

A method for testing the assumption of phylogenetic independence in comparative data

Ehab Abouheif*

*Department of Ecology and Evolution, State University of New York at Stony Brook,
Stony Brook, NY 11794-5245, USA*

ABSTRACT

When making a comparison between two or more traits across taxa, the assumption of phylogenetic independence must be tested. Empirically determining the validity of this assumption before and after applying a phylogenetically based comparative method (PCM) can provide researchers with a rigorous methodological approach for analysing comparative data. This approach can help resolve current debates regarding whether it is always appropriate to apply a PCM, and whether PCMs are in fact successful in accounting for all of the historical non-independence in comparative data. To use this methodological approach, however, a generally applicable statistical diagnostic that can test the assumption of phylogenetic independence in comparative data is required. In this context, I present the application of a statistical diagnostic called the ‘test for serial independence’. This diagnostic can be applied to almost all currently employed PCMs, including Felsenstein’s (1985) independent contrasts. Furthermore, I demonstrate the application of the test for serial independence by analysing three data sets from the literature. The results of these analyses show that this diagnostic can successfully detect different degrees of phylogenetic autocorrelation in small and large phylogenies as well as in different types of phenotypic characters. The challenges and difficulties associated with applying the proposed methodological approach and the test for serial independence are discussed.

Keywords: comparative method, independent contrasts, phylogeny, phylogenetic autocorrelation, test for serial independence.

INTRODUCTION

The degree to which phenotypic characters are shaped by phylogenetic history is a central issue in ecology and evolutionary biology (Gould and Lewontin, 1979; Felsenstein, 1985; McKittrick, 1993; Gould, 1997). Over the last 15 years, a great deal of attention has been dedicated to the fact that species and higher taxa are part of a hierarchically structured phylogeny, and thus may not be biologically and statistically independent from one another when making comparisons across taxa (for reviews see Harvey and Pagel, 1991; Miles and Dunham, 1993; Martins and Hansen, 1996). As a consequence, several statistical

* Address all correspondence to Ehab Abouheif, Department of Zoology, Duke University, Durham, NC 27708-0325, USA. e-mail: abouheif@duke.edu

methods have been created to incorporate phylogenies and account for the historical non-independence among taxa, such as the independent contrasts (Felsenstein, 1985) and phylogenetic autocorrelation (Cheverud *et al.*, 1985) techniques. It has become abundantly clear that, if the assumption of phylogenetic independence within a particular taxonomic group is violated, then the application of traditional statistical analyses to that group will suffer from two major problems: (1) confidence limits on the statistics used will be spuriously narrow, in which case one may claim that a relationship between the variables under study exists when, in fact, it may not (i.e. a type I error); (2) the parameters of interest (e.g. regression or correlation coefficients) may be estimated inaccurately (Felsenstein, 1985; Harvey and Pagel, 1991; Martins and Garland, 1991). Therefore, in cases where the assumption of phylogenetic independence is violated, one should always apply a phylogenetically based comparative method (PCM) to account for these statistical problems.

The key to successfully applying any PCM, however, depends upon the accuracy of both the phylogenetic hypothesis and the proposed model of evolutionary change (Felsenstein, 1988). Previous simulation studies have shown that, if one applies a given PCM in which the evolutionary model, branch lengths, or topology of the phylogeny are inaccurate, then the PCM can perform poorly and may not account for all of the statistical non-independence contained within the data (Martins and Garland, 1991; Björklund, 1994; Gittleman and Luh, 1994; Purvis *et al.*, 1994; Díaz-Uriarte and Garland, 1996, 1998). Furthermore, not all traits are going to be strongly correlated to their phylogenetic history (Hansen and Martins, 1996), and applying a PCM when one or more of its assumptions is being violated can actually create problems of statistical non-independence among taxa when none existed initially (Gittleman and Luh, 1994; Gittleman *et al.*, 1996). These and other problems have caused considerable debate as to whether it is always appropriate to apply PCMs, and whether PCMs can always remove the non-independence in comparative data (Björklund, 1994; Harvey *et al.*, 1995; Westoby *et al.*, 1995; Díaz-Uriarte and Garland, 1996; Ricklefs, 1996; Ricklefs and Starck, 1996; Price, 1997; Abouheif, 1998).

These problems can be largely circumvented, however, if the assumption of phylogenetic independence is treated as an empirical issue (Gittleman *et al.*, 1996). The assumption of phylogenetic independence should be tested *before* one applies a PCM to determine whether the traits of interest are significantly correlated to their phylogenetic history. If the traits of interest are significantly correlated to phylogeny, then a PCM must be used. If, however, the traits of interest are not significantly correlated to phylogeny, then traditional (ahistorical) statistical analyses can safely be applied (Ackerly and Donoghue, 1998).

The assumption of phylogenetic independence should also be tested *after* the comparative data have been transformed by a PCM, such as independent contrasts, to ensure that the PCM has in fact accounted for all of the historical non-independence in the data. If the PCM fails to account for all of the historical non-independence, then the assumptions (i.e. the branch lengths, model of evolutionary change, and phylogeny used) of the PCM should be critically evaluated or an alternative PCM should be applied. Thus, empirically testing the assumption of phylogenetic independence before and after applying a PCM can provide a rigorous and methodological approach for analysing comparative data. The flow chart in Fig. 1 outlines this approach, and is designed to guide researchers to the most efficient and accurate solution to a comparative problem.

This methodological approach (Fig. 1), however, requires a generally applicable statistical diagnostic that can actually test the assumption of phylogenetic independence before or

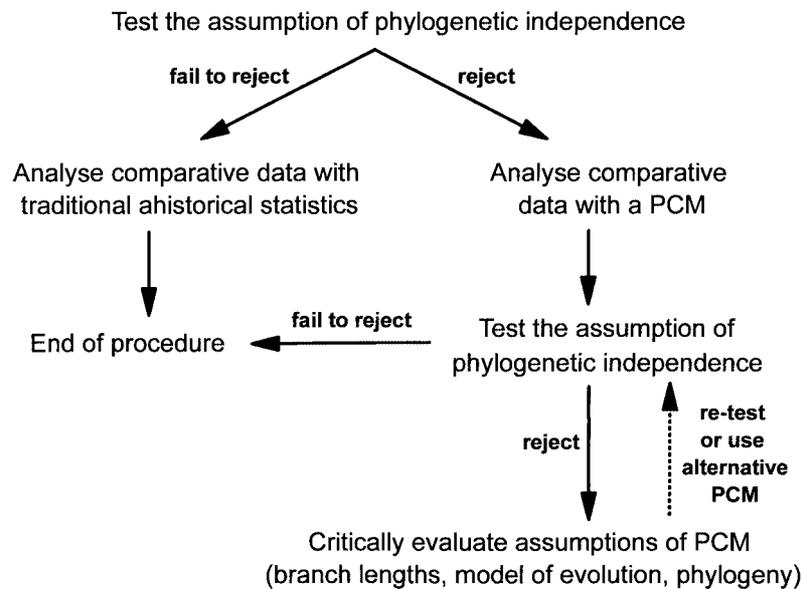


Fig. 1. Flow chart to identify the best course of action when analysing comparative data. Testing the assumption of independence before applying a phylogenetically based comparative method (PCM) can allow the identification of traits which are not phylogenetically autocorrelated, whereas testing the assumption of independence after the application of a PCM allows the identification of cases where the PCM has failed to account for the historical non-independence of taxonomic data points. If the latter occurs, this methodological approach advocates a re-evaluation of the assumptions of the PCM (i.e. model of evolution change, branch lengths, or phylogeny), and the re-test of the assumption of independence until it is satisfied or an alternative PCM is applied.

after the application of any PCM. My objective here is to present the application of a statistical diagnostic, called the ‘test for serial independence’, to comparative data. This diagnostic is generally applicable to almost all PCMs, and can be used to test the assumption of phylogenetic independence at any stage of the methodological approach outlined in Fig. 1. These types of statistical diagnostics are not only valuable as a method for approaching comparative problems, but are also the key to understanding the degree to which phylogenetic history has shaped the evolution of phenotypic characters (Gittleman *et al.*, 1996; Diniz-Filho *et al.*, 1998).

PREVIOUS APPROACHES

The need to treat the assumption of independence in comparative biology as an empirical issue was recognized by Gittleman and Kot (1990), who applied a statistical diagnostic called Moran’s *I* (Moran, 1950) to directly test the assumption of phylogenetic independence. Moran’s *I*, a statistic principally used to measure spatial autocorrelation (Cliff and Ord, 1981), is a diagnostic that has been applied to measure the degree of dependence among species data points at various taxonomic or phylogenetic distances. Moran’s *I*, however, is not compatible with the independent contrasts technique (Felsenstein, 1985),

and performs poorly when applied to small data sets (Martins, 1996a; Diniz-Filho *et al.*, 1998). Furthermore, Moran's *I* requires information on branch lengths or taxonomic distances to accurately measure the degree of phylogenetic autocorrelation within a series of specified distance classes along a phylogeny. As a diagnostic, this has two disadvantages: (1) branch length and taxonomic distance information is often unavailable or inaccurate; and (2) accurately specifying distance classes for a given trait along a phylogeny assumes that one can accurately estimate the number and location of evolutionary changes occurring for that trait along the specified phylogeny.

Other diagnostics, such as the nested ANOVA technique (Bell, 1989), and examining the distribution of independent contrasts (Garland *et al.*, 1992; Martins, 1994), are indirect measures of the degree of dependence among species data points, in the sense that they do not propose to test the statistical significance of whether or not the data points are independent of one another. Thus, they are only of heuristic value and cannot be applied to the methodological approach in Fig. 1. Regardless of the disadvantages, the above diagnostics are valuable as heuristic tools (Martins and Hansen, 1996).

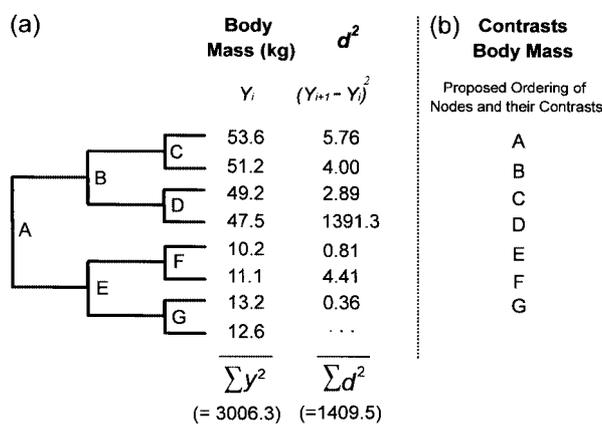
THE TEST FOR SERIAL INDEPENDENCE

The test for serial independence (TFSI), originally proposed by von Neumann *et al.* (1941), was designed to measure the degree of non-randomness in a sequence of continuous variates. It can detect self-similarity or non-random alternation among adjacent observations, and can be used for the purpose of testing the assumption of phylogenetic independence in comparative data (Fig. 1). If the observations are ordered in a sequence along the tips of a given phylogeny, then the test can potentially detect positive or negative phylogenetic autocorrelation. Positive phylogenetic autocorrelation is the self-similarity of adjacent observations caused by phylogenetic descent, whereas negative phylogenetic autocorrelation is the non-random alternation of observations among sister taxa that may be caused by convergent evolution.

The TFSI is a parametric test that assumes normality in its underlying distribution (Zar, 1984). It is not dependent on a particular model of evolutionary change, and does not require information on branch lengths, but does assume that the phylogenetic topology reflecting the evolutionary relationships among the organisms of interest is known. If the TFSI is applied to measure the degree of phylogenetic autocorrelation among species or any higher taxa, it assumes that within-species variation is negligible in comparison to the level of among-species variability.

The TFSI is based upon the sum of the successive squared differences between observations $\Sigma d^2 = \Sigma (Y_{i+1} - Y_i)^2$. If the observations are independent from each other, then the sum of the successive squared differences will be twice the sum of squares $\Sigma y^2 = \Sigma (Y_i - \bar{Y})^2$, such that the expected ratio between these two quantities $\eta = \Sigma d^2 / \Sigma y^2$ will approximate 2.0 (Sokal and Rohlf, 1995). If there is positive phylogenetic autocorrelation (self-similarity), then the variance of the successive differences will be less than if the observations were randomly ordered and, consequently, η will be less than 2.0. Conversely, if the observations show negative phylogenetic autocorrelation (non-random alternation), then the variance of the successive differences will be greater than if the observations were randomly ordered, and η will be greater than 2.0. Young (1941) tabulated critical values for the TFSI in the form of a 'C-statistic', which equals $(1 - \eta/2)$, and can be used for tests of significance. This table can be found in Rohlf and Sokal (1995, Table HH).

A hypothetical example (Fig. 2a) will clarify the application and logic behind the TFSI. Given a symmetrical eight-taxon tree (with its nodes labelled A to G), and body mass data for these taxa, the null and alternate hypotheses for applying the TFSI to these hypothetical data are: H_0 = body mass across these eight taxa are serially independent (i.e. there is no phylogenetic autocorrelation) or H_A = body mass across these eight taxa are serially dependent (i.e. there is phylogenetic autocorrelation). Body mass data are then ordered in a sequence along the tips of the hypothetical eight-taxon tree and the sum of squares ($\sum y^2$) and the sum of the squared successive differences ($\sum d^2$) are calculated. By visually inspecting the ordered sequence of body masses on the phylogenetic tree in Fig. 2a, it is clear that those belonging to clade B have a much higher mean than those belonging to clade E. Thus, one should suspect that there is positive phylogenetic autocorrelation in this data set. Indeed, the ratio between the sum of squared successive differences and sum of squares (i.e. η) is well below 2.0. This deviation from the expected ratio is statistically significant because the calculated value of the C-statistic is greater than the critical value, yielding a $P < 0.01$. There is significant phylogenetic autocorrelation in this hypothetical data set of body mass observations, and thus the data must be transformed by applying a PCM.



$$\eta = \sum d^2 / \sum y^2 = 0.469$$

$$\text{Calculated C} = 1 - \eta/2 = 0.766$$

$$\text{Critical C} = 0.587$$

$$P < 0.01$$

Fig. 2. Hypothetical example of applying the test for serial independence (TFSI) to comparative data. (a) Y_i represents hypothetical body mass observations, and d^2 represents the squared difference between successive observations. By calculating these two quantities, this example demonstrates how the TFSI can detect significant phylogenetic autocorrelation. (b) The TFSI can also be applied to Felsenstein's (1985) independent contrasts by ordering the nodes and their contrasts as shown.

The independent contrasts technique (Felsenstein, 1985) is a PCM that has become a standard procedure for estimating the evolutionary regression or correlation coefficient between phenotypic characters. It uses a phylogeny to transform the original observations into a set of $n - 1$ standardized difference scores called ‘contrasts’. These independent contrasts are theoretically free of the influence of phylogeny, and can be used in conjunction with any statistical procedure (Garland *et al.*, 1992). To test the assumption that a given set of independent contrasts is not phylogenetically autocorrelated, one simply has to order the nodes and their contrasts in the sequence shown in Fig. 2b, and apply the TFSI as already described.

Although the above application of the TFSI is relatively straightforward, there remains one important step. We must correct for the arbitrary representation of the topology of phylogenetic trees. Any branches connected to a particular node in a phylogeny can be rotated freely, which means that there are $2^{(n-1)}$ possible ways in which the order of the branches can be represented. For example, a tree with four taxa has eight possible ways of being represented, whereas a tree with 50 taxa has 5.6×10^{14} ways. This, of course, will affect the way in which observations are ordered into a sequence along a phylogenetic tree. For example, if the order of body mass observations and their contrasts in Fig. 2 are labelled along the original representation of the tree (Fig. 3a), then one can see how randomly rotating nodes within the original representation can potentially change the order of the sequence of observations and their contrasts along the tree (e.g. Figs 3b and 3c). This problem can be resolved, however, by incorporating the TFSI into a randomization procedure where it is conducted on a subset of all possible representations of a given topology.

Applying the TFSI to a subset of all possible representations of a given topology will yield a distribution of C-statistics ($1 - \eta/2$). From this distribution, the mean C-statistic is chosen as the observed value. To test the significance of this observed mean C-statistic, it is compared to a null hypothesis sampling distribution, which is estimated from the same data that are used to calculate the observed mean C-statistic. The null hypothesis sampling distribution is estimated by randomly shuffling the original data so that taxa are randomly placed on the tips of the original phylogenetic topology. For each random shuffle of the original data, the TFSI is once again applied to a subset of all possible representations of the given topology to obtain a distribution of ‘randomized’ C-statistics, from which a randomized mean C-statistic is calculated. This procedure is done repeatedly until a distribution of randomized mean C-statistics is obtained. Comparing the observed mean C-statistic to the estimated null hypothesis sampling distribution of randomized mean C-statistics can then determine whether the observed mean C-statistic is improbable enough to reject the null hypothesis that there is no phylogenetic autocorrelation in the data.

APPLYING THE TFSI TO REAL DATA

I analysed three comparative data sets from the literature to demonstrate the application of the TFSI, as well as to demonstrate some of the conclusions which are possible when the TFSI is used to follow the methodological approach in Fig. 1. To conduct the TFSI, I used a program called ‘Phylogenetic Independence’ by J. Reeve and E. Abouheif. This program can be downloaded from my World Wide Web page, <http://life.bio.sunysb.edu/ee/ehab>, or is available from the author on request. For each of the three data sets, I calculated the observed and randomized mean C-statistics by randomly rotating the nodes within a given

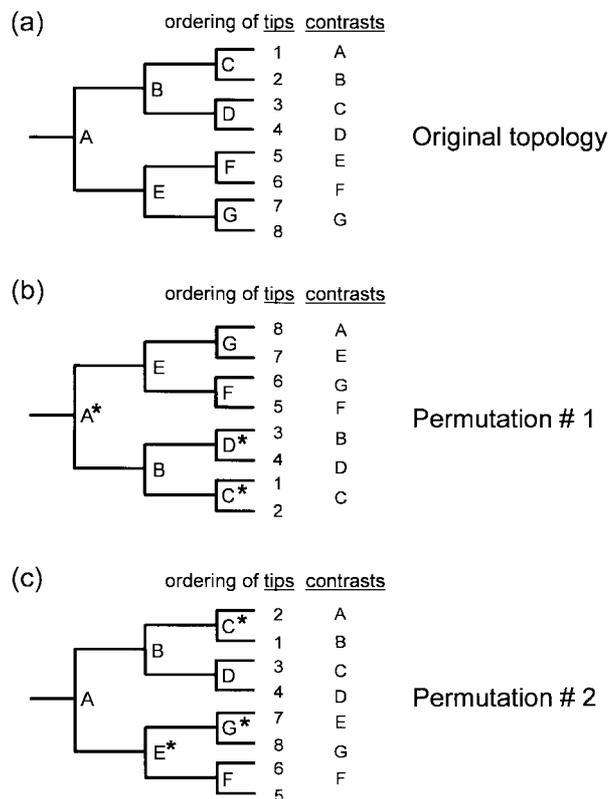


Fig. 3. The arbitrary representation of phylogenetic topologies. Because the topology of any phylogeny can be represented $2^{(n-1)}$ different ways, the way in which observations are ordered in a sequence along the tips of a phylogeny is affected. Asterisks indicate that a particular node has been rotated. The tips and nodes of the phylogenies have been labelled to show how the ordering of individual observations and their contrasts is affected by these rotated nodes.

phylogenetic topology 2999 times (in each of the 2999 iterations, each individual node in the given topology had a probability of 0.5 of being randomly rotated). I randomly shuffled the original data 999 times to estimate the null hypothesis sampling distribution. It is important to note that, because the null hypothesis sampling distribution is now estimated from the original data, the TFSI no longer assumes normality in its underlying distribution. I used one-tailed probabilities to assess the statistical significance of the TFSI because *a priori* there is good reason to predict that the direction of the autocorrelation in the three traits I analysed, if any, will be positive – that is, self-similarity due to phylogenetic descent (Harvey and Pagel, 1991; Rice and Gaines, 1994).

For each of the three data sets, I applied the TFSI to tip data (i.e. the original observations) and to standardized independent contrasts (Felsenstein, 1985). I calculated standardized independent contrasts using the PDTREE program (in the PDAP package by T. Garland Jr., J.A. Jones, A.W. Dickerman, P.E. Midford and R. Díaz-Uriarte). When calculating contrasts between the values at the tips or nodes, the direction of subtraction is entirely arbitrary (Garland *et al.*, 1992), and thus I conducted the TFSI on the absolute

value of the standardized contrasts. If significant phylogenetic autocorrelation among a given set of standardized contrasts was detected using the TFSI, I checked the adequate standardization of the contrasts and, if needed, I applied the appropriate branch length transformations as proposed by Garland *et al.* (1992). In such cases, I then re-tested the assumption of phylogenetic independence among the contrasts to check if the branch length transformations actually worked.

For comparison, I also calculated Moran's I for the tip data using the MORAN program (in the Phylogenetic Autocorrelation package by H.-K. Luh, J.L. Gittleman and M. Kot). Using Moran's I , tip data were judged to be significantly autocorrelated if Z_r had values greater than 1.96 at any of the specified phylogenetic distance classes along a given tree (Gittleman and Kot, 1990).

Example 1: Body mass in 49 species of carnivores and ungulates

I analysed Garland and co-workers' (1993) data set of body masses across 49 species of carnivores and ungulates, using their tip data and phylogenetic topology. The TFSI showed that the tip data were significantly phylogenetically autocorrelated (Fig. 4a; $P = 0.001$), but, as expected, after applying the independent contrasts technique, the TFSI detected no significant phylogenetic autocorrelation among the contrasts (Fig. 4b; $P = 0.828$). Body mass is a trait that tends to be strongly correlated to its phylogenetic history (Gittleman *et al.*, 1996; Abouheif and Fairbairn, 1997), and thus the results of the TFSI in this example conform to our expectation. Moran's I also detected significant phylogenetic autocorrelation for the tip data ($Z_r = 3.633$), further supporting these conclusions.

Example 2: Growth rate in 18 species of plethodontid salamanders

For this second example, I analysed Sessions and Larson's (1987) data set of growth rate across 18 species of plethodontid salamanders, using their tip data and phylogenetic topology. The TFSI showed no significant phylogenetic autocorrelation among the tip data (Fig. 5a; $P = 0.756$), or among the contrasts (Fig. 5b; $P = 0.736$). Once again, these conclusions are further supported by Moran's I ($Z_r = -0.081$). This example is interesting because it reinforces the notion that not all traits are correlated to their phylogenetic history and require transformation by a PCM (Gittleman *et al.*, 1996). Clearly, although we do expect most traits to be phylogenetically autocorrelated, there are specific evolutionary scenarios for which we expect little correlation between a trait and its phylogenetic history. For example, Hansen and Martins (1996) have shown that there will be little phylogenetic autocorrelation if, over evolutionary time, a trait has been exposed to a regime of stabilizing selection in a fluctuating environment. In general, however, we expect traits to be weakly correlated to phylogeny when their response to selection is faster in comparison to their rate of speciation (Martins and Hansen, 1996). The TFSI can serve as a valuable tool to identify such situations.

Example 3: Ejaculatory incorporation into ovaries across 34 species of Drosophila

Finally, I analysed Pitnick and co-workers' (1997) data set of ejaculatory incorporation into ovaries across 34 species of *Drosophila*, using their tip data and phylogenetic topology. In this case, the TFSI detected significant phylogenetic autocorrelation for the tip data (Fig. 6a; $P = 0.001$), as well as among the contrasts (Fig. 6b; $P = 0.028$). These conclusions

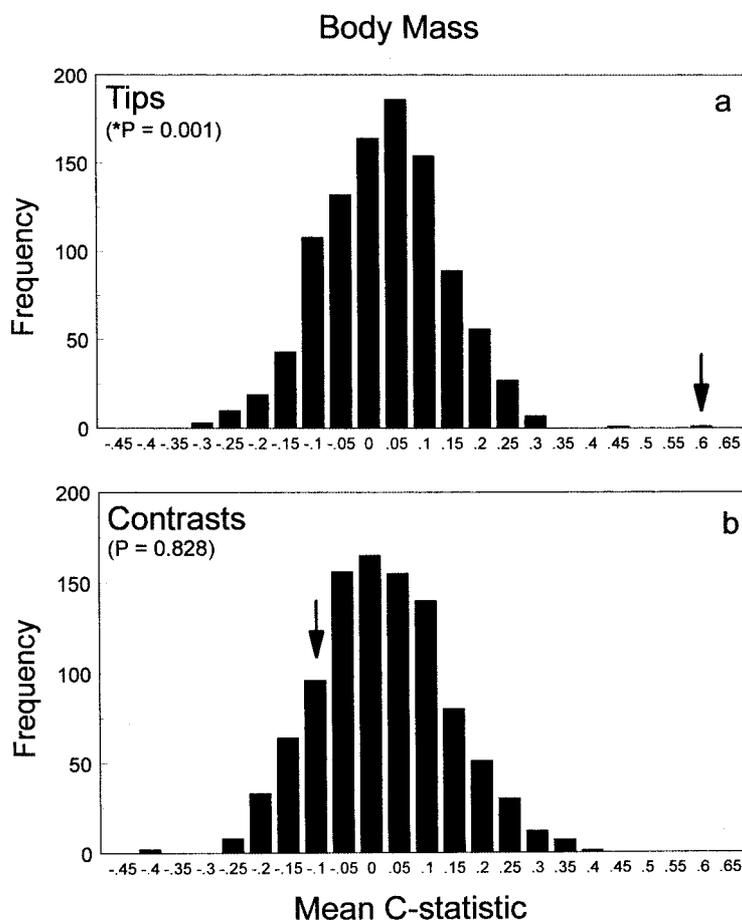


Fig. 4. The test for serial independence applied to Garland and co-workers' (1993) data set of body mass across 49 species of ungulates and carnivores. The arrows indicate the position of the observed mean C-statistic relative to the null hypothesis sampling distribution of randomized mean C-statistics. An asterisk indicates that the P -value of the observed mean C-statistic is statistically significant at an alpha of 0.05. The frequency distribution in (a) represents the mean C-statistics calculated from the body mass data along the tips of the phylogeny (tips), whereas the frequency distribution in (b) represents the mean C-statistics calculated from independent contrasts.

are further supported by Moran's I ($Z_r = 2.559$) and the diagnostic plots proposed by Garland *et al.* (1992). I detected a significant positive trend between the absolute value of the standardized contrasts and their branch lengths, indicating that the contrasts were not adequately standardized. After logarithmically transforming (\log_{10}) the branch lengths and recalculating the contrasts, however, the degree of phylogenetic autocorrelation among the contrasts was no longer significant, but remained high relative to the other data sets examined.

We should expect such cases to occur frequently. Because the independent contrasts method assumes that characters have evolved according to a Brownian motion model of

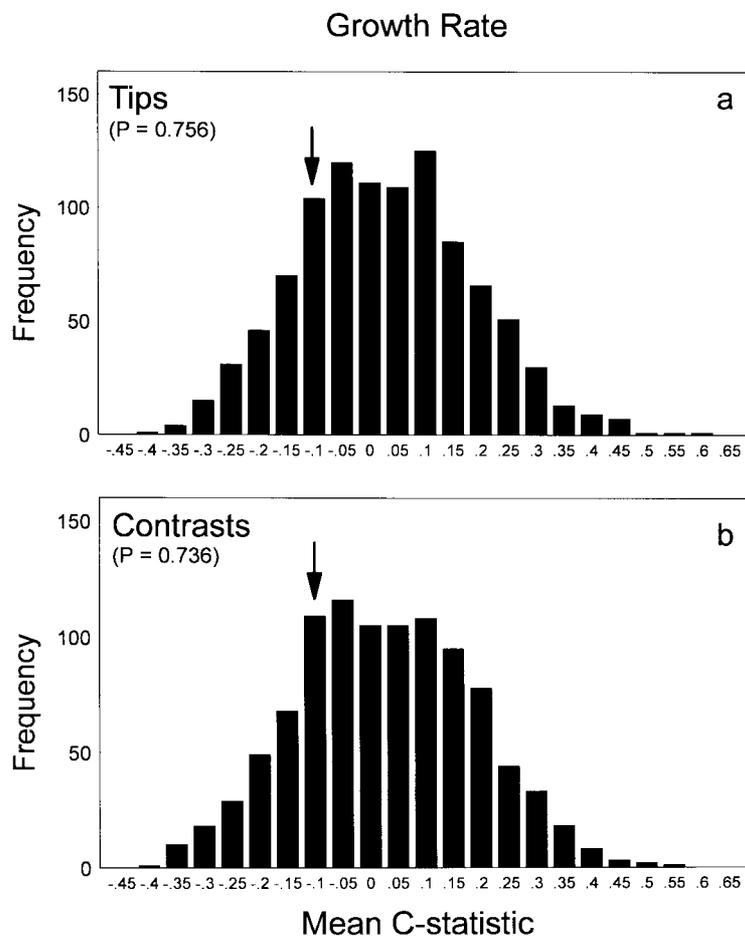


Fig. 5. The test for serial independence applied to Sessions and Larson's (1987) data set of growth rate across 18 species of plethodontid salamanders. The arrows indicate the position of the observed mean C-statistic relative to the null hypothesis sampling distribution of randomized mean C-statistics. An asterisk indicates that the P -value of the observed mean C-statistic is statistically significant at an alpha of 0.05. The frequency distribution in (a) represents the mean C-statistics calculated from the growth rate data along the tips of the phylogeny (tips), whereas the frequency distribution in (b) represents the mean C-statistics calculated from independent contrasts.

evolutionary change, any significant departure from this assumption may cause contrasts to become non-independent from one another (Díaz-Uriarte and Garland, 1996, 1998). For example, Price (1997) argued that, if some clades have higher rates of evolution than other clades, or if large evolutionary changes in a trait occur in the lower part of the phylogeny followed by smaller changes in the upper part, then contrasts from some parts of the tree are expected to be more similar to each other than other parts. Thus, as is evident from this example and previous simulation studies (Díaz-Uriarte and Garland, 1996, 1998), it is very important to apply the TFSI after using a PCM, and critically evaluate the standardization of the contrasts as proposed by Garland *et al.* (1992).

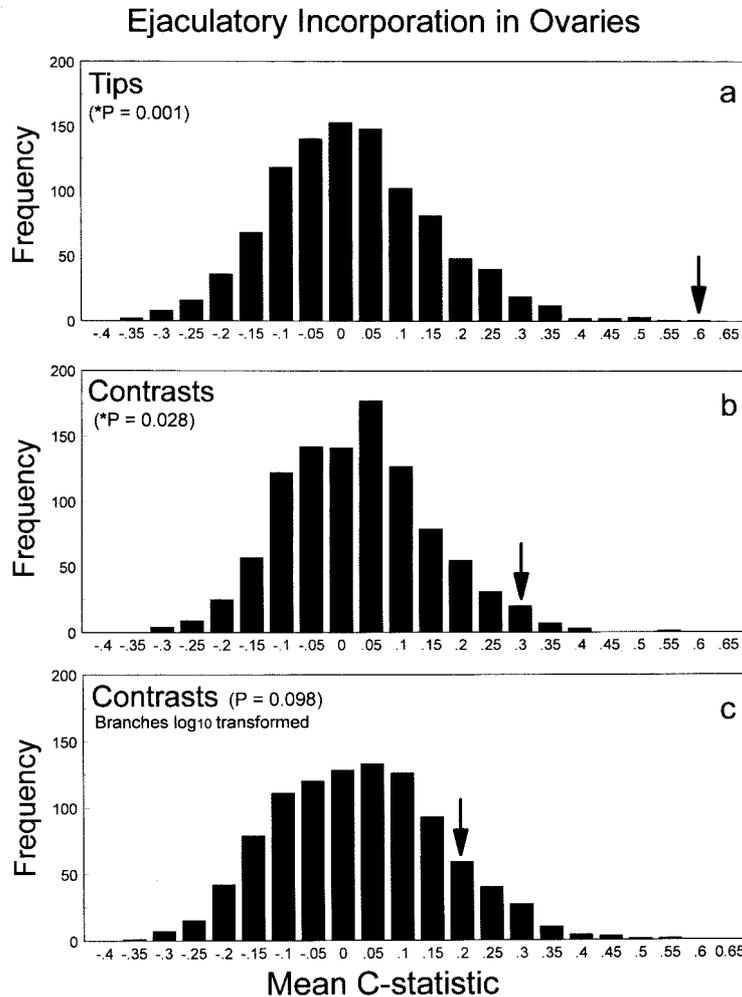


Fig. 6. The test for serial independence applied to Pitnick and co-workers' (1997) data set of ejaculatory incorporation into ovaries across 34 species of *Drosophila*. The arrows indicate the position of the observed mean C-statistic relative to the null hypothesis sampling distribution of randomized mean C-statistics. An asterisk indicates that the *P*-value of the observed mean C-statistic is statistically significant at an alpha of 0.05. The frequency distribution in (a) represents the mean C-statistics calculated from the ejaculatory incorporation data along the tips of the phylogeny (tips), whereas the frequency distribution in (b) represents the mean C-statistics calculated from independent contrasts. The frequency distribution in (c) represents mean C-statistics calculated from independent contrasts that were adequately standardized by logarithmically transforming the branch lengths in the phylogeny.

DIFFICULTIES AND CHALLENGES

The three examples shown above demonstrate that the TFSI can successfully detect different degrees of phylogenetic autocorrelation in both small and large phylogenies, as well as in different types of phenotypic characters. These examples also show the importance of

applying the TFSI in conjunction with the methodological approach in Fig. 1 to test the assumption of phylogenetic independence before and after the application of a PCM.

Martins (1996a) has criticized the general use of statistical diagnostics to test the assumption of phylogenetic independence on the grounds that they may be at high risk of committing type II errors (i.e. claiming that the data are independent, when, in actual fact, they are not), especially when used on data sets with small sample sizes or with incompletely resolved phylogenies. Although this is a valid criticism, it should not be taken to mean that one should altogether abandon the methodological approach in Fig. 1, and proceed to rationalize whether it is safe to always ignore or to always apply PCMs. The assumption of phylogenetic independence is clearly an empirical issue. If, in a particular situation, one is concerned about the statistical power of the TFSI, then one should be conservative and set *a priori* the probability at $P = 0.1$, in which case the test becomes more statistically liberal, and the probability of rejecting the null hypothesis (i.e. that there is no phylogenetic autocorrelation in the data) is increased. Furthermore, testing the assumption of phylogenetic independence after a PCM has been applied to transform comparative data, is just as important as it is beforehand.

One of the major strengths of the TFSI as a statistical diagnostic, is that it does not assume a model of evolutionary change, and does not require branch lengths. This makes the TFSI a practical tool for allowing researchers to determine whether or not the comparative data are significantly phylogenetically autocorrelated without incurring the costs associated with assuming that the branch lengths and the model of evolutionary change are known and accurate. In this context, it will be important to determine through simulations how well the TFSI performs when applied to data that have been simulated under different models of evolutionary change, and to data that have been simulated along phylogenies with different branch lengths and divergence times.

Successful application of the TFSI does, however, depend on the veracity of the phylogenetic topology used to conduct the test. If several equally likely or parsimonious topologies exist, then one should conduct the TFSI on each one of them to test the sensitivity of the conclusions to the topology used (Donoghue and Ackerly, 1996). In such cases, researchers should report the mean and range of *P*-values obtained from the test. Moreover, it is often the case that phylogenetic topologies are either incomplete or partially resolved. The TFSI can be applied in these situations by randomizing the order of taxa within a polytomy in the same manner that one randomly rotates the orientation of the nodes outside the polytomy. For example, imagine an extreme case where the TFSI is conducted on a star phylogeny. The order of all taxa will be randomized, and there will be no phylogenetic autocorrelation in the data – that is, one assumes that the data are historically independent of one another (Felsenstein, 1985). Importantly, incomplete or partially resolved phylogenetic topologies reduce the statistical power of the TFSI, as well as any PCM, and thus one should be conservative in these cases (Purvis and Garland, 1993; Losos, 1994; Martins, 1996b). Again, further studies are required to determine how statistically robust and powerful the TFSI is in such circumstances.

Finally, the statistical procedures described in this paper, for adopting von Neumann and co-workers' (1941) TFSI to test the assumption of phylogenetic independence among continuously valued characters, can also be used to adopt the 'runs test' (Sokal and Rohlf, 1995) to test the assumption of phylogenetic independence among discretely valued characters. To conduct the runs test, one simply has to assign a character state – for example, the presence (+) or absence (–) of a particular trait – to each taxon in the

comparative data set. One then has to order these '+' and '-' signs according to the phylogeny, and perform a runs test in the same manner that one would perform the TFSI (Figs 2 and 3). If there is no phylogenetic autocorrelation in the sample, then the '+' and '-' signs should be in random sequence. Conversely, if the data are phylogenetically autocorrelated, then there will more or fewer runs of '+' and '-' signs than would be expected by chance alone.

To conclude, the assumption of independence is one of the most basic and fundamental assumptions in standard statistical analyses. By explicitly incorporating phylogenies and treating this assumption as an empirical issue to be tested, we not only increase the rigour with which we approach comparative analyses, but gain insight into the magnitude and patterns of the historical influences of phylogeny on the evolution of phenotypic traits.

ACKNOWLEDGEMENTS

This project would not have been possible without the computer programming expertise and help of J. Reeve. I would also like to thank A. Chi and M. Rosenberg for technical assistance. D. Adams, D.J. Fairbairn, M. Rosenberg, J. Stone, W. AlGharaibeh, J. Felsenstein, F.J. Rohlf, R. Sokal and T. Garland Jr. provided many insightful discussions on this topic. I am indebted to D. Adams, B. Rannala, J. Stone, M. Rosenberg, T. Garland Jr., E. Martins, J. Gittleman and R. Sokal for reading earlier versions of this manuscript and providing valuable comments. The paper is dedicated to the graduate students of Ecology and Evolution at SUNY Stony Brook for fostering a critical environment in which to test and encourage new ideas. This work was supported by a graduate fellowship from the Fonds pour la Formation de Chercheurs et l'Aides a la Recherche (FCAR) from the Quebec Government.

REFERENCES

- Abouheif, E. 1998. Random trees and the comparative method: A cautionary tale. *Evolution*, **52**: 1197–1204.
- Abouheif, E. and Fairbairn, D.J. 1997. A comparative analysis of allometry for sexual size dimorphism: Assessing Rensch's rule. *Am. Nat.*, **149**: 540–562.
- Ackerly, D.D. and Donoghue, M.J. 1998. Leaf size, sapling allometry, and Corner's rules: Phylogeny and correlated evolution in maples (*Acer*). *Am. Nat.*, **152**: 767–791.
- Bell, G. 1989. A comparative method. *Am. Nat.*, **133**: 553–571.
- Björklund, M. 1994. The independent contrast method in comparative biology. *Cladistics*, **10**: 425–433.
- Cheverud, J.M., Dow, M.M. and Leutenegger, W. 1985. The quantitative assessment of phylogenetic constraints in comparative analyses: Sexual dimorphism in body weight among primates. *Evolution*, **39**: 1335–1351.
- Cliff, A.D. and Ord, J.K. 1981. *Spatial Processes: Models and Applications*. London: Pion.
- Díaz-Uriarte, R. and Garland, T., Jr. 1996. Testing hypotheses of correlated evolution using phylogenetically independent contrasts: Sensitivity to deviations from Brownian motion. *Syst. Biol.*, **45**: 27–47.
- Díaz-Uriarte, R. and Garland, T., Jr. 1998. Effects of branch length errors on the performance of phylogenetically independent contrasts. *Syst. Biol.*, **47**: 654–672.
- Diniz-Filho, J.A.F., de Sant' Ana, C.E.R. and Bini, L.M. 1998. An eigenvector method for estimating phylogenetic inertia. *Evolution*, **52**: 1247–1262.
- Donoghue, M.J. and Ackerly, D.D. 1996. Phylogenetic uncertainties and sensitivity analyses in comparative biology. *Phil. Trans. R. Soc. Lond. B*, **351**: 1241–1249.

- Felsenstein, J. 1985. Phylogenies and the comparative method. *Am. Nat.*, **125**: 1–15.
- Felsenstein, J. 1988. Phylogenies and quantitative characters. *Ann. Rev. Ecol. Syst.*, **19**: 445–471.
- Garland, T., Jr., Harvey, P.H. and Ives, A.R. 1992. Procedures for the analysis of comparative data using phylogenetically independent contrasts. *Syst. Biol.*, **41**: 18–31.
- Garland, T., Jr., Dickerman, A.W., Janis, C.M. and Jones, J.A. 1993. Phylogenetic analysis of covariance by computer simulation. *Syst. Biol.*, **42**: 265–292.
- Gittleman, J.L. and Kot, M. 1990. Adaptation: Statistics and a null model for estimating phylogenetic effects. *Syst. Zool.*, **39**: 227–241.
- Gittleman, J.L. and Luh, H.-K. 1994. Phylogeny, evolutionary models and comparative methods: A simulation study. In *Phylogenetics and Ecology* (P. Eggleton and D. Vane-Wright, eds), pp. 103–122. London: Academic Press.
- Gittleman, J.L., Anderson, C.G., Kot, M. and Luh, H.-K. 1996. Phylogenetic lability and rates of evolution: A comparison of behavioral, morphological and life history traits. In *Phylogenies and the Comparative Method in Animal Behaviour* (E.P. Martins, ed.), pp. 166–205. Oxford: Oxford University Press.
- Gould, S.J. 1997. The exaptive excellence of spandrels as a term and prototype. *Proc. Natl. Acad. Sci. USA*, **94**: 10750–10755.
- Gould, S.J. and Lewontin, R.C. 1979. The spandrels of San Marco and the Panglossian paradigm: A critique of the adaptationist program. *Proc. R. Soc. Lond. B*, **205**: 581–598.
- Hansen, T.F. and Martins, E.P. 1996. Translating between microevolutionary process and macroevolutionary patterns: The correlation structure of interspecific data. *Evolution*, **50**: 1404–1417.
- Harvey, P.H. and Pagel, M.D. 1991. *The Comparative Method in Evolutionary Biology*. Oxford: Oxford University Press.
- Harvey, P.H., Read, A.F. and Nee, S. 1995. Why ecologists need to be phylogenetically challenged. *J. Ecol.*, **83**: 535–536.
- Losos, J.B. 1994. An approach to the analysis of comparative data when a phylogeny is unavailable or incomplete. *Syst. Biol.*, **43**: 117–123.
- Martins, E.P. 1994. Estimating the rate of phenotypic evolution from comparative data. *Am. Nat.*, **144**: 193–209.
- Martins, E.P. 1996a. Phylogenies, spatial autoregression, and the comparative method: A computer simulation test. *Evolution*, **50**: 1750–1765.
- Martins, E.P. 1996b. Conducting phylogenetic comparative studies when the phylogeny is not known. *Evolution*, **50**: 12–22.
- Martins, E.P. and Garland, T., Jr. 1991. Phylogenetic analyses of the correlated evolution of continuous characters: A simulation study. *Evolution*, **45**: 534–557.
- Martins, E.P. and Hansen, T.F. 1996. The statistical analysis of interspecific data: A review and evaluation of phylogenetic comparative methods. In *Phylogenies and the Comparative Method in Animal Behaviour* (E.P. Martins, ed.), pp. 22–75. Oxford: Oxford University Press.
- McKittrick, M.C. 1993. Phylogenetic constraint in evolutionary time: Has it any explanatory power. *Ann. Rev. Ecol. Syst.*, **24**: 307–330.
- Miles, D.B. and Dunham, A.E. 1993. Historical perspectives in ecology and evolutionary biology: The use of phylogenetic comparative analyses. *Ann. Rev. Ecol. Syst.*, **24**: 587–619.
- Moran, P.A.P. 1950. Notes on continuous stochastic phenomena. *Biometrika*, **37**: 17–23.
- Pitnick, S., Spicer, G.S. and Markow, T. 1997. Phylogenetic examination of female incorporation of ejaculate in *Drosophila*. *Evolution*, **51**: 833–845.
- Price, T. 1997. Correlated evolution and independent contrasts. *Phil. Trans. R. Soc. Lond. B*, **352**: 519–528.
- Purvis, A. and Garland, T., Jr. 1993. Polytomies in comparative analyses of continuous characters. *Syst. Biol.*, **42**: 569–575.
- Purvis, A., Gittleman, J.L. and Luh, H.-K. 1994. Truth or consequences: Effects of phylogenetic accuracy on two comparative methods. *J. Theor. Biol.*, **167**: 293–300.

- Rice, W.R. and Gaines, S.D. 1994. 'Heads I win, tails you lose': Testing directional alternative hypotheses in ecological and evolutionary research. *Trends Ecol. Evol.*, **9**: 235–237.
- Ricklefs, R.E. 1996. Phylogeny and ecology. *Trends Ecol. Evol.*, **11**: 229–230.
- Ricklefs, R.E. and Starck, J.M. 1996. Applications of phylogenetically independent contrasts: A mixed progress report. *Oikos*, **77**: 167–172.
- Rohlf, F.J. and Sokal, R.R. 1995. *Statistical Tables*. New York: W.H. Freeman.
- Sessions, S.K. and Larson, A. 1987. Developmental correlates of genome size in plethodontid salamanders and their implications for genome evolution. *Evolution*, **41**: 1239–1251.
- Sokal, R.R. and Rohlf, F.J. 1995. *Biometry*. New York: W.H. Freeman.
- von Neumann, J., Kent, R.H., Bellinson, H.R. and Hart, B.I. 1941. The mean square successive difference. *Ann. Math. Stat.*, **12**: 153–162.
- Westoby, M., Leishman, M.R. and Lord, J.M. 1995. On misinterpreting the 'phylogenetic correction'. *J. Ecol.*, **83**: 531–534.
- Young, L.C. 1941. On randomness in ordered sequences. *Ann. Math. Stat.*, **12**: 293–300.
- Zar, J.H. 1984. *Biostatistical Analysis*. Englewood Cliffs, NJ: Prentice-Hall.