Euclidean Nature of Phylogenetic Distance Matrices

DAMIEN M. DE VIENNE1,2,*, GABRIELA AGUILETA1,2, AND SÉBASTIEN OLLIER3

1Département Génétique et Ecologie Évolutions, unité Ecologie, Systématique et Évolution, Université Paris-Sud 11, UMR 8079 UPS-CNRS-AGROPARISTECH, 91405 Orsay cedex, France;
2Bioinformatics and Genomics program, Center for Genomic Regulation (CRG), C/Dr. Aiguadé, 88, Barcelona 08003, Spain; E-mail: gabriela.aguileta@crg.es; and
3Département Biodiversité Systématique et Évolution, unité Ecologie, Systématique et Évolution, Université Paris-Sud 11, UMR 8079 UPS-CNRS-AGROPARISTECH, 91405 Orsay cedex, France; E-mail: sebastien.ollier@u-psud.fr.

*Correspondence to be sent to: Bioinformatics and Genomics program, Center for Genomic Regulation (CRG), C/Dr. Aiguadé, 88, Barcelona 08003, Spain; E-mail: damien.de-vienne@crg.es.

Received 22 November 2010; reviews returned 11 January 2011; accepted 7 March 2011

Associate Editor: Cécile Ané

Abstract.—Phylogenies are fundamental to comparative biology as they help to identify independent events on which statistical tests rely. Two groups of phylogenetic comparative methods (PCMs) can be distinguished: those that take phylogenies into account by introducing explicit models of evolution and those that only consider phylogenies as a statistical constraint and aim at partitioning trait values into a phylogenetic component (phylogenetic inertia) and one or multiple specific components related to adaptive evolution. The way phylogenetic information is incorporated into the PCMs depends on the method used. For the first group of methods, phylogenies are converted into variance–covariance matrices of traits following a given model of evolution such as Brownian motion (BM). For the second group of methods, phylogenies are converted into distance matrices that are subsequently transformed into Euclidean distances to perform principal coordinate analyses. Here, we show that simply taking the elementwise square root of a distance matrix extracted from a phylogenetic tree ensures having a Euclidean distance matrix. This is true for any type of distances between species (patristic or nodal) and also for trees harboring multifurcating nodes. Moreover, we illustrate that this simple transformation using the square root imposes less geometric distortion than more complex transformations classically used in the literature such as the Cailliez method. Given the Euclidean nature of the elementwise square root of phylogenetic distance matrices, the positive semidefinitiveness of the phylogenetic variance–covariance matrix of a trait following a BM model, or related models of trait evolution, can be established. In that way, we build a bridge between the two groups of statistical methods widely used in comparative analysis. These results should be of great interest for ecologists and evolutionary biologists performing statistical analyses incorporating phylogenies. [Biodiversity estimation; coevolution; cospeciation; functional diversity; multivariate analysis; phylogeny.]

As emphasized by Felsenstein (1985), “phylogenies are fundamental to comparative biology; there is no doing it without taking them into account.” The reason for this statement is straightforward: evolutionarily closely related species share many similarities by descent that could confound comparative studies if they were not considered (Harvey and Pagel 1991). A number of phylogenetic comparative methods (PCMs) have been developed in the last 25 years in order to evaluate the degree of correlated evolution between two traits, to reconstruct the ancestral states of characters, to estimate the degree of phylogenetic inertia, or to evaluate the rate of phenotypic evolution (Martins and Hansen 1997). These statistical methods can be divided into two groups: those that take into account phylogenies by introducing explicit models of evolution (e.g., phylogenetic independent contrasts [PIC], phylogenetic generalized least-square [PGLS]) and those that only consider phylogenies as a statistical constraint and aim at quantifying the amount of phylogenetic inertia present in the data (e.g., phylogenetic autoregressive model [ARM], phylogenetic eigenvectors regression [PVR] method). Among the latter, the use of phylogenetic distance matrices constitutes one of the most common strategies to introduce phylogenetic information into a statistical framework. Often, these matrices are required to be Euclidean for a standard treatment of distances that facilitates further statistical analyses.

Among the first group of PCM methods, the PIC method of Felsenstein (1985) is the more widely used PCM. Independent contrasts are calculated as differences (“contrasts”) between the character values of pairs of species and/or nodes, divided by the square root of the sum of their branch lengths. This method assumes that the data have resulted from an evolutionary process similar to Brownian motion (BM) (i.e., random walk). These contrasts can then be used in regression, analysis of variance, analysis of covariance, or any other appropriate analysis. The PIC method was improved later to take polytomies ( Purvis and Garland 1993), sampling error and within-species variation into account (Felsenstein 2008). The PGLS model (Martins and Hansen 1997) is an extension of the independent contrasts method that allows for other evolutionary scenarios to be considered. PIC and PGLS do not involve distance matrices but define a phylogenetic covariance matrix of a trait following a model of evolution. In the case of BM, this matrix contains in its diagonal the sums of the branch lengths between each taxon and the root (i.e., the height of each taxon in the tree) and in the off-diagonal the covariances between pairs of taxa (i.e., the height of their most recent common ancestor). Note that this is also
true for other models of trait evolution on transformed branch lengths such as OU (Uhlenbeck and Ornstein 1930), lambda (Pagel 1999), or ACDC (Blomberg et al. 2003). This matrix, sometimes called Σ (Rohlf 2001), C (Garland and Ives 2000; O’Meara et al. 2006) or V (Martins and Hansen 1997) is positive semidefinite, and even though this property is widely accepted in the statistical literature, it has not been noted in biological journals.

The second group of methods aims at partitioning trait values into a phylogenetic component (i.e., inherited) and one or multiple-specific components related to adaptive evolution. Methods for estimating this phylogenetic component (also referred to as “phylogenetic inertia,” see Blomberg and Garland 2002 for a definition of the concept) include the phylogenetic autoregressive model (ARM; Cheverud et al. 1985) or a maximum likelihood–based method (Lynch 1991). However, Lynch’s (1991) maximum likelihood–based method is computationally intensive and difficult to apply to real data, and the phylogenetic ARM (Cheverud et al. 1985) has very poor statistical performance when data have little phylogenetic correlation or when the number of species studied is too small to detect it (Martins 1996; Diniz-Filho et al. 1998). Diniz-Filho et al. (1998) proposed a new method to estimate and correct for phylogenetic inertia, called the phylogenetic eigenvectors regression method (PVR). In this method, a principal coordinate analysis (PCO; Gower 1966) is performed on a double-centered phylogenetic distance matrix (extracted from a phylogenetic tree or the aligned sequences directly), resulting in eigenvalues (that represent the variance explained by each principal coordinate) and eigenvectors (representing the phylogenetic information) on which trait values of interest can be regressed. Diniz-Filho et al. (1998) proposed to keep only a subset of the eigenvectors obtained by PCO by reference to a broken-stick model (Barton and David 1956; Frontier 1976), thereby eliminating the problem of negative eigenvalues obtained when phylogenetic distance matrices are not Euclidean. However, as suggested by Desdevises et al. (2003), “there is no particular reason why a broken-stick model would preferably select principal coordinates that are of importance for the explanation of the dependent variables.” It is indeed accepted that the first eigenvalues (i.e., higher values) represent deep splits in the phylogeny whereas the smaller values represent splits near the tips of the tree (Martins and Housworth 2002; Ané 2008). Removing some eigenvalues and eigenvectors thus implies some loss of potentially important information for the comparative analysis. Desdevises et al. (2003) proposed an amelioration of the PVR method, taking into account all principal coordinates and thus requiring the conversion of non-Euclidean phylogenetic distance matrices into Euclidean ones prior to the analysis. Another method using PCO on phylogenetic distance matrices was developed by Legendre et al. (2002) to test for the independence between host and parasite, genetic or patristic distances. This test is implemented in the program ParaFit and is widely used in the so-called cophylogenetic literature. In this case too, phylogenetic distance matrices need to be Euclidean prior to the principal coordinate analysis and, if they are not, they need to be artificially modified to satisfy this condition. Surprisingly, despite the success of the PVR and the ParaFit methods and the requirement of a Euclidean distance matrix to perform the principal coordinate analysis underlying it, some important issues have not, as far as we know, been addressed: (i) the relative proportion of Euclidean versus non-Euclidean phylogenetic distance matrices obtained from random trees; and (ii) the distortion imposed by the methods typically used to convert non-Euclidean distances into Euclidean ones. Even more surprising is the fact that studies using PVR and the modification proposed by Desdevises et al. (2003) usually do not even mention the method employed (when necessary) to convert non-Euclidean phylogenetic distance matrices into Euclidean ones. Note that the Euclidean transformation concerns only distances in non-ultrametric trees, as the patristic distances in ultrametric trees (i.e., where all tips are at the same distance from the root) are already Euclidean (Pavoine, Ollier, Pontier 2005). Alternatively, non-ultrametric trees can be rendered ultrametric by relaxing the global clock assumption or by post hoc tree transformation (e.g., penalized likelihood rate smoothing, see: http://loco.biosci.arizona.edu/rls/).

Here we propose a new way to obtain Euclidean phylogenetic distance matrices. It only requires taking the elementwise square root of the patristic phylogenetic distance matrix. We demonstrate that this “new distance” is always Euclidean and that this property is also true for unrooted trees and for trees harboring polytomies.

Moreover, this simple way to obtain Euclidean phylogenetic distance matrices is more appropriate than usual methods because it limits the distortion of the original distances. For example, the Cailliez (1983) method introduces distortion by adding a constant to all elements in the distance matrix, whereas other methods introduce distortion by simply removing negative eigenvalues (similar to what is done in Diniz-Filho et al. 1998). This property gives a strong justification for the present work, as distance matrices extracted from phylogenetic trees are frequently used in comparative studies using the PVR method and in cophylogenetic studies using ParaFit. Moreover, our demonstration of the Euclidean nature of the square root of phylogenetic distance matrices requires that the phylogenetic covariance matrix of a trait following a BM (or similar) model of evolution is positive semidefinite. This result is well known in the statistical literature, but it has not been acknowledged in biological journals. In that way, we build a bridge between the two groups of statistical methods widely used in comparative analysis.

In the present paper, we show first that the phylogenetic covariance matrix, C, of a trait following a BM model of evolution is positive semidefinite. This is also true for other models of trait evolution with
transformed branch lengths. Second, we prove that the elementwise square root of any phylogenetic distance matrix $D$ is Euclidean. Third, we illustrate by simulations that the distance matrix $D$ is not typically Euclidean. Finally, we underline that the simple square root of $D$ results in a matrix with less distortion than other popular methods of transforming the pairwise distance matrix to render it Euclidean.

**Proofs**

**Background**

**Euclidean distance matrix.**—A distance matrix $D = [d_{ij}]$ is said to be Euclidean if one can find $n$ points $M_k$ ($k = 1, \ldots, n$) in a multidimensional space such that the respective Euclidean distance between $M_i$ and $M_j$ (calculated with the classical Euclidean metric) generates all the $d_{ij}$ values (Gower and Legendre 1986). The classical Euclidean distance between two points in a multidimensional space, also referred to as the “ordinary distance,” is the distance that one can measure with a ruler. Thus follows that for any Euclidean distance matrix $D$, for any two points $M_i$ and $M_j$, the distance separating them can be computed by

\[d^2_{ij} = \|M_i - M_j\|^2 = \|M_i\|^2 + \|M_j\|^2 - 2M_i \cdot M_j\]  \hspace{1cm} (1)

where $\| \|$ symbolizes the Euclidean norm and $\bullet$ the scalar product.

When a Euclidean distance matrix $D = [d_{ij}]$ satisfies $d_{ij} \leq \max(d_{ik}, d_{kj}) \forall i, j, k$, $D$ is also ultrametric. Ultrametric trees always produce Euclidean distance matrices (Pavoine, Ollier, Pontier 2005) but for non-ultrametric trees, nothing guarantees that the dissimilarities $d_{ij}$ observed between species generate a Euclidean distance matrix $D$.

**Property of a Euclidean distance matrix.**—The distance matrix $D$ will be Euclidean if and only if the matrix $H$ with general term $h_{ij} = M_i \bullet M_j = -\frac{1}{2} [d_{ij}]^{\bullet \bullet}$ is positive semidefinite (Gower 1982). This is equivalent to saying that $H$ only has positive eigenvalues. Note that $[ \bullet \bullet ]$ represents the double-centred matrix.

**Demonstration of the Euclidean Nature of the Square Root of a Phylogenetic Distance**

Distances between leaves in a rooted phylogenetic tree $d^R_{\text{tr}}$ can be given with respect to the distance between each leaf (tip) and the root of the tree. Consider the tree in (Fig. 1). The distance between leaves $a$ and $e$ can be computed by

\[d^R_{ae} = c_{aa} + c_{ee} - 2c_{ae}\]

where $c_{aa}$ and $c_{ee}$ are the distances between $a$ and the root of the tree and $e$ and the root of the tree, respectively, and $c_{ae}$ is the distance between the more recent common ancestor (MRCA) of $a$ and $e$ and the root of the tree (Fig. 1). Note that $c_{ae}$ may also be referred to as $t_{ae}$ in the specialized literature.

**If we generalize this, we can say that for any two leaves $i$ and $j$ in a tree, the distance between them is given by:**

\[d^R_{ij} = c_{ii} + c_{jj} - 2c_{ij},\]  \hspace{1cm} (2)

if we set $c_{ij} = M_i \bullet M_j$, then

\[d^R_{ij} = \|M_i\|^2 + \|M_j\|^2 - 2M_i \bullet M_j.\]  \hspace{1cm} (3)

Consequently, by comparing Equations (1) and (2), we see that $d^R_{ij}$ acts as the square of a distance between two points in a Euclidean space, so that $\sqrt{d^R_{ij}}$ acts like a distance between two points in such space. However, as explained above, $D^* = \sqrt{d^R_{ij}}$ will be a Euclidean distance matrix if and only if the matrix $C$ (that represents the distances between the MRCA of pairs of leaves and the root of the tree) with the general term $c_{ij} = M_i \bullet M_j$ is positive semidefinite.

**Proof that $C$ is Positive Semidefinite**

The demonstration of the positive semidefinite nature of $C$ is straightforward because $C$ is a variance covariance matrix. Let $C = c_{ij}$ be the variance covariance matrix of a vector $Y$ of random variables following a BM model of evolution ($Y$‘s represent the trait values associated to each species). Without loss of generality, we may only consider the case when the BM has mean 0.

Then

\[C = E[YY']\].

Showing that $C$ is positive semidefinite consists in proving that for any vector $v$, $v' C v$ is positive or null. We thus calculate $v' C v$ as follows:

\[v' C v = v' E[YY'] v = E[v'YY' v] = E((v' Y)^2) \geq 0,\]

$v' C v$ is always positive because it is the expectation of $(v' Y)^2$.

The direct consequence of this (see Equation 3) is that a phylogenetic distance matrix $D$ with general term $d^R_{ij}$ acts as the square of a distance between two points in a
Euclidean space (see Equations (1) and (3)), so that the elementwise square root of a distance matrix \( D^* = \sqrt{d_{ij}} \) extracted from a phylogenetic tree is always Euclidean. Note, moreover, that this is also true for trees with multifurcating nodes. It is also true for a tree without root: all that is needed is to define a root arbitrarily. Finally, the demonstration also works for distances calculated by the number of nodes separating the leaves (nodal distances). Indeed, this is equivalent to saying that all the branches in the tree have branch lengths of 1 and simply removing 1 to the values in the \( C \) matrix.

**Examples**

In order to illustrate the importance of proposing a new way to convert phylogenetic distances into Euclidean ones, we focused on three aspects of Euclidean phylogenetic distances. We first look at the proportion of random trees for which distance matrices are Euclidean when they are obtained by the current commonly used methods, as a function of tree size (number of leaves). We then examine, for the current commonly used methods, how far from being Euclidean are the extracted distance matrices as a function of tree size. Finally, we illustrate that the current commonly used methods introduce more distortion compared with the square root transformation.

**Methods**

We performed the simulations with R (v2.10.1). We generated random trees using the function \texttt{rtree} from the “ape” package (Paradis et al. 2004) that generates trees by randomly splitting the edges and distributes the branch lengths according to a uniform distribution (on the interval from 0 to 1 as proposed by the default settings of the \texttt{rtree} function). We converted trees to phylogenetic distance matrices using a uniform distribution and with the \texttt{cophenetic.phylo} (“ape” package). Finally, we converted non-Euclidean distance matrices into Euclidean distance matrices with the \texttt{cailliez} function (“ade4” package, Chessel et al. 2004; Dray et al. 2007) for the Cailliez (1983) method and with the \texttt{quasieuclid} function (“ade4” package) for the method that removes negative eigenvalues. Note that the \texttt{rtree} function from the “ape” package was chosen for generating random trees because of its ease of use, but any other way of generating random trees could have been a good choice as well, as long as the trees generated were not ultrametric.

**Proportion of Non-Euclidean Phylogenetic Distances**

In order to illustrate how random phylogenetic trees are rarely associated with Euclidean distance matrices, we randomly generated 10,000 trees for each number of leaves from 3 to 100, and calculated the proportion of associated phylogenetic distance matrices that were Euclidean. We plotted this proportion as a function of the number of leaves in the trees (Fig. 2).

It appears that for trees with more than 20 leaves the probability of observing Euclidean distance matrices is almost nil, emphasizing the need to develop new and appropriate methods to convert phylogenetic distances into Euclidean ones. 

**Distance from Euclidean**

Most random trees produce non-Euclidean distances. But how far from Euclidean are they and how do they correlate to the size of the trees? In order to answer these questions, we randomly generated 500 trees with 5 to 100 leaves (5, 10, 15, . . . , 100 leaves). For each tree size, we transformed all trees into distance matrices and we computed (i) the constant that had to be added to each matrix to render it Euclidean (Cailliez method) and (ii) the number of negative eigenvalues that were present in the diagonalized matrices. These two values give an indication on how distant from Euclidean the matrices are. We then calculated the mean Cailliez constant and the mean number of negative eigenvalues for each tree size (Fig. 3).

The mean Cailliez constant that had to be added to render the trees Euclidean ranged from 0.02 (SD = 0.052) for trees with 5 leaves to 6.2 (SD = 0.69) for trees with 100 leaves, whereas the mean number of negative eigenvalues ranged from 1.2 (SD = 0.4) for trees with 5 leaves to 9.79 (SD = 1.62) for trees with 100 leaves (Fig. 3). There was a correlation between these values such that high mean numbers of negative eigenvalues were associated with high mean Cailliez constants. These values were also clearly correlated with the size of the trees, where distance matrices extracted from large trees were
FIGURE 3. Mean Cailliez constant (○) and mean number of negative eigenvalues (●) when rendering non-Euclidean phylogenetic distance matrices Euclidean as a function of the number of leaves (from 5 to 100) in randomly generated trees (x-axis). A total of 500 trees were generated for each number of leaves. Vertical bars represent standard deviation.

farther from being Euclidean than distance matrices extracted from small trees (Fig. 3).

**Distortions Imposed by the Different Transformations**

We want to illustrate here the fact that the Cailliez method, and the method consisting in removing the negative eigenvalues to render phylogenetic distances Euclidean, impose more distortion on the data than does simply computing the elementwise square root of the distance matrix.

We generated 10,000 random trees with 50 leaves (leaves were named “A,” “B,” “C,” etc.) and systematically computed the ratio between the distance separating Species A to Species B and the distance separating Species A to Species C. We then rendered the distance matrices Euclidean by (i) using the Cailliez method, (ii) removing the negative eigenvalues, and (iii) calculating the elementwise square root of the distance matrix, and we computed the new ratio $d_{AB}'/d_{AC}'$ and compared it with the initial one $d_{AB}/d_{AC}$.

If the method used for rendering the distances Euclidean does not create too much distortion, we expect to find a good correlation between $d_{AB}/d_{AC}$ and $d_{AB}'/d_{AC}'$. On the contrary, a method that creates more distortion will give a bad correlation.

For the square root method, the expected correlation is known:

$$\frac{d_{AB}'}{d_{AC}'} = \sqrt{\frac{d_{AB}}{d_{AC}}}.$$

For the Cailliez method, however, it depends on the constant added, and for the method consisting in removing the negative eigenvalues it depends on the number of negative eigenvalues present in the initial diagonalized matrix.

Figure 4 shows the correlations obtained for the three methods. While the squaring method we propose shows a clear correlation (the simple square of the original ratio), the two other methods show diffuse points. If we focus on initial ratios of 2 (dashed vertical line in Fig. 4), we see that over the 10,000 random trees, when this ratio is met, the ratio after conversion is always the same using the square root method (and it is $\sqrt{2}$) whereas it can be very variable with the two other methods. In other words, the ratios after conversion into Euclidean distances are not independent of the trees themselves and depend on the overall topology of the rest of the tree, except when using the simple square root. This is particularly important when comparing multiple trees for a set of species as is the case in phylogenomic analyses. Indeed, species that are at the same distance in two trees can end up at very different distances after the conversion of the associated phylogenetic distance matrices into Euclidean ones, either by the Cailliez method or the method that removes the negative eigenvalues. This problem is absent when the square root method proposed here is used.
Discussion

Implication of the Study for Phylogenetic Comparative Methods

We have shown that (i) the C matrix that gives in the diagonal the height of each species to the root of the tree and in the off-diagonal the height of the MRCA of pairs of species is positive semidefinite and that (ii) the elementwise square root of a phylogenetic distance matrix is always Euclidean.

These results have implications for both groups of PCMs, namely those that take into account phylogenies by introducing explicit models of evolution (such as PIC) and those that only consider phylogenies as a statistical constraint and aim at quantifying the amount of phylogenetic inertia present in the data (such as PVR). In the first group, recognizing that the C matrix is also the variance–covariance matrix of a trait following a BM model, our study recognizes the positive semidefinite nature of such a matrix. This is widely accepted in the literature, but as far as we know, this is the first time a formal proof is given, at least in a biological journal. Moreover, recognizing the link between our study and PCMs introducing explicit models of evolution such as PIC also gives another interpretation of the elementwise square root of the phylogenetic distance matrix. Indeed, if C is the variance–covariance matrix of a trait Y following a BM model, then D (the phylogenetic distance matrix) represents the variance of the difference \( Y_i - Y_j \) for a trait Y evolving under the neutral BM model. Considering the elementwise square root of D is thus meaningful as this new matrix can be seen as the standard deviation of the difference \( Y_i - Y_j \) for a trait Y evolving under the neutral BM model. This is also true for similar models using transformed branch lengths (e.g., Ornstein–Uhlenbeck processes; Uhlenbeck and Ornstein 1930).

For the second group of PCMs, the implication of this work is direct. By using the elementwise square root of the phylogenetic distance in order to get a Euclidean distance matrix, (i) one ensures to minimize the distortion of the data, compared with other transformation methods used to render distance matrices Euclidean, and (ii) it is thus possible to keep all the information carried by the phylogenetic distance matrix to perform the Principal Coordinate analysis, as proposed by Desdevises et al. (2003), without having to remove the negative eigenvalues as is done in Diniz-Filho et al. (1998).

Impact of the Study in Other Fields of Research

The use of Euclidean phylogenetic distance matrices is not restricted to phylogenetic comparative methods. Desdevises et al. (2002) developed a test of host–parasite cospeciation, ParaFit, that requires as input two Euclidean distance matrices that are then converted into principal coordinate matrices by PCO prior to the statistical test for cospeciation.

The need for Euclidean phylogenetic distance matrices is also prevalent in a very active field of research today: the measure of biodiversity in conservation biology. In particular, Rao’s quadratic entropy (Rao 1982) is a popular biodiversity index that outperforms many other indexes available for this purpose, combining abundance and distance among entities (e.g., species, alleles), and handling more than one trait (Botta-Dukát 2005). For example, some authors consider information on dissimilarities between species such as their taxonomic or phylogenetic distances to measure the originality of a species (Pavoine, Ollier, Dufour 2005) or to maximize the genetic diversity (Pavoine, Ollier, Pontier 2005). In that context, it is necessary that Rao’s index satisfies the concavity property, an essential feature of measures of biological diversity (Lande 1996). This condition is satisfied if and only if the distance between entities (species) is Euclidean (Lau 1985). Until now, the use of dissimilarities between species has thus been restricted to taxonomy or ultrametric trees which are naturally Euclidean (see, e.g., Pavoine, Ollier, Dufour 2005; Pavoine, Ollier, Pontier 2005). As the Euclidean nature of phylogenetic distance matrices is essential for the use of quadratic entropy to characterize genetic diversity, the square root transformation introduced in this article will be very valuable in this field as well.

Conclusions

We demonstrated that the elementwise square root of a distance matrix between species extracted from a rooted phylogenetic tree by computing either patrictic or nodal distances is always Euclidean, and that this property is true for trees with multifurcations. We illustrated the fact that this canonical way of obtaining Euclidean distances for this kind of data caused less distortion than the transformations commonly used (Cailliez method or suppression of negative eigenvalues) and was homogeneous among multiple phylogenetic distance matrices. We also noted that the variance–covariance matrix of a trait following a BM model of evolution was positive semidefinite. These results should be of great interest for the large community of researchers working with comparative phylogenetic methods but also in other fields of research such as cophylogenetic studies and studies on biodiversity measure.

Acknowledgments

We would like to thank Dr Cécile Ané and two anonymous reviewers for their constructive comments and suggestions on this manuscript.

Funding

D.M.D.V. acknowledges post-doctoral grants from the CNRS, the University Paris-Sud 11 and the Chaire Modélisation Mathématique et Biodiversité, VEOLIA-MNHN-Ecole Polytechnique. G.A. acknowledges post doctoral grants from the CNRS and the University Paris-Sud 11.
REFERENCES


