A semiparametric method to test for correlated evolution in a phylogenetic context

Luke J. Harmon[1] **and Liam J. Revell**[2]

[1]**Department of Biological Sciences, University of Idaho, Moscow, ID, USA**
[2]**Department of Biology, University of Massachusetts Boston, Boston, MA, USA**

Corresponding author:
Luke J. Harmon[1]

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 substantial growth in sophistication and complexity in recent years, among the most commonly-employed phylogenetic comparative method continues to be a 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 semiparametric method that relaxes this assumption by testing for the evolutionary correlation between variables based on contrast ranks, and then obtains 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 is 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 hypothesis studied by researchers using statistical comparative methods is the simple question of whether one trait is correlated with another – while accounting for the non-independence of the species that make up our data points due to their shared histories (Felsenstein 1985; reviewed in Nunn 2011; 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 a Wiener process) is a continuous-time stochastic diffusion process 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 (Felsenstein 1973; Felsenstein 2004; 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, 2004; Harmon 2019). Likewise, under phylogenetic generalized least squares regression (PGLS, Grafen 1989; Rohlf 2001), we assume that either the data themselves, or their residual errors, are autocorrelated in a manner consistent with Brownian motion evolution– or a closely-related process (e.g., Pagel 1999; Garland and Ives 2000; Rohlf 2001; Revell 2010). Although starting from different assumptions, PICs and PGLS are

equivalent (Blomberg et al. 2012). Though there are a number of circumstances in which continuously valued phenotypic traits might genuinely evolve via a Brownian process (e.g., Lynch 1990), violations of the Brownian model are likely be common in real data sets (Blomberg et al. 2003; 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; Rohlf 2001; Stone 2011).

Nonetheless, 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 violations of the Brownian model (e.g., Garland 1992; Blomberg et al. 2003; Freckleton and Harvey 2006). For example, if the generative model of the data 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 phylogenetic contrast), this suggests that there may be either more or less evolutionary change along long branches of the phylogeny (depending on the sign of the correlation) than would have been expected under a Brownian process (Garland 1992). We might then attempt various transformations of the branch lengths of the tree so that our contrasts no longer violate this test (Garland 1992).

Another similar and closely-related diagnostic is 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 assumption holds for our data.

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 Brownian motion has been 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; Revell 2010). In particular, we may proceed to fit our correlation assuming a different evolutionary model, such as the Ornstein-Uhlenbeck or a speciational 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; Revell 2010). Nevertheless, 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 set or tree to fit the assumptions of our analysis!

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 precisely this kind of robust approach. Among existing comparative methods, nonparametric approaches are already available for the analysis of discrete traits (e.g., Maddison 1990), and for morphometric shape data (Adams 2014). To date, however, the only non- or semiparametric comparative method designed specifically for measuring the correlation of continuously valued characters 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 fully 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, that we call the phylogenetic rank correlation (PRC) test. This test, which is based on the correlation of ranks of independent contrasts on a phylogenetic tree, seems to be robust to substantial deviations from a strict assumption of Brownian

motion evolution of our traits, but nonetheless has power comparable to standard independent contrasts and much greater than the CST. Consequently, we argue that the PRC test may provide a robust alternative to fully parametric methods, such as PGLS and independent contrasts, when the Brownian motion evolution model is clearly violated for our data (or their residual errors, Rohlf 2001; Revell 2010).

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, X and Y. These contrasts each consist of two numeric values (a contrast for X and a contrast for Y), and they should number $\mathbf{C}_1$, $\mathbf{C}_2$, through $\mathbf{C}_m$ for m internal nodes in the tree (Figure 1).

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. Thus, for example, if contrast $\mathbf{C}_1$ consists of the numeric values $[-0.6, 1.9]$, we would create an additional contrast in our data set, $\mathbf{C}_{m+1}$, that was equal to $[0.6, -1.9]$ (Figure 1), and then proceed to do the same for all m contrasts in the tree. This reflection is necessary due to the arbitrary direction of subtraction chosen when calculating contrasts, as discussed below.

3. Create ranks independently for the numeric values of the contrasts for each variable, X and Y (Figure 1). In the case of tied values, we should assign the mean rank to all contrast values in the tie.

4. Calculate the rank correlation test statistic. Here, we use Spearman's rank correlation coefficient (Spearman's ρ, Spearman 1904), but we could have also use Kendall's τ (Kendall 1938).

5. Generate a null distribution for the test statistic via randomization of the contrasts. To do this, we must permute the numeric values of the undoubled independent contrasts for one variable, X or Y (it doesn't matter which). Then we reflect the contrasts, create new ranks, and calculate the test statistic. (Note that the ranks of the variable that is not permuted will not change!) We repeat this permutation procedure a large number of times (say, 999 or 9,999), each time recalculating the test statistic (ρ or τ) to obtain a null distribution.

6. Compare the test statistic with the null distribution from permutation to obtain a P-value of the test. We calculate our P-value as $2\times$ the fraction 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 might require some additional comment or explanation. For step (1.), standard-ized 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 (2.), 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 or Y will become negative, as well as the converse.) Reflecting contrasts around the origin ensures that our analysis is unchanged by node rotations of the tree. This procedure is likewise conceptually equivalent to forcing our regression line through the point $[0,0]$, 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 (4.), any nonparametric test statistic could be used for this test (see, e.g., Wasserman 2006). For simplicity, we have chosen to use 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

Species	Traits	
	x	y
A	1.5	-0.2
B	2.1	-2.1
C	4.1	1.1
D	3.5	0.2

Tree branch lengths: 1.0 (C_1), 2.5, 1.5 (C_3), 2.0 (C_2)

	Raw contrast		Contrast sign		Std. contrast		Std. contrast + reflected		Std. contrast rank	
	x	y	x	y	x	y	x	y	x	y
C_1	-0.6	1.9	−	+	-0.42	1.34	-0.42	1.34	2	6
C_2	0.6	0.9	+	+	0.30	0.45	0.30	0.45	4	4
C_3	-2.0	-1.8	−	−	-0.85	-0.77	-0.85	-0.77	1	2
							0.42	-1.34	5	1
							-0.30	-0.45	3	3
							0.85	0.77	6	5

Contrast sign test statistic: w = 2
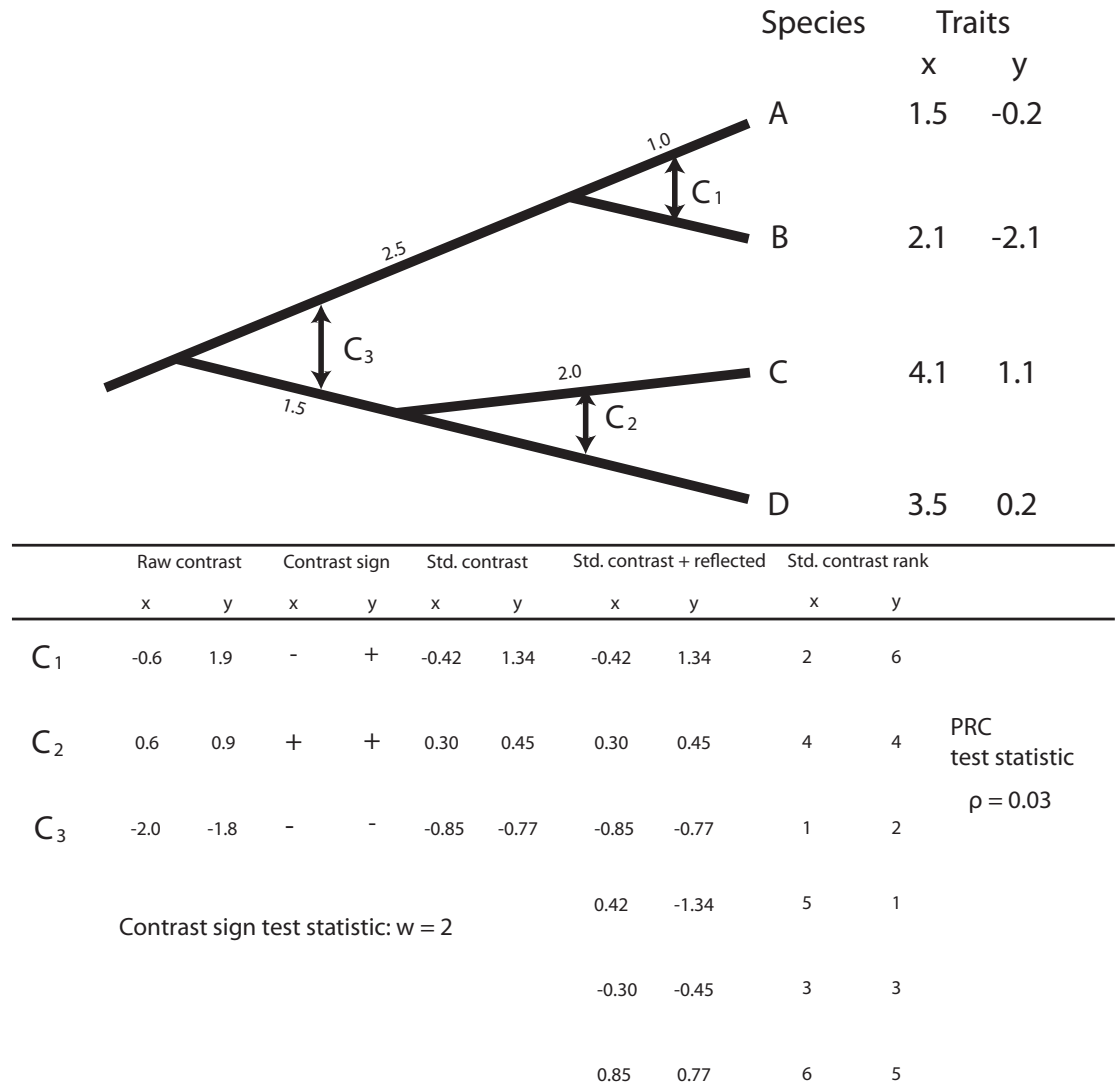
PRC test statistic

$\rho = 0.03$

Figure 1. Illustration of the method of phylogenetic rank contrasts, PRC. First, we compute raw independent contrasts in the standard way. For a bifurcating tree with N tips there should be $m = N - 1$ such contrasts, in which m is also the number of nodes in the tree. Next, we standardize the contrasts by their expected variance under Brownian motion following Felsenstein (1985). We duplicate all pairs of contrasts for each of our two variable, but swap the signs (i.e., rotating around the origin, $[0,0]$). Next, we convert all contrasts (and their reflections) to ranks. Finally, we calculate the test-statistic, in this case, Spearman's ρ. A P-value for ρ is obtained via permutation of the contrasts for either X or Y (it doesn't matter which). See main text for more details.

149 (Kendall 1938) is another viable option that could be substituted with no other changes to the described
150 methodology.

151   We provide a worked example of the calculations involved in this test in Figure 1. Software to carry
152 out the PRC test, written in the R statistical computing language (R Core Team 2024), will be made
153 publicly available as part of the popular *geiger* R phylogenetics package (Harmon et al. 2008; Pennell et
154 al. 2014).

3 STATISTICAL PROPERTIES OF THE METHOD

156 To explore the statistical properties of our semiparametric evolutionary correlation test, the PRC, we
157 measured type I error and power of the test under a range of circumstances. Then we compared both to
158 the same measures for both non-phylogenetic (i.e., OLS or 'standard' non-phylogenetic correlation) and
159 related phylogenetic comparative methods (independent contrasts, contrasts with a statistical distribution
160 obtained via permutation, and the contrasts sign test of Ackerly 2000).

161   To measure type I error we simulated data for X and Y under Brownian motion evolution and with no
162 evolutionary correlation between the two traits. For this simulation, we first began by generating 1,000
163 pure-birth (i.e., Yule) phylogenetic trees containing from $N = 10$ to 100 taxa on ten taxon intervals (i.e.,
164 $N = 10, 20, 30$, etc.), an exercise that resulted in a total of $9 \times 1,000 = 9,000$ simulated trees. These trees
165 were rescaled to have a total height of 1.0, and each of X and Y were simulated (independently, as they
166 are uncorrelated) with a Brownian motion rate of $\sigma^2 = 1.0$.

167   Using this set of 9,000 trees and simulated datasets for X and Y, we then proceeded to fit a standard
168 (OLS) regression ("Standard"), a contrasts regression ("IC"), a contrasts regression but in which a P-value
169 for hypothesis testing was obtained via random permutation ("IC Perm"), the contrasts sign test ("CST"),
170 and our phylogenetic rank correlation ("PRC"). We decided to include a permutation test based on
171 phylogenetically independent contrasts to ensure that our measured type I error and power for the PRC
172 were due to the use of a rank correlation, rather than for having substituted a permutation distribution for
173 the parametric distribution typically used in phylogenetic correlation with independent contrasts. For each
174 analysis we computed a P-value, and then we estimated the type I error rate as the fraction of P-values
175 from each set of simulation conditions and analysis with values equal to or smaller than 5%.

176   Results from this analysis are given in Figure 2a. Consistent with expectations (e.g., Revell 2010), we
177 found high type I error of standard, non-phylogenetic correlation for phylogenetic data simulated under
178 Brownian motion (Figure 2a), at a rate that increased with increasing numbers of taxa, N (Figure 2a). By
179 contrast, when the assumption of Brownian motion evolution held true, all phylogenetic analyses had type
180 I error rates close to the nominal level of $\alpha = 0.05$ (Figure 2a).

181   We next proceeded to simulate uncorrelated data for X and Y, but this time under an Ornstein-
182 Uhlenbeck, rather than Browian, process. The Ornstein-Uhlenbeck process is a stochastic process similar
183 to Brownian motion, but in which there exists a tendency to revert towards some central value of the
184 trait (Hansen 1997; Butler and King 2004). (As such Ornstein-Uhlenbeck is sometimes referred to as
185 Brownian motion with a 'rubber-band,' in which we can imagine an elastic rubber bands that tends to
186 pull the stochastic process towards its tether-point.) The strength of this tendency is determined by the
187 model parameter α (not to be confused with our nominal type I error rate, also denoted as α). Higher
188 values of α correspond to a greater tendency to revert towards the central value. The parameter α can
189 also be understood in terms of "phylogenetic half-life" (Hansen et al. 2008), $\ln(2)/\alpha$, which is equal to
190 the average time (under the process) that a trait should require to evolve from its ancestral value halfway
191 to the optimum of the Ornstein-Uhlenbeck process. Conversely, a value of $\alpha = 0$ (and consequently a
192 phylogenetic half-life of $\ln(2)/\alpha = \infty$) corresponds to Brownian motion evolutionary change.

193   Using only the set of 1,000 simulated trees with $N = 50$ from our prior analysis, we simulated data
194 for X and Y under an Ornstein-Uhlenbeck model with values of $\alpha = 0.1, 0.2, 0.5, 1, 2, 5, 10, 20, 50$, and
195 100. This resulted in a total of $10 \times 1,000 = 10,000$ additional simulated datasets. We then repeated the
196 same non-phylogenetic and phylogenetic correlation analyses as were described in the prior section.

197   Results for this test are shown in Figure 2b. In general, we found that type I error of the non-
198 phylogenetic (i.e., standard) correlation was highest for low α and decreased with increasing α (Figure
199 2b). This is fairly unsurprising because low α corresponds closely to Brownian motion in which a
200 standard correlation is unreliable; while (by contrast) very high α tends to reduce the autocorrelation
201 of species data which, in the limit as $\alpha \to \infty$, will lose all phylogenetic autocorrelation altogether and
202 thus satisfy the assumptions of standard OLS. Conversely, type I error of all phylogenetic methods was
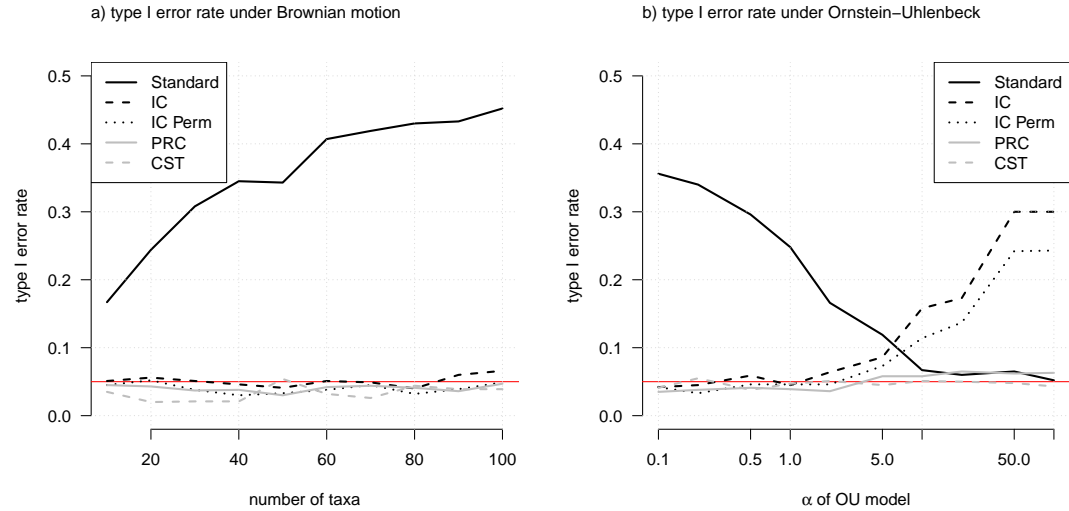
Figure 2. Results of the analysis of type I error. (a) Type I error of both a standard (non-phylogenetic) correlation, and various phylogenetic correlations, when data for *X* and *Y* were simulated under a Brownian motion model, for various total number of taxa, *N*. (b) Type I error for standard and phylogenetic correlation tests when data were simulated under an Ornstein-Uhlenbeck process, for various values of α. Legend labels are as follows: "Standard," non-phylogenetic correlation; "IC," correlation test based on contrasts; "IC Perm," same as "IC," but in which a null distribution of the test statistic was obtained via permutation of the contrasts; "PRC," phylogenetic rank correlation (of this article); and "CST," contrasts sign test of Ackerly (2000). See main text for additional details.

close 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 smallest for the CST and PRC (of this article), and highest for standard independent contrasts and permutation-based contrasts regression (Figure 2b). In particular, type I error at the very *highest* level of α was actually slightly below 5% for the CST, 6.3% for the PRC, and around 30% for contrasts regression.

Taken together, these results indicate that our semiparametric contrasts rank correlation method (PRC) 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 can have much lower type I error than parametric methods when the strict assumption of Brownian motion evolution is violated by our data (Figure 2b).

In addition to measuring type I error, we also wanted to test the power of the PRC method to detect a statistically significant correlation where one exists. To this end, we commenced by simulating data for *X* and *Y* using exactly the same conditions as for Figure 2, described above, but using an evolutionary correlation between the traits of $r = 0.5$. Since the simulations in every other way paralleled our type I error test, the total number of data sets generated for this analysis was $10,000 + 9,000 = 19,000$. Power for each simulation condition and method was scored as the fraction of P-values than were equal to or fell below 5%. The results for this analysis are given in Figure 3.

When *X* and *Y* were simulated under Brownian motion, we found that standard contrasts (IC) and contrasts with a statistical test via permutation (IC Perm) had the highest power; however, these methods were followed closely by our semiparametric PRC (Figure 3a). Ackerly's (2000) CST 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 standard contrasts (IC), a permutation test using contrasts (IC Perm), standard (non-phylogenetic) correlation, and our phylogenetic rank correlation (PRC) have the highest power – much higher than the power of the CST by Ackerly (2000; Figure 3b).

Lastly, to explore the influence of effect size on statistical power, we used the same set of 1,000 simulated trees employed for the previous analysis in which the number of taxa $N = 50$. We first simulated correlated Brownian motion evolution on each set of trees while varying the evolutionary correlation
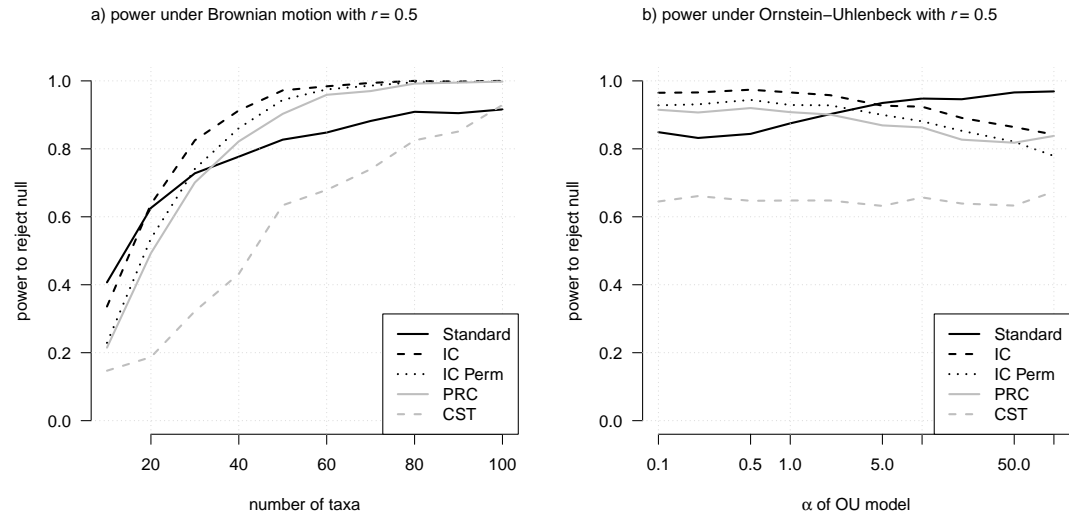
Figure 3. Results of the analysis of power for a true underlying evolution correlation between X and Y of $r = 0.5$. (a) Power of both a standard (non-phylogenetic) correlation, and various phylogenetic correlations, when data for X and Y were simulated under a Brownian motion model, for various total number of taxa, N. (b) Power of standard and phylogenetic correlation tests when data were simulated under an Ornstein-Uhlenbeck process, for various values of α. Legend as in Figure 2.

coefficient, r, between $r = 0$ and 1.0 on the interval 0.1 ($r = 0$, 0.1, 0.2, and so on). We next repeated the same simulation, but using Ornstein-Uhlenbeck with $\alpha = 10$ for simulation. In total, we thus generated a total of $11 \times 1,000 \times 2 = 22,000$ simulated data sets for this analysis.

For each simulation condition, we measured power as the fraction of statistical tests from each simulation condition rejecting the null hypothesis of no correlation (although, of course, this was just a re-measurement of the type I error for $r = 0$). Results for these analyses are shown in Figure 4. In general, when data were simulated under Brownian motion, our PRC test had highly similar power to contrasts regression (Figure 4a). The CRT had much lower power over all simulation conditions (Figure 4a). Standard, non-phylogenetic correlation had high "power" for low values of r; however, this actually reflects the much higher type I error of non-phylogenetic methods that we showed in Figure 2a. By contrast, for higher values of r the power of non-phylogenetic analyses falls substantially below both the contrasts method and our PRC. Likewise, for data simulated under an Ornstein-Uhlenbeck model with $\alpha = 10$, the PRC method of this article exhibited power across all effect sizes that was substantively higher than the CRT and comparable to both standard (non-phylogenetic) and contrasts-based correlation methods (Figure 4b). Consistent with our analysis of Figure 2b, however, the PRC also showed *lower* "power" (reflecting its lower type I error, and thus a desirable attribute of the method) for the lowest r values of $r = 0$ and 0.1 (Figure 4b).

4 DISCUSSION

Phylogenetic comparative methods have become an indispensable tool of evolutionary research (reviewed in O'Meara 2012; Harmon 2019; Revell and Harmon 2022), and are increasingly employed across a wide variety of different research fields, from community ecology, to cancer biology, to infectious disease epidemiology (e.g., Peterson et al. 2021; Gao et al. 2022; Magalis et al. 2024). Even as they grow in complexity and sophistication, a simple test of the evolutionary correlation of two variables (while accounting for statistical non-independence arising from common descent) remains among the most popular phylogenetic comparative analyses. Typical methods for measuring the evolutionary correlation between traits, however, rely on an assumption of Brownian motion evolution of the variables or their residual error (Felsenstein 1985; Garland et al. 1992).

In this contribution, we introduce a new semiparametric phylogenetic rank correlation (PRC) test that seeks to relax this assumption. We have characterized this method as "semiparametric" because it
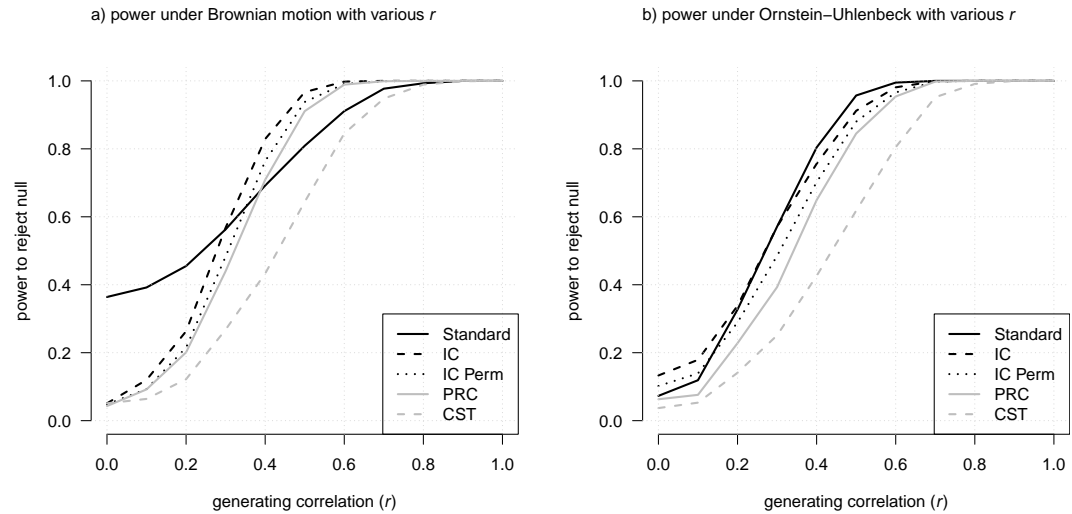
a) power under Brownian motion with various *r*   b) power under Ornstein–Uhlenbeck with various *r*

Figure 4. Results of an analysis of power, for various evolutionary correlations between *X* and *Y*. (a) Power of both a standard (non-phylogenetic) correlation, and various phylogenetic correlations, when data for *X* and *Y* were simulated under a Brownian motion model, for evolutionary correlations, *r*. (b) Power of standard and phylogenetic correlation tests when data were simulated under an Ornstein-Uhlenbeck process with $\alpha = 10$, for various values of the evolutionary correlation, *r*. Legend is as in Figures 2 and 3.

is not truly distribution-free. Indeed, the method begins via the calculation of standardized independent contrasts (see Figure 1). Standardization is undertaken by dividing each raw contrast by its expected variance, a step that implicitly assumes our traits have evolved under a Brownian process (Felsenstein 1985)! Nonetheless, we assert, and the results of our simulation analyses show, that by converting these contrasts to ranks, the PRC method substantially lessens the effects of violations of this assumption for testing correlations between characters (Figures 2, 3, 4). This is probably because by transforming the standardized contrasts into ranks, our assumption of Brownian evolution effectively changes into the much weaker assumption that longer branches of the tree should contain more evolution of the trait (on average) than do shorter branches. Though this is likely to be true across a much broader swathe of evolutionary processes than is strict Brownian motion evolution, it's not difficult to envision circumstances in which this assumption (too) is violated. Indeed, we would predict that strong heterogeneity in the rate of evolution (e.g., O'Meara et al. 2006) will lead to elevated type I error of the PRC, just as it would to all of the other phylogenetic methods explored in this article. In a case were substantial heterogeneity of the rate of evolution is thought to exist, we would recommend a method designed to explicitly account for this variation (e.g., Revell and Collar 2009; Caetano and Harmon 2018).

5 CONCLUSIONS

Phylogenetic comparative methods are often employed to measure the correlation between continuously-valued species characteristics. Here, we provide a robust semiparametric correlation test that has fewer assumptions than do traditional approaches, but higher power than competing non- or semiparametric methods (e.g., Ackerly 2000). We propose that this method be considered under condition in which the statistical significance of an evolutionary correlation between traits is of interest, but where the satisfaction of model assumptions cannot be assured.

6 CODE AND DATA AVAILABILITY

All analyses of this article were undertaken in R (R Core Team 2024), using the contributed R packages *ape* (Paradis and Schliep 2019), *geiger* (Pennell et al. 2014), and *phytools* (Revell 2024). Parallelization of simulations was accomplished using the *foreach* package (Microsoft and Weston 2022). Manuscript drafts of the article were 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) can be found at https://github.com/liamrevell/Harmon-and-Revell.NonParametricPCM.

7 ACKNOWLEDGMENTS

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.

Blomberg, S. P., J. G. Lefevre, J. A. Wells, and M. Waterhouse. 2012. Independent contrasts and PGLS regression estimators are equivalent. Systematic Biology 61:382–391. Oxford University Press (OUP).

Butler, M. A., and A. A. King. 2004. Phylogenetic comparative analysis: A modeling approach for adaptive evolution. The American Naturalist 164:683–695. University of Chicago Press.

Caetano, D. S., and L. J. Harmon. 2018. Estimating correlated rates of trait evolution with uncertainty. Systematic Biology 68:412–429. Oxford University Press (OUP).

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. 2004. Inferring phylogenies. Sinauer associates, Sunderland, MA.

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).

Gao, Y., J. Gaither, J. Chifman, and L. Kubatko. 2022. A phylogenetic approach to inferring the order in which mutations arise during cancer progression. PLOS Computational Biology 18:e1010560. 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 62:1965–1977. 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. 2008. GEIGER: Investigating evolutionary radiations. Bioinformatics 24:129–131. Oxford University Press (OUP).

338 Harvey, P. H., and M. D. Pagel. 1991. The comparative method in evolutionary biology. Oxford University
339   Press, Oxford, UK.
340 Hunt, G. 2006. Fitting and comparing models of phyletic evolution: Random walks and beyond. Paleobi-
341   ology 32:578–601. The Paleontological Society.
342 Kendall, M. G. 1938. A new measure of rank correlation. Biometrika 30:81–93. Oxford University Press
343   (OUP).
344 Lynch, M. 1990. The rate of morphological evolution in mammals from the standpoint of the neutral
345   expectation. Am. Nat. 136:727–741.
346 Maddison, W. P. 1990. A method for testing the correlated evolution of two binary characters: Are gains
347   or losses concentrated on certain branches of a phylogenetic tree? Evolution 44:539–557. Society for
348   the Study of Evolution.
349 Magalis, B. R., A. Riva, S. Marini, M. Salemi, and M. Prosperi. 2024. Novel insights on unraveling
350   dynamics of transmission clusters in outbreaks using phylogeny-based methods. Infection, Genetics
351   and Evolution 124:105661. Elsevier BV.
352 Martins, E. P., and T. F. Hansen. 1997. Phylogenies and the comparative method: A general approach to
353   incorporating phylogenetic information into the analysis of interspecific data. The American Naturalist
354   149:646–667. University of Chicago Press.
355 Microsoft, and S. Weston. 2022. Foreach: Provides foreach looping construct.
356 Nunn, C. L. 2011. The comparative approach in evolutionary anthropology and biology. University of
357   Chicago Press.
358 O'Meara, B. C. 2012. Evolutionary inferences from phylogenies: A review of methods. Annual Review
359   of Ecology, Evolution, and Systematics 43:267–285. Annual Reviews.
360 O'Meara, B. C., C. Ané, M. J. Sanderson, and P. C. Wainwright. 2006. Testing for different rates of
361   continuous trait evolution using likelihood. Evolution 60:922–933. Wiley.
362 Pagel, M. 1999. The maximum likelihood approach to reconstructing ancestral character states of discrete
363   characters on phylogenies. Systematic Biology 48:612–622. Oxford University Press (OUP).
364 Paradis, E., and K. Schliep. 2019. Ape 5.0: An environment for modern phylogenetics and evolutionary
365   analyses in R. Bioinformatics 35:526–528.
366 Pennell, M. W., J. M. Eastman, G. J. Slater, J. W. Brown, J. C. Uyeda, R. G. FitzJohn, M. E. Alfaro, and
367   L. J. Harmon. 2014. Geiger v2.0: An expanded suite of methods for fitting macroevolutionary models
368   to phylogenetic trees. Bioinformatics 30:2216–2218. Oxford University Press (OUP).
369 Pennell, M. W., R. G. FitzJohn, W. K. Cornwell, and L. J. Harmon. 2015. Model adequacy and the
370   macroevolution of angiosperm functional traits. Am. Nat. 186:E33–E50. University of Chicago Press
371   Chicago, IL.
372 Peterson, K., M. Ruffley, and C. E. Parent. 2021. Phylogenetic diversity and community assembly in a
373   naturally fragmented system. Ecology and Evolution 11:18066–18080. Wiley.
374 R Core Team. 2024. R: A language and environment for statistical computing. R Foundation for Statistical
375   Computing, Vienna, Austria.
376 Revell, L. J. 2010. Phylogenetic signal and linear regression on species data. Methods Ecol. Evol.
377   1:319–329. Wiley.
378 Revell, L. J. 2024. phytools 2.0: An updated R ecosystem for phylogenetic comparative methods (and
379   other things). PeerJ 12:e16505. PeerJ.
380 Revell, L. J., and D. C. Collar. 2009. Phylogenetic analysis of the evolutionary correlation using likelihood.
381   Evolution 63:1090–1100. Wiley.
382 Revell, L. J., and L. J. Harmon. 2022. Phylogenetic comparative methods in R. Princeton University
383   Press, Princeton, NJ.
384 Rohlf, F. J. 2001. Comparative methods for the analysis of continuous variables: Geometric interpretations.
385   Evolution 55:2143–2160. Wiley.
386 RStudio Team. 2020. RStudio: Integrated development environment for R. RStudio, PBC., Boston, MA.
387 Spearman, C. 1904. The proof and measurement of association between two things. The American
388   Journal of Psychology 15:72. University of Illinois Press.
389 Stone, E. A. 2011. Why the phylogenetic regression appears robust to tree misspecification. Systematic
390   Biology 60:245–260. Oxford University Press (OUP).
391 Uyeda, J. C., T. F. Hansen, S. J. Arnold, and J. Pienaar. 2011. The million-year wait for macroevolutionary
392   bursts. Proc. Natl. Acad. Sci. U. S. A. 108:15908–15913.

393 Wasserman, L. 2006. All of nonparametric statistics. Springer Science & Business Media, New York, NY.

394 Xie, Y. 2023. bookdown: Authoring books and technical documents with R markdown.

395 Xie, Y. 2016. bookdown: Authoring books and technical documents with R markdown. Chapman;
396   Hall/CRC, Boca Raton, Florida.

397 Xie, Y., J. J. Allaire, and G. Grolemund. 2018. R markdown: The definitive guide. Chapman; Hall/CRC,
398   Boca Raton, Florida.

399 Xie, Y., C. Dervieux, and E. Riederer. 2020. R markdown cookbook. Chapman; Hall/CRC, Boca Raton,
400   Florida.