

Calculating independent contrasts for the comparative study of substitution rates

John J. Welch^{a,b,*}, David Waxman^b

^a*Institute of Evolutionary Biology, School of Biological Sciences, University of Edinburgh, West Mains Rd., Edinburgh EH9 3JT, UK*

^b*Centre for the Study of Evolution, School of Life Sciences, University of Sussex, Falmer, Brighton BN1 9QG, UK*

Received 13 September 2007; received in revised form 17 December 2007; accepted 20 December 2007

Available online 27 December 2007

Abstract

Phylogenetic comparative methods have been used to study patterns of correlated evolution between biological traits of all kinds, and are increasingly used to identify predictors of the rate of DNA substitution. Substitution rate differs from most traits studied in that it cannot be observed directly, but must be inferred from substitutions accrued over a long period of time. Studying a mathematical model of trait and rate evolution, we show that the special nature of substitution rates can lead to a severe loss of power for standard comparative methods. We further show how strategies designed to maximise power, by increasing the number of data points, can have the opposite effect when substitution rate is involved. We then propose two modifications of the standard methods that can mitigate these problems. First, we show how pre-existing tests for homogeneity of variance can be used to identify and exclude those data from which changes in substitution rate cannot be reliably inferred. Second, we show that power can be increased by comparing substitution rate with the time average of the predictor trait along the history of the lineage, and introduce a maximum likelihood estimator of this quantity.

© 2008 Elsevier Ltd. All rights reserved.

Keywords: Comparative methods; Molecular evolution; Trait evolution

1. Introduction

As the availability of DNA sequence data grows, the causes of molecular evolution are increasingly open to empirical investigation. One important goal is to identify those aspects of a population's biology that influence its rate of nucleotide substitution. A large number of biological factors have been proposed as determinants of the substitution rate—from generation time and metabolic rate, to speciation rate and environmental energy—and large-scale data sets are being used to test the various hypotheses (Bromham, 2003; Barraclough and Savolainen, 2001; Bromham et al., 1996; Webster et al., 2003; Davies et al., 2004; Gillooly et al., 2005; Thomas et al., 2006;

Davies and Savolainen, 2006; Wright et al., 2006; Fontanillas et al., 2007).

Rigorous tests of the association between rate and trait rely on phylogenetic comparative methods—a set of well-developed procedures for studying correlated evolution between traits (Felsenstein, 1973; Grafen, 1989; Harvey and Pagel, 1991; Garland et al., 1992; Freckleton, 2000; Freckleton and Harvey, 2006). These methods were developed because comparative data are unlikely to meet the assumptions of standard statistical tests of association, such as correlation or regression, and themselves rely on model-based assumptions about the process of trait evolution over the phylogeny (Harvey and Pagel, 1991).

In the present work, we show that phylogenetic comparative methods face a severe loss of power when substitution rate is studied. The problems arise from the simple fact that substitution rates cannot be observed directly, even in principle, but must be estimated from the number of substitutions that actually occur over a given period of divergence. This is shown to violate the

*Corresponding author at: Institute of Evolutionary Biology, School of Biological Sciences, University of Edinburgh, West Mains Rd., Edinburgh EH9 3JT, UK. Tel.: +44 131 650 8683; fax: +44 131 650 6564.

E-mail addresses: J.J.Welch@ed.ac.uk (J.J. Welch), D.Waxman@sussex.ac.uk (D. Waxman).

assumptions on which the comparative methods rely, and in a way that can mislead the standard diagnostic tests of those assumptions. We then introduce some modifications of the standard methods that can mitigate the problems introduced by substitution rate data.

We begin by outlining the phylogenetic comparative method on which the succeeding analysis is based.

2. Phylogenetically independent contrasts

The approach discussed here is a simplified version of the method of independent contrasts (Felsenstein, 1985; Harvey and Pagel, 1991). Using a phylogeny of the relevant taxa, we begin by choosing phylogenetically independent sister pairs, that is, a set of pairs for which the lineages connecting each pair are non-overlapping. Then, for each trait, a data point is taken to be the difference in the log transformed trait values for the pair. These data points are then standardised by dividing each by some measure of the divergence separating the pair, usually the square root of the age of the pair's most recent common ancestor. These procedures aim to generate a set of data points that are statistically independent, with expectations of zero and constant variances. The points can therefore be used, for example, in a regression forced through the origin (Garland et al., 1992). The sole difference between this approach and the full method of independent contrasts, is our failure to include contrasts between reconstructed trait values at the internal nodes of the phylogeny (Felsenstein, 1985). The justification for excluding such contrasts is given in the discussion.

The procedures just described rely on three assumptions about the process of trait evolution. Each of these assumptions is associated with customised diagnostic tests, which together form an integral part of the method (Garland et al., 1992; Freckleton, 2000; Freckleton and Harvey, 2006).

The first assumption is that the evolutionary changes in each trait took place independently in each lineage in the phylogeny—i.e., an increase in one lineage made an increase in any other lineage neither more nor less likely. This assumption may be violated if the lineages were subject to strong ecological interactions (Price, 1997), and so Freckleton and Harvey (2006) introduced a number of tests of the independence of evolutionary changes.

The second assumption is that, on the scale of measurement chosen, the evolutionary changes in each trait took place independently of its initial state—i.e., large changes in trait value were no more or less likely in lineages with larger ancestral trait values. This assumption is the rationale for the log transformation of the trait values, because proportional changes are more likely than absolute changes to have the desired property. Violation of the second assumption will lead to inhomogeneity of variance amongst contrasts, and therefore to the unequal weighting of points in the statistical test. As suggested by Freckleton (2000), the validity of the assumption can be tested by

plotting the magnitude of the contrasts against the mean transformed trait value for each pair, and then testing for any correlation between the two. A significant trend would indicate that a new transformation of the trait measurements should be attempted, or that a non-parametric test be used.

The third assumption, also relating to homogeneity of variance, is that lineages tend to evolve gradually over time. This is the rationale for standardising each contrast with the square root of its divergence time, because more distantly related lineages are expected to show more variance in evolutionary outcomes. Garland et al. (1992) suggested testing this assumption by correlating the magnitude of the standardised contrasts with their standardisation factors. Any trends in the plots could be removed by transforming the standardisation factors. But if a suitable correction cannot be found, then only a sign test can be used on the data, as even a rank correlation test will be invalid (Grafen, 1989).

3. Brownian motion model of trait evolution

To appreciate why the procedures described above may be successful when applied to measurable traits, but compromised when applied to substitution rates, let us introduce the model of trait evolution on which the methods were premised, namely, geometric Brownian motion (Felsenstein 1973, 1985; Harvey and Pagel, 1991).

Consider a single pair of species labelled $i = 1$ and 2 , which shared a common ancestor a time t in the past. These species are each characterised by a pair of traits X and Y which are the subject of the comparative analysis. Let $X_i(t)$ and $Y_i(t)$ denote the values of the traits in species i at time t , while X_0 and Y_0 denote the trait values of their common ancestor (Fig. 1). Now assume that the log transformed trait values evolved from their ancestral states via a random walk. The two traits may evolve in a

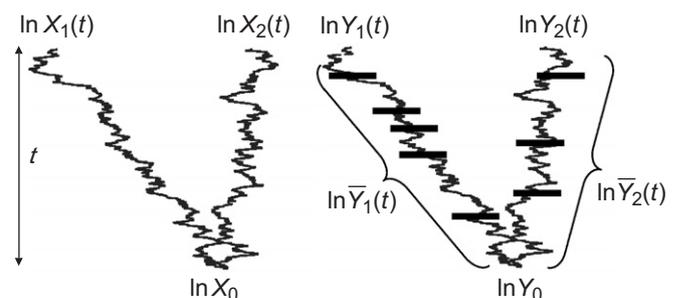


Fig. 1. Schematic representation of quantities involved in calculating independent contrasts. Shown are the evolutionary trajectories of two traits, X and Y in a pair of sister lineages evolving over a time t from a common ancestor. The mode of evolution is geometric Brownian motion (that is, the log transformed trait values undergo a random walk). Trait X is always assumed to be measurable, in that X_1 and X_2 could be direct observations. If trait Y is assumed to be the substitution rate, only the time-averaged values \bar{Y}_1 and \bar{Y}_2 can be estimated indirectly from the number of substitutions that have occurred (represented by horizontal bars).

correlated manner along each lineage, but the random walks will be uncorrelated with any other lineage. These assumptions lead to the expressions

$$\ln X_i(t) = \ln X_0 + \sigma_X W_{X_i}(t), \quad (1)$$

$$\ln Y_i(t) = \ln Y_0 + \sigma_Y \left[\sqrt{1 - \rho^2} W_{Y_i}(t) + \rho W_{X_i}(t) \right]. \quad (2)$$

Here $W_{X_i}(t)$ and $W_{Y_i}(t)$ are Wiener processes that represent the cumulative changes along a random walk (see Appendix A), while σ_X and σ_Y parameterise the evolutionary rates of the two traits. The parameter ρ is the correlation between the traits, and is generally the quantity of greatest interest.

3.1. Contrast properties for measurable traits

Before we consider the case where Y is the rate of substitution, let us briefly review results when both X and Y are traits such as body size, which can be directly measured in extant populations (Felsenstein, 1985). The unstandardised contrasts for the two traits are defined as

$$\Delta X \stackrel{\text{def}}{=} \ln X_1(t) - \ln X_2(t), \quad (3)$$

$$\Delta Y \stackrel{\text{def}}{=} \ln Y_1(t) - \ln Y_2(t). \quad (4)$$

It is straightforward to show that these contrast variables will be normally distributed, with expected values of zero ($E[\Delta X] = E[\Delta Y] = 0$), and variances given by

$$\text{Var}[\Delta X] = 2 \text{Var}[\ln X_i(t)] = 2\sigma_X^2 t, \quad (5)$$

$$\text{Var}[\Delta Y] = 2 \text{Var}[\ln Y_i(t)] = 2\sigma_Y^2 t. \quad (6)$$

Because evolutionary changes were assumed to be independent on a logarithmic scale, the ancestral states, X_0 and Y_0 , appear in neither the mean nor the variance of the variables. Accordingly, a set of such contrasts, from non-overlapping lineages, will be statistically independent, and so expected to pass the diagnostic tests of Freckleton (2000) and Freckleton and Harvey (2006).

However, due to the gradual trait evolution, contrast variance will increase with the divergence time of the pair, and so be heterogeneous across the data set. But as this increase is linear, dividing each contrast by \sqrt{t} removes all time dependence. We will use an asterisk to denote the suitably standardised contrasts,

$$\Delta X^* \stackrel{\text{def}}{=} \Delta X / \sqrt{t}, \quad (7)$$

$$\Delta Y^* \stackrel{\text{def}}{=} \Delta Y / \sqrt{t}, \quad (8)$$

which have zero means, and constant variances. Accordingly, the magnitude of these standardised contrasts should show no correlation with \sqrt{t} (Garland et al., 1992).

The single most important quantity is the squared correlation of the contrasts, i.e., the quantity for which the r^2 of a linear regression is an estimator. This quantity

is simply

$$\text{Corr}^2(\Delta X^*, \Delta Y^*) = \rho^2 \quad (9)$$

and so a sample of contrasts may be used to estimate the strength of association between traits X and Y .

4. Substitution rate as the response variable

Let us now assume that $Y_i(t)$ is the rate of nucleotide substitution per site. In this case, $Y_i(t)$ must be estimated from substitution counts. In most real analyses, of course, the substitution count cannot be observed either, but must be estimated from the comparison of homologous DNA sequences. But as the problems associated with this inference have been comprehensively treated in the phylogenetics literature, they will be neglected here. Accordingly, let us treat as an observation the substitution count along the lineage leading to species i , and denote this quantity $N_i(t)$. The number of substitutions measured will depend not on the substitution rate at the single point in time when the measurement is made, but rather on the time average of the substitution rate during the complete period of divergence. In other words, $N_i(t)$ will be a function not of $Y_i(t)$, but of $\bar{Y}_i(t)$, its average value between times 0 and t :

$$\bar{Y}_i(t) \stackrel{\text{def}}{=} \frac{1}{t} \int_0^t Y_i(s) ds \quad (10)$$

(Fig. 1). Because substitution is a stochastic process, the actual count, $N_i(t)$, must be treated as a random variable even for a fixed value of $\bar{Y}_i(t)$. But the expected value of $N_i(t)$ follows by definition, as the product of the number of sites in the sequence, L , the time over which the substitutions occur, t , and the expected rate over that time period:

$$E[N_i(t)] = E[\bar{Y}_i(t)]Lt. \quad (11)$$

To obtain the complete distribution of $N_i(t)$, we assume that substitutions occur by a Poisson process (although relaxing this assumption should make little qualitative difference to the results).

Our substitution rate contrasts will be defined by analogy with Eq. (4):

$$\Delta N \stackrel{\text{def}}{=} \ln N_1 - \ln N_2 \quad (12)$$

(where we do not display the time-dependence of the N_i for brevity). Contrasts of this kind have been used in the literature (Bromham et al., 1996; Thomas et al., 2006; Fontanillas et al., 2007), but so have other variable types, such as the contrast in the untransformed branch lengths, standardised by their mean value:

$$\Delta N_{alt} \stackrel{\text{def}}{=} \frac{N_1 - N_2}{(N_1 + N_2)/2} \quad (13)$$

(see, e.g., Davies et al., 2004; Davies and Savolainen, 2006). (Also quite similar is the contrast variable of Wright et al. (2006) which can be written as $(N_1 - N_2)/\max(N_1, N_2)$.)

While Eqs. (12) and (13) look quite different, both can be written in terms of the branch length ratio, N_1/N_2 and are very close in value if this ratio is within a reasonable range:

$$\Delta N = \ln(N_1/N_2) \simeq \frac{2(N_1/N_2 - 1)}{N_1/N_2 + 1} = \Delta N_{alt}. \quad (14)$$

As such, results presented below should apply approximately to both variable types.

4.1. Substitution rates contrast properties

In Appendix A, we derive expressions for the variance of the substitution rate contrasts, and the squared correlation of the standardised trait and rate contrasts. Introducing the notation $\Delta \bar{Y} \stackrel{\text{def}}{=} \ln \bar{Y}_1(t) - \ln \bar{Y}_2(t)$ to denote the unmeasurable contrasts in time-averaged rate, these results can be written as

$$\text{Var}[\Delta N] \approx \text{Var}[\Delta \bar{Y}](1+f), \quad (15)$$

$$\text{Corr}^2(\Delta X^*, \Delta N^*) \approx \frac{\text{Corr}^2(\Delta X^*, \Delta \bar{Y}^*)}{(1+f)}, \quad (16)$$

where

$$f \approx E_{t,Y_0} \left[\frac{e^{\sigma_Y^2 t/12} - 3}{LY_0 t \sigma_Y^2 t} \right] \quad (17)$$

$$\approx E_{t,Y_0} \left[E \left[\frac{1}{N_i(t)} \right] \frac{2}{\text{Var}[\Delta \bar{Y}]} \right] \quad (18)$$

with $E_{t,Y_0}[\cdot]$ denoting the expectation over the full set of divergence times and ancestral rate values characterising the data set. Comparing Eqs. (15) and (16), with the equivalent expressions for measurable traits, Eqs. (6) and (9), indicates two differences, and these reflect the two major complications introduced by substitution rates. First, there is the new factor $(1+f)$, which appears because rate must be inferred from substitution counts which will fluctuate around their expected value. Second, there is the appearance of $\Delta \bar{Y}$, rather than ΔY , which reflects the fact that substitution counts depend on the average rate over the divergence (Fig. 1). We now consider these two complications in turn, and suggest ways of mitigating their effects.

4.2. Identifying inaccurately measured contrasts

The factor $(1+f)$ quantifies the uncertainty inherent in measuring rate changes. The extent of this uncertainty depends on whether sufficient substitutions have been observed given the size of the rate changes that have taken place (note, from Eq. (18), that large values of f result from short branches, i.e., large $E[1/N_i(t)]$, and/or small changes in rate, i.e., low $\text{Var}[\Delta \bar{Y}]$).

Eq. (16) demonstrates that such uncertainty can reduce the inferred r^2 , and thereby the power of the comparative analysis. Fig. 2 quantifies this reduction in r^2 , and shows

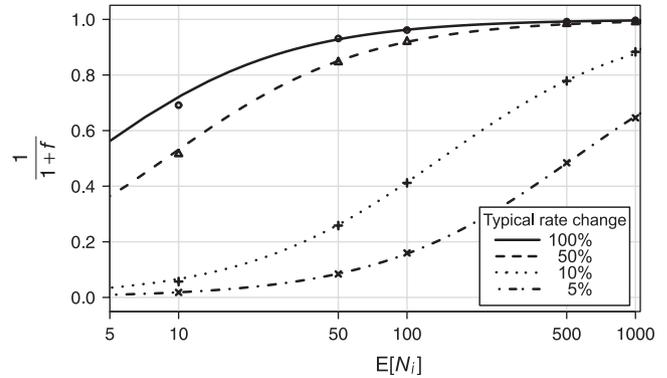


Fig. 2. The proportional reduction in the r^2 value attributable to stochastic fluctuations in the substitution rate (Eqs. (16) and (17)), is plotted against the expected number of substitutions (Eq. (11)), shown on a logarithmic scale. Curves are shown for four different values of the “typical” change in substitution rate. This typical value is defined in terms of the absolute rate contrast, such that a typical difference of 5% implies $E[|\Delta \bar{Y}|] = \ln 1.05$, a 10% difference implies $E[|\Delta \bar{Y}|] = \ln 1.10$, and so on.

This is related to contrast variance via $E[|\Delta \bar{Y}|] \approx \sqrt{2 \text{Var}[\Delta \bar{Y}]/\pi}$. The underlying model parameters, $LY_0 t$ and $\sigma_Y^2 t$, were derived from these values using Eqs. (16) and (17). Points were generated by simulating 1000 individual contrasts for each parameter combination, and then calculating the realised reduction in squared correlation: $\text{Corr}^2(\Delta X^*, \Delta \bar{N}^*)/\text{Corr}^2(\Delta X^*, \Delta Y^*)$.

that it can be drastic. For this reason, if we are to obtain an accurate estimate of the strength of association between rate and trait, we must select a set of contrasts for which f is negligible. Eq. (17) implies that this can be done by choosing deeper comparisons, and longer or faster portions of sequence. But what we really require is a method of determining whether f really is negligible for our chosen data set, i.e., whether our sequences are sufficiently long, or our comparisons sufficiently deep given the strictly unmeasurable rate changes under investigation.

A solution to this problem lies in the diagnostic tests for heterogeneity of variance proposed by Garland et al. (1992) and Freckleton (2000). To see the effects of f on contrast variance, consider the first factor of Eq. (15), which is well approximated by

$$\text{Var}[\Delta \bar{Y}] \simeq \frac{2\sigma_Y^2 t}{3}. \quad (19)$$

So if f is small, then contrast variance will increase roughly linearly with time, just as in the standard case, Eq. (6). However, when f is not small, there will be a strong opposing tendency for variance to decrease with time. Intuitively, this is because when branch lengths are small, as they will be for shallow comparisons, just one or two substitutions can make a large difference to the inferred rate, and this makes rate estimates highly variable. The practical consequence is that a data set characterised by substantial uncertainty should show a steep negative trend in the graphical test of Garland et al. (1992) (Fig. 3a). But this trend should affect only the youngest contrasts, after which the variance should level out (a failure for the

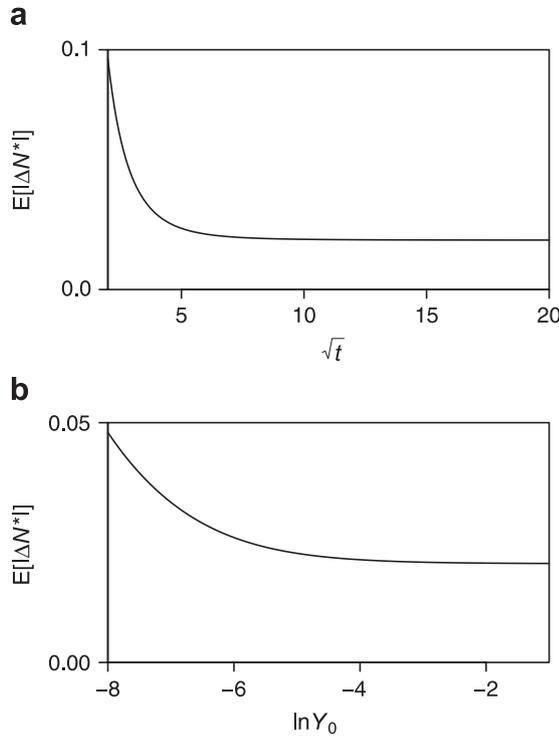


Fig. 3. The expected behaviour of diagnostic tests for heterogeneity of variance, when the method of independent contrasts is applied to substitution rates. (a) Depicts the test of Garland et al. (1992). (b) The test of Freckleton (2000) (in which the ancestral rate is estimated from the mean value for the pair). Parameters were set at $\sigma_Y^2 = 10^{-3}$ and $L = 3000$, with $Y_0 = 10^{-3}$ for (a) and $t = 15$ for (b).

variance to do so might indicate that such uncertainty affects all contrasts in the set).

As such, a method of meeting the assumptions of the regression, and increasing our chances of accurately estimating the true effect size, is to test for a negative trend as suggested by Garland et al. (1992), and if found, exclude the youngest contrasts until the trend disappears. Depending on the phylogeny, data from excluded contrasts could then be reused to create a smaller number of new, deeper contrasts. In this way, the test of Garland et al. (1992) can be used to establish a minimum contrast depth appropriate for the data under consideration.

The value of f might also be non-negligible if the data set contains comparisons with particularly slow ancestral substitution rates (Eq. (17)). These contrasts can be detected and excluded in a similar way using the test of Freckleton (2000), where again, they are expected to create a transient negative trend in the plots (Fig. 3b).

Recall that with standard traits, inhomogeneity of variance would be dealt with not by excluding points, but by transforming the data or standardisation factors. This will not always be an appropriate response for substitution rates. The reason is that substitution counts, unlike most standard traits, are the result of two stochastic processes taking place simultaneously, namely, rate evolution and the substitution process (this is apparent in Eq. (17) from the

two ‘time-scales’ defined by $\sigma_Y^2 t$ and $LY_0 t$). The result is that a contrast may have a high variance for two quite different reasons: because rate has evolved substantially, or because too few substitutions have occurred for the rate change to be reliably measured. Giving points of both kinds equal weight, or even including the unreliable points in the test, is likely to reduce power. This conclusion applies even if the majority of pairs are sufficiently deep, as is evident from Eq. (18), where the expectation of a reciprocal will be strongly influenced by the inclusion of a few small values.

While, in general, any reduction in sample size is undesirable, it is clear that maximising the number of data points by including many pairs with short molecular branch lengths, may reduce the power of the analysis overall.

4.3. Time averaging the predictor trait

If $f \ll 1$ for our contrasts, then the sole difference from a standard comparative analysis is that rate can be estimated only as a time average over the history of the lineage. This has little effect on the time-dependence of the contrast variance (Eqs. (6) and (19)). However, for the contrast correlation, we find

$$\text{Corr}^2(\Delta X^*, \Delta \bar{Y}^*) \approx (3/4)\rho^2 \quad (20)$$

(Appendix A) and so even if changes in substitution rate have been measured with high accuracy for all points, the r^2 will still tend to underestimate the true strength of association between rate and trait.

One possible solution to this problem is to inflate inferred r^2 values by 4/3, but this work-around is certain to increase false positive error for data deviating from the model assumptions. An alternative and preferable solution would be to obtain accurate estimates of the time-averaged value of the predictor variable,

$$\overline{\ln X}_i(t) \stackrel{\text{def}}{=} \frac{1}{t} \int_0^t \ln X_i(s) ds \quad (21)$$

and then derive contrasts from these estimates, rather than from the tip measurements:

$$\Delta \bar{X} \stackrel{\text{def}}{=} \overline{\ln X}_1(t) - \overline{\ln X}_2(t). \quad (22)$$

This is desirable because we find

$$\text{Corr}(\Delta \bar{X}^*, \Delta \bar{Y}^*) \approx \rho^2 \quad (23)$$

(Appendix A) and so if $\Delta \bar{X}$ could be accurately estimated for each of our contrasts, then the power of the test would approach that obtained with measurable traits (Eq. (9)).

Whether the strategy of estimating the time-averaged contrast $\Delta \bar{X}$ is worthwhile, depends on the quantity of data available. To see this, in Appendix B we derive a formal maximum likelihood estimator (MLE) of $\Delta \bar{X}$ using the evolutionary model of Eq. (1). We then show that if each estimate is obtained solely from a pair of tip measurements

then Eq. (20) still holds, and so the method leads to no increase in statistical power.

However, the method can increase power if additional data are available. The ideal data would be trait measurements from each of the pair's sister lineages, taken at intervals since their split. Measurements approximating this ideal scenario might be available in some cases, e.g., from the fossil record, and can be easily incorporated into the MLE. Less obviously, estimates of $\Delta\bar{X}$ can also be improved if contemporaneous measurements are available from multiple descendants of the pair's most recent common ancestor. Of course, in a standard comparative analysis, such measurements would be used to define new pairs, thereby increasing the number of data points. But as discussed above, when substitution rates are the subject of the analysis, it may be preferable to exclude many shallow comparisons, and replace them with a smaller number of deeper pairs. With the estimator introduced here, measurements from excluded pairs can be exploited to estimate $\Delta\bar{X}$ for the deeper pairs.

Using additional data to estimate a time-averaged contrast also has other benefits. For example, the method should increase power to detect correlated evolution if DNA sequence data are available only for clade members with recently evolved and anomalous trait values (perhaps domesticated species, or island endemics). The method can also be used if sequence data and trait measurements are available for different members of a clade, increasing the number of data points available.

We have made publicly available an implementation of the estimator of $\Delta\bar{X}$ in the R system (R core development team, 2006). In addition to the trait measurements, the software requires a tree with branch lengths, and this may contain non-contemporaneous tips, and unresolved polytomies (full details in Appendix B).

5. Worked example

To illustrate the points above, we now carry out a complete comparative analysis on a simulated data set of 50 phylogenetically independent species pairs. Each analysis included calculation of independent contrasts, diagnostic tests for heterogeneity of variance, and a linear regression forced through the origin.

To generate the data, trait values for each pair were simulated under the Brownian motion model of Eqs. (1) and (2). To parameterise the model, we assumed that traits X and Y evolve in a strongly correlated manner, with $\rho^2 = 0.25$. The age of the most recent common ancestor of each pair was then drawn at random from a uniform distribution: $\sqrt{t} \sim U[2, 6]$, and the other important parameters were set at $\sigma_Y^2 = 10^{-3}$ and $LY_0 = 3$ for all pairs (see Fig. 3a).

We first carried out a standard comparative analysis, assuming that both X and Y were directly measurable traits. Using the divergence times for each pair, we calculated 50 standardised contrasts (Eqs. (7) and (8)), for which the diagnostic tests of Garland et al. (1992) and

Table 1
Regression tests on simulated data

Regression	n	r^2	Adj. r^2	p -Value
(i) ΔY^* on ΔX^*	50	0.26	0.25	0.0001 ²
(ii) ΔN^* on ΔX^*	50	0.02	0.00	0.2819
(iii) ΔN on ΔX^*	50	0.03	0.01	0.2103
(iv) ΔN^* on ΔX^*	23	0.22	0.18	0.0208 ¹
(v) ΔN^* on $\Delta\bar{X}^*$	23	0.26	0.22	0.0119 ¹

Note: (i) Both variables measurable traits. (ii) Response variable is estimated substitution rate. (iii) Rate contrasts left unstandardised to homogenise variance (Fig. 4b). (iv) Shallow comparison pairs, with $\sqrt{t} < 4$, excluded from analysis to homogenise variance (Fig. 4a). (v) Predictor variable is time-averaged trait value estimated from eight contemporaneous tip measurements.

¹ $p < 0.05$.

² $p < 0.005$.

Freckleton (2000), indicated no heterogeneity of variance. A linear regression of ΔY^* on ΔX^* detected a highly significant relationship between the traits (see Table 1, line (i)), with the adjusted r^2 accurately estimating the true parameter value ($r^2 = \rho^2 = 0.25$).

Next, using the same simulated trait values, we assumed that Y was the rate of DNA substitution, and simulated molecular branch lengths for each of the 50 pairs (Eqs. (10) and (11)). We then calculated standardised rate contrasts from these substitution counts (Eq. (12)), and repeated the regression analysis. The regression now failed to detect any significant relationship between rate and trait (Table 1, line (ii)). However, as shown in Fig. 4a, the test of Garland et al. (1992) indicated significant heterogeneity of variance for the rate contrasts, which tended to decrease in magnitude with the depth of the pair (this was confirmed with a Kendall's rank correlation test: $p < 10^{-6}$).

If our response variable were a typical measurable trait, then an appropriate response to this heterogeneity would be to leave rate contrasts unstandardised (Harvey and Purvis, 1991; Garland et al., 1992). As shown in Fig. 4b, this strategy does remove the linear trend in the contrast magnitudes (rank correlation $p > 0.05$). However, regression using these unstandardised contrasts again fails to detect any relationship between rate and trait (Table 1, line (iii)). In fact, what Fig. 4b shows is not genuine homogeneity of variance, but rather a U-shaped pattern, with a decrease in variance for the shallower pairs balanced by an increase for the deeper pairs; and the negative trend for the shallower pairs tells us that rate changes cannot be reliably inferred from these data.

Accordingly, let us return to Fig. 4a. Visual inspection suggests that homogeneity of variance might be achieved by excluding contrasts with, say, $\sqrt{t} < 4$. Excluding these shallow points (shown as unfilled circles) again removes the significant correlation between ΔN^* and \sqrt{t} ($p > 0.05$), but also reduces sample size by more than half. Nevertheless, a regression test on this subset of points does successfully detect the strong association between rate and trait (Table 1,

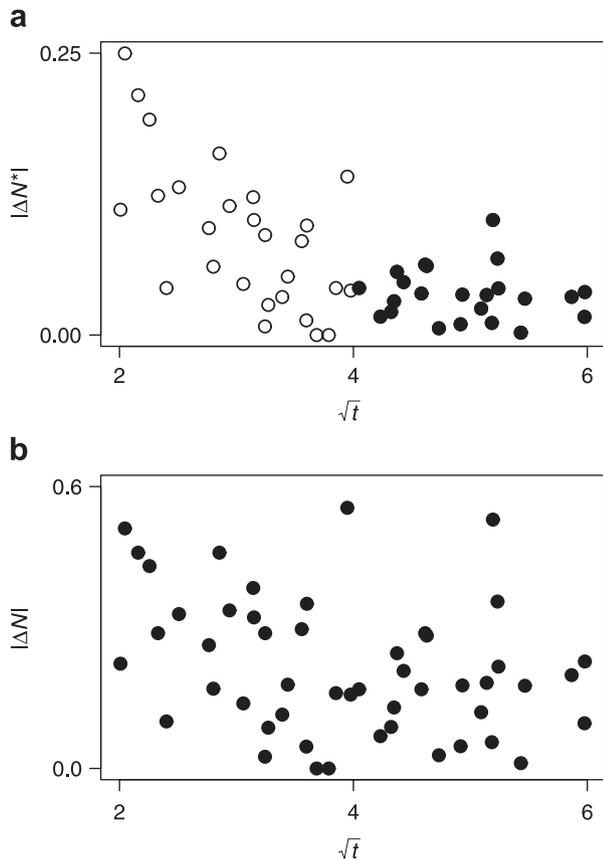


Fig. 4. The diagnostic test of Garland et al. (1992) applied to substitution rate contrasts generated by simulation. (a) Shows standardised contrasts, with shallow pairs (with $\sqrt{t} < 4$) shown as unfilled circles. Excluding these pairs removes the negative trend evident in the plot. (b) Shows unstandardised contrasts which also show no linear trend. For details see text, and Table 1.

line (iv)), a relationship which had been obscured by the inclusion of the shallow pairs (lines (ii) and (iii)).

Finally, to demonstrate the effects of time-averaging the predictor trait, we assumed that each comparison pair defined a clade of eight species for which contemporaneous measurements of trait X were available. Bifurcations in the subtrees defining each pair were assumed to occur at times $t/2$ and $t/4$, and the remaining six X values simulated accordingly. We then used the eight simulated values, combined with the dated phylogeny, to estimate $\Delta\bar{X}^*$ for each pair, using the computer software described in Appendix B. Carrying out a regression test using these time-averaged estimates in place of the standard comparisons of paired tip values, increased the power of the analysis further (Table 1, line (v)), and although the increase was not huge, the adjusted r^2 once more approached the true parameter value ($r^2 = 0.22 \approx \rho^2 = 0.25$).

6. Discussion

Several studies have tested the robustness of phylogenetic comparative methods to the inevitable violation of

their assumptions (Grafen, 1989; Purvis et al., 1994; Diaz-Uriarte and Garland, 1996, 1998). We have shown that when substitution rate is involved, there can be a substantial reduction in statistical power, even if the evolutionary model and dated phylogeny have been correctly specified.

One reason for this reduction in power is that substitution counts are a cumulative record of evolutionary change, and so reflect the time-averaged rate over the history of the lineage, rather than its current value (Fig. 1; Eq. (20)). To address this problem, we have introduced a method of using multiple tip measurements to estimate time-averaged trait values—quantities that are directly comparable to the substitution counts (Eq. (23)).

A second, and potentially more serious problem is that stochastic fluctuations in substitution number may prevent the accurate estimation of the underlying changes in rate (Eqs. (16)–(18); Fig. 2). The extent of this problem cannot be known in advance, because the number of substitutions needed to infer a rate change will depend on the size of the change in question. Our proposed solution to this problem is to use tests for inhomogeneous variance (Garland et al., 1992; Freckleton, 2000; Figs. 3 and 4) to establish a minimum contrast depth, and/or a minimum ancestral rate, that is appropriate for the lineages and sequence under consideration. The unreliable points can then be excluded from the analysis (although the constituent data may be used to construct deeper pairs, or to estimate the time-averaged trait value for an existing pair).

Table 1 shows that both procedures can increase power in some cases, but the arguments for excluding data are equivocal, because power is a function of both effect size and sample size. It is, however, well established that maximising the number of data points in a comparative analysis is not the same thing as maximising statistical power (Felsenstein, 1985; Barraclough et al., 1998). Furthermore, in one respect, a principled reduction in sample size is already part of standard practice in the comparative study of substitution rates—namely, the exclusion of contrasts between reconstructed states at internal nodes (Bromham et al., 1996; Barraclough and Savolainen, 2001; Davies et al., 2004; Thomas et al., 2006; Davies and Savolainen, 2006; Wright et al., 2006; Fontanillas et al., 2007). This certainly reduces sample size, because internodal contrasts allow $n - 1$ data points to be generated from n tips, rather than $n/2$ or fewer when sister pairs only are used (Felsenstein, 1985). But internodal contrasts for substitution rates are problematic. For example, for measurable traits, the weighted averages used to generate internodal contrasts need not represent reconstructions of ancestral states, and analyses can succeed even if these weighted averages deviate substantially and systematically from the true ancestral states (Grafen, 1989; Oakley and Cunningham, 2000). By contrast, the internal branch lengths of a molecular phylogeny are unavoidably reconstructions of ancestral states, and so, unlike tip measurements, should reflect

systematic changes in the early history of the clade. Other, equally serious problems are the node density artefact (Fitch and Beintema, 1990; Webster et al., 2003; Davies and Savolainen, 2006; Venditti et al., 2006), and the circularities arising when divergence dates are themselves estimated from molecular data (for example, the confounding of rate and date in molecular dating may lead to systematic errors of sign for internodal, or nested root-to-tip contrasts, but not for sister pairs, where the two lineages share the same period of divergence). For these reasons, comparative studies of substitution rates have generally excluded internodal contrasts. We have argued that an analogous trade-off between sample size and the reliability of each data point applies to sister pairs.

Results presented here might help to explain some contradictory results that have appeared recently in the literature. For example, Thomas et al. (2006) used a large data set of sister pair contrasts to argue that body mass had no effect on the rate of substitution across the invertebrates, while Fontanillas et al. (2007) used a smaller data set of deeper comparisons and found that such an effect was present. Methods introduced here should allow us to assess whether this discrepancy reflects something real about the biology (e.g., different evolutionary processes acting at different taxonomic levels), or simply a lack of power in the shallower comparisons (Fontanillas et al., 2007). In any case, in the light of the present results, it may be worth revisiting some other failures to reject null hypotheses that have appeared in the literature.

Acknowledgements

Many thanks to Lindell Bromham, Rob Lanfear, Jess Thomas, Eric Fontanillas, Andrew Rambaut and Fraser Lewis for helpful comments and discussions. J.J.W. is supported by BBSRC Grant DO17750 awarded to Andrew Rambaut.

Appendix A. Derivation of results in main text

Consider first the substitution rate contrast, Eq. (12), assuming that the substitution process is Poisson, but leaving unspecified the model of rate evolution. The moments of the random variable $\ln N_i$ (written without time dependence for brevity) are undefined, but we can consider a series expansion about the mean, \bar{N}_i :

$$\begin{aligned} \ln N_i &= \ln \bar{N}_i + \ln \left(1 + \frac{N_i - \bar{N}_i}{\bar{N}_i} \right) = \ln \bar{N}_i \\ &+ \frac{N_i - \bar{N}_i}{\bar{N}_i} - \frac{(N_i - \bar{N}_i)^2}{2\bar{N}_i^2} + \dots \end{aligned} \quad (\text{A.1})$$

from which we obtain approximations for the expected value of $\ln N_i$, and of its squared value, $\ln^2 N_i$, conditional on the time-averaged rate, \bar{Y}_i (Eq. (10)):

$$E[\ln N_i | \bar{Y}_i] \simeq \ln \bar{N}_i - \text{CV}^2(N_i)/2, \quad (\text{A.2})$$

$$E[\ln^2 N_i | \bar{Y}_i] \simeq \ln^2 \bar{N}_i + [1 - \ln(\bar{N}_i)]\text{CV}^2(N_i), \quad (\text{A.3})$$

where $\bar{N}_i \equiv E[N_i | \bar{Y}_i]$ and $\text{CV}^2(N_i) = \text{Var}[N_i | \bar{Y}_i] / \bar{N}_i^2$. These approximations will be accurate only if $\text{CV}^2(N_i)$ is small, which will hold if branches contain many substitutions. In terms of model parameters, it follows by definition that $E[N_i | \bar{Y}_i] = \bar{Y}_i L t$, while the value of $\text{CV}^2(N_i)$, follows from the Poisson assumption, which implies that $\text{Var}[N_i | \bar{Y}_i] = E[N_i | \bar{Y}_i]$, and so

$$\text{CV}^2(N_i) = \frac{1}{\bar{Y}_i L t}. \quad (\text{A.4})$$

Eq. (A.4) will underestimate the true quantity if substitution is overdispersed, in which case, from Eqs. (A.2) and (A.3), the influence of stochastic fluctuation in substitution number will also be underestimated.

Let us now define a random variable Z to capture the evolution of substitution rate:

$$Z \equiv \ln(\bar{Y}_i / Y_0). \quad (\text{A.5})$$

Unconditional forms of Eqs. (A.2) and (A.3) can be written in terms of this variable, and these can be used to derive the contrast variance and covariance:

$$\begin{aligned} \text{Var}[\Delta N] &= 2 \text{Var}[\ln N_i] = 2E[\ln^2 N_i] - 2E^2[\ln N_i] \\ &\simeq 2\text{Var}[Z] + 2 \frac{E[e^{-Z}(1 + E[Z] - Z)]}{LY_0 t} \\ &\quad - \frac{1}{2} \left(\frac{E[e^{-Z}]}{LY_0 t} \right)^2. \end{aligned} \quad (\text{A.6})$$

$$\begin{aligned} \text{Cov}[\Delta X, \Delta N] &= E[\Delta X \Delta N] = 2E[\ln X_i \ln N_i] \\ &\simeq 2E \left[\ln X_i \times \left(Z - \frac{e^{-Z}}{2LY_0 t} \right) \right]. \end{aligned} \quad (\text{A.7})$$

A.1. The Brownian motion model

The Wiener process, $W(t)$, represents the position of a particle undergoing Brownian motion along one dimension, starting at the origin at time $t = 0$. The model of Eqs. (1) and (2) contains independent copies of a Wiener process, $W_{X_i}(t)$ and $W_{Y_i}(t)$, labelled by a subscript to distinguish them. These processes are Gaussian random functions of time with the properties

$$E[W_{X_i}(t)] = E[W_{Y_j}(t)] = 0,$$

$$\begin{aligned} E[W_{X_i}(t_1) W_{X_j}(t_2)] &= E[W_{Y_i}(t_1) W_{Y_j}(t_2)] \\ &= \min(t_1, t_2) \delta_{ij}, \end{aligned}$$

$$E[W_{X_i}(t_1) W_{Y_j}(t_2)] = 0, \quad (\text{A.8})$$

where $\min(t_1, t_2)$ denotes the minimum of t_1 and t_2 and δ_{ij} denotes a Kronecker delta ($\delta_{ij} = 1$ if $i = j$ and is zero otherwise).

A.2. Combined results

Given the Brownian motion model, the variable Z from Eq. (A.5), is defined as

$$Z = \ln \left[t^{-1} \int_0^t ds \exp(\sigma_Y W(s)) \right], \quad (\text{A.9})$$

where

$$W(s) = \sqrt{1 - \rho^2} W_Y(s) + \rho W_X(s). \quad (\text{A.10})$$

The probability density of this variable has been derived by Comtet et al. (1998), but this is difficult to work with analytically, and even numerically (Ishiyama, 2005). However, we can make progress by assuming $\sigma_Y^2 t \ll 1$. Then, defining $\overline{W}^n = t^{-1} \int_0^t W^n(s) ds$, we have

$$\begin{aligned} Z &= \ln(1 + \sigma_Y \overline{W} + \sigma_Y^2 \overline{W}^2/2 + O(\sigma_Y^3 \overline{W}^3)) \\ &\simeq \sigma_Y \overline{W} + \sigma_Y^2 (\overline{W}^2 - \overline{W}^2)/2. \end{aligned} \quad (\text{A.11})$$

The density of this truncated variable is found to be approximately normal, meaning that we can use the following approximation:

$$\psi_Z(z) \simeq \frac{1}{\sqrt{2\pi(\sigma_Y^2 t/3)}} \exp\left(-\frac{(z - \sigma_Y^2 t/12)^2}{2(\sigma_Y^2 t/3)}\right), \quad \sigma_Y^2 t \ll 1. \quad (\text{A.12})$$

(Explicit corrections to all of the following results can also be found by retaining additional terms in the expansion Eq. (A.12).)

Using Eqs. (A.6) and (A.12), we now find

$$\begin{aligned} \text{Var}[\Delta N] &\simeq (2/3)\sigma_Y^2 t + \frac{2(1 + \sigma_Y^2 t/3)e^{\sigma_Y^2 t/12}}{LY_0 t} - \left(\frac{e^{\sigma_Y^2 t/12}}{LY_0 t}\right)^2 \\ &= (2/3)\sigma_Y^2 t \left[1 + \frac{e^{\sigma_Y^2 t/12}}{LY_0 t} (1 + 3/\sigma_Y^2 t) \right] + O((LY_0 t)^{-2}) \end{aligned} \quad (\text{A.13})$$

from which Eqs. (15), (17) and (19) follow directly. Similarly, from Eq. (A.7), we find

$$\begin{aligned} \text{Cov}[\Delta X, \Delta N] &\simeq 2E \left[\sigma_X W_X(t) \left(Z - \frac{1 - Z}{2LY_0 t} \right) \right] \\ &= 2\sigma_X \left(1 + \frac{1}{2LY_0 t} \right) E[W_X(t)Z] \\ &\simeq 2\sigma_X \sigma_Y \left(1 + \frac{1}{2LY_0 t} \right) E[W_X(t)\overline{W}] \\ &= \sigma_X \sigma_Y t \rho \left(1 + \frac{1}{2LY_0 t} \right). \end{aligned} \quad (\text{A.14})$$

Squaring this result, and excluding terms of order $(LY_0 t)^{-2}$, we can then use Eqs. (5) and (A.13) to yield:

$$\begin{aligned} \text{Corr}^2(\Delta X^*, \Delta N^*) &= \frac{1}{t} \frac{\text{Cov}^2[\Delta X, \Delta N]}{\text{Var}[\Delta X]\text{Var}[\Delta N]} \\ &\approx \frac{(3/4)\rho^2 \left(1 + \frac{1}{LY_0 t} \right)}{\left[1 + \frac{e^{\sigma_Y^2 t/12}}{LY_0 t} (1 + 3/\sigma_Y^2 t) \right]} \end{aligned} \quad (\text{A.15})$$

from which Eq. (16) follows as a further approximation.

A.3. Punctuational evolution

In addition to the true correlation, ρ , the results for Brownian motion evolution can be expressed in terms of just two compound parameters, $LY_0 t$ and $\sigma_Y^2 t$, which characterise the two stochastic processes of substitution and rate evolution. To understand the effects of the substitution process in isolation, it is useful to consider a model of punctuational evolution, in which rate evolution takes place solely at speciation events (e.g., Harvey and Purvis, 1991). This is equivalent to defining the following model of trait evolution in place of Eqs. (1) and (2):

$$\ln X_i = \ln X_0 + \sigma_X \xi_{X_i}, \quad (\text{A.16})$$

$$\ln Y_i = \ln Y_0 + \sigma_Y [\sqrt{1 - \rho^2} \xi_{Y_i} + \rho \xi_{X_i}], \quad (\text{A.17})$$

where ξ_{X_i} and ξ_{Y_i} are independent Gaussian random variables, with mean zero, and variance unity. Under this punctuational model, the variable Z , Eq. (A.5), is normally distributed, with mean zero and variance σ_Y^2 (i.e., with no dependence on time). Using this form of Z in Eqs. (A.6) and (A.7) shows that Eqs. (15), (16) and (18) still hold, but that the factor of 3/4 in Eq. (20) no longer applies. The continued dependence on time via $LY_0 t$, shows that for substitution rate contrasts, unless $f \approx 0$ holds for the data set, a punctuational pattern of evolution cannot be accommodated by setting all branch lengths in the phylogeny to a fixed value, as it can for measurable traits (Harvey and Purvis, 1991).

A.4. Time averaging both variables

Returning to the Brownian motion model, consider now the time average of $\ln X_i(t)$, Eq. (21). In terms of the Wiener process, this variable is

$$\overline{\ln X}_i \equiv t^{-1} \sigma_X \int_0^t W_X(s) ds \quad (\text{A.18})$$

which, with no approximation, is normally distributed, with mean zero and variance $\sigma_X^2 t/3$. Using Eq. (A.9), the covariance of this variable and the contrast in time-averaged rate is found to be

$$\begin{aligned} \text{Cov}[\Delta \overline{\ln X}^*, \Delta \ln \overline{Y}^*] \\ = E[\Delta \overline{\ln X}^* \Delta \ln \overline{Y}^*] \end{aligned}$$

$$\begin{aligned}
&= \frac{2\sigma_X}{t^2} \int_0^t ds E[W_X(s)Z] \\
&= \frac{2\sigma_X\sigma_Y\rho}{t^2} E \left[\frac{\int_0^t ds_1 \int_0^t ds_2 \min(s_1, s_2) e^{\sigma_Y W_1(s_2)}}{\int_0^t e^{\sigma_Y W_1(s_2)} ds_2} \right] \\
&= \frac{2\sigma_X\sigma_Y\rho}{t^2} E \left[\frac{\int_0^t ds_1 s_1 (t - s_1/2) e^{\sigma_Y W(s_1)}}{\int_0^t e^{\sigma_Y W(s_2)} ds_2} \right]. \quad (\text{A.19})
\end{aligned}$$

Calculating this expectation to leading order in σ_Y yields the approximation $\text{Cov}[\Delta \ln \bar{X}^*, \Delta \ln \bar{Y}^*] \approx 2\sigma_X\sigma_Y\rho/3$ from which Eq. (23) follows. This result approximates the exact covariance if the time averages of both variables are taken after their log transformation:

$$\begin{aligned}
&\text{Cov}[\Delta \ln \bar{X}^*, \Delta \ln \bar{Y}^*] \\
&= 2 \frac{\sigma_X\sigma_Y}{t^3} \int_0^t ds_1 \int_0^t ds_2 E[W_X(s_1)W(s_2)] \\
&= 4 \frac{\sigma_X\sigma_Y}{t^3} \rho \int_0^t ds_1 \int_0^{s_1} ds_2 s_2 \\
&= \frac{2\sigma_X\sigma_Y\rho}{3}. \quad (\text{A.20})
\end{aligned}$$

Appendix B. MLE of time-averaged trait values

Here, an MLE is derived for the difference in time-averaged trait values along a pair of lineages, where the time average is taken across the complete period of divergence since their most recent common ancestor (Eq. (22)). The estimator, like the method of independent contrasts itself, relies on the Brownian motion model of Eq. (1), and so for real data, an appropriate transformation of the trait measurements must be chosen, and the adequacy of the model tested as described in the main text.

In this appendix, for notational brevity, we will use x_i to denote the suitably transformed trait value at the time of measurement, such that $x_i \equiv \ln X_i(t)$ in the notation of the main text. We will also measure time in units of the expected variance of change in the traits (Garland et al., 1992) which is equivalent here to scaling time by σ_X^2 . Note that any scaling of time makes no difference to the key result of Eq. (B.12), and alters Eq. (B.14) by a multiplicative constant.

B.1. Likelihood surface of the subtree

Consider a pair of species which constitute a single phylogenetically independent comparison. The most recent common ancestor of this pair defines a clade containing a total of n tips, where $n \geq 2$. The estimator introduced here makes use of data solely from within this clade, that is, from the subtree whose root is the most recent common ancestor of the comparison species pair. The data required are measured trait values from the n tips, and topology and divergence dates for the subtree. The subtree can contain unresolved polytomies, and non-contemporaneous tips, by

which means measurements from putative ancestors of the extant taxa can be exploited in the estimation.

Let us use x_1, x_2, \dots, x_n to denote the transformed trait values for each of the tips, and collect these n values into an n -dimensional column vector.

$$\mathbf{x} = (x_1, x_2, \dots, x_n)^T, \quad (\text{B.1})$$

where the superscript T denotes matrix transpose. Given the model of trait evolution, the likelihood surface of these trait values across the tree, $l(\mathbf{x})$, is a multivariate normal (Felsenstein, 1973)

$$l(\mathbf{x}) \propto \exp\left(-\frac{1}{2}(\mathbf{x} - x_0\mathbf{1})^T \mathbf{T}^{-1}(\mathbf{x} - x_0\mathbf{1})\right), \quad (\text{B.2})$$

where x_0 is the generally unknown common ancestral state of the n tips (i.e., the transformed trait value at the root of the subtree), $\mathbf{1} = (1, 1, \dots, 1)^T$ is an n -dimensional column vector of ones, and \mathbf{T} is the variance–covariance matrix. The matrix entries, defined via $t_{ij} = E[(x_i - x_0)(x_j - x_0)]$, are simply the period of divergence common to species i and j , or, equivalently, the length of divergence from the root of the subtree to the most recent common ancestor of the two tips (Felsenstein, 1973). Using this likelihood surface, estimators of the ancestral state, x_0 , have been derived by Maddison (1991) and Schluter et al. (1997).

B.2. The time-averaged trait contrast

The quantity of interest here is the contrast in time-averaged trait values between the comparison pair. Labelling these taxa 1 and 2, this quantity can be written in terms of the entries of \mathbf{T} as follows:

$$\Delta \bar{x} \equiv t_{11}^{-1} \int_0^{t_{11}} x_1(s) ds - t_{22}^{-1} \int_0^{t_{22}} x_2(s) ds \quad (\text{B.3})$$

(see Eqs. (21) and (22)). It follows from Eq. (A.18), that the probability density of this variable is normal, with mean zero, and variance given by

$$\text{Var}[\Delta \bar{x}] \equiv \sigma_{\Delta \bar{x}}^2 = \frac{t_{11} + t_{22}}{3}. \quad (\text{B.4})$$

The likelihood surface of the tip values, conditional on the value of the time-averaged contrast, is then the multivariate normal

$$l(\mathbf{x}|\Delta \bar{x}) = \frac{\exp\left(-\frac{1}{2}[(\mathbf{x} - x_0\mathbf{1})^T, \Delta \bar{x}] \mathbf{V}^{-1}[(\mathbf{x} - x_0\mathbf{1})^T, \Delta \bar{x}]^T\right)}{\int dy^n \exp\left(-\frac{1}{2}[(\mathbf{y} - x_0\mathbf{1})^T, \Delta \bar{x}] \mathbf{V}^{-1}[(\mathbf{y} - x_0\mathbf{1})^T, \Delta \bar{x}]^T\right)}, \quad (\text{B.5})$$

where the variance–covariance matrix is

$$\mathbf{V} = \begin{pmatrix} \mathbf{T} & \mathbf{c} \\ \mathbf{c}^T & \sigma_{\Delta \bar{x}}^2 \end{pmatrix}. \quad (\text{B.6})$$

The n -dimensional column vector \mathbf{c} , contains the covariances between the tip values and the time-averaged contrasts. It can be shown that the elements of \mathbf{c} can be

written in terms of elements of the matrix \mathbf{T} as follows:

$$c_i \equiv E[(x_i - x_0)\Delta\bar{x}] = t_{i1} - \frac{t_{i1}^2}{2t_{11}} - t_{i2} + \frac{t_{i2}^2}{2t_{22}}. \quad (\text{B.7})$$

B.3. The estimator

Having specified all of the elements of the matrix \mathbf{V} , to derive an MLE of $\Delta\bar{x}$ we differentiate the log likelihood function, Eq. (B.5), by the unknown quantities, $\Delta\bar{x}$ and x_0 , then set to zero and solve for $\Delta\bar{x}$. If we denote the elements of the inverted matrix as

$$\mathbf{V}^{-1} = \begin{pmatrix} \mathbf{M} & \mathbf{n} \\ \mathbf{n}^T & k \end{pmatrix} \quad (\text{B.8})$$

then the equations to be solved are found to be

$$\frac{d \ln l(\mathbf{x}|\Delta\bar{x})}{d\Delta\bar{x}} = -\mathbf{1}^T \mathbf{M}(\mathbf{x} - x_0 \mathbf{1}) - \mathbf{1}^T \mathbf{n} \Delta\bar{x} = 0, \quad (\text{B.9})$$

$$\frac{d \ln l(\mathbf{x}|\Delta\bar{x})}{dx_0} = (\mathbf{x} - x_0 \mathbf{1})^T \mathbf{n} + \Delta\bar{x} \mathbf{n}^T \mathbf{M}^{-1} \mathbf{n} = 0, \quad (\text{B.10})$$

which have the solution

$$\Delta\bar{x} = \frac{[(\mathbf{1}^T \mathbf{n}) \mathbf{1}^T \mathbf{M} - (\mathbf{1}^T \mathbf{M} \mathbf{1}) \mathbf{n}^T] \mathbf{x}}{(\mathbf{1}^T \mathbf{M} \mathbf{1})(\mathbf{n}^T \mathbf{M}^{-1} \mathbf{n}) - (\mathbf{1}^T \mathbf{n})^2}. \quad (\text{B.11})$$

This gives us a MLE for $\Delta\bar{x}$ in terms of the n measured trait values collected in the vector \mathbf{x} , and an n -dimensional column vector.

$$\widehat{\Delta\bar{x}} \equiv \mathbf{b}^T \mathbf{x}. \quad (\text{B.12})$$

where the vector \mathbf{b} is defined via Eq. (B.11). Note that we have

$$\mathbf{b}^T \mathbf{1} = 0 \quad (\text{B.13})$$

and so shifting all x values by the same amount: $x_i \rightarrow x_i + \text{const}$, leaves $\widehat{\Delta\bar{x}}$ unchanged. This is a consequence of the contrast variable's independence of the ancestral trait value. A further consequence is that we can derive the expected variance of estimator by replacing \mathbf{x} in Eq. (B.12) by $\mathbf{x}' \equiv \mathbf{x} - E[\mathbf{x}]$, because $E[\mathbf{x}] = \text{const} \times \mathbf{1}$. We then find

$$\text{Var}[\widehat{\Delta\bar{x}}] = E[\widehat{\Delta\bar{x}}^2] = E[\mathbf{b}^T \mathbf{x}' \mathbf{x}'^T \mathbf{b}] = \mathbf{b}^T \mathbf{T} \mathbf{b} = \frac{\mathbf{n}^T \mathbf{T} \mathbf{n}}{(\mathbf{n}^T \mathbf{M}^{-1} \mathbf{n})^2}. \quad (\text{B.14})$$

As such, contrast variance can be standardised as follows:

$$\widehat{\Delta\bar{x}}^* = \frac{\mathbf{b}^T \mathbf{x}}{\sqrt{\mathbf{b}^T \mathbf{T} \mathbf{b}}}. \quad (\text{B.15})$$

B.4. Only two tip measurements available

If the only measurements available are for the pair themselves, i.e., if $n = 2$, then it follows from Eq. (B.13)

that the estimate of the time-averaged contrast will be equal to a constant multiplied by the traditional contrast in tip values, i.e., $\widehat{\Delta\bar{x}} \propto (x_1 - x_2) = \Delta x$ (the constant of proportionality is $2/3$ if both tips are of the same age, such that $t_{11} = t_{22}$). Furthermore, it follows from Eq. (B.15) that the standardised contrasts will be exactly equal in value whether or not the time-averaging estimator is used: $\widehat{\Delta\bar{x}}^* = \Delta x^*$. This establishes that the method outlined in this appendix is wholly redundant unless additional data are available for the clade defined by the pair, that is, unless $n > 2$.

B.5. Estimates when comparison pair lack measurements

The method above can also be used to derive estimators of $\Delta\bar{x}^*$ when one or both of the comparison species pair lack measurements, that is, if one or both of x_1 and x_2 are unknown. This is done by replacing Eq. (B.5) with a density function explicitly conditioned on the one or two missing measurements, and then taking derivatives with respect to the three or four unknown quantities, before solving for $\Delta\bar{x}$. The results, expressible via a partition of the matrix \mathbf{M} , are lengthy and so not given here. These estimators give good results in general, but the problem becomes ill-conditioned for clades with only a small number of measured tips, and so sensitivity to rounding error should be checked in such cases.

B.6. The software

The estimators described above have been implemented in freely-available software written for the R system (R core development team, 2006), making use of algorithms for processing phylogenetic trees contained in the APE package (Paradis et al., 2004). The software takes as input a phylogenetic tree with branch lengths, in Phylip format (also known as Newick format), and suitably transformed trait measurements for some, though not necessarily all, of the tips within the clade defined by the comparison pair specified. The software is available for download from <http://tree.bio.ed.ac.uk/>.

References

- Barraclough, T.G., Nee, S., Harvey, P.H., 1998. Sister group analysis in identifying correlates of diversification. *Evol. Ecol.* 12, 751–754.
- Barraclough, T.G., Savolainen, V., 2001. Evolutionary rate and species diversity in flowering plants. *Evolution* 55, 677–683.
- Bromham, L., 2003. Molecular clocks and explosive radiations. *J. Mol. Evol.* 57, S13–S20.
- Bromham, L., Rambaut, A., Harvey, P.H., 1996. Determinants of rate variation in mammalian DNA sequence evolution. *J. Mol. Evol.* 43, 610–621.
- Comtet, A., Monthus, C., Yor, M., 1998. Exponential functionals of Brownian motion and disordered systems. *J. Appl. Probab.* 35, 255.
- Davies, T.J., Savolainen, V., Chase, M.W., Moat, J., Barraclough, T.G., 2004. Environmental energy and evolutionary rates in flowering plants. *Proc. R. Soc. London B* 271, 2195–2200.

- Davies, T.J., Savolainen, V., 2006. Neutral theory, phylogenies, and the relationship between phenotypic change and evolutionary rates. *Evolution* 60, 476–483.
- Diaz-Uriarte, R., Garland Jr., T., 1996. Testing hypotheses of correlated evolution using phylogenetically independent contrasts: sensitivity to deviations from Brownian motion. *Syst. Biol.* 45, 27–47.
- Diaz-Uriarte, R., Garland Jr., T., 1998. Effects of branch length errors on the performance of phylogenetically independent contrasts. *Syst. Biol.* 47, 654–672.
- Felsenstein, J., 1973. Maximum-likelihood estimation of evolutionary trees from continuous characters. *Am. J. Hum. Genet.* 25, 471–492.
- Felsenstein, J., 1985. Phylogenies and the comparative method. *Am. Nat.* 125, 1–15.
- Fitch, W.M., Beintema, J.J., 1990. Correcting parsimonious trees for unseen nucleotide substitutions: the effect of dense branching as exemplified by ribonuclease. *Mol. Biol. Evol.* 7, 438–443.
- Fontanillas, E., Welch, J.J., Thomas, J.A., Bromham, L., 2007. The influence of body size and net diversification rate on molecular evolution during the radiation of animal phyla. *BMC Evol. Biol.* 7, 95 doi:10.1186/1471-2148-7-95.
- Freckleton, R.P., 2000. Phylogenetic tests of ecological and evolutionary hypotheses: checking for phylogenetic independence. *Funct. Ecol.* 14, 129–134.
- Freckleton, R.P., Harvey, P.H., 2006. Detecting non-Brownian trait evolution in adaptive radiations. *PLoS* 4, e373.
- Garland Jr., T., Harvey, P.H., Ives, A.R., 1992. Procedures for the analysis of comparative data using phylogenetically independent contrasts. *Syst. Biol.* 41, 18–32.
- Gillooly, J.F., Allen, A.P., West, G.B., Brown, J.H., 2005. The rate of DNA evolution: effects of body size and temperature on the molecular clock. *Proc. Natl Acad. Sci. USA* 102, 140–145.
- Grafen, A., 1989. The phylogenetic regression. *Philos. Trans. R. Soc. B* 326, 119–157.
- Harvey, P.H., Pagel, M.D., 1991. *The comparative method in evolutionary biology*. Oxford University Press, Oxford.
- Harvey, P.H., Purvis, A., 1991. Comparative methods for explaining adaptations. *Nature* 351, 619–624.
- Ishiyama, K., 2005. Methods for evaluating density functions of exponential functionals represented as integrals of geometric Brownian motion. *Methodol. Comput. Appl. Probab.* 7, 271–283.
- Maddison, W.P., 1991. Squared-change parsimony reconstructions of ancestral states for continuous-valued characters on a phylogenetic tree. *Syst. Zool.* 40, 304–314.
- Oakley, T.H., Cunningham, C.W., 2000. Independent contrasts succeed where ancestor reconstruction fails in a known bacteriophage phylogeny. *Evolution* 54, 397–405.
- Paradis, E., Claude, J., Strimmer, K., 2004. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* 20, 289–290.
- Price, T., 1997. Correlated evolution and independent contrasts. *Philos. Trans. R. Soc. B* 352, 519–529.
- Purvis, A., Gittleman, J.L., Luh, H.-K., 1994. Truth or consequences: effects of phylogenetic accuracy on two comparative methods. *J. Theor. Biol.* 167, 293–300.
- R core development team, 2006. R: a language and environment for statistical computing, V 2.4. See (<http://www.R-project.org/>).
- Schluter, D., Price, T., Mooers, A.O., Ludwig, D., 1997. Likelihood of ancestor states in adaptive radiation. *Evolution* 51, 1699–1711.
- Thomas, J.E., Welch, J.J., Woolfit, M., Bromham, L., 2006. There is no universal molecular clock for invertebrates, but rate variation does not scale with body size. *Proc. Natl Acad. Sci. USA* 103, 7366–7371.
- Venditti, C., Meade, A., Pagel, M., 2006. Detecting the node-density artefact in phylogeny reconstruction. *Syst. Biol.* 55, 637–643.
- Webster, A.J., Payne, R.J.H., Pagel, M., 2003. Molecular phylogenies link rates of evolution and speciation. *Science* 301, 478.
- Wright, S., Keeling, J., Gillman, L., 2006. The road from Santa Rosalia: a faster tempo of evolution in tropical climates. *Proc. Natl Acad. Sci. USA* 103, 7718–7722.