

Available online at www.sciencedirect.com



Theoretical Population Biology

Theoretical Population Biology 73 (2008) 79-91

www.elsevier.com/locate/tpb

Testing for phylogenetic signal in phenotypic traits: New matrices of phylogenetic proximities

Sandrine Pavoine^{a,*}, Sébastien Ollier^b, Dominique Pontier^a, Daniel Chessel^a

^aLaboratoire de Biométrie et Biologie Evolutive, UMR CNRS 5558, Université de Lyon, Université Lyon 1, France ^bUniversité Paris-Sud 11, Faculté des Sciences d'Orsay, Unité Ecologie, Systématique et Evolution, Département Biodiversité UMR-8079 UPS-CNRS-ENGREF, France

> Received 2 March 2007 Available online 12 October 2007

Abstract

Abouheif adapted a test for serial independence to detect a phylogenetic signal in phenotypic traits. We provide the exact analytic value of this test, revealing that it uses Moran's I statistic with a new matrix of phylogenetic proximities. We introduce then two new matrices of phylogenetic proximities highlighting their mathematical properties: matrix A which is used in Abouheif test and matrix M which is related to A and biodiversity studies. Matrix A unifies the tests developed by Abouheif, Moran and Geary. We discuss the advantages of matrices A and M over three widely used phylogenetic proximity matrices through simulations evaluating power and type-I error of tests for phylogenetic autocorrelation. We conclude that A enhances the power of Moran's test and is useful for unresolved trees. Data sets and routines are freely available in an online package and explained in an online supplementary file. \bigcirc 2007 Elsevier Inc. All rights reserved.

Keywords: Autocorrelation; Computer simulation; Cyclic permutation; Cyclic ordering; Geary's *c* statistic; Moran's *I* statistic; Permutation test; Phenotypic trait; Phylogenetic diversity; Phylogenetic signal

1. Introduction

In the last decades, phylogeny has been more and more often recognized as a potential confounding factor when comparing the states of traits among several species. It is now widely accepted that given the phylogenetic links among species, species values may not be independent data so that the phylogenetic context should be taken into account when assessing the statistical significance of crossspecies patterns (Martins and Hansen, 1997).

In this context, Abouheif (1999) introduced a diagnostic test for phylogenetic signal in comparative data. It derives from a test for serial independence (TFSI) developed originally by von Neumann et al. (1941) in a non-

E-mail address: pavoine@mnhn.fr (S. Pavoine).

phylogenetic context. The TFSI detects dependences in a sequence of continuous observations by comparing the average squared differences between two successive observations to the sum of all successive squared differences. Abouheif (1999) proposed an adaptation of this test for phylogenies, by remarking that any single-tree topology can be represented in T different ways. Each representation is obtained by rotating the nodes within a phylogenetic topology, that is to say by permuting the branches connected to nodes in the original phylogeny (Figs. 1A and B). Such a rotation procedure results in changes in the ranking of the tips without changing the topology. Each rotation of the nodes results in a specific sequence of the species and thus in a specific sequence of the values taken by these species for a phenotypic trait under study providing therefore a sequence of continuous observations. The TFSI's statistic can thus be calculated for each rotation. Abouheif's (1999) test consists of taking C_{mean} as a statistic, the mean of TFSI's statistics calculated on all (or merely a random subset of all) the T possible representations of the tree topology.

^{*}Corresponding author. Present address: Sandrine Pavoine, UMR 5173 MNHN-CNRS-P6 'Conservation des espèces, restauration et suivi des populations' Muséum National d'Histoire Naturelle, CRBPO, 55 rue Buffon, 75005 Paris, France. Fax: + 33140793835.

^{0040-5809/\$ -} see front matter \odot 2007 Elsevier Inc. All rights reserved. doi:10.1016/j.tpb.2007.10.001



Fig. 1. Description of Abouheif's statistic: (A) Hypothetical phylogenetic tree with four species together with the values taken by the four species for a theoretical trait. These values are given by a Cleveland (1994) dot plot. The scale is horizontal; (B) The set of equivalent representations of the tree topology with the corresponding C_i values; (C) The matrix A associated with the hypothetical phylogeny; (D) Values taken by the TFSI's statistic C_{mean} and Moran's statistic applied to A. These two indices are equal.

In this paper, we revisit the test introduced by Abouheif (1999), demonstrating that its corresponding statistic uses a Moran's (1948) I statistic, initially developed for spatial analyses but introduced in phylogenetic analyses by Gittleman and Kot (1990). After the background on the

analyses of spatial autocorrelation, we demonstrate that Abouheif's (1999) test unifies two schools of thought developed around Moran's (1948) and Geary's (1954) work (Cliff and Ord, 1981). We propose redefining Abouheif's statistic for positioning it among other measures developed in other contexts especially in conservation biology. In addition to a mathematical formalism, this redefinition allows us to provide a biological interpretation of this statistic, previously based on node rotations, a process leading to an approximate value of a more explicit statistic. We introduce then a new phylogenetic proximity matrix (Matrix A) derived from Abouheif's statistic. It does not rely on branch lengths of the phylogeny; rather, it focuses on topology. We define this matrix analytically for all phylogenies (resolved or unresolved), independently of the trait under study. It has excellent statistical features that are presented and discussed compared to three phylogenetic proximity matrices previously used in comparative studies. Once again the analytic definition of this matrix allows us to place it among other measures developed in both evolutionary biology and conservation biology. This formalisation leads us to introduce a second matrix (M), based on May (1990)'s propositions for measuring the taxonomic distinctiveness of a set of species. The performances of Moran's test with A and M in terms of power and Type-I error are compared with the performances of Moran's test used with previously defined matrices of phylogenetic proximities.

2. Material and methods

2.1. Data

First a simple, theoretical data set (Fig. 1) contains four hypothetical species and will serve to facilitate the description of Abouheif's statistics. A total of 22 trees were used for computer simulations. We first defined trees according to the following models: the symmetric model which provides trees with equal branch length among nodes and 2^n tips, where *n* is the number of bifurcations; the comb-like model which generates trees in which the tips are spread out like the teeth of a comb; Yule model which assumes that all species are equally likely to speciate. Trees have been generated from the three models with the following numbers of species: 8, 16, 32, 64, 128, 256. Four real trees have also been added to the simulations. The objective was to obtain a variety of tree shapes. All both are available in the packages ade4 and ape from the R Core Development Team 2007. We considered three small sized phylogenies: a 17-taxa maple phylogeny (data 'maples' in ade4 obtained from Ackerly and Donoghue (1998)), a 18-taxa lizard phylogeny (data 'lizards' in ade4 obtained from Bauwens and Díaz-Uriarte (1997)), a 19-taxa bird phylogeny (data 'procella' in ade4 obtained from Bried et al. (2003)). Finally we applied simulations to a 137-taxa bird phylogeny to study statistical performance on large sample sizes (data 'bird.families' in ape obtained from Sibley and Ahlquist (1990)).

Phenotypic trait data were generated using an Ornstein– Uhlenbeck (OU) process with functions available in the package ouch from the R Core Development Team 2007. The phenotype is held near a fixed optimum by a force measured by a parameter named α . When $\alpha \approx 0$, the OU process approximates the Brownian motion model. When α increases, the data become more and more independent. We scaled branch lengths on all phylogenies so that the maximum length from root to tips is equal to 1, and let α vary from 0, 1, 2 to 10 (Diniz-Filho, 2001). The OU model involves two other parameters. The first one (θ) is the value of the optima. We fixed $\theta = 0$. The second parameter (σ) measures the standard deviation expected at each generation due to random evolution along the branches of the phylogeny. We fixed $\sigma = 1$. For the comb-like model, the simulations were performed on the ultrametric tree (pendant edges regularly increases along the comb leading to the alignment of the tips on one line).

2.2. Background: discrepencies between Moran's and Geary's statistics in spatial analyses

Several tests for phylogenetic signal use methods created for the analysis of spatial data (Cheverud and Dow, 1985; Cheverud et al., 1985; Gittleman and Kot, 1990; Rohlf, 2001). Several measures of spatial autocorrelation exist in the literature. They quantify the degree to which the value of a quantitative trait in a location is correlated to its value in neighboring locations. Here we focus on two widely used statistics: Moran's (1948) *I* and Geary's (1954) *c*. We present below these two statistics and show how they can be used to measure phylogenetic autocorrelation.

In the measurement of spatial autocorrelation, the first step is the description of the neighboring relationships by means of a graph. The next step is the translation of these defined relationships into a matrix of neighboring relationships Δ , whose general term δ_{ij} is equal to 1 if individuals *i* and *j* are neighbors and 0 if they are not. Let $\mathbf{R} = \text{diag}$ $\{r_1, r_2, ..., r_n\}$ be the diagonal matrix containing the row sums $r_i = \sum_{j=1}^n \delta_{ij}$ for Δ . Denote $\mathbf{x} = (x_1, x_2, ..., x_n)^t$ the values taken by a given trait *X* for each of the individuals, \bar{x} the average value and $\mathbf{z} = (z_1, z_2, ..., z_n)^t$ the corresponding standardized values of the trait *X*:

$$z_i = (x_i - \bar{x}) / \sqrt{\frac{1}{n} \sum_{k=1}^n (x_k - \bar{x})^2}.$$

Moran's I and Geary's c are defined as

$$I = \frac{\mathbf{z}^t \Delta \mathbf{z}}{\sum_{i=1}^n \sum_{j=1}^n \delta_{ij}} \quad \text{and} \quad c = \frac{n-1}{n} \frac{\mathbf{z}^t (\mathbf{R} - \Delta) \mathbf{z}}{\sum_{i=1}^n \sum_{j=1}^n \delta_{ij}}.$$

In a phylogenetic context, instead of individuals we have taxa, say species. Because the phylogenetic links among species depends on ancestry and evolutionary time, a matrix of phylogenetic proximity is usually not reduced to binary values (1 = neighbor, 0 = not neighbor, Gittleman and Kot, 1990). In fact defining such binary values would mean to define a finite number of clades in the trees, and declare that all species from a single clade are neighbor and two species belonging to two different clades are not neighbors, which would considerably reduce the possibilities for defining phylogenetic proximities. Therefore, we use the generalized version of Moran's and Geary's statistics (Upton and Fingleton, 1985) replacing the binary matrix Δ with a proximity matrix $\mathbf{W} = [w_{ij}]$, where $w_{ij} \ge 0$ and $w_{ij} = w_{ji}$.

We will test the null hypothesis (H_0) of no phylogenetic autocorrelation under the assumption that the observations are random independent drawings from one (or separate identical) population(s) with unknown distribution function. A non-parametric test is defined. For a given trait X, the observed values $(x_1, x_2, ..., x_n)$ are randomly permuted around the species, while W is kept unchanged. It is assumed that each species is equally likely to receive a value from $(x_1, x_2, ..., x_n)$. For each permutation, the statistic I (respectively c) is calculated. The proportion of randomized I (respectively c) higher (respectively lower) than the observed I (respectively c) indicates whether the observed I (respectively c) is improbable enough to reject the null hypothesis that there is no phylogenetic autocorrelation in the data. The choice of matrix W is not without consequences. It implies a model of evolution and will influence the power of the test.

These two statistics I and c either in their original Δ or in their generalized W form are at the core of two schools of thought. The first school advocates the advantages of Geary's c statistic, stating for example that c is more sensitive to the absolute differences between pairs of values, whereas I is more sensitive to extreme values (Jumar et al., 1977). The second school advocates the advantages of Moran's I statistic, mainly because Cliff and Ord (1981) demonstrated that I is more powerful than c, that is to say the probability of rejecting H_0 given that H_1 is effectively true is higher when I is used rather than c. Actually, the segregation in two distinct schools relies more on mathematical properties than on their consequences for the results of the tests. When applied to real data, the results obtained either by I or by c are very similar (Cliff and Ord, 1981, p. 170; Upton and Fingleton, 1985).

Thioulouse et al. (1995) provided some reconciliation between these two statistics, by applying two modifications. First Geary's statistic considers variance computed with 1/(n-1) rather than 1/n whereas Moran's statistic uses the latter. Therefore Thioulouse et al. suggested to unify the use of variances by dividing Geary's index by (n-1)/nleading to

$$c_n = \frac{\mathbf{z}^t (\mathbf{R} - \mathbf{W}) \mathbf{z}}{\sum_{i=1}^n \sum_{j=1}^n w_{ij}},$$

while *I* is unchanged. Secondly they introduced a vector of neighborhood weights to standardize the data: let $r_i^* = \sum_{j=1}^n w_{ij} / \sum_{i=1}^n \sum_{j=1}^n w_{ij}$ be the weight attributed to species *i*,

$$z_{i} = \left(x_{i} - \sum_{j=1}^{n} r_{j}^{*} x_{j}\right) / \sqrt{\sum_{k=1}^{n} r_{k}^{*} \left(x_{k} - \sum_{j=1}^{n} r_{j}^{*} x_{j}\right)^{2}}$$

leading to I^* and c_n^* . In a phylogenetic context, we will call the weights r_i^* "relatedness weights". Thioulouse et al.

(1995) demonstrated that these two modifications unify Moran's and Geary's concepts with the simple relationship $I^* + c_n^* = 1$. The consequence is that the tests based on either I^* or c_n^* are identical and the two measures are complementary: I^* measures local correlation while c_n^* measures local variation. However, the introduction of these relatedness weights r_i^* is not without consequences for the results of the tests. Indeed, the weights r_i^* are higher for species close in average from all others in the tree, which is characterized by more internal nodes, speciation events between them and the root node, that is to say, species belonging to species-rich clades. As a result, the isolated couples of species in a phylogenetic tree have much less influence on test results, while, because they are isolated from the rest of the tree, their similarities with one another together with their deferences from the rest of the species would reinforce a phylogenetic signal.

We will show in the three next sections that a formal definition of Abouheif's (1999) test unifies Abouheif's, Moran's, and Geary's tests, this time giving all species equal weights.

2.3. Presentation of Abouheif's test

Abouheif's (1999) Cmean statistic is defined as

$$C_{mean} = 1 - \frac{\frac{1}{T} \sum_{i=1}^{T} C_i}{2 \sum_{j=1}^{n} (x_j - \bar{x})^2},$$

where $C_i = \sum_{j=1}^{n-1} (x_{K_i(j)} - x_{K_i(j+1)})^2$. In this formula $x_{K_i}(j)$ denotes the observed phenotypic trait for the species $K_i(j)$. Imagine that the tree topology is displayed with all the tips aligned with one column as in Figs. 1A and B. The function K_i represents the tips' order for a given representation *i* of the tree topology from the top of the column to its bottom. For example, on the first topology of Fig. 1B at the top lefthand corner, the first species, "species a", is placed at the third position in the column of tips' letters, that is to say $K_1(3) = 1$. In the 12th which is the last topological representation at the bottom right-hand corner of this Fig. 1B, "species a" is placed at the second position, that is to say $K_{12}(2) = 1$. When the number of tips increases, it quickly becomes impractical to manage all the possible representations. Abouheif (1999) considers then an approximate solution by sampling a subset of 2999 random rotations of the topology (with rotated nodes), using a program called 'Phylogenetic Independence' developed by J. Reeve and E. Abouheif and available at http:// biology.mcgill.ca/faculty/abouheif/. The statistic (C_{mean}) serves to test for the null hypothesis of no phylogenetic autocorrelation (hypothesis H_0). The non-parametric test of H_0 proposed by Abouheif (1999) is close to those proposed for Moran's and Geary's statistics (Cliff and Ord, 1981). It consists of randomly permuting the original values $(x_1, x_2, ..., x_n)$ 999 times so that the species' values are randomly placed on the tips of the original phylogenetic topology. For each permutation, the statistic C_{mean} is

calculated. This procedure is done repeatedly until a distribution of C_{mean} is obtained. The number of randomized C_{mean} higher than the observed C_{mean} indicates whether the observed C_{mean} is improbable enough to reject the null hypothesis that there is no phylogenetic autocorrelation in the data.

2.4. Abouheif's test turns out to be a Moran's test

We discovered that Abouheif's test turns out to be an application of Moran's test with a special phylogenetic proximity matrix A (Fig. 1C and D). We show below that C_{mean} can be rewritten as a Moran's *I* statistic. We demonstrate that the test proposed by Abouheif leads to a new matrix of phylogenetic proximity. Indeed, the C_{mean} can be rewritten as

$$C_{mean} = 1 - \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} (x_i - x_j)^2}{2 \sum_{i=1}^{n} (x_i - \bar{x})^2},$$

where a_{ij} is the general term of a phylogenetic proximity matrix A. Each off-diagonal term a_{ij} is equal to the frequency of rotations of the nodes which put species *j* just behind species *i*. The values of the diagonal terms of A do not change C_{mean} because $a_{ii}(x_i-x_i)^2 = 0$, whatever a_{ii} . We are thus free to choose the diagonal values that we think most appropriate. We choose that each diagonal term a_{ii} is equal to the frequency of representations which put species *i* at one extremity of the sequence of the tips, i.e. after all the other species. This choice leads to $\sum_{j=1}^{n} a_{ij} = 1$. Matrix A is thus symmetric and each row, as well as each column, has a sum equal to 1. With matrix A the relatedness weights are all equal to $r_i^* = 1/n$ so that means and variances in Abouheif's test are unweighted. This matrix A has the following interesting properties. It is a $n \times n$ matrix of components a_{ij} satisfying $a_{ij} = a_{ji}$, $a_{ij} > 0$ and $\sum_{i=1}^{n} a_{ij} = 1$. By verifying this property, matrix A is said to be "doubly stochastic".

The row weights associated with **A** are thus all equal to 1/n and $\sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} = n$. Consequently, it can be shown that

$$C_{mean} = 1 - \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} (z_i - z_j)^2}{2n}$$
$$= 1 - \frac{\sum_{i=1}^{n} z_i^2 - \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} z_i z_j}{n}$$

and thus

$$C_{mean} = 1 - \frac{\mathbf{z}^{t}(\mathbf{I}_{n} - \mathbf{A})\mathbf{z}}{n} = 1 - \frac{\mathbf{z}^{t}(\mathbf{I}_{n})\mathbf{z}}{n} + \frac{\mathbf{z}^{t}\mathbf{A}\mathbf{z}}{n}$$
$$= 1 - \frac{\mathbf{z}^{t}(\mathbf{I}_{n})\mathbf{z}}{n} + \frac{\mathbf{z}^{t}\mathbf{A}\mathbf{z}}{\sum_{i=1}^{n}\sum_{j=1}^{n}a_{ij}},$$

where \mathbf{I}_n denotes the identity matrix. Given that $\mathbf{z}^t(\mathbf{I}_n)\mathbf{z} = n$,

$$C_{mean} = \frac{\mathbf{z}^{t} \mathbf{A} \mathbf{z}}{\sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij}}$$

As a result, C_{mean} is equal to Moran's *I* when one choose matrix **A** as a proximity matrix (see Appendix A for more explanations). Because **A** is doubly stochastic, using **A** in Moran's *I* leads to $I + c_n = 1$.

2.5. Description of the new matrix of phylogenetic proximity (A)

For unrooted trees, the rotation of nodes in a tree is called cyclic permutation. For a given cyclic permutation the arrangement of the set of leaves is called "cyclic ordering". Cyclic permutations are studied to compare trees and to analyze tree metrics (functions defining distances between tips, Semple and Steel, 2004).

We can show now that matrix **A** has quite a simple analytic expression for all phylogenies, whether resolved or not (demonstration in the Appendix **B**):

$$a_{ij} = \frac{1}{\prod_{p \in P_{ij}} dd_p}$$
, for $i \neq j$ and $a_{ii} = \frac{1}{\prod_{p \in P_{iRoot}} dd_p}$.

For a species *i*, a_{ii} is therefore the inverse of the product of the number of branches descending from each node from the species to the root. For a couple (*i*, *j*), a_{ij} is the inverse of the product of the number of branches descending from each node in the path connecting *i* and *j*.

Now that **A** has been analytically resolved, the test proposed by Abouheif can be computed without referring to the cyclic permutation.

Furthermore, we presented above the diagonal terms a_{ii} as the residuals of a processus to render matrix **A** doubly stochastic, but they are more than that. They have a biological meaning: they measure originality *sensu* Pavoine et al. (2005). The originality of a species provides a single-species measure of cladistic distinctiveness. It measures how evolutionarily isolated a species is relative to other members (tips) of a phylogenetic tree. The more a species is in average distant from the other, the more it is original. May's (1990) provided such an index of originality whose formula is exactly a_{ii} except the product is replaced by a sum:

$$m_{ii} = \frac{1}{\sum_{p \in P_{iRoot}} dd_p}.$$

The first consequence of this relation between matrix **A** and the indices of phylogenetic originality is that matrix **A** contains elements developed for ecological studies, evolutionary studies and conservation biology, when the need for establishing bridges between these three disciplines has been pointed out as urgently necessary. The second consequence is that we can now define a second matrix (**M**) of phylogenetic proximities, where

$$m_{ij} = \frac{1}{\sum_{p \in P_{ij}} dd_p}$$
, for $i \neq j$ and $m_{ii} = \frac{1}{\sum_{p \in P_{iRoot}} dd_p}$.

Note that **M** is not doubly stochastic.

We provide then a formalization of Abouheif's test both in mathematical and biological terms.

Redefinition of Abouheif Test: The Abouheif's test is a Moran's test with a specified matrix of phylogenetic proximity **A**:

- The non-diagonal values of **A**, *a_{ij}*, measure the proximity between two species *i* and *j* and are equal to the inverse of the product of the number of branches descending from each interior node in the path connecting *i* to *j*.
- The diagonal values of **A**, *a_{ii}*, measure the originality of species *i* as the inverse of the product of the number of branches descending from each interior node in the path connecting *i* to the root of the tree.

Semple and Steel (2004) introduced a calculation under the same realms as matrix **A** but for unrooted trees: for unrooted trees, the proportion of circular orderings for which *j* immediately follows *i* is $\lambda_{ij} = \prod_{p \in Q_{ij}} (dd_p)^{-1}$ for $i \neq j$,

where Q_{ij} is the set of interior nodes in the path connecting *i* and *j*. Fixing $\lambda_{ii} = 0$, we obtain a matrix of phylogenetic proximity for unrooted tree. Denote Λ the matrix $[\lambda_{ij}]$. By definition, Λ is doubly stochastic.

Semple and Steel (2004) discovered an interesting property for Λ : let's d_{ij} be the sum of the branch lengths in the path connecting *i* and *j*, then

$$PD = \sum_{ij} \lambda_{ij} d_{ij}$$

is the sum of all the branch lengths on the tree, which is a measure of phylogenetic diversity (Faith, 1992), used in conservation biology.

Proposition. Consider $d_{ij}^* = d_{ij}$ for $i \neq j$ and $d_{ii}^* = \sum_{j|P_{ij}=1}^{j} \{Root\}a_{jj}d_{ij}dd_{Root}(dd_{Root}-1)$, then $PD = \sum_{ij}a_{ij}d_{ij}^*$. (Proof in Appendix C).

Roughly speaking, d_{ii}^* concerns for species *i* "what happens on the other side of the root".

Once again, this result contributes to filling in the gap between ecology, evolutionary biology and conservation biology.

2.6. Estimating the effect of the matrix of phylogenetic proximity in Moran's test

For each simulated tree, we calculated the matrix **A**. We chose to compare the use of matrix **A** in Moran's test with four other matrices of proximities among pairs of species, which were proposed for trees with equal branch lengths. The first one is the new matrix **M**. With the second one, named **B**, the proximity between the two species is the number of internal nodes, or taxonomic levels from the first common node to the root. When branch lengths are available and summed rather than counting the numbers of nodes, **B** is the matrix of phylogenetic variance–covariance given a Brownian motion model of character changes

(e.g. Felsenstein, 1985). For the third matrix, denoted C, Cheverud and Dow (1985) first defined the 4-point metric distance d_{ii} between the two species by the number of internal nodes connecting the two species to a common ancestor. According to Cheverud and Dow (1985) and Cheverud et al. (1985), we should consider in the phylogenetic tree a level at which species are said unrelated, and if two species are unrelated, a zero is entered in matrix C. We considered that the species connected only at the root node are unrelated, but other choices could be made. for example Cheverud et al. (1985) truncated the tree at the family level in a taxonomy. The proximity is then defined as $1/d_{ii}$ if $d_{ii} \neq 0$, and 0 elsewhere. The particularity of this matrix is the zero values on the diagonal, while the proximity of a given species with itself should be maximum. Matrix C was used by Cheverud and Dow (1985) in the phylogenetic autoregressive method which they developed for distinguishing between the phylogenetic effect and the specific effect on variation in trait values: $v = \rho \mathbf{C} v + \varepsilon$, where v is the normalized vector of observed data, ρ is called the "autoregressive coefficient". For the fourth matrix called **D**, we apply the exponent α proposed by Gittleman and Kot (1990; see Martins and Hansen, 1996) to Cheverud and Dow (1985) proximity matrix. In the phylogenetic autoregressive method, the values of both α and ρ were obtained by maximum likelihood or more correctly by least squares (Rohlf, 2001). The objective of using α was to improve the performance of the phylogenetic autoregressive method developed by Cheverud and Dow (1985) and Cheverud et al. (1985) in distinguishing between the phylogenetic effect ($\rho \mathbf{C} y$) and the specific effect (ɛ) on variation in trait values. Owing to Martins and Hansen (1996) another objective was to best stretch and shrink the phylogeny. Here, we used the value of α which maximizes Mantel (1967) correlation among A and D. One of our objectives is to highlight the high congruence among A and **D**. We also studied the effect of α by varying its value from 0.1 to 3 using a step equal to 0.1, from 3 to 10 using a step of 1 and from 10 to 100 using a step of 10.

Moran's tests were performed on the 22 trees, containing four real trees, six symmetric trees, six comb-like trees and six yule-model trees. We performed first Type I error tests. For a tree with *n* species, 1000 data sets of size *n* were independently and randomly drawn from a normal distribution N(0,1). We applied a Moran test for each simulated data set and measured the type I error as the percentage of significant tests at the nominal 5% level. We performed then a series of power tests. For a tree with *n* species, and a fixed value of the constraining force, 1000 data sets of size *n* were simulated from a OU process, as indicated in Section 2.1. The power was then measured as the percentage of significant tests at the nominal 5% level.

For each tree and each simulated trait, Moran's tests are realized with the statistic I^* including species relatedness weights, and with 1000 random permutations of the values $(x_1, x_2, ..., x_n)$ around the species. All computations and graphical displays were carried out using R Core Development Team 2007, with both pre-programmed and personal routines. The data and routines for performing Moran's test are available in the 'ade4' package at http://lib.stat.cmu.edu/R/CRAN/ (Chessel et al., 2004). Instruction guidelines are available in the supplementary file.

3. Results

We use the following notation for the rest of the text: MT_A , MT_M , MT_B , MT_C , MT_D denotes Moran's test when I is measured with matrices A, M, B, C, D, respectively. What emerges from these simulations (Fig. 2) is that except for the comb-like model, the use of matrix **B** in Moran's I provides a less powerful test. With the Yule model, our simulations highlight a far lower power of I when it is used with **B**. With 256 species and a constraining force $\alpha = 10$ for the OU process, the power of MT_B is estimated equal to 0.368, while the estimated power of the four other tests varies from 0.962 to 0.985. With most of the trees, the power of MT_A and MT_D are very close, and for most of the trees except the comb-like trees, the power of Moran's test increases in the following order: $MT_B < MT_M < MT_C <$ $(MT_A \approx MT_D)$. For the comb-like model, the power of Moran's test increases in the inverse order: $MT_A <$ $MT_{D} < MT_{C} < MT_{M} < MT_{B}$, but the power of the five tests are in that case very close. (Fig. 3)

The type I error of all tests are close to the nominal 5% level (Table 1).

The five matrices differ in how they value a high proximity and a low proximity and the gap between them. The coefficients of variation (CV) for non-diagonal values of these five matrices are in average (over the 22 trees): 3.56 for matrix A, 2.58 for D, 1.40 for C, 0.97 for B, and 0.93 for M. Matrix D and, above all, matrix A provides the most contrasting values (high CV). Their CV highly increases with the number of species (Fig. 4). This difference in CV appears in the graphical representations of the matrices (Fig. 5). Furthermore, one can observe in Fig. 5 that the values near the diagonal for A and D (which are the estimates of the proximities among close species) are not identical but very close to each others. Despite **D** provides more contrast between very close species and less related species, only matrix A also provides clear distinctions among the proximities of less related species (values far from the diagonal). Note also that one of the differences between A, M and B, C, D is that A and M never consider that species connected only at the root node of the tree are not related.

Regarding the effect of an exponent on **C**, let $MT_{C\alpha}$ be the Moran's test used with C^{α} . For all trees studied, the power of $MT_{C\alpha}$ is first enhanced when α increases from 0.1 to reach a maximum for a value of α comprised between 1 and 2 in all our examples (Fig. 3). Then the power regularly decreases. In all cases, **D** is at or very close to the maximum power of $MT_{C\alpha}$ over α .

We highlighted that the diagonal terms of matrix **A** are measures of originality (single-species measure of cladistic

distinctiveness) sensu Pavoine et al. (2005). For all our 22 trees (Fig. 6), the rank correlation between the diagonal values a_{ii} of matrix **A** and May (1990)'s index (diagonal values of matrix **M**), one of the main indices of species originality, is equal to 1. The difference between the formulas of the a_{ii} and the m_{ii} is the use of a product instead of a sum. The consequence is that the a_{ii} decreases more quickly with the number of nodes between *i* and the most original species are more emphasized (Fig. 6). By definition, a_{ii} is equal to the frequency of representations which put species *i* at one extremity of the sequence of the tips, i.e. after all the other species. Consequently, another advantage of the a_{ii} is that $\sum_{i=1}^{n} a_{ii} = 1$ while $\sum_{i=1}^{n} m_{ii}$ depends on the tree shape and size.

4. Discussion

4.1. Rediscovering the link between Moran's I and Geary's c

Because A is a doubly stochastic matrix, when using A as a proximity matrix, Moran's I is equal to Abouheif's C_{mean} and to one minus Geary's c_n , thus the three statistical methods are brought back together in the same theoretical framework. In that case, the two statistics I and c_n provide complementary information. The statistic I measures the local autocorrelation, which is the degree to which related species are close from each other in a given trait, and the statistic c_n measures the local variability, that is to say the degree to which related species differ from each other.

4.2. Improving Abouheif's test and giving it new purposes

We improved Abouheif's test by providing its exact analytical value, while it was previously calculated by an approximate algorithm. We centered the test on a new matrix of phylogenetic proximities (A). This mathematical formalization leads to a clear biological definition of the Abouheif test: Abouheif test measures the proximity between two taxa *i* and *j* as the inverse of the number of branches descending from each interior node in the path connecting *i* to *j* the root of the tree. The proximity depends thus on the interior nodes in the path connecting *i* to *j*, and was previously approximated by a technical approach referring to the percentage of time *i* and *j* were found next to each other in the set of all cyclic permutations of the tree. The reference to cyclic permutations was thus unnecessary and can be thought as technical, even makeship job and disturbing because a cyclic permutation does not change topology. Despite that, cyclic permutations were here at the foundation of the three matrices A, M and Λ . Research on cyclic permutations has already proved useful for studies on tree metric and tree reconstitutions (Semple and Steel, 2004). A tree is mechanically embedded in a plane which implicitly demands to choose a way of arranging tips (a circular ordering of the tips). The existence of a finite number of



Fig. 2. Power tests for (A) the symmetric model, (B) the comb-like model, (C) the Yule model and (D) the observed, real trees. Simulations were done separately for each sample size from n = 8 to 256 species. The legends for the line symbols, drawing and color (black or gray) for the five matrices of phylogenetic proximity (A, M, B, C and D) are given in the box on the bottom left-hand corner. The power is given in the ordinate axis as a function of alpha, the constraining force of the OU process. Note that the powers obtained with A and D are very close so that the curves for A and D are often superimposed.

distinct circular orderings for a tree (Semple and Steel, 2004) constitutes one of the properties of the tree object which merits our attention. Abouheif's intuition can

reinforce an underestimated interest of the research on cyclic permutations for evolutionary and biological conservation studies.



Fig. 3. Effect of the exponent γ in \mathbf{C}^{γ} proposed by Gittleman and Kot (1990) on the power of Moran's test used with \mathbf{C}^{γ} for (A) the symmetric model, (B) the comb-like model and (C) the Yule model. The black square indicates the position of $D = \mathbf{C}^{\beta}$ where $\operatorname{cor}(\mathbf{C}^{\beta}, \mathbf{A}) = \max_{\gamma} [\operatorname{cor}(\mathbf{C}^{\gamma}, \mathbf{A})]$. The open circle indicates the position of **C**. The graphs are given for different values of the constraining force from $\alpha = 0$ to 10. The precise value of α for a given graph is given on the bottom right-hand corner of the graph.

Table 1 Average type I error for Moran's test at the nominal 5% level (with standard deviation in brackets)

	Symmetric	Comb-like	Yule	Real-trees
A	4.68 (0.44)	4.53 (0.22)	4.65 (0.61)	4.44 (0.59)
Μ	4.32 (0.59)	5.35 (1.21)	5.15 (0.43)	4.88 (0.35)
В	4.83 (0.32)	4.25 (0.96)	4.57 (0.68)	5.05 (0.99)
С	4.78 (0.41)	4.93 (0.88)	4.77 (0.76)	4.68 (0.81)
D	4.83 (0.63)	5.15 (0.72)	4.57 (0.55)	4.75 (0.83)
	. ,		. ,	

All values are given as percentages. The matrix used with Moran's test is given in the first column. For the symmetric, the comb-like and the Yule models, values are averaged over the six sample sizes. For the real trees, values are averaged over the four phylogenies considered.

4.3. Comparison between A, M and currently used proximity matrices

The following kinds of methods have been recommended when suspecting phylogenetic dependence. A first group of methods aims to test for a phylogenetic signal in each of the phenotypic traits under study, whatever the shape of this signal for example by measuring a correlation among sister-species (Gittleman and Kot, 1990). The second group of methods aims to describe the link between a phylogenetic tree and the states of one or several traits, either by searching at which level(s) in a phylogenetic tree, one can detect phylogenetic signal with methods such as



Fig. 4. Coefficients of variation (CV) for the five matrices **A**, **M**, **B**, **C** and **D** for (A) the symmetric model, (B) the comb-like model and (C) the Yule model. The legends for the line symbols, drawing and color (black or gray) for the five matrices of phylogenetic proximity (A, M, B, C and D) are given in the boxes on the right side of the figure. Each graph provides CV as a function of an index of the number of species on a logarithmic scale.



Fig. 5. Differences among the five matrices of phylogenetic proximities for (A) the symmetric model, (B) the comb-like model and (C) the Yule model. The name of the matrices is given in boxes on the right-hand corner of the figures. Each value in the matrix is represented by a square. The larger the square the higher is the value. Species are ordered by rows and columns according to the phylogenetic tree which is given.

nested ANOVA (Crook, 1965; Clutton-Brock and Harvey, 1977, 1979, 1984), correlogram (Gittleman and Kot, 1990; Rohlf, 2001, 2006), and orthonormal transform (Giannini, 2003; Ollier et al., 2006) or by modeling the evolution of the trait supposing that the real causes involving phylogenetic inertia and adaptation are precisely known (Martins

and Hansen, 1996; Blomberg et al., 2003; Bonsall and Mangel, 2004; Bonsall, 2006). A critical step in all these approaches is the proper specification of a phylogenetic proximity matrix. Indeed, many possibilities exist depending on whether branch lengths are known and on the model of macroevolution used, for examples pure neutral model



Fig. 6. The diagonal values of matrix \mathbf{A} are measures of phylogenetic originality: (A) Representation by Cleveland (1994) dot plots of the links between the diagonal values of \mathbf{A} and the diagonal values of \mathbf{M} (May's (1990) index) for the comb-like model, the Yule model, and the four real trees considered in the text. The congruence among the diagonal values of matrix \mathbf{A} and May's index is high but the differences in the originality values among species are higher for the diagonal values of matrix \mathbf{A} than for May's index, which leads in (B) to a parabolic shape for the relationship among the diagonal values of matrix \mathbf{A} and the values obtained from May's (1990) index may be small but never null.

(Brownian motion), stabilizing selection (Ornstein–Uhlenbeck), directional selection, accelerating and decelerating Brownian evolution (ACDC) (Blomberg et al., 2003) and more complex models (Mooers et al., 1999) and it is difficult to choose among them (Hansen and Martins, 1996).

In this context, matrix A constitutes a useful alternative. In this paper, we give the exact analytic expression of this matrix. We prove that it is a symmetric and doubly stochastic proximity matrix and because it is doubly stochastic, it unifies Moran's and Geary's statistics. We showed through the power tests, that there may be a strong effect of the choice of the proximity matrix in Moran's non-parametric test. Two main criticisms were raised to Abouheif's test: it does not rely on branch length and it is used due to technical reasons rather than because it relies on a precise model of character changes. By formalizing Abouheif's test we redefined it and clarified its biological foundations. Regarding the absence of branch length, if branch lengths are available, one should use them and have higher power, to the extent that they are accurate. Even if focusing on nodes means using a rather unlikely model in which branch lengths are assumed to be equal, the tree topology is one out of the two key components of this history.

Unlike **B**, the matrices **A** and **M** we introduced and the matrices **C** and **D** suggested by Cheverud and Dow (1985) and Gittleman and Kot (1990) are specially designed for topologies. The advantage of matrices **A** and **D** over **B**, **C**

and **M** is that they provide more contrasted species proximity estimates, which enhances the power of the tests. Our results showed that most of the time matrix **A** fits data sets better than existing matrices of phylogenetic proximities and when it does not fit better, it fits almost as well as other matrices. Matrix **A** and **M** also have the advantage of correcting proximities for unresolved trees. Indeed, for these two matrices the product or sum of the number of branches descending from nodes are counted instead of the simple number of nodes (see, May, 1990).

Gittleman and Kot (1990) proposed to add an exponent to matrix **C** to best-fit data. This led us to introduce matrix **D** with an exponent which maximizes its correlation with matrix **A**. We studied the effect of the exponent on **C**. Let $MT_{C\alpha}$ be the Moran's test used with C^{α} . For all trees studied, **D** was at or very close to the maximum power of $MT_{C\alpha}$ over α . This result reinforces the interest of **A** as a powerful matrix.

4.4. Matrix A, M and conservation biology

We emphasized the statistical advantages gained by using matrix **A** as a matrix of phylogenetic proximities among species. Actually the advantages gained by using matrix **A** are both statistical and biological, because matrix **A** has a biological meaning which has not been explored with previous matrices of phylogenetic proximity. The diagonal values of matrix **A** are important for two reasons: first because they give the doubly stochastic property to the whole matrix **A**, unifying then Moran's, Geary's and Abouheif's tests, and second because they are measures of cladistic originality, that is to say a_{ii} is a measure of how evolutionarily isolated species *i* is relative to other members of the phylogenetic tree under study. As we highlighted, the diagonal values of matrix **A** have the same capacity as May's index (diag(**M**)) to measure cladistic originality. They are even more segregating, giving a larger range of values from the less to the most original species. Consequently, the diagonal values of matrix **A** can be used as a powerful alternative to current indices of cladistic originality when designing conservation priorities.

Moran's test only uses the non-diagonal values of **A** and **M**. The challenge now is to evaluate the use of these complete two matrices in more complex analyses, such as multivariate analysis (cf. Thioulouse et al. (1995) for ordination under spatial autocorrelation) and comparative analyses.

In conclusion, the appealing qualities of matrix **A** are that it unifies Abouheif's, Moran's and Geary's tests; it increases the power of Moran's test; its diagonal values measures cladistic originality and it contributes to filling in the gap between evolutionary biology and conservation biology. The interest of matrix **A** for a wide variety of taxa and traits, and for a larger range of evolutionary and ecological issues has still to be proved but our first results are encouraging.

Acknowledgments

The authors would like to thank Michael Bonsall (University of Oxford, UK), Margaret Evans (University of Yale, USA), Sébastien Devillard, Clément Calenge, Stéphane Dray and Thibaut Jombart (University of Lyon, France), Patrick James Degeorges (French ministery of ecology and sustainable development) for their helpful comments on a first draft of this paper.

Appendix A. Abouheif's test is a Moran's test

We demonstrated in the text that

$$C_{mean} = \frac{\mathbf{z}^t \mathbf{A} \mathbf{z}}{\sum_{i=1}^n \sum_{j=1}^n a_{ij}}$$

Let us decompose matrix **A** into two additive components: $\Lambda_{\mathbf{A}}$ contains the diagonal values of **A**, and $\mathbf{M}_{\mathbf{A}}$ the non-diagonal values (with 0 on the diagonal); hence $\mathbf{A} = \Lambda_{\mathbf{A}} + \mathbf{M}_{\mathbf{A}}$. Moran's statistics was defined for matrices with zeros down the principal diagonal. Consider $S_A = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij}$ and $S_{M_A} = 2\sum_{i=1}^{n-1} \sum_{j>i}^{n} a_{ij}$, then

$$C_{mean} = \frac{\mathbf{z}^t \mathbf{\Lambda}_A \mathbf{z}}{S_A} + \frac{S_{M_A}}{S_A} \frac{\mathbf{z}^t \mathbf{M}_A \mathbf{z}}{S_{M_A}}.$$

During the randomization test, when the observed values $(x_1, x_2, ..., x_n)$ are randomly permuted around the species and S_{M_A}/S_A is a constant. Consequently,

 C_{mean} depends on two components: one linked to the originalities of the species and one linked to the proximities of couples of species.

This demonstration is still true for matrix M.

Appendix B. Analytical expression of the proximity matrix A

We propose to give explicitly the analytical expression of matrix **A**. These assertions will be illustrated by way of a simple theoretical example (Fig. 1).

We will consider the following terminology concerning the phylogenetic tree. The root is the common ancestor to the sets L and N of all the l contemporary taxons (OTUs: operational taxonomic units) and *n* interior nodes, respectively (HTUs: hypothetical taxonomic units). Branches emanate from the root and nodes, tracing the course of evolution. The definition of A is independent of branch length and depends only on the topology of the tree. The minimum spanning path between two taxonomic units (nodes or tips) defines an ordered set of nodes. For example in Fig. 1 A, the set $P_{ab} = \{A\}$ is associated with the path (a, A, b) that spans species a and b. Similarly, the set $P_{aRoot} = \{A, B\}$ defines the hypothetical ancestors associated to the path (a, A, B, Root) between species a and the root of the tree. We denote DD_i the set of direct descendants for node *i*, including hypothetical descendants at interior nodes and species at tips. For example, in Fig. 1A, $DD_A = \{a, b\}$ and $DD_B = \{A, c, d\}$. Let $dd_i = card(DD_i)$ be the number of direct descendants for node *i* ($dd_A = 2$ and $dd_B = 3$). The total number of consistent representations of the tree topology is defined by the product of rotations associated with each node: $T = \prod_{i \in N} dd_i!$, where dd_i denotes the number of direct

descendants of each node $i \in N$.

Among the set of equivalent representations of the phylogeny (Fig. 1B), there is at least one representation that puts tip j just behind tip i. Consider such a representation and observe what happens if a rotation occurs at one node $p \in N$. If $p \notin P_{ij}$, all the $dd_p!$ rotations associated with that node do not disturb the respective position of i and j. On the other hand, if $p \in P_{ij}$, only $(dd_p-1)!$ out of the $dd_p!$ possible rotations do not disturb the respective positions that put tip j just behind tip i is equal to the set of equivalent representations that put tip i just behind j, we deduce the analytical expression a_{ij} of the matrix A (Fig. 1C):

$$a_{ij} = \frac{I_{ij}}{T} = \frac{\prod_{p \in P_{ij}} (dd_p - 1)! \prod_{p \in N - P_{ij}} dd_p!}{\prod_{p \in N} dd_p!} = \frac{1}{\prod_{p \in P_{ij}} dd_p}$$

From this expression, we can deduce the diagonal values:

$$a_{ii}=1-\sum_{j=1,j\neq i}^n a_{ij}.$$

Therefore, the C_{mean} statistic is equal to Moran's statistic applied to the so-defined matrix A (Fig. 1D).

Appendix C. Proof for $PD = \sum_{ij} a_{ij} d_{ij}^*$

Because of the presence of a root in the tree, if $P_{ij} \neq \{Root\}, a_{ij} = \lambda_{ij}, \text{ and if } P_{ij} = \{Root\} a_{ij} = \lambda_{ij}/dd_{Root}$ and $a_{ii} = a_{ii}a_{ii}dd_{Root}$.

$$PD = \sum_{ij} \lambda_{ij} d_{ij}$$

= $\sum_{ij|P_{ij} \neq \{Root\}} a_{ij} d_{ij} + \sum_{ij|P_{ij} = \{Root\}} a_{ij} dd_{Root} d_{ij}$
= $\sum_{ij} a_{ij} d_{ij} + \sum_{ij|P_{ij} = \{Root\}} a_{ij} (dd_{Root} - 1) d_{ij}$
= $\sum_{ij} a_{ij} d_{ij} + \sum_{ij|P_{ij} = \{Root\}} a_{ii} a_{jj} dd_{Root} (dd_{Root} - 1) d_{ij}$
= $\sum_{ij} a_{ij} d_{ij} + \sum_{i} a_{ii} \sum_{j|P_{ij} = \{Root\}} a_{jj} d_{jj} dd_{Root} (dd_{Root} - 1).$

Consider $d_{ij}^* = d_{ij}$ for $i \neq j$ and $d_{ii}^* = \sum_{j \mid P_{ij} = ij} \{Root\}a_{ij}d_{ij}d_{Root}(dd_{Root} - 1)$ then $PD = \sum_{ij}a_{ij}d_{ij}^*$

For dichotomous trees,

$$PD = \sum_{ij} a_{ij}d_{ij} + \sum_{i} a_{ii} \sum_{j|P_{ij} \in \{Root\}} 2a_{ij}d_{ij}$$

Appendix D. Supplementary materials

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.tpb. 2007.10.001.

References

- Abouheif, E., 1999. A method for testing the assumption of phylogenetic independence in comparative data. Evol. Ecol. Res. 1, 895-909.
- Ackerly, D.D., Donoghue, M.J., 1998. Leaf size, sappling allometry, and Corner's rules: phylogeny and correlated evolution in Maples (Acer). Am. Nat. 152, 767-791.
- Bauwens, D., Díaz-Uriarte, R., 1997. Covariation of life-history traits in lacertid lizards: a comparative study. Am. Nat. 149, 91-111.
- Blomberg, S.P., Garland, T., Ives, A.R., 2003. Testing for phylogenetic signal in comparative data: behavioral traits are more liable. Evolution 57. 717-745.
- Bonsall, M.B., 2006. The evolution of anisogamy: the adaptative significance of damage, repair and mortality. J. Theor. Biol. 238, 198-210.
- Bonsall, M.B., Mangel, M., 2004. Life-history trade-offs and ecological dynamics in the evolution of longevity. Proc. Roy. Soc. Lond. Ser. B 271. 1143-1150.
- Bried, J., Pontier, D., Jouventin, P., 2003. Mate fidelity in monogamous birds: a re-examination of the Procellariiformes. Anim. Behav. 65, 235-246.
- Chessel, D., Dufour, A.-B., Thioulouse., J., 2004. The ade4 package -I- One-table methods. R News 4, 5-10.
- Cheverud, J.M., Dow, M.M., 1985. An autocorrelation analysis of genetic variation due to lineal fission in social groups of rhesus macaques. Am. J. Phys. Anthrop. 67, 113-121.

- Cheverud, J.M., Dow, M.M., Leutenegger, W., 1985. The quantitative assessment of phylogenetic constraints in comparative analyses: sexual dimorphism in body weight among primates. Evolution 39, 1335-1351.
- Cleveland, W.S., 1994. The Elements of Graphing Data. AT&T Bell Laboratories, Murray Hill, New Jersey.
- Cliff, A.D., Ord, J.K., 1981. Spatial Processes. Model & Applications. Pion, London, UK.
- Clutton-Brock, T.H., Harvey, P.H., 1977. Primate ecology and social organization. J. Zool. Lond. 183, 1-39.
- Clutton-Brock, T.H., Harvey, P.H., 1979. Comparison and adaptation. Proc. Natl. Acad. Sci. USA 205, 547-565.
- Clutton-Brock, T.H., Harvey, P.H., 1984. Comparative approaches to investigating adaptation. In: Krebs, J.R., Davies, N.B. (Eds.), Behavioral Ecology: An Evolutionary Approach. Blackwell Press, Oxford, England, pp. 7-29.
- Crook, J.H., 1965. The adaptive significance of avian social organization. Symp. Zool. Soc. Lond. 14, 181-218.
- Diniz-Filho, J.A.F., 2001. Phylogenetic autocorrelation under distinct evolutionary processes. Evolution 55, 1104-1109.
- Faith, D.P., 1992. Conservation evaluation and phylogenetic diversity. Biol. Conserv. 61, 1-10.
- Felsenstein, J., 1985. Phylogenies and the comparative method. Am. Nat. 125. 1-15.
- Geary, R.C., 1954. The contiguity ratio and statistical mapping. Inc. Stat. 5. 115-145.
- Giannini, N.P., 2003. Canonical phylogenetic ordination. Syst. Biol. 52, 684-695
- Gittleman, J.L., Kot, M., 1990. Adaptation: statistics and a null model for estimating phylogenetic effects. Syst. Zool. 39, 227-241.
- Hansen, T.F., Martins, E.P., 1996. Translating between microevolutionary process and macroevolutionary patterns: the correlation structure of interspecific data. Evolution 50, 1404-1417.
- Jumar, P.A., Thistle, D., Jones, M.L., 1977. Detecting two-dimensional spatial structure in biological data. Oecologia 28, 109-123.
- Mantel, N., 1967. The detection of disease clustering and a generalized regression approach. Cancer Res. 27, 209-220.
- Martins, E.P., Hansen, T.F., 1996. The statistical analysis of interspecific data: a review and evaluation of phylogenetic comparative methods. In: Martins, E.P. (Ed.), Phylogenies and the Comparative Method in Animal Behavior. Oxford University Press, Oxford, pp. 22-27.
- Martins, E.P., Hansen, T.F., 1997. Phylogenies and the comparative method: a general approach to incorporating phylogenetic information into the analysis of interspecific data. Am. Nat. 149, 646-667.
- May, R.M., 1990. Taxonomy as destiny. Nature 347, 129-130.
- Mooers, A.O., Vamosi, S.M., Schluter, D., 1999. Using phylogenies to test macroevolutionary hypotheses of trait evolution in cranes (Gruinae). Am. Nat. 154, 249-259.
- Moran, P.A.P., 1948. The interpretation of statistical maps. J. Roy. Stat. Soc. B 10, 243-251.
- Ollier, S., Couteron, P., Chessel, D., 2006. Orthonormal transform to decompose the variance of a life-history trait across a phylogenetic tree. Biometrics 62, 417-477.
- Pavoine, S., Ollier, S., Dufour, A.B., 2005. Is the originality of a species measurable? Ecol. Lett. 8, 579-586.
- Rohlf, F.J., 2001. Comparative methods for the analysis of continuous variables: geometric interpretations. Evolution 55, 2143-2160.
- Semple, C., Steel, M., 2004. Cyclic permutations and evolutionary trees. Advances in Applied Mathematics 32, 669-680.
- Sibley, C.G., Ahlquist, J.E., 1990. Phylogeny and Classification of Birds: A Study in Molecular Evolution. Yale University Press, New Haven.
- Thioulouse, J., Chessel, D., Champely, S., 1995. Multivariate analysis of spatial patterns: a unified approach to local and global structures. Environ. Ecol. Stat. 2, 1-14.
- Upton, G., Fingleton, B., 1985. Spatial Data Analysis by Example. Wiley, Chichester
- Von Neumann, J., Kent, R.H., Bellinson, H.R., Hart, B.I., 1941. The mean square successive difference. Ann. Math. Stat. 12, 153-162.