Syst. Biol. 73(1):140–157, 2024

https://doi.org/10.1093/sysbio/syad070
Advance Access Publication November 30, 2023

Robust Phylogenetic Regression

Richard Adams[1,2,*], Zoe Cain[3], Raquel Assis[4,5,†], and Michael DeGiorgio[4,†]

[1]*Department of Entomology and Plant Pathology, University of Arkansas, Fayetteville, AR, USA,*
[2]*Agricultural Statistics Laboratory, University of Arkansas, Fayetteville, AR, USA,*
[3]*Department of Biological and Environmental Sciences, Georgia College, Milledgeville, GA, USA,*
[4]*Department of Electrical Engineering and Computer Science, Florida Atlantic University, Boca Raton, FL, USA and*
[5]*Institute for Human Health and Disease Intervention, Florida Atlantic University, Boca Raton, FL, USA*
[*]*Correspondence to be sent to: Department of Entomology and Plant Pathology, University of Arkansas, Fayetteville, AR, USA; E-mail: adamsrh@uark.edu.*
†*Equal contributions.*

Received 25 August 2022; reviews returned 16 November 2023; accepted 28 November 2023
Associate Editor: Josef Uyeda

Abstract.—Modern comparative biology owes much to phylogenetic regression. At its conception, this technique sparked a revolution that armed biologists with phylogenetic comparative methods (PCMs) for disentangling evolutionary correlations from those arising from hierarchical phylogenetic relationships. Over the past few decades, the phylogenetic regression framework has become a paradigm of modern comparative biology that has been widely embraced as a remedy for shared ancestry. However, recent evidence has shown doubt over the efficacy of phylogenetic regression, and PCMs more generally, with the suggestion that many of these methods fail to provide an adequate defense against unreplicated evolution—the primary justification for using them in the first place. Importantly, some of the most compelling examples of biological innovation in nature result from abrupt lineage-specific evolutionary shifts, which current regression models are largely ill equipped to deal with. Here we explore a solution to this problem by applying robust linear regression to comparative trait data. We formally introduce robust phylogenetic regression to the PCM toolkit with linear estimators that are less sensitive to model violations than the standard least-squares estimator, while still retaining high power to detect true trait associations. Our analyses also highlight an ingenuity of the original algorithm for phylogenetic regression based on independent contrasts, whereby robust estimators are particularly effective. Collectively, we find that robust estimators hold promise for improving tests of trait associations and offer a path forward in scenarios where classical approaches may fail. Our study joins recent arguments for increased vigilance against unreplicated evolution and a better understanding of evolutionary model performance in challenging—yet biologically important—settings. [Brownian motion; gene expression; linear regression; phylogenetics; quantitative traits; trait evolution.]

Since Darwin's time, biologists have struggled to understand the evolutionary dynamics among organisms and their traits that have collectively shaped present-day biodiversity. One of the most interesting research avenues in modern comparative biology is unreplicated evolution (Fig. 1), a widespread phenomenon whereby species tend to covary according to the hierarchical structure of their phylogenetic relationships (Felsenstein 1985; Grafen 1989; Martins and Hansen 1997; Pagel 1997, 1999; Rohlf 2001). Consequently, related species and their traits do not represent independent observations, violating a common assumption of statistical tests. Failure to account for correlations arising from phylogenetic relationships can therefore bias inferences toward incorrect conclusions with high confidence (Felsenstein 1985; Maddison and FitzJohn 2015; Uyeda et al. 2018), such that it is now broadly acknowledged that one must consider the phylogenetic background upon which organisms and their traits have evolved when testing hypotheses about correlated trait evolution.

Nearly 40 years ago, Felsenstein (1985) proposed a simple yet elegant solution to this problem: phylogenetic regression, in which the statistical model is informed by tree structure. In doing so, he established phylogenetic comparative methods (PCMs) as the *de facto* standard in the field, ushering in a new era of tree thinking in comparative biology (Carvalho et al. 2005; Huey et al. 2019). The principles of phylogenetic regression were further clarified and expanded upon with the application of generalized least squares (Grafen 1989; Martins and Hansen 1997; Pagel 1997, 1999; Rohlf 2001), which accounts for statistical dependence by explicitly modeling the correlated error structure that results from shared inheritance. Over the past few decades, phylogenetic regression has become a paradigm of modern comparative biology, inspiring a wealth and diversity of offspring approaches for studying a myriad of biological hypotheses and questions under the PCM umbrella (Harvey and Pagel 1991; Blomberg et al. 2003; Felsenstein 2004; O'Meara et al. 2006; Revell et al. 2008; Beaulieu et al. 2012; Pennell and Harmon 2013). Importantly, a major goal of phylogenetic regression, and PCMs more generally, is to test evidence of correlated trait evolution, and thus, these methods have been widely embraced as a remedy for statistical nonindependence of species when studying adaptation (Doughty 1996).

However, recent studies have sown doubt over the reliability of phylogenetic regression, with evidence suggesting that PCMs may not be the panacea as long

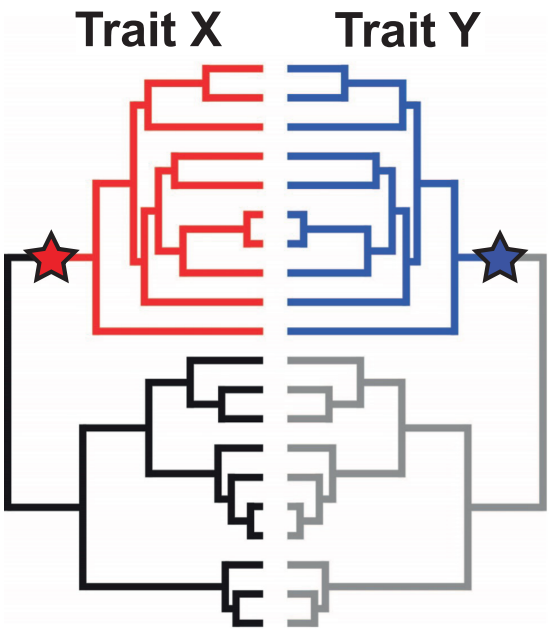

FIGURE 1. Unreplicated evolution and shared ancestry can lead to false associations between traits. In this example, observed values for traits X and Y are not independent across species due to the hierarchical structure of the sampled species. A single shift in the trait distribution (marked by stars) affects a large number of species due to their shared ancestry, leading to a false signal of correlated evolution if not properly accounted for. Figure inspired by Figure 2 of Uyeda et al. (2018).

hoped (Maddison and FitzJohn 2015; Uyeda et al. 2018). Ironically, many of these methods seemingly fail to protect against statistical dependency of trait data (Fig. 1 in Uyeda et al. 2018)—the reason they were developed in the first place. Deeper introspection into the justification and philosophical underpinnings of PCMs led to a recent call for a complete "rethinking" of the current paradigm (Uyeda et al. 2018), suggesting that biologists may be seriously overestimating evidence of trait associations (Maddison and FitzJohn 2015). At the heart of these concerns is the realization that widely held assumptions about trait evolution may not always reflect reality. In particular, a fundamental assumption of classical phylogenetic regression is that evolution proceeds more or less as a continuous process that can be approximated using Brownian motion (BM; Cavalli-Sforza and Edwards 1967; Felsenstein 1973) or related models that extend BM principles (Lande 1979; Hansen 1997; Pagel 1999; Blomberg et al. 2003; Harmon et al. 2010). Yet a breadth of macroevolutionary data suggests that biodiversity has been profoundly shaped by abrupt evolutionary shifts and discontinuities that often act in a lineage-specific manner, violating core PCM assumptions (Fig. 1; Schluter 2000; Uyeda et al. 2011, 2017; Slater and Pennell 2014; Landis and Schraiber 2017).

Scenarios of "evolution by jumps" were first hypothesized to describe rapid and dramatic phenotypic shifts in response to changes in the adaptive landscape (e.g., new environments, empty niches, and key innovations), which often manifest as lineage-specific novelties (Goldschmidt 1940; Simpson 1944). Whereas more realistic models of trait shifts have been proposed in several contexts (Eastman et al. 2011; Bartoszek et al. 2012; Uyeda and Harmon 2014; Clavel et al. 2015; Bastide et al. 2018a), phylogenetic regression is typically implemented under the assumption of continuous trait change in the absence of such phenomena. Current techniques are thus ill-equipped to deal with the dynamics observed in nature (O'Meara 2012; Pennell and Harmon 2013; Garamszegi 2014; Mazel et al. 2016), yielding systematic error that results from the inability of models to distinguish true statistical associations from instantaneous evolutionary shifts (Maddison and FitzJohn 2015; Uyeda et al. 2018).

In this study, we explore a new solution to this problem: the application of robust regression (Huber 2004; Yu and Yao 2017) for testing statistical trait associations. Mirroring the history of PCMs, "robust statistics" emerged during the 1980s as a new branch of statistics born out of concern for the rigor of classic techniques in the presence of model violations and outliers (Huber 2004; Yu and Yao 2017). Yet, save for a handful of examples (Slater and Pennell 2014; Arbour and López-Fernández 2016; Puttick 2018), robust methods have been largely overlooked in comparative trait studies, such that there has been an almost singular focus on using classical regression models to infer trait associations. Here we introduce robust phylogenetic regression to the PCM toolkit with 4 linear estimators, and we evaluate their performances across an array of statistically challenging and yet biologically important scenarios of rapid, lineage-specific evolutionary shifts known to mislead traditional phylogenetic regression. Following the protocol of a previous study that revealed biases in classical PCMs under such conditions (Uyeda et al. 2018), we examine the application of robust phylogenetic regression to traits that have experienced episodes of instantaneous jumps in trait space. To investigate the properties of these estimators for phylogenetic regression, we probe their behavioral characteristics in an array of simulated and empirical examples.

Materials and Methods

The Linear Model, Comparative Trait Data, and Phylogenetic Regression

Linear regression is arguably the most commonly applied statistical technique in biology (Sokal and Rohlf 1995; Ford 2000; Queen et al. 2002), as well as in the sciences more generally (Montgomery et al. 2012; Seber and Lee 2012), for studying relationships between variables. The familiar linear regression equation can be written as

$$\mathbf{y} = \boldsymbol{X}\boldsymbol{\beta} + \varepsilon, \quad (1)$$

where $\mathbf{y} \in \mathbb{R}^n$ is an n-dimensional vector of the response variable, $\boldsymbol{X} \in \mathbb{R}^{n \times (1+p)}$ is an $n \times (1 + p)$ design matrix (first column corresponds to the intercept term

as a vector $1 \in \mathbb{R}^n$ with the value of one for each element) containing n measurements for each of p predictor variables, $\boldsymbol{\beta} \in \mathbb{R}^{(1+p)}$ denotes the vector of unknown model parameters, and $\boldsymbol{\varepsilon} \in \mathbb{R}^n$ represents the residuals, or errors in predicting the response variable $\mathbf{y}$. Given a dataset $\mathcal{D} = \{\boldsymbol{y}, \boldsymbol{X}\}$, linear regression seeks to approximate $\boldsymbol{\beta}$ by finding the optimal estimates $\hat{\boldsymbol{\beta}}$ that minimize a cost function summarizing the overall magnitude of the residuals $\boldsymbol{\varepsilon} = \mathbf{y} - \hat{\mathbf{y}}$, which are computed as the difference between the observed response values $\mathbf{y}$ and the model predictions $\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$. Importantly, accuracy of the estimates $\hat{\boldsymbol{\beta}}$ and associated tests of significance (Rencher and Schaalje 2008) depend on fundamental assumptions of the model (Poole and O'Farrell 1971; Montgomery et al. 2012; Seber and Lee 2012; Mundry 2014). One such assumption of ordinary least-squares (OLS) regression is that the errors ε are independently and identically distributed (i.i.d.) as normal random variables, such that $\boldsymbol{\varepsilon} \mid \boldsymbol{X} \sim \mathcal{N}\left(0,\ \sigma^2 \boldsymbol{I}\right)$ follows a multivariate normal distribution, where $\mathbf{0} \in \mathbb{R}^n$ is an n-dimensional column vector representing the expected value of zero for each element in $\boldsymbol{\varepsilon}$, $\sigma^2 \in \mathbb{R}$ specifies a constant and unknown variance, and $\boldsymbol{I} \in \mathbb{R}^{n \times n}$ is the $n \times n$ identity matrix.

Phylogenetic regression strives to fit Equation 1 to test for a statistical relationship between a response trait $\mathbf{y}$ and one or more predictor traits X measured in $\mathcal{D}$ while explicitly considering the phylogenetic background upon which they evolved. Accounting for phylogeny requires correcting for the assumption of OLS that the error terms ε are uncorrelated. Researchers face 2 options for this task: phylogenetic independent contrasts (PIC; Felsenstein 1985) and phylogenetic generalized least squares (PGLS; Grafen 1989; Martins and Hansen 1997; Pagel 1997, 1999; Rohlf 2001), including the PGLS-based phylogenetic transformation. Both PIC and PGLS recognize that traits measured across related species are not statistically independent and use different strategies to address this issue. PIC computes a series of statistically independent contrasts according to the algorithm of Felsenstein (1985), yielding a newly transformed dataset $\mathcal{D}_{\text{PIC}} = \left\{\mathbf{y}_{\text{PIC}}, \boldsymbol{X}_{\text{PIC}}\right\}$, where $\mathbf{y}_{\text{PIC}} \in \mathbb{R}^{m-1}$ and $\boldsymbol{X}_{\text{PIC}} \in \mathbb{R}^{(m-1)\times(1+p)}$ compose the collection of $m - 1$ contrasts in m species for the response trait $\mathbf{y}$ and predictor traits $\boldsymbol{X}$, respectively. Regression is then performed on $\mathcal{D}_{\text{PIC}}$ through the origin, such that the intercept $\beta_0 = 0$ (Felsenstein 1985).

PGLS expands upon PIC by explicitly modeling a covariance structure Σ into the residual error, such that $\boldsymbol{\varepsilon} \mid \boldsymbol{X} \sim \mathcal{N}(0, \boldsymbol{\Sigma})$, where $\mathbf{0} \in \mathbb{R}^m$ and $\boldsymbol{\Sigma} \in \mathbb{R}^{m \times m}$ is the $m \times m$ phylogenetic covariance matrix (Felsenstein 1973; O'Meara et al. 2006) specified according to a particular tree and assumed evolutionary model (Grafen 1989; Martins and Hansen 1997; Pagel 1997, 1999; Rohlf 2001). Following the established OLS transformation of GLS (Judge and Griffiths 1985; Kariya and Kurata 2004; Rencher and Schaalje 2008), PGLS can also be implemented as a projection of traits through a phylogenetic transformation matrix $\boldsymbol{P} = (\boldsymbol{U}\boldsymbol{W}^{\frac{1}{2}}\boldsymbol{U}^t)^{-1} \in \mathbb{R}^{m \times m}$, where the eigenvectors $\boldsymbol{U} \in \mathbb{R}^{m \times m}$ and m associated eigenvalues along the diagonal of matrix $\boldsymbol{W} \in \mathbb{R}^{m \times m}$ are obtained from eigendecomposition of $\boldsymbol{\Sigma} = \boldsymbol{U}\boldsymbol{W}\boldsymbol{U}^{-1}$ (Garland Theodore and Ives 2000; Adams 2014; Adams and Collyer 2018). The original trait values can then be transformed by $\boldsymbol{P}$ to yield $\mathcal{D}_{\text{PGLS}} = \left\{\mathbf{y}_{\text{PGLS}}, \boldsymbol{X}_{\text{PGLS}}\right\}$, where $\mathbf{y}_{\text{PGLS}} = \boldsymbol{P}\mathbf{y}$ and $\mathbf{X}_{\text{PGLS}} = \boldsymbol{P}\boldsymbol{X}$. By rescaling branch lengths based on trait evolution, modeling the residual error as a function of Σ allows flexibility to the extent of "phylogenetic signal" present in the data (Grafen 1989; Mundry 2014).

PGLS estimates are identical to those of classical OLS in the absence of signal, while phylogenetic covariance is corrected to an appropriate degree with intermediate signal (Grafen 1989; Mundry 2014). PIC can also adapt these considerations (Felsenstein 1985; Symonds and Blomberg 2014), though this strategy is less common in practice. Importantly, PIC and PGLS regression estimates, or coefficients, are equivalent under BM evolution (Grafen 1989; Garland Theodore and Ives 2000; Rohlf 2001; Blomberg et al. 2012; Symonds and Blomberg 2014). PGLS uses the phylogenetic covariance matrix to project the trait data, such that its coefficients can be visualized alongside the original trait measurements (Blomberg et al. 2012; Symonds and Blomberg 2014).

The Ubiquitous Least-Squares Estimator

Almost without exception, PIC (i.e., using $\mathcal{D}_{\text{PIC}}$) and PGLS (i.e., using $\mathcal{D}_{\text{PGLS}}$) regression are conducted using the classical least-squares (L2) estimator (Gauss 1809). The L2 estimator can be defined as

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \left(y_i - [\boldsymbol{X}\boldsymbol{\beta}\ ]_i\right)^2 = \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \varepsilon_i^2, \# \quad (2)$$

where y_i denotes the known response and $[\boldsymbol{X}\boldsymbol{\beta}]_i$ is the predicted response given parameter vector $\boldsymbol{\beta}$ of the ith contrast of $n = m - 1$ total contrasts for PIC regression, or of the ith species of $n = m$ total species for PGLS regression, with $\varepsilon_i = y_i - [\boldsymbol{X}\boldsymbol{\beta}]_i$ therefore measuring the ith residual for both PIC and PGLS regression. Equation 2 also leads to the closed-form solution for the parameter estimates $\hat{\boldsymbol{\beta}}$ according to the normal equations $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}\mathbf{y}$, where $\boldsymbol{X}^{\mathrm{T}}$ is the transpose of $\boldsymbol{X}$, which can represent either $\boldsymbol{X}_{\text{PIC}}$ or $\boldsymbol{X}_{\text{PGLS}}$, and the -1 superscript indicates the matrix inverse. The L2 estimator makes no assumptions about the validity of the linear model (Equation 1); it simply finds the regression coefficients that provide the best linear fit to the data by minimization with Equation 2. Another way of viewing the L2 estimator is that it minimizes the squared Euclidian distance, or ℓ_2-norm, between the known values of $\mathbf{y}$ and the model predictions $\hat{\mathbf{y}}$. The ℓ_k-norm of a vector $\mathbf{v}$ is defined as $\|\mathbf{v}\|_k = \left(|v_0|^k + |v_1|^k + \cdots + |v_n|^k\right)^{\frac{1}{k}}$ for finite integer $k > 0$. In general, higher indices k place more emphasis on larger values in $\mathbf{v}$. For example, the ℓ_2-norm emphasizes larger values in $\mathbf{v}$ (i.e., potential outliers) relative to the ℓ_1-norm.

Robust Phylogenetic Regression

Here we introduce "robust phylogenetic regression" to the PCM toolkit with the linear M, L1 (least absolute deviation), S, and MM estimators. Each uses a different approach (described in detail below) to obtain a solution for the parameter estimates $\hat{\boldsymbol{\beta}}$. In particular, it is well-established that the standard L2 estimator is highly sensitive to outliers, and these estimators replace least squares with robust criteria (Yu and Yao 2017). Specifically, the sensitivity of an estimator to outliers can be represented by its breakdown point, a threshold for the maximum proportion of outliers that will yield unbiased estimates (Yohai 1987). The L2 estimator has a breakdown point of $1/n$, which tends to zero as the sample size n increases, meaning that even a single unusual observation can exert strong influence (Rousseeuw and Yohai 1984). Thus, an estimator with a larger breakdown point will have higher resistance to outliers. Furthermore, it is important that an estimator possesses high efficiency or precision relative to the L2 estimator when model assumptions are met. An efficient estimator also requires fewer observations to achieve the highest possible precision. Thus, robust estimators strive to achieve high breakdown point, efficiency, or both (Donoho and Huber 1983; Yu and Yao 2017).

M estimators are a class of estimators that incorporate appropriate weighting functions (Huber 1973, 1992), which have been used to detect early bursts of trait evolution in the presence of outlier taxa with posterior predictive model checks (Slater and Pennell 2014; Arbour and López-Fernández 2016). M estimators can be viewed as a generalization of maximum likelihood estimation (hence "M estimation") with a form of weighted regression that seeks to downweigh the influence of large residuals that may bias inferences (Yu and Yao 2017). An iteratively reweighted least squares procedure is used to estimate the weights as (Holland and Welsch 1977)

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \rho\left(\frac{\varepsilon_i}{\hat{\sigma}}\right), \tag{3}$$

where $\rho(t)$ is a robust loss function giving the weights of each residual ε_i, and $\hat{\sigma}$ is a scale estimate that is commonly chosen to be the median absolute deviation, as it is a robust measure of dispersion. Observations that do not deviate substantially from model predictions are effectively applied weights of one, and conversely, larger residuals are downweighed in the process. In particular, the solution to the L2 estimator is found when $\rho(t) = t^2$.

We can also obtain the L1 estimator (Koenker and Basset Jr. 1978) by setting $\rho(t) = |t|$, which minimizes the sum of the absolute values of the residuals as

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} |\varepsilon_i|. \tag{4}$$

The L1 estimator minimizes the ℓ_1-norm between the observed and predicted responses, which is designed to be more robust than the L2 estimator because Equation 4 minimizes the sum of absolute residual values $|\varepsilon_i|$, rather than the sum of squared residuals ε_i^2, as in Equation 2. Hence, the L1 estimator de-emphasizes outlier residuals that may lead to large squared distances with high leverage, or unusual values of the predictor traits $\boldsymbol{X}$ (Rousseeuw and Yohai 1984), which have also been shown to bias classical L2-based phylogenetic regression (Uyeda et al. 2018). However, like the L2 estimator, the L1 estimator has a breakdown point of $1/n$, which tends toward zero as n increases, and thus can be sensitive to high-leverage outliers (Maronna et al. 2019). In general, M estimators with monotone influence functions have a breakdown point of $1/n \to 0$ as $n \to \infty$, which results in a lack of immunity to large outliers (Maronna et al. 2019).

S estimators are based on estimators of scale (Rousseeuw and Yohai 1984) and seek to minimize the dispersion of the residuals defined as

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \hat{\sigma}(\varepsilon_1(\boldsymbol{\beta}), \varepsilon_2(\boldsymbol{\beta}), \ldots, \varepsilon_n(\boldsymbol{\beta})), \tag{5}$$

where $\varepsilon_i(\boldsymbol{\beta}) = \varepsilon_i$ and $\hat{\sigma}(\varepsilon_1(\boldsymbol{\beta}), \varepsilon_2(\boldsymbol{\beta}), \ldots, \varepsilon_n(\boldsymbol{\beta}))$ are the scaled M estimator defined as the solution to

$$\frac{1}{n}\sum_{i=1}^{n} \rho\left(\frac{\varepsilon_i(\boldsymbol{\beta})}{\hat{\sigma}}\right) = \delta\#, \tag{6}$$

for any given values of parameters $\boldsymbol{\beta}$, where δ is taken to be $E[\rho(\varepsilon)]$. S estimators can achieve a high breakdown point of 0.5 with an asymptotic efficiency of 0.29 (Maronna et al. 2019), and were first presented based on their proposed "invulnerability" to outliers and contaminated data (Rousseeuw and Yohai 1984).

MM estimators were first proposed by Yohai (1987), and now represent one of the most popular robust regression techniques (Yu and Yao 2017). MM estimation implements a 3-step procedure: (1) an initial estimate $\hat{\boldsymbol{\beta}}^{(0)}$ is obtained using an estimator with high breakpoint but potentially low efficiency, (2) a robust M estimate of the scale $\hat{\sigma}$ of the residuals is obtained from $\hat{\boldsymbol{\beta}}^{(0)}$, and (3) a final M estimate $\hat{\boldsymbol{\beta}}$ is found using $\hat{\boldsymbol{\beta}}^{(0)}$ as a starting point (Yu and Yao 2017). That is, the first step results in an initial estimate of the regression coefficients that are obtained using a consistent and robust yet not strictly efficient estimator (Yohai 1987). In the second step, residuals from the first step are used as input to obtain the error scales $\hat{\sigma}$. Lastly, the third step computes a final estimate of the regression coefficients using M estimation. MM estimation is statistically consistent and asymptotically normal when errors are normally distributed (Yohai 1987). In practice, the initial estimate can be obtained via an S estimator, followed by 2 successive rounds of M estimation to estimate $\hat{\sigma}$ and the final $\hat{\boldsymbol{\beta}}$. Typically, an MM estimate with a bisquare function defined as

$$\rho(x) = \begin{cases} \left(\frac{k^2}{6}\right)\left(1 - \left[1 - \left(\frac{x}{k}\right)^2\right]^3\right) & \text{for } |x| \le k \\ k^2/6 & \text{for } |x| > k \end{cases}$$

where $k = 4.685\sigma$ and the standard deviation of the errors σ is estimated from a robust S estimator (initial S estimate also starting from a bisquare function; Gross 1977) and efficiency of 0.85 is recommended in practice (Yu and Yao 2017), which can yield an asymptotic breakdown point of 0.5. Thus, each robust estimator has a unique cost function that decreases its sensitivity to outliers relative to the classical L2 estimator. Specifically, robust estimation is achieved by M through downweighing large residuals, L1 through minimizing the sum of absolute values of residuals, S through minimizing the dispersion of residuals, and MM through employing multiple steps of S and M estimation.

For our analyses, we leveraged a rich library of robust regression software packages within the R statistical environment (R Core Team 2013). Specifically, we used the *rlm* function with *method="M"* in the MASS package (Ripley 2015) for M estimation with iterated reweighted least squares and Huber's weighting scheme, the *lad* function in the L1PACK package (Osorio et al. 2017) for L1 estimation, the *lmRob* function with option *estim="Initial"* in the ROBUST package (Maechler 2014) for S estimation, and the *lmRob* function in the ROBUST package for MM estimation. Classical L2 estimation was conducted using the *lm* function in base R. These functions have been wrapped into a new open-source R package, ROBRT (ROBust Regression on Trees), for conducting robust regression within a coordinated framework for studies of comparative trait evolution.

In the next few subsections, we describe a series of simulations to evaluate each of these approaches with respect to (1) robustness to phylogenetic outliers and (2) statistical power to detect true trait associations. To evaluate robustness, we simulated evolutionary scenarios targeting an understanding of resistance to instantaneous evolutionary shifts that can drive false positive rates when traits are not statistically associated (i.e., zero covariance among traits). To assess statistical power, we simulated 2 traits that were indeed associated with one another (i.e., nonzero covariance among traits). Last, we conducted a more complex simulation analysis to understand the performance of each estimator for detecting true positives in the presence of outliers (both shifts and nonzero trait covariance). We follow these simulation-based studies with empirical case studies that include analyses of gene expression levels in mammals (Brawand et al. 2011), propagule size (PS) in invasive invertebrates (Makino and Kawata 2019), and muscle fiber characteristics in lizards (Scales et al. 2009). For each of these empirical analyses, we sought to mirror the protocol used in their respective original publications, except for the application of robust phylogenetic estimators.

Simulations With Evolutionary Shifts

We investigated the performance of robust regression for challenging—yet biological important—scenarios of rapid and unreplicated evolutionary change known to bias inferences (Uyeda et al. 2018). To do so, we first compared performances of the 4 robust estimators and the classical L2 estimator on datasets simulated across a wide range of increasing model misspecification severity. In these scenarios, we set the true regression coefficients in $\boldsymbol{\beta}$ to zero to represent uncorrelated (statistically independent) trait evolution, allowing us to evaluate the false positive rate for detecting correlated trait evolution of each estimator.

Specifically, we generated bivariate datasets $\mathcal{D} = \{\mathbf{y}, \mathbf{x}\}$ containing measurements of 2 statistically independent traits $\mathbf{y} \in \mathbb{R}^m$ and $\mathbf{x} \in \mathbb{R}^m$ simulated under a model of evolutionary shifts (termed "shift" model) for a phylogeny of m species (Fig. 2a,b). In our shift model, we assumed that $\mathbf{y}$ and $\mathbf{x}$ shifted simultaneously, but with differing magnitudes, at the same time and location (node in the tree), while all other parameters remained stationary throughout the tree (Eastman et al. 2013). We modified the classical BM process (Felsenstein 1973) to model the evolution of 2 independent traits by incorporating a pair of shift magnitudes $\delta = (\delta_{\mathbf{y}}, \delta_{\mathbf{x}}) \in \mathbb{R}^2$ with standard evolutionary variances corresponding to BM rates $\sigma^2 = \left(\sigma^2_{\mathbf{y}}, \sigma^2_{\mathbf{x}}\right) \in \mathbb{R}^2$ (O'Meara et al. 2006; Revell 2008) for traits $\mathbf{y}$ and $\mathbf{x}$. The shift magnitudes $\boldsymbol{\delta}$ were applied at the beginning of a single branch directly subtending the root that splits the tree into 2 clades of equal size (vertical dashes in Fig. 2a,b), which is predicted to yield high-leverage outliers that bias phylogenetic regression (Uyeda et al. 2018). To model variability in the shift magnitudes, we randomly sampled values of $\boldsymbol{\delta}$ across simulations from a bivariate normal distribution with zero mean, zero covariance (independent δ_y and δ_x), and variances that are scalar multiples of the BM rates $\boldsymbol{\sigma}^2$. That is, the simulated shift magnitudes are distributed as $\delta \sim \mathcal{N}(0, s\boldsymbol{\sigma}^2\boldsymbol{I})$, where $0 = (0,0)$, and $s\boldsymbol{\sigma}^2\boldsymbol{I}$ represents the product of a scalar variance $s > 0$, the BM rates $\boldsymbol{\sigma}^2$, and the 2×2 identity matrix $\boldsymbol{I}$. Throughout our simulations, we used a standard BM rate of one for each trait by setting $\boldsymbol{\sigma}^2 = (1,1)$ and the function *sim.char* from the GEIGER (Pennell et al. 2014) R package to simulate the ancestral (preshift) trait values with a mean state of zero for both $\mathbf{y}$ and $\mathbf{x}$. In these scenarios, we expect the regression model to be misspecified because the expectation of $\mathbf{y}$ no longer adheres to the null hypotheses due to a lineage-specific shift (i.e., dashes in Fig. 2).

Our simulation protocol generally followed the approach of Uyeda et al. (2018). First, we selected a shift variance $s \in \left\{10^{-2}, 10^{-1}, \ldots, 10^{5}\right\}$. Second, we used s to generate 2 independent values of the shift magnitudes $\delta \sim \mathcal{N}\left(0, s\boldsymbol{\sigma}^2\boldsymbol{I}\right)$. Third, we simulated a trait dataset $\mathcal{D}$ under a shift model with $\boldsymbol{\delta}$. Fourth, we transformed $\mathcal{D}$ into $\mathcal{D}_{\text{PIC}}$ for PIC regression and $\mathcal{D}_{\text{PGLS}}$ for PGLS regression. Last, we employed each of the 5 estimators (L2, M, L1, S, and MM) to perform PIC and PGLS regression on their transformed trait datasets. Given its popularity in comparative trait studies (Mazel et al. 2016), we used the default options for the classical t test with significance level $\alpha = 0.05$ and $n = m - 1$ and $n = m - 2$ degrees

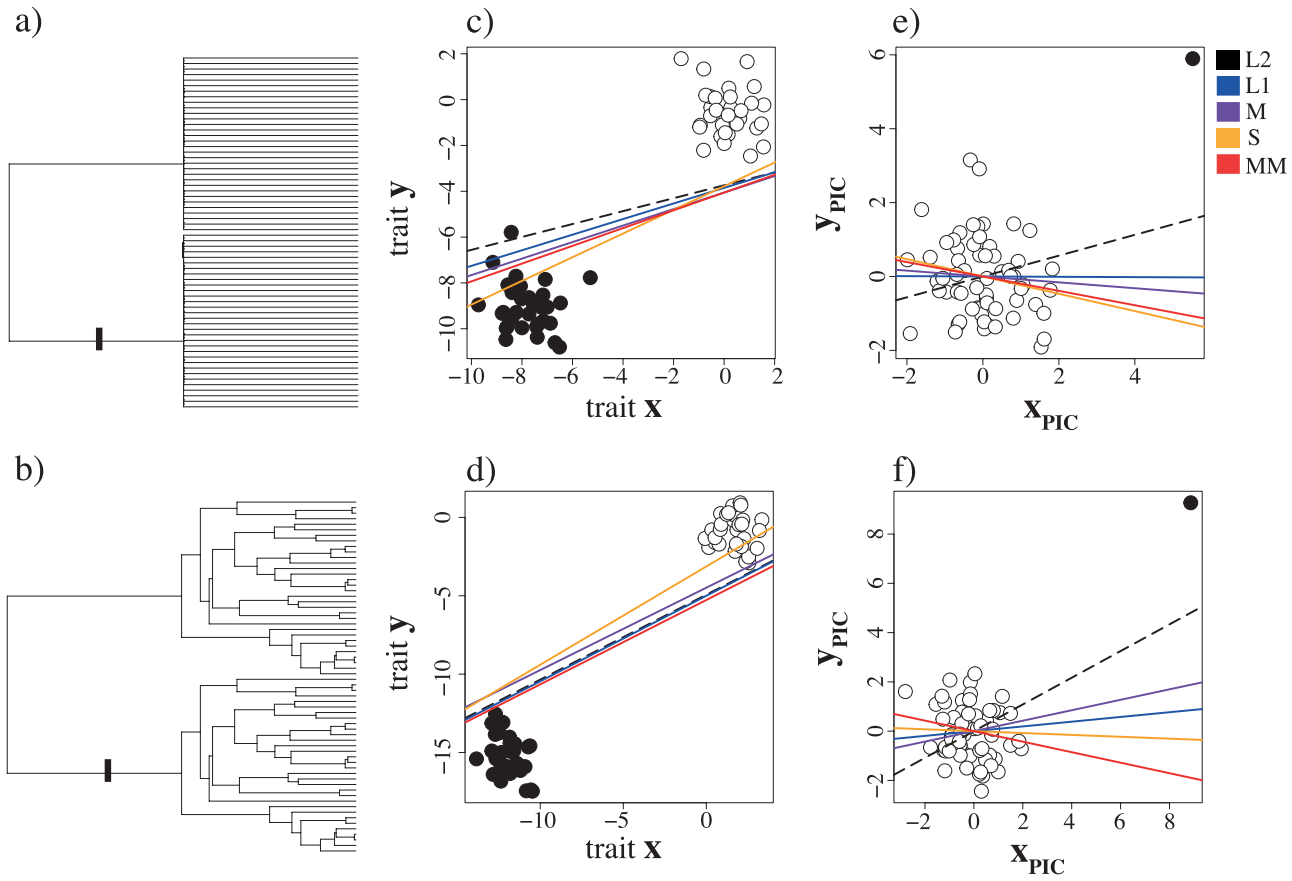

FIGURE 2. Characterizing differences among 5 linear estimates (L2, M, L1, S, and MM) for phylogenetic regression applied to statistically independent traits y and x. Data were simulated under the shift model with variance $s = 10^2$ according to Felsenstein's worst-case scenario (a) and a simulated balanced pure birth tree (b), with shift locations marked by dashes in corresponding trees. Covariation between traits y and x was assayed with PGLS (c and d) regression, and between transformed traits y_{PIC} and X_{PIC} with PIC regression (e and f). For graphical purposes, results shown in panels (c) and (d) are visualized in the original trait space—an advantage of PGLS (Garland Theodore et al. 1992). Examples of corresponding diagnostic plots are shown in Supplementary Fig. S1.

of freedom for PIC and PGLS regression, respectively. For each test, our goal was to evaluate the null hypothesis that the slope coefficient $\beta = 0$. We explored the performance of each estimator across diverse phylogenetic backgrounds using randomly generated trees of varying sizes corresponding to the number of species m. For each tree size $m \in \{64, 128, 256, 512, 1024\}$, we applied the scripts of Uyeda et al. (2018) to generate 100 balanced bifurcating trees by first simulating a pair of equally sized subtrees with $m/2$ species under a pure birth model, and subsequently joining them together at the root (Fig. 2a,b). For each of the 100 trees, we simulated 100 replicate datasets according to the protocol described above for each value of $s \in \{10^{-2}, 10^{-1}, \ldots, 10^{5}\}$.

Simulations With Correlated Trait Evolution

In addition to assaying false positive rates with simulations based on our shift model (Fig. 2a,b), we conducted a series of analyses to evaluate the statistical power (i.e., true positive rate) of each estimator for recovering a true relationship between 2 traits **y** and **x**. For these studies, we simulated trait data using a correlated trait model of bivariate BM evolution (Revell and Harmon 2008) that varied the strength of the statistical association between the traits $\rho \in \{0.1, 0.2, \ldots, 1.0\}$. Specifically, we modeled the evolution of the traits **y** and x with among-trait covariance matrix

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

Each element of $\boldsymbol{R}$ corresponds to a particular evolutionary covariance term, with the first row and column corresponding to the response trait **y**, and the second row and column corresponding to the predictor trait **x**. The antidiagonal values ρ denote the degree of covariance between **y** and **x**, and the diagonal elements indicate BM rates of 1 for both **y** and **x** (i.e., $\sigma^2_{\mathbf{y}} = \sigma^2_{\mathbf{x}} = 1$).

We first selected the true among-trait covariance from $\rho \in \{0.1, 0.2, \ldots, 1.0\}$. Next, we simulated traits **y** and **x** according to a bivariate BM model with among-trait covariance matrix $\boldsymbol{R}$ containing ρ and an ancestral state of zero. Last, we performed PIC and PGLS regression with each of the 5 estimators (L2, M, L1, S, and MM) to assess parameter estimates and significance

of the relationship between traits **y** and **x**. Simulations under this correlated trait evolution model did not include evolutionary shifts in the trait space, i.e., $\boldsymbol{\delta}$ = (0,0). Mirroring the shift model simulations, we generated 100 balanced Yule trees using scripts provided by Uyeda et al. (2018), as well as 100 replicate datasets for each tree while varying the number of species $m \in \{64, 128, 256, 512, 1024\}$. To supplement these analyses, we also conducted an array of simulations that included both shifts and true associations (nonzero ρ) between the 2 traits. We used the same simulation protocol as described above for the shift-only simulations, but included 3 different values of the trait covariance $\rho \in \{0.10, 0.25, 0.50\}$ with $m = 256$ species.

Empirical Analyses

We first applied simple phylogenetic regression to an empirical gene expression dataset from 11 female and male tissues in 8 mammals and chicken (Brawand et al. 2011). Specifically, we obtained normalized gene expression abundance measurements computed in reads per kilobase of exon model per million mapped reads (RPKM; Mortazavi et al. 2008) from female and male brain (whole brain without cerebellum), female and male cerebellum, female and male heart, female and male kidney, female and male liver, and testis in human (*Homo sapiens*), chimpanzee (*Pan trogodytes*), gorilla (*Gorilla gorilla*), orangutan (*Pongo pygmaeus abelii*), macaque (*Macaca mulatta*), mouse (*Mus musculus*), opossum (*Monodelphis domestica*), platypus (*Ornithorhynchus anatinus*), and chicken (*Gallus gallus*; Brawand et al. 2011). Some combinations of tissues and species in this dataset contain RPKM values from multiple replicates (e.g., 5 for human male brain; Brawand et al. 2011). Though distributions of RPKM values across genes in this dataset are approximately normally distributed (Supplementary Fig. S2), we chose to be conservative and use the median RPKM across replicates in such scenarios. We restricted our analysis to the most conservative 5321 1:1 orthologs, or those with constitutive exons that aligned across all species in the dataset (Brawand et al. 2011). Along with this expression dataset, we downloaded the original ultrametic phylogenetic tree constructed by Brawand et al. (2011), and scaled the tree depth to unit height. We investigated the statistical performance of the 5 estimators (L2, M, L1, S, and MM) with simple regression in 3 experimental settings: expression in female brain ~ male brain, female heart ~ male heart, and female kidney ~ male kidney. For each experiment, we conducted PIC regression based on log-transformed RPKM values across the nine species, and explored relationships between conditions via evaluations of statistical significance (*P*-values) and estimated slope coefficients $\hat{\boldsymbol{\beta}}$ at each gene in the dataset.

We next investigated the application of the 5 estimators to multiple regression scenarios with $p > 1$ predictor traits. For this demonstration, we focused on an example dataset obtained from a recent study that sought to test for genomic correlates with PS across a phylogenetic sampling of invasive and non-native invertebrates (Makino and Kawata 2019). We obtained the phylogenetic tree and dataset from the original study, which included measurements for PS (response trait) and 3 predictor traits: genome size (GS), proportion of duplicate genes, and species status (invasive or non-invasive) for 34 species. Following the protocol provided by Makino and Kawata (2019), we log-transformed PS and GS, and subsequently computed PIC transformations of all traits using the original phylogenetic tree. Using the PIC-transformed traits, we fit 4 alternative regression models and compared relative model fit using AIC (Akaike 1973); these analyses were conducted separately for each of the 5 estimators (L1, M, L1, S, and MM) to explore potential differences in relative model fit.

Lastly, we used a dataset on fast-twitch glycolytic and slow oxidative muscle fiber proportions in a clade of 22 lizards (Scales et al. 2009) to demonstrate differences between regression using the classical L2 estimator and the 4 robust estimators. Fast-twitch glycolytic fibers produce higher force and power, but tend to fatigue easily compared to slow oxidative fibers. Thus, we were interested in evaluating the relationship between these 2 muscle fiber types in a phylogenetic context. For this example, we downloaded a dataset and phylogeny (Scales et al. 2009) that was used by Uyeda et al. (2019) to illustrate the pitfalls of current PCMs. For each species, we computed PICs for 2 traits: proportion of fast glycolytic fibers and proportion of slow oxidative fibers. Next, we performed phylogenetic regression with L2, M, L1, S, and MM estimators to test for a linear relationship between these traits.

Results

Demonstrating Robust Phylogenetic Regression

Our simulations under the shift model (Fig. 2) represented opportunities to assess the sensitivity of each regression estimator (L2, M, S, and MM) to model violations because the 2 traits **y** and **x** were always statistically independent of one another. That is, the true slope coefficient of the linear model (Equation 1) was $\beta = 0$ in these scenarios, and our hope was that robust estimators would therefore fail to reject the null hypothesis of uncorrelated evolution by returning a nonsignificant *P*-value. We illustrated the application of robust phylogenetic regression in 2 familiar examples: Felsenstein's worst-case scenario (Fig. 2a) and a simple bifurcating tree (Fig. 2b). In both cases, all estimators applied to PGLS-transformed values displayed false positive trait associations ($\hat{\boldsymbol{\beta}} > 0$; Fig. 2c,d), because species experiencing the ancestral trait shift (black circles) reside in a different space than those without the shift (white circles). In contrast, all robust estimators (M, L1, S, and MM) showed decreased sensitivity with PIC regression to the single high-leverage outlier (black circle;

Fig. 2e,f), which represents the contrast generated by the branch with the ancestral trait shift, whereas the remaining contrasts cluster together (white circles). However, the L2 estimator is misled by this outlier contrast, exhibiting false positive trait associations ($\hat{\beta} > 0$; Fig. 2e,f). These examples hint that the application of robust regression estimators to PIC-transformed values may provide a solution to erroneous trait associations in the presence of such model violations.

Robust Regression Is Less Sensitive to Evolutionary Model Violations

Across our simulations, we found that all 4 robust estimators (L1, S, M, and MM) were less sensitive to model violations in the form of instantaneous evolutionary shifts than the classical L2 estimator (Fig. 3), though this result was primarily restricted to PIC regression (Fig. 3a,c, and e). Specifically, as we increased the severity

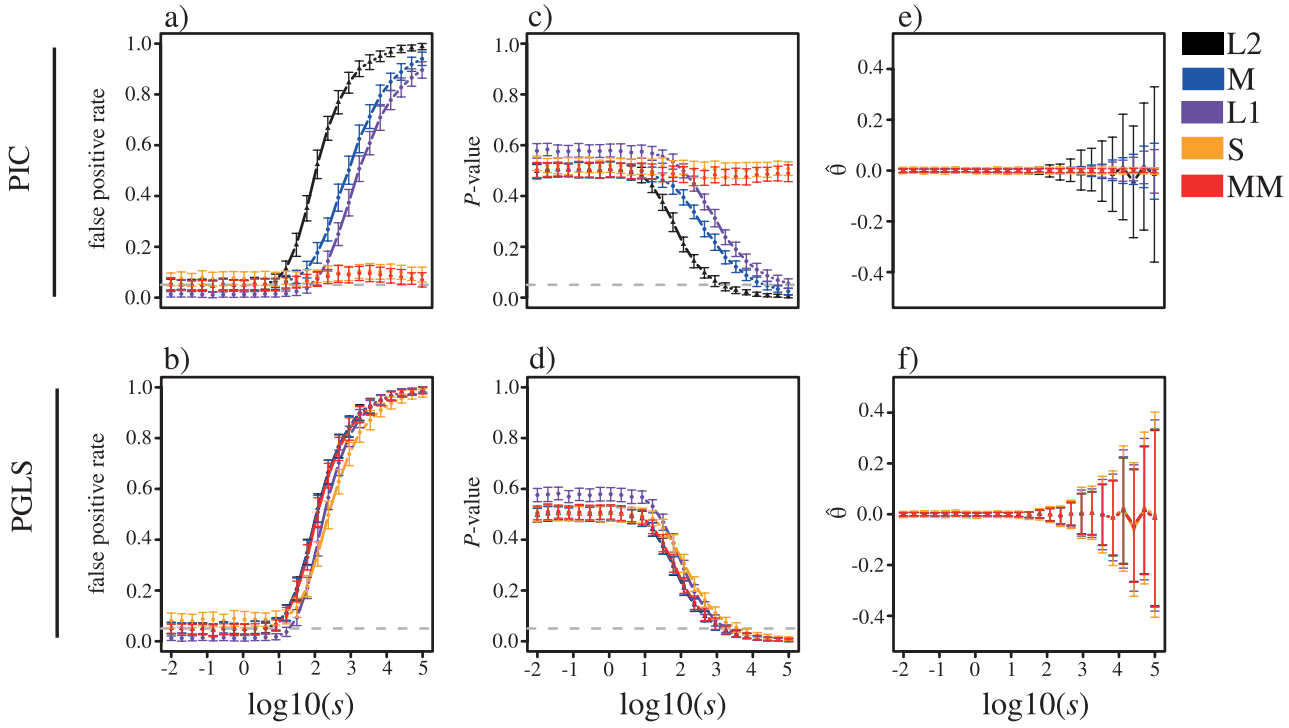

FIGURE 3. Investigating the impacts of evolutionary model violations for uncorrelated traits simulated under a shift model with variance s in $m = 256$ species. Depicted are the false-positive rate measured as the proportion of replicates with $P \leq 0.05$ (a and b), mean P-value (c and d), and estimate $\hat{\beta}$ of the slope coefficient β (e and f) plotted as a function of s for each of the 5 linear estimators (L2, M, L1, S, and MM) with PIC (top) and PGLS (bottom) regression, respectively.

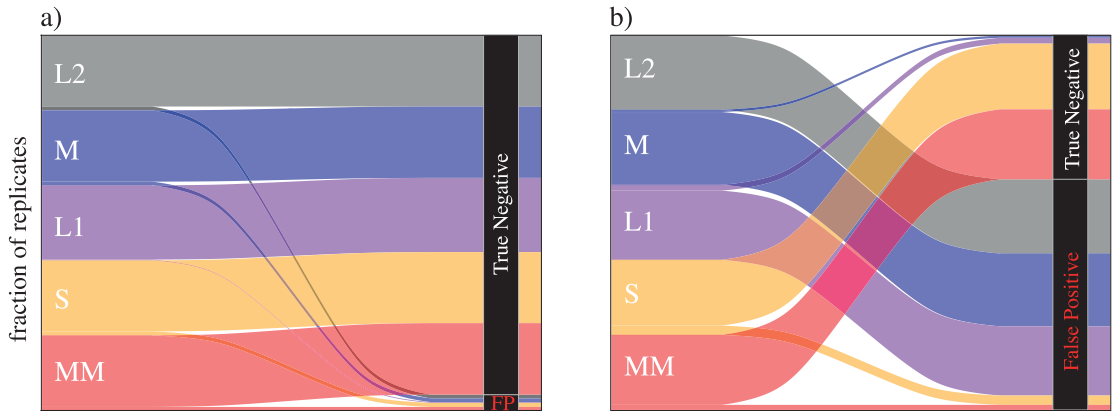

FIGURE 4. Alluvial plots showing the relative fractions of replicates resulting in either a true negative ($P > 0.05$) or false positive ($P \leq 0.05$) for applications of the 5 linear estimators (L2, M, L1, S, and MM) with PIC regression to simulations without any model violations (a; $s = 0$) and with strong model violations (b; $s = 10^5$). Moving from left to right in each panel illustrates the shift in the relative fraction of significant replicates for each of the 5 estimators (top to bottom).

of model violation in our simulations by increasing shift variance s, the false-positive rate of the L2 estimator with PIC regression increased, whereas the robust estimators displayed improved and varying levels of resistance (Fig. 3a). S and MM estimators were the most resistant to high values of s, whereas the M and L1 estimators exhibited slightly higher false-positive rates, but nonetheless fared better than the L2 estimator (Fig. 3a). The mean P-value across replicates tracked these differences between estimators as s increased in the simulations (Fig. 3c). Estimates of the slope parameter $\hat{\beta}$ appeared to be unbiased (means centered on zero) for these estimators, but with higher variance as s increased in our simulations (Fig. 3e). Increasing the number of species underscored the benefits of the 4 robust estimators over the L2 estimator with PIC (Supplementary Fig. S3), but again not with PGLS (Supplementary Fig. S4), regression. In all cases, our results corroborate those of Uyeda et al. (2018), who found that classical L2-based regression is highly sensitive to instantaneous evolutionary shifts.

The major advantage of employing robust estimators, particularly S and MM, with PIC regression is summarized by alluvial plots showing their relative fractions of true negatives ($P > 0.05$) and false positives ($P \leq 0.05$) across simulated replicates (Fig. 4). For simulations with no instantaneous shifts ($s = 0$), and hence no model violations, the L2 and all 4 robust estimators exhibited low false-positive rates (Fig. 4a). In contrast, for simulations with strong instantaneous shifts ($s = 10^5$), all estimators had high false-positive rates except for the S and MM estimators, which controlled the false-positive rates (Fig. 4b). Our results suggest that in the presence of uncorrelated evolution between traits and uncertainty in their past dynamics, S and MM estimators with PIC regression represent appropriate statistical models for the PCM toolkit.

Robust Regression Can Identify Correlated Trait Evolution

In addition to assaying robustness to model violations, we evaluated statistical power (true-positive rate) of the 5 estimators (L2, L1, S, M, and MM) to detect true trait relationships (i.e., $\rho > 0$ under the correlated trait evolution model with PIC and PGLS regression (Fig. 5). Statistical power was highest and comparable with the L2, M, and MM estimators, intermediate with the L1 estimator, and lowest with the S estimator for both PIC (Fig. 5a) and PGLS (Fig. 5b) regression. Unlike our sensitivity to outlier analyses under the shift model (Fig. 4), we did not find major differences in statistical power of the 5 estimators when comparing PIC to PGLS regression (Fig. 5a vs. b), though the S estimator demonstrated slightly higher power with PGLS regression. As expected, power to detect true trait correlations increased as ρ increased, with the S estimator achieving

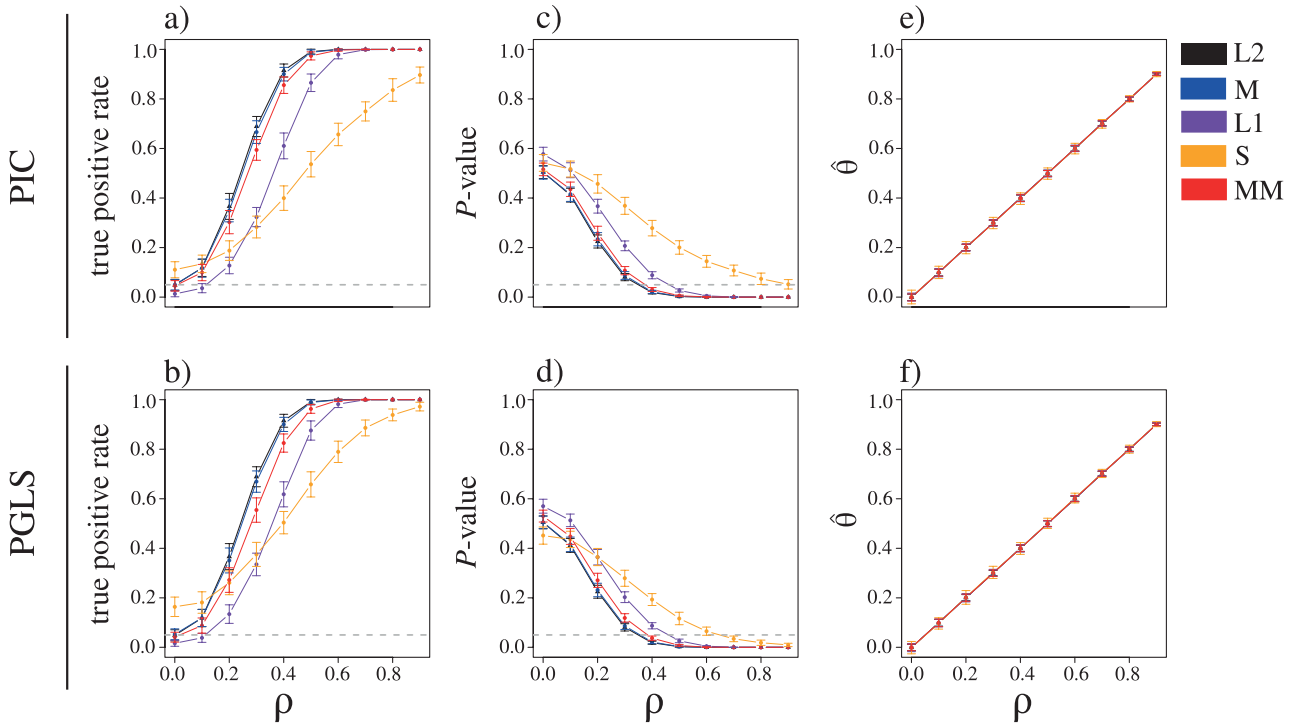

FIGURE 5. Investigating the impacts of the degree of true trait association for correlated traits simulated with among-trait covariance ρ in $m = 256$ species. Depicted are the true-positive rate measured as the proportion of replicates with $P \leq 0.05$ (a and b), mean P-value (c and d), and estimate $\hat{\beta}$ of the slope coefficient β (e and f) plotted as a function of ρ for each of the 5 linear estimators (L2, M, L1, S, and MM) with PIC (top) and PGLS (bottom) regression.

lower power relative to other estimators for moderate to high values of ρ. Mimicking our power analyses, mean *P*-value decreased with increasing ρ for all estimators, falling below 0.05 for ρ > 0 .4 with L2, M, and MM estimators, for ρ > 0.5 with the L1 estimator, and for ρ > 0.7 and 0.9 with the S estimator for PGLS and PIC regression, respectively (Fig. 5c,d). Similar to our results under the shift model, the correlated trait evolution model produced unbiased estimates of the slope coefficient $\hat{\boldsymbol{\beta}}$, such that they were on average equal to the value of ρ for both PIC (Fig. 5e) and PGLS (Fig. 5f) regression. Moreover, variability surrounding the estimates $\hat{\boldsymbol{\beta}}$ tended to be small, regardless of the true value of trait correlation ρ (Fig. 5e,f). Furthermore, all performance results exhibited by the 5 estimators under the correlated trait evolution model held across a wide range of sample sizes (Supplementary Figs. S5 and S6).

To better understand the impact of ρ on the power to detect true trait correlation, we generated alluvial plots depicting the relative fractions of false negatives ($P > 0.05$) and true positives ($P \leq 0.05$) for applications of the 5 linear estimators with PIC regression to simulated replicates with true evolutionary covariances of ρ > 0.3, 0.6, and 0.9 (Fig. 6). When true evolutionary covariance was weak (ρ > 0.3), power was <0.5 for all methods, with the L1 and S estimators demonstrating the lowest powers (Fig. 6a). In contrast, when evolutionary covariance was moderately strong (ρ > 0.6), all methods had powers of 1.0 or close to 1.0, except for the S estimator, which displayed a false negative rate of 0.34 (Fig. 6b). In the extreme setting of strong evolutionary covariance (ρ > 0.9), all estimators had high power, with each achieving a power of 1.0 except for the S estimator, for which the false negative rate decreased to 0.10 (Fig. 6c).

Building on these results, we evaluated the performance of the 5 estimators for detecting true positives in the presence of evolutionary outliers under more complex simulation scenarios including both shifts and nonzero trait covariance (Fig. 7; Supplementary Fig. S7). Receiver-operating characteristic curves illustrate clear differences among the 5 estimators for detecting trait associations in the presence of shifts (Fig. 7); these results are particularly apparent with larger trait covariances (e.g., bottom row of Fig. 7). As evolutionary shifts increase in magnitude, the L2 estimator tends to lose power (lighter to darker colors; Fig. 7a–c) for a given false positive rate. Yet the S and MM estimators are less affected by outliers (Fig. 7j–o), with MM demonstrating particularly high power at low false positive rates for larger trait covariances. Altogether, these results suggest that application of the MM estimator with PIC regression yields both the greatest robustness to false signatures of trait correlation and the highest power to detect trait associations.

Application of Robust Regression to Empirical Data

Our simulation experiments highlighted the varying robustness and statistical power of the 5 estimators (L2, M, L1, S, and MM) across an array of scenarios with and without trait correlation. Hence, we next explored whether application of these estimators to an empirical dataset would yield different conclusions about trait evolution. To this end, we examined the agreement between predictions employing robust estimators and the classical L2 estimator with PIC regression applied to a dataset composing expression measurements of 5615 genes in 11 tissues from nine species (Brawand et al. 2011; see Materials and Methods for details). We chose this dataset because adaptation often proceeds through regulatory modifications that alter spatial or temporal expression levels of genes (King and Wilson 1975; Wray et al. 2003; Carroll 2005; Jones et al. 2012; Mack et al. 2018), with recent studies showing that such changes can occur as rapid evolutionary shifts (Barua and Mikheyev 2020; Hamann et al, 2021). Thus, these spatial gene expression data represent an excellent setting to showcase the empirical performance of robust estimators.

We first investigated whether there were global differences in the abilities of the 5 estimators to detect correlated expression evolution by using simple regression to assay relationships between the expression measurements of genes in the same tissue in females and males (Supplementary Fig. S8). Indeed, distributions of *P*-values varied across estimators, and were

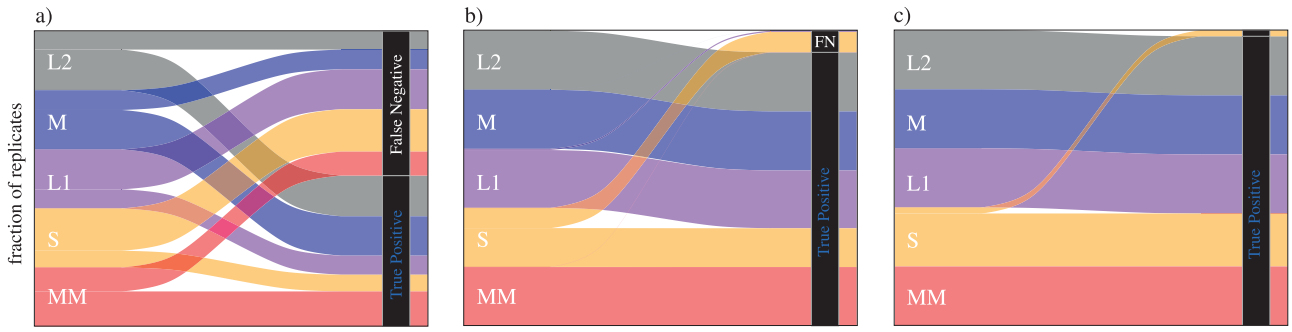

FIGURE 6. Alluvial plots showing the relative fractions of replicates resulting in either a true positive ($P \leq 0.05$) or false negative ($P > 0.05$) for applications of the 5 linear estimators (L2, M, L1, S, and MM) with PIC regression to simulations with true evolutionary covariance ρ = 0.3 (a), ρ = 0.6 (b), and ρ = 0.9 (c). Moving from left to right in each panel illustrates the shift in the relative fraction of significant replicates for each of the 5 estimators (top to bottom).

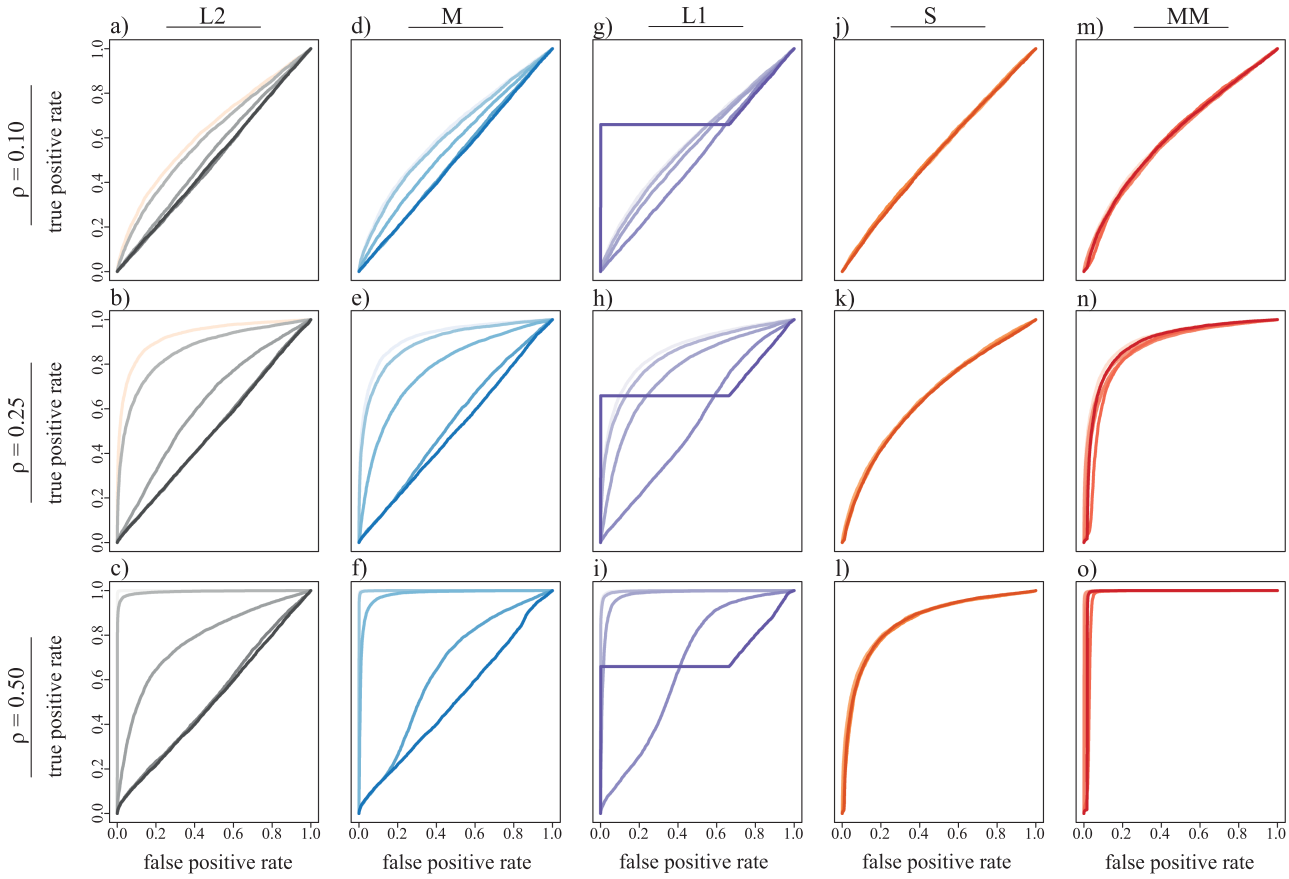

FIGURE 7. Receiver-operating characteristic curves in complex settings that include the joint scenario of a shift s and nonzero trait covariance $\rho \in \{0.1,0.25,0.50\}$ with $m = 256$ species. False positive rate (x-axis) is obtained from the P-value distribution of simulation replicates with a given shift s and $\rho = 0$, and true-positive rate (y-axis) is obtained from the P-value distribution of simulation replicates with the same shift s and $\rho = 0$. Curves are shown for each value of $\log_{10}(s) \in \{-2, -3, \ldots, 5\}$ for L2 (a-c), M (d-f), L1 (g-i), S (j-l), and MM (m-o) estimators. Light-to-dark shading of curves indicates their shift magnitude measured by $\log_{10}(s)$, with the lightest shading representing the smallest shift of $\log_{10}(s) = -2$, and the darkest shade representing the largest shift of $\log_{10}(s) = 5$.

smallest with the S estimator, then with MM, M, L2, and finally L1 estimators (Supplementary Fig. S8a–c). Because this coarse overview of P-value distributions across estimators ignores information about how often the P-value differs for one estimator when controlling for the same tested gene, we also computed differences in gene-specific P-values between each of the robust estimators and the L2 estimator. Comparison of the resulting distributions among estimators revealed that P-value differences with the L2 estimator were smallest and least variable for the M estimator, and largest and most variable for the L1 estimator (Supplementary Fig. S8d–f). Moreover, relative to the L2 estimator, P-values were smaller for the L1 estimator and larger for the S and MM estimators (Supplementary Fig. S8d–f). Similarly, we calculated differences in estimates of the slope, which measures the magnitude and direction of the expression relationship, between each of the robust estimators and the L2 estimator. Comparison of these distributions demonstrated that slope differences with the L2 estimator were small for all robust estimators on average, though tended to be least variable for M estimation (Supplementary Fig. S8g–i).

To more finely examine alignment between the 4 robust estimators and the L2 estimator, we created alluvial plots to compare the proportion of genes at which a given robust estimator agrees or disagrees with the L2 estimator rejecting ($P \leq 0.05$) or not rejecting ($P > 0.05$) the null hypothesis of no relationship between female and male expression in heart, kidney, and brain (Supplementary Fig. S9). As expected from our broader analysis (Supplementary Fig. S8), the M estimator tended to agree with the L2 estimator, with a skew toward the M estimator rejecting the null hypothesis when the L2 estimator does not reject it more often than not rejecting the null hypothesis when the L2 estimator rejects it (Supplementary Fig. S9a). The behavior of the L1 estimator was similar (Supplementary Fig. S9b), whereas the S and MM estimators demonstrated the same skew but also often did not agree with the L2 estimator (Supplementary Fig. S9c,d).

As a final investigation into the differences between the robust and classical L2 estimators, we performed case studies of genes for which robust estimators disagreed with the L2 estimator in each of the 3 tissues. We first considered one gene in each tissue for which

the null hypothesis was rejected by the L2 estimator, but not by at least one robust estimator (Figs. 8a–c). In such cases, the L2 estimator was potentially misled into identifying spurious correlated gene expression by a single high-leverage outlier (Fig. 8a,c). Both S and MM estimators did not reject the null hypothesis when the L2 estimator rejected the null hypothesis for *TP53I11* (ENSG00000175274) in heart (Fig. 8a), *NAV1* (ENSG00000134369) in kidney (Fig. 8b), and *DIPK1B* (ENSG00000165716) in brain (Fig. 8c). *TP53I11* encodes a protein predicted to negatively regulate cell population proliferation (The Alliance of Genome Resources Consortium 2019), *NAV1* encodes a protein hypothesized to be involved in neuronal development and regeneration (O'Leary et al. 2016), and *DIPK1B* encodes a transmembrane protein whose function is unknown (O'Leary et al. 2016). Thus, there does not appear to be a clear biological reason for correlation of any of these genes between females and males in the tissues in which they were uncovered by L2-based PIC regression.

Next, we examined one gene in each tissue for which the null hypothesis was rejected by at least one robust estimator, but not by the L2 estimator (Fig. 8d–f). Again, in all cases, the L2 estimator appeared to be swayed by a high-leverage outlier. Specifically, the L2 estimator did not reject the null hypothesis, whereas both S and MM estimators provided evidence of a significant relationship ($P \leq 0.05$) for female and male expression of 3 genes: *CENPI* (ENSG00000102384) in heart (Fig. 8d), *BAMBI* (ENSG00000095739) in kidney (Fig. 8e), and *REEP1* (ENSG00000068615) in brain (Fig. 8f). Intriguingly, *CENPI* encodes a key mitotic protein involved in gonad development and gametogenesis (O'Leary et al. 2016), which may provide a basis for correlations between female and male expression. *BAMBI* encodes a transmembrane glycoprotein that limits signaling of the TGF-beta gene family during early embryogenesis (O'Leary et al. 2016) and is associated with diabetes insipidus and childhood-onset diabetes mellitus (Awata et al. 2000; El-Shanti et al. 2000), perhaps explaining its correlation in female and male kidney. *REEP1* encodes a mitochondrial protein that enhances the cell surface expression of odorant receptors (O'Leary et al. 2016) and is associated with several neurodegenerative diseases (Züchner et al. 2006; Beetz et al. 2012; Kanwal et al. 2021), which may account for its correlation in female and male brain. Thus, all genes appear to have some biological support for either their sex-specific or tissue-specific covariation.

We investigated the application of robust estimators to the multiple regression problem by evaluating the association between propagule (egg or offspring) size and 3 predictors (GS, proportion of duplicate genes, and invasiveness status) in 34 invertebrate species (Makino and Kawata 2019). Specifically, we considered 4 models describing the relationship between PS and all subsets of at least 2 predictors, applied the 5 linear estimators (L2, M, L1, S, and MM) with PIC regression to fit these models to the data, and used AIC (Akaike 1973) to assess model fit (Table 1). The best-fit model for the L2, M, and L1 estimators was M3, which contained the proportion of duplicate genes and invasiveness status as its 2 predictors. However, the best-fit model for the S and MM estimators was M2, which instead contained the proportion of duplicate genes and GS as its 2 predictors. Moreover, the ranking of models was identical for the L2, M, and L1 estimators, whereas it was slightly different for the S and MM estimators. These differences are insightful, as the S and MM estimators yielded the smallest false positive rates in our study (Figs. 3, 4, and Supplementary Fig. S3), with the MM estimator also demonstrating comparable power to the L2 estimator in detecting true trait associations (Figs. 5, 6 and Supplementary Fig. S5). Furthermore, it is interesting that the best-fit model for all estimators contains the proportion of duplicate genes as a predictor, as gene duplication has long been hypothesized as a major driver of phenotypic diversity and evolutionary change (Ohno 1970).

Lastly, our application of robust estimators to the lizard muscle fiber dataset revealed substantial differences in inferred significance and coefficient estimates (Supplementary Fig. S10) for the linear model. While all 5 regression algorithms estimated a negative relationship between the proportion of fast glycolytic fibers and slow oxidative fibers ($\hat{\boldsymbol{\beta}} < 0$), only the S and MM estimators returned evidence of a significant relationship (orange and red lines; Supplementary Fig. S10).

Discussion

Unreplicated evolution and shared ancestry are banes of comparative biologists that can confound inferences of correlated trait evolution. Robust phylogenetic regression provides a solution to this problem by building a better defense against model violations and statistical outliers commonly encountered in trait data. Here we sought to build upon the PCM framework by introducing and exploring the application of a suite of new robust linear estimators for phylogenetic regression. Our analyses revealed that these robust estimators hold up to their namesake in both Felsenstein's (1985) worst-case scenario and a simple bifurcating tree (Fig. 2). Despite deliberate use of challenging evolutionary and statistical scenarios (Uyeda et al. 2018), several robust estimators demonstrated improvements over the classical L2 estimators in many contexts (Figs. 3–6 and Supplementary Figs.S3-S6), with particularly favorable performance observed for the S and MM estimators with PIC regression (Figs. 3–6, Supplementary Figs. S3 and S5). Specifically, these estimators tended to not be misled by model violations and outliers into detecting false signals of trait associations (Figs. 3, 4, and Supplementary Fig. S3), and also demonstrated high statistical power for detecting true trait associations

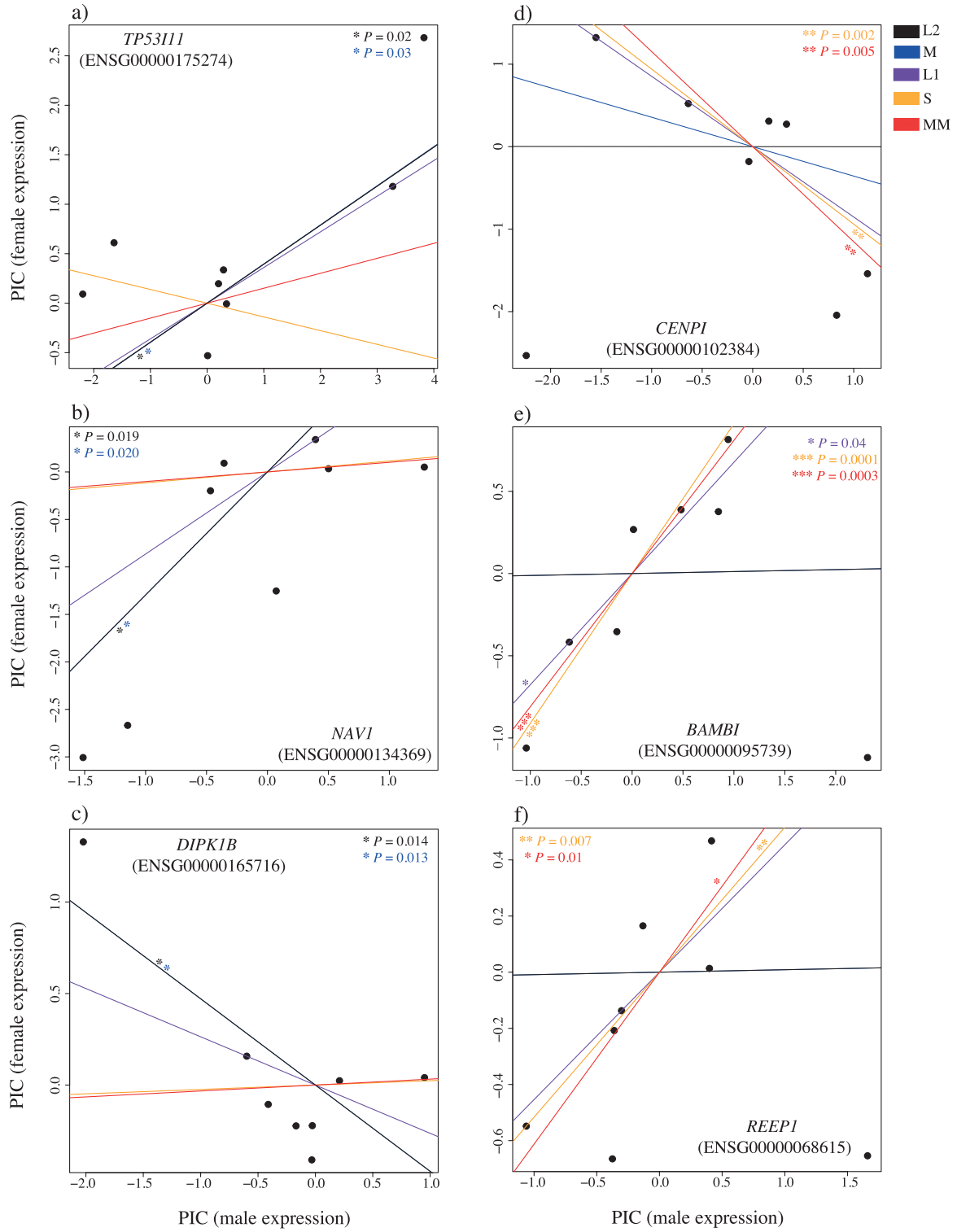

FIGURE 8. Examples highlighting differences between the L2 estimator and 4 robust estimators with PIC regression for testing relationships between expression in female and male heart (a and d), kidney (b and e), and brain (c and f) tissues. Stars indicate statistical significance ($P \leq 0.05$).

(Figs. 5, 6, Supplementary Figs.S5 and S6). In contrast, and in strong agreement with previous studies (Uyeda et al. 2018), we find that classical L2-based phylogenetic regression, which has long remained the *de facto* approach for evaluating trait correlations, suffers high systematic error in many scenarios.

TABLE 1. Model selection for phylogenetic regression applied to evaluate the relationship between propagule size (PS) and 3 predictors—genome size (GS), proportion of duplicate genes (PD), and invasiveness status (IS)—in 34 invertebrate species. The best-fit model is bolded for each estimator

Estimator	Model	AIC	Model ranking
L2	M1: PS ~ GS + PD + IS	96.20	2
	M2: PS ~ GS + PD	107.44	3
	M3: PS ~ PD + IS	**95.79**	**1**
	M4: PS ~ GS + IS	122.94	4
M	M1: PS ~ GS + PD + IS	96.40	2
	M2: PS ~ GS + PD	107.98	3
	M3: PS ~ PD + IS	**96.06**	**1**
	M4: PS ~ GS + IS	123.13	4
L1	M1: PS ~ GS + PD + IS	94.32	2
	M2: PS ~ GS + PD	101.35	3
	M3: PS ~ PD + IS	**92.65**	**1**
	M4: PS ~ GS + IS	115.85	4
S	M1: PS ~ GS + PD + IS	126.62	3
	M2: PS ~ GS + PD	**125.32**	**1**
	M3: PS ~ PD + IS	125.97	2
	M4: PS ~ GS + IS	138.86	4
MM	M1: PS ~ GS + PD + IS	137.25	2
	M2: PS ~ GS + PD	**122.46**	**1**
	M3: PS ~ PD + IS	138.33	3
	M4: PS ~ GS + IS	134.89	4

The best-fit model is bolded for each estimator.

One pervasive trend in our analyses was the contrasting robustness of phylogenetic regression based on PIC and PGLS principles (Figs. 2, 3, Supplementary Figs.S3 and S4). PIC and PGLS regression are mathematically equivalent for L2 estimation (Blomberg et al. 2012), and yet, the performance of robust estimators differed markedly between the 2 approaches, particularly in the presence of outliers (Figs. 3, Supplementary Figs. S3 and S4). While we observed considerable advantages in using robust estimators with PIC regression (Fig. 3a), these benefits were less apparent with PGLS regression (Fig. 3b). There may be several reasons for these discrepancies, and our analyses suggest that the manner by which traits are projected may hold clues. Our simple demonstration of phylogenetic regression in Fig. 2 offers insights into differences when using robust estimators with PGLS (Fig. 2c,d) vs. PIC (Fig. 2e,f). By transforming the traits into a set of independent contrasts, PIC regression has effectively one fewer input observation (i.e., $n = m - 1$ contrasts), which is predicted to reduce significance when testing a null hypothesis by increasing the standard error of the test statistic. However, PGLS regression with the L2 estimator includes an intercept term, reducing the degrees of freedom, such that the statistical tests are equivalent. Importantly, these contrasts collectively express all variation present among a set of m species related by a phylogeny (Felsenstein 1985), and by doing so, encode information about ancestral traits and the phylogenetic locations of contrasts that may represent outliers (e.g., dark points in Fig. 2e,f). Thus, clustering of PICs tended to isolate the single outlying contrast representing the specific branch location of the model violation itself (e.g., marked dashes in Fig. 2a,b correspond to dark points in Figs. 2e,f). In contrast, trait projections based on the PGLS transformation yielded 2 large partitions corresponding to 2 clusters: species that did (dark points in Fig. 2c,d) and that did not (white points in Fig. 2c,d) experience the model violation in their phylogenetic history. That is, PGLS transformations separate the trait values into 2 clusters as a simple binary response to the presence of the model violation in the lineage history of a set of species, whereas PIC effectively distinguishes the presence of a single distinctive outlier (Fig. 2c vs. 2e). PIC regression therefore allowed the robust estimators to take full advantage of the information present within trait contrasts, improving resistance to model violations; this can be observed when comparing estimated relationships of L2 with M, L1, S, and MM estimators in Fig. 2.

It is important to note that we primarily focused our simulations on a single shift affecting many species in a tree, and it remains to be seen whether these results hold for more complex scenarios with multiple evolutionary shifts as well as with other complicated aspects of trait evolution. As an initial exploration of more complex settings, we expanded our simulations to include both shifts and true trait associations (Fig. 7 and Supplementary Fig. S7), uncovering evidence that L2, L1, and M estimators may become overconfident under conditions of extreme trait shifts, and that their estimates of regression coefficients were subject to downward bias and high variance as the magnitude of the shift increased (Supplementary Fig. S7). In contrast, S and MM estimators showed no apparent trends with increasing trait shifts while retaining unbiased coefficient estimates (Supplementary Fig. S7). Also consistent with our prior findings, the MM estimator had a higher true positive rate than the S estimator, while remaining unaffected by potential false association signals due to trait shifts (Fig. 7). Taken together, our findings reinforce the advantages of robust regression with the

MM estimator across scenarios explored in this study, while also highlighting the utility of comparing significance and parameter estimates across estimators as the behavior of these measures differs depending on the particular regression model used.

Throughout our analyses, perhaps most apparent were the advantages of using the MM estimator with PIC regression, which achieved the highest observed resistance to evolutionary outliers (Fig. 3 and Supplementary Fig. S3) while retaining sufficient power for detecting true trait associations (Fig. 5 and Supplementary Fig. S5). Specifically, whereas MM-based PIC regression was largely unaffected by the presence of very large magnitude model violations in extreme evolutionary scenarios (Fig. 3a), it displayed comparable power to classical L2-based PIC regression for detecting true trait associations (Fig. 5). Thus, application of MM-based PIC regression may serve as a single best general-purpose strategy for assaying trait associations. Furthermore, among the 4 robust estimators explored here (M, L1, S, and MM), the L1 estimator appeared to be the least robust, and yet still outperformed the L2 estimator in many cases (Fig. 3 and Supplementary Fig. S3). Collectively, these findings underscore improvements of robust phylogenetic regression compared to classical L2-based approaches, demonstrating that robust estimators may be used *in lieu* of the classical L2 estimator to provide more reliable inferences of correlated trait evolution. In this sense, robust regression represents a much-needed answer to ongoing questions and debates about trait inferences in the presence of evolutionary model violations.

Our empirical studies of gene expression in 3 female and male tissues also provided support for improved inferences with robust estimators compared to standard L2-based phylogenetic regression. In particular, we showed that there were often differences between genes inferred as correlated in female and male tissues by robust and the L2 estimator (Supplementary Figs. S8 and S9), and that such differences may be indicative of their varying sensitivities to outliers (Fig. 8a–c) or powers for detecting true relationships (Fig. 8d–f). For example, evaluation of female and male expression in heart tissue resulted in rejection of the null hypothesis by the L2 estimator but not by either S or MM estimators for the *TP53I11* gene (Fig. 8a), and conversely, in rejection of the null hypothesis by both S and MM estimators but not by the L2 estimator for the *CENPI* gene (Fig. 8d). In both cases, these differences were driven by a single outlier that either reduced the robustness (Fig. 8a) or power (Fig. 8d) of the L2 estimator, highlighting the sensitivity of classical L2-based phylogenetic regression to model violations and underscoring the benefits of robust estimators in this context. Consistent with this observation, *TP53I11* does not appear to be related to sex or heart tissue, whereas *CENPI* is involved in sex-specific functions (O'Leary et al. 2016) that may help explain its apparent correlation in female and male tissues. Hence, this example highlights the utility of robust phylogenetic regression in detecting biologically relevant trait associations when classical methods may fail in the presence of model violations. When considering the implications of these findings, we found the following quote from Tukey (1975) to be quite fitting: "It is perfectly proper to use both classical and robust/resistance methods routinely, and only worry when they differ enough to matter. But when they differ, you should think hard." Thus, contrasting the results of different linear estimators (both robust and classical) may be fruitful if trait data are suspect of model violations.

Though robust phylogenetic regression presents a promising solution to the problem of unreplicated evolutionary jumps, we neither claim that it is the only solution, nor that it completely solves this problem. However, our results do suggest that robust regression is likely to be helpful in many scenarios of unreplicated evolutionary events, and therefore represents a step toward a common solution. Uyeda et al. (2018) argued for a philosophical unification of phylogenetic natural history and a priori hypothesis testing (Maddison et al. 2007; Alfaro et al. 2009; Stadler 2011; FitzJohn 2012; Rabosky 2014); because robust regression decreases sensitivity to phylogenetic outliers, this approach may help achieve both goals by accounting for hidden unreplicated evolutionary events in the tree when testing for trait associations. Specifically, large differences between classical L2 phylogenetic regression and robust methods may be indicative of interesting evolutionary phenomena, represented by high-leverage outliers within the ancestral history of a focal clade of interest (Fig. 2). Whereas we examined the performance of robust phylogenetic regression in a few key settings, future investigations will be necessary to further our understanding of its sensitivity and power in the presence of a diversity of model violations and evolutionary scenarios. In particular, phylogenetic regression is typically conducted under the assumption of BM evolution (or similar models) in the absence of shifts, and thus, recent contributions to multivariate models of trait evolution and shifts that include both between-species correlations (covariance due to shared species histories) and between-trait correlations (covariance due to shared trait functions) hold particular promise (Uyeda and Harmon 2014; Bastide et al. 2017, 2018a; Duchen et al. 2017). Strategies such as Bayesian mixture models (Uyeda and Harmon 2014; Uyeda et al. 2017, 2018) and the implementation of heavy-tailed distributions for modeling error terms while fitting regression models (Landis et al. 2013; Elliot and Mooers 2014) have also been suggested to deal with these issues (Uyeda et al. 2018), as well as goodness-of-fit tests and graphical models (Höhna et al. 2014, 2016). Future work will be useful for evaluating different types of model violations and scenarios of non-BM evolution, non-bifurcating trees (Bastide et al. 2018b), and more complex evolutionary shifts (Duchen et al. 2017; Bastide et al. 2018a; Mitov et al. 2019). Collectively, our findings join arguments for increased vigilance against unreplicated evolution and shared ancestry, and a better understanding of phylogenetic modeling

assumptions (Revell 2010; Ho and Ané 2014; Mundry 2014; Maddison and FitzJohn 2015; Uyeda et al. 2018; Ives 2019), setting the stage for a shift in method development in this important area.

SUPPLEMENTARY MATERIAL

Data available from the Dryad Digital Repository: https://doi.org/10.5061/dryad.xpnvx0kn1

ACKNOWLEDGMENTS

The authors would like to acknowledge the use of services provided by Research Computing at Florida Atlantic University. This work was also supported by start-up funds provided by the University of Arkansas.

FUNDING

This work was supported by National Science Foundation grants DEB-1949268, DEB-2001059, DBI-2130666, and BCS-2001063, the National Institutes of Health grants R35GM142438 and R35GM128590. This research was also supported by start-up funds provided by the University of Arkansas and the Arkansas High Performance Computing Center, which is funded through multiple National Science Foundation grants and the Arkansas Economic Development Commission.

DATA AVAILABILITY

We have implemented a suite of functions for conducting both simple and multiple robust phylogenetic regression using PIC and PGLS in the R package ROBRT (ROBust Regression on Trees), which also includes several example applications in the tutorial. ROBRT includes functions for computing each of the 4 robust estimators explored in this study (M, L1, S, and MM), as well as the standard L2 estimator. ROBRT is freely available for open use by the community (https://github.com/radamsRHA/ROBRT), and requires the dependencies APE (Paradis et al. 2004) and PHYTOOLS (Revell 2012). ROBRT also includes functionality for assessing diagnostic plots with phylogenetic regression.

REFERENCES

Adams D.C. 2014. A method for assessing phylogenetic least squares models for shape and other high-dimensional multivariate data. Evolution 68:2675–2688.

Adams D.C., Collyer M.L. 2018. Phylogenetic ANOVA: group-clade aggregation, biological challenges, and a refined permutation procedure. Evolution 72:1204–1215.

Akaike H. 1973. A new look at statistical model identification. IEEE Trans. Automat. Contr. 19:716–723.

Alfaro M.E., Santini F., Brock C., Alamillo H., Dornburg A., Rabosky D.L., Carnevale G., Harmon L.J. 2009. Nine exceptional radiations plus high turnover explain species diversity in jawed vertebrates. Proc. Natl. Acad. Sci. U.S.A. 106:13410–13414.

The Alliance of Genome Resources Consortium. 2019. Allianice of Genome Resources Portal: unified model organism research platform. Nucleic Acids Res. 48:D650–D658.

Arbour J.H., López-Fernández H. 2016. Continental cichlid radiations: functional diversity reveals the role of changing ecological opportunity in the Neotropics. Proc. Biol. Sci. 283:20160556.

Awata T., Inoue K., Kurihara S., Ohkubo T., Inoue I., Abe T., Takino H., Kanazawa Y., Katayama S. 2000. Missense variations of the gene responsible for Wolfram syndrome (WFS1/wolframin) in Japanese: possible contribution of the Arg456His mutation to type 1 diabetes as a nonautoimmune genetic basis. Biochem. Biophys. Res. Commun. 268:612–616.

Bartoszek K., Pienaar J., Mostad P., Andersson S., Hansen T.F. 2012. A phylogenetic comparative method for studying multivariate adaptation. J. Theor. Biol. 314:204–215.

Barua A., Mikheyev A.S. 2020. Toxin expression in snake venom evolves rapidly with constant shifts in evolutionary rates. Proc. Biol. Sci. 287:20200613.

Bastide P., Ané C., Robin S., Mariadassou M. 2018a. Inference of adaptive shifts for multivariate correlated traits. Syst. Biol. 67:662–680.

Bastide P., Mariadassou M., Robin S. 2017. Detection of adaptive shifts on phylogenies by using shifted stochastic processes on a tree. J. R. Stat. Soc. Ser. B 79:1067–1093.

Bastide P., Solís-Lemus C., Kriebel R., William Sparks K., Ané C. 2018b. Phylogenetic comparative methods on phylogenetic networks with reticulations. Syst. Biol. 67:800–820.

Beaulieu J.M., Jhwueng D., Boettiger C., O'Meara B.C. 2012. Modeling stabilizing selection: expanding the Ornstein–Uhlenbeck model of adaptive evolution. Evol. Int. J. Org Evol. 66:2369–2383.

Beetz C., Pieber T.R., Hertel N., Schabhüttl M., Fischer C., Trajanoski S., Graf E., Keiner S., Kurth I., Wieland T., Varga R.-E., Timmerman V., Reilly M.M., Strom T.M., Auer-Grumbach M. 2012. Exome sequencing identifies a REEP1 mutation involved in distal hereditary motor neuropathy type V. Am. J. Hum. Genet. 91:139–145.

Blomberg S.P., Garland T. Jr, Ives A.R. 2003. Testing for phylogenetic signal in comparative data: behavioral traits are more labile. Evolution 57:717–745.

Blomberg S.P., Lefevre J.G., Wells J.A., Waterhouse M. 2012. Independent contrasts and PGLS regression estimators are equivalent. Syst. Biol. 61:382–391.

Brawand D., Soumillon M., Necsulea A., Julien P., Csárdi, G, Harrigan P., Weier M., Liechti A., Aximu-Petri A., Kircher M., Albert F.W. 2011. The evolution of gene expression levels in mammalian organs. Nature. 478:343–348.

Carroll S.B. 2005. Evolution at two levels: on genes and form. PLoS Biol. 3:e245.

Carvalho P., Diniz-Filho J.A.F., Bini L.M. 2005. The impact of Felsenstein's "Phylogenies and the comparative method" on evolutionary biology. Scientometrics 62:53–66.

Cavalli-Sforza L.L., Edwards A.W.F. 1967. Phylogenetic analysis models and estimation procedures. Am. J. Hum. Genet. 19:233–257.

Clavel J., Escarguel G., Merceron G. 2015. mvMORPH: an R package for fitting multivariate evolutionary models to morphometric data. Methods Ecol. Evol. 6:1311–1319.

Donoho D.L., Huber P.J. 1983. The notion of breakdown point. Erich L. Lehmann 157184:157184.

Doughty P. 1996. Statistical analysis of natural experiments in evolutionary biology: comments on recent criticisms of the use of comparative methods to study adaptation. Am. Nat. 148:943–956.

Duchen P., Leuenberger C., Szilágyi S.M., Harmon L., Eastman J., Schweizer M., Wegmann D. 2017. Inference of evolutionary jumps in large phylogenies using Lévy processes. Syst. Biol. 66:950–963.

Eastman J.M., Alfaro M.E., Joyce P., Hipp A.L., Harmon L.J. 2011. A novel comparative method for identifying shifts in the rate of character evolution on trees. Evol. Int. J. Org Evol. 65:3578–3589.

Eastman J.M., Wegmann D., Leuenberger C., Harmon L.J. 2013. Simpsonian "evolution by jumps" in an adaptive radiation of Anolis lizards. ArXiv:1305.4216.

Elliot M.G., Mooers A.O. 2014. Inferring ancestral states without assuming neutrality or gradualism using a stable model of continuous character evolution. BMC Evol. Biol. 14:1–15.
El-Shani H., Lidral A.C., Jarrah N., Druhan L., Ajlouni K. 2000. Homozygosity mapping identifies an additional locus for Wolfram syndrome on chromosome 4q. Am. J. Hum. Genet. 66:1229–1236.
Felsenstein J. 1973. Maximum-likelihood estimation of evolutionary trees from continuous characters. Am. J. Hum. Genet. 25:471–492.
Felsenstein J. 1985. Phylogenies and the comparative method. Am. Nat. 125:1–15.
Felsenstein J., Felenstein J. 2004. Inferring phylogenies. Sunderland, Massachusetts: Sinauer Associates.
FitzJohn R.G. 2012. Diversitree: comparative phylogenetic analyses of diversification in R. Methods Ecol. Evol. 3:1084–1092.
Ford E.D. 2000. Scientific method for ecological research. Cambridge, MA: Cambridge University Press.
Garamszegi L.Z. 2014. Modern phylogenetic comparative methods and their application in evolutionary biology: concepts and practice. Heidelberg: Springer.
Garland Theodore J., Harvey P.H., Ives A.R. 1992. Procedures for the analysis of comparative data using phylogenetically independent contrasts. Syst. Biol. 41:18–32.
Garland Theodore J., Ives A.R. 2000. Using the past to predict the present: confidence intervals for regression equations in phylogenetic comparative methods. Am. Nat. 155:346–364.
Gauss C.F. 1809. Theoria motus corporum coelestium. Hamburg: Perthes.
Goldschmidt R. 1940. The material basis of evolution. New Haven: Yale University Press.
Grafen A. 1989. The phylogenetic regression. Philos. Trans. R. Soc. London, Ser. B 326:119–157.
Gross A.M. 1977. Confidence intervals for bisquare regression estimates. J. Am. Stat. Assoc. 72:341–354.
Hamann E., Pauli C.S., Joly-Lopez Z., Groen S.C., Rest J.S., Kane N.C., Purugganan M.D., Franks S.J. 2021. Rapid evolutionary changes in gene expression in response to climate fluctuations. Mol. Ecol. 30:193–206.
Hansen T.F. 1997. Stabilizing selection and the comparative analysis of adaptation. Evolution 51:1341–1351.
Harmon L.J., Losos J.B., Jonathan Davies T., Gillespie R.G., Gittleman J.L., Bryan Jennings W., Kozak K.H., McPeek M.A., Moreno-Roark F., Near T.J. 2010. Early bursts of body size and shape evolution are rare in comparative data. Evol. Int. J. Org. Evol. 64:2385–2396.
Harvey P.H., Pagel M.D. 1991. The comparative method in evolutionary biology. Oxford: Oxford University Press.
Ho L.S.T., Ané C. 2014. Intrinsic inference difficulties for trait evolution with Ornstein-Uhlenbeck models. Methods Ecol. Evol. 5:1133–1146.
Höhna S., Heath T.A., Boussau B., Landis M.J., Ronquist F., Huelsenbeck J.P. 2014. Probabilistic graphical model representation in phylogenetics. Syst. Biol. 63:753–771.
Höhna S., Landis M.J., Heath T.A., Boussau B., Lartillot N., Moore B.R., Huelsenbeck J.P., Ronquist F. 2016. RevBayes: Bayesian phylogenetic inference using graphical models and an interactive model-specification language. Syst. Biol. 65:726–736.
Holland P.W., Welsch R.E. 1977. Robust regression using iteratively reweighted least-squares. Commun. Stat. - Theory Methods 6:813–827.
Huber P.J. 1973. Robust regression: asymptotics, conjectures and Monte Carlo. Ann. Stat. 1:799–821.
Huber P.J. 1992. Robust estimation of a location parameter. Breakthroughs in statistics. New York. Springer. 492–518.
Huber P.J. 2004. Robust statistics. New York. John Wiley & Sons.
Huey R.B., Garland T. Jr, Turelli M. 2019. Revisiting a key innovation in evolutionary biology: Felsenstein's "phylogenies and the comparative method". Am. Nat. 193:755–772.
Ives A.R. 2019. R s for Correlated Data: phylogenetic models, LMMs, and GLMMs. Syst. Biol. 68:234–251.
Jones F.C., Grabherr M.G., Chan Y.F., Russell P., Maceli E., et al. 2012. The genomic basis of adaptive evolution in threespine sticklebacks. Nature 484:55–61.
Judge G.G., Griffths W.E., Hill E.C., Lutkepöhl H. Lee T.C. 1985. The Theory and Practice of Econometrics, 2nd edn. New York, Wiley.
Kanwal S., Choi Y.J., Lim S.O., Choi H.J., Park J.H., Nuzhat R., Khan A., Perveen S., Choi B.-O., Chung K.W. 2021. Novel homozygous mutations in Pakistani families with Charcot-Marie-Tooth disease. BMC Med. Genomics 14:174.
Kariya T., Kurata H. 2004. Generalized least squares. England: John Wiley & Sons.
King M.C., Wislon A.C. 1975. Evolution at two levels in humans and chimpanzees. Science 1884184:107–116.
Koenker R., Bassett G. Jr. 1978. Regression quantiles. Econometrica 46:33–50.
Lande R. 1979. Quantitative genetic analysis of multivariate evolution, applied to brain: body size allometry. Evolution 33:402–416.
Landis M.J., Schraiber J.G. 2017. Pulsed evolution shaped modern vertebrate body sizes. Proc. Natl. Acad. Sci. USA. 114:13224–13229.
Landis M.J., Schraiber J.G., Liang M. 2013. Phylogenetic analysis using Lévy processes: finding jumps in the evolution of continuous traits. Syst. Biol. 62:193–204.
Mack K.L., Ballinger M.A., Phifer-Rixey M., Nachman M.W. 2018. Gene regulation underlies environmental adaptation in house mice. Genome Res. 28:1636–1645.
Maddison W.P., FitzJohn R.G. 2015. The unsolved challenge to phylogenetic correlation tests for categorical characters. Syst. Biol. 64:127–136.
Maddison W.P., Midford P.E., Otto S.P. 2007. Estimating a binary character's effect on speciation and extinction. Syst. Biol. 56:701–710.
Maechler M. 2014. CRAN task view: robust statistical methods: Basic Robust Statistics.
Makino T., Kawata M. 2019. Invasive invertebrates associated with highly duplicated gene content. Mol. Ecol. 28:1652–1663.
Maronna R.A., Martin R.D., Yohai V.J., Salibián-Barrera M. 2019. Robust statistics: theory and methods (with R). New York. John Wiley & Sons.
Martins E.P., Hansen T.F. 1997. Phylogenies and the comparative method: a general approach to incorporating phylogenetic information into the analysis of interspecific data. Am. Nat. 149:646–667.
Mazel F., Davies T.J., Georges D., Lavergne S., Thuiller W., Peres-Neto P.R. 2016. Improving phylogenetic regression under complex evolutionary models. Ecology 97:286–293.
Mitov V., Bartoszek K., Stadler T. 2019. Automatic generation of evolutionary hypotheses using mixed Gaussian phylogenetic models. Proc. Natl. Acad. Sci. USA. 116:16921–16926.
Montgomery D.C., Peck E.A., Vining G.G. 2012. Introduction to linear regression analysis. New York. John Wiley & Sons.
Mortazavi A., Williams B.A., McCue K., Schaeffer L., Wold B. 2008. Mapping and quantifying mammalian transcriptomes by RNA-seq. Nat. Methods 5:621–628.
Mundry R. 2014. Statistical issues and assumptions of phylogenetic generalized least squares. Modern phylogenetic comparative methods and their application in evolutionary biology. Berlin: Springer; p. 131–153.
O'Leary N.A., Write M.W., Brister J.R., Ciufo S., Haddad D., et al. 2016. Reference sequence (RefSeq) databaset at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res. 44:D733–D745.
O'Meara B.C. 2012. Evolutionary inferences from phylogenies: a review of methods. Annu. Rev. Ecol. Evol. Syst. 43:267–285.
O'Meara B.C., Ané C., Sanderson M.J., Wainwright P.C. 2006. Testing for different rates of continuous trait evolution using likelihood. Evolution 60:922–933.
Ohno, S. 1970. Evolution by gene duplication. Springer-Verlag.
Osorio F., Wolodzko T., Osorio M.F. 2017. Package "L1pack..". https://vps.fmvz.usp.br/CRAN/web/packages/L1pack/L1pack.pdf
Pagel M. 1997. Inferring evolutionary processes from phylogenies. Zoologica Scripta 26:331–348.
Pagel M. 1999. Inferring the historical patterns of biological evolution. Nature 401:877–884.
Paradis E., Claude J., Strimmer K. 2004. APE: analyses of phylogenetics and evolution in R language. Bioinformatics 20:289–290.
Pennell M.W., Eastman J.M., Slater G.J., Brown J.W., Uyeda J.C., FitzJohn R.G., Alfaro M.E., Harmon L.J. 2014. geiger v2 0: an expanded suite of methods for fitting macroevolutionary models to phylogenetic trees. Bioinformatics 30:2216–2218.

Pennell M.W., Harmon L.J. 2013. An integrative view of phylogenetic comparative methods: connections to population genetics, community ecology, and paleobiology. Ann. N. Y. Acad. Sci. 1289:90–105.
Poole M.A., O'Farrell P.N. 1971. The assumptions of the linear regression model. Trans. Inst. Br. Geogr 52:145–158.
Puttick M.N. 2018. Mixed evidence for early bursts of morphological evolution in extant clades. J. Evol. Biol. 31:502–515.
Queen J.P., Quinn G.P., Keough M.J. 2002. Experimental design and data analysis for biologists. Massachusetts: Cambridge University Press.
Rabosky D.L. 2014. Automatic detection of key innovations, rate shifts, and diversity-dependence on phylogenetic trees. PLoS One 9:e89543.
Rencher A.C., Schaalje G.B. 2008. Linear models in statistics. New Jersey, USA: John Wiley & Sons.
Revell L.J. 2008. On the analysis of evolutionary change along single branches in a phylogeny. Am. Nat. 172:140–147.
Revell L.J. 2010. Phylogenetic signal and linear regression on species data. Methods Ecol. Evol. 1:319–329.
Revell L.J. 2012. phytools: an R package for phylogenetic comparative biology (and other things). Methods Ecol. Evol. 3:217–223.
Revell L.J., Harmon L.J. 2008. Testing quantitative genetic hypotheses about the evolutionary rate matrix for continuous characters. Evol. Ecol. Res. 10:311–331.
Revell L.J., Harmon L.J., Collar D.C. 2008. Phylogenetic signal, evolutionary process, and rate. Syst. Biol. 57:591–601.
Ripley B. 2015. MASS: Support Functions and Datasets for Venables and Ripley's MASS. R package version 7.3-45.
Rohlf F.J. 2001. Comparative methods for the analysis of continuous variables: geometric interpretations. Evolution 55:2143–2160.
Rousseeuw P., Yohai V. 1984. Robust regression by means of S-estimators. Robust and nonlinear time series analysis. New York. Springer. p. 256–272.
Scales J.A., King A.A., Butler M.A. 2009. Running for your life or running for your dinner: what drives fiber-type evolution in lizard locomotor muscles? Am. Nat. 173:543–553.
Schluter D. 2000. The ecology of adaptive radiation. UK: OUP Oxford.
Seber G.A.F., Lee A.J. 2012. Linear regression analysis. New Jersey, USA: John Wiley & Sons.
Simpson G.G. 1944. Tempo and mode in evolution. New York. Columbia University Press.
Slater G.J., Pennell M.W. 2014. Robust regression and posterior predictive simulation increase power to detect early bursts of trait evolution. Syst. Biol. 63:293–308.
Sokal R.R., Rohlf F.J. 1995. Biometry. New York, USA: Macmillan.
Stadler T. 2011. Mammalian phylogeny reveals recent diversification rate shifts. Proc. Natl. Acad. Sci. USA. 108:6187–6192.
Symonds M.R.E., Blomberg S.P. 2014. A primer on phylogenetic generalised least squares. Modern phylogenetic comparative methods and their application in evolutionary biology. Berlin Heidelberg: Springer, chap. 5, pp. 105–130.
Team R.C. 2013. R: A language and environment for statistical computing.
Tukey, J.W. 1975. "Useable resistant/robust techniques of analysis." In Proceedings of the First ERDA Statistical Symposium, Los Alamos, NM.
Uyeda J.C., Hansen T.F., Arnold S.J., Pienaar J. 2011. The million-year wait for macroevolutionary bursts. Proc. Natl. Acad. Sci. USA. 108:15908–15913.
Uyeda J.C., Harmon L.J. 2014. A novel Bayesian method for inferring and interpreting the dynamics of adaptive landscapes from phylogenetic comparative data. Syst. Biol. 63:902–918.
Uyeda J.C., Pennell M.W., Miller E.T., Maia R., McClain C.R. 2017. The evolution of energetic scaling across the vertebrate tree of life. Am. Nat. 190:185–199.
Uyeda J.C., Zenil-Ferguson R., Pennell M.W. 2018. Rethinking phylogenetic comparative methods. Syst. Biol. 67:1091–1109.
Wray G.A., Hahn M.W., Abouheif E., Baljoff J.P., Pizer M., Rockman M.V., Romano L.A. 2003. The evolution of transcriptional regulation in eukaryotes. Mol. Biol. Evol. 20:1377–1419.
Yohai V.J. 1987. High breakdown-point and high efficiency robust estimates for regression. Ann. Stat. 15:642–656.
Yu C., Yao W. 2017. Robust linear regression: a review and comparison. Commun. Stat. Comput. 46:6261–6282.
Züchner S., Wang G., Tran-Viet K.-N., Nance M.A., Gaskell P.C., Vance J.M., Ashley-Koch A.E., Pericak-Vance M.A. 2006. Mutations in the novel mitochondrial protein REEP1 cause hereditary spastic paraplegia type 31. Am. J. Hum. Genet. 79:365–369.