non-zero patterns of correlation (matrices 6–25), PROTEST showed equivalent or greater power than the Mantel test for detecting association between data matrices (Fig. 5). For sample size 10, PROTEST and the Mantel test present comparable power; however as sample size increases (20 and 30 observations) PROTEST outperforms the Mantel test for all correlation matrices with intermediate degrees of association. In some cases the dif-

ference in power between the two tests was between 20% and 30% (e.g., matrices 8 and 9 for sample size 30). For matrices having high degrees of association, both tests were extremely powerful (Fig. 5). Using a PCA to summarize within-matrix patterns led to similar or enhanced power in most situations. The largest differences were provided by tests based on matrices 8, 9, 17 and 18 (Fig. 5) that present low degrees of association (Table 1).

### Procrustes superimposition plot

The two first axes from the PCA of the lizard morphological data (Losos 1990) accounted for 92% of the total variation, whereas the PCA on the performance data accounted for 98% of the total variance in the first two axes. Based on these two components, the PROTEST revealed highly significant concordance between data sets ( $m_{12}=0.51$ ,  $P\leq 0.001$ ). To simplify the interpretation of the Procrustean superimposition plot, residuals between configurations after optimal fit are represented by straight lines connecting each species between data sets (Fig. 6a). Species such as *A. stratulus* have small residuals indicating a close match between morphology and performance in contrast to species having larger residuals such as *A. pulchellus*.

In order to specify the components of morphological and performance variation, we used the morphological



**Fig. 6a, b** The Procrustes superimposition of *Anolis* species based on the two first morphology (reference) and performance principal components (raw data from Losos 1990). **a** Procrustes superimposition plot of morphological (solid squared) and performance components (end point of solid lines). Solid lines represent Procrustes residuals from both configurations. Species codes represent the two first letters of their names: *A. cristatellus*, *A. cuvieri*, *A. evermanni*, *A. garmani*, *A. grahami*, *A. gundlachi*, *A. krugi*, *A. lineatopus*, *A. opalinus*, *A. poncensis*, *A. pulchellus*, *A. sagrei*, *A. stratulus* and *A. valencienni*. **b** Directions of variation of morphological and performance variables with their first two principal components. Inset plot shows correlations of performance variables with their first two principal components

(Fig. 6b) and performance (inset in Fig. 6b) eigenvectors (i.e., correlations) in the first two principal components. Because now the performance components are rotated towards the morphological components, we also rotated the performance eigenvectors according to the optimal rotational matrix (inset plot in Fig. 6b) so that one can identify differences in species performances back in the original performance space (Fig. 6a). Each variable is represented by a vector, the length of which is proportional to the correlation of that variable with each morphological or performance PCA axis (Fig. 6b). All morphological variables have positive correlations with their PC-1, indicating that this component can be interpreted as a general variable of size, whereas PC-2 may be useful in examining size-scaled variation (i.e., shape). For instance, *A. cuvieri* and *A. garmani* are larger in size than *A. stratulus* and *A. opalinus*; however these latter species have relatively long forelegs and greater numbers of lamella than *A. cuvieri* and *A. garmani* (Fig. 6a, b; see also Losos 1990: Table 1). In terms of performance (inset in Fig. 6b), sprinting ability is negatively correlated with the ability of jumping and pulling (but see Losos 1990) so that *A. garmani* and *A. gundlachi* tend to jump and pull better than *A. stratulus* and *A. valencienni* which have a better sprinting ability (Fig. 6a and inset in Fig. 6b).

Once the components of variation within the two sets are identified, one can also specify which components of morphological variation provide a better match with performance. At this time, each performance variable was correlated with the first two morphological components.

Interpreting the correlation between all variables and the morphological components (Fig. 6b), two patterns can be recognized: (1) larger species (PC-1) tend to jump and pull better than smaller ones, which in turn tend to run faster than larger ones; and (2) shape variation does not seem to be highly associated to differences in performance abilities, although some species with relatively longer tails will tend to run slower (e.g., *A. gundlachi*). According to the direction of change of residuals, most of performance residuals are related to shape differences (PC-2) rather than size variation (Fig. 6a), indicating that in fact size is more correlated with performance than shape. In fact, when partitioning the total  $m_{12}=0.59$  into the size (i.e., PC-1  $m_{12}=0.02$ ) and shape (i.e., PC-2  $m_{12}=0.57$ ) contributions, the latter presents a considerably larger component of residual variation. This result corroborates our previously finding in which size (PC-1) is a better predictor of performance than shape (PC-2) (Fig. 6b). Specifically, the largest residuals are associated with *A. pulchellus* and *A. grahami*, suggesting that their shape provides the poorest predictions of their performance. Species with relatively long tails have lower sprinting capabilities in general (Fig. 6a), but *A. pulchellus* provides a better sprint than predicted by its tail whereas *A. grahami* exhibits the opposite relationship, being less efficient in sprinting than predicted by the relative size of its tail.

### Partial PROTEST

The first four positive eigenvalues from the PCoA of Spielman's (1973) genetic distance matrix accounted for 74% of the total variation. In order to keep matrices with same dimensionality, we also retained the four first PCoA axes for the anthropomorphic (80%) and geographic (100%) distance matrices. Note that for geographic data (i.e., straight line distances) the two first axes should account for 100% of the variation; however when data present measurement error, which was the case here, the representation spans through more than two axes. To partition out the variation due to common geographic variation, partial PROTEST was conducted by regressing each of the four anthropomorphic and genetic PCoA axes on the four retained geographic PCoA axes. The test indicated that there is a significant association between genetic and anthropomorphic variation after the spatial contribution is partitioned out ( $m_{12}=0.65$ ,  $P\leq 0.001$ ).

## Discussion

### Type I error rates and power

Our simulation results revealed that PROTEST is as powerful or more powerful than the Mantel test in most

situations. Large differences were achieved by intermediate degrees of association, where PROTEST provides greater power than the Mantel test. We are aware of only few studies that have attempted to evaluate type I error and/or power of the Mantel test (Dietz 1983; Manly 1986; Smith 1998). In all these studies, including ours, the correlation between data sets was manipulated to generate a gradient from low to high association between matrices. These studies simulated data where the correlation within data matrices was set to zero (independent random variables), and the variables were generated having either normal (Dietz 1983) or uniform distributions (Manly 1986). More recently, Smith (1998) generated data where the within-matrix correlation was varied, but either set to zero or extremely high at  $r=0.8$ . Therefore, our simulations were more extensive in creating a larger number of more varied scenarios where the structure within- and between-matrices was more finely examined (Table 1) relative to other studies. We also evaluated the power where multivariate analysis (e.g., PCA) is used as a prelude to performing the tests, a common procedure among evolutionary biologists and ecologists when using the Mantel test (e.g., Jackson and Harvey 1989; Douglas and Matthews 1992; Rodrigues and Lewis 1997; Douglas et al. 1999). The results show some benefit in terms of power for both tests when using PCA, especially for matrices with low degrees of association, indicating that it may be advantageous to reduce data dimensionality by applying an ordination prior to testing.

In our comparisons involving the Mantel test, we restricted our evaluations to the measure of Euclidean distance due to its broad use in evolutionary and ecological applications, especially in studies related to spatial analysis (e.g., Jackson and Harvey 1989; Oden and Sokal 1992; Rodrigues and Lewis 1997; Manly 1997). Additionally, previous studies have tested the power of the Mantel test using straight Euclidean distance (Dietz 1983; Manly 1986; Smith 1998). However, different distance measures and types of data might provide different results for the Mantel test (Jackson 1995). Recently, Dutilleul et al. (2000) showed that the squared Euclidean distance provided a better association between data sets than simple Euclidean distances. Based on their findings we recalculated our power estimates for the Mantel test using squared Euclidean distances as the association metric for generating pairwise distance matrices. Using the raw data matrices and sample size of 30 observations for our correlation matrices, we found an increase in power compared to our original estimates for the Mantel test, but the squared measure had always equivalent or significantly lower power when compared to PROTEST. The impact of other choices of distances on the power of the Mantel test may be an avenue for future investigation.

We suggest that PROTEST has greater power because it measures association using the raw observations, or their ordination solutions, whereas the Mantel test measures association between distance matrices. Dutilleul et al. (2000) showed that the conclusions based on one

approach are not necessarily consistent with the other. They explored some bivariate examples where the Pearson correlation was zero, but the standardized Mantel statistic was either negative or positive (e.g., Fig. 1). Because  $m_{12}=1-r^2$ , the significance of the Procrustes statistics and the Pearson correlation will always match. In addition, Dutilleul et al. (2000) simulated samples following different bivariate correlation levels to show that the Mantel test has lower power when compared to a parametric and a randomization test based on the Pearson correlation. Due to the inconsistencies found between the Mantel and Pearson statistics, Dutilleul et al. (2000) concluded that interpretations based on a space of distances may not always be validly applied back to the space of the original data, which has been the general avenue when interpreting results based on the Mantel statistics. However, it is often the case that one is interested in choosing a particular distance measure which may emphasize, or diminish, particular attributes that are considered to be important or unimportant. For example, inverse-distance matrices emphasize local-scale relationships between data matrices, while down-weighting large-scale relationships. In these cases, the representation among objects given by principal coordinate analysis, prior to Procrustes analysis, also reflects the properties of any particular distance measure selected.

#### Procrustes superimposition as an analytical tool

In addition to the statistical performance of PROTEST over the Mantel test, another advantage in applying Procrustes analysis is the graphical solution provided (i.e., Procrustean superimposition plot) because the patterns of concordance between data sets can be displayed and individual observations assessed separately (Fig. 6a and also see Jackson 1995). For instance, we showed that the association between morphology and performance among West Indian *Anolis* lizards (Losos 1990) is largely due to their size rather than shape differences, implying that the evolution in performance may have been acquired through changes in size rather than allometric variation. In addition to visual inspection, these Procrustes residuals can be used to analyze other important sources of variation that may be important in explaining the evolution of morphology and performance. For example, the larger residuals encountered could be related to behavioral and ecological differences that these species (i.e., *A. pulchellus* and *A. grahamani*, Fig. 6) have compared to species that show smaller residuals between morphology and performance.

In addition, Procrustes residuals can be also related to other sources of information, becoming useful in many designs of post hoc analysis such as regression or analysis of variance, cluster analysis, ordinations, and even additional Procrustes analysis (e.g., Klingenberg and McIntyre 1998). These designs provide further tools when searching for additional factors in defining the association between evolutionary data sets (e.g., Smith



et al. 1997; Monteiro 1999). In these cases, Procrustes residuals should be considered as analogous to geometric regression residuals (model II regression; Sokal and Rohlf 1995) where both data sets are subjected to similar measurement errors. To avoid this condition we used multiple regressions (model I regression) to factor out the effect of geographic position when conducting the partial PROTEST on the anthropomorphic and genetic data sets of Spielman (1973). The use of Procrustes residuals as a means of factoring out the spatial component, instead, would have assumed that geography and the other two data sets have the same measurement errors. This is likely not the case since geographic coordinates can be measured more accurately.

In the present study, we have compared the statistical performance only for the case of matching two data sets because the Mantel test is restricted to the direct comparison of two data matrices at a time. In order to compare three or more matrices, the Mantel test imposes solutions where the variation in the two target matrices related to other matrices being analyzed is factored out, just as we have done. Another contribution of Procrustes fitting not shared by many other methods (but see Kettenring 1971; Krzanowski and Marriot 1994) is that several multivariate data sets can be analyzed concurrently. We do not show this approach here, but it can be achieved by simply rotating each individual matrix to the same reference or to a consensus (i.e., average) configuration. Our final remark relates to the fitting technique used in this study. Although we show the very good performance of the least-squares Procrustes method, this is the simplest rotational approach available. Because of the ample use of Procrustes analysis in morphometrics, several other strategies, like resistant-fit techniques (see Siegel and Benson 1982; Rohlf and Slice 1990), were developed and provide even more robust assessments of matrix concordance. This robustness is due to downweighting the influence of unusual data points with high residuals, and potentially will generate greater power and improved interpretations regarding the congruence of multivariate data sets. There are many possible avenues for expanding the applications of Procrustes analysis and our study is an initiative to show ecologists and evolutionary biologists the enhanced statistical performance of PROTEST and the use of the Procrustes method as an analytical tool for assessing the congruence between multivariate data sets.

**Acknowledgements** We would like to thank P. Legendre, K. Somers, A. Stewart-Oaten and an anonymous reviewer for their comments on the manuscript. P. Legendre and K. Somers also made suggestions regarding our simulation protocol and computational details; P. Dutilleul for discussions on Mantel versus Pearson's correlation; W.J. Krzanowski for discussions regarding training data sets and Procrustes analysis; D. Slice and J. Rohlf for comments on Procrustes analysis; J. Olden, who also PROTESTed along the way with us, provided support during the development of this study. PROTEST software is available on <http://www.zoo.utoronto.ca/jackson/software/>. Funding for this project was provided by a CNPq Doctoral Fellowship to P.R.P.-N. and an NSERC operating grant to D.A.J.

## References

- Bookstein FL (1996) Biometrics, biomathematics and the morphometric synthesis. *Bull Math Biol* 58:313–365
- Dietz EJ (1983) Permutation tests for association between two distance matrices. *Syst Zool* 32:21–26
- Douglas ME, Endler JA (1982) Quantitative matrix comparisons in ecological and evolutionary investigations. *J Theor Biol* 99:777–795
- Douglas ME, Matthews WJ (1992) Does morphology predict ecology? Hypothesis testing within a freshwater stream fish assemblage. *Oikos* 65:213–224
- Douglas ME, Minckley WL, DeMarais BD (1999) Did vicariance mold phenotypes of western North American fishes? Evidence from *Gila* river Cyprinids. *Evolution* 53: 238–246
- Dutilleul P, Stockwell JD, Frigon D, Legendre P (2000) The Mantel test versus Pearson's correlation analysis: assessment of the differences for biological and environmental studies. *J Agric Biol Environ Stud* 5:131–150
- Ganeshanandam S, Krzanowski WJ (1990) Error-rate estimation in two-group discriminant analysis using linear discriminant function. *J Stat Comput Simul* 36: 157–175
- Gower JC (1971) Statistical methods of comparing different multivariate analyses of the same data. In: Hodson FR, Kendall DG, Tautu P (eds) *Mathematics in the archaeological and historical sciences*. Edinburgh University Press, Edinburgh, pp 138–149
- Gower JC, Legendre P (1986) Metric and Euclidean properties of dissimilarity coefficients. *J Class* 3:5–48
- Heo M, Gabriel KR (1998) A permutation test of association between configurations by means of the RV coefficient. *Comm Stat – Simul* 27:843–856
- Jackson DA (1993) Stopping rules in principal component analysis: a comparison of heuristic and statistical approaches. *Ecology* 74:2204–2216
- Jackson DA (1995) PROTEST: a Procrustean randomization test of community environment concordance. *Écoscience* 2:297–303
- Jackson DA, Harvey HH (1989) Biogeographic associations in fish assemblages: local vs. regional processes. *Ecology* 70: 1472–1484
- Jackson DA, Harvey HH (1993) Fish and benthic invertebrates: community concordance and community-environment relationships. *Can J Fish Aquat Sci* 50:2641–2651
- Jackson DA, Somers KM (1989) Are probability estimates from the permutation model of Mantel's test stable? *Can J Zool* 67:766–769
- Jenkins DG, Buikema AL (1998) Do similar communities develop in similar sites? A test with zooplankton structure and function. *Ecol Monogr* 68:421–443
- Kettenring JR (1971) Canonical analysis of several sets of variables. *Biometrika* 58:433–451
- Klingenberg CP, McIntyre GS (1998) Geometric morphometrics of developmental instability: analyzing patterns of fluctuating asymmetry with Procrustes methods. *Evolution* 52:1363–1375
- Krzanowski WJ, Marriot FHC (1994). *Multivariate analysis. Part 1. Distributions, ordination and inference*. Arnold, London
- Legendre P, Legendre L (1998) *Numerical ecology*, 2nd English edn. Elsevier, Amsterdam
- Losos JB (1990) Ecomorphology, performance capability, and scaling of West Indian *Anolis* lizards: an evolutionary analysis. *Ecol Monogr* 60:368–388
- Losos JB, de Queiroz K (1997) Evolutionary consequences of ecological release in Caribbean *Anolis* lizards. *Biol J Linn Soc* 61:459–483
- Malhotra A, Thorpe RS, Black H, Daltry JC, Wuster W (1996) Relating geographic patterns to phylogenetic processes. In: Harvey PH, Brown AJL, Smith JM, Lee S (eds) *New uses for new phylogenies*. Oxford University Press, Oxford, pp 187–202
- Manly BFJ (1986) Randomization and regression methods for testing for associations with geographical, environmental and biological distances between populations. *Res Pop Ecol (Kyoto)* 28:201–218

- Manly BFJ (1997) Randomization, bootstrap and Monte Carlo. Methods in biology, 2nd edn. Chapman and Hall, London
- Mantel NA (1967) The detection of disease clustering and a generalized regression approach. *Can Res* 27:209–220
- McArdle BH, Anderson MJ (2001) Fitting multivariate models to community data: a comment on distance-based redundancy analysis. *Ecology* 82:290–297
- Monteiro LR (1999) Multivariate regression models and geometric morphometrics: the search for causal factors in the analysis of shape. *Syst Biol* 48:192–199
- Oden NL, Sokal RR (1992) An investigation of three-matrix permutation test. *J Class* 9:275–290
- Olden JD, Jackson DA, Peres-Neto PR (2001) Spatial isolation and fish communities in drainage lakes. *Oecologia* 127:572–585
- Peres-Neto PR, Marques F (2000) When are random data not random or is the PTP test useful? *Cladistics* 16:420–424
- Peres-Neto PR, Olden JD (2001) Assessing the robustness of randomization tests: examples from behavioural studies. *Anim Behav* 61:79–86
- Rodríguez MA, Lewis WM Jr (1997) Structure of fish assemblages along environmental gradients in floodplain lakes of the Orinoco river. *Ecol Monogr* 67:109–128
- Rohlf FJ, Slice DE (1990) Extensions of the Procrustes method for the optimal superimposition of landmarks. *Syst Zool* 39:40–59
- Rosa CA, Morais PB, Santos SR, Peres-Neto PR, Mendonça-Hagler LC, Hagler AN (1995) Yeast associated with different plant resources in Sandy Coastal Plains of Southeastern Brazil. *Mycol Res* 99:1047–1054
- Siegel AF, Benson RH (1982) A robust comparison of biological shapes. *Biometrics* 38:341–350
- Smith DR, Crespi BJ, Bookstein FL (1997) Fluctuating asymmetry in the honey bee, *Apis mellifera*: effect of ploidy and hybridization. *J Evol Biol* 10:551–574
- Smith EP (1998) Randomization methods and the analysis of multivariate ecological data. *Environmetrics* 9:37–51
- Smouse PE, Long JC, Sokal RR (1986) Multiple regression and correlation extensions of the Mantel test of matrix correspondence. *Syst Zool* 35:627–632
- Sokal RR, Rohlf FJ (1995) Biometry, 3d edn. Freeman, New York
- Spielman RS (1973) Differences among Yanomama Indian villages: do the patterns of allele frequencies, anthropometrics and map locations correspond? *Am J Phys Anthropol* 39:461–480
- Thorpe RSM, Black H, Malhotra A (1996) Matrix correspondence of the DNA phylogeny if the Tenerife Lacertid elucidata both historical causes and morphological adaptation. *Syst Biol* 45:335–343
- Van Damme R, Aerts P, Vanhooydonck B (1998) Variation in morphology, gait characteristics and speed of locomotion in two populations of lizards. *Biol J Linn Soc* 63:409–427