

How to measure and test phylogenetic signal

Tamara Münkemüller^{1*}, Sébastien Lavergne¹, Bruno Bzeznik^{2,3}, Stéphane Dray³, Thibaut Jombart⁴, Katja Schiffers¹ and Wilfried Thuiller¹

¹Laboratoire d'Ecologie Alpine, UMR CNRS 5553, University Joseph Fourier, BP 53, 38041 Grenoble Cedex 9, France; ²CIMENT project, UMR CNRS 5553, University Joseph Fourier, BP 53, 38041 Grenoble Cedex 9, France; ³Université de Lyon, F-69000, Lyon; Université Lyon 1; CNRS, UMR5558, Laboratoire de Biométrie et Biologie Evolutive, F-69622, Villeurbanne, France; and ⁴Department of Infectious Disease Epidemiology, MRC Centre for Outbreak Analysis and Modelling, Imperial College, St Mary's Campus, Norfolk Place, London W2 1PG, UK

Summary

1. Phylogenetic signal is the tendency of related species to resemble each other more than species drawn at random from the same tree. This pattern is of considerable interest in a range of ecological and evolutionary research areas, and various indices have been proposed for quantifying it. Unfortunately, these indices often lead to contrasting results, and guidelines for choosing the most appropriate index are lacking.

2. Here, we compare the performance of four commonly used indices using simulated data. Data were generated with numerical simulations of trait evolution along phylogenetic trees under a variety of evolutionary models. We investigated the sensitivity of the approaches to the size of phylogenies, the resolution of tree structure and the availability of branch length information, examining both the response of the selected indices and the power of the associated statistical tests.

3. We found that under a Brownian motion (BM) model of trait evolution, Abouheif's C_{mean} and Pagel's λ performed well and substantially better than Moran's I and Blomberg's K . Pagel's λ provided a reliable effect size measure and performed better for discriminating between more complex models of trait evolution, but was computationally more demanding than Abouheif's C_{mean} . Blomberg's K was most suitable to capture the effects of changing evolutionary rates in simulation experiments.

4. Interestingly, sample size influenced not only the uncertainty but also the expected values of most indices, while polytomies and missing branch length information had only negligible impacts.

5. We propose guidelines for choosing among indices, depending on (a) their sensitivity to true underlying patterns of phylogenetic signal, (b) whether a test or a quantitative measure is required and (c) their sensitivities to different topologies of phylogenies.

6. These guidelines aim to better assess phylogenetic signal and distinguish it from random trait distributions. They were developed under the assumption of BM, and additional simulations with more complex trait evolution models show that they are to a certain degree generalizable. They are particularly useful in comparative analyses, when requiring a proxy for niche similarity, and in conservation studies that explore phylogenetic loss associated with extinction risks of specific clades.

Key-words: assembly rules, comparative analysis, evolutionary community ecology, niche similarity, phylogenetic niche conservatism, trait evolution

Introduction

The interplay between ecological and evolutionary processes is increasingly recognized to shape the distribution of species in space and time. In addition, larger and more detailed phylogenies containing signatures of past evolutionary processes that

led to contemporary biodiversity are becoming more rapidly available. As a result, many studies now use these phylogenies to account for the relatedness of species and the resulting dependencies of observations, for example, in the fields of comparative analyses, community ecology and macro-ecology (as reviewed by Blomberg, Garland & Ives 2003; Lavergne *et al.* 2010). One central concept in these studies is the statistical non-independence among species trait values because of their phylogenetic relatedness (Felsenstein 1985; Revell, Harmon &

*Correspondence author. E-mail: tamara.muenkemueller@ujf-grenoble.fr

Collar 2008). This non-independence can be measured by the ‘phylogenetic signal’, hereafter defined as the ‘tendency for related species to resemble each other more than they resemble species drawn at random from the tree’ (Blomberg & Garland 2002, p. 905). Phylogenetic signal has been used to investigate questions in a wide range of research areas: How strongly are certain traits correlated with each other (Felsenstein 1985)? Which processes drive community assembly (Webb *et al.* 2002)? Are niches conserved along phylogenies (Losos 2008) and is vulnerability to climate change clustered in the phylogeny (Thuiller *et al.* 2011)?

Along with the different types of applications, a variety of indices has been proposed to measure and test for phylogenetic signal in a quantitative trait (cf. Table 1 and Table A1 in the Appendix for a selection of common indices and tests). Among these are indices that were originally developed within the context of spatial autocorrelation and have later been adapted to phylogenetic applications (e.g. Moran’s *I*, Moran 1950; Gittleman & Kot 1990; Pavoine *et al.* 2008; Revell, Harmon & Collar 2008). These phylogenetic autocorrelation indices have in common that their calculated values are not originally designed to offer a quantitative interpretation (Li, Calder & Cressie 2007 show this in a spatial context). Other indices explicitly relate to a Brownian motion (BM) model of trait evolution (Martins 1996; Pagel 1999; Blomberg, Garland & Ives 2003) and are designed to allow for the comparison of observed values among different phylogenies (Blomberg, Garland & Ives 2003). In the BM model, trait evolution follows a random walk along the branches of the phylogenetic tree, with the variance in the distribution of trait values being directly proportional to branch length. To test the null hypothesis of no phylogenetic signal, the observed value of the focal index can be compared with values expected under random

trait distribution. Random trait distributions can either be numerically simulated by random permutations of the trait values among the tips of the phylogenetic tree or can be derived analytically, for example, by assuming a chi-square distribution for likelihood ratio tests (cf. Table 1 and Materials and methods section).

Even though all indices have been developed to quantify and test for phylogenetic signal, they are calculated following different approaches. Consequently, all of these indices measure different aspects of phylogenetic signal and have been shown to respond differently to inaccurate phylogenetic information, low sample sizes and the absence of branch length information (Blomberg, Garland & Ives 2003; Cavender-Bares, Keen & Miles 2006). However, in the literature, they are used for the same ecological questions, and guidelines for selecting the most appropriate method are missing. To make the best use of these indices, it is essential to assess how estimates of strength and tests of phylogenetic signal are influenced by different properties of the data (Revell, Harmon & Collar 2008). Ultimately, in each specific situation, an educated decision on which index to use is necessary. Here, we compare four indices which have been commonly used in evolutionary ecology studies: Moran’s *I* (Gittleman & Kot 1990; applied e.g. in Nabout *et al.* 2010), Abouheif’s C_{mean} (Abouheif 1999; applied e.g. in Thuiller *et al.* 2011), Blomberg’s *K* (Blomberg, Garland & Ives 2003; applied e.g. in Krasnov, Poulin & Mouillot 2011) and Pagel’s λ (Pagel 1999; applied e.g. in Thuiller *et al.* 2011). Moran’s *I* (Gittleman & Kot 1990) and Abouheif’s C_{mean} (Abouheif 1999) are autocorrelation indices and are not based on an evolutionary model. The resulting values do not offer any quantitative interpretation when comparing values between different phylogenetic trees because the expected value of the statistic under the assumed model is unknown a priori. However, stronger deviations from zero indicate stronger relationships between trait values and the phylogeny. Blomberg’s *K* (Blomberg, Garland & Ives 2003) and Pagel’s λ (Pagel 1999) assume a BM model of trait evolution. For both indices, a value close to zero indicates phylogenetic independence and a value of one indicates that species’ traits are distributed as expected under BM. In most cases, the upper limit of Pagel’s λ is close to one (see Materials and methods for details), while Blomberg’s *K* can take higher values indicating stronger trait similarity between related species than expected under BM. All four indices have been shown to perform well for the specific aspects and range of phylogenetic signal they were developed for. However, for the typical applications in evolutionary ecology, some indices are more appropriate than others. The aim of our comparison is to provide guidelines for an adequate choice.

To develop these guidelines, we create synthetic data using numerical simulations to control for different strength of expected phylogenetic signal and compare both the response of the selected indices and the power of the associated statistical tests. Furthermore, we investigate the sensitivity of the four indices to the size of phylogenies (small vs. very large species number), the resolution of tree structure (phylogenies with and without polytomies) and the availability of branch length

Table 1. Phylogenetic signal indices and associated tests used in this study

	Approach	Directly model based?	Branch length considered?	Commonly applied test
Abouheif’s C_{mean}	Autocorrelation	No	No	Permutation
Moran’s <i>I</i>	Autocorrelation	No	Yes ³	Permutation
Pagel’s λ	Evolutionary	Yes	Yes	Maximum likelihood
Blomberg’s <i>K</i>	Evolutionary	Yes	Yes	–
Blomberg’s test ¹	Evolutionary	Yes ²	Yes	Permutation test

¹The test statistic for Blomberg’s test is not Blomberg’s *K* but the variance of standardized phylogenetic independent contrasts (see Materials and methods for more details).

²When calculated based on phylogenetic independent contrasts, it assumes Brownian motion; when it is based on generalized least squares, it depends on how the variance–covariance matrix of species dissimilarities is built from branch lengths (commonly this is performed under the assumption of Brownian motion).

³Only true if the definition of the weighting matrix is based on phylogenetic distances.

estimates (phylogenies with branch length information vs. phylogenies with uniform branch lengths). Finally, we account for more complex models of trait evolution such as Ornstein–Uhlenbeck processes and models that slow-down or speed-up the rate of trait evolution over evolutionary time.

Materials and methods

We used simulated data to explore the behaviour of the four different indices under study. We compared the estimated values of phylogenetic signal and the power of the associated tests for different topologies of the trees (number of tips, presence of polytomies and availability of branch length information) and increasingly strong BM. All calculations were performed within R (R Development Core Team 2011). In the following, we first introduce the four indices in more detail and then describe the simulations.

PHYLOGENETIC SIGNAL INDICES

Moran's I

Moran's I was originally introduced as a measure of spatial autocorrelation (Moran 1950). Gittleman & Kot (1990) adopted it for the use in phylogenetic analyses. They refer to it as an autocorrelation coefficient describing the relation of cross-taxonomic trait variation to phylogeny. The estimator is given as:

$$\hat{I} = \frac{n}{S_0} \frac{\sum_{i=1}^N \sum_{j=1}^N v_{ij} (y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=1}^N (y_i - \bar{y})^2},$$

where y_i is the trait value of species i and \bar{y} the average trait value. The heart of this statistic is the weighting matrix $\mathbf{V} = [v_{ij}]$ where v_{ij} describes the phylogenetic proximity between species i and j . The sum of all pairwise weights is S_0 . Moran's I is very flexible, because different types of proximities can be used to describe the phylogenetic information (e.g. Pavoine *et al.* 2008). In this study, proximities were computed as the inverse of the patristic distances, with v_{ii} equal to zero (package *adephylo*, Jombart, Balloux & Dray 2010). Moran's I was then estimated with the function *abuouheif.moran* (package *adephylo*).

Abouheif's C_{mean}

Abouheif's C_{mean} tests for serial independence is based on the sum of the successive squared differences between trait values of neighbouring species (Abouheif 1999). As there exist multiple ways to present the order of branches in a phylogenetic tree, Abouheif suggested C_{mean} as the mean value of a random subset of all possible representations. Pavoine *et al.* (2008) provided an exact analytical value of the test. They demonstrated that it uses Moran's I statistic with a new matrix of phylogenetic proximities, which does not relate to branch length but focuses on topology and has a non-zero diagonal (Pavoine *et al.* 2008). We estimated Abouheif's C_{mean} with the function *abuouheif.moran* and the method *oriAbouheif* for the proximity matrix (package *adephylo*).

Pagel's λ

Pagel's λ was introduced as a scaling parameter for the phylogeny and measures phylogenetic dependence of observed trait data (Pagel

1999; Freckleton, Harvey & Pagel 2002). Under the assumption of a pure Brownian model of evolution, the phylogenetic relationships of species uniquely define the expected covariance matrix of their traits. However, whenever additional factors, unrelated to the phylogenetic history, have an impact on trait evolution, the influence of the phylogeny needs to be down-weighted. The coefficient λ defines this weight and is fitted to observed data such that it scales the Brownian phylogenetic covariances down to the actually observed ones. In other words, λ is the transformation of the phylogeny that ensures the best fit of trait data to a BM model. Pagel's λ can adopt values larger than one (traits of related species are more similar than expected under BM) but in practice the upper limit is restricted because the off-diagonal elements in the variance–covariance matrix cannot be larger than the diagonal elements (Freckleton, Harvey & Pagel 2002). We estimated Pagel's λ with the function *fitContinuous* (package *geiger*), which is based on likelihood optimization.

Blomberg's K

Blomberg's K expresses the strength of phylogenetic signal as the ratio of the mean squared error of the tip data (MSE_0) measured from the phylogenetic corrected mean and the mean squared error based on the variance–covariance matrix derived from the given phylogeny under the assumption of BM (MSE , Blomberg, Garland & Ives 2003). In a case in which the similarity of trait values is well predicted by the phylogeny, MSE will be small and thus MSE_0/MSE large. To make the resulting value comparable to other trees with different sizes and shapes, this ratio is standardized by the analytically derived expectation for the ratio under BM evolution. K is computed as:

$$K = \text{observed} \frac{\text{MSE}_0}{\text{MSE}} / \text{expected} \frac{\text{MSE}_0}{\text{MSE}}.$$

We estimated Blomberg's K with the function *phylosignal* (package *picante*, Kembel *et al.* 2010).

SIMULATIONS

Phylogenetic trees

We simulated ultrametric phylogenetic trees with n species (tips), with n ranging from 20 to 500. To account for phylogenies with and without polytomies, we started by creating a basic tree with $n/2$ tips and afterwards added the missing $n/2$ species to these basic trees. The basic trees were pure-birth, stochastic phylogenies with a branching rate of 0.05 (function *birthdeath.tree* in package *geiger*, Harmon *et al.* 2008). We added the missing $n/2$ species by first randomly drawing a tip from the basic tree for each new species (with replacement, so that potentially several species could be added to one tip), second removing the final branches leading to the selected tips, and third replacing these branches either with terminal polytomies or with pure-birth stochastic phylogenies containing the former tips of the basic tree and the new species. The root-to-tip branch length of the basic tree equalled the one of the final tree. This way, phylogenetic trees with polytomies and pure-birth stochastic phylogenetic trees had the same number of species, comparable root-to-tip branch length and (besides the polytomies) comparable structures.

Trait evolution with variable strength of Brownian motion

We simulated continuous traits with different strengths of BM. The weighting factor w determined the strength of BM and thus the expected strength of phylogenetic signal (Fig. 1, traitgrams show

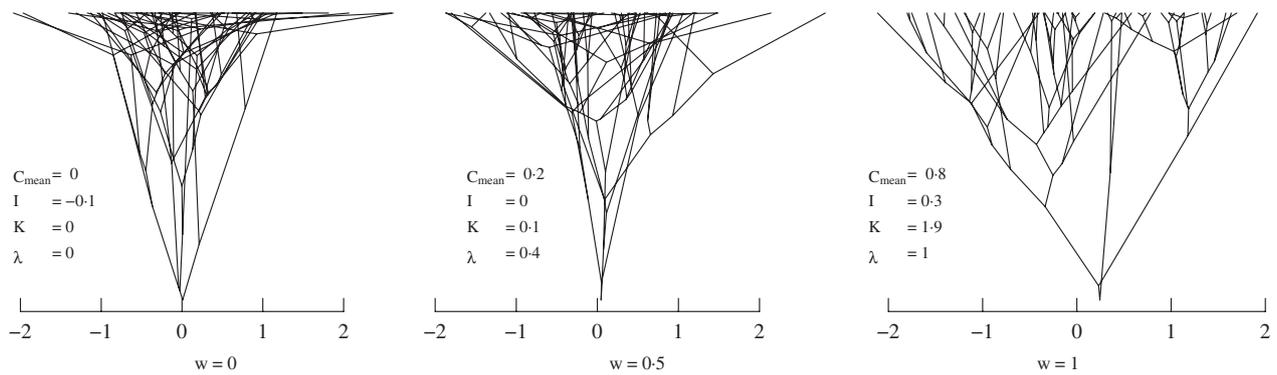


Fig. 1. Traitgram for different weighting factors (w) for the Brownian motion component. Traitgrams arrange species along a continuous trait axis (the x-axis) and connect them with their underlying phylogenetic tree (time on the y-axis). This way, the degree of line crossings in the branches that connect species with their ancestors gives an intuitive picture of phylogenetic signal: The more the lines cross, the more randomly is the trait distributed. As an example, the values of the different selected indices in relation to w are displayed at the bottom of each traitgram.

the effect of increasing w on the dynamic evolution of trait values along the phylogenies). We calculated traits as the weighted sum of two components: $\text{trait} = w \text{trait}_{\text{BM}} + (1 - w) \text{trait}_{\text{rand}}$, with w ranging from 0 to 1. The first component, trait_{BM} , is a vector of trait values created with a BM process of trait evolution with a root value of zero and a standard deviation of 0.1 (function `rTraitCont` in package `ape`, Paradis, Claude & Strimmer 2004). The second component, $\text{trait}_{\text{rand}}$, is the randomly shuffled vector trait_{BM} . We calculated the final trait values by z -standardizing the weighted sums. The idea behind this procedure is to simulate a continuum of trait evolution with pure BM vs. complete randomness as the two extremes. This allowed us to investigate how phylogenetic signal indices and tests respond to continuously increasing strength of BM in the trait evolution.

Note that the use of a trait evolution model with the weighting factor w is very similar to generating trait values using a Pagel's λ model (Appendix A2). This has two important consequences: First, this study does not independently test the performance of Pagel's λ but rather the values that we calculate for Pagel's λ index and test provide a baseline to which the other methods can be compared. Second, the difference in variance scaling between the two models leads to a s-shaped relationship between w and estimates of phylogenetic signal (see Appendix A2 and Fig. A5 for more detail).

More complex models of trait evolution

In addition to this sensitivity analysis, we ran some more complex models of trait evolution. We accounted for Ornstein–Uhlenbeck models and models that slow-down or speed-up the rate of character evolution over evolutionary time. The Ornstein–Uhlenbeck model describes a random walk with a central tendency. The slow-down or speed-up models correspond to evolutionary rates that decrease or increase as a function of evolutionary time since the root node of the tree. These simulations and the theoretical expectations for results are described in more detail in Appendix A1.

Estimation of phylogenetic signal

Not all methods used in this study can handle polytomies. Blomberg's randomization test fails when it is based on independent contrasts. Thus, we randomly resolved the polytomies by arbitrarily transforming all multichotomies into a series of dichotomies with zero length

branches (function `multi2di` in package `ape`). To account for phylogenies with and without branch length information, we either kept the branch length simulated in the tree evolution process or set all branch lengths in the phylogenetic tree to unity. Afterwards, we applied the four chosen indices of phylogenetic signal (Fig. 1).

We divided the values we report for Moran's I (Gittleman & Kot 1990) and Abouheif's C_{mean} (Abouheif 1999) by their maximum possible value to give the observed values a common upper limit among the simulation scenarios. The maximum possible value depends on the underlying proximity matrix, \mathbf{D} (equals \mathbf{V} for Moran's I), and is given as $(n/\mathbf{1}'\mathbf{D}\mathbf{1}) \lambda_{\text{max}}$, where λ_{max} is the first eigenvalue of the symmetric matrix $\mathbf{\Omega} = (\mathbf{I} - \mathbf{1}\mathbf{1}'/n)\mathbf{D}(\mathbf{I} - \mathbf{1}\mathbf{1}'/n)$, \mathbf{I} is the identity matrix and $\mathbf{1}$ is a vector of ones (De Jong, Sprenger & Vanveen 1984; Dray, Legendre & Peres-Neto 2006).

Testing the null hypothesis of random trait variation

We tested the ability of Moran's I , Abouheif's C_{mean} and Blomberg's K to detect deviations from random trait variation using randomization tests. In these tests, we randomly permuted the observed trait values across the tips of the tree and computed the focal indices based on the new, randomized trait pattern. Repeating this procedure, a large number of times yielded a distribution of the focal indices under random trait variation. To compare the values of the tested indices for the observed traits with these random distributions, we extracted their quantiles. Significant deviation from random expectations is indicated by quantiles larger than 0.95 for a significance level of 0.05.

Blomberg, Garland & Ives (2003) suggested to use a different approach to test for phylogenetic signal complementary to the proposed K value for quantitatively characterizing phylogenetic signal. One implementation of this test is via phylogenetically independent contrasts as described in the seminal paper by Felsenstein (1985). The idea is to use the variance of standardized phylogenetic independent contrasts (PICs scaled by branch length) computed for the observed traits on the focal phylogenetic tree. If closely related species tend to have similar trait values (i.e. in the presence of phylogenetic signal), the variance of the standardized contrast will tend to be low. It can be assessed whether it is significantly lower than expected under random trait variation by applying the randomization tests described above. The test can be implemented via generalized least square techniques or via phylogenetically independent contrasts (Blomberg,

Garland & Ives 2003). We used the function `phylosignal` (package `picante`), which is based on independent contrasts for Blomberg's randomization test.

In theory, Pagel's λ could be tested via randomization as well. However, for large phylogenies, the calculation of λ is time-consuming, making this approach impractical for our comparative analysis. We thus followed another approach and compared the likelihood of a model accounting for the observed lambda with the likelihood of a model that assumes phylogenetic independence. To compute these likelihoods in each case, we fit the lambda model using generalized least square models accounting for phylogenetic dependencies by incorporating a correlation structure (function `gls` and `corPagel` in package `ape`). Using likelihood ratio test statistics, we compared models weighting this correlation structure with the observed λ with models assuming a λ equal to zero (i.e. no phylogenetic signal). We then compared these likelihood ratio test statistics to chi-square distributions. An alternative way would have been to compute the likelihoods with the functions `fitContinuous` or `phylosig` (package `phytools`, Revell 2012).

Experimental design

For the comparative analysis of different methods for detecting phylogenetic signal, we developed a full factorial design including phylogenies with polytomies vs. without polytomies, phylogenies with vs. without branch length information, sample sizes of 20, 50, 100, 250 and 500 species and increasing the strength of BM from $w = 0$ to $w = 1$ by steps of 0.1. This gave rise to 220 different scenarios. Each scenario was replicated 100 times. Within each scenario, tests of the null model were based on 1000 randomizations of the data. We report rejection rate of the null hypothesis (random variation of the traits) and observed values of the different indices for all scenarios.

Model-based sensitivity analyses

We used generalized additive models (GAMs) to further analyse the main effects and the two-way interaction effects (interaction of two variables) in the simulation experiments. Response variables were the observed values of phylogenetic signal and for the tests of phylogenetic signal the P -values (i.e. quantiles of observed values in null model distributions), respectively. Explanatory variables were phylogeny size, polytomies, branch length information and strength of BM. The full models used splines for smoothing the effect of the strength of BM and phylogeny size as main and interaction effects. Transformations of the response variable and degrees of freedom for the splines were chosen based on visual analysis of the residuals (for details see footnotes of Tables A2 in the Appendix). The significant influence of the explanatory variables on the response variables was tested by model comparison of the full model and a model missing the focal variable. It is important to note that the P -values of these tests should be treated with caution in simulation studies. Even if true effect sizes are very small, given enough simulations, P -values will eventually become significant. However, given equal numbers of simulations, the differences between P -values can be used to compare different scenarios. In addition to P -values, we reported effect sizes to describe the effect of the explanatory variables on the change in the response variables (Nakagawa & Cuthill 2007). Effect sizes were computed as the coefficients of variation of the response (averaged over repetitions) among the groups defined by the explanatory variables. GAMs were obtained using the function `gam` in R (package `gam`, Hastie 1992).

Results

MEASURING PHYLOGENETIC SIGNAL

None of the four tested indices of phylogenetic signal responded linearly to the here-applied weighting factor for the strength of BM in trait variance (Fig. 2, see also results of GAMs in Table A2a). However, transforming the weighting factor linearizes these relationships (Appendix A2, Fig. A5 and not shown results). As expected, for Pagel's λ and Blomberg's K , the mean value under pure BM was one.

Variation among different simulation runs for one scenario was rather small for the values of Pagel's λ , Abouheif's C_{mean} and Moran's I when phylogenies were sufficiently large (larger than 50 species, Fig. 2, see also Table A2b). The main difference among them was that Pagel's λ showed the highest uncertainty for intermediate strength of BM (especially for small phylogenies), while Abouheif's C_{mean} and Moran's I were most uncertain for strong BM. Blomberg's K showed many outliers with values below but also well above one for strong BM.

The number of species in the phylogeny had a strong influence on the estimated phylogenetic signal (Fig. 2, see also Table A2a). We found that Pagel's λ was the only index for which the mean value did not respond to an increasing number of species. While Abouheif's C_{mean} tended to increase, the response of Moran's I was hump shaped and Blomberg's K tended to decrease (Fig. 2, Table A2a). Uncertainty in Pagel's λ was the least affected by species number, while Abouheif's C_{mean} and Blomberg's K slightly improved and Moran's I strongly improved for larger phylogenies (Table A2b). The existence of polytomies did neither influence mean observed values of the indices nor their uncertainty (Fig. 3, see also Table A2a,b). Similarly, the effect of branch length information was small (Fig. 3). Abouheif's C_{mean} remained unaffected as it ignores branch length information. Pagel's λ and Moran's I increased slightly, while Blomberg's K increased more strongly and got more uncertain when branch length information was missing (Fig. 3, see also Table A2a,b).

TESTING FOR PHYLOGENETIC SIGNAL

Pagel's λ had the smallest type I error for all sizes of phylogenies when testing observed phylogenetic signal against random expectations. Indeed, only in 1% of the simulations was a random trait misidentified as showing significant phylogenetic signal (Fig. 4, see printed values in the plots). With increasing departure from random trait evolution, Pagel's λ , Abouheif's C_{mean} and Moran's I quickly gained power, that is, correctly identified phylogenetic signal. Even for moderate BM ($w > 0.5$), phylogenetic signal was identified in almost all simulations, indicating low type II errors (Fig. 4, see also results of GAMs in Table A2c). In contrast, Blomberg's test showed a much higher type II error for intermediate strength of BM. Only when traits exhibited strong phylogenetic signal ($w > 0.8$) did all simulations significantly reject the null hypothesis of absence of phylogenetic signal. For very small

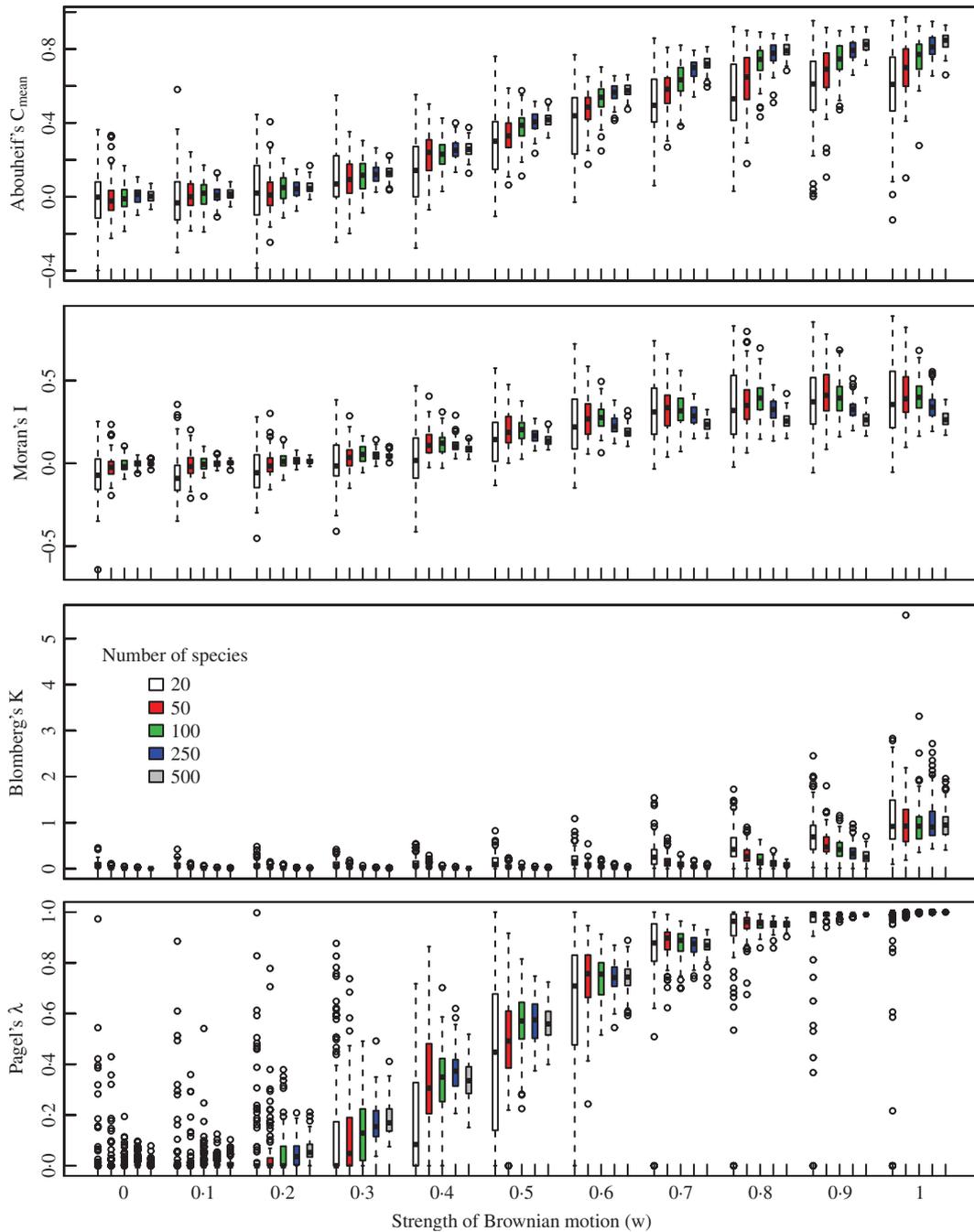


Fig. 2. Response of phylogenetic signal indices to increasing strength of Brownian motion for different sample sizes (shown are scenarios with branch length information and no polytomies).

phylogenies (20 species), all approaches had high type II errors.

We further compared the variation among the different simulation runs for one scenario by plotting the P -values of the observed values given the null distributions (Fig. A1). As expected, the uncertainty in P -value estimates was highest for more random trait distributions in the phylogeny (low w , see also Table A2d). Tests associated with Pagel's λ , Abouheif's C_{mean} and Moran's I showed reduced uncertainty already for $w = 0.3$, whereas Blomberg's test only became more certain

for $w > 0.6$. For moderate BM in the range of $w = 0.3\text{--}0.5$, power for Pagel's λ , Abouheif's C_{mean} and Moran's I strongly depended on the sizes of phylogenies and showed increasing power for increasing species number. Blomberg's test performance did respond much less to increasing the size of the phylogeny (Fig. 4, Table A2c).

None of the tests was affected strongly by the presence of polytomies (Fig. A2). Similarly, the effect of incorporating branch length information was negligible (Fig. A2 and Table A2c). Blomberg's test responded slightly positively to missing

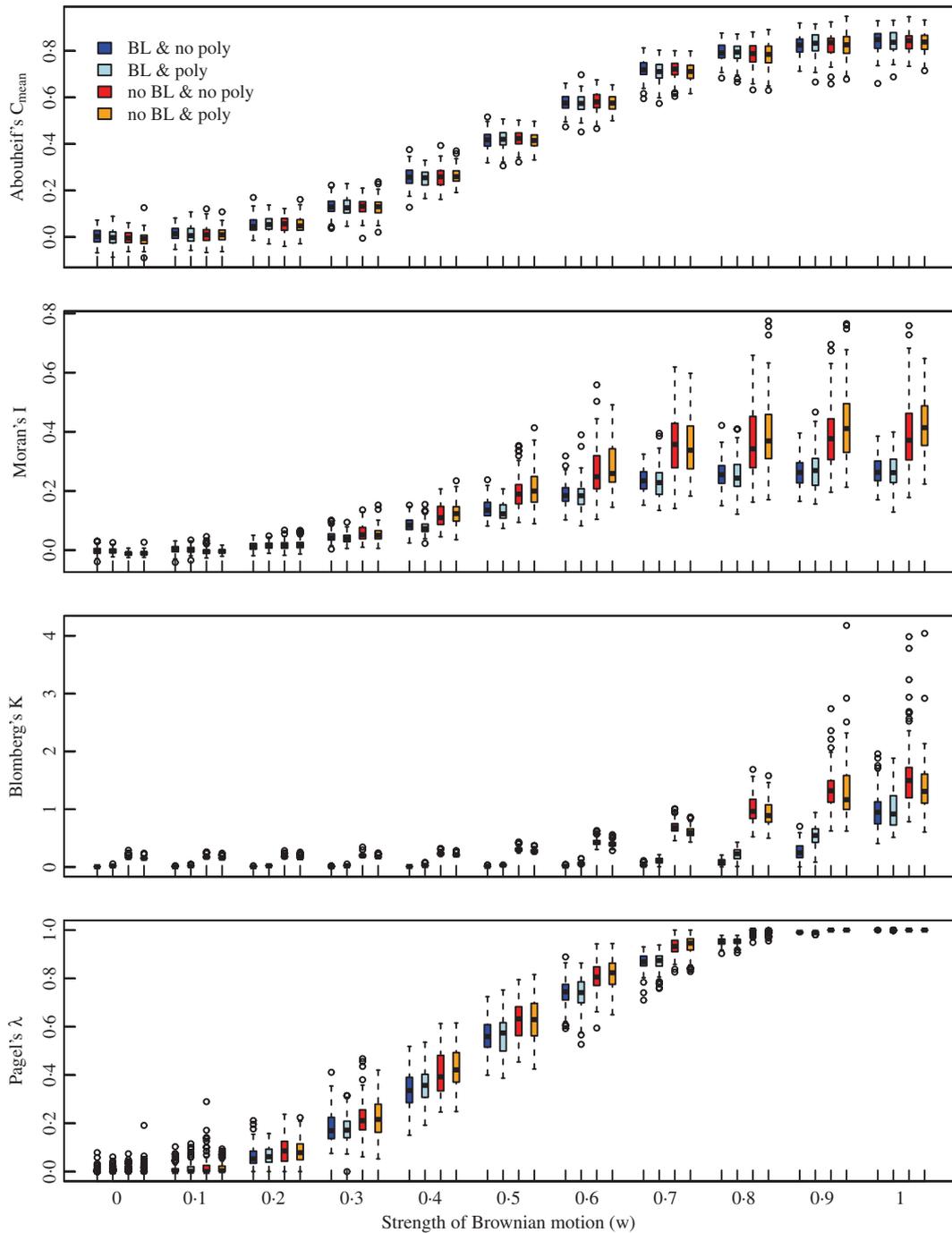


Fig. 3 Response of phylogenetic signal indices to polytomies and branch length information (shown are scenarios for 500 species)

branch length, that is, phylogenetic signal was detected for lower w and with higher accuracy (Fig. A2).

Pairwise correlations between the P -values of the different methods ranged from moderate to high but exhibited relatively strong variability (Fig. A3). Abouheif's C_{mean} P -values correlated most strongly with those of Moran's I , followed by correlations of Pagel's λ with Abouheif's C_{mean} and Pagel's λ with Moran's I . Again, correlations with P -values of Blomberg's test were the weakest. Here, we plotted both the PIC variance test suggested by Blomberg and the randomization procedure

based on Blomberg's K . Our results show that both approaches give almost equal results even when comparing phylogenies with very different numbers of species (Fig. A3).

PHYLOGENETIC INDEX AND TEST PERFORMANCE IN AN OVERVIEW

We used GAMs to further explore the main effects and the two-way interaction effects (interaction of two variables) of increasing strength of BM, the number of species, polytomies

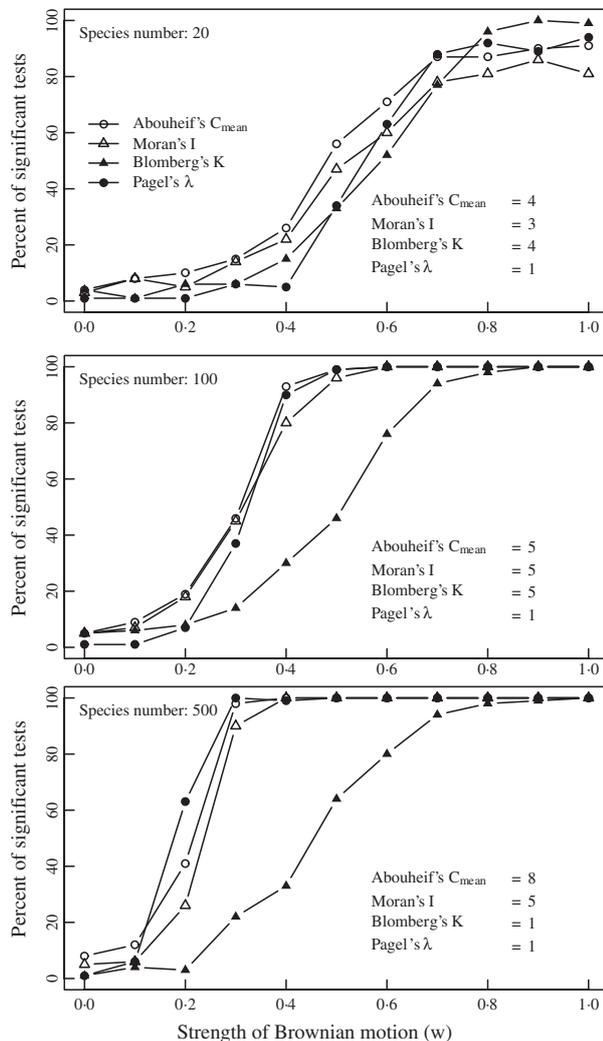


Fig. 4. Response of phylogenetic signal tests to increasing strength of Brownian motion for different sample sizes (shown are scenarios with branch length information and no polytomies). Figures refer to the rejection rate for the null hypothesis that there is no phylogenetic signal. This describes the type I error for w equal to zero (for each index, we indicate the number of significant tests obtained among the 100 repetitions, e.g. for 20 species Pagel's λ identified once a significant deviation from a random pattern even so the pattern originated from a random shuffling of trait values) and tests power for Brownian motion with $w > 0$.

and branch length. Comparisons of the influence of these variables on phylogenetic indices and tests are presented in the Appendix (see Materials and methods and Table A2). Table 2 gives a qualitative summary of the most important findings for measuring and testing phylogenetic signal and provides a quickly accessible validation of the different methods. Overall, Pagel's λ and Abouheif's C_{mean} fulfilled most of the criteria for good index and test performance (indicated in Table 2 by 'yes' and '0' for good and moderate performance). Both methods differ in certain aspects of index performance (especially the response to increasing strength of BM and dependence on the size of the phylogeny) but much less in test performance. Moran's I performs less well than Pagel's λ and Abouheif's

C_{mean} as an index and a test. Blomberg's K performed less well for these data simulated under the assumption of BM, especially when BM was weak.

MORE COMPLEX MODELS OF TRAIT EVOLUTION

All indices and associated tests showed either the theoretically expected functional relationships with the parameters of the more complex trait evolution models or no response (see Fig. 5, Appendix A1 for more details on theoretical expectations and Fig. A4). Under the Ornstein–Uhlenbeck model, changes in patterns of phylogenetic signal were strongest (Fig. 5). This was the only model under which the null model expectation could not always be rejected (Appendix, Fig. A4). Pagel's λ showed higher uncertainty than the other methods especially for intermediate values of phylogenetic signal. Under the κ -model, trends were strongest for Blomberg's K and Pagel's λ . However, the range of change was much smaller for Pagel's λ than for Blomberg's K . Under the δ -model, only Blomberg's K showed a clear trend. In sum, comparing the indices of phylogenetic signal with each other, we observed the strongest effect sizes for Blomberg's K , followed by Abouheif's C_{mean} and Pagel's λ . Moran's I showed only small changes and great overlap of estimates between different scenarios (Fig. 5). The test of Pagel's λ responded most sensitive (Appendix, Fig. A4).

Discussion

The phenotypic trait values of extant species are shaped by their evolutionary history (Harvey & Pagel 1991). Thus, even if this dependence may be blurred by the progression of time, phylogenetic dependence of trait distribution should be considered ubiquitous in the living world (Blomberg, Garland & Ives 2003). This awareness has – beginning 25 years ago – revolutionized the scientific area of comparative analysis. In his seminal paper, Felsenstein (1985) pointed out that because of their phylogenetic relationships, species cannot be regarded as independent data points in statistical approaches of comparative biology. Hereby, he gave impulse to a multitude of conceptual and applied studies about how to infer the statistical non-independence among species trait values because of their phylogenetic relatedness by quantifying the pattern of phylogenetic signal in these traits.

Since then, it has become clear that the level of phylogenetic dependence can vary strongly among investigated phylogenies and even clades, often being significantly reduced when contrasted against the expectations from the standard model of BM (Revell, Harmon & Collar 2008). The divergence from BM expectations can result from a number of different reasons relating to the underlying evolutionary processes, such as fluctuations in the rate of evolution over time (Pagel 1999), directional or stabilizing selection (Revell, Harmon & Collar 2008; Ackerly 2009) or measurement error (Freckleton, Harvey & Pagel 2002; Ives, Midford & Garland 2007; Felsenstein 2008). However, variability in estimated phylogenetic signal could also stem from the statistical tools, that is,

Table 2. Main characteristics of the response of phylogenetic signal indices (a) and tests (b) to increasing strength of Brownian motion (BM), increasing species number, polytomies and branch length information. This table summarizes findings from visual analyses and statistical models (GAMs are described in more detail in the Appendix); 'yes' indicates that the characteristic is fulfilled, 'no' a lack of the specified characteristic ($P < 0.001$ and effect size > 0.2 , cf. Appendix) and '0' an intermediate response ($P < 0.05$ and effect sizes > 0.1 , cf. Appendix)

	Abouheif's C_{mean}	Moran's I	Pagel's λ	Blomberg's K
(a) Measuring phylogenetic signal				
Discrimination of increasing BM (visual validation)	Yes	0/yes	Yes	0
Constant uncertainty under increasing BM	No ¹	No ¹	No ²	No ¹
Constant under increasing N	0 ¹	No ²	Yes	0 ³
Constant uncertainty under increasing N	0 ³	No ³	Yes	0 ³
Constant under polytomies	Yes	Yes	Yes	Yes
Constant uncertainty under polytomies	Yes	Yes	Yes	Yes
Constant under missing branch length	NA	Yes	Yes	No ¹
Constant uncertainty under missing branch length	NA	0 ¹	Yes	No ¹
(b) Testing for phylogenetic signal				
Discrimination of increasing BM (visual validation)	Yes ⁴	Yes ⁴	Yes ⁴	Yes ⁵
Constant uncertainty under increasing BM	No ³	No ³	No ²	No ³
Constant under increasing N	0 ³	No ³	0 ³	0 ³
Constant uncertainty under increasing N	0 ³	No ³	0 ³	0 ³
Constant under polytomies	Yes	Yes	Yes	Yes
Constant uncertainty under polytomies	Yes	Yes	Yes	Yes
Constant under missing branch length	NA	Yes	Yes	0 ¹
Constant uncertainty under missing branch length	NA	Yes	Yes	0 ¹

¹Upwards trend; ²Hump-shaped; ³Downwards trend; ⁴Early; ⁵Late.

the indices used for measuring phylogenetic signal and the power of the associated tests. This is because the different approaches capture different aspects of phylogenetic signal and their values thus can differ greatly, impeding comparison and straightforward interpretation. Here, we compared four of the most widely used approaches to provide guidelines on the choice of an index and an associated test, and on a critical interpretation of the results.

PERFORMANCE OF INDICES AND TESTS UNDER DIFFERENT CONDITIONS

Our results show that Abouheif's C_{mean} and Pagel's λ performed well, both as measures of phylogenetic signal and when tested against expectations of random trait distribution. However, as our trait evolution model is very similar to a λ -model of trait evolution (see Appendix A2) Pagel's λ is predestined to perform well. Thus, our sensitivity analysis does not provide an independent test for Pagel's λ but rather a baseline for comparison with other metrics. While the significance levels of Abouheif's C_{mean} , Moran's I and Pagel's λ were highly correlated, our simulation scenarios showed that Pagel's λ and Blomberg's K can lead to divergent conclusion even for time-independent simulations and a phylogenetic signal not larger than one. This seems surprising as earlier findings from similar simulation models found concordant results for the two methods (Revell, Harmon & Collar 2008).

As predicted, the results on Moran's I and Abouheif's C_{mean} attested that it is not reliable to quantitatively compare their estimates of phylogenetic signal among different phylogenies. In our simulations, the mean value of Moran's I showed a hump-shaped relation with increasing sample size even when

the underlying strength of BM was identical. In contrast, the mean value of Abouheif's C_{mean} increased with increasing sample size. The latter is because the considered distances depend only on topology and are calculated in a way that disproportionately increases long distances (species pairs that split early in the phylogenetic tree) in comparison to short distances (species pairs that split late in the phylogenetic tree) when the number of species is increasing. Because Abouheif's C_{mean} is merely a Moran's I test with a particular phylogenetic distance metric, the choice of the phylogenetic metric used to compute these indices (matrix \mathbf{V}) is likely to impact the results. These results also demonstrate that both indices depend on the structure and size of the phylogeny and that their values cannot be quantitatively compared.

As expected, all four tested methods showed less uncertainty with increasing size of phylogenies. This effect was much stronger for the estimates of indices than for the tests against phylogenetic independence. This result is consistent with earlier studies showing that, for example, Moran's I performs poorly when applied to small sample sizes (Diniz-Filho, De Sant-Ana & Bini 1998). However, type I errors depend not only on the sample size but also on the strength of phylogenetic dependence. While Moran's I is very variable at low sample sizes for highly phylogenetically structured traits, Pagel's λ is most unreliable at small sample sizes and moderate phylogenetic dependence (note, however, that Pagel's λ cannot greatly exceed one by definition, see Materials and methods for details).

All methods (except Abouheif's C_{mean} , which does not include branch length information) showed slightly higher phylogenetic signal when branch length information was missing but no consistent trends with respect to the accuracy of the

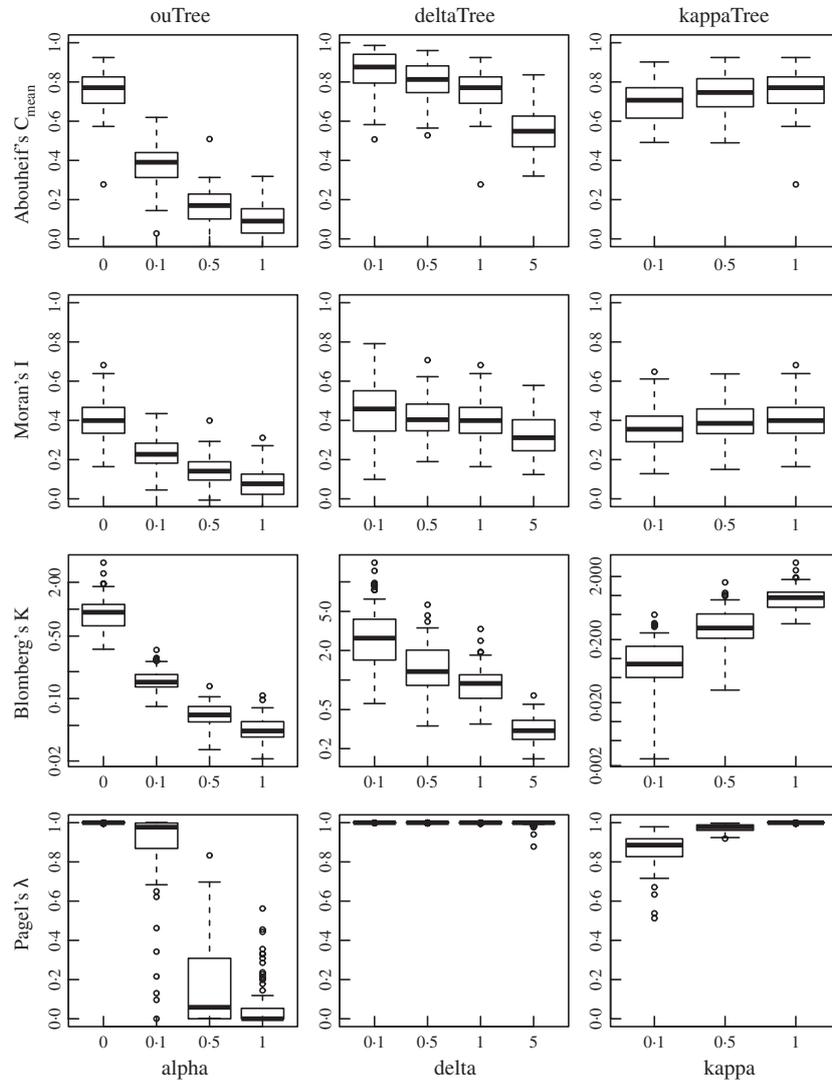


Fig. 5. Response of phylogenetic signal indices to increasing values of the parameters for different tree transformations (shown are scenarios with 100 species, with branch length information and no polytomies). OuTree corresponds to evolution under an Ornstein–Uhlenbeck model, that is, a random walk model with a central tendency with strength α ($\alpha = 0$ is Brownian motion, BM); deltaTree simulates a slow-down or speed-up in the rate of character evolution through time ($\delta = 1$ is BM, $\delta > 1$ is speed-up, $\delta < 1$ is slow-down); kappaTree simulates ‘speciational’ models ($\kappa = 1$ is BM, $\kappa = 0$ is a speciational model).

results were identified. Surprisingly, Blomberg’s test additionally revealed a slightly increased type II error when branch length information was available. Given other sources of uncertainty, the effects of the availability of branch length information were negligible however. This finding is congruent with earlier studies concluding that autocorrelation methods are reasonably robust to missing branch length information (Martins 1996). It has to be noted that our simulations are not only based on traits that evolved (partly) under BM but also on tree structures that evolved under a uniform, time-homogeneous birth process. Consequently, branch lengths are exponentially distributed and less biased than typically observed in nature. The question of how important branch length information is for different types of trees (e.g. with strongly skewed branch length distributions) remains open. For example, we would expect branch length to be more informative if different

molecular clocks or different selective regimes exist in different parts of the tree. Polytomies had very small effects on Blomberg’s methods and on Moran’s I . Abouheif’s C_{mean} and Pagel’s λ were not affected at all. While this finding is supported by earlier studies (Martins 1996), polytomies that occur deeper in the phylogenetic structure, soft polytomies or polytomies including more species may affect results more strongly (Davies *et al.* in press).

Some complementary simulations with more complex models of trait evolution confirmed most of our results and conclusions (Fig. 5, Appendix A1, Fig. A4). None of the different methods showed unexpected trends. However, some methods did respond very weakly and with high uncertainty to changes in niche conservatism (parameter α in the Ornstein–Uhlenbeck model), slowed-down trait evolution over evolutionary time (parameter δ) and speciation events (parameter κ). Overall,

Abouheif's C_{mean} and Moran's I responded most conservatively, while Pagel's λ and especially Blomberg's K were more sensitive. These differences in sensitivity can lead to different conclusions regarding the question whether a parameter change leads to a change in the pattern of phylogenetic signal. It is difficult to judge which index is most correct, or in other words which is the 'right' sensitivity, because theoretical expectations refer to the direction of change but not to the strength of these changes (see Appendix A1 for theoretical expectations). However, Blomberg's K outperformed the other indices in identifying small differences in niche evolution processes that were not related to the strength of BM.

In practice, our results indicate that Blomberg's K is difficult to interpret when applied to traits that developed under BM. This is because for empirical data, test results depend on only one phylogeny and Blomberg's K shows very high variability. Thus, large sample sizes would be required for testing for phylogenetic signal using Blomberg's K when the underlying trait evolution process follows BM. However, our additional simulations also reveal an advantage of this sensitivity to slight changes in the phylogenetic distribution of traits: On average, Blomberg's K is very well suited to capture theoretically predicted changes in phylogenetic signal. Overall, this highlights how important the implicit assumption of a trait evolution model is for calculating phylogenetic signal.

GUIDELINES FOR MEASURING, TESTING AND INTERPRETING PHYLOGENETIC SIGNAL

The results of our sensitivity analysis suggest that Abouheif's C_{mean} is a well performing method for measuring and testing phylogenetic signal under the set of investigated situations. Similarly, Pagel's λ performed well for the more complex models of trait evolution (the good performance of Pagel's λ in the sensitivity analysis needs to be interpreted with caution as it was predestined by the chosen model of trait evolution). Under the assumption that traits evolved following a BM process, the choice of one over another merely depends on the question under investigation and on the nature of the expected phylogenetic signal. In the past, implementations of Pagel's λ were very slow, and nonparametric randomization tests were therefore unfeasible for high numbers of species. However, this problem has become less severe as now an optimized implementation in R language is available (cf. Appendix Table A1).

Many studies not only aim at attesting the presence of phylogenetic signal (i.e. a significant test result) but also at estimating the strength of phylogenetic signal (i.e. the effect size, Nakagawa & Cuthill 2007). However, as discussed above, the use of Abouheif's C_{mean} is restricted to comparisons among different traits in the same phylogeny and therefore not suited as an effect size measure. In contrast, Pagel's λ and Blomberg's K may also be used to compare values across different phylogenies even though phylogenetic signal indices always depend on phylogenies and specific data characteristics and comparisons may thus be hindered by noise. Of these two, only Blomberg's K can capture a phylogenetic signal much stronger than

expected under BM because the range of Pagel's λ is restricted. One field where the effect size of phylogenetic signal becomes important is in studies of community assembly. Here, a sufficiently high level of phylogenetic signal is a prerequisite to allow drawing macro-ecological conclusions on assembly rules on the basis of phylogenetic diversity by assuming that phylogenetic distance can be used as a proxy for niche similarity (Webb *et al.* 2002; Gilbert & Webb 2007). However, as our simulations show, even small deviations from random patterns can result in significant results. These deviations most probably are too small to allow using phylogenetic distance as a proxy for species' niche similarity. A similar argument holds for comparative analyses. In these analyses, too strong patterns of phylogenetic signal need to be removed from the data to assure that data points are statistically independent from each other. Phylogenetic independent contrasts are commonly used to achieve such correction (Felsenstein 1985). But to assess the strength of phylogenetic signal and resulting dependence of data points, one needs an estimate of effect size. Finally, when using simulations to compare the effect of different models of trait evolution on phylogenetic signal, effect size measures seem to be more reliable than the number of significant test results. This is because increasing the sample size increases the number of significant test results. Increasing sample size, that is, the number of repetitions, is fairly easy in simulation models and without much costs. Thus, the number of significant test results is not very meaningful.

Beyond these considerations, it has been argued that approaches with an explicit assumption of an evolutionary model offer the advantage of having a straightforward evolutionary interpretation, while autocorrelation approaches show better robustness to inaccurate phylogenetic information and impose less restrictive assumptions (Gittleman & Kot 1990; Martins 1996). However, phylogenetic signal can be the result of a multitude of evolutionary or non-evolutionary processes (Revell, Harmon & Collar 2008; Ackerly 2009). It is therefore challenging to use estimates of phylogenetic signal for making inferences about the underlying processes, which shaped observed patterns (see discussion on phylogenetic niche conservatism, Appendix A1, Revell, Harmon & Collar 2008). This puts the advantage of offering evolutionary interpretation of some phylogenetic signal indices into perspectives. We argue that inferring evolutionary processes from phylogenetic signal is only possible when the measure of the latter is performed under the clear assumption of a specific trait evolution model (Cooper, Jetz & Freckleton 2010).

Because Abouheif's C_{mean} and Pagel's λ show good behaviour in statistical respects and will in general have comparable biological interpretability, the choice of the method should be mainly driven by the necessity of estimating effect size. Additionally, one should consider expectations with regard to the strength of phylogenetic signal, practical run time considerations and possibly slight differences in performance with respect to specific features of the phylogeny such as the size of the tree and uncertainties about its topology and about branch lengths. When the underlying process of trait evolution does not follow BM, Blomberg's K may be equally well suited.

To account for this uncertainty because of the dependence of results on the underlying model of niche evolution and to better understand the data, we suggest that analyses of phylogenetic signal should be complemented by graphical exploration of the data and further investigation (Ollier, Couteron & Chessel 2006). A particularly interesting question lies in the identification of regions of the tree exhibiting the strongest phylogenetic signal. Indeed, the assumption that patterns of correlation between trait values and phylogenetic relatedness are constant in a phylogeny is often biologically unrealistic (Harvey & Pagel 1991; O'Meara *et al.* 2006). This is especially true in large phylogenies, which are paradoxically the most powerful to detect a significant phylogenetic signal. A promising approach to address this question consists in decomposing observed phylogenetic patterns across multiple phylogenetic scales, using methods such as phylogenetic autocorrelograms (Gittleman & Kot 1990), orthograms (Ollier, Couteron & Chessel 2006) or phylogenetic eigenvectors (Covain *et al.* 2008), branch lengths transformations (Pagel 1999) or decompositions of trait diversity across nodes (Pavoine, Bagueette & Bonsall 2010; see Table A1b). However, the adequateness of phylogenetic eigenvector regression for accounting for phylogenetic non-independence among taxa has recently been questioned (Freckleton, Cooper & Jetz 2011; Adams & Church 2011).

FROM SIMULATED DATA TO REALITY

Experimental studies with simulated data are useful for investigating the sensitivity of phylogenetic methods to violations of their assumptions (Rohlf 2001). However, they can only represent a limited selection of all possible implementations. In our study, all comparisons for the sensitivity analysis are based on data simulated with a stronger or weaker influence of a BM process. In the Appendix, we analysed some additional models with more complex evolutionary processes resulting from Ornstein–Uhlenbeck processes (Felsenstein 1988; Butler & King 2004), speed-up and slowed-down evolutionary rate models and speciation models (Pagel 1999). However, we did not explore the full parameter ranges and the full range of models suggested in the literature to simulate trait evolution. We argue that our experimental setting is justified because validating indices and tests demands clear quantitative expectations for the phylogenetic signal. However, simulations under more complex evolutionary models can show very complex patterns even without comparing different indices (Revell, Harmon & Collar 2008), and in these cases, no clear quantitative expectations for the strength of phylogenetic signal would exist. Moreover, the 22 000 simulations underlying our sensitivity analysis already represent the current limits of computing power offered by a high-performance computer grid. Finally, even our relatively simple simulations enabled us to observe considerable discrepancies between standard indices of phylogenetic signal and allowed us to provide valuable guidelines for their application.

Our extensive sensitivity analysis relied on the assumption of a Brownian model of evolution. Interestingly, additional

simulations involving more complex models of evolution confirmed most of the obtained results. While this suggests that our conclusions are fairly general, some care should still be taken when analysing phylogenetic comparative data. The question whether results from simulation studies generalize to specific field applications where the underlying trait evolution model is unknown cannot be resolved finally.

One aspect worth considering in future studies based on simulated data is the shape of the phylogenetic trees. In our experiments, we simulated tree shapes under a Yule process, which lead to trees being more balanced than trees typically observed in nature. It would be interesting to explore whether results change when contrasted against more arbitrary trees (e.g. fully balanced or comb-shaped trees) and real trees. Less balanced phylogenetic trees may be especially problematic for approaches that ignore branch length (e.g. Abouheif's C_{mean}), because trait values expected under BM differ more strongly when branch lengths are more extreme. Furthermore, we would expect stronger outlier effects in imbalanced trees for all branch length-based approaches. This is because of the strong leverage effect trait values with greater average distance to the other species in the tree have.

Another important aspect is process and measurement uncertainty. All indices and tests of phylogenetic signal require good estimates of the phylogeny and trait values for the organisms under scrutiny (Rohlf 2001). Errors in the tree topology and in the estimation of species mean trait values are likely to bias the calculated phylogenetic signal (Freckleton, Harvey & Pagel 2002; Blomberg, Garland & Ives 2003; Ives, Midford & Garland 2007; Felsenstein 2008). Phylogenetic reconstructions are usually based on limited genetic information and are therefore uncertain. In the past, the most likely phylogeny was often chosen from a range of possible trees. Nowadays, Bayesian approaches become increasingly used and allow for considering samples of most likely phylogenies to account and evaluate this uncertainty (e.g. Lopez-Vaamonde *et al.* 2006). Similarly, trait measurements are subject to inevitable measurement errors and possible biases of the sampling designs. An additional and important source of uncertainty, when using mean trait values for species, comes from intraspecific trait variability (Ives, Midford & Garland 2007; Felsenstein 2008; Albert *et al.* 2010).

Conclusions

Pagel's λ , an approach based on a BM process of trait evolution, and Abouheif's C_{mean} , an autocorrelation measure, were shown overall to perform best given that the underlying evolutionary model is random or follows BM. Pagel's λ performs better for discriminating random and BM patterns of trait distribution in the phylogeny but is computationally more demanding than Abouheif's C_{mean} . The strongest argument for Pagel's λ is that it provides a reliable effect size measure besides testing for phylogenetic signal. Blomberg's K did perform least well in our sensitivity analysis (trait evolution under more or less BM) especially when considering not the mean trend but sensitivity to noise as a measure of performance. This

indicates that it measures an aspect of phylogenetic signal that differs from the other studied methods. However, Blomberg's K has shown to be a good choice for simulation studies with simulated data where trends are in the focus of interest. Scenarios can be repeated and thus sensitivity of Blomberg's K to small changes in phylogenetic trait distribution is more a virtue than a problem because it allows to detect subtle changes in phylogenetic signal where other methods would tend to fail. We challenge the view that Pagel's λ may have a more straightforward evolutionary interpretation than Abouheif's C_{mean} , because in practice our ability to infer processes from patterns of phylogenetic signal is very limited and critically depends on the assumed underlying evolutionary model. Therefore, measuring phylogenetic signal is most valuable for studies aiming to identify a pattern, that is, for comparative analyses and for studies requiring a proxy for species' niche similarity.

Acknowledgements

We thank an anonymous reviewer who very kindly brought key discussion points that greatly improved our paper and provided the basis for the R-code presented now in the Appendix section A2. This work was funded by the French 'Agence Nationale de la Recherche' with the projects DIVERSITALP (ANR-07-BDIV-014) and EVORANGE (ANR-09-PEXT-011), and by the European Commission's FP6 ECOCHANGE project (GOCE-CT-2007-036866). The computations presented in this paper were performed using the CIMENT infrastructure (<https://ciment.ujf-grenoble.fr>), which is supported by the Rhône-Alpes region (GRANT CPER07_13 CIRA: <http://www.ci-ra.org>). KS was supported by the European Union (IEF Marie-Curie Fellowship 252811).

References

- Abouheif, E. (1999) A method for testing the assumption of phylogenetic independence in comparative data. *Evolutionary Ecology Research*, **1**, 895–909.
- Ackerly, D. (2009) Conservatism and diversification of plant functional traits: evolutionary rates versus phylogenetic signal. *Proceedings of the National Academy of Sciences*, **106**, 19699–19706.
- Adams, C.A. & Church, O.C. (2011) The evolution of large-scale body size clines in Plethodon salamanders: evidence of heat-balance or species-specific artifact? *Ecography*, **34**, 1067–1075.
- Albert, C.H., Thuiller, W., Yoccoz, N.G., Soudant, A., Boucher, F., Saccone, P. & Lavorel, S. (2010) Intraspecific functional variability: extent, structure and sources of variation. *Journal of Ecology*, **98**, 604–613.
- Blomberg, S.P. & Garland, T. (2002) Tempo and mode in evolution: phylogenetic inertia, adaptation and comparative methods. *Journal of Evolutionary Biology*, **15**, 899–910.
- Blomberg, S.P., Garland, T. & Ives, A.R. (2003) Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution*, **57**, 717–745.
- Butler, M.A. & King, A.A. (2004) Phylogenetic comparative analysis: a modeling approach for adaptive evolution. *American Naturalist*, **164**, 683–695.
- Cavender-Bares, J., Keen, A. & Miles, B. (2006) Phylogenetic structure of floridian plant communities depends on taxonomic and spatial scale. *Ecology*, **87**, S109–S122.
- 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.
- Cooper, N., Jetz, W. & Freckleton, R.P. (2010) Phylogenetic comparative approaches for studying niche conservatism. *Journal of Evolutionary Biology*, **23**, 2529–2539.
- Covain, R., Dray, S., Fisch-Muller, S. & Montoya-Burgos, J.I. (2008) Assessing phylogenetic dependence of morphological traits using co-inertia prior to investigate character evolution in Loricariinae catfishes. *Molecular Phylogenetics and Evolution*, **46**, 986–1002.
- Davies, T.J., Boardman Kraft, N.J., Salamin, N. & Wolkovich, E.M. (in press) Incompletely resolved phylogenetic trees inflate estimates of phylogenetic conservatism. *Ecology*, <http://www.esajournals.org/doi/abs/10.1890/11-1360.1>.
- De Jong, P., Sprenger, C. & Vanveen, F. (1984) On extreme values of Moran's I and Geary's c. *Geographical Analysis*, **16**, 17–24.
- Diniz-Filho, J.A.F., De Sant'ana, C.E.R. & Bini, L.M. (1998) An eigenvector method for estimating phylogenetic inertia. *Evolution*, **52**, 1247–1262.
- Dray, S., Legendre, P. & Peres-Neto, P.R. (2006) Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). *Ecological Modelling*, **196**, 483–493.
- Felsenstein, J. (1985) Phylogenies and the comparative method. *American Naturalist*, **125**, 1–15.
- Felsenstein, J. (1988) Phylogeny and quantitative characters. *Annual Review of Ecology and Systematics*, **19**, 445–471.
- Felsenstein, J. (2008) Comparative methods with sampling error and within-species variation: contrasts revisited and revised. *American Naturalist*, **171**, 713–725.
- Freckleton, R.P., Cooper, N. & Jetz, W. (2011) Comparative methods as a statistical fix: the dangers of ignoring an evolutionary model. *American Naturalist*, **178**, E10–E17.
- Freckleton, R.P., Harvey, P.H. & Pagel, M. (2002) Phylogenetic analysis and comparative data: a test and review of evidence. *American Naturalist*, **160**, 712–726.
- Geary, R.C. (1954) The contiguity ratio and statistical mapping. *The Incorporated Statistician*, **5**, 115–145.
- Gilbert, G.S. & Webb, C.O. (2007) Phylogenetic signal in plant pathogen-host range. *Proceedings of the National Academy of Sciences*, **104**, 4979–4983.
- Gittleman, J.L. & Kot, M. (1990) Adaptation – statistics and a null model for estimating phylogenetic effects. *Systematic Zoology*, **39**, 227–241.
- Harmon, L.J., Weir, J.T., Brock, C.D., Glor, R.E. & Challenger, W. (2008) GEIGER: investigating evolutionary radiations. *Bioinformatics*, **24**, 129–131.
- Harvey, P.H. & Pagel, M. (1991) *The Comparative Method in Evolutionary Biology*. Oxford University Press, Oxford.
- Hastie, T.J. (1992) Generalized additive models. *Statistical Models in S* (eds J.M. Chambers & T.J. Hastie), pp. 249–308. Wadsworth & Brooks-Cole, Pacific Grove, California.
- Ives, A.R., Midford, P.E. & Garland, T. (2007) Within-species variation and measurement error in phylogenetic comparative methods. *Systematic Biology*, **56**, 252–270.
- Jombart, T., Balloux, F. & Dray, S. (2010) adephylo: new tools for investigating the phylogenetic signal in biological traits. *Bioinformatics*, **26**, 1907–1909.
- Kembel, S.W., Cowan, P.D., Helmus, M.R., Cornwell, W.K., Morlon, H., Ackerly, D.D., Blomberg, S.P. & Webb, C.O. (2010) Picante: R tools for integrating phylogenies and ecology. *Bioinformatics*, **26**, 1463–1464.
- Krasnov, B.R., Poulin, R. & Mouillot, D. (2011) Scale-dependence of phylogenetic signal in ecological traits of ectoparasites. *Ecography*, **34**, 114–122.
- Lavergne, S., Mouquet, N., Thuiller, W. & Ronce, O. (2010) Biodiversity and climate change: integrating evolutionary and ecological responses of species and communities. *Annual Review of Ecology, Evolution and Systematics*, **41**, 321–350.
- Li, H.F., Calder, C.A. & Cressie, N. (2007) Beyond Moran's I: testing for spatial dependence based on the spatial autoregressive model. *Geographical Analysis*, **39**, 357–375.
- Lopez-Vaamonde, C., Wikstrom, N., Labandeira, C., Godfray, H.C.J., Goodman, S.J. & Cook, J.M. (2006) Fossil-calibrated molecular phylogenies reveal that leaf-mining moths radiated millions of years after their host plants. *Journal of Evolutionary Biology*, **19**, 1314–1326.
- Losos, J.B. (2008) Phylogenetic niche conservatism, phylogenetic signal and the relationship between phylogenetic relatedness and ecological similarity among species. *Ecology Letters*, **11**, 995–1003.
- Lynch, M. (1991) Methods for the analysis of comparative data in evolutionary biology. *Evolution*, **45**, 1065–1080.
- Martins, E.P. (1996) Phylogenies, spatial autoregression, and the comparative method: a computer simulation test. *Evolution*, **50**, 1750–1765.
- Moran, P.A.P. (1950) Notes on continuous stochastic phenomena. *Biometrika*, **37**, 17–23.
- Nabout, J.C., Terribile, L.C., Bini, L.M. & Diniz-Filho, J.A.F. (2010) Phylogenetic autocorrelation and heritability of geographic range size, shape and position of fiddler crabs, genus *Uca* (Crustacea, Decapoda). *Journal of Zoological Systematics and Evolutionary Research*, **48**, 102–108.

- Nakagawa, S. & Cuthill, I.C. (2007) Effect size, confidence interval and statistical significance: a practical guide for biologists. *Biological Reviews*, **82**, 591–605.
- O’Meara, B.C., Ane, C., Sanderson, M.J. & Wainwright, P.C. (2006) Testing for different rates of continuous trait evolution using likelihood. *Evolution*, **60**, 922–933.
- Ollier, S., Couteron, P. & Chessel, D. (2006) Orthonormal transform to decompose the variance of a life-history trait across a phylogenetic tree. *Biometrics*, **62**, 471–477.
- Pagel, M. (1999) Inferring the historical patterns of biological evolution. *Nature*, **401**, 877–884.
- Paradis, E., Claude, J. & Strimmer, K. (2004) APE: analyses of phylogenetics and evolution in R language. *Bioinformatics*, **20**, 289–290.
- Pavoine, S., Baguette, M. & Bonsall, M.B. (2010) Decomposition of trait diversity among the nodes of a phylogenetic tree. *Ecological Monographs*, **80**, 485–507.
- Pavoine, S., Ollier, S., Pontier, D. & Chessel, D. (2008) Testing for phylogenetic signal in phenotypic traits: new matrices of phylogenetic proximities. *Theoretical Population Biology*, **73**, 79–91.
- R Development Core Team (2011) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Revell, L.J. (2012) Phytools: phylogenetic tools for comparative biology (and other things). *Methods in Ecology and Evolution*, doi: 10.1111/j.2041-210X.2011.00169.x
- Revell, L.J., Harmon, L.J. & Collar, D.C. (2008) Phylogenetic signal, evolutionary process, and rate. *Systematic Biology*, **57**, 591–601.
- Rohlf, F.J. (2001) Comparative methods for the analysis of continuous variables: geometric interpretations. *Evolution*, **55**, 2143–2160.
- Thuiller, W., Lavergne, S., Roquet, C., Boulangéat, I. & Araujo, M.B. (2011) Consequences of climate change to the Tree of Life in Europe. *Nature*, **470**(53), 1–534.
- Webb, C.O., Ackerly, D.D., McPeck, M.A. & Donoghue, M.J. (2002) Phylogenies and community ecology. *Annual Review of Ecology and Systematics*, **33**, 475–505.
- Wiens, J.J. & Graham, C.H. (2005) Niche conservatism: integrating evolution, ecology, and conservation biology. *Annual Review of Ecology, Evolution and Systematics*, **36**, 519–539.

Received 17 May 2011; accepted 6 February 2012

Handling Editor: Luke Harmon

Supporting Information

Additional Supporting Information may be found in the online version of this article.

Appendix S1. More complex models of trait evolution.

Appendix S2. Similarity of our simulations to a λ model of trait evolution.

Table S1. Phylogenetic signal indices and tests (a) and related measures (b).

Table S2. Influence of explanatory variables on response variables. (a) Phylogenetic signal indices. (b) Standard deviations of phylogenetic signal indices. (c) Phylogenetic signal tests. (d) Standard deviations of phylogenetic signal tests.

Fig. S1. Response of phylogenetic signal tests to increasing strength of Brownian motion for different sample sizes.

Fig. S2. Response of phylogenetic signal tests to polytomies and branch length information.

Fig. S3. Correlation of phylogenetic signal tests for different N.

Fig. S4. Response of phylogenetic signal tests to increasing values of the parameters for different tree transformations.

Fig. S5. Comparison of phylogenetic signal values against increasing strength of Brownian motion for the model of trait evolution applied in this paper (w model) and a λ model of trait evolution.

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials may be re-organised for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.