Testing for correlation between traits under directional evolution

Manuela Royer-Carenzi[1] and Gilles Didier[2]

[1] Aix Marseille Univ, CNRS, Centrale Marseille, I2M, Marseille, France
[2] IMAG, Univ Montpellier, CNRS, Montpellier, France

May 24, 2019

Abstract

Being confounding factors, directional trends are likely to make two quantitative traits appear as spuriously correlated. By determining the probability distributions of independent contrasts when traits evolve following Brownian motions with linear trends, we show that the standard independent contrasts can not be used to test for correlation in this situation. We propose a multiple regression approach which corrects the bias caused by directional evolution.

We show that our approach is equivalent to performing a Phylogenetic Generalized Least Squares (PGLS) analysis with tip times as covariables by providing a new and more general proof of the equivalence between PGLS and independent contrasts methods.

Our approach is assessed and compared with three previous correlation tests on data simulated in various situations and overall outperforms all the other methods. The approach is next illustrated on a real dataset to test for correlation between hominin cranial capacity and body mass.

Keywords: directional evolution; correlation tests; Brownian motion; independent contrasts.

1 Introduction

Testing for correlation between two traits is a natural question which has been widely studied, notably in a comparative biology context (Groussin and Gouy 2011; Marchini et al. 2014; Grabowski et al. 2015; Will et al. 2017; Zhao et al. 2017). Correlation tests may concern any kind of traits: phenotypic, genetic or other. For instance, Seligmann (2018, 2019) observed a correlation between the syntheny level of poxviruses with amoeban mitogenome and their genome size. Assessing the correlation between two traits measured on several species cannot be performed by directly computing the Pearson correlation coefficient on the traits values since these values are not independent but related through the evolutionary relationships of the species involved (Diaz-Uriarte and Garland 1996). This point raises questions about how to interpret a correlation between two traits in a phylogenetic context. Actually, since all the observed taxa are assumed to have evolved from a single ancestral taxa, stating that two traits shared by the taxa are correlated can have only one significance, which is that the respective evolutions of these traits are correlated one with the other (Harvey and Pagel 1991). Assumptions about traits evolution are thus essential in order to disentangle the dependency structure of their extant values, and eventually to be able to study their correlation by correcting biases due to their evolutionary relationships (Martins and Garland 1991; Garland and Adolph 1994; Diaz-Uriarte and Garland 1996; Martins 1996; Oakley and Cunningham 2000).

A widely used approach for testing correlation between traits on a phylogenetic tree is the independent contrasts method of Felsenstein (1985) which extracts independent quantities from the tip values of the traits in order to estimate their correlation. The rationale behind this approach is that if two traits follow two correlated Brownian motions then their matched independent contrasts are realizations of independent and identically distributed pairs of Gaussian random variables correlated with the same correlation as the Brownian motions. Therefore, testing for correlation through independent contrasts is perfectly founded if one assumes that the two traits to compare follow Brownian motions along the phylogenetic tree, thus a neutral evolution for both of them (Felsenstein 1988). A strongly related method, called Phylogenetic Generalized Least Squares (PGLS, Grafen 1989) addresses the same question in an equivalent way. We extend and give formal proofs of results about the equivalence between

independent contrasts and PGLS methods for regression analysis stated in Garland and Ives (2000); Rohlf (2001); Blomberg et al. (2012).

Both independent contrasts and PGLS approaches make the strong assumption that evolution of the considered traits is neutral. There are many situations where this assumption is not granted, for instance when evolution is driven by ecological pressures. Arnold and Moncrieff (1994) showed that several lizard species developed the same discrete traits in approximately the same order when adapting to a same environmental condition. In the same way, quantitative traits may evolves toward optima or following a general trend. This type of non-neutral evolution, referred to as adaptative and directional evolution, cannot be modeled by using standard Brownian processes. Adaptation, i.e., evolution toward optima, is generally modeled with Ornstein-Uhlenbeck processes or more complex models (Cressler et al. 2015). For instance, Hansen et al. (2008) consider traits evolving according Ornstein-Uhlenbeck processes with optima which themselves evolve following a Brownian process. Directional evolution is generally modeled with arithmetic Brownian motions, i.e., Brownian process with linear trends. Note that after a change in the optimum during the adaptation process, traits under selection may look under directional evolution until getting close to their new optima. In this work, we focus on the directional evolution case. More precisely, we address the question of how to detect correlation between two traits when at least one of them is under directional evolution. Among the numerous examples of evolutionary tendencies in the evolution of phenotypic or genetic traits, Cope's rule, which predicts that the body size of species tends to increase over evolutionary time, has gained considerable empirical support (Kingsolver and Pfennig 2004; Van Valkenburgh et al. 2004; Hone and Benton 2005; Hone et al. 2005; Hunt and Roy 2006; Bokma et al. 2016). Beside the Cope's rule, there are strong evidence that traits of some clades have evolved following trends at certain periods. For instance, there are evidence of increase in the Mysticetes body size (Slater et al. 2010) and in the microsatellite size in Maize (Vigouroux et al. 2003). Last, insular dwarfism and gigantism phenomena provide numerous examples of directional evolution (Lomolino 2005). We emphasize that an evolutionary tendency in increasing or decreasing the body size of species implies that most of their morphological measures follow the same trend, and would be systematically tested as significantly correlated (Yule 1926; Entorf 1997; Deng 2015).

In order to study the correlation between traits under directional evolution, we first determine the probability distribution of the independent contrasts of a trait which evolves following a Brownian motion with a linear trend. The form of these distributions shows that testing for correlation between two traits under directional evolution through independent contrasts makes no sense. We propose an alternative approach based on a multiple regression which includes time as explanatory variable in order to correct the bias due to a linear trend. We show that our approach is equivalent to performing a PGLS analysis by adding the tip times as covariable in the regression.

A previous approach to correct the trend effect on the independent contrasts and to test for correlation between traits under directional evolution was proposed in Elliott (2015). Its general idea is to "center" the independent contrasts with regard to the trend (this method is detailed below). A thorough study of the Elliott's (2015) approach shows that it does not satisfy the regression assumptions. Note that if the phylogenetic tree supporting the evolution of the traits is ultrametric, our new method and that of Elliott (2015) are both equivalent to the independent contrasts method.

We simulated evolution of correlated and uncorrelated quantitative traits with and without trend on hominin phylogenetic tree in order to assess and to compare our new correlation test and three previous ones, namely the standard correlation test on the tips values of the traits, the correlation between independent contrasts of Felsenstein (1985) and that between the directional contrasts proposed by Elliott (2015). Simulation results shows that the new test is the most accurate as soon as one of the traits is under directional evolution. Despite its statistical flaws, the approach suggested by Elliott (2015) performs almost as well as our approach.

Last, the approach was applied on a real dataset in order to test for correlation between the logarithms of hominin body mass and cranial capacity, among which the logarithm of cranial capacity shows a significant positive trend. Our test concluded to a significant correlation between the logarithms of these two traits.

R-scripts implementing the correlation tests and the simulations performed for this work are available at `https://github.com/gilles-didier/Correlation`.

The rest of the paper is organized as follows. The independent contrasts and their distributions when traits follow Brownian motions with linear trends are presented in Section 2. Section 3 recalls three previous correlation tests on phylogenetic data and introduces a new one based on a multiple regression

between independent contrasts which includes time as explanatory variable. The four correlation tests are assessed and compared on simulated data in Section 4. Last, our test is applied to check correlation between hominin cranial capacity and body mass in Section 5.

2 Independent contrasts

2.1 Phylogenetic trees - Notations

In all what follows, we assume that the evolutionary history of the species is known and given as a rooted binary phylogenetic tree $\mathcal{T}$ with branch lengths. Our typical tree $\mathcal{T}$ contains $2n+1$ nodes among which n are internal nodes. According to the Felsenstein's (1985) convention, the nodes are indexed in the following way:

- index 0 for the root,
- indices 1 to $n-1$ for the other internal nodes,
- indices n to $2n$ for the tips.

For all nodes k, we put

- $r(k)$ and $\ell(k)$ for the two direct descendants of k, if k is an internal node,
- $a(k)$ for the direct ancestor of k, if k is not the root,
- v_k for the length of the branch ending at k,
- t_k for the (absolute) time of k, which is the sum of the branch lengths of the path relying the root to k (both included), and
- z_k for the value of the trait at node k, which is defined only if k is a tip.

2.2 Felsenstein's (1973) algorithm

Independent contrast were introduced in Felsenstein (1973) in order to compute the probability of the tip values of a quantitative trait evolving on a phylogenetic tree under the assumption that this trait follows a Brownian motion. This method extracts a series of realizations of independent Gaussian variables from the tip values by iteratively considering differences between terminal sister taxa and by replacing them with a single terminal taxa, while modifying the length of the branch that it ends and associating it with an artificial trait value computed from those of the two terminal sister taxa. Namely, the method recursively computes a new branch length $\overline{\mathrm{v}}_k$ and an "artificial" trait value $\overline{\mathrm{z}}_k$ for all nodes k of $\mathcal{T}$ in the following way.

Definition 1 (Felsenstein 1973)**.** *Let $\mathcal{T}$ be a phylogeny and $(z_k)_{n\leq k\leq 2n}$ be the values of a quantitative trait only known at the tips of $\mathcal{T}$. Under the notations of Section 2.1 and for all nodes k of $\mathcal{T}$, the quantities $\overline{\mathrm{v}}_k$, $\overline{\mathrm{z}}_k$ are recursively defined as*

$$\overline{\mathrm{v}}_k = \begin{cases} v_k & \text{if } k \text{ is a tip,} \\ v_k + \dfrac{\overline{\mathrm{v}}_{r(k)}\overline{\mathrm{v}}_{\ell(k)}}{\overline{\mathrm{v}}_{r(k)} + \overline{\mathrm{v}}_{\ell(k)}} & \text{otherwise, and} \end{cases}$$

$$\overline{\mathrm{z}}_k = \begin{cases} z_k & \text{if } k \text{ is a tip,} \\ \dfrac{\overline{\mathrm{v}}_{\ell(k)}\overline{\mathrm{z}}_{r(k)} + \overline{\mathrm{v}}_{r(k)}\overline{\mathrm{z}}_{\ell(k)}}{\overline{\mathrm{v}}_{r(k)} + \overline{\mathrm{v}}_{\ell(k)}} & \text{otherwise.} \end{cases}$$

For all internal nodes k of $\mathcal{T}$, the (standardized) independent contrast u_k is then defined as

$$u_k = \frac{\overline{\mathrm{z}}_{r(k)} - \overline{\mathrm{z}}_{\ell(k)}}{\sqrt{\overline{\mathrm{v}}_{r(k)} + \overline{\mathrm{v}}_{\ell(k)}}}.$$

Contrasts can be computed thanks to the `pic` function in the `ape` R-package (Paradis et al. 2004).

2.3 Independent contrasts as random variables

By construction, under an evolutionary model for quantitative traits, the tips values $(z_k)_{n\leq k\leq 2n}$ are realizations of random variables $Z_n, \ldots, Z_{2n}$. In this context, the artificial trait values $(\overline{z}_k)_{0\leq k\leq 2n}$ computed during the independent contrasts procedure are also realizations of random variables which can be expressed from $Z_n, \ldots, Z_{2n}$ and so are the independent contrasts.

Definition 2. *Let $(Z_k)_{n\leq k\leq 2n}$ be the random variables associated to the tip values of a trait. For all nodes k of $\mathcal{T}$, the random variable $\overline{\mathrm{Z}}_k$ is defined as*

$$\overline{\mathrm{Z}}_k = \begin{cases} Z_k & \text{if } k \text{ is a tip,} \\ \dfrac{\overline{\mathrm{v}}_{\ell(k)}\overline{\mathrm{Z}}_{r(k)} + \overline{\mathrm{v}}_{r(k)}\overline{\mathrm{Z}}_{\ell(k)}}{\overline{\mathrm{v}}_{r(k)} + \overline{\mathrm{v}}_{\ell(k)}} & \text{otherwise.} \end{cases}$$

And for all internal nodes k of $\mathcal{T}$, the random variable U_k is defined as

$$U_k = \frac{\overline{\mathrm{Z}}_{r(k)} - \overline{\mathrm{Z}}_{\ell(k)}}{\sqrt{\overline{\mathrm{v}}_{r(k)} + \overline{\mathrm{v}}_{\ell(k)}}}.$$

In plain English, for all nodes k, the random variable $\overline{\mathrm{Z}}_k$ corresponds to the value associated to k by the computation presented in Definition 1. In the same way, for all internal nodes k, the contrast u_k is a realization of the random variable U_k.

As evolutionary models of traits, we shall consider below either the Brownian Motion (BM) model or the Brownian Motion with linear trend, also known as the Arithmetic Brownian Motion (ABM) model. Namely, the ABM model with parameters (x_0, μ, σ^2), i.e., initial value x_0, trend μ and variance σ^2, is the process $(X_t)_{t>0}$ defined as

$$\mathrm{d}X_t = \mu \mathrm{d}t + \sigma \mathrm{d}W_t \qquad \text{and} \qquad X_0 = x_0,$$

where $(W_t)_{t>0}$ is the Wiener process (Grimmett and Stirzaker 2001). For all times t and s, the increments $X_{t+s} - X_t$ are independent Gaussian random variables with law $\mathcal{N}\left(\mu s, \sigma^2 s\right)$.

Since a BM model is nothing but an ABM model with trend parameter $\mu = 0$, we shall write results and properties in the ABM case only. Basically, any property or result granted for ABM also holds for BM.

The ABM process with parameters (x_0, μ, σ^2) running on the phylogenetic tree $\mathcal{T}$ starts at the root of $\mathcal{T}$ with the value x_0, then evolves independently on each branch of $\mathcal{T}$ by splitting at each internal node into two independent and identical processes both starting from the value of the process at this node and eventually ends at the tips of $\mathcal{T}$. It allows to model traits under directional evolution, e.g., following the Cope's rule (Kingsolver and Pfennig 2004; Van Valkenburgh et al. 2004; Hone and Benton 2005).

Theorem 3. *Let $(Z_k)_{n\leq k\leq 2n}$ be the random variables associated to the tip values of a trait following the ABM model with parameters (x_0, μ, σ^2) on $\mathcal{T}$. For all nodes k of $\mathcal{T}$, the random variable $\overline{\mathrm{Z}}_k$ can be written as*

$$\overline{\mathrm{Z}}_k = x_0 + \mu(t_k + \gamma_k) + \sigma\left(W_{t_k} + \Psi_k(W)\right),$$

where

$$\gamma_k = \begin{cases} 0 & \text{if } k \text{ is a tip,} \\ \dfrac{\overline{\mathrm{v}}_{\ell(k)}(\gamma_{r(k)} + v_{r(k)}) + \overline{\mathrm{v}}_{r(k)}(\gamma_{\ell(k)} + v_{\ell(k)})}{\overline{\mathrm{v}}_{r(k)} + \overline{\mathrm{v}}_{\ell(k)}} & \text{otherwise, and} \end{cases}$$

and $\Psi_k(W)$ is a linear combination of increments of the form $W_{t_i} - W_{t_{a(i)}}$, recursively defined as

$$\Psi_k(W) = \begin{cases} 0 & \text{if } k \text{ is a tip,} \\ \dfrac{\overline{\mathrm{v}}_{\ell(k)}\left(W_{t_{r(k)}} - W_{t_k} + \Psi_{r(k)}(W)\right) + \overline{\mathrm{v}}_{r(k)}\left(W_{t_{\ell(k)}} - W_{t_k} + \Psi_{\ell(k)}(W)\right)}{\overline{\mathrm{v}}_{r(k)} + \overline{\mathrm{v}}_{\ell(k)}} & \text{otherwise.} \end{cases}$$

For all internal nodes k of $\mathcal{T}$, the random variable U_k can be written as

$$U_k = \mu h_k + \frac{\sigma}{\sqrt{\overline{\mathrm{v}}_{r(k)} + \overline{\mathrm{v}}_{\ell(k)}}}\Phi_k(W), \tag{1}$$

where

$$h_k = \begin{cases} 0 & \text{if } k \text{ is a tip,} \\ \dfrac{v_{r(k)} + \gamma_{r(k)} - v_{\ell(k)} - \gamma_{\ell(k)}}{\sqrt{\overline{\mathrm{v}}_{r(k)} + \overline{\mathrm{v}}_{\ell(k)}}} & \text{otherwise, and} \end{cases} \tag{2}$$

$$\Phi_k(W) = (W_{t_{r(k)}} - W_{t_k}) + \Psi_{r(k)}(W) - (W_{t_{\ell(k)}} - W_{t_k}) - \Psi_{\ell(k)}(W).$$

Random variables $\Psi_k(W)$ and $\Phi_k(W)$ have Gaussian distributions $\mathcal{N}\left(0, \frac{\overline{\mathrm{v}}_{r(k)}\overline{\mathrm{v}}_{\ell(k)}}{\overline{\mathrm{v}}_{r(k)}+\overline{\mathrm{v}}_{\ell(k)}}\right)$ and $\mathcal{N}\left(0, \overline{\mathrm{v}}_{r(k)} + \overline{\mathrm{v}}_{\ell(k)}\right)$ respectively.

Proof. Appendix A. □

Corollary 4. *Under the ABM model with parameters (x_0, μ, σ^2), for all internal nodes k of $\mathcal{T}$, the independent contrast random variables U_k are independent from one another and Gaussian distributed with*

$$U_k \sim \mathcal{N}\left(\mu h_k, \sigma^2\right).$$

Proof. The proof of the independence of the independent contrasts under the ABM model follows the same arguments as the proof of the same property in the BM case provided in Felsenstein (1973). The form of the distribution of the independent contrasts is a direct consequence of Theorem 3. □

Proposition 5. *Let $\mathcal{T}$ be an ultrametric tree and T be the length of the path from the root to the tips. For all nodes k of $\mathcal{T}$, we have that $\gamma_k = T - t_k$ and $h_k = 0$.*

Proof. Appendix B. □

2.4 Correlation and independent contrasts

Independent contrasts are mainly used in order to test for correlation between two quantitative traits known only at the tips of a phylogenetic tree (Felsenstein 1985). The rationale behind this approach is that if two quantitative traits follow two correlated Brownian motions then their standardized independent contrasts are realizations of independent and identically distributed pairs of Gaussian random variables with the same correlation as the correlated Brownian motions. The transpose of a matrix or a vector D is noted D'.

Definition 6. *Let $(X_t^A)_{t\geq 0}$ and $(X_t^B)_{t\geq 0}$ two ABM models with parameters $(x_0^A, \mu_A, \sigma_A^2)$ and $(x_0^B, \mu_B, \sigma_B^2)$ respectively and let $(W_t^A)_{t\geq 0}$ and $(W_t^B)_{t\geq 0}$ be two Wiener processes such that*

$$X_t^A = x_0^A + \mu_A t + \sigma_A W_t^A \qquad \text{and} \qquad X_t^B = x_0^B + \mu_B t + \sigma_B W_t^B.$$

The processes $(X_t^A)_{t\geq 0}$ and $(X_t^B)_{t\geq 0}$ are correlated with correlation coefficient ρ if for all $t \geq 0$, the random vector $(W_t^A, W_t^B)'$ has covariance matrix $t\Sigma$ where

$$\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}.$$

Any pair of Wiener processes $(W_t^A, W_t^B)_{t\geq 0}$ with correlation matrix Σ can be obtained from two independent Wiener processes $(\widetilde{W}_t^A)_{t\geq 0}$ and $(\widetilde{W}_t^B)_{t\geq 0}$ by applying the Cholesky decomposition on their covariance matrix Σ (Gupta 2013). Namely, by decomposing Σ as $\Sigma = L\,L'$ where

$$L = \begin{pmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{pmatrix},$$

the random vector $(W_t^A, W_t^B)' = L(\widetilde{W}_t^A, \widetilde{W}_t^B)'$ has covariance matrix Σ.

In sum, two ABM processes $(X_t^A)_{t\geq 0}$ and $(X_t^B)_{t\geq 0}$ with parameters (x_0^A, μ_A, σ_A) and (x_0^B, μ_B, σ_B) are correlated with correlation coefficient ρ if and only if they can be written as

$$X_t^A = x_0^A + \mu_A t + \sigma_A \widetilde{W}_t^A \qquad \text{and} \qquad X_t^B = x_0^B + \mu_B t + \sigma_B(\rho\widetilde{W}_t^A + \sqrt{1-\rho^2}\widetilde{W}_t^B),$$

where $(\widetilde{W}_t^A)_{t\geq 0}$ and $(\widetilde{W}_t^B)_{t\geq 0}$ are two independent Wiener processes.

Theorem 7. *Let $(X_t^A)_{t\geq 0}$ and $(X_t^B)_{t\geq 0}$ be two ABM processes running on $\mathcal{T}$ with parameters $(x_0^A, \mu_A, \sigma_A^2)$ and $(x_0^B, \mu_B, \sigma_B^2)$ respectively and $(Z_k^A)_{n\leq k\leq 2n}$ and $(Z_k^B)_{n\leq k\leq 2n}$ be the random variables associated to the values of these processes at the tips of $\mathcal{T}$. The processes $(X_t^A)_{t\geq 0}$ and $(X_t^B)_{t\geq 0}$ are correlated with correlation coefficient ρ if and only if for all internal nodes k of $\mathcal{T}$, the independent contrast random variables U_k^A and U_k^B computed from $(Z_k^A)_{n\leq k\leq 2n}$ and $(Z_k^B)_{n\leq k\leq 2n}$ are correlated with correlation coefficient ρ.*

Proof. Let us assume that the ABM processes $(X_t^A)_{t\geq 0}$ and $(X_t^B)_{t\geq 0}$ of parameters $(x_0^A, \mu_A, \sigma_A^2)$ and $(x_0^B, \mu_B, \sigma_B^2)$ respectively are correlated with correlation coefficient ρ. It is equivalent to say that there exist two independent Wiener processes $(\widetilde{W}_t^A)$ and $(\widetilde{W}_t^B)$ such that

$$X_t^A = x_0^A + \mu_A t + \sigma_A \widetilde{W}_t^A \qquad \text{and} \qquad X_t^B = x_0^B + \mu_B t + \sigma_B(\rho \widetilde{W}_t^A + \sqrt{1-\rho^2}\widetilde{W}_t^B).$$

By applying Equation 1 of Theorem 3 to X_t^A and X_t^B, we then have that for all internal nodes k, the independent contrast random variables U_k^A and U_k^B can be written as

$$U_k^A = \mu_A h_k + \frac{\sigma_A}{\sqrt{\overline{\mathrm{v}}_{r(k)} + \overline{\mathrm{v}}_{\ell(k)}}} \Phi_k(\widetilde{W}^A), \quad \text{and} \tag{3}$$

$$U_k^B = \mu_B h_k + \frac{\sigma_B \rho}{\sqrt{\overline{\mathrm{v}}_{r(k)} + \overline{\mathrm{v}}_{\ell(k)}}} \Phi_k(\widetilde{W}^A) + \frac{\sigma_B\sqrt{1-\rho^2}}{\sqrt{\overline{\mathrm{v}}_{r(k)} + \overline{\mathrm{v}}_{\ell(k)}}} \Phi_k(\widetilde{W}^B). \tag{4}$$

Since Theorem 3 ensures that $\mathrm{Var}\left(\Phi_k(\widetilde{W}^A)\right) = \overline{\mathrm{v}}_{r(k)} + \overline{\mathrm{v}}_{\ell(k)}$, it follows that

$$\mathrm{Cov}\left(U_k^A, U_k^B\right) \quad = \quad \frac{\sigma_A\sigma_B\rho}{\overline{\mathrm{v}}_{r(k)} + \overline{\mathrm{v}}_{\ell(k)}} \times \mathrm{Var}\left(\Phi_k(\widetilde{W}^A)\right) = \sigma_A\sigma_B\rho,$$

thus $\mathrm{Cor}\left(U_k^A, U_k^B\right) = \rho$ for all internal nodes k of $\mathcal{T}$, which ends the proof. □

3 Correlation tests

In practical situations, evaluating or testing for correlation between two random variables is performed by considering a series of a certain number n of independent realizations of this pair of random variables. Let us remark that sampling independently n times a pair of random variables is equivalent to draw a joint sample of n independent and identically distributed pairs of random variables with the same correlation between the random variables of all the pairs. Standard regression analysis can thus be applied in this last case.

We emphasize the fact that performing a correlation test on a joint sample of n pairs of random variables which are not identically distributed makes absolutely no sense, even if each pair has the same correlation. Moreover, if we do not have independence between the pairs of random variables, then assumptions required by correlation tests are violated.

Linear regression is a usual tool for studying the association between two variables. Statistic analysis of regression requires additional assumptions, which are referred to as the *key assumptions* in Fox (2015), namely the constancy of the error variance, the Gaussianity of the errors, their null mean, and their independence. In the multiple linear regression, testing for correlation between the response variable and one of the regressors is performed by testing the nullity of the corresponding regression coefficient. Under the key assumptions, this test of nullity is based on the fact that, by assuming that the coefficient is null, the ratio of the ordinary least squares estimate of this coefficient to its standard deviation follows a Student distribution with a number of degrees of freedom equal to the difference between the number of samples and the number of regressors (including the intercept if there is one, Fox 2015). In the case of the simple linear regression with intercept, testing the nullity of the regression coefficient corresponding to the slope is equivalent to performing a Pearson's correlation test between the response and the explanatory variables (Kendall and Stuart 1961, p985).

Let A and B be two traits, $\mathbf{z}_A = (z_k^A)_{n\leq k\leq 2n}$ and $\mathbf{z}_B = (z_k^B)_{n\leq k\leq 2n}$ be their tip-value vectors (of dimension $n+1$) and $\mathbf{u}_A = (u_k^A)_{0\leq k<n}$ and $\mathbf{u}_B = (u_k^B)_{0\leq k<n}$ be the corresponding independent contrast vectors (of dimension n) and $\mathbf{1}$ be for the vector with all entries equal to 1 (its dimension depending on the context). Below, we shall present several ways of testing for correlation between A and B.

3.1 Standard regression (SR)

The most basic way to test for correlation between two traits A and B is to consider the standard Pearson's correlation test obtained from the linear equation:

$$\mathbf{z}_B = a_{\text{SR}}\mathbf{1} + b_{\text{SR}}\mathbf{z}_A + \boldsymbol{\varepsilon}_{\text{SR}}, \tag{5}$$

which will be referred to as the SR method. In the SR method, the Pearson's correlation test between traits A and B amounts to testing for the nullity of the coefficient b_{SR}.

Let

$$\begin{bmatrix} \widehat{a}_{\text{SR}} \\ \widehat{b}_{\text{SR}} \end{bmatrix} = \left(\begin{bmatrix} \mathbf{1}' \\ \mathbf{z}_A' \end{bmatrix} \begin{bmatrix} \mathbf{1} & \mathbf{z}_A \end{bmatrix} \right)^{-1} \begin{bmatrix} \mathbf{1}' \\ \mathbf{z}_A' \end{bmatrix} \mathbf{z}_B,$$

be the vector of ordinary least square estimates of the coefficients a_{SR} and b_{SR}. The variance estimate of b_{SR} is $\widehat{\sigma}^2_{\text{SR}}(\mathbf{z}_A'\mathbf{z}_A)^{-1}$ where

$$\widehat{\sigma}^2_{\text{SR}} = \frac{(\mathbf{z}_B - \widehat{a}_{\text{SR}}\mathbf{1} - \widehat{b}_{\text{SR}}\mathbf{z}_A)'(\mathbf{z}_B - \widehat{a}_{\text{SR}}\mathbf{1} - \widehat{b}_{\text{SR}}\mathbf{z}_A)}{n-1}$$

is the residual variance estimate. Under the key assumptions of the regression model, testing the nullity of the coefficient b_{SR} is performed thanks to the fact that if $b_{\text{SR}} = 0$ then the ratio of the coefficient estimate $\widehat{b}_{\text{SR}}$ to its standard error $\sqrt{\widehat{\sigma}^2_{\text{SR}}(\mathbf{z}_A'\mathbf{z}_A)^{-1}}$ follows a Student distribution with $n-1$ degrees of freedom. Unfortunately the key assumptions of the regression model are not granted here since entries of the error vector $\boldsymbol{\varepsilon}_{\text{SR}}$ are not sampled from independent and identically distributed random variables because of the evolutionary relationships between the species involved (Harvey and Pagel 1991). Though one expects correlation tests with the SR method to be inaccurate in a phylogenetic context (except if $\mathcal{T}$ is a star tree), the SR method is included in the study in order to be used as a basis of comparison.

3.2 Independent contrasts (IC)

The usual way to cope with the evolutionary dependency between the tip values of the traits is to consider their independent contrasts (Felsenstein 1985), referred to as the IC method below, which is based on the regression through origin between the independent contrasts according to the equation:

$$\mathbf{u}_B = b_{\text{IC}}\mathbf{u}_A + \boldsymbol{\varepsilon}_{\text{IC}}. \tag{6}$$

The IC method is widely used and has been assessed in several works (Grafen 1989; Martins and Garland 1991; Pagel 1993; Garland and Adolph 1994; Martins 1996; Diaz-Uriarte and Garland 1996). In the IC approach, testing correlation between traits A and B amounts to testing for the nullity of the coefficient b_{IC} in the regression through the origin. The ordinary least square estimate $\widehat{b}_{\text{IC}} = (\mathbf{u}_A'\mathbf{u}_A)^{-1}\mathbf{u}_A'\mathbf{u}_A$ of b_{IC} has variance $\widehat{\sigma}^2_{\text{IC}}(\mathbf{u}_A'\mathbf{u}_A)^{-1}$ where

$$\widehat{\sigma}^2_{\text{IC}} = \frac{(\mathbf{u}_B - \widehat{b}_{\text{SR}}\mathbf{u}_A)'(\mathbf{u}_B - \widehat{b}_{\text{SR}}\mathbf{u}_A)}{n-1}.$$

Under the key assumptions of the regression analysis, if $b_{\text{IC}} = 0$ then the ratio of $\widehat{b}_{\text{IC}}$ to its standard deviation follows a Student distribution with $n-1$ degrees of freedom.

Testing for the nullity of b_{IC} by using this property is theoretically founded if one assumes that both traits A and B evolve following a BM model since Corollary 4 and Theorem 7 ensures that their respective independent contrasts are realizations of independent and identically distributed centered Gaussian random variables $(U_k^A)_{0\le k<n}$ and $(U_k^B)_{0\le k<n}$ with the same correlation as the two traits. Putting ρ, σ_A^2 and σ_B^2 for this correlation and the variances of $(U_k^A)_{0\le k<n}$ and $(U_k^B)_{0\le k<n}$, respectively, and considering a Cholesky decomposition of their covariance matrix, we get that the random contrasts U_k^B can be written as

$$U_k^B = \frac{\sigma_B\rho}{\sigma_A}U_k^A + \sigma_B\sqrt{1-\rho^2}\mathcal{E}_k,$$

where the terms $(\mathcal{E}_k)_{0\le k<n}$ are independent centered standard Gaussian variables. This shows that the key assumptions of the regression analysis are well granted. The correlation between the independent

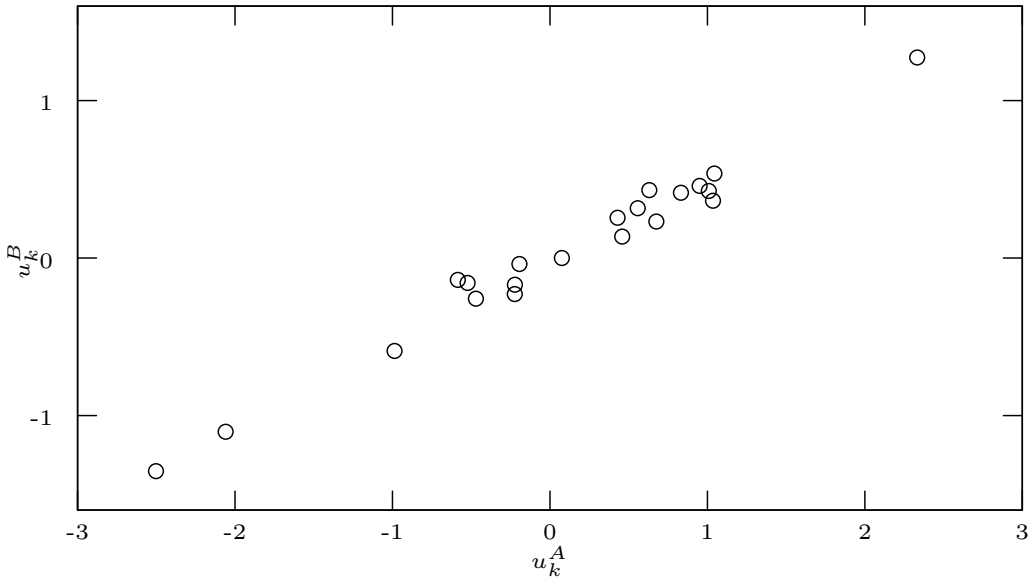

Figure 1: Independent contrasts of two traits A and B simulated on the tree of Figure 2 under two uncorrelated ABM models with trends $\mu_A = 0.5$ and $\mu_B = 0.2$ respectively.

contrast random variables $(U_k^A)_{0\leq k<n}$ and $(U_k^B)_{0\leq k<n}$, thus between the traits A and B, can be assessed by regression through origin between the contrast series $(u_k^A)_{0\leq k<n}$ and $(u_k^B)_{0\leq k<n}$.

In the case where at least one of the two traits to compare follows an ABM model with a non-zero trend, Theorem 7 still ensures that their respective independent contrasts are correlated with the same correlation as the traits. Unfortunately, since from Corollary 4 their independent contrasts are no longer identically distributed but depend on k (we have $U_k^A \sim \mathcal{N}\left(\mu_A h_k, \sigma_A^2\right)$ and $U_k^B \sim \mathcal{N}\left(\mu_B h_k, \sigma_B^2\right)$ with $h_k \neq 0$ in the general case), it makes no sense to test for their correlation through a joint sample.

Figure 1 illustrates how directional trends may make independent contrasts computed from two uncorrelated traits look strongly correlated. This was expected since spurious correlations due to a common dependency on a third factor is a classical phenomenon (Yule 1926; Entorf 1997; Deng 2015).

3.3 Directional contrasts (DC)

A first approach to correct the independent contrasts when at least one of the two traits to compare evolves with a linear trend was proposed in Elliott (2015). The general idea of Elliott's (2015) approach is to center the independent contrasts in order to make them identically distributed. To this end, Elliott (2015) defined the β-directional contrasts $d_k(\beta)$. The formal definition of the β-directional contrasts is recalled in Appendix D in which we show that, for all internal nodes k, the β-directional contrasts $d_k(\beta)$ of Elliott (2015) are equal to the β-centered contrasts $c_k(\beta)$ defined as:

$$c_k(\beta) = u_k - \beta h_k,$$

which have a direct interpretation with regards to the formalism of Section 2.

Let us first remark that under an ABM model with trend μ, Theorem 3 ensures that the μ-centered/μ-directional contrasts are well independent and identically distributed thus could be used in correlation tests. Unfortunately, obtaining these corrected contrasts requires to have the trend parameter μ. More exactly, for testing correlation between two traits A and B, their respective trends μ_A and μ_B have to be known, but they are *a priori* unknown in practical situations. Elliott (2015) proposed to use instead their estimates $\widehat{\mu}_A$ and $\widehat{\mu}_B$ (Equation C2) and to consider the correlation between the (estimated) directional contrasts $d_k^A(\widehat{\mu}_A)$ and $d_k^A(\widehat{\mu}_B)$, where $d_k^A(\widehat{\mu}_A) = u_k^A - \widehat{\mu}_A h_k$ and $d_k^A(\widehat{\mu}_B) = u_k^B - \widehat{\mu}_B h_k$ for all internal nodes k of $\mathcal{T}$.

Regression between estimated directional contrasts will be referred to as the DC method. Namely, by putting $\mathbf{d}_A$ and $\mathbf{d}_B$ for the vectors of directional contrasts $(d_k^A(\widehat{\mu}_A))_{0\leq k<n}$ and $(d_k^A(\widehat{\mu}_B))_{0\leq k<n}$, it is

based on the following linear equation:

$$\mathbf{d}_B = a_{\text{DC}}\mathbf{1} + b_{\text{DC}}\mathbf{d}_A + \boldsymbol{\varepsilon}_{\text{DC}}. \tag{7}$$

Let

$$\begin{bmatrix} \widehat{a}_{\text{DC}} \\ \widehat{b}_{\text{DC}} \end{bmatrix} = \left(\begin{bmatrix} \mathbf{1}' \\ \mathbf{d}_A' \end{bmatrix} \begin{bmatrix} \mathbf{1} & \mathbf{d}_A \end{bmatrix} \right)^{-1} \begin{bmatrix} \mathbf{1}' \\ \mathbf{d}_A' \end{bmatrix} \mathbf{d}_B,$$

be the vector of ordinary least square estimates of the coefficients a_{DC} and b_{DC} and

$$\widehat{\sigma}^2_{\text{DC}} = \frac{(\mathbf{d}_B - \widehat{a}_{\text{DC}}\mathbf{1} - \widehat{b}_{\text{DC}}\mathbf{d}_A)'(\mathbf{d}_B - \widehat{a}_{\text{DC}}\mathbf{1} - \widehat{b}_{\text{DC}}\mathbf{d}_A)}{n-2},$$

be the residual variance (vectors $\mathbf{d}_A$ and $\mathbf{d}_B$ have dimension n). Testing for correlation between traits A and B amounts to testing for the nullity of the parameter b_{DC}, which, under the key assumptions of the regression model, is performed thanks to the fact that, by assuming that $b_{\text{DC}} = 0$, the ratio of the coefficient estimate $\widehat{b}_{\text{DC}}$ to its standard error $\sqrt{\widehat{\sigma}^2_{\text{DC}}(\mathbf{d}_A'\mathbf{d}_A)^{-1}}$ follows a Student distribution with $n-2$ degrees of freedom.

Unfortunately, since the trends $\widehat{\mu}_A$ and $\widehat{\mu}_B$ are estimated from independent contrasts, the estimated directional contrasts are neither independent, nor identically distributed under ABM models (Appendix D). Applying standard correlation tests on estimated directional contrasts is not founded from a statistical point view.

3.4 Multiple Regression (MR)

If traits A and B follows two ABM models $(x_0^A, \mu_A, \sigma_A^2)$ and $(x_0^B, \mu_B, \sigma_B^2)$ then, for all internal nodes k of $\mathcal{T}$, both independent contrast random variables U_k^A and U_k^B depend on the same explanatory variable h_k (Equation 1). As shown in Figure 1, this dependence on a common factor is likely to cause a systematic correlation between the random variables U_k^A and U_k^B. Neutralizing this spurious correlation requires to include the common explanatory variable h_k in the regression (Yule 1926; Deng 2015). By combining Equations 3 and 4, we get that

$$U_k^B = \frac{\sigma_B\rho}{\sigma_A}U_k^A + \left(\mu_B - \frac{\sigma_B\rho}{\sigma_A}\mu_A\right)h_k + \frac{\sigma_B\sqrt{1-\rho^2}}{\sqrt{\overline{\mathrm{v}}_{r(k)} + \overline{\mathrm{v}}_{\ell(k)}}}\Phi_k(\widetilde{W}^B).$$

This suggests to consider the multiple regression through origin between contrasts by including h_k as co-variable, i.e., to consider the equation:

$$\mathbf{u}_B = b_{\text{MR}}\mathbf{u}_A + c_{\text{MR}}\mathbf{h} + \boldsymbol{\varepsilon}_{\text{MR}}, \tag{8}$$

which will be referred to as the MR method. The multiple regression procedure is statistically sound here since the entries of the error vector are sampled from random variables $\left(\frac{\sigma_B\sqrt{1-\rho^2}}{\sqrt{\overline{\mathrm{v}}_{r(k)}+\overline{\mathrm{v}}_{\ell(k)}}}\Phi_k(\widetilde{W}^B)\right)_{0\le k<n}$ which are independent and Gaussian distributed with mean zero and constant variance $\sigma_B^2(1-\rho^2)$ under the current assumptions. The vector of ordinary least square estimates of b_{MR} and c_{MR} is

$$\begin{bmatrix} \widehat{b}_{\text{MR}} \\ \widehat{c}_{\text{MR}} \end{bmatrix} = \left(\begin{bmatrix} \mathbf{u}_A' \\ \mathbf{h}' \end{bmatrix} \begin{bmatrix} \mathbf{u}_A & \mathbf{h} \end{bmatrix} \right)^{-1} \begin{bmatrix} \mathbf{u}_A' \\ \mathbf{h}' \end{bmatrix} \mathbf{u}_B,$$

and the variance of the estimator of b_{MR} is $\sqrt{\widehat{\sigma}^2_{\text{MR}}(\mathbf{u}_A'\mathbf{u}_A)^{-1}}$ where

$$\widehat{\sigma}^2_{\text{MR}} = \frac{(\mathbf{u}_B - \widehat{b}_{\text{MR}}\mathbf{u}_A - \widehat{c}_{\text{MR}}\mathbf{h})'(\mathbf{u}_B - \widehat{b}_{\text{MR}}\mathbf{u}_A - \widehat{c}_{\text{MR}}\mathbf{h})}{n-2}.$$

Equations 3 and 4 show that if traits A and B follows two ABM models $(x_0^A, \mu_A, \sigma_A^2)$ and $(x_0^B, \mu_B, \sigma_B^2)$ with correlation ρ, the random contrasts U_k^B can be written as

$$U_k^B = \frac{\sigma_B\rho}{\sigma_A}U_k^A + \left(\mu_B - \frac{\sigma_B\rho}{\sigma_A}\mu_A\right)h_k + \frac{\sigma_B\sqrt{1-\rho^2}}{\sqrt{\overline{\mathrm{v}}_{r(k)} + \overline{\mathrm{v}}_{\ell(k)}}}\Phi_k(\widetilde{W}^B),$$

where the random error variables $\left(\frac{\sigma_B\sqrt{1-\rho^2}}{\sqrt{\bar{v}_{r(k)}+\bar{v}_{\ell(k)}}}\Phi_k(\widetilde{W}^B)\right)_{0\le k<n}$ are independent and Gaussian distributed with mean zero and constant variance $\sigma_B^2(1-\rho^2)$. The key assumptions of the regression analysis are thus granted if traits A and B follow two ABM models. Testing for correlation between traits A and B can be performed by testing for the nullity of coefficient b_{MR} thanks to the fact that if $b_{\text{MR}} = 0$ then the ratio of the coefficient estimate $\widehat{b}_{\text{MR}}$ to its standard deviation follows a Student distribution with $n-2$ degrees of freedom.

It is worth pointing out that the MR test does not require to estimate neither trend μ_A nor trend μ_B.

3.5 Relation with PGLS method

PGLS method was introduced in Grafen (1989) and further studied in Martins and Garland (1991); Pagel (1997); Martins and Hansen (1997). It is a generalized least squares method specifically designed to take into account the phylogenetic dependency of the regression errors. This dependency relies on evolutionary assumptions. In particular under the BM model, the dependency structure is exactly the same as for the IC method which is based on the same model. Namely, the covariance matrix of the tip random variables of $\mathcal{T}$ has the form $\sigma^2\Sigma$ where σ^2 is the variance of the Brownian model and Σ is the matrix indexed on the tips of $\mathcal{T}$ such that for all pairs of tips (i,j), the entry (i,j) is the total time between the root of $\mathcal{T}$ and the most recent common ancestor of i and j. The PGLS approach is based on the following linear equation

$$\mathbf{z}_B = a_{\text{PGLS}}\mathbf{1} + b_{\text{PGLS}}\mathbf{z}_A + \boldsymbol{\varepsilon}_{\text{PGLS}},$$

which looks the same as that of the SR method but the error vector $\boldsymbol{\varepsilon}_{\text{PGLS}}$ is now assumed to be sampled from a centered Gaussian vector with covariance matrix proportional to Σ. The vector of general least square estimates of the coefficients a_{PGLS} and b_{PGLS} is

$$\begin{bmatrix} \widehat{a}_{\text{PGLS}} \\ \widehat{b}_{\text{PGLS}} \end{bmatrix} = \left(\begin{bmatrix} \mathbf{1}' \\ \mathbf{z}_A' \end{bmatrix} \Sigma^{-1} \begin{bmatrix} \mathbf{1} & \mathbf{z}_A \end{bmatrix} \right)^{-1} \begin{bmatrix} \mathbf{1}' \\ \mathbf{z}_A' \end{bmatrix} \Sigma^{-1}\mathbf{z}_B,$$

and the residual variance estimate is

$$\widehat{\sigma}^2_{\text{PGLS}} = \frac{(\mathbf{z}_B - \widehat{a}_{\text{PGLS}}\mathbf{1} - \widehat{b}_{\text{PGLS}}\mathbf{z}_A)'\Sigma^{-1}(\mathbf{z}_B - \widehat{a}_{\text{PGLS}}\mathbf{1} - \widehat{b}_{\text{PGLS}}\mathbf{z}_A)}{n-1}.$$

Under the PGLS assumptions, testing the nullity of the coefficient b_{PGLS} is performed thanks to the fact that, by assuming that $b_{\text{PGLS}} = 0$, the ratio of the coefficient estimate $\widehat{b}_{\text{PGLS}}$ to its standard error follows a Student distribution with $n-1$ degrees of freedom.

The strong relation between PGLS and IC approaches is well known (Garland and Ives 2000; Rohlf 2001). Blomberg et al. (2012) provided a formal proof that, in the simple regression case, least square estimates of the regression coefficient of the explanatory variable is exactly the same with the IC as with the PGLS methods, namely that $\widehat{b}_{\text{PGLS}} = \widehat{b}_{\text{IC}}$. In Appendix E, we prove the same result in the multiple regression case and show that the variance of the least square estimates of the coefficients are also the same with the IC and the PGLS methods.

Theorem 8. *Let $\mathbf{z}_0$, $\mathbf{z}_1$,... $\mathbf{z}_p$ be the tip value vectors of traits or covariables (e.g., tip times, environmental variables ...) and $\mathbf{u}_0$, $\mathbf{u}_1$,... $\mathbf{u}_p$ be the corresponding independent contrast vectors, $\widehat{\boldsymbol{\beta}}_{IC}$ be the vector of the ordinary least square coefficient estimates from the linear equation though origin*

$$\mathbf{u}_0 = \begin{bmatrix} \mathbf{u}_1 & \dots & \mathbf{u}_p \end{bmatrix} \boldsymbol{\beta}_{IC} + \boldsymbol{\varepsilon}_{IC},$$

and let $\widehat{\boldsymbol{\beta}}_{PGLS}$ be the vector of the generalized least square coefficient estimates from the linear equation with intercept

$$\mathbf{z}_0 = \begin{bmatrix} \mathbf{1} & \mathbf{z}_1 & \dots & \mathbf{z}_p \end{bmatrix} \begin{bmatrix} \alpha_{PGLS} \\ \boldsymbol{\beta}_{PGLS} \end{bmatrix} + \boldsymbol{\varepsilon}_{PGLS},$$

where the error vector $\boldsymbol{\varepsilon}_{PGLS}$ is assumed to be a realization of a centered Gaussian vector with covariance matrix proportional to Σ (the covariance matrix associated to the tree $\mathcal{T}$). The vectors $\widehat{\boldsymbol{\beta}}_{IC}$ and $\widehat{\boldsymbol{\beta}}_{PGLS}$

as well as covariance matrices of the corresponding least squares estimators are equal. Moreover, the degrees of freedom involved in the nullity tests of their coefficients are both equal to $n-p$ with the two approaches.

Proof. Appendix E. □

Theorem 8 directly implies that testing for correlation between two traits with IC and PGLS is completely equivalent, even by considering others traits or covariables.

In order to show the relation between the MR and the PGLS approaches, let us add the tip times as an explanatory covariable in the PGLS regression, i.e., let us consider the linear equation

$$\mathbf{z}_B = a_{\text{PGLSt}}\mathbf{1} + b_{\text{PGLSt}}\mathbf{z}_A + c_{\text{PGLSt}}\mathbf{t} + \boldsymbol{\varepsilon}_{\text{PGLSt}}, \tag{9}$$

where $\mathbf{t} = (t_k)_{n \leq k \leq 2n}$ is the tip time vector. The vector of general least square estimates of the coefficients a_{PGLSt}, b_{PGLSt} and c_{PGLSt} is

$$\begin{bmatrix} \widehat{a}_{\text{PGLSt}} \\ \widehat{b}_{\text{PGLSt}} \\ \widehat{c}_{\text{PGLSt}} \end{bmatrix} = \left(\begin{bmatrix} \mathbf{1}' \\ \mathbf{z}_A' \\ \mathbf{t}' \end{bmatrix} \Sigma^{-1} \begin{bmatrix} \mathbf{1} & \mathbf{z}_A & \mathbf{t} \end{bmatrix} \right)^{-1} \begin{bmatrix} \mathbf{1}' \\ \mathbf{z}_A' \\ \mathbf{t}' \end{bmatrix} \Sigma^{-1}\mathbf{z}_B.$$

Testing for the nullity of b_{PGLSt} is performed by considering the ratio of $\widehat{b}_{\text{PGLSt}}$ to its standard deviation in the Student distribution with $n-2$ degrees of freedom.

We prove in Appendix E that the variables $(h_k)_{0 \leq k < n}$ are the independent contrasts of the tip times. Theorem 8 then implies that both the least square estimates of the regression coefficient associated to trait A and their variances are exactly the same with the MR method and with the PGLSt method.

In short, the IC and PGLS (resp. the MR and PGLSt) methods are interchangeable to test for correlation between two continuous traits under neutral or directional evolution.

An important point is that computations of the PGLS method require to inverse the covariance matrix accounting for the phylogenetic dependencies, which has cubic time complexity with respect to the size of the tree, whereas the IC and MR methods takes advantage of the tree structure of the phylogenetic dependencies in order to perform the same computations in linear time.

3.6 Ultrametric trees

Proposition 5 states that if $\mathcal{T}$ is ultrametric then $h_k = 0$ for all internal nodes k of $\mathcal{T}$. This implies that Equations 6 and 8 turn out to be exactly the same in this case. In other words, on an ultrametric phylogenetic tree, the IC and MR methods are totally equivalent to test for correlation. Moreover, since in an ultrametric tree, the maximum likelihood estimator of the trend returns always 0 (again because $h_k = 0$ for all internal nodes k of $\mathcal{T}$, cf Equation C2), Equation 7 is the same as Equations 6 and 8. In sum, the IC, DC and MR methods are equivalent on ultrametric trees. Using the IC method is statistically founded here since independent contrasts satisfy the requirements of correlations tests in the ultrametric case.

Note that non-ultrametric phylogenetic trees arise in several situations. In particular, phylogenetic trees containing fossil taxa (with or without extant taxa) are not ultrametric (e.g., Laurin 2004; Heim et al. 2015). The ultrametric character relies on the evolutionary model used to infer the trees. For instance, the speciational model which somehow assumes a same unitary branch length all along the tree generally provides non-ultrametric trees (Knouft and Page 2003; Moen 2006; Laurin et al. 2012). Measuring branch lengths in terms of genetic changes (Moen 2006) or scaling branch lengths by their own evolution rates in heterogeneous models (Baker et al. 2015, 2016) instead of considering their geological ages also lead to non-ultrametric trees even if all the taxa are extant.

4 Simulation study

In this section, we shall assess and compare the four correlation tests presented in Section 3:

- SR: standard regression of tips values (Equation 5),
- IC: regression through origin of independent contrasts (Equation 6),

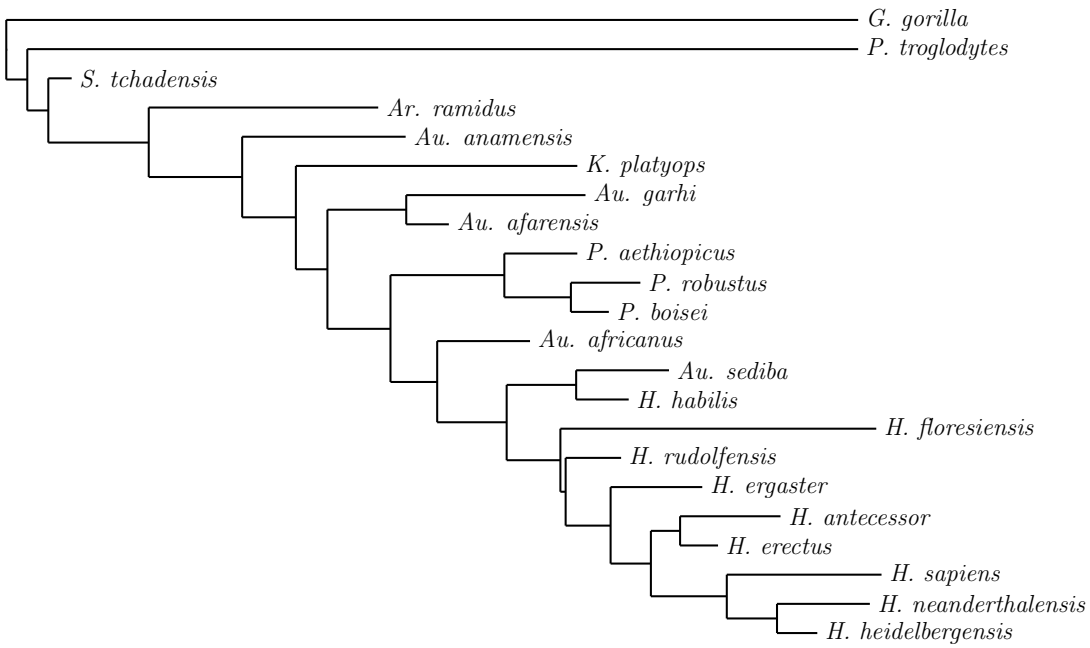

Figure 2: Hominin phylogeny (Dembo et al. 2015).

- DC: regression of directional contrasts (Equation 7),
- MR: multiple regression through origin of independent contrasts with h_k as co-variable (Equation 8).

4.1 Simulation and evaluation protocol

We simulated the evolution of two quantitative traits A and B under various conditions, i.e., under BM and ABM models with several sets of parameters and several levels of correlation between A and B. The simulated evolution runs on the hominin phylogenetic tree displayed in Figure 2 (Dembo et al. 2015).

Although the ABM model has three parameters (x_0, μ, σ^2), we only vary the trend parameter μ in the simulations. The parameter x_0 just translates the whole evolution process, which has no effect on the correlation of a trait with another. Multiplying both the trend and the standard deviation of an ABM model with a constant just results in multiplying the values of the ABM process with same constant (i.e., what actually matters is the ratio of the trend to the standard deviation).

The four correlation tests were next assessed in terms of type I error, i.e., with regard to their ability to not falsely reject the null hypothesis, the null hypothesis being that the traits are uncorrelated, in the case where the traits to compare are actually uncorrelated. We also displays ROC plots of the tests for summarizing their ability to distinguish between correlated and uncorrelated traits (Zhou et al. 2011). Plots of type I error were obtained by simulating uncorrelated traits and by plotting the proportion of simulations for which the null hypothesis was rejected *versus* the level of risk (each test associates to a simulation, a level of risk between 0 and 1, accounting for the chance that this simulation satisfies the null hypothesis). ROC curves were obtained by simulating both negative (i.e., uncorrelated) population and positive (i.e., correlated) population and by plotting for all levels of risk the proportion of true negatives *versus* the proportion of false positives detected by each test. We simulated 50 000 evolutions of correlated and uncorrelated pairs of traits for each plot.

4.2 Correlation tests between two traits under neutral evolution

We first simulated two traits evolving under neutral evolution (i.e., under the BM model) with correlation levels 0 and 0.5 (results obtained with correlation 0.7 are provided in the supplementary information).

Figure 3-Left displays the proportion of type I error at all level of rejection α obtained from 50,000 simulations of two uncorrelated traits under the BM model with variance 0.09. We do observe that both the IC and the MR methods are perfect in the sense that they both rejected the null hypothesis at the

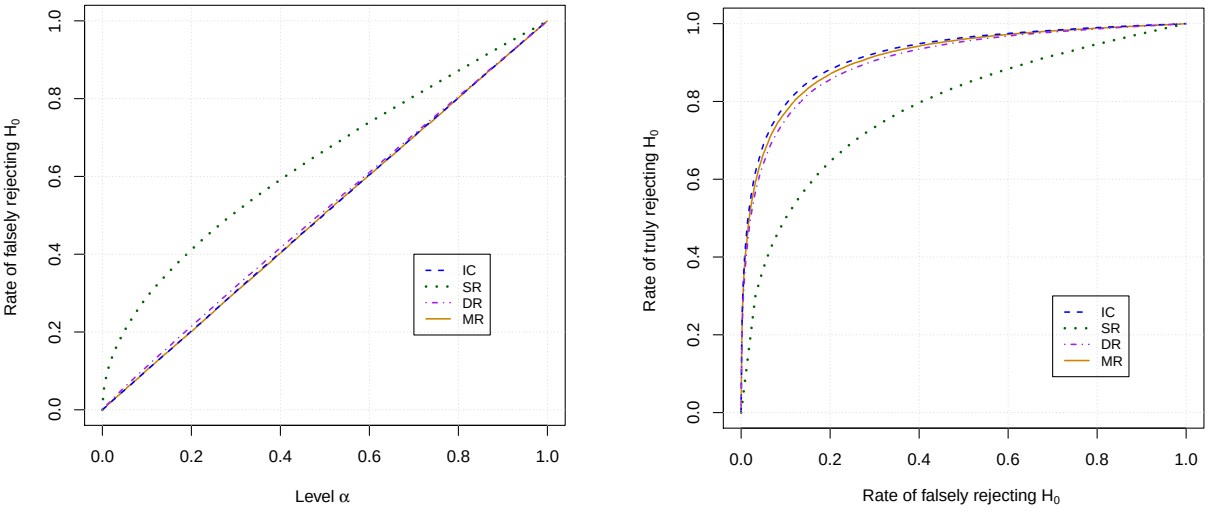

Figure 3: Left: Rates of false rejection of the null hypothesis at level α *vs* α when both traits A and B follow the BM model with variance 0.09. Right: ROC plots of the correlation tests obtained from two simulated traits under the BM model with variance 0.09. Negative population is simulated with uncorrelated traits and positive population with traits correlated with correlation 0.5.

exact level α required (both plots of IC and MR completely overlap with the diagonal in Figure 3-Left). The DC method is close to perfect but tends to reject the null hypothesis a little bit more than it should do. Last, as expected, the worst performance comes from the SR method.

The p-values obtained from testing for the simulations are expected to be uniformly distributed with support $[0, 1]$. We used Kolmogorov-Smirnov (K-S) test in order to check this point. We observed that p-values obtained when testing for the correlation of independent traits follow an uniform distribution both for the IC method (Kolmogorov-Smirnov test, p-value=0.537) and for the MR method (K-S test, p-value=0.832). This is the case neither for the DC nor for the SR methods (K-S test, p-values smaller than 10^{-8}). In sum, under the BM model, only the IC and the MR methods have the behavior expected from a statistical test.

The ROC plots of the tests with a positive population simulated under the same BM model, but with a correlation 0.5 between the traits, are displayed Figure 3-Right. It shows that under the BM model, the most accurate test is IC but MR and DC tests have close performances. As expected the less accurate test is SR.

4.3 Correlation tests between a trait under neutral evolution and a trait under directional evolution

We consider here the mixed situation where one of the traits follows a neutral evolution, here simulated under the BM model with variance 0.09 and the other one follows a directional evolution, here simulated under the ABM model with trend 0.5 and variance 0.09.

Figure 4-Left shows that the behavior of the type I error with regard to the level of rejection α is essentially the same as in the case of two traits under neutral evolution for all the methods. Moreover, p-values obtained here still follow an uniform distribution with support $[0, 1]$ both for the IC method (K-S test, p-value=0.345) and for the MR method (K-S test, p-value=0.586). This is the case neither for the DC nor for the SR methods (K-S test, p-values smaller than 10^{-8}).

The ROC plots displayed in Figure 4-Right shows that performances of the IC and the SR tests are significantly lower than that of the MR and DC tests. The MR method is slightly more accurate than the DC test.

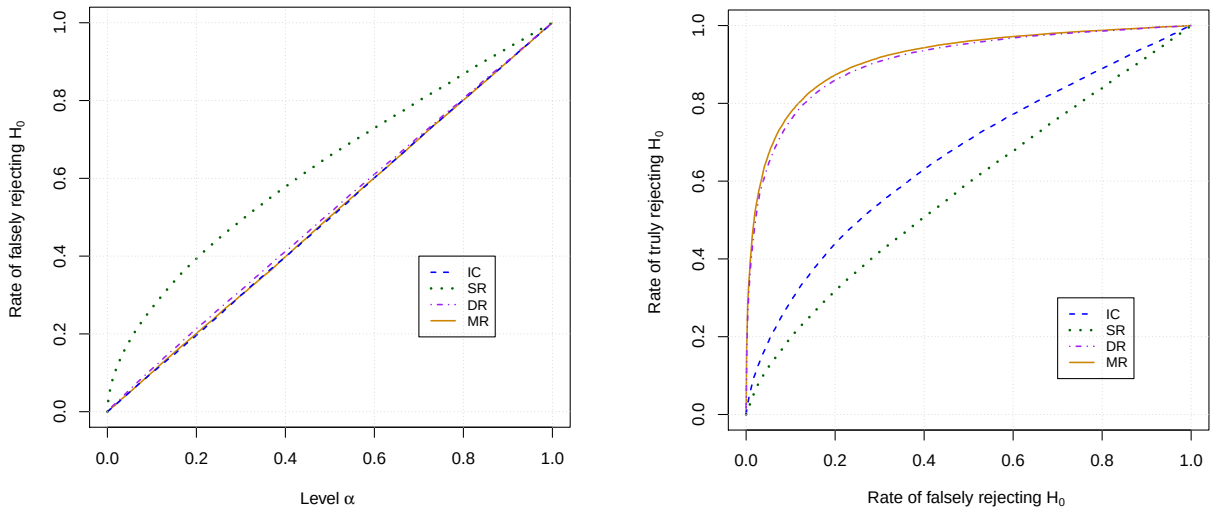

Figure 4: Left: Rates of false rejection of the null hypothesis at level α *vs* α when trait A follows the BM model with variance 0.09 and trait B follows the ABM model with trend 0.5 and variance 0.09. Right: ROC plots of the correlation tests obtained from two simulated traits under a BM model and an ABM model with trend 0.5 respectively, both with variance 0.09. Negative population is simulated with uncorrelated traits and positive population with traits correlated with correlation 0.5.

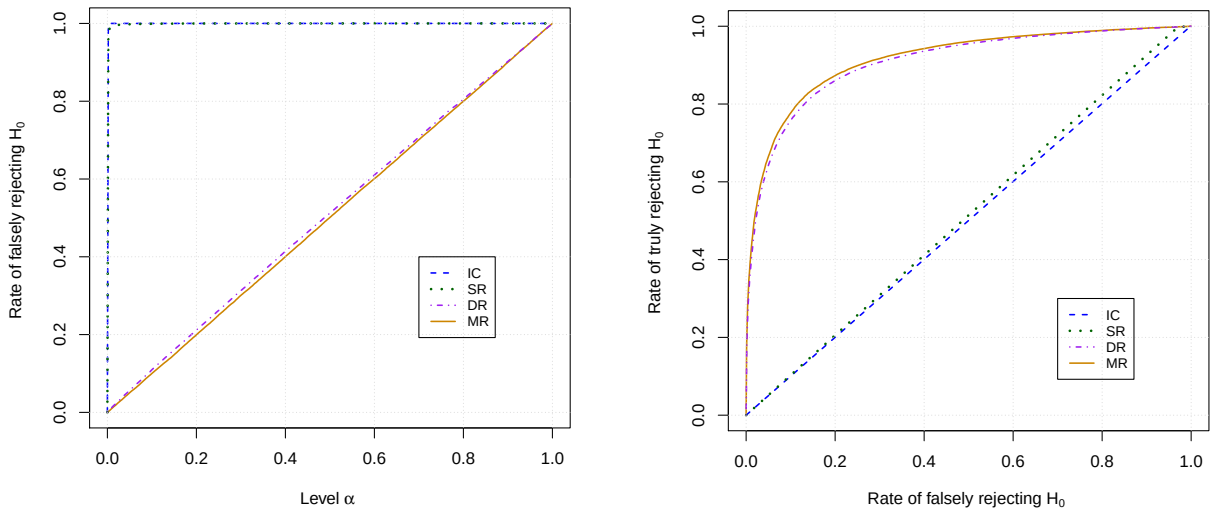

Figure 5: Left: Rates of false rejection of the null hypothesis at level α *vs* α when both traits A and B follow the ABM model with variance 0.09, and trend 0.5 and 1 respectively. Right: ROC plots of the correlation tests obtained from two simulated traits under ABM models with trend 0.5 and 1 respectively, both with variance 0.09. Negative population is simulated with uncorrelated traits and positive population with traits correlated with correlation 0.5.

4.4 Correlation tests between two traits under directional evolution

In the case where the two traits are under directional evolution (here trait A has trend 0.5, and trait B has trend 1 both with variance 0.09), the rate of type I error of both the SR and the IC methods becomes maximal (Fig. 5-Left). In plain English, the SR and the IC method systematically reject the hypothesis that the traits are uncorrelated, even when they are uncorrelated. This behavior clearly prevents us to use the IC and the SR methods to detect correlation between traits under directional evolution. Still with regard to type I errors, performances of the MR and DC methods are essentially the same as in the case of neutral evolution or in the "mixed" case. The MR method looks perfect and the DC method still tends to reject the null hypothesis a little bit more than it should do. Taking a closer look on the p-values of the tests, we observe that those of the IC method no longer follow an uniform distribution (K-S test, p-value below 10^{-9}) and so do those of the SR and DC methods. Only the MR method has the expected behavior of a test in this situation (K-S test, p-value=0.935).

Figure 5-Right displays the ROC plots of the SR, IC, DC and MR tests. The performance of the IC test is not better than a random guess and that of the SR test is almost as bad. The accuracy of the MR and DC tests is essentially the same as in the preceding case.

4.5 Discussion

As expected, the less accurate test is SR in all the situations. Overall, we observe that despite the flaws in its statistical properties, the DC method performs generally well with regards to the two criteria considered, whatever the trends of the traits. However, the DC method is always outperformed by the MR method. The "historical" IC method is outperformed by both the MR and the DC methods as soon as one of the traits is under directional evolution. It has the best performance from the ROC criteria only when the two traits evolve under the BM model, which is not very surprising since it corresponds exactly to the assumptions of this method, but the accuracy of the MR and DC methods is very close. The additional figures obtained from a greater variety of parameters and provided in the Supplementary Information lead to the same observations.

The simulations suggest to first test for the presence of a non-zero trend on each trait to compare, for instance by using the method of Appendix C, then to use the MR method if at least one of the traits shows a significant trend, and to use the standard IC method only if the two traits are under neutral evolution.

5 Correlation between hominin cranial capacity and body mass

5.1 Data

Evolution of hominin cranial capacity and body mass was studied in numerous works (Kappelman 1996; Henneberg 1998; Wood and Collard 1999; Leonard et al. 2003; Falk et al. 2005; Weber et al. 2005; Martin et al. 2006; Young 2006; Snodgrass et al. 2009; Montgomery et al. 2010; Potts 2011; Shultz et al. 2012; Schoenemann 2013; Hofman 2014; Grabowski et al. 2015; Grabowski 2016; Argue et al. 2017; Will et al. 2017; Du et al. 2018).

Our study is based on the hominin phylogenetic tree summarizing the best trees obtained in the dated Bayesian analysis of Dembo et al. (2015, Fig. 1), which is displayed in Figure 2. We combined data from several articles in order to get the body mass and the cranial capacity of as many species as possible, namely from Kappelman (1996, Table 1), Wood and Collard (1999, Table 3), Leonard et al. (2003, Table 3), Young (2006, Table 1), Schoenemann (2013, Tables 8.1 and 8.2), Grabowski et al. (2015, Table 4), Will et al. (2017, Table 4) and Du et al. (2018, Elec. Supp.). We excluded data associated to ambiguously identified species and to juvenile specimens. We finally averaged all the collected cranial capacities and body masses by species in order to obtain the data displayed in Table 1. We excluded *H. floresiensis* for calibrating our models, because being an outlier (Weber et al. 2005; Martin et al. 2006; Falk et al. 2007; Argue et al. 2017), this species over-influenced the results. Each time that a data required in an analysis was missing, we did not consider the corresponding species in this analysis. In particular, the correlation study pertains only to species for which both cranial capacity and body mass are known.

Species	Cranial capacity (cm^3)	Body mass (g)
Ar. ramidus	300.0	38,067
Au. afarensis	436.4	38,680
Au. africanus	457.1	31,260
Au. anamensis	–	46,300
Au. garhi	450.0	–
Au. sediba	420.0	26,485
G. gorilla	520.2	120,500
H. antecessor	1,218.3	–
H. erectus	982.9	58,274
H. ergaster	840.0	58,164
H. floresiensis	417.0	27,500
H. habilis	580.3	35,782
H. heidelbergensis	1,214.5	80,440
H. neanderthalensis	1,470.2	79,573
H. rudolfensis	752.0	45,597
H. sapiens	1,391.8	64,224
K. platyops	425.0	–
P. aethiopicus	418.2	37,666
P. boisei	509.2	45,971
P. robustus	527.5	36,124
P. troglodytes	387.7	53,011
S. tchadensis	365.0	–

Table 1: Cranial capacity and body mass of species of the phylogenetic tree of Figure 2.

We considered the logarithms of cranial capacity and body mass data such as in Kappelman (1996); Henneberg (1998); Leonard et al. (2003); Snodgrass et al. (2009); Navarrete et al. (2011); Du et al. (2018). Taking the logarithm of quantitative trait values is quite usual since it accounts for the fact that for instance, an increase of 100 g does not have the same significance for an organism of 1 kg as for a organism of 100 kg. From a statistical point of view, log-transformation is a particular case of Box-Cox transformations which tend to stabilize the variance. It is also sometimes used to approach Gaussian behavior required by Brownian evolution models (Legendre and Desdevises 2009).

We applied diagnostic tests on residuals after log-transformation in order to check for least squares regression validity conditions. Namely, we use Durbin-Watson's test to detect autocorrelation at lag 1 (Durbin and Watson 1950, 1951, 1971); Harrison-McCabe's test to detect heteroscedasticity (Harrison and McCabe 1979) and Jarque-Bera's test to confirm normality (Jarque and Bera 1987).

5.2 Evolutionary trends of hominin cranial capacity and body mass

Several works agree with the fact that hominin brain size increased through evolution (Henneberg 1998; Montgomery et al. 2010; Navarrete et al. 2011; Potts 2011; Shultz et al. 2012; Hofman 2014). Henneberg (1998) found a significant correlation between the log-transformed cranial capacity and the fossil age through a direct "non-phylogenetic" regression approach. Applying the same approach on our dataset, we also detected a positive evolutionary trend (p-value=0.003), but least squares regression conditions are violated (Harrison-McCabe's test, p-value=0.001), certainly because of the phylogenetic relationships between tips values. Conversely, the hominin cranial capacities fulfill the conditions of the phylogenetic trend detection test presented introduced in Appendix C (Durbin-Watson's test, p-value=0.422; Harrison-McCabe's test, p-value=0.604; Jarque-Bera's test, p-value=0.646). The test of Appendix C concludes to a positive trend (p-value=0.012) which amounts to multiplying the cranial capacity by 1.2 per Ma.

Several studies conclude that the body mass data increased during evolution by using non phylogenetic approaches, i.e., without taking into account the evolutionary relationships between the species (Henneberg 1998; Will et al. 2017). Considering a direct "non phylogenetic" regression of the logarithm

of the body mass of our dataset led to detect a positive evolutionary trend (p-value=0.002), whereas, by taking into account the evolutionary relationships between species, the detection test of Appendix C did not conclude to a significant trend on the logarithm of hominin body mass (p-value=0.072). Both the non phylogenetic regression and our detection test satisfy the regression assumptions (Durbin-Watson's test, p-value=0.243 and p-value=0.227; Harrison-McCabe's test, p-value=0.066 and p-value=0.796; Jarque-Bera's test, p-value=0.953 and p-value=0.865 respectively).

Our results are consistent with those of Montgomery et al. (2010), who also found a positive trend in cranial capacity and no significative trend in the body size evolution of hominids. Finally, testing for correlation between the logarithms of hominin cranial capacity and body mass falls in a situation close to that of Section 4.3, in which we compared a trait simulated under neutral evolution with a trait simulated under directional evolution.

5.3 Correlation between hominin cranial capacity and body mass

We applied the MR method in order to test for correlation between the logarithms of the hominin body mass and cranial capacity. These data fulfill the requirements of our correlation test (Durbin-Watson's test, p-value=0.819; Harrison-McCabe's test, p-value=0.690; Jarque-Bera's test, p-value=0.461). Durbin-Watson's and Harrison-McCabe's tests require to order the residuals with respect to the explanatory variable, which is straightforward for single regressions, but not for multiple regressions. Following Fan and Huang (2001), we ordered the residuals with respect to most informative linear combination of the explanatory variables obtained from principal component analysis.

It is thus allowed to apply the MR test, which concluded that the logarithms of hominin cranial capacity and body size are significantly correlated (p-value=0.027). In plain English, even by taking into account possible evolutionary trends, the logarithms of these two traits do not change independently. This result is consistent with Grabowski (2016), who showed that the evolution of hominin body size and that of the cranial capacity are related.

References

Adams, D. C. and Collyer, M. L. (2015). Permutation tests for phylogenetic comparative analyses of high-dimensional shape data: What you shuffle matters. *Evolution*, 69(3):823–829.

Argue, D., Groves, C., Lee, M., and Jungers, W. (2017). The affinities of *Homo floresiensis* based on phylogenetic analyses of cranial, dental and postcranial characters. *J. Hum. Evol.*, 107:107–133.

Arnold, E. N. and Moncrieff, C. (1994). Do ecological analogues assemble their common features in the same order? an investigation of regularities in evolution, using sand-dwelling lizards as examples. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 344(1309):277–290.

Baker, J., Meade, A., Pagel, M., and Venditti, C. (2015). Adaptative evolution toward larger size in mammals. *Am. Nat.*, 112(16):5093–5098.

Baker, J., Meade, A., Pagel, M., and Venditti, C. (2016). Positive phenotypic selection inferred from phylogenies. *Biological Journal of the Linnean Society*, 118(1):95–115.

Blomberg, S., Lefevre, J., Wells, J., and Waterhouse, M. (2012). Independent Contrasts and PGLS Regression Estimators are Equivalent. *Syst. Biol.*, 61(3):382–391.

Bokma, F., Godinot, M., Maridet, O., Ladevèze, S., Costeur, L., Solé, F., Gheerbrant, E., Peigné, S., Jacques, F., and Laurin, M. (2016). Testing for Depéret's Rule (Body Size Increase) in Mammals using Combined Extinct and Extant Data. *Syst. Biol.*, 65(1):98–108.

Cressler, C. E., Butler, M. A., and King, A. A. (2015). Detecting Adaptive Evolution in Phylogenetic Comparative Analysis Using the Ornstein–Uhlenbeck Model. *Systematic Biology*, 64(6):953–968.

Dembo, M., Matzke, J., Mooers, A. O., and Collard, M. (2015). Bayesian analysis of a morphological supermatrix sheds light on controversial fossil hominin relationships. *Proc. R. Soc. B*, 282:1–9.

Deng, A. (2015). A Primer on Spurious Statistical Significance in Time Series Regressions. *Economics Committee Newsletter*, 14(1):25–36.

Diaz-Uriarte, R. and Garland, T. (1996). Testing hypotheses of correlated evolution using phylogenetically independent contrasts: sensitivity to deviations from brownian motion. *Syst. Biol.*, 45(1):27–47.

Du, A., Zipkin, A., Hatala, K., Renner, E., Baker, J., Bianchi, S., Bernal, K., and Wood, B. (2018). Pattern and process in hominin brain size evolution are scale-dependent. *Proc. R. Soc. B*, 285:1–8.

Durbin, J. and Watson, G. (1950). Testing for Serial Correlation in Least Squares Regression I. *Biometrika*, 37:409–428.

Durbin, J. and Watson, G. (1951). Testing for Serial Correlation in Least Squares Regression II. *Biometrika*, 38:159–178.

Durbin, J. and Watson, G. (1971). Testing for Serial Correlation in Least Squares Regression III. *Biometrika*, 58:1–19.

Elliott, M. (2015). Identical inferences about correlated evolution arise from ancestral state reconstruction and independent contrasts. *J. Theor. Biol.*, 364:321–325.

Entorf, H. (1997). Random walks with drifts: Nonsense regression and spurious fixed-effects estimation. *Journal of Econometrics*, 80:287–296.

Falk, D., Hildebolt, C., Smith, K., Morwood, M., Sutikna, T., Brown, P., Jatmiko, S., Saptomo, E., Brunsden, B., and Prior, F. (2005). The Brain of LB1, *Homo floresiensis*. *Science*, 308(5719):1–4.

Falk, D., Hildebolt, C., Smith, K., Morwood, M., Sutikna, T., Jatmiko, S., Saptomo, E., Imhof, H., Seidler, H., and Prior, F. (2007). Brain shape in human microcephalics and *Homo floresiensis*. *PNAS*, 104(7):2513–2518.

Fan, J. and Huang, L. (2001). Goodness-of-Fit Tests for Parametric Regression Models. *J. Am. Statist. Assoc.*, 96(454):674–688.

Felsenstein, J. (1973). Maximum-Likelihood Estimation of Evolutionary Trees from Continuous Characters. *Am. J. Hum. Genet.*, 25(5):471–492.

Felsenstein, J. (1985). Phylogenies and the comparative methods. *Am. Nat.*, 125:1–15.

Felsenstein, J. (1988). Phylogenies and quantitative characters. *Annual Review of Ecology and Systematics*, 19(1):445–471.

Fox, J. (2015). *Applied regression analysis and generalized linear models*. Sage Publications.

Garland, T. and Adolph, S. (1994). Why Not to Do Two-Species Studies: Limitations on Inferring Adaptation. *Physiological Zoology*, 67(4):797–828.

Garland, T. and Ives, A. (2000). Using the Past to Predict the Present: Confidence Intervals for Regression Equations in Phylogenetic Comparative Methods. *Am. Nat.*, 155(3):346–364.

Grabowski, M. (2016). Bigger brains led to bigger bodies? the correlated evolution of human brain and body size. *Curr. Anthropol.*, 57(2):174–196.

Grabowski, M., Hatala, K., Jungers, W., and Richmond, B. (2015). Body mass estimates of hominin fossils and the evolution of human body size. *J. Hum. Evol.*, 85:75–93.

Grafen, A. (1989). The phylogenetic regression. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 326(1233):119–157.

Grimmett, G. and Stirzaker, D. (2001). *Probability and Random Processes*. Oxford university press.

Groussin, M. and Gouy, M. (2011). Adaptation to Environmental Temperature Is a Major Determinant of Molecular Evolutionary Rates in Archaea. *Mol. Biol. Evol.*, 28(9):2661–2674.

Gupta, A. (2013). *Risk Management and Simulation.* Chapman & Hall, CRC Press.

Hansen, T. F., Pienaar, J., and Orzack, S. H. (2008). A comparative method for studying adaptation to a randomly evolving environment. *Evolution*, 62(8):1965–1977.

Harrison, M. and McCabe, B. (1979). A Test for Heteroscedasticity Based on Ordinary Least Squares Residuals. *J. Am. Statist. Assoc.*, 74(366):494–499.

Harvey, P. and Pagel, M. (1991). *The comparative methods in evolutionary biology.* Oxford Series in Ecology and Evolution. Oxford Univ. Press, Oxford, England.

Heim, N., Knope, M., Schaal, E., Wang, S., and Payne, J. (2015). Cope's rule in the evolution of marine animals. *Science*, 347:867–870.

Henneberg, M. (1998). Evolution of the human brain: is bigger better? *Clin. Exp. Pharmacol. Physiol.*, 25(9):745–749.

Hofman, M. (2014). Evolution of the human brain: when bigger is better. *Front. Neuroanat.*, 8(15):1–12.

Hone, D. and Benton, M. (2005). The evolution of large size: how does Cope's Rule work? *Trends Ecol. Evol.*, 20(1):4–6.

Hone, D., Keesey, T., Pisani, D., and A., P. (2005). Macroevolutionary trends in the Dinosauria: Cope's rule. *J. Evol. Biol.*, 18:587–595.

Hunt, G. and Roy, K. (2006). Climate change, body size evolution, and Cope's Rule in deep-sea ostracodes. *Proceedings of the National Academy of Science*, 103(5):1347–1352.

Jarque, C. and Bera, A. (1987). A test for normality of observations and regression residuals. *International Statistical Review*, 55(2):163–172.

Kappelman, J. (1996). The evolution of body mass and relative brain size in fossil hominids. *J. Hum. Evol.*, 30:243–276.

Kendall, M. and Stuart, A. (1961). *The advanced theory of statistics – Inference and relationship*, volume 2. Griffin & Co, London.

Kingsolver, J. and Pfennig, D. (2004). Individual-level selection as a cause of Cope's Rule of phylogenetic size increase. *Evolution*, 58:1608–1612.

Knouft, J. and Page, L. (2003). The evolution of body size in extant groups of north american freshwater fishes: Speciation, size distributions, and cope's rule. *Am. Nat.*, 161:413–421.

Laurin, M. (2004). The evolution of body size, Cope's rule and the origin of amniotes. *Syst. Biol.*, 53(4):594–622.

Laurin, M., Gussekloo, S. W. S., Marjanović, D., Legendre, L., and Cubo, J. (2012). Testing gradual and speciational models of evolution in extant taxa: the example of ratites. *Journal of Evolutionary Biology*, 25(2):293–303.

Legendre, P. and Desdevises, Y. (2009). Independent contrasts and regression through the origin. *J. theor. Biol.*, 259(4):727–743.

Leonard, W., Robertson, M., Snodgrass, J., and Kuzawa, C. (2003). Metabolic correlates of hominid brain evolution. *Comp. Biochem. Physiol A*, 136:5–15.

Lomolino, M. V. (2005). Body size evolution in insular vertebrates: generality of the island rule. *Journal of Biogeography*, 32(10):1683–1699.

Marchini, M., Sparrow, L., Cosman, M. N., Dowhanik, A., Krueger, C., Hallgrimsson, B., and Rolian, C. (2014). Impacts of genetic correlation on the independent evolution of body mass and skeletal size in mammals. *BMC Evol. Biol.*, 14(1):258.

Martin, R., Maclarnon, A., Phillips, J., Dussubieux, L., Williams, P., and Dobyns, W. (2006). Comment on "The Brain of LB1, *Homo floresiensis*". *Science*, 312(5776):999.

Martins, E. (1996). Phylogenies, spatial autocorrelation, and the comparative method: a computer simulation test. *Evolution*, 50(5):1750–1765.

Martins, E. and Garland, T. (1991). Phylogenetic analyses of the correlated evolution of continuous characters: a simulation study. *Evolution*, 45(3):534–557.

Martins, E. and Hansen, T. (1997). Phylogenies and the comparative method: a general approach to incorporating phylogenetic information into the analysis of interspecific data. *American Naturalist*, 149:646–667.

Moen, D. (2006). Cope's rule in cryptodiran turtles: do the body sizes of extant species reflect a trend of phyletic size increase? *J. Evol. Biol.*, 19(4):1210–1221.

Montgomery, S., Cappellini, I., Barton, R., and Mundi, N. (2010). Reconstructing the ups and downs of primate brain evolution: implications for adaptative hypotheses and *Homo floresiensis*. *BMC Biology*, 8(9):1–19.

Navarrete, A., Van Schaik, C., and Isler, K. (2011). Energetics and the evolution of human brain size. *Nature*, 480:91–94.

Oakley, T. H. and Cunningham, C. W. (2000). Independent contrasts succeed where ancestor reconstruction fails in a known bacteriophage phylogeny. *Evolution*, 54(2):397–405.

Pagel, M. (1993). Seeking the evolutionary regression coefficient: An analysis of what comparative methods measure. *J. theor. Biol.*, 164:191–205.

Pagel, M. (1997). Inferring evolutionary processes from phylogenies. *Zool. scr.*, 26(4):331–348.

Paradis, E., Claude, J., and Strimmer, K. (2004). APE: analyses of phylogenetics and evolution in R language. *Bioinformatics*, 20:289–290.

Potts, R. (2011). Big brains explained. *Nature*, 480:43–44.

Rohlf, F. (2001). Comparative methods for the analysis of continuous variable: geometric interpretations. *Evolution*, 55(11):2143–2160.

Schoenemann, P. (2013). Hominid Brain Evolution. In *A Companion to Paleoanthropology*, pages 136–164. Blackwell Publishing Ltd, Oxford, begun, d.r. edition.

Seligmann, H. (2018). Giant viruses as protein-coated amoeban mitochondria? *Virus Research*, 253:77 – 86.

Seligmann, H. (2019). Giant viruses: spore-like missing links between rickettsia and mitochondria? *Annals of the New York Academy of Sciences*, 0(0).

Shultz, S., Nelson, E., and Dunbar, R. (2012). Hominin cognitive evolution: identifying patterns and processes in the fossil and archaeological record. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, 367(1599):2130–2140.

Slater, G. J., Price, S. A., Santini, F., and Alfaro, M. E. (2010). Diversity versus disparity and the radiation of modern cetaceans. *Proceedings of the Royal Society of London B: Biological Sciences*.

Snodgrass, J., Leonard, W., and Robertson, M. (2009). The Energetics of Encephalization in Early Hominids. In Hublin, J. and Richards, M., editors, *The Evolution of Hominin Diets*, Vertebrate Paleobiology and Paleoanthropology, pages 15–29. Springer, Dordrecht.

Van Valkenburgh, B., Wang, X., and Damuth, J. (2004). Cope's Rule, Hypercarnivory, and extinction in North American Canids. *Science*, 306(5693):101–104.

Vigouroux, Y., Matsuoka, Y., and Doebley, J. (2003). Directional Evolution for Microsatellite Size in Maize. *Molecular Biology and Evolution*, 20(9):1480–1483.

Weber, J., Czarnetzki, A., and Pusch, C. (2005). Comment on "The Brain of LB1, *Homo floresiensis*". *Science*, 310(5746):236.

Will, M., Pablo, A., and Stock, J. (2017). Long-term patterns of body mass and stature evolution within the hominin lineage. *R. Soc. open sci*, 4(171339).

Wood, B. and Collard, M. (1999). The human genus. *Science*, 284:65–71.

Young, P. (2006). Body Mass Estimates and Encephalization Quotients: A Fresh Look at the Australopithecines and *Homo habilis*. *CRS Quaterly*, 42(4):217–226.

Yule, G. (1926). Why do we Sometimes get Nonsense-Correlations between Time-Series?–A Study in Sampling and the Nature of Time-Series. *J. Roy. Statist. Soc.*, 89:1–63.

Zhao, T., Liu, D., and Li, Z. (2017). Correlated evolution of sternal keel length and ilium length in birds. *PeerJ*, PeerJ 5:e3622.

Zhou, X.-H., Obuchowski, N., and McClish, D. (2011). *Statistical methods in diagnostic medicine*. Wiley Series in Probability and Statistics. J. Wiley And Sons, New York, 2nd edition.

A Distribution of independent contrasts – Proof of Theorem 3

Let us first prove by induction that for all nodes k of $\mathcal{T}$, we have that

$$\begin{aligned}
\overline{\mathrm{Z}}_k &= x_0 + \mu(t_k + \gamma_k) + \sigma\left(W_{t_k} + \Psi_k(W)\right), \quad \text{with} \\
\gamma_k &= \begin{cases} 0 & \text{if } k \text{ is a tip,} \\ \dfrac{\overline{\mathrm{v}}_{\ell(k)}\big(\gamma_{r(k)} + v_{r(k)}\big) + \overline{\mathrm{v}}_{r(k)}\big(\gamma_{\ell(k)} + v_{\ell(k)}\big)}{\overline{\mathrm{v}}_{r(k)} + \overline{\mathrm{v}}_{\ell(k)}} & \text{otherwise, and} \end{cases} \\
\Psi_k(W) &= \begin{cases} 0 & \text{if } k \text{ is a tip,} \\ \dfrac{\overline{\mathrm{v}}_{\ell(k)}\Big(W_{t_{r(k)}} - W_{t_k} + \Psi_{r(k)}(W)\Big) + \overline{\mathrm{v}}_{r(k)}\Big(W_{t_{\ell(k)}} - W_{t_k} + \Psi_{\ell(k)}(W)\Big)}{\overline{\mathrm{v}}_{r(k)} + \overline{\mathrm{v}}_{\ell(k)}} & \text{otherwise.} \end{cases}
\end{aligned}$$

In the base case, i.e., when k is a tip, the property is granted since from Definition 2, we have that

$$\begin{aligned}
\overline{\mathrm{Z}}_k &= Z_k \\
&= x_0 + \mu t_k + \sigma W_{t_k} \\
&= x_0 + \mu(t_k + \gamma_k) + \sigma\left(W_{t_k} + \Psi_k(W)\right), \quad \text{with } \gamma_k = 0 \text{ and } \Psi_k(W) = 0.
\end{aligned}$$

Let k be an internal node and let us assume that the property holds for its direct descendants $r(k)$ and $\ell(k)$. From Definition 2, we have that

$$\begin{aligned}
\overline{\mathrm{Z}}_k &= \frac{\overline{\mathrm{v}}_{\ell(k)}\overline{\mathrm{Z}}_{r(k)} + \overline{\mathrm{v}}_{r(k)}\overline{\mathrm{Z}}_{\ell(k)}}{\overline{\mathrm{v}}_{r(k)} + \overline{\mathrm{v}}_{\ell(k)}} \\
&= \frac{\overline{\mathrm{v}}_{\ell(k)}x_0 + \overline{\mathrm{v}}_{r(k)}x_0}{\overline{\mathrm{v}}_{r(k)} + \overline{\mathrm{v}}_{\ell(k)}} + \mu\left(\frac{\overline{\mathrm{v}}_{\ell(k)}}{\overline{\mathrm{v}}_{r(k)} + \overline{\mathrm{v}}_{\ell(k)}}(t_{r(k)} + \gamma_{r(k)}) + \frac{\overline{\mathrm{v}}_{r(k)}}{\overline{\mathrm{v}}_{r(k)} + \overline{\mathrm{v}}_{\ell(k)}}(t_{\ell(k)} + \gamma_{\ell(k)})\right) \\
&\quad + \sigma\left(\frac{\overline{\mathrm{v}}_{\ell(k)}}{\overline{\mathrm{v}}_{r(k)} + \overline{\mathrm{v}}_{\ell(k)}}\left(W_{t_{r(k)}} + \Psi_{r(k)}(W)\right) + \frac{\overline{\mathrm{v}}_{r(k)}}{\overline{\mathrm{v}}_{r(k)} + \overline{\mathrm{v}}_{\ell(k)}}\left(W_{t_{\ell(k)}} + \Psi_{\ell(k)}(W)\right)\right) \\
&= x_0 + \mu\left(\frac{\overline{\mathrm{v}}_{\ell(k)}}{\overline{\mathrm{v}}_{r(k)} + \overline{\mathrm{v}}_{\ell(k)}}(t_k + v_{r(k)} + \gamma_{r(k)}) + \frac{\overline{\mathrm{v}}_{r(k)}}{\overline{\mathrm{v}}_{r(k)} + \overline{\mathrm{v}}_{\ell(k)}}(t_k + v_{\ell(k)} + \gamma_{\ell(k)})\right) \\
&\quad + \sigma\left(W_{t_k} + \frac{\overline{\mathrm{v}}_{\ell(k)}}{\overline{\mathrm{v}}_{r(k)} + \overline{\mathrm{v}}_{\ell(k)}}\left(W_{t_{r(k)}} - W_{t_k} + \Psi_{r(k)}(W)\right) + \frac{\overline{\mathrm{v}}_{r(k)}}{\overline{\mathrm{v}}_{r(k)} + \overline{\mathrm{v}}_{\ell(k)}}\left(W_{t_{\ell(k)}} - W_{t_k} + \Psi_{\ell(k)}(W)\right)\right) \\
&= x_0 + \mu(t_k + \gamma_k) + \sigma\left(W_{t_k} + \Psi_k(W)\right),
\end{aligned}$$

by setting

$$\gamma_k = \frac{\overline{\mathrm{v}}_{\ell(k)}(\gamma_{r(k)} + v_{r(k)}) + \overline{\mathrm{v}}_{r(k)}(\gamma_{\ell(k)} + v_{\ell(k)})}{\overline{\mathrm{v}}_{r(k)} + \overline{\mathrm{v}}_{\ell(k)}}, \quad \text{and}$$

$$\Psi_k(W) = \frac{\overline{\mathrm{v}}_{\ell(k)}\left(W_{t_{r(k)}} - W_{t_k} + \Psi_{r(k)}(W)\right) + \overline{\mathrm{v}}_{r(k)}\left(W_{t_{\ell(k)}} - W_{t_k} + \Psi_{\ell(k)}(W)\right)}{\overline{\mathrm{v}}_{r(k)} + \overline{\mathrm{v}}_{\ell(k)}},$$

and the property holds for all nodes k of $\mathcal{T}$.

Proving the second point of the theorem is direct since from Definition 2, we have that

$$\begin{aligned} U_k &= \frac{\overline{\mathrm{Z}}_{r(k)} - \overline{\mathrm{Z}}_{\ell(k)}}{\sqrt{\overline{\mathrm{v}}_{r(k)} + \overline{\mathrm{v}}_{\ell(k)}}} \\ &= \mu \frac{v_{r(k)} + \gamma_{r(k)} - v_{\ell(k)} - \gamma_{\ell(k)}}{\sqrt{\overline{\mathrm{v}}_{r(k)} + \overline{\mathrm{v}}_{\ell(k)}}} + \frac{\sigma}{\sqrt{\overline{\mathrm{v}}_{r(k)} + \overline{\mathrm{v}}_{\ell(k)}}} \Phi_k(W), \end{aligned}$$

where

$$\begin{aligned} \Phi_k(W) &= W_{t_{r(k)}} + \Psi_{r(k)}(W) - W_{t_{\ell(k)}} - \Psi_{\ell(k)}(W) \\ &= (W_{t_{r(k)}} - W_{t_k}) + \Psi_{r(k)}(W) - (W_{t_{\ell(k)}} - W_{t_k}) - \Psi_{\ell(k)}(W), \end{aligned}$$

By construction, for all nodes k, $\Psi_k(W)$ is a linear combination of independent centered Gaussian random variables of the form $W_{t_i} - W_{t_{a(i)}}$ where i is a descendant of k. It follows that $\Psi_k(W)$ is a centered Gaussian variable, which is independent from any random variable $W_{t_j} - W_{t_{a(j)}}$ if j is not a descendant of k. Since $(W_t)_{t>0}$ is the Wiener process, we have that

$$\mathrm{Var}(W_{t_i} - W_{t_{a(i)}}) = t_i - t_{a(i)} = v_i$$

for all nodes i of $\mathcal{T}$(Grimmett and Stirzaker 2001).

Let us prove by induction that

$$\mathrm{Var}\left(\Psi_k(W)\right) = \begin{cases} 0 & \text{if } k \text{ is a tip,} \\ \frac{\overline{\mathrm{v}}_{r(k)}\overline{\mathrm{v}}_{\ell(k)}}{\overline{\mathrm{v}}_{r(k)} + \overline{\mathrm{v}}_{\ell(k)}} & \text{otherwise,} \end{cases}$$

which is equivalent to say that $\mathrm{Var}\left(\Psi_k(W)\right) + v_k = \overline{\mathrm{v}}_k$ from Definition 2. It is basically true in the base case where k is a tip. If k is an internal node, by assuming that the induction assumption holds for its direct descendants $r(k)$ and $\ell(k)$, we have that

$$\begin{aligned} &\mathrm{Var}\left(\Psi_k(W)\right) \\ &= \frac{\overline{\mathrm{v}}_{\ell(k)}^2\left(\mathrm{Var}\left(W_{t_{r(k)}} - W_{t_k}\right) + \mathrm{Var}\left(\Psi_{r(k)}(W)\right)\right) + \overline{\mathrm{v}}_{r(k)}^2\left(\mathrm{Var}\left(W_{t_{\ell(k)}} - W_{t_k}\right) + \mathrm{Var}\left(\Psi_{\ell(k)}(W)\right)\right)}{(\overline{\mathrm{v}}_{r(k)} + \overline{\mathrm{v}}_{\ell(k)})^2} \\ &= \frac{\overline{\mathrm{v}}_{\ell(k)}^2\left(v_{r(k)} + \mathrm{Var}\left(\Psi_{r(k)}(W)\right)\right) + \overline{\mathrm{v}}_{r(k)}^2\left(v_{\ell(k)} + \mathrm{Var}\left(\Psi_{\ell(k)}(W)\right)\right)}{(\overline{\mathrm{v}}_{r(k)} + \overline{\mathrm{v}}_{\ell(k)})^2} \\ &= \frac{\overline{\mathrm{v}}_{\ell(k)}^2\overline{\mathrm{v}}_{r(k)} + \overline{\mathrm{v}}_{r(k)}^2\overline{\mathrm{v}}_{\ell(k)}}{(\overline{\mathrm{v}}_{r(k)} + \overline{\mathrm{v}}_{\ell(k)})^2} = \frac{\overline{\mathrm{v}}_{r(k)}\overline{\mathrm{v}}_{\ell(k)}}{\overline{\mathrm{v}}_{r(k)} + \overline{\mathrm{v}}_{\ell(k)}}, \end{aligned}$$

which proves the form of $\mathrm{Var}\left(\Psi_k(W)\right)$.

In the same way, $\Phi_k(W)$ is a linear combination of independent centered Gaussian random variables of the form $W_i - W_{a(i)}$ where i is a descendant of k, with variance

$$\begin{aligned} \mathrm{Var}\left(\Phi_k(W)\right) &= \mathrm{Var}\left(W_{t_{r(k)}} - W_{t_k}\right) + \mathrm{Var}\left(\Psi_{r(k)}(W)\right) + \mathrm{Var}\left(W_{t_{\ell(k)}} - W_{t_k}\right) + \mathrm{Var}\left(\Psi_{\ell(k)}(W)\right) \\ &= v_{r(k)} + \mathrm{Var}\left(\Psi_{r(k)}(W)\right) + v_{\ell(k)} + \mathrm{Var}\left(\Psi_{\ell(k)}(W)\right) = \overline{\mathrm{v}}_{r(k)} + \overline{\mathrm{v}}_{\ell(k)}. \end{aligned}$$

B Ultrametric trees - Proof of Proposition 5

Let $\mathcal{T}$ be an ultrametric tree and T be the total path-length/time from its root to its tips.

We shall prove by induction that $\gamma_k = T - t_k$ and $h_k = 0$ for all nodes k of $\mathcal{T}$. The property is basically true if k is a tip, our base case. Let k be an internal node and let us assume that the property holds for its direct descendants $r(k)$ and $\ell(k)$. From Theorem 3, we have that

$$\begin{aligned}
\gamma_k &= \frac{\overline{\mathrm{v}}_{\ell(k)}(\gamma_{r(k)} + v_{r(k)}) + \overline{\mathrm{v}}_{r(k)}(\gamma_{\ell(k)} + v_{\ell(k)})}{\overline{\mathrm{v}}_{r(k)} + \overline{\mathrm{v}}_{\ell(k)}} \\
&= \frac{\overline{\mathrm{v}}_{\ell(k)}(T - t_{r(k)} + v_{r(k)}) + \overline{\mathrm{v}}_{r(k)}(T - t_{\ell(k)} + v_{\ell(k)})}{\overline{\mathrm{v}}_{r(k)} + \overline{\mathrm{v}}_{\ell(k)}} \\
&= \frac{\overline{\mathrm{v}}_{\ell(k)}(T - t_k) + \overline{\mathrm{v}}_{r(k)}(T - t_k)}{\overline{\mathrm{v}}_{r(k)} + \overline{\mathrm{v}}_{\ell(k)}} = T - t_k, \quad \text{and,} \\
h_k &= \frac{v_{r(k)} + \gamma_{r(k)} - v_{\ell(k)} - \gamma_{\ell(k)}}{\sqrt{\overline{\mathrm{v}}_{r(k)} + \overline{\mathrm{v}}_{\ell(k)}}} \\
&= \frac{(v_{r(k)} + T - t_{r(k)}) - (v_{\ell(k)} + T - t_{\ell(k)})}{\sqrt{\overline{\mathrm{v}}_{r(k)} + \overline{\mathrm{v}}_{\ell(k)}}} \\
&= \frac{-t_k + t_k}{\sqrt{\overline{\mathrm{v}}_{r(k)} + \overline{\mathrm{v}}_{\ell(k)}}} = 0, \quad \text{which ends the proof.}
\end{aligned}$$

C Trend estimation and detection

Equation 1 shows that under an ABM model with trend μ, the independent contrasts U_k can be written as the product of μ with the corresponding temporal variable h_k, plus an independent, centered Gaussian term of constant variance with respect to k. This suggests to estimate the trend μ as the slope of the following linear equation

$$\mathbf{u} = c_{\mathrm{hR}}\mathbf{h} + \boldsymbol{\varepsilon}_{\mathrm{hR}} \tag{C1}$$

where, from Equation 1, the entries of the error vector $\boldsymbol{\varepsilon}_{\mathrm{hR}}$ are samples of random variables

$$\frac{\sigma \Phi_k(W)}{\sqrt{\overline{\mathrm{v}}_{r(k)} + \overline{\mathrm{v}}_{\ell(k)}}},$$

which are independent and Gaussian distributed with mean zero and variance σ^2 under the current assumptions.

From Equation C1, the linear regression estimator of μ is

$$\widehat{\mu} = \frac{\sum_{j=0}^{n-1} u_j h_j}{\sum_{j=0}^{n-1} h_j^2}. \tag{C2}$$

Again from Equation C1, testing for a trend in the evolution of trait under an ABM model can be performed by testing for the nullity of the slope. In order to assess the accuracy of this trend test (referred to as the "hR test") with regard to the standard regression of the tips values with respect to their times (referred to as the "SR test"), we simulated evolution of quantitative traits with and without trend on the tree of Figure 2 and plot the type I error rate against the rejection level and the ROC plots of these two tests. Results are displayed in Figure C1. The hR test clearly outperforms the SR test. Figure C1-Left shows that the SR test rejects the null hypothesis more than it should do whereas the hR test rejects it at the exact level required. Moreover, their ROC plots show that the hR test better discriminates between traits simulated with and without trend than the SR test (Fig. C1-Right).

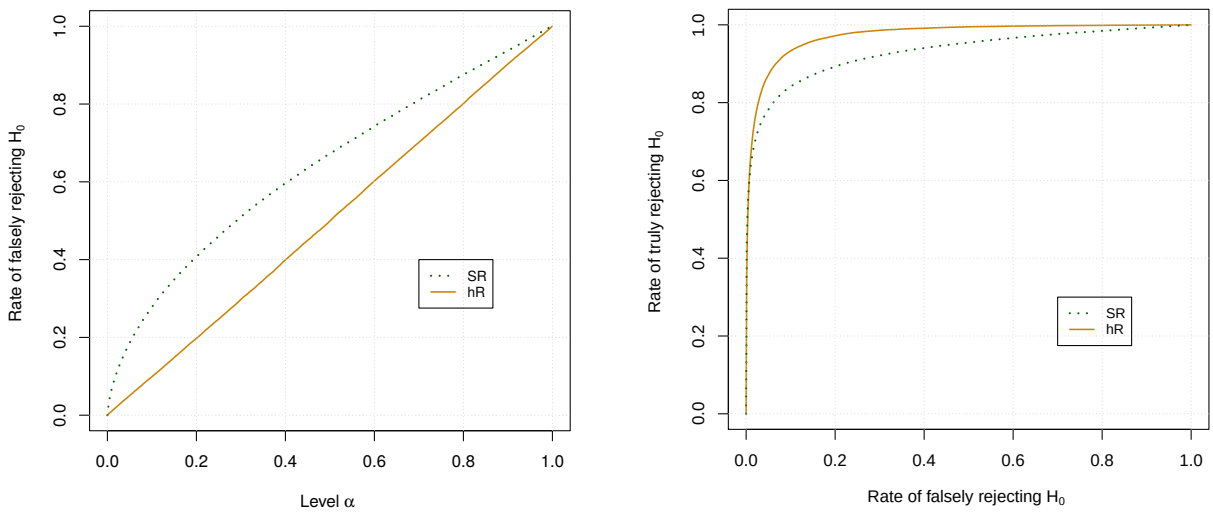

Figure C1: Left: Rates of false rejection of the null hypothesis (i.e., no trend) at level α *vs* α on traits simulated under the BM model with variance 0.09. Right: ROC plots of the trend detection tests obtained from simulated traits under the BM model for negative population (i.e., under H_0) and under the ABM model with trend 0.2 for the positive population, both with variance 0.09.

D Directional contrasts

Equivalence between centered and directional contrasts

Elliott (2015) associated to all nodes k of $\mathcal{T}$ and all values β the quantity $e_k(\beta)$ defined as

$$e_k(\beta) = \begin{cases} z_k & \text{if } k \text{ is a tip,} \\ \dfrac{\overline{\mathrm{v}}_{\ell(k)}(e_{r(k)}(\beta) - \beta v_{r(k)}) + \overline{\mathrm{v}}_{r(k)}(e_{\ell(k)}(\beta) - \beta v_{\ell(k)})}{\overline{\mathrm{v}}_{r(k)} + \overline{\mathrm{v}}_{\ell(k)}} & \text{otherwise,} \end{cases}$$

where the modified branch lengths $(\overline{\mathrm{v}}_k)_{0 \leq k \leq 2n}$ are given in Definition 1.

For all internal nodes k of $\mathcal{T}$, Elliott (2015) then defined the *(β-)directional contrast* as

$$d_k(\beta) = \frac{e_{r(k)}(\beta) - e_{\ell(k)}(\beta)}{\sqrt{\overline{\mathrm{v}}_{r(k)} + \overline{\mathrm{v}}_{\ell(k)}}} - \beta \frac{v_{r(k)} - v_{\ell(k)}}{\sqrt{\overline{\mathrm{v}}_{r(k)} + \overline{\mathrm{v}}_{\ell(k)}}}.$$

Let us start to prove by induction that for all nodes k of $\mathcal{T}$, we have that $e_k(\beta) = \overline{\mathrm{z}}_k - \beta\gamma_k$.

The equality is basically true in the base cases, since if k is a tip of $\mathcal{T}$, we have that $\gamma_k = 0$ and $e_k(\beta) = z_k = \overline{\mathrm{z}}_k$.

Let k be an internal node of $\mathcal{T}$ and let us assume that the equality holds for its two direct descendants $r(k)$ and $\ell(k)$. We have

$$\begin{aligned} e_k(\beta) &= \frac{\overline{\mathrm{v}}_{\ell(k)}(e_{r(k)}(\beta) - \beta v_{r(k)}) + \overline{\mathrm{v}}_{r(k)}(e_{\ell(k)}(\beta) - \beta v_{\ell(k)})}{\overline{\mathrm{v}}_{r(k)} + \overline{\mathrm{v}}_{\ell(k)}} \\ &= \frac{\overline{\mathrm{v}}_{\ell(k)}(\overline{\mathrm{z}}_{r(k)} - \beta\gamma_{r(k)}) + \overline{\mathrm{v}}_{r(k)}(\overline{\mathrm{z}}_{\ell(k)} - \beta\gamma_{\ell(k)})}{\overline{\mathrm{v}}_{r(k)} + \overline{\mathrm{v}}_{\ell(k)}} - \beta \frac{\overline{\mathrm{v}}_{\ell(k)} v_{r(k)} + \overline{\mathrm{v}}_{r(k)} v_{\ell(k)}}{\overline{\mathrm{v}}_{r(k)} + \overline{\mathrm{v}}_{\ell(k)}} \\ &= \frac{\overline{\mathrm{v}}_{\ell(k)}\overline{\mathrm{z}}_{r(k)} + \overline{\mathrm{v}}_{r(k)}\overline{\mathrm{z}}_{\ell(k)}}{\overline{\mathrm{v}}_{r(k)} + \overline{\mathrm{v}}_{\ell(k)}} - \beta \frac{\overline{\mathrm{v}}_{\ell(k)}(v_{r(k)} + \gamma_{r(k)}) + \overline{\mathrm{v}}_{r(k)}(v_{\ell(k)} + \gamma_{\ell(k)})}{\overline{\mathrm{v}}_{r(k)} + \overline{\mathrm{v}}_{\ell(k)}} \\ &= \overline{\mathrm{z}}_k - \beta\gamma_k, \end{aligned}$$

which proves the equality for all nodes of $\mathcal{T}$.

Last, we have that

$$\begin{aligned}
d_k(\beta) &= \frac{e_{r(k)}(\beta) - e_{\ell(k)}(\beta)}{\sqrt{\overline{\mathrm{v}}_{r(k)} + \overline{\mathrm{v}}_{\ell(k)}}} - \beta \frac{v_{r(k)} - v_{\ell(k)}}{\sqrt{\overline{\mathrm{v}}_{r(k)} + \overline{\mathrm{v}}_{\ell(k)}}} \\
&= \frac{\overline{\mathrm{v}}_{\ell(k)}\left(\overline{\mathrm{z}}_{r(k)} - \beta\gamma_{r(k)}\right) + \overline{\mathrm{v}}_{r(k)}\left(\overline{\mathrm{z}}_{\ell(k)} - \beta\gamma_{\ell(k)}\right)}{\overline{\mathrm{v}}_{r(k)} + \overline{\mathrm{v}}_{\ell(k)}} - \beta \frac{\overline{\mathrm{v}}_{\ell(k)} v_{r(k)} + \overline{\mathrm{v}}_{r(k)} v_{\ell(k)}}{\overline{\mathrm{v}}_{r(k)} + \overline{\mathrm{v}}_{\ell(k)}} \\
&= \frac{\overline{\mathrm{v}}_{\ell(k)}\overline{\mathrm{z}}_{r(k)} + \overline{\mathrm{v}}_{r(k)}\overline{\mathrm{z}}_{\ell(k)}}{\overline{\mathrm{v}}_{r(k)} + \overline{\mathrm{v}}_{\ell(k)}} - \beta \frac{\overline{\mathrm{v}}_{\ell(k)}\left(v_{r(k)} + \gamma_{r(k)}\right) + \overline{\mathrm{v}}_{r(k)}\left(v_{\ell(k)} + \gamma_{\ell(k)}\right)}{\overline{\mathrm{v}}_{r(k)} + \overline{\mathrm{v}}_{\ell(k)}} \\
&= u_k - \beta h_k \\
&= c_k(\beta).
\end{aligned}$$

In plain English, for all internal nodes k and all values β, the β-directional contrast $d_k(\beta)$ is equal to the β-centered contrast $c_k(\beta)$.

Estimated directional contrasts

In order to correct the trend effect, Elliott (2015) proposed to consider the $\widehat{\mu}$-directional contrasts where $\widehat{\mu}$ is the estimated trend (Equation C2).

Let B be the random variable associated to the estimated trend. From Equation C2, we have that

$$B = \frac{\sum_{j=0}^{n-1} h_j U_j}{\sum_{j=0}^{n-1} h_j^2}.$$

From Corollary 4, if the trait follows the ABM model of parameters (x_0, μ, σ^2), then for all internal nodes k of $\mathcal{T}$, the independent contrast random variables U_k are Gaussian distributed with

$$U_k \sim \mathcal{N}\left(\mu h_k, \sigma^2\right),$$

which implies that B follows the Gaussian distribution $\mathcal{N}\left(\mu, \sigma^2 / \sum_{k=0}^{n-1} h_k^2\right)$ under the ABM model (x_0, μ, σ^2).

For all internal nodes k of $\mathcal{T}$, the random variable $D_k(B)$ associated to the kth estimated directional contrast is defined as

$$\begin{aligned}
D_k(B) &= U_k - B h_k \\
&= U_k - \frac{\sum_{j=0}^{n-1} h_j U_j}{\sum_{j=0}^{n-1} h_j^2} h_k \\
&= \frac{U_k \sum_{\substack{j=0 \\ j \neq k}}^{n-1} h_j^2 - h_k \sum_{\substack{j=0 \\ j \neq k}}^{n-1} h_j U_j}{\sum_{j=0}^{n-1} h_j^2}.
\end{aligned}$$

Corollary 4 implies that the estimated directional contrast random variables $D_k(B)$ are Gaussian distributed with

$$D_k(B) \sim \mathcal{N}\left(0, \sigma^2 \left[1 - \frac{h_k^2}{\sum_{j=0}^{n-1} h_j^2}\right]\right).$$

Moreover, since the independent contrast random variables U_k are independent from one another, we have that for all $i \neq k$

$$\mathrm{Cov}\left(D_i(B), D_k(B)\right) = \frac{-h_k h_i \sigma^2}{\sum_{j=0}^{n-1} h_j^2}.$$

Since $h_k \neq 0$ in the general case, the estimated directional contrast random variables are neither identically distributed nor independent.

E IC and PGLS regressions – Proof of Theorem 8

In all what follows, $\mathbf{0}$ (resp. $\mathbf{1}$) denotes the column vector with all entries equal to 0 (resp. to 1; their dimensions depending on the context) and for all numbers N, $\mathbf{I}_N$ is the identity matrix of dimension $N \times N$. The transpose of a matrix or a vector A is noted A'. We recall that n is the number of internal nodes of $\mathcal{T}$ which thus has $n+1$ tips.

A matrix presentation of the Felsenstein's (1973) algorithm

Let us sketch a matrix presentation of the Felsenstein's (1973) algorithm which iteratively computes the following variables for all nodes k of $\mathcal{T}$. By putting $\mathcal{T}_k$ for the subtree of $\mathcal{T}$ rooted at k and n_k for its number of internal nodes, let

- $\mathbf{g}_k$ be the vector of dimension $n_k + 1$ that is such that the "artificial" trait value of node k, i.e., $\overline{\mathrm{z}}_k$ of Definition 1, is obtained by multiplying $\mathbf{g}'_k$ with the tip value vector of $\mathcal{T}_k$,
- Q_k be the $n_k \times (n_k + 1)$ matrix giving the independent contrasts of the subtree $\mathcal{T}_k$ from its tip value vector, and,
- δ_k be the increment applied to the branch ending by k (i.e., $\overline{\mathrm{v}}_k - v_k$ in Definition 1).

By initializing $\mathbf{g}_k$ to the vector [1], Q_k to the 0×0 "empty" matrix and δ_k to 0 for all tips k of $\mathcal{T}$, these variables are recursively computed for all internal nodes k of $\mathcal{T}$ with direct descendants $\ell(k)$ and $r(k)$ by setting:

- $\mathbf{g}'_k = \left[\begin{array}{cc} \frac{(v_{r(k)}+\delta_{r(k)})\mathbf{g}'_{\ell(k)}}{v_{\ell(k)}+\delta_{\ell(k)}+v_{r(k)}+\delta_{r(k)}} & \frac{(v_{\ell(k)}+\delta_{\ell(k)})\mathbf{g}'_{r(k)}}{v_{\ell(k)}+\delta_{\ell(k)}+v_{r(k)}+\delta_{r(k)}} \end{array} \right]$,
- $Q_k = \left[\begin{array}{cc} Q_{\ell(k)} & 0 \\ 0 & Q_{r(k)} \\ \frac{\mathbf{g}'_{\ell(k)}}{\sqrt{v_{\ell(k)}+\delta_{\ell(k)}+v_{r(k)}+\delta_{r(k)}}} & \frac{-\mathbf{g}'_{r(k)}}{\sqrt{v_{\ell(k)}+\delta_{\ell(k)}+v_{r(k)}+\delta_{r(k)}}} \end{array} \right]$,
- $\delta_k = \frac{(v_{\ell(k)}+\delta_{\ell(k)})(v_{r(k)}+\delta_{r(k)})}{v_{\ell(k)}+\delta_{\ell(k)}+v_{r(k)}+\delta_{r(k)}}$.

It is straightforward to prove by induction that $\mathbf{g}'_k\mathbf{1} = 1$ for all nodes k of $\mathcal{T}$. Let us prove by induction that $Q'_kQ_k\mathbf{1} = \mathbf{0}$ for all nodes k of $\mathcal{T}$. The property is basically true for all tips. Let us assume that k is an internal node and that its direct descendants $\ell(k)$ and $r(k)$ both satisfy the property. We have

$$Q'_kQ_k = \left[\begin{array}{cc} Q'_{\ell(k)}Q_{\ell(k)} + \frac{\mathbf{g}_{\ell(k)}\mathbf{g}'_{\ell(k)}}{v_{\ell(k)}+\delta_{\ell(k)}+v_{r(k)}+\delta_{r(k)}} & \frac{-\mathbf{g}_{\ell(k)}\mathbf{g}'_{r(k)}}{v_{\ell(k)}+\delta_{\ell(k)}+v_{r(k)}+\delta_{r(k)}} \\ \frac{-\mathbf{g}_{r(k)}\mathbf{g}'_{\ell(k)}}{v_{\ell(k)}+\delta_{\ell(k)}+v_{r(k)}+\delta_{r(k)}} & Q'_{r(k)}Q_{r(k)} + \frac{\mathbf{g}_{r(k)}\mathbf{g}'_{r(k)}}{v_{\ell(k)}+\delta_{\ell(k)}+v_{r(k)}+\delta_{r(k)}} \end{array} \right],$$

thus

$$Q'_kQ_k\mathbf{1} = \left[\begin{array}{c} Q'_{\ell(k)}Q_{\ell(k)}\mathbf{1} + \frac{\mathbf{g}_{\ell(k)}\mathbf{g}'_{\ell(k)}\mathbf{1}}{v_{\ell(k)}+\delta_{\ell(k)}+v_{r(k)}+\delta_{r(k)}} - \frac{\mathbf{g}_{\ell(k)}\mathbf{g}'_{r(k)}\mathbf{1}}{v_{\ell(k)}+\delta_{\ell(k)}+v_{r(k)}+\delta_{r(k)}} \\ \frac{-\mathbf{g}_{r(k)}\mathbf{g}'_{\ell(k)}\mathbf{1}}{v_{\ell(k)}+\delta_{\ell(k)}+v_{r(k)}+\delta_{r(k)}} + Q'_{r(k)}Q_{r(k)}\mathbf{1} + \frac{\mathbf{g}_{r(k)}\mathbf{g}'_{r(k)}\mathbf{1}}{v_{\ell(k)}+\delta_{\ell(k)}+v_{r(k)}+\delta_{r(k)}} \end{array} \right] = \mathbf{0},$$

since $\mathbf{g}'_{r(k)}\mathbf{1} = \mathbf{g}'_{\ell(k)}\mathbf{1} = 1$ from above and $Q'_{\ell(k)}Q_{\ell(k)}\mathbf{1} = Q'_{r(k)}Q_{r(k)}\mathbf{1} = \mathbf{0}$ from the induction hypothesis, which proves that $Q'_kQ_k\mathbf{1} = \mathbf{0}$ for all nodes k of $\mathcal{T}$.

Let us set $\mathbf{g}' = \mathbf{g}'_r$, $Q = Q_r$ and $\delta = \delta_r$ where r is the root of $\mathcal{T}$ and let Σ be the covariance matrix associated to $\mathcal{T}$, i.e., for all pairs of tips (i, j) of $\mathcal{T}$, the (i, j)-entry of Σ is the total time between the root and the most recent common ancestor of i and j. Felsenstein (1973) showed that by assuming that a trait follows a Brownian process with variance σ^2 and by putting $\mathbf{Z}$ for the random vector of its tip values, the contrasts, i.e., the entries of $Q\mathbf{Z}$, are independent centered Gaussian variables with variance σ^2 and that $\mathbf{g}'\mathbf{Z}$ is a centered Gaussian variable with variance $\sigma^2\delta$ which is independent from all the contrasts. It follows that if $\mathbf{Z}$ is a Gaussian vector with covariance matrix $\sigma^2\Sigma$ then $Q\mathbf{Z}$ is a n-Gaussian vector with covariance matrix $\sigma^2\mathbf{I}_n$ and, by defining the $(n + 1) \times (n + 1)$ matrix $P = \left[\begin{array}{c} Q \\ \frac{\mathbf{g}'}{\sqrt{\delta}} \end{array} \right]$, that $P\mathbf{Z}$ is a $(n + 1)$-Gaussian vector with covariance matrix $\sigma^2\mathbf{I}_{n+1}$. Since P is invertible (the standard presentation of the Felsenstein's (1973) algorithm shows that P can be written as a product of invertible matrices), this implies $P'P = \Sigma^{-1}$ and that

$$\Sigma^{-1}\mathbf{1} = P'P\mathbf{1} = Q'Q\mathbf{1} + \frac{\mathbf{g}\mathbf{g}'\mathbf{1}}{\delta} = \frac{\mathbf{g}}{\delta}, \tag{E1}$$

which itself directly implies that

$$(\mathbf{1}'\Sigma^{-1}\mathbf{1})^{-1} = \delta. \tag{E2}$$

Moreover, since Σ^{-1} is symmetric, Equation E1 implies that $\mathbf{g}\mathbf{g}' = \delta^2\Sigma^{-1}\mathbf{1}\mathbf{1}'\Sigma^{-1}$ and that

$$\Sigma^{-1} - \delta\Sigma^{-1}\mathbf{1}\mathbf{1}'\Sigma^{-1} = Q'Q. \tag{E3}$$

Note that matrix P plays the same role as the *phylogenetic transformation matrix* in Adams and Collyer (2015) and as the matrix $\mathbf{D}$ defined in a complete different way in Garland and Ives (2000, p349).

Multiple regression with IC and PGLS

Let $\mathbf{y}$ and X be respectively a vector of tip values of a trait and a matrix of tip values of p traits or co-variables (e.g., the time, an environmental variable etc.). We shall show that testing for correlation between $\mathbf{y}$ and any column of X leads to the same result with the IC and the PGLS methods.

In the IC regression and under the notations of subsection above, we consider the following linear equation

$$Q\mathbf{y} = QX\boldsymbol{\beta}_{\text{IC}} + \boldsymbol{\varepsilon}_{\text{IC}},$$

where the error vector $\boldsymbol{\varepsilon}_{\text{IC}}$ is a realization of the centered Gaussian random vector with covariance matrix proportional to $\mathbf{I}_n$. Vector $Q\mathbf{y}$ has dimension n and there are p regressors. The vector of estimated regression coefficients is

$$\widehat{\boldsymbol{\beta}}_{\text{IC}} = \left(X'Q'QX\right)^{-1} X'Q'Q\mathbf{y}$$

and the estimator variance of the ith coefficient $[\boldsymbol{\beta}_{\text{IC}}]_i$ is $\widehat{\sigma}^2_{\text{IC}}\left[(X'Q'QX)^{-1}\right]_{i,i}$ where the residual variance estimate $\widehat{\sigma}^2_{\text{IC}}$ is

$$\widehat{\sigma}^2_{\text{IC}} = \frac{(\mathbf{y} - X\widehat{\boldsymbol{\beta}}_{\text{IC}})'Q'Q(\mathbf{y} - X\widehat{\boldsymbol{\beta}}_{\text{IC}})}{n-p}.$$

Under the key regression assumptions and that the ith coefficient $[\boldsymbol{\beta}_{\text{IC}}]_i$ is null, the statistics

$$\frac{\left[\widehat{\boldsymbol{\beta}}_{\text{IC}}\right]_i}{\sqrt{\widehat{\sigma}^2_{\text{IC}}\left[(X'Q'QX)^{-1}\right]_{i,i}}}$$

follows the Student distribution with $n-p$ degrees of freedom.

In the PGLS regression, the linear equation is

$$\mathbf{y} = \begin{bmatrix} \mathbf{1} & X \end{bmatrix} \begin{bmatrix} \alpha_{\text{PGLS}} \\ \boldsymbol{\beta}_{\text{PGLS}} \end{bmatrix} + \boldsymbol{\varepsilon}_{\text{PGLS}},$$

where the error vector is a realization of the centered Gaussian vector with covariance matrix proportional to $\Sigma = (P'P)^{-1}$. Vector $\mathbf{y}$ has dimension $(n+1)$ and there are $p+1$ regressors (including the intercept). The vector of estimated regression coefficients is

$$\begin{bmatrix} \widehat{\alpha}_{\text{PGLS}} \\ \widehat{\boldsymbol{\beta}}_{\text{PGLS}} \end{bmatrix} = \left(\begin{bmatrix} \mathbf{1}' \\ X' \end{bmatrix} \Sigma^{-1} \begin{bmatrix} \mathbf{1} & X \end{bmatrix}\right)^{-1} \begin{bmatrix} \mathbf{1}' \\ X' \end{bmatrix} \Sigma^{-1}\mathbf{y}, \quad \text{and,}$$

and the estimator variance of the ith coefficient $[\boldsymbol{\beta}_{\text{PGLS}}]_i$ is

$$\widehat{\sigma}^2_{\text{PGLS}} \left[\left(\begin{bmatrix} \mathbf{1}' \\ X' \end{bmatrix} \Sigma^{-1} \begin{bmatrix} \mathbf{1} & X \end{bmatrix}\right)^{-1}\right]_{i,i}$$

where the residual variance estimate $\widehat{\sigma}^2_{\text{IC}}$ is

$$\widehat{\sigma}^2_{\text{PGLS}} = \frac{\left(\mathbf{y} - \begin{bmatrix} \mathbf{1} & X \end{bmatrix} \begin{bmatrix} \widehat{\alpha}_{\text{PGLS}} \\ \widehat{\boldsymbol{\beta}}_{\text{PGLS}} \end{bmatrix}\right)' \Sigma^{-1} \left(\mathbf{y} - \begin{bmatrix} \mathbf{1} & X \end{bmatrix} \begin{bmatrix} \widehat{\alpha}_{\text{PGLS}} \\ \widehat{\boldsymbol{\beta}}_{\text{PGLS}} \end{bmatrix}\right)}{n-p}.$$

Under the regression assumptions and that the ith coefficient $[\boldsymbol{\beta}_{\text{PGLS}}]_i$ is null, the statistics

$$\frac{\left[\widehat{\boldsymbol{\beta}}_{\text{PGLS}}\right]_i}{\sqrt{\widehat{\sigma}^2_{\text{PGLS}}\left[\left(\begin{bmatrix}\mathbf{1}' \\ X'\end{bmatrix}\Sigma^{-1}\begin{bmatrix}\mathbf{1} & X\end{bmatrix}\right)^{-1}\right]_{i,i}}}$$

follows the Student distribution with $n-p$ degrees of freedom.

In order to prove that the IC and the PGLS methods are equivalent to test for correlation in a multiple regression context, we shall establish that the three following properties hold:

1. $\left(\begin{bmatrix}\mathbf{1}' \\ X'\end{bmatrix}\Sigma^{-1}\begin{bmatrix}\mathbf{1} & X\end{bmatrix}\right)^{-1} = \begin{bmatrix} a & \mathbf{b}' \\ \mathbf{c} & (X'Q'QX)^{-1}\end{bmatrix}$ for a real a and two n-vectors $\mathbf{b}$ and $\mathbf{c}$,
2. $\widehat{\boldsymbol{\beta}}_{\text{PGLS}} = \widehat{\boldsymbol{\beta}}_{\text{IC}}$,
3. $\widehat{\sigma}^2_{\text{PGLS}} = \widehat{\sigma}^2_{\text{IC}}$.

From the block matrix inversion formula and Equations E1, E2 and E3, we get that

$$\left(\begin{bmatrix}\mathbf{1}' \\ X'\end{bmatrix}\Sigma^{-1}\begin{bmatrix}\mathbf{1} & X\end{bmatrix}\right)^{-1} = \begin{bmatrix}\mathbf{1}'\Sigma^{-1}\mathbf{1} & \mathbf{1}'\Sigma^{-1}X \\ X'\Sigma^{-1}\mathbf{1} & X'\Sigma^{-1}X\end{bmatrix}^{-1} = \begin{bmatrix}\delta^{-1} & \mathbf{1}'\Sigma^{-1}X \\ X'\Sigma^{-1}\mathbf{1} & X'\Sigma^{-1}X\end{bmatrix}^{-1}$$

$$= \begin{bmatrix} \delta + \delta^2\mathbf{1}'\Sigma^{-1}X\left(X'\Sigma^{-1}X - \delta X'\Sigma^{-1}\mathbf{1}\mathbf{1}'\Sigma^{-1}X\right)^{-1}X'\Sigma^{-1}\mathbf{1} & -\delta\mathbf{1}'\Sigma^{-1}X\left(X'\Sigma^{-1}X - \delta X'\Sigma^{-1}\mathbf{1}\mathbf{1}'\Sigma^{-1}X\right)^{-1} \\ -\delta\left(X'\Sigma^{-1}X - \delta X'\Sigma^{-1}\mathbf{1}\mathbf{1}'\Sigma^{-1}X\right)^{-1}X'\Sigma^{-1}\mathbf{1} & \left(X'\Sigma^{-1}X - \delta X'\Sigma^{-1}\mathbf{1}\mathbf{1}'\Sigma^{-1}X\right)^{-1}\end{bmatrix}$$

$$= \begin{bmatrix} \delta + \mathbf{g}'X\left(X'Q'QX\right)^{-1}X'\mathbf{g} & -\mathbf{g}'X\left(X'Q'QX\right)^{-1} \\ -\left(X'Q'QX\right)^{-1}X'\mathbf{g} & \left(X'Q'QX\right)^{-1}\end{bmatrix}, \quad \text{which proves Property 1.}$$

The vector $\widehat{\boldsymbol{\beta}}_{\text{PGLS}}$ of the PGLS regression coefficient estimates without the intercept is obtained by multiplying the second line of the block matrix above with the column $\begin{bmatrix}\mathbf{1}'\Sigma^{-1}\mathbf{y} \\ X'\Sigma^{-1}\mathbf{y}\end{bmatrix} = \begin{bmatrix}\frac{\mathbf{g}'}{\delta}\mathbf{y} \\ X'\Sigma^{-1}\mathbf{y}\end{bmatrix}$:

$$\begin{aligned}
\widehat{\boldsymbol{\beta}}_{\text{PGLS}} &= -\left(X'Q'QX\right)^{-1}X'\frac{\mathbf{g}\mathbf{g}'}{\delta}\mathbf{y} + \left(X'Q'QX\right)^{-1}X'\Sigma^{-1}\mathbf{y} \\
&= \left(X'Q'QX\right)^{-1}X'\left(\Sigma^{-1} - \frac{\mathbf{g}\mathbf{g}'}{\delta}\right)\mathbf{y} \\
&= \left(X'Q'QX\right)^{-1}X'Q'Q\mathbf{y} = \widehat{\boldsymbol{\beta}}_{\text{IC}}, \quad \text{which proves Property 2.}
\end{aligned}$$

In the same way, the intercept estimate of PGLS is

$$\begin{aligned}
\widehat{\alpha}_{\text{PGLS}} &= \left(\delta + \mathbf{g}'X\left(X'Q'QX\right)^{-1}X'\mathbf{g}\right)\frac{\mathbf{g}'}{\delta}\mathbf{y} - \mathbf{g}'X\left(X'Q'QX\right)^{-1}X'\Sigma^{-1}\mathbf{y} \\
&= \mathbf{g}'\mathbf{y} + \mathbf{g}'X\left(X'Q'QX\right)^{-1}X'\left(\frac{\mathbf{g}\mathbf{g}'}{\delta} - \Sigma^{-1}\right)\mathbf{y} \\
&= \mathbf{g}'\mathbf{y} - \mathbf{g}'X\left(X'Q'QX\right)^{-1}X'Q'Q\mathbf{y} = \mathbf{g}'(\mathbf{y} - X\widehat{\boldsymbol{\beta}}_{\text{IC}}).
\end{aligned}$$

The residual variance estimate of PGLS is given by

$$\begin{aligned}
\widehat{\sigma}^2_{\text{PGLS}} &= \frac{(\mathbf{y} - \mathbf{1}\widehat{\alpha}_{\text{PGLS}} - X\widehat{\boldsymbol{\beta}}_{\text{PGLS}})'\Sigma^{-1}(\mathbf{y} - \mathbf{1}\widehat{\alpha}_{\text{PGLS}} - X\widehat{\boldsymbol{\beta}}_{\text{PGLS}})}{n-p} \\
&= \frac{(\mathbf{y} - \mathbf{1}\mathbf{g}'(\mathbf{y} - X\widehat{\boldsymbol{\beta}}_{\text{IC}}) - X\widehat{\boldsymbol{\beta}}_{\text{IC}})'\Sigma^{-1}(\mathbf{y} - \mathbf{1}\mathbf{g}'(\mathbf{y} - X\widehat{\boldsymbol{\beta}}_{\text{IC}}) - X\widehat{\boldsymbol{\beta}}_{\text{IC}})}{n-p} \\
&= \frac{(\mathbf{y} - X\widehat{\boldsymbol{\beta}}_{\text{IC}})'\Sigma^{-1}(\mathbf{y} - X\widehat{\boldsymbol{\beta}}_{\text{IC}}) - (\mathbf{y} - X\widehat{\boldsymbol{\beta}}_{\text{IC}})'\mathbf{g}\mathbf{1}'\Sigma^{-1}\mathbf{1}\mathbf{g}'(\mathbf{y} - X\widehat{\boldsymbol{\beta}}_{\text{IC}})}{n-p} \\
&= \frac{(\mathbf{y} - X\widehat{\boldsymbol{\beta}}_{\text{IC}})'\left(\Sigma^{-1} - \frac{\mathbf{g}\mathbf{g}'}{\delta}\right)(\mathbf{y} - X\widehat{\boldsymbol{\beta}}_{\text{IC}})}{n-p} \\
&= \frac{(\mathbf{y} - X\widehat{\boldsymbol{\beta}}_{\text{IC}})'Q'Q(\mathbf{y} - X\widehat{\boldsymbol{\beta}}_{\text{IC}})}{n-p} = \widehat{\sigma}^2_{\text{IC}}, \quad \text{which proves Property 3,}
\end{aligned}$$

and ends to prove that testing for correlation with IC and PGLS is equivalent.

Multiple regression with time as co-variable

Let us start by showing that for all internal nodes k, h_k is the phylogenetic contrast of the tip times associated to k. By putting $\bar{t}_k$ for the "artificial" time reconstructed at the internal node k by the Felsenstein's (1973) algorithm, it is straightforward to prove by induction that $\bar{t}_k = t_k + \gamma_k$ for all nodes k of $\mathcal{T}$. For all internal nodes k with direct descendants $r(k)$ and $\ell(k)$, we have that

$$h_k = \frac{v_{r(k)} + \gamma_{r(k)} - v_{\ell(k)} - \gamma_{\ell(k)}}{\sqrt{\bar{\mathrm{v}}_{r(k)} + \bar{\mathrm{v}}_{\ell(k)}}} = \frac{\bar{t}_{r(k)} - \bar{t}_{\ell(k)}}{\sqrt{\bar{\mathrm{v}}_{r(k)} + \bar{\mathrm{v}}_{\ell(k)}}},$$

which is well the contrast of the tip times associated to node k. It follows that the vector $\mathbf{h} = (h_k)_{1 \leq k < n}$ is obtained by multiplying the vector $\mathbf{t} = (t_k)_{n \leq k \leq 2n}$ of tip times by Q, i.e., $\mathbf{h} = Q\mathbf{t}$.

Let us consider two traits A and B and their tip-value vectors $\mathbf{z}_A = (z_k^A)_{n \leq k \leq 2n}$ and $\mathbf{z}_B = (z_k^B)_{n \leq k \leq 2n}$. The linear equation

$$Q\mathbf{z}_B = Q \begin{bmatrix} \mathbf{z}_A & \mathbf{t} \end{bmatrix} \begin{bmatrix} b_{\mathrm{MR}} \\ c_{\mathrm{MR}} \end{bmatrix} + \boldsymbol{\varepsilon}_{\mathrm{MR}} = b_{\mathrm{MR}}\mathbf{u}_A + c_{\mathrm{MR}}\mathbf{h} + \boldsymbol{\varepsilon}_{\mathrm{MR}},$$

where the error vector $\boldsymbol{\varepsilon}_{\mathrm{MR}}$ is a realization of a centered Gaussian vector with covariance matrix proportional to $\mathbf{I}_n$, corresponds to the regression considered in the MR method (Equation 8). We showed in the section above that testing for correlation between traits A and B with this equation is equivalent to testing for correlation between A and B with the linear equation

$$\mathbf{z}_B = \begin{bmatrix} \mathbf{1} & \mathbf{z}_A & \mathbf{t} \end{bmatrix} \begin{bmatrix} a_{\mathrm{PGLSt}} \\ b_{\mathrm{PGLSt}} \\ c_{\mathrm{PGLSt}} \end{bmatrix} + \boldsymbol{\varepsilon}_{\mathrm{PGLSt}},$$

where the error vector $\boldsymbol{\varepsilon}_{\mathrm{PGLSt}}$ is a realization of a centered Gaussian vector with covariance matrix proportional to Σ which corresponds to testing for correlation between A and B with tip times as covariables by the PGLS method.