

Appendix from R. P. Freckleton et al., ‘Relating Traits to Diversification: A Simple Test’ (Am. Nat., vol. 172, no. 1, p. 000)

Diversification Model and Variance in the Average Value of the Predictor

The logistic model (eq. [2] or [3]) is a nonlinear function of the predictor x , such that variance in x will affect the mean value of λ (or μ). To illustrate this, a Taylor series approximation of $\lambda(x)$ about \bar{x} yields

$$\lambda(x) = \lambda(\bar{x}) + f'(x)(x - \bar{x}) + \frac{f''(x)}{2}(x - \bar{x})^2 + \text{higher-order terms.}$$

Taking expectations gives

$$\overline{\lambda(x)} = \lambda(\bar{x}) + \frac{f''(x)}{2} \text{Var}(x) + \text{higher-order terms.}$$

So the difference between the mean value of λ and the value of λ predicted at the mean value of x depends on the nonlinearity in the function relating the two, as measured by the second (and higher) differentials.

For equation (2), the mean value of λ is approximated by (assuming terms higher than second order are negligible)

$$\overline{\lambda(x)} \approx (1 + \beta^2) \text{Var}(x) \lambda(\bar{x}).$$

Thus, the mean speciation rate is higher than that predicted at the average trait value by a factor proportional to the variance in x and the square of β . Notably, however, the difference in this case is simply a multiplicative factor, and the underlying relationship is of the same form as equation (2).

The effects of this need not be so simple. For instance, consider the more complex logistic model (e.g., Paradis 2005):

$$\lambda_i = \frac{1}{1 + \exp(-\beta x_i - \alpha)}.$$

As shown in figure A1, the nonlinear averaging effect can be considerable for this model. Two values of β are illustrated, $\beta = 1$ and $\beta = 4$. It is clear that increasing the variance in x can have a marked effect on the mean realized rate of speciation compared with the value predicted by the relationship for a given x . The reason for this is that the nonlinear terms in the above equations are large relative to the values of the function at mean x . The variance in x is expected to increase with time under a Brownian model. Consequently, as phylogenies grow, the relationship between node depth and x would be expected to become progressively shallower and more linear.

Illustration of the Method

We illustrate the basis for the test we propose before going on to explore its behavior in simulated data. Figure A2A and A2B shows two hypothetical traits for the same phylogeny. The values at the tips are the states of each trait, together with the node depth for each tip. In the first case, there is a positive correlation between node density and trait values (fig. A2B), whereas in the second case, there is no correlation (fig. A2C). The phylogenetically corrected correlations for the two data sets are $r = 0.81$, $n = 30$ (fig. 1B) and $r = 0.36$, $n =$

30 (fig. A2C). Note that in figure A2, many species inevitably share the same node depth. Consequently, in presenting the results of this type of analysis, we have found it useful to present the average trait value (together with error bars) across all species sharing the same node depth in order to more clearly discern patterns (e.g., see figs. 2, A1). The statistical test, however, utilizes the raw data as in figure A2.

Analysis of Hypothetical Data Using Nested Comparisons

Here we outline by means of a simple example that the analysis developed by Isaac et al. (2003) does not estimate the simple model for diversification in species-level phylogenies that we or Paradis (2005) use. What we do is to show that if we assume a simple linear relationship between speciation rate and traits, none of the metrics proposed by Isaac et al. (2003) recover that relationship, even if there is no error in the data. On the other hand, we show that the method we suggest is able to recover the relationship exactly. The conclusion we draw is that the null model assumed in the method by Isaac et al. (2003) is rather different from that assumed here and that, in application of either method, careful consideration should be given to the expected form of relationship between traits and diversification.

Our starting point is a simple linear model of diversification, such that the rate of lineage splitting (λ) per unit time is a linear function of a trait T :

$$\lambda = a + bT. \quad (\text{A1})$$

Figure A1 shows a simple hypothetical example of a phylogeny of six extant species that might have evolved according to such a rule. There is a simple linear association between the trait T and the number of speciation events between the root and the tips (e.g., as seen in fig. 1 in the main text, resulting from the more complex simulation). We emphasize that we do not regard this as a realistic model for the evolution of a group of species; however, we use this simple phylogeny to demonstrate our point, since it is straightforward to visualize how the various metrics behave throughout the tree and because it accords with equation (A1).

For this simple example, we calculated contrasts in T (using reconstructed values, denoted X) as well as contrasts in node density, as outlined in the text. We also calculated four of the statistics of Isaac et al. (2003). These are (1) absolute rate difference (ARD), (2) absolute difference in clade size ΔS , (3) relative rate difference (RRD), and (4) phylogenetic diversity index (PDI). A full definition and justification for each as a potential measure is given by Isaac et al. (2003). It is clear in figure A3 that none of these yields a linear relationship, although in three of the cases, the relationship is positive.

Figure A4 generalizes the simple model in figure A3 to a group of N species evolved in such a way. Figure A4 also shows how it is possible to generate analytical expressions for the various metrics for diversification as well as the contrasts generated from T for this dummy phylogeny. On the basis of this, we can generate relationships between the contrasts in the trait and the rate statistics for phylogenies of varying sizes (fig. A5). As shown in figure A5, the relationship between ΔX and the rate statistics are nonlinear and vary greatly. The ARD index performs very poorly and even varies negatively with ΔX . The RRD associates reasonably for small phylogenies but, as is clear, performs poorly for trees with much greater than 100 tips. It is highly nonlinearly related to ΔX , yielding a much weaker correlation. The PDI is even more nonlinearly related to ΔX ; indeed, it is effectively unrelated to ΔX for most of the range of ΔX .

The consequence of the nonlinearity in figure A5 is threefold. First, the indices cannot be used to estimate b ; that is, they cannot be used to infer underlying parameters. Second, the power of these indices will be lowered owing to the nonlinearity since the R^2 for analyses assuming linear relationships will always be lower than for the nonlinear model. Finally, the bias will increase with N , with the consequence that power will be weakened as the size of the phylogeny is increased.

Of course, the model in figures A3 and A4 is an abstraction; however, the lessons from this should apply to any phylogeny. The problem is that none of the metrics of Isaac et al. (2000) estimate λ or $\Delta\lambda$; therefore, any phylogeny would be subject to such an effect.

R Code

R code is available as a downloadable file and from the corresponding author on request for running example analyses and simulations to determine Type I error rates of single predictor analyses on real data.[†] As outlined in

[†] Code that appears in the *American Naturalist* has not been peer reviewed, nor does the journal provide support.

the main text, this is probably an important first step when conducting multiple predictor analyses since it is difficult to envisage how to run a single simulation study that could cover all eventualities. This code is being continually updated; we recommend contacting R. P. Freckleton for the latest version (r.freckleton@sheffield.ac.uk).

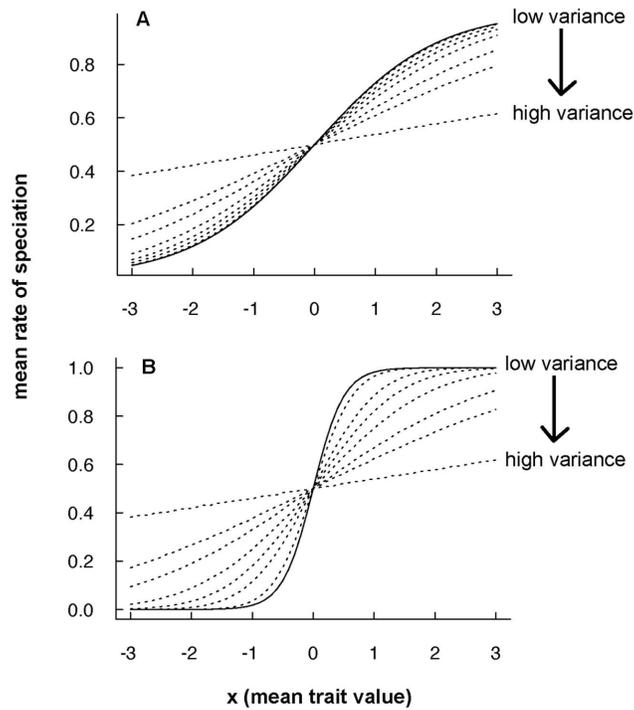


Figure A1: Relationship between the mean rate of speciation and traits for different levels of trait variance in a nonlinear model. This figure illustrates that variation in the value of a trait along the branches of a phylogeny contributes to the mean rate of speciation if traits are correlated with speciation and if the relationship between speciation and trait values is nonlinear. The function shown is a logistic function (for details, see Paradis 2005 and text) with moderate nonlinearity (A; $\beta = 1$) and strong nonlinearity (B; $\beta = 4$). The effects of increasing variance in trait values are shown.

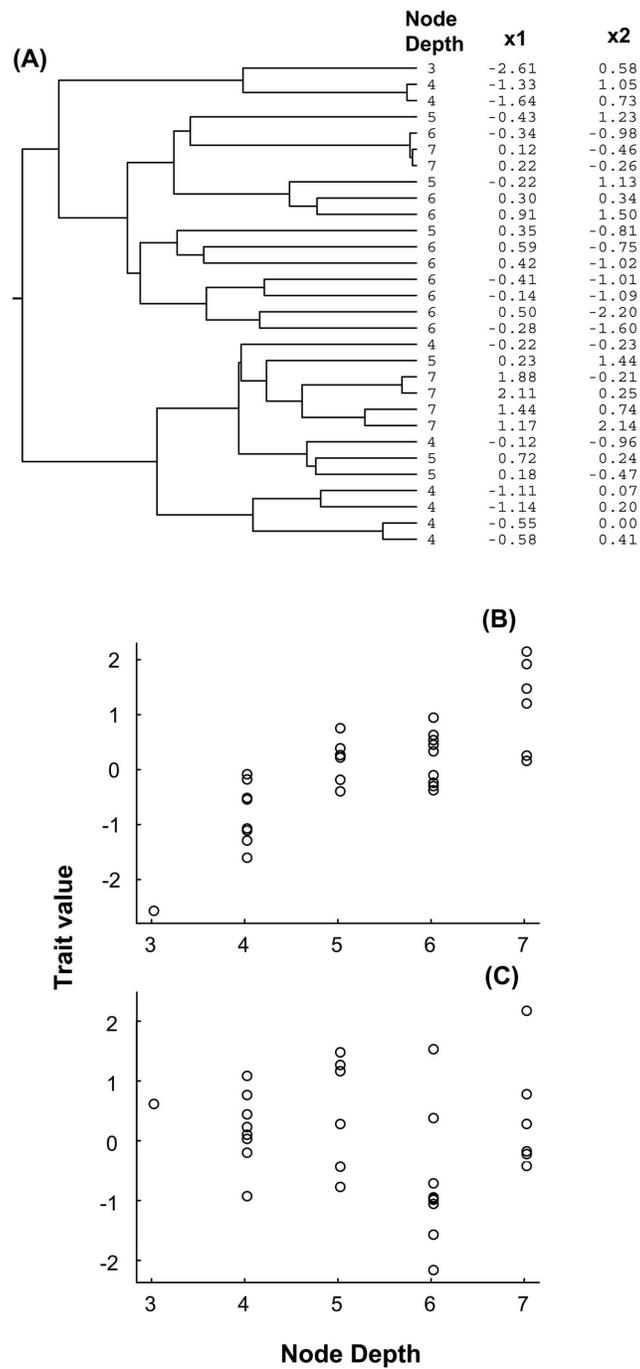


Figure A2: Example of the application of the test described in the text to hypothetical data. *A*, A phylogeny, together with node depth (the number of nodes linking each tip to the root of the phylogeny) and with two traits. *B*, Correlation of node depth with x_1 . *C*, Correlation of node depth with x_2 .

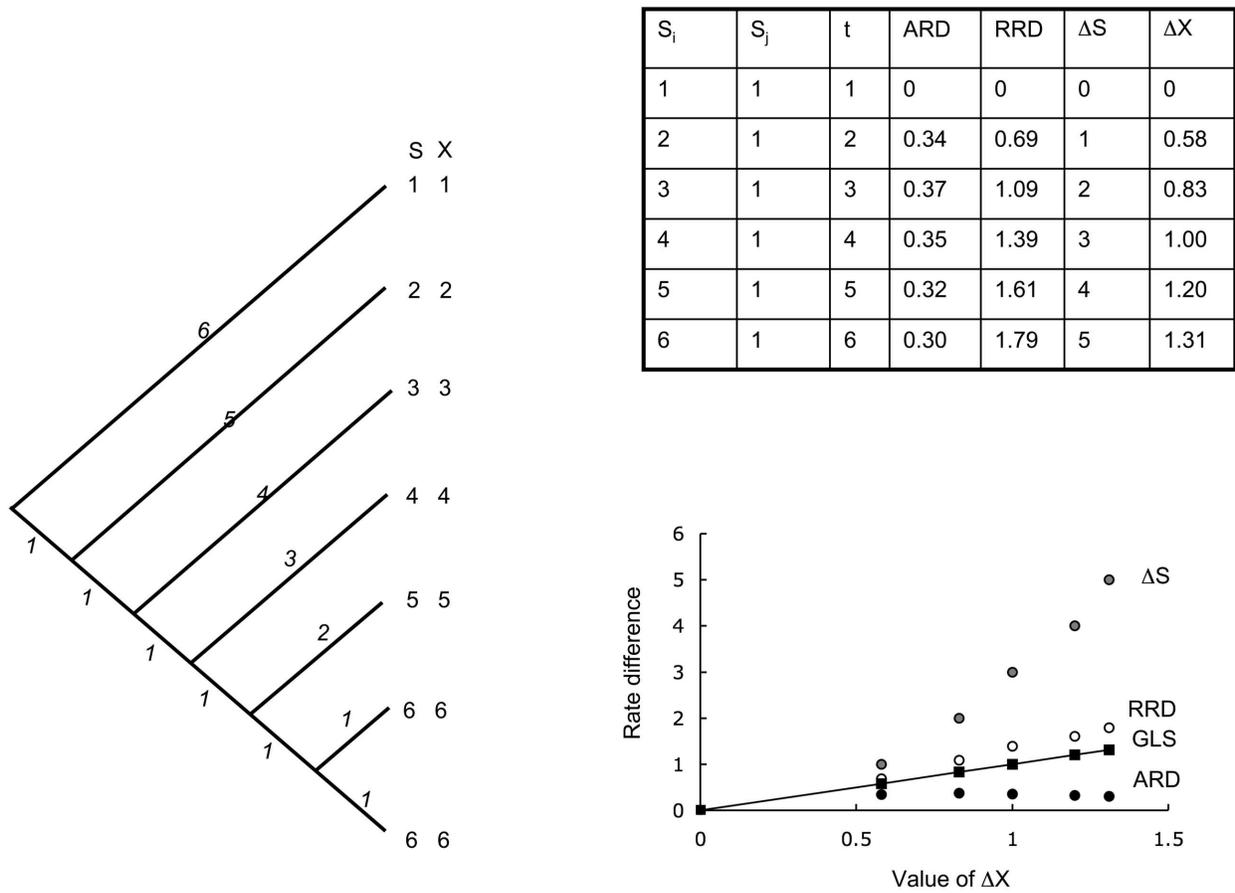


Figure A3: Hypothetical phylogeny and the state of a trait X that is correlated with speciation. The numbers in italic are branch lengths. The table shows the details of the calculations used to generate contrasts. The parameters S_i and S_j are the node densities of the branches subtended from each node in the phylogeny. The parameter t is the branch length; then shown are the contrasts in X and the values of the other diversification metrics. The graph shows the relationship between the trait contrast and the rate contrasts. The true relationship is represented by the line, which is exactly retrieved by correlating the trait contrasts with contrasts in node density (points denoted *GLS*).

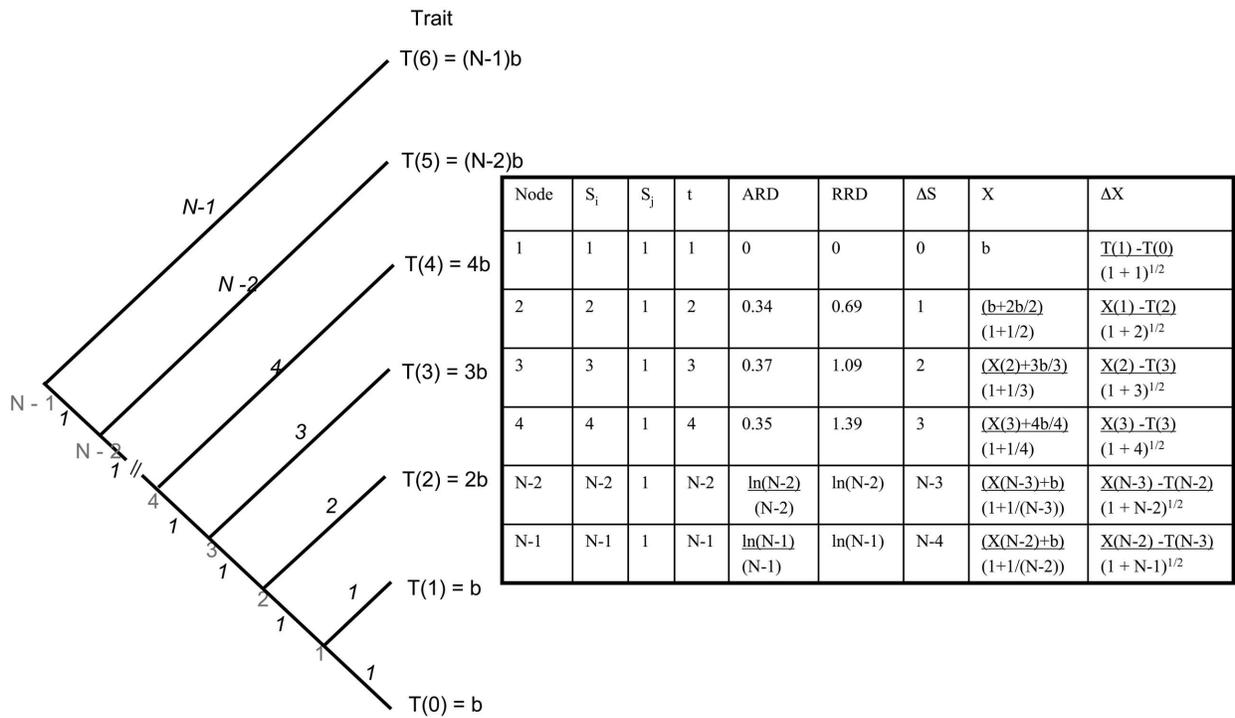


Figure A4: Generalized hypothetical phylogeny and the state of a trait X that is correlated with speciation. The numbers in italic are branch lengths. The table shows the details of the calculations used to generate contrasts. The parameters S_i and S_j are the node densities of the branches subtended from each node in the phylogeny. The parameter t is the branch length; then shown are the contrasts in X and the values of the other diversification metrics. The table presents an algorithm for iterating this model for a phylogeny of any size.

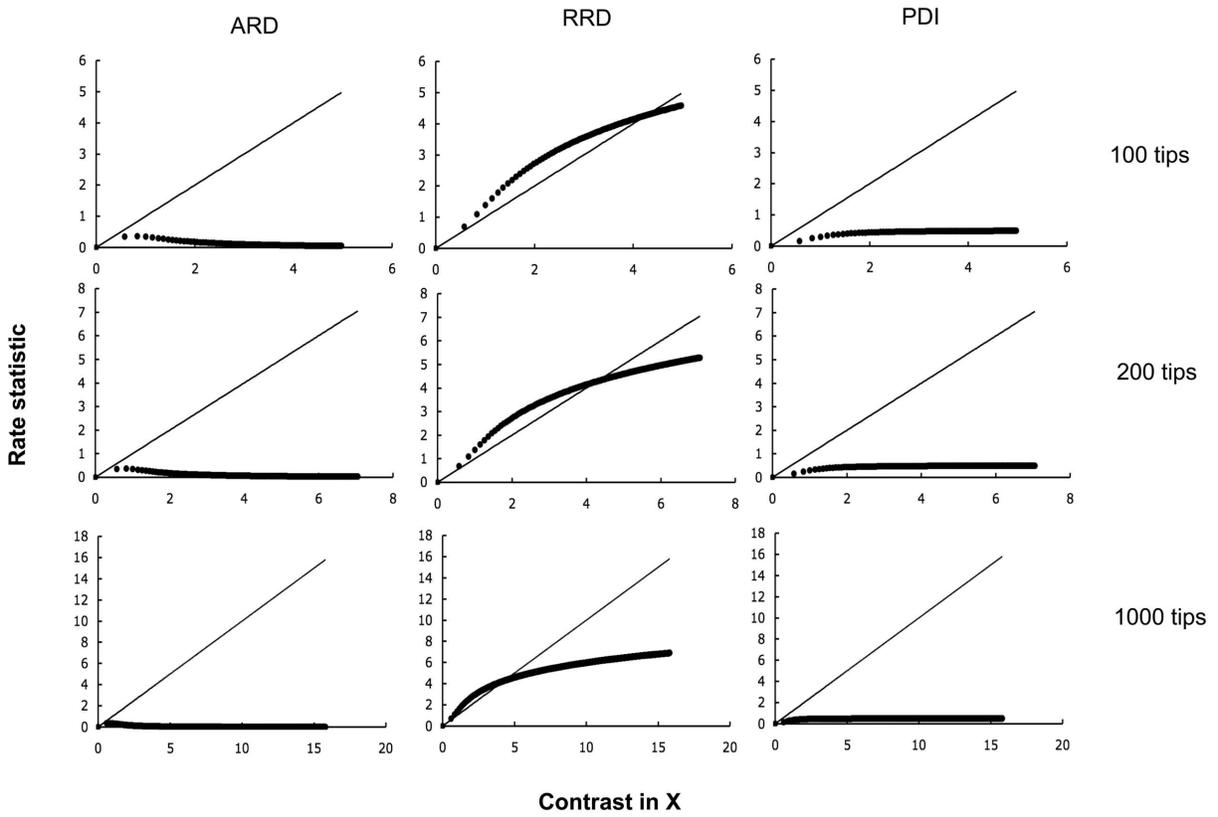


Figure A5: The algorithm in figure A4 was used to generate estimates of three diversification rate statistics and plotted against trait contrasts for phylogenies differing in size. The true relationship between rate of diversification and trait value is shown by the line.