# Enhancing Phylogenetic Analysis: Integrating Nonparametric Methods for Addressing Abrupt Evolutionary Shifts

Zheng-Lin Chen, Hong-Ji Guo, Rui Huang and Deng-Ke Niu*

MOE Key Laboratory for Biodiversity Science and Ecological Engineering, College of Life Sciences, Beijing Normal University, Beijing, 100875, China

*Correspondence author:

Deng-Ke Niu, College of Life Sciences, Beijing Normal University, Beijing, 100875, China.

dkniu@bnu.edu.cn, dengkeniu@hotmail.com

## Abstract

Phylogenetic comparative methods are crucial for analyzing cross-species data while accounting for evolutionary relationships. Traditional methods like Felsenstein's phylogenetically independent contrasts (PIC) and phylogenetic generalized least squares (PGLS) assume gradual trait changes, which can be problematic with abrupt evolutionary shifts. This study evaluates the performance of 19 statistical methods, including nonparametric PIC, in detecting relationships between traits X1 and X2 under simulated evolutionary scenarios with and without abrupt trait jumps. Our results show that nonparametric PIC methods, specifically Spearman correlation analysis of PIC, maintain high accuracy and robustness across various shift magnitudes, effectively reducing false positives. In scenarios without evolutionary shifts, nonparametric methods perform comparably to traditional phylogenetic comparative methods. These findings underscore the robustness and simplicity of nonparametric approaches, particularly in handling data with outliers and non-normal distributions, making them valuable tools for evolutionary biologists. Despite their advantages, nonparametric methods have been largely overlooked in favor of linear regression models. We call for the integration of nonparametric methods into advanced phylogenetic comparative analyses to enhance accuracy and reliability in understanding evolutionary processes.



## Keywords

Phylogenetic comparative methods; Phylogenetically independent contrasts (PIC); Phylogenetic generalized least squares (PGLS); Nonparametric methods; Evolutionary shifts; Trait correlation; Spearman correlation; Brownian motion model; Robustness in evolutionary analysis.

## Introduction

When analyzing cross-species data, species that are phylogenetically closer typically exhibit more similar traits due to their shared ancestry. This phylogenetic structure implies that samples are not independent but exhibit autocorrelation (Felsenstein 1985). Ignoring this autocorrelation and using conventional statistical methods can lead to erroneous conclusions (Revell 2010). Consequently, evolutionary biologists have proposed various phylogenetic comparative methods to address this issue (Grafen 1989, Lynch 1991, Garamszegi 2014).

The earliest method was Felsenstein's phylogenetically independent contrasts (PIC) (Felsenstein 1985). This method uses trait changes along branches (contrasts) and performs ordinary linear regression on these contrasts to determine trait relationships. Building on PIC, the phylogenetic generalized least squares (PGLS) method was proposed, which assumes a variance-covariance matrix for the residuals in the generalized linear model to account for phylogenetic autocorrelation among samples (Martins and Hansen 1997, Pagel 1997, Pagel 1999, Rohlf 2001). This consideration reduces the occurrence of false positives to some extent. Although some PGLS models (such as Pagel’s lambda, Ornstein–Uhlenbeck, and Early-burst) can handle heterogeneous evolutionary rates (Hansen 1997, Freckleton et al. 2002, Harmon et al. 2010, Ho and Ane 2014), they fundamentally assume gradual trait changes (Landis et al. 2013). This assumption can be problematic when dealing with abrupt evolutionary shifts caused by singular events.

During biological evolution, certain branches may experience rare and dramatic trait changes (Maddison and FitzJohn 2015). Increasingly, studies and data have shown that abrupt evolutionary shifts have a significant impact on biodiversity evolution (Landis and Schraiber 2017, Gao and Wu 2022, Smith et al. 2023, Sumner et al. 2023). A typical example of this is Felsenstein’s worst-case scenario (Felsenstein 1985, Uyeda et al. 2018), where two traits undergo a significant jump near the root of the phylogenetic tree while

evolving independently on other branches. Even if this dramatic change occurred only once, it can still lead to erroneous conclusions of significant correlation between the two traits using ordinary linear regression on contrasts or PGLS models (Uyeda et al. 2018).

Recently, more studies have focused on identifying jump events in trait evolution and providing more reasonable explanations for trait evolution. Uyeda et al. (2018) underscore the need for models like the singular events model, which can accommodate the complexities and unpredictabilities inherent in evolutionary biology. Some studies suggest that jump events cause trait changes to exhibit a “fat-tails” distribution, using Levy processes to handle this distribution (Landis et al. 2013, Duchen et al. 2017). Others propose that jump events lead to outliers, using robust regression or robust estimators to reduce model sensitivity to violations (Slater and Pennell 2014, Adams et al. 2024), thereby obtaining the true relationships between traits. These methods typically involve more complex models.

Dramatic changes in some evolutionary branches can result in several outliers in the contrast dataset, distorting the statistical results of parametric tests. Non-parametric tests, which measure the strength and direction of association between two variables without assuming a specific distribution, can effectively handle data with outliers, avoiding false positives. Therefore, we recommend using nonparametric phylogenetic correction to analyze relationships between traits with jump evolution events. Specifically, combining the existing PIC method with non-parametric correlation tests; after obtaining the trait contrasts, their normality is tested, and based on the results, either parametric or non-parametric correlation analysis is chosen. Our simulation data analysis shows that incorporating non-parametric tests into PIC correlation analysis can significantly reduce the probability of false positives in the presence of strong shifts. Additionally, when shifts are weak or absent, results are comparable to those obtained using PGLS and traditional PIC methods.

# Materials and Methods

## Simulated Data with Trait Jumps

First, we simulated a balanced binary phylogenetic tree with 128 species (*n*=128). We then simulated two traits (X1 and X2) on this tree. On one branch diverging from the root node, both traits underwent a jump, with the jump magnitude taking values of 0.01, 0.1, 1, 5, 10, 25, 125, and 250. Traits on other branches evolved under the Brownian motion (BM) model with a variance of 1. Each jump gradient was simulated 1000 times, totaling 8000 simulations. Although the two traits were designed to be independent, significant correlations occasionally appeared, likely due to noise occurring in the same direction.

## Simulated Data without Trait Jumps

We simulated a balanced binary phylogenetic tree with 128 species (*n*=128) and then simulated two traits (X1 and X2) on this tree with the relationship X2 = X1 + e, where e represents a noise term. By adjusting the variance of the noise term e, we varied the correlation strength between X1 and X2. Referencing Revell (2010), our simulations included the following scenarios:

1. **BM&BM+Norm**: X1 evolves under Brownian motion (variance = 1), and e follows a normal distribution (mean = 0, variance = d).
2. **Norm&Norm+BM**: X1 follows a normal distribution (mean = 0, variance = 1), and e evolves under Brownian motion (variance = d).
3. **BM1&BM1+BM2**: X1 evolves under Brownian motion (variance = 1), and e evolves under Brownian motion (variance = d).
4. **Norm1&Norm1+Norm2**: X1 follows a normal distribution (mean = 0, variance = 1), and e follows a normal distribution (mean = 0, variance = d).

The d takes values of 0.0001, 0.01, 1, 4, 16, 64, 256, and 1024 in all the four scenarios. For each scenario and value of d, we conducted 1000 replicate simulations, totaling 32,000 simulations.

**Model Evaluation Criteria**

We aimed to compare the performance of different methods in avoiding false-positive relationships between X1 and X2. A robust method should balance the avoidance of false positives with the ability to detect rare but genuine correlations arising from random noise. Thus, an effective method should minimize false positives while not overlooking rare meaningful correlations due to random noise.

We evaluated methods by comparing their predicted trait correlations to the true evolutionary relationships between X1 and X2 inferred from traditional correlation tests between ΔX1/L and ΔX2/L, where L is branch length. If ΔX1/L and ΔX2/L followed a normal distribution, Pearson correlation was used; otherwise, Spearman rank tests were conducted. Correct results were those aligning with this standard.

**Statistical Methods Compared**

We compared 19 statistical methods for their performance in examining the evolutionary relationship between X1 and X2 amid an abrupt evolutionary jump. These methods included:

- **Pearson-Correlation-Trait**: Pearson correlation analysis of trait values.
- **Spearman-Correlation-Trait**: Spearman rank correlation analysis of trait values.
- **Selected-Correlation-Trait**: Either Pearson or Spearman correlation based on normality test results of trait values.
- **Pearson-Correlation-PIC**: Pearson correlation analysis of PIC values.
- **Spearman-Correlation-PIC**: Spearman rank correlation analysis of PIC values.

- **Selected-Correlation-PIC**: Either Pearson or Spearman rank correlation based on normality test results of PIC values.

Additionally, we included 13 PGLS (Phylogenetic Generalized Least Squares) methods:

- **PGLS-BrownianMotion**: PGLS using the Brownian Motion model (Felsenstein 1985).
- **PGLS-Lambda-X1Dependent**: PGLS using Pagel's Lambda model (Pagel 1999) with X1 as the dependent variable.
- **PGLS-Lambda-X2Dependent**: PGLS using Pagel's Lambda model with X2 as the dependent variable.
- **PGLS-Lambda-SelectedDependent**: PGLS using Pagel's Lambda model with the dependent variable selected based on stronger phylogenetic signal.
- **PGLS-OU-Random-X1Dependent**: PGLS using the Ornstein-Uhlenbeck (OU) random model (Hansen 1997) with X1 as the dependent variable.
- **PGLS-OU-Random-X2Dependent**: PGLS using the OU random model with X2 as the dependent variable.
- **PGLS-OU-Random-SelectedDependent**: PGLS using the OU random model with the dependent variable selected based on stronger phylogenetic signal.
- **PGLS-OU-Fixed-X1Dependent**: PGLS using the OU fixed model with X1 as the dependent variable.
- **PGLS-OU-Fixed-X2Dependent**: PGLS using the OU fixed model with X2 as the dependent variable.
- **PGLS-OU-Fixed-SelectedDependent**: PGLS using the OU fixed model with the dependent variable selected based on stronger phylogenetic signal.
- **PGLS-EarlyBurst-X1Dependent**: PGLS using the Early Burst model (Harmon et al. 2010) with X1 as the dependent variable.

- **PGLS-EarlyBurst-X2Dependent**: PGLS using the Early Burst model with X2 as the dependent variable.
- **PGLS-EarlyBurst-SelectedDependent**: PGLS using the Early Burst model with the dependent variable selected based on stronger phylogenetic signal.

The Pagel's Lambda, OU, and Early Burst models interpret and adjust for evolutionary processes differently based on the roles of the variables, which means switching dependent and independent variables in these models may lead to conflicting results (Chen et al. 2023). Therefore, three methods were derived from each of these models concerning the choice of dependent and independent variables. The BM model, named PGLS-BrownianMotion, being simpler and more symmetric in its treatment of variables, is unaffected by such switches, leading to consistent results regardless of which variable is dependent or independent. Thus, we only used X1 as the dependent variable in the PGLS regression analysis with the BM model.

**Programs and Packages**

All simulations and analyses were performed in the R environment (version 4.0.2) (R Core Team 2020) on a Linux system. The geiger *package* (version 2.0.11) (Pennell et al. 2014) was used for simulations. Normality tests were conducted using the Shapiro-Wilk test (Royston 1992) via the *shapiro.test()* function from the *stats* package (version 4.3.1). Pearson and Spearman correlation analyses were performed using the *cor.test()* function from the *stats* package. PICs were calculated using the *pic()* function from the ape package (version 5.8) (Paradis and Schliep 2019). All PGLS regressions were conducted using the *phylolm()* function from the *phylolm* package (version 2.6.2) (Ho and Ane 2014). Phylogenetic signals (Pagel's λ) were estimated using the *phylosig()* function from the *phytools* package (version 1.0-3) (Revell 2012).

# Results

## Nonparametric PIC Analysis Reduces False Positives Amid Strong Shifts

We evaluated the accuracy of different statistical methods in detecting the relationship between traits X1 and X2 under an evolutionary scenario involving a single branch shift. Table 1 presents the accuracy of these methods across various shift magnitudes. Figure 1 illustrates the impact of these shifts, showcasing the outcomes of three simulations with shift values of 1, 5, and 10.

The results indicate that traditional correlation methods applied to trait values show lower accuracy, particularly at lower shift magnitudes, compared to phylogenetic comparative methods. However, at higher shift magnitudes, their performance becomes similar. Except for the two nonparametric PIC methods, the accuracy of all other methods generally decreases as shift magnitude increases. Notably, the Early Burst model showed strong performance for larger shift magnitudes. Pearson-Correlation-PIC and PGLS-BrownianMotion provided similar results, indicating consistency in accounting for phylogenetic relationships under a Brownian motion pattern.

Strikingly, nonparametric Spearman-Correlation-PIC and Selected-Correlation-PIC exhibited consistent high performance across all shift magnitudes. This result suggests nonparametric PIC methods are more robust and less affected by evolutionary shifts, particularly because nonparametric methods are effective at handling data with outliers and non-normal distributions. However, Selected-Correlation-PIC did not show higher accuracy than Spearman-Correlation-PIC, indicating that the choice between Pearson and Spearman based on normality does not significantly increase accuracy.

Moreover, selecting the dependent variable with a stronger phylogenetic signal in the PGLS models did not result in higher accuracy.

Table 1. Accuracy of Methods for Examining X1-X2 Relationship with Single Branch Shift

| Methods | Shifts | | | | | |
|---|---|---|---|---|---|---|
| | 0.1 | 1 | 5 | 10 | 25 | 125 |
| Pearson-Correlation-Trait | 42.9% | 53.1% | 4.5% | 4.2% | 4.9% | 3.4% |
| Spearman-Correlation-Trait | 41.9% | 53.0% | 4.5% | 4.2% | 4.9% | 3.4% |
| Selected-Correlation-Trait | 43.0% | 53.7% | 4.5% | 4.2% | 4.9% | 3.4% |
| Pearson-Correlation-PIC | 93.1% | 93.6% | 80.2% | 28.6% | 4.9% | 3.4% |
| Spearman-Correlation-PIC | 94.3% | 92.9% | 92.6% | 93.8% | 92.0% | 92.7% |
| Selected-Correlation-PIC | 93.1% | 93.4% | 90.6% | 93.2% | 92.0% | 92.7% |
| PGLS-BrownianMotion | 93.2% | 93.6% | 80.5% | 29.2% | 4.9% | 3.4% |
| PGLS-Lambda-X1Dependent | 93.4% | 93.3% | 80.1% | 28.7% | 4.9% | 3.4% |
| PGLS-Lambda-X2Dependent | 92.9% | 93.5% | 80.2% | 29.0% | 4.9% | 3.4% |
| PGLS-Lambda-SelectedDependent | 93.1% | 93.5% | 80.3% | 29.1% | 4.9% | 3.4% |
| PGLS-OU-Random-X1Dependent | 93.7% | 93.4% | 76.9% | 26.8% | 4.9% | 3.4% |
| PGLS-OU-Random-X2Dependent | 93.4% | 93.0% | 77.4% | 26.8% | 4.9% | 3.4% |
| PGLS-OU-Random-SelectedDependent | 93.6% | 93.8% | 79.2% | 27.6% | 4.9% | 3.4% |
| PGLS-OU-Fixed-X1Dependent | 93.7% | 93.4% | 76.9% | 26.8% | 4.9% | 3.4% |
| PGLS-OU-Fixed-X2Dependent | 93.4% | 93.0% | 77.4% | 26.8% | 4.9% | 3.4% |
| PGLS-OU-Fixed-SelectedDependent | 93.1% | 93.6% | 80.5% | 29.2% | 4.9% | 3.4% |
| PGLS-EarlyBurst-X1Dependent | 93.2% | 93.6% | 90.7% | 83.7% | 28.2% | 3.4% |
| PGLS-EarlyBurst-X2Dependent | 93.3% | 93.6% | 90.0% | 84.5% | 28.2% | 3.4% |
| PGLS-EarlyBurst-SelectedDependent | 93.3% | 93.6% | 91.4% | 85.5% | 28.2% | 3.4% |

In a balanced binary phylogenetic tree with 128 species, both traits X1 and X2 experienced a sudden change on one of the branches diverging from the root node. The traits on all other branches evolved according to the Brownian motion model with a variance of 1. Each shift gradient was simulated 1000 times. Due to space limitations, we present only six shifts in this table. Please refer to

Supplementary Table S1 for all results of the eight shift gradients. For detailed definitions of the methods listed in this table and the criteria used to evaluate the accuracy of the models, please refer to the Materials and Methods section.

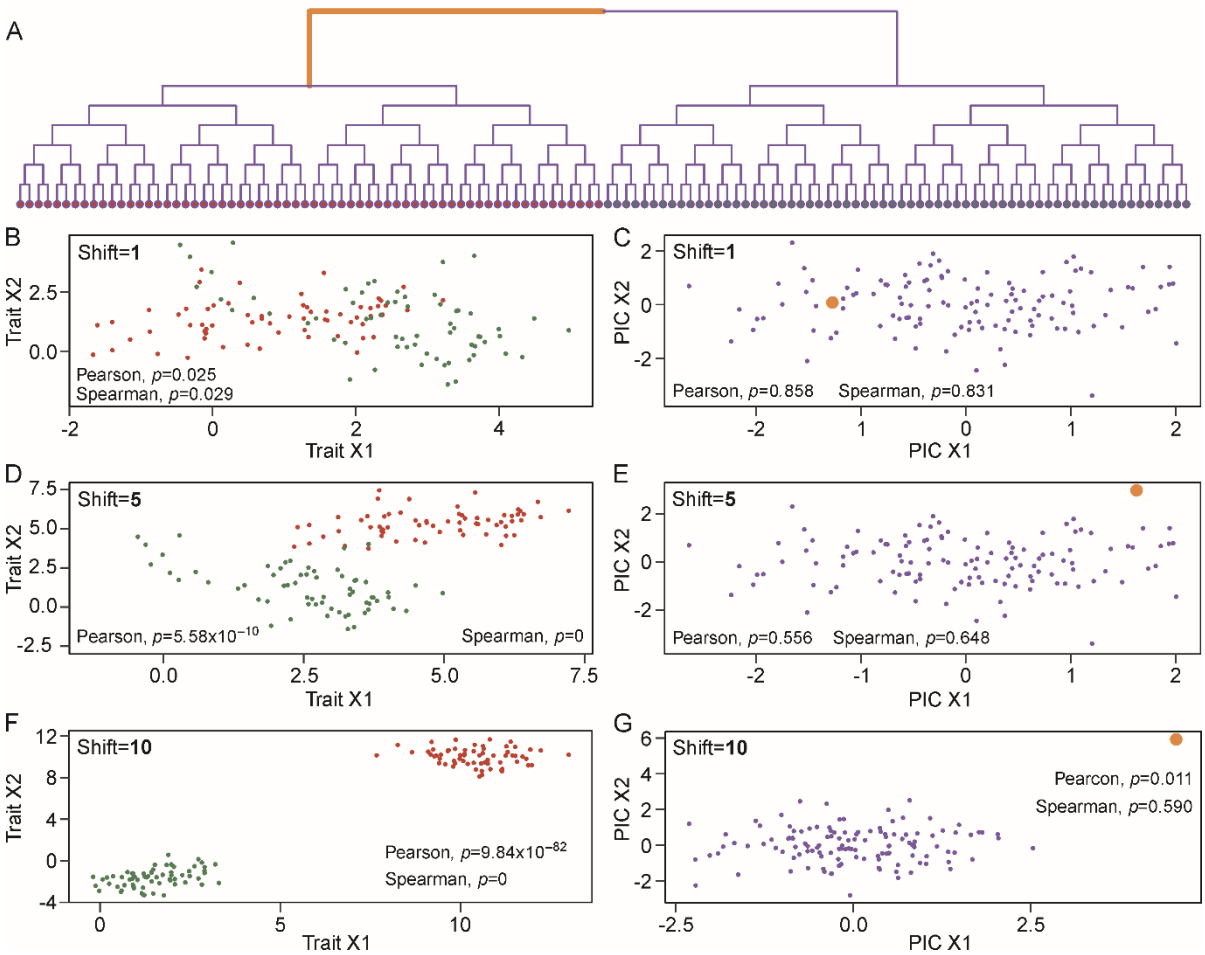


Figure 1. Effects of Large Shifts on PIC and Correlation Analysis. (A) A balanced binary phylogenetic tree with 128 species, with the orange thick line indicating a large evolutionary shift. (B, C) Simulation outcomes with shift = 1. (D, E) Simulation outcomes with shift = 5. (F, G) Simulation outcomes with shift = 10. The simulations demonstrate that large evolutionary shifts result in outliers (marked by enlarged orange dots in C, E, and G) in PIC (Phylogenetically Independent Contrasts). The figure highlights the performance of traditional correlation methods and Pearson correlation of PIC under different shift magnitudes. Traditional correlation methods consistently produce false positive results. Pearson correlation of PIC yields correct results when the shift is small but leads to false-positive results as the shift magnitude increases.

**Comparable Accuracy of Nonparametric PIC Analysis Without Evolutionary Shifts**

Nonparametric methods, which do not rely on data distribution assumptions, are generally robust to outliers but less sensitive to small changes, potentially overlooking subtle relationships. In the above simulated dataset with trait jumps, there is typically no significant correlation between X1 and X2. The disadvantage of nonparametric methods in this context is not severe. Our concern is that nonparametric PIC analysis may be considerably less accurate when dealing with datasets where X1 and X2 are significantly correlated in most cases. For this reason, we conducted 28,000 replicate simulations across four scenarios, varying the correlation strength between X1 and X2 by adjusting the noise term variance.

Table 2 and Supplementary Table S2 present the accuracy of these methods across scenarios and variances. Phylogenetic comparative methods generally showed higher accuracy than traditional correlation analyses, particularly when noise variance was large. Among the 16 phylogenetic comparative methods, the two nonparametric methods had accuracy comparable to the other methods. Selected-Correlation-PIC performed slightly better than Spearman-Correlation-PIC in some scenarios. These findings highlight the robust performance of nonparametric methods in various scenarios.

In PGLS analyses with the Lambda or OU model, the choice of dependent and independent variables significantly affected results when noise variance was high (Supplementary Table S2). In the most extreme case, where the noise variance is 16 in the scenario Norm&Norm+BM, PGLS-OU-Random-X1Dependent and PGLS-OU-Random-X2Dependent show very divergent accuracy, 32.9% vs. 92.4%, highlighting the need for caution in choosing the dependent variable when the causal relationship between analyzed traits is unclear.

Table 2. Accuracy of Methods for Examining X1-X2 Relationship Without Shift

| Scenarios and Methods | Variance | | | | | |
|---|---|---|---|---|---|---|
| | 1 | 4 | 16 | 64 | 256 | 1024 |
| BM1&BM1+BM2 | | | | | | |
| Pearson-Correlation-Trait | 97.7% | 85.4% | 61.8% | 46.6% | 45.5% | 52.6% |
| Pearson-Correlation-PIC | 100.0% | 100.0% | 80.0% | 71.9% | 86.6% | 91.6% |
| Spearman-Correlation-PIC | 100.0% | 100.0% | 75.7% | 69.1% | 85.6% | 91.1% |
| Selected-Correlation-PIC | 100.0% | 100.0% | 79.3% | 71.2% | 86.4% | 91.3% |
| PGLS-median-accuracy | 100.0% | 100.0% | 80.9% | 71.4% | 86.1% | 91.4% |
| BM&BM+Norm | | | | | | |
| Pearson-Correlation-Trait | 99.8% | 75.6% | 31.0% | 39.8% | 69.7% | 82.1% |
| Pearson-Correlation-PIC | 99.8% | 75.5% | 53.6% | 73.0% | 84.1% | 85.8% |
| Spearman-Correlation-PIC | 99.8% | 75.3% | 52.5% | 72.4% | 82.4% | 84.8% |
| Selected-Correlation-PIC | 99.8% | 75.3% | 53.0% | 72.3% | 82.3% | 84.9% |
| PGLS-median-accuracy | 99.8% | 75.5% | 53.2% | 73.5% | 83.8% | 85.7% |
| Norm&Norm+BM | | | | | | |
| Pearson-Correlation-Trait | 100.0% | 82.3% | 31.2% | 14.1% | 24.1% | 59.6% |
| Pearson-Correlation-PIC | 100.0% | 100.0% | 93.6% | 41.3% | 31.3% | 61.5% |
| Spearman-Correlation-PIC | 100.0% | 100.0% | 88.7% | 35.8% | 29.7% | 61.2% |
| Selected-Correlation-PIC | 100.0% | 100.0% | 89.3% | 36.9% | 29.9% | 61.4% |
| PGLS-median-accuracy | 100.0% | 100.0% | 92.8% | 39.6% | 30.6% | 61.5% |
| Norm1&Norm1+Norm2 | | | | | | |
| Pearson-Correlation-Trait | 100.0% | 99.9% | 69.5% | 58.0% | 74.9% | 81.2% |
| Pearson-Correlation-PIC | 100.0% | 98.9% | 66.9% | 57.4% | 65.8% | 70.0% |
| Spearman-Correlation-PIC | 100.0% | 99.8% | 66.2% | 58.7% | 72.0% | 78.1% |
| Selected-Correlation-PIC | 100.0% | 99.7% | 65.9% | 58.6% | 71.7% | 78.1% |
| PGLS-median-accuracy | 100.0% | 99.7% | 71.9% | 58.6% | 74.0% | 78.8% |

In a balanced binary phylogenetic tree with 128 species, traits X1 and X2 were simulated with the relationship X2=X1+e, where e represents a noise term. There are no sudden changes on any branches; the traits evolved according to the four scenarios. Due to space limitations, we present only six variances of e and the median accuracy of the PGLS methods in this table. Please refer to Supplementary Table S2 for more variances and the accuracy of additional methods. For detailed definitions of the methods listed in this table and the details of the four scenarios, please refer to the Materials and Methods section.

# Disscussion

## Comparing Traditional (Phylogenetic) Methods for Detecting Trait Correlations

Our study evaluated the performance of various statistical methods in detecting relationships between traits X1 and X2 under different evolutionary scenarios. We found that nonparametric PIC methods, particularly Spearman-Correlation-PIC, demonstrated exceptional robustness and consistency across various shift magnitudes, effectively reducing false positives. In scenarios without evolutionary shifts, nonparametric methods maintained comparable accuracy to traditional phylogenetic comparative methods.

The consistent high performance of nonparametric PIC methods across different shift magnitudes highlights their robustness in handling data with outliers and non-normal distributions. This robustness is crucial for evolutionary biologists, as it ensures more reliable detection of true evolutionary relationships without being misled by abrupt shifts or noise in the data. Furthermore, the comparable accuracy of nonparametric methods in scenarios without evolutionary shifts suggests their versatility and reliability across various evolutionary contexts.

## Comparison with Other Advanced Methods Dealing with Evolutionary Shifts

In recent years, several advanced methods have been developed to manage abrupt, lineage-specific evolutionary shifts, such as singular events models and robust linear regression (Landis et al. 2013, Slater and Pennell 2014, Duchen et al. 2017, Uyeda et al. 2018, Adams et al. 2024). Compared to Spearman's rank correlation analysis of PIC, these methods typically involve more complex models. Spearman's rank correlation analysis does not require normality or linearity assumptions, simplifying data preparation and analysis. Its simplicity and ease of understanding offer distinct advantages, particularly for biologists with varying levels of statistical training.

A simpler method is easier to grasp conceptually compared to more complex methods like robust phylogenetic regression. Spearman's rank correlation is based on ranking data, an intuitive concept for most scientists, measuring monotonic relationships without assuming data distribution, making it straightforward to understand and interpret. Simplified methods reduce the risk of misinterpretation from complex statistical techniques, ensuring more transparent and interpretable results.

**Rationale for Using ΔX1/L and ΔX2/L**

Some readers might question why we did not use the absence of correlation as the standard for the "Simulated Data with Trait Jumps" and the presence of correlation for the "Simulated Data without Trait Jumps." While this approach might seem intuitive given the simulation parameters, it does not fully capture the complexities of evolutionary processes. Using ΔX1/L and ΔX2/L offers several advantages:

1. **Addressing Evolutionary Complexity**: This approach considers the rate of trait change per unit time along each branch, providing a more nuanced understanding of the evolutionary relationship. It accounts for independent changes on each branch, aligning more closely with the evolutionary model's assumptions.

2. **Avoiding Oversimplification**: Simply assuming no correlation or correlation based on initial simulation parameters may oversimplify evolutionary processes. Evolutionary shifts, noise, and other factors introduce complexities that are better captured by analyzing ΔX1/L and ΔX2/L.
3. **Consistency Across Methods**: Traditional correlation tests between ΔX1/L and ΔX2/L provide a consistent criterion, allowing direct comparison across different methods and avoiding biases from initial simulation assumptions.
4. **Handling Noise and Outliers**: This method accounts for random noise and outliers, inherent in biological data, ensuring the evaluation criterion reflects realistic data scenarios.

In summary, our choice of evaluation criterion using ΔX1/L and ΔX2/L provides a more reliable and consistent standard for comparing the performance of different statistical methods in detecting trait correlations. This approach aligns with our goal of accurately reflecting true evolutionary relationships between traits.

**Limitations of Spearman's Rank Correlation Analysis of PIC**

While Spearman's rank correlation analysis of PIC offers several advantages, it also has limitations. It cannot capture non-monotonic relationships and can be less powerful than parametric methods when data meet normality and linearity assumptions, potentially reducing sensitivity in detecting true associations. However, our simulations indicate this limitation may not be critical, as nonparametric PIC methods maintained comparable accuracy in scenarios without evolutionary shifts (Table 2 and Supplementary Table S2). Moreover, while robust to outliers, Spearman's rank correlation can be influenced by tied ranks, common in biological data. Unlike regression analysis, correlation analysis does not allow modeling complex relationships, such as multiple predictors or interactions between

variables. Therefore, while valuable for initial analyses, it should be complemented with more sophisticated methods for detailed evolutionary process inferences.

## Conclusion

Our findings highlight the critical importance of selecting appropriate methods for phylogenetic comparative analyses. Nonparametric PIC methods, particularly Spearman rank correlation analysis of PIC, demonstrate significant advantages due to their simplicity, robustness, and ease of understanding. These methods are especially valuable for evolutionary biologists with diverse backgrounds and varying levels of statistical expertise. Despite their benefits, nonparametric methods have been largely overlooked by evolutionary biologists in favor of linear regression models. Our study emphasizes the need for integrating nonparametric methods into advanced phylogenetic comparative analyses. This integration can enhance the accuracy and reliability of trait correlation analyses across diverse datasets and better accommodate the complexities and nuances of evolutionary processes.

**Data availability**

The codes to generate the results and raw data generated in this study are presented in Supplementary Data.zip.

**Supplementary material**

Supplementary Tables S1, S2, and Supplementary Data.zip are submitted as Supplementary materials online of this manuscript.

**Acknowledgements**

This study was funded by the National Natural Science Foundation of China (grant number 31671321). Many sentences in this article were written with the assistance of ChatGPT

(OpenAI 2024). The authors thoroughly reviewed and revised each sentence and take full responsibility for the language and content of the entire article.

**Author Contributions**

DKN and ZLC conceived the research hypothesis; ZLC, HJG and RH performed the simulations and analyzed the data. DKN and ZLC drafted and revised the manuscript. All authors have read and agree with the final version and the submission.

**References**

Adams R., Cain Z., Assis R., DeGiorgio M. 2024. Robust phylogenetic regression. Syst. Biol., 73:140-157.

Chen Z.-L., Guo H.-J., Niu D.-K. 2023. Dependent variable selection in phylogenetic generalized least squares regression analysis under Pagel's lambda model. bioRxiv:2023.2005.2021.541623.

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

Felsenstein J. 1985. Phylogenies and the comparative method. Am. Nat., 125:1-15.

Freckleton R.P., Harvey P.H., Pagel M. 2002. Phylogenetic analysis and comparative data: A test and review of evidence. Amer. Natur., 160:712-726.

Gao Y., Wu M. 2022. Microbial genomic trait evolution is dominated by frequent and rare pulsed evolution. Science Advances, 8:eabn1916.

Garamszegi L.Z. 2014. Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology: Concepts and Practice. Berlin, Springer.

Grafen A. 1989. The phylogenetic regression. Philos. Trans. R. Soc. Lond. B Biol. Sci., 326:119-157.

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., Purvis A., Ricklefs R.E., Schluter D., Schulte Ii J.A., Seehausen O., Sidlauskas B.L., Torres-Carvajal O., Weir J.T., Mooers A.Ø. 2010. Early bursts of body size and shape evolution are rare in comparative data. Evolution, 64:2385-2396.

Ho L.S.T., Ane C. 2014. A linear-time algorithm for Gaussian and non-Gaussian trait evolution models. Syst. Biol., 63:397-408.

Landis M.J., Schraiber J.G. 2017. Pulsed evolution shaped modern vertebrate body sizes. Proc Natl Acad Sci U S A, 114:13224-13229.

Landis M.J., Schraiber J.G., Liang M. 2013. Phylogenetic analysis using Levy processes: finding jumps in the evolution of continuous traits. Syst Biol, 62:193-204.

Lynch M. 1991. Methods for the analysis of comparative data in evolutionary biology. Evolution, 45:1065-1080.

Maddison W.P., FitzJohn R.G. 2015. The unsolved challenge to phylogenetic correlation tests for categorical characters. Syst. Biol., 64:127-136.

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. Amer. Natur., 149:646-667.

OpenAI. 2024. ChatGPT (June 14 version) [Large language model]. https://www.openai.com/.

Pagel M. 1997. Inferring evolutionary processes from phylogenies. Zool. Scr., 26:331-348.

Pagel M. 1999. Inferring the historical patterns of biological evolution. Nature, 401:877-884.

Paradis E., Schliep K. 2019. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics, 35:526-528.

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.

R Core Team. 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

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.

Rohlf F.J. 2001. Comparative methods for the analysis of continuous variables: geometric interpretations. Evolution, 55:2143-2160.

Royston P. 1992. Approximating the Shapiro-Wilk W-test for non-normality. Statistics and Computing, 2:117-119.

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.

Smith E.G., Surm J.M., Macrander J., Simhi A., Amir G., Sachkova M.Y., Lewandowska M., Reitzel A.M., Moran Y. 2023. Micro and macroevolution of sea anemone venom phenotype. Nat. Commun., 14:249.

Sumner S., Favreau E., Geist K., Toth A.L., Rehan S.M. 2023. Molecular patterns and processes in evolving sociality: lessons from insects. Philos Trans R Soc Lond B Biol Sci, 378:20220076.

Uyeda J.C., Zenil-Ferguson R., Pennell M.W. 2018. Rethinking phylogenetic comparative methods. Syst. Biol., 67:1091-1109.