

A Semiparametric Method to Test for Correlated Evolution in a Phylogenetic Context

Luke J. Harmon¹ and Liam J. Revell²

¹Department of Biological Sciences, University of Idaho, Moscow, ID, USA

²Department of Biology, University of Massachusetts Boston, Boston, MA, USA

Corresponding author:

Luke J. Harmon¹

Email address: lukeh@uidaho.edu

ABSTRACT

Phylogenetic comparative methods are a broad suite of approaches for studying trait and species diversification using evolutionary trees. In spite of their extraordinary growth in sophistication and complexity in recent years, among the most commonly-employed phylogenetic comparative method is the simple measure of the evolutionary correlation between variables, while accounting for the statistical non-independence in our data that arises from common descent. The standard parametric phylogenetic approach for measuring the evolutionary correlation between continuously-valued characters assumes a model called Brownian motion for the evolution of our traits. Here, we introduce a new semi-parametric method which relaxes this assumption by testing for the evolutionary correlation between variables based on ranks, and then obtaining a null distribution on the test statistic via random permutation. We show that this approach has reasonable statistical properties: type I error close to the nominal level, and power that's similar to fully parametric methods. We conclude by comparing our new method to related approaches.

1 INTRODUCTION

Statistical phylogenetic comparative methods have become an invaluable tool in contemporary evolutionary research (Felsenstein 1985; Revell and Harmon 2022). This group of methodologies comprises a large and growing set of approaches for studying increasingly sophisticated evolutionary hypotheses of trait and species diversification using phylogenetic trees (reviewed in Harvey and Pagel 1991; Nunn 2011; O'Meara 2012; Harmon 2019; Revell and Harmon 2022). Nonetheless, among the most common evolutionary hypotheses studied by researchers is the simple question of whether one trait is correlated with another, while accounting for the statistical non-independence of the species that make up our data points due to their shared histories (Felsenstein 1985; reviewed in Harmon 2019; Revell and Harmon 2022).

When measuring and undertaking a hypothesis test of this nature, phylogenetic comparative methods typically assume a particular underlying model of trait evolution, such as the Brownian motion model (Felsenstein 1973; Felsenstein 1985; O'Meara et al. 2006). Brownian motion (also called the Weiner process) is a continuous-time stochastic diffusion in which the expected value is constant and the variance increases as a linear function of time (multiplied by the Brownian motion rate, typically denoted σ^2 , O'Meara et al. 2006; Harmon 2019; Revell and Harmon 2022). In particular, under the phylogenetically independent contrasts algorithm of Felsenstein (1985), both the internal node values in the contrasts algorithm, and the variances of the (non-standardized) contrasts are calculated assuming a Brownian evolutionary process on the tree (Felsenstein 1985; Harmon 2019). Likewise, under phylogenetic generalized least squares regression (Grafen 1989), we assume that either the data themselves or their residual errors are autocorrelated in a manner consistent with Brownian motion evolution or a close relative (e.g., Pagel 1999; Garland and Ives 2000; Rohlf 2001; Revell 2010). Though there are a number of circumstances under which continuously-valued phenotypic traits might genuinely evolve via a Brownian

process (e.g., Lynch 1990), violations of the Brownian model are likely to be common in real data sets (Freckleton and Harvey 2006; Hunt 2006; Estes and Arnold 2007; Uyeda et al. 2011). Brownian motion has nonetheless persisted as one of the most commonly-used trait evolution models for studying evolutionary correlations, largely due to its simplicity, tractability, and apparent robustness to deviations from model assumptions under some circumstances (e.g., Garland et al. 1993; Stone 2011).

A number of different approaches have been devised to identify violations of Brownian motion in our data while measuring evolutionary correlations, and then to account for these deviations statistically.

First, we can apply various diagnostic tests for independent contrasts that are sensitive to a variety of violations of the Brownian model (e.g., Garland 1992; Blomberg et al. 2003; Freckleton and Harvey 2006). For example, if the true model is Brownian motion, then standardized independent contrasts of Felsenstein (1985) should be independently and identically distributed (i.i.d.). If, on the other hand, standardized contrasts are correlated with their expected variances (that is, the amount of evolutionary branch length that subtends each contrast), this suggests that there may be more or less evolutionary change along long branches of the tree (depending on the sign of the correlation), than would have been expected under a Brownian process. We might then attempt various transformations of the branch lengths of the tree so that our contrasts no longer violate the test (Garland 1992).

Another similar and closely-related diagnostic is one that's referred to as the node-height test (Freckleton and Harvey 2006). Here, we measure for a relationship between our standardized contrasts and the height above the root of the node over which each contrast was calculated (Freckleton and Harvey 2006). Once again, this correlation should be statistically indistinguishable from zero if the Brownian motion model assumptions hold.

Both the measures of Garland (1992) and Freckleton and Harvey (2006)'s node-heights test can also be interpreted as measurements of model adequacy. If they fail (that is, if we reject a null hypothesis of no correlation between the standardized contrast and its variance or node height), then this suggests that our underlying assumption of Brownian motion may be invalid (Pennell et al. 2015).

Under circumstances in which this is true (i.e., when Brownian motion is rejected via a diagnostic test), we might then proceed to consider one or more of a relatively limited set of alternative trait evolution models (e.g., Pagel 1999). In particular, we might proceed to fit our correlation assuming a different evolutionary scenario that is nonetheless Gaussian, such as the Ornstein-Uhlenbeck or a speciation trait evolution models (e.g., Hansen 1997; Martins and Hansen 1997). Alternatively, we could transform the branches of the phylogenetic tree or the elements of the variance-covariance matrix derived from the tree (Grafen 1989; Garland 1992; Pagel 1999). Nonetheless, there is no general solution to the problem of evolutionary model assumption violations, and on occasion there may simply be no way to coax our data sets or tree to fit the particular assumptions of our analyses.

In statistics, nonparametric tests can sometimes provide robustness in cases where models are not known or provide poor fits to data (e.g., Wasserman 2006). As comparative methods are employed to tackle a wider and wider range of problems, from genomics to cancer biology to evolutionary studies at the broadest scale, there is a growing need for methods that are robust to model violations. Among existing comparative methods, nonparametric methods are already available for the analysis of discrete characters (e.g., Maddison 1990), and for morphometric shape data (Adams 2014). To date, however, the only available nonparametric comparative method designed specifically for measuring the correlation of continuously-valued traits is Ackerly's (2000) contrast sign test (CST). This test involves comparing the signs (positive or negative) of phylogenetically independent contrasts for two characters (Ackerly 2000). If the two characters are evolving independently, then the number of pairs of contrasts with the same sign should follow a binomial distribution with probability of success $p = 0.5$. The CST provides a robust alternative to parametric comparative methods, but is not commonly used. Its primary drawback is that the test can have very low power, in part because it relies only on the contrast signs, and discards all information about their relative magnitudes.

Here we describe a new semiparametric comparative method, the phylogenetic rank correlation (PRC) test. This test, which is based on the correlation of ranks of independent contrasts on a phylogenetic tree, is robust under several models of trait evolution, and has power comparable to standard independent contrasts and much greater than the CST. We argue that the PRC test thus may provide a robust alternative to parametric methods such as PGLS and independent contrasts.

2 DESCRIPTION OF THE METHOD

The PRC test has a total of six steps (Figure 1), as follows.

1. Calculate standardized independent contrasts (Felsenstein 1985) for two continuously distributed traits.
2. Reflect the contrasts, multiplying both the X and Y value of each pair of contrasts in the two traits by -1 . The data will now include two ‘copies’ of each X, Y pair of contrasts, one with the signs of each contrast reversed.
3. Create ranks independently for each variable. For tied scores, assign the mean rank to all tied contrasts.
4. Calculate the test statistic. Here, we use Spearman’s rank correlation coefficient (Spearman’s ρ , Spearman 1904), but we could also use Kendall’s τ (Kendall 1938).
5. Generate the null distribution of the test statistic via randomization of the contrasts. To do this, permute the independent contrasts for one variable. Then, create new ranks, and calculate the test statistic. Repeat this permutation procedure a large number of times (say, 9,999) to generate a null distribution. (We can either permute the doubled contrasts or permute the original contrasts and double them – the result will be the same.)
6. Compare the test statistic with the null distribution from permutation to obtain a P-value of the test, calculated as $2 \times$ the portion of permuted test-statistic values (for a two-tailed test) that are equally or more extreme than our measured value of ρ (or τ).

Several of these steps require some additional comment or explanation.

For step one, standardized contrasts are calculated. The calculation of these contrasts, as originally described by Felsenstein (1985), assumes a Brownian motion model of evolution; however, since the standardized contrasts are to be converted into ranks, we maintain that violating this assumption will have a much smaller effect on their subsequent analysis, than in a fully parametric approach (see below). In practice, by converting contrasts into ranks, we have assumed only that contrasts subtending longer branches of the tree tend to be larger in absolute magnitude. This assumption is compatible with a broad range of evolutionary scenarios, as long as variance among species tends to increase through time (see Estes and Arnold 2007).

In step two, reflecting the contrasts is critical because the direction of any particular contrast is arbitrary. (This can be understood via the thought experiment in which we rotate descendant edges around any node of the tree. Positively-valued contrasts in X and Y will become negative, and the reverse.) Reflecting contrasts across the axes ensures that our analysis is unchanged by rotating the nodes in the tree. This procedure is the statistical equivalent of forcing the regression through the origin, which is required when testing for correlations of independent contrasts (Garland et al. 1992). We do not have to worry about inflating degrees of freedom by doubling the number of data points because statistical significance is to be determined by permutations.

Finally, in step four, any nonparametric test statistic could be used for this test (see, e.g., Wasserman 2006). We used Spearman’s rank correlation (Spearman’s ρ), which is simply the Pearson product-moment correlation of the ranks of the two variables (Spearman 1904). Kendall’s τ is another viable option (Kendall 1938).

We provide a worked example of the calculations involved in this test in Figure 1. Software to carry out this test, written in the R statistical computing language (R Core Team 2024) and will be made publicly available as part of the GEIGER R package (Harmon et al. 2007; Pennell et al. 2014).

3 STATISTICAL PROPERTIES OF THE METHOD

To explore the statistical properties of our semiparametric method, we measured type I error and power, and compared both to the same measures for both non-phylogenetic (i.e., OLS or ‘standard’ regression) and related phylogenetic comparative methods (independent contrasts, contrasts with a statistical distribution obtained via permutation, and the contrasts sign test of Ackerly 2000).

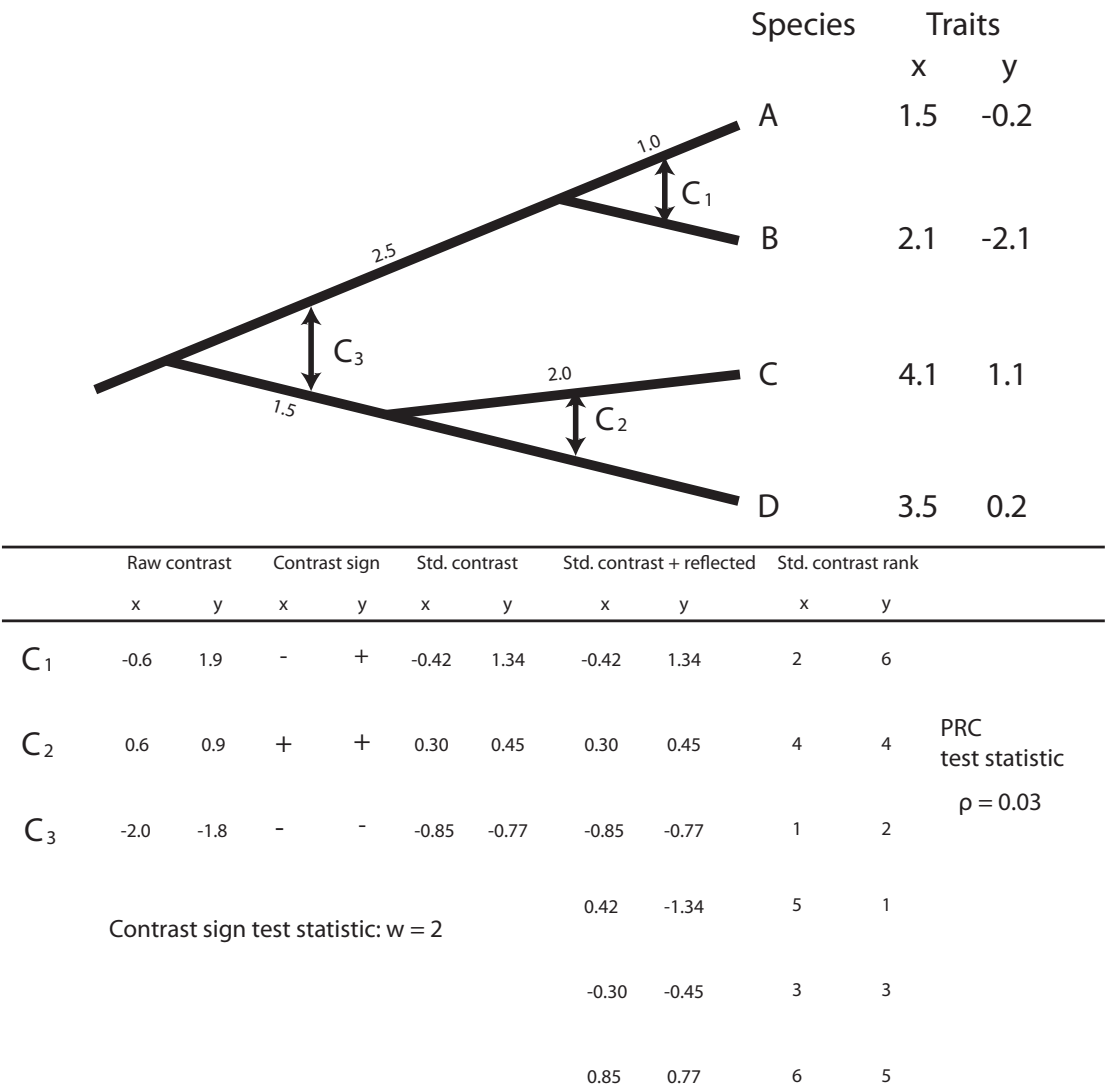


Figure 1. Illustration of the method of phylogenetic rank contrasts, PRC. First, compute the raw contrasts. For a bifurcation tree with N tips there should be $N - 1$ such contrasts. Next, standardize the contrasts by their expected variance in the typical way following Felsenstein (1985). Duplicate all pairs of contrasts for each of our two variable but swapping the signs. Convert all contrasts (and their reflections) to ranks. Calculate the test-statistic. See main text for more details.

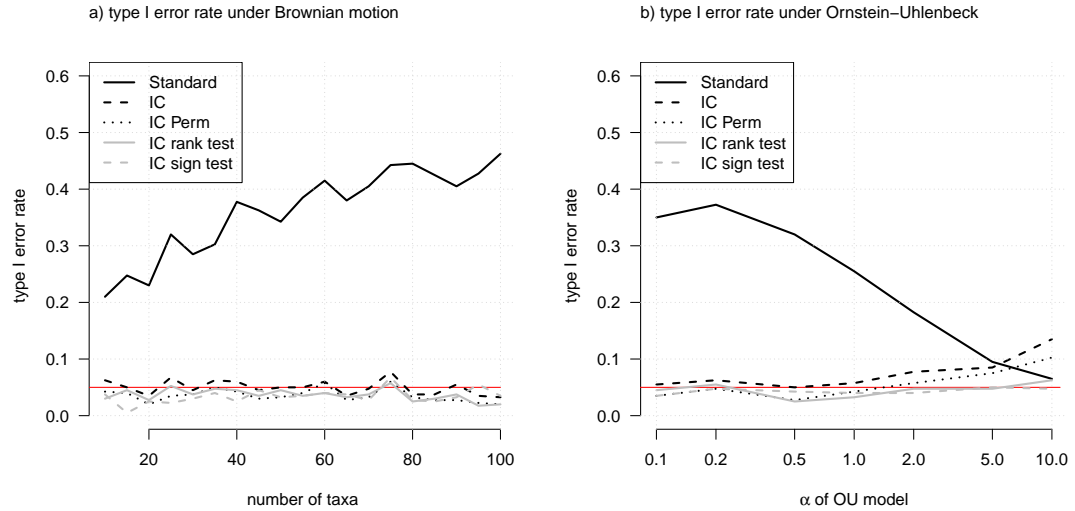


Figure 2. Type I error results.

To measure type I error we began by simulating data for X and Y under Brownian motion evolution and with no evolutionary correlation between the two traits. For this simulation, we first generated 400 pure-birth (i.e., Yule) phylogenetic trees containing from 10 to 100 taxa on 5 taxa intervals (i.e., 10, 15, 20, etc.). These trees were rescaled to a total height of 1.0, and each of X and Y were simulated (independently, as they are uncorrelated) with a Brownian motion rate of $\sigma^2 = 1.0$. We fit a standard (OLS) regression, a contrasts regression, a contrasts regression but in which a P-value was obtained via random permutation, the contrasts sign test, and our phylogenetic rank correlation. For each analysis we computed a P-value and we estimated the type I error rate as the fraction of P-values from each analysis ≤ 0.05 . Results from this test are given in Figure 2a. Consistent with prior work, we found high type I error of OLS regression for phylogenetic data simulated under Brownian motion (Figure 2a), at a rate that increased with increasing numbers of taxa, N (Figure 2a). By contrast, all phylogenetic analyses had type I error rates close to the nominal level of $\alpha = 0.05$ (Figure 2a).

Using simulated trees from the prior analysis with $N = 50$, we next proceeded to simulated uncorrelated data for X and Y , but this time under an Ornstein-Uhlenbeck, rather than Brownian, process. The Ornstein-Uhlenbeck process is a stochastic property similar to Brownian motion, but in which the process has a tendency to revert towards a central value. The strength of this tendency is determined by the parameter α . Higher values of α correspond to a greater tendency to revert towards the central value. The parameter α can also be understood in terms of “phylogenetic half-life” (Hansen et al. 2008), $\ln(2)/\alpha$, which measures the time a trait is expected to evolve from its ancestral value to the optimum of the Ornstein-Uhlenbeck process. Conversely, a value of $\alpha = 0$ (and consequently a phylogenetic half-life of $\ln(2)/\alpha = \infty$) corresponds to Brownian motion evolutionary change. We simulated our data for X and Y for values of $\alpha = 0.1, 0.2, 0.5, 1, 2, 5$, and 10, and then repeated the same non-phylogenetic and phylogenetic correlation analyses as described previously. Results for this test are shown in Figure 2b. In general, we found that type I error of the phylogenetic regression was highest for low α and decreased with increasing α (Figure 2b). Conversely, type I error of all phylogenetic methods was close to or equal to the nominal level of $\alpha = 0.05$ for the lowest values of (Ornstein-Uhlenbeck) α , but increased for higher α . This increase, however, was lowest for the contrasts sign test and contrasts rank test (of this article), and highest for standard independent contrasts and permutation-based contrasts regression (Figure 2b).

Taken together, these results indicate that our semiparametric contrasts rank correlation method has type I error under Brownian motion that is virtually identical to existing parametric and non-parametric phylogenetic evolutionary correlation tests (Figure 2a). Our results also show, however, that the phylogenetic rank correlation test has lower type I error when the assumption of Brownian motion evolution is violated by our data (Figure 2b).

In addition to measuring type I error, we also tested the power of the method to detect a statistically

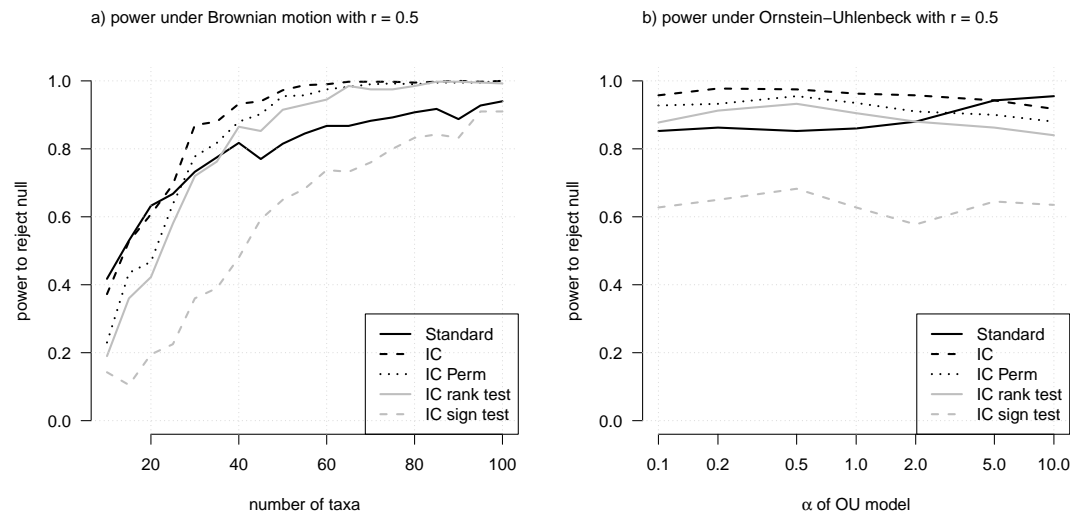


Figure 3. Power analysis results.

significant correlation where one existed. To this end, we simulated data for X and Y using that same simulation conditions as for Figure 2, described above, but in which the data arose under an evolutionary correlation between the traits of $r = 0.5$. The results for this analysis are shown in Figure 3.

When X and Y were simulated under Brownian motion, we found that standard contrasts and contrasts with a statistical test via permutation had the highest power; however, they were followed closely by our semiparametric phylogenetic rank correlation (Figure 3a). Ackerly (2000)'s method had the lowest power of all methods (Figure 3a). When our correlated data for X and Y were simulated under Ornstein-Uhlenbeck a similar pattern repeats in which contrasts, a permutation test using contrasts, standard (non-phylogenetic) correlation, and our phylogenetic rank correlation have the highest power – much higher than the power of Ackerly (2000)'s contrast sign test (Figure 3b).

4 DISCUSSION

We have characterized this method as “semiparametric” because it is not truly distribution-free. One key component of the method involves calculating standardized independent contrasts, which assumes that the characters are evolving under a Brownian motion model. However, by converting these contrasts to ranks, this method greatly lessens the effects of violations of this assumption for testing correlations between characters. In fact, to preserve contrast rank order, one must assume that changes on each branch are independent, and that there is more change on longer branches in the tree than along short branches. Violations of this assumption will lead to increased type I error of this test. One example of violations that would have negative impacts on all tests described here (standard regression, independent contrasts, CST, and PST) is rate heterogeneity across the tree (O’Meara et al., 2006). For example, if evolution is faster along certain branches in the tree, nearby contrasts values and signs will be correlated, and all methods relying on their independence will exhibit inflated type I error.

5 CONCLUSIONS

Since the model of evolution underlying continuous characters is frequently unknown, and may depart from Brownian motion, we recommend using this method to test correlations between characters. If standard contrasts are well-behaved, then standard techniques are still useful for estimating slopes and other parameter values. However, this semiparametric test provides a method for generating robust P-values across a range of models. The technique is nearly as powerful as independent contrasts but much more robust to deviations from the underlying model of evolution.

6 DATA AVAILABILITY

This article was written in Rmarkdown (Xie et al. 2018, 2020; Allaire et al. 2023), and developed with the help of both *bookdown* (Xie 2016, 2023) and the posit Rstudio IDE (RStudio Team 2020). All data and markdown code necessary to exactly rebuild the submitted version of this article (including its analyses and figures) are available at <https://github.com/liamrevell/Revell-and-Harmon-NonParametricPCM>.

7 ACKNOWLEDGMENTS

We thank J. J. Kolbe, J. B. Losos, D. Schluter, R. E. Glor, B. Banbury, T. Hagey, C. Brock, and H. Alamillo for providing helpful feedback on this paper.

LITERATURE CITED

- Ackerly, D. D. 2000. Taxon sampling, correlated evolution, and independent contrasts. *Evolution* 54:1480–1492. Wiley.
- Adams, D. C. 2014. A method for assessing phylogenetic least squares models for shape and other high-dimensional multivariate data. *Evolution* 68:2675–2688. Wiley.
- Allaire, J., Y. Xie, J. McPherson, J. Luraschi, K. Ushey, A. Atkins, H. Wickham, J. Cheng, W. Chang, and R. Iannone. 2023. Rmarkdown: Dynamic documents for R.
- Blomberg, S. P., T. Garland, and A. R. Ives. 2003. Testing for phylogenetic signal in comparative data: Behavioral traits are more labile. *Evolution* 57:717–745. Wiley.
- Estes, S., and S. J. Arnold. 2007. Resolving the paradox of stasis: Models with stabilizing selection explain evolutionary divergence on all timescales. *Am. Nat.* 169:227–244.
- Felsenstein, J. 1973. Maximum-likelihood estimation of evolutionary trees from continuous characters. *American Journal of Human Genetics* 25:471–492.
- Felsenstein, J. 1985. Phylogenies and the comparative method. *Am. Nat.* 125:1–15.
- Freckleton, R. P., and P. H. Harvey. 2006. Detecting non-brownian trait evolution in adaptive radiations. *PLoS Biol.* 4:e373. Public Library of Science (PLOS).
- Garland, Jr., and A. R. Ives. 2000. Using the past to predict the present: Confidence intervals for regression equations in phylogenetic comparative methods. *Am. Nat.* 155:346–364. University of Chicago Press.
- Garland, T., Jr. 1992. Rate tests for phenotypic evolution using phylogenetically independent contrasts. *Am. Nat.* 140:509–519.
- Garland, T., A. W. Dickerman, C. M. Janis, and J. A. Jones. 1993. Phylogenetic analysis of covariance by computer simulation. *Syst. Biol.* 42:265–292. Oxford Academic.
- Garland, T., P. H. Harvey, and A. R. Ives. 1992. Procedures for the analysis of comparative data using phylogenetically independent contrasts. *Systematic Biology* 41:18–32. Oxford University Press (OUP).
- Grafen, A. 1989. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences* 326:119–157. The Royal Society.
- Hansen, T. F. 1997. Stabilizing selection and the comparative analysis of adaptation. *Evolution* 51:1341–1351. Wiley.
- Hansen, T. F., J. Pienaar, and S. H. Orzack. 2008. A comparative method for studying adaptation to a randomly evolving environment. *Evolution* ???–??? Wiley.
- Harmon, L. J. 2019. Phylogenetic comparative methods: Learning from trees. *Ecoevorxiv*.
- Harmon, L. J., J. T. Weir, C. D. Brock, R. E. Glor, and W. Challenger. 2007. GEIGER: Investigating evolutionary radiations. *Bioinformatics* 24:129–131. Oxford University Press (OUP).
- Harvey, P. H., and M. D. Pagel. 1991. *The comparative method in evolutionary biology*. Oxford University Press, Oxford, UK.
- Hunt, G. 2006. Fitting and comparing models of phyletic evolution: Random walks and beyond. *Paleobiology* 32:578–601. The Paleontological Society.
- Kendall, M. G. 1938. A new measure of rank correlation. *Biometrika* 30:81–93. Oxford University Press (OUP).
- Lynch, M. 1990. The rate of morphological evolution in mammals from the standpoint of the neutral expectation. *Am. Nat.* 136:727–741.

- 262 Maddison, W. P. 1990. A method for testing the correlated evolution of two binary characters: Are gains
263 or losses concentrated on certain branches of a phylogenetic tree? *Evolution* 44:539–557. Society for
264 the Study of Evolution.
- 265 Martins, E. P., and T. F. Hansen. 1997. Phylogenies and the comparative method: A general approach to
266 incorporating phylogenetic information into the analysis of interspecific data. *The American Naturalist*
267 149:646–667. University of Chicago Press.
- 268 Nunn, C. L. 2011. *The comparative approach in evolutionary anthropology and biology*. University of
269 Chicago Press.
- 270 O’Meara, B. C. 2012. Evolutionary inferences from phylogenies: A review of methods. *Annual Review*
271 *of Ecology, Evolution, and Systematics* 43:267–285. *Annual Reviews*.
- 272 O’Meara, B. C., C. Ané, M. J. Sanderson, and P. C. Wainwright. 2006. Testing for different rates of
273 continuous trait evolution using likelihood. *Evolution* 60:922–933. Wiley.
- 274 Pagel, M. 1999. The maximum likelihood approach to reconstructing ancestral character states of discrete
275 characters on phylogenies. *Systematic Biology* 48:612–622. Oxford University Press (OUP).
- 276 Pennell, M. W., J. M. Eastman, G. J. Slater, J. W. Brown, J. C. Uyeda, R. G. FitzJohn, M. E. Alfaro, and
277 L. J. Harmon. 2014. Geiger v2.0: An expanded suite of methods for fitting macroevolutionary models
278 to phylogenetic trees. *Bioinformatics* 30:2216–2218. Oxford University Press (OUP).
- 279 Pennell, M. W., R. G. FitzJohn, W. K. Cornwell, and L. J. Harmon. 2015. Model adequacy and the
280 macroevolution of angiosperm functional traits. *Am. Nat.* 186:E33–E50. University of Chicago Press
281 Chicago, IL.
- 282 R Core Team. 2024. *R: A language and environment for statistical computing*. R Foundation for Statistical
283 Computing, Vienna, Austria.
- 284 Revell, L. J. 2010. Phylogenetic signal and linear regression on species data. *Methods Ecol. Evol.*
285 1:319–329. Wiley.
- 286 Revell, L. J., and L. J. Harmon. 2022. *Phylogenetic comparative methods in R*. Princeton University
287 Press, Princeton, NJ.
- 288 Rohlf, F. J. 2001. Comparative methods for the analysis of continuous variables: Geometric interpretations.
289 *Evolution* 55:2143–2160. Wiley.
- 290 RStudio Team. 2020. *RStudio: Integrated development environment for R*. RStudio, PBC., Boston, MA.
- 291 Spearman, C. 1904. The proof and measurement of association between two things. *The American*
292 *Journal of Psychology* 15:72. University of Illinois Press.
- 293 Stone, E. A. 2011. Why the phylogenetic regression appears robust to tree misspecification. *Systematic*
294 *Biology* 60:245–260. Oxford University Press (OUP).
- 295 Uyeda, J. C., T. F. Hansen, S. J. Arnold, and J. Pienaar. 2011. The million-year wait for macroevolutionary
296 bursts. *Proc. Natl. Acad. Sci. U. S. A.* 108:15908–15913.
- 297 Wasserman, L. 2006. *All of nonparametric statistics*. Springer Science & Business Media, New York, NY.
- 298 Xie, Y. 2023. *Bookdown: Authoring books and technical documents with r markdown*.
- 299 Xie, Y. 2016. *Bookdown: Authoring books and technical documents with R markdown*. Chapman;
300 Hall/CRC, Boca Raton, Florida.
- 301 Xie, Y., J. J. Allaire, and G. Golemund. 2018. *R markdown: The definitive guide*. Chapman; Hall/CRC,
302 Boca Raton, Florida.
- 303 Xie, Y., C. Dervieux, and E. Riederer. 2020. *R markdown cookbook*. Chapman; Hall/CRC, Boca Raton,
304 Florida.