

Classification: Biological Sciences – Evolution

**A Standardized Effect Size for Evaluating and Comparing the Strength of Phylogenetic
Signal**

Dean C. Adams^{1*}, Erica K. Baken^{1,2}, and Michael L. Collyer²

¹Department of Ecology, Evolution, and Organismal Biology, Iowa State University, Ames, Iowa, USA.
²Department of Science, Chatham University, Pittsburgh, Pennsylvania, USA.
*Correspondence: Dean C. Adams. Department of Ecology, Evolution, and Organismal Biology, Iowa State
University, Ames, Iowa, USA. 515-294-3834. dcadams@iastate.edu

15 July, 2020

words	characters
2553	16162

Abstract

Macroevolutionary studies frequently characterize the phylogenetic signal in phenotypes, and wish to compare the strength of that signal across traits. However, analytical tools for such comparisons have largely remained underdeveloped. Here we evaluate the efficacy of one commonly used parameter (Pagel's λ) to estimate the strength of phylogenetic signal in phenotypic traits, and evaluate the degree to which λ correctly identifies known levels of phylogenetic signal. We find that λ behaves as a Bernoulli random variable, and that estimates are increasingly skewed at larger and smaller input levels of phylogenetic signal. Further, the precision of λ in estimating actual levels of phylogenetic signal is often inaccurate, and biological interpretations of the strength of phylogenetic signal based on λ are therefore compromised. We propose a standardized effect size based on κ , (Z_κ), which measures the strength of phylogenetic signal more reliably than does λ , and places that signal on a common scale for statistical comparison. We develop tests based on Z_κ to provide a mechanism for formally comparing the strength of phylogenetic signal across datasets, in much the same manner as effect sizes may be used to summarize patterns in quantitative meta-analysis. Our approach extends the phylogenetic comparative toolkit to address hypotheses that compare the strength of phylogenetic signal between various phenotypic traits, even when those traits are found in different evolutionary lineages or have different units or scales.

Keywords: phylogenetic signal, macroevolution, lambda, kappa

Significance Statement

Evolutionary biologists wish to quantify and compare the strength of phylogenetic signal across traits, but analytical tools for these comparisons are generally lacking. Here we develop a standardized effect size based on κ (Z_κ), which measures the strength of phylogenetic signal on a common statistical scale, and provides a mechanism for formally comparing the strength of phylogenetic signal across datasets. Additionally, we find that a commonly used parameter (Pagel's λ) is unsuitable for this purpose. Our procedure enables biologists to quantitatively address hypotheses that compare the strength of phylogenetic signal between various phenotypic traits, even when those traits are found in different evolutionary lineages or have different units or scales.

Introduction

Investigating macroevolutionary patterns of trait variation requires a phylogenetic perspective, because the shared ancestry among species violates the assumption of independence among trait values that is common for statistical tests (1, 2). Accounting for this evolutionary non-independence is the purview of *phylogenetic comparative methods* (PCMs): a suite of analytical tools that condition trends in the data on the phylogenetic relatedness of observations (3–12). These methods are predicated on the notion that phylogenetic signal – the tendency for closely related species to display similar trait values – is present in cross-species datasets (1, 13, 14). Indeed, under numerous evolutionary models, phylogenetic signal is to be expected, as stochastic character change along the hierarchical structure of the tree of life generates trait covariation among related taxa (1, 14, 15).

Several analytical tools have been developed to quantify phylogenetic signal in phenotypic datasets (13, 14, 16–19), and their statistical properties – namely type I error rates and statistical power – have been investigated to determine under what conditions phylogenetic signal can be detected (15, 18, 20–25). One of the most widely used methods for characterizing phylogenetic signal is Pagel’s λ (13), which transforms the lengths of the internal branches of the phylogeny to improve the fit of data to the phylogeny via maximum likelihood (13, 26). When incorporated in PGLS, λ serves as a tuning parameter (ranging from $0 \rightarrow 1$) which is optimized via log-likelihood profiling while evaluating the covariation between the dependent and independent variables, given the phylogeny (13, 26). To infer whether phylogenetic signal differs from no signal or a Brownian motion model of evolutionary divergence, the observed model fit using $\hat{\lambda}$ may be statistically compared to that using $\lambda = 0$ or $\lambda = 1$ via likelihood ratio tests (26–28) or confidence limits (29). Another widely used measure is Blomberg’s κ (14), which measures the ratio of observed trait variation to the amount expected under Brownian motion. Blomberg’s κ can be treated as a test statistic by employing a permutation test to generate its sampling distribution (14, 18) for determining whether significant phylogenetic signal is present in data. Both λ and κ seem intuitive to interpret, as a value of 0 for both corresponds to no phylogenetic signal, while a value of 1 corresponds to the amount of phylogenetic signal expected under Brownian motion. Thus, it is tempting to regard both λ and κ as descriptive statistics that measure the relative strength of phylogenetic signal, providing an estimate of its magnitude for comparison.

The appeal of Pagel’s λ and Blomberg’s κ as descriptive statistics is that they provide a basis for interpreting “weak” versus “strong” phylogenetic signal; i.e., small versus large values of $\hat{\lambda}$ or κ , respectively, in a

comparative sense (30–32). Nonetheless, an important question that has yet to be considered is whether these statistics are or can be converted to effect sizes for comparative analyses across datasets? To be statistics representing phylogenetic signal, they should have reliable distributional properties, which could be revealed with simulation experiments. For instance, as a proportional random variable bounded by 0 and 1, we might expect that $\hat{\lambda}$ follows a Bernoulli distribution (**add ref**); i.e., branch lengths in a tree are scaled proportionally to the probability that data arise from a BM process. Given a known λ value used to generate random data on a tree, we would also expect that the mean of an empirical sampling distribution of $\hat{\lambda}$ would approximately equal λ ; the dispersion of $\hat{\lambda}$ would be largest at intermediate values of λ , $\hat{\lambda}$ would be predictable over the range of λ with respect to treesize; the distribution of $\hat{\lambda}$ would be symmetric at intermediate values of λ and more skewed toward values of 0 or 1; and that the distribution of $\hat{\lambda}$ will be more platykurtic at intermediate values of λ , becoming more leptokurtic toward 0 and 1 (**add same ref**). Prior work (20) seems to support some of these conjectures, based superficially on statistical moments for a given tree size (mean, variance, skewness, and kurtosis; see Fig. 2 of (20)). However, because the “strength of Brownian motion” was simulated as a varied weighted-average of data simulated on trees with $\lambda = 0$ and $\lambda = 1$ and not prescribed values of λ (20), interpretation of these patterns is challenging.

By contrast, Blomberg’s κ , which is positively ubounded, might be expected to follow a normal distribution (**add ref**). Thus we might expect that κ is symmetrically distributed across different strengths of phylogenetic signal, for any λ used to generate data. This attribute seemed less reasonable based on the simulations performed by Münkemüller et al. (20), as distributions were postively skewed, suggesting the Blomberg’s κ might not behave as a statistic that follows a noraml distribution. However, because their simulations used a weighted combination of simulated phylogenetic signal strengths, strong inferences are not possible (and distributional attributes were not the intended result of their simulations). For both Pagel’s λ or Blomberg’s κ , evaluation of statistical moments across a range of λ used to generate data would be valuable for adjudicating the relibality of these statistics. Furthermore, these are statistics that appear to have expected values that vary with tree size (20), making comparisons across studies challenging. Therefore, transformation of these statistics into Z -scores in the same simulation experiments would allow evaluation of the efficacy of each statistic to yield effect sizes that could be used for comparative anlyses.

Here we used simulation experiments to compare the distributional attributes of $\hat{\lambda}$ and Blomberg’s κ , plus their effect sizes (Z -scores), across a range of tree size and phylogenetic signal strength. We find that estimates of $\hat{\lambda}$ are increasingly skewed at larger and smaller input levels of phylogenetic signal, vary widely for a given

input value of λ , and that the precision of $\hat{\lambda}$ is not constant across its range. By contrast, estimates of κ are more consistent across tree sizes, and are normally distributed across the range of input levels of λ . We then propose an effect size based on κ (Z_κ), which provides consistent estimates of the strength of phylogenetic signal across tree sizes, and facilitates quantitative comparisons of the relative strength of phylogenetic signal across datasets.

Results

Lambda (λ) estimates of phylogenetic signal are inaccurate. Computer simulations revealed that for $\hat{\lambda}$, the distributional expectations of a Bernoulli variable were mostly upheld. First, the mean value of $\hat{\lambda}$ increased as λ increased (Fig. 1 black line); the standard deviation of $\hat{\lambda}$ was largest at intermediate values of λ and smallest at extreme values (Fig. 1 red line); and the distributions of $\hat{\lambda}$ were increasing skewed at extreme values of λ but rather normal at intermediate values (Fig. 1 blue line). For small tree sizes, it was also clear that distributions were more platykurtic at intermediate values of $\hat{\lambda}$. Second, $\hat{\lambda}$ was a negatively-biased estimate of λ for small tree sizes, as the mean value was consistently less than the input λ value. Third, standard deviations of $\hat{\lambda}$ were negatively associated with tree size, and for trees of 128 species or less, $\hat{\lambda}$ was quite variable, except for cases when λ was near or equal to 1. As such its statistical properties make it unsuitable as an effect size for measuring the strength of phylogenetic signal.

Kappa (κ) estimates of phylogenetic signal are more stable. Results for κ demonstrated better statistical properties. First, mean values increased consistently with λ irrespective of tree size (Fig. 2 black line); the standard deviation of κ was consistent across tree sizes (Fig. 2 red line); and κ was normally distributed regardless of λ or tree size (Fig. 2 blue line). The standard deviation of κ increased with λ , somewhat tracking the mean, but was always less than the corresponding means. This is perhaps unsurprising, as κ is bound by 0, and was never large for small values of λ . Finally, distributions of κ displayed slight skewing only at small tree sizes and when λ was near or equal to 1. This differs from the results of (20), where skewing appears to be from combining random values generated independently.

Effect sizes from κ (Z_κ) better characterize phylogenetic signal. We derived effect sizes (Z-scores) for both λ and κ (see Methods) and evaluated them relative to known phylogenetic signal. Both Z_λ and Z_κ were associated with input phylogenetic signal (λ), indicating that both statistics captured the observed signal (Fig. 3). However, effect sizes from $\hat{\lambda}$ made little sense, as they were more strongly

associated with tree size than the actual phylogenetic signal in the data (Fig. 3). By contrast, Z_κ was much more consistent across tree sizes, and exhibited a much stronger association with phylogenetic signal strength as compared to tree size (Fig. 3). Thus between the two statistics, Z_κ is a more reliable measure of the strength of phylogenetic signal, and may be used to compare levels of phylogenetic signal across datasets.

A test statistic (\hat{Z}_{12}) allows meaningful comparisons across datasets. We derived a two-sample test statistic (\hat{Z}_{12}) for statistically comparing the strength of phylogenetic signal across datasets (see Methods). To demonstrate its utility, we compared Z_κ for two ecologically-relevant traits in plethodontid salamander (Fig. 4): surface area to volume ratios (SA:V) and relative body width ($\frac{BW}{SVL}$) (33, 34). While both contained significant phylogenetic signal, tests based on \hat{Z}_{12} revealed that the degree of phylogenetic signal was significantly stronger in SA:V (Fig. 4). Biologically, this observation corresponds with the fact that tropical species – which form a monophyletic group within plethodontids – display greater variation in SA:V which covaries with disparity in their climatic niches (34).

Discussion

address hypotheses that compare the strength of phylogenetic signal between various phenotypic traits, even when those traits are found in different evolutionary lineages or have different units or scales

Methods

Derivation of effect sizes. Statistically, a standardized effect size may be found as:

$$Z_\theta = \frac{\theta_{obs} - E(\theta)}{\sigma_\theta} \quad (1)$$

where θ_{obs} is the observed test statistic, $E(\theta)$ is its expected value under the null hypothesis, and σ_θ is its standard error (35–37). Typically, θ_{obs} and σ_θ are estimated from the data, while $E(\theta)$ is obtained from the distribution of θ derived from parametric theory. However, recent advances in resampling theory (38–41) have shown that $E(\theta)$ and σ_θ may also be obtained from an empirical sampling distribution of θ obtained from permutation procedures.

Formalizing the suggestion of Adams and Collyer (42), an effect size for κ may be found as:

$$Z_{\kappa} = \frac{\kappa_{obs} - \hat{\mu}_{\kappa}}{\hat{\sigma}_{\kappa}}, \quad (2)$$

where κ_{obs} is the observed phylogenetic signal, and $\hat{\mu}_{\kappa}$ and $\hat{\sigma}_{\kappa}$ are the mean and standard deviation of the empirical sampling distribution of κ obtained via permutation. The empirical sampling distribution of κ is first transformed via Box-Cox to better adhere to the assumption of normality.

For λ deriving an effect size is more challenging, it does not have a sampling distribution from which the standard error (and thus confidence intervals) may be obtained. Thus, its confidence intervals are generated for the values of λ that intersect the log-likelihood profile for corresponding percentiles of the χ^2 distribution used to compare the putative appropriate model to a null model with $\lambda = 0$ [add ref: MLC thinks Boettiger paper?]. As such, an effect size for λ may be found as:

$$|Z_{\lambda}| = \sqrt{\chi_{\hat{\lambda}}^2} \quad (3)$$

where, $\hat{\lambda}$ is the maximized likelihood value of λ and $\chi_{\hat{\lambda}}^2$ is the likelihood ratio statistic for the value. Note that alternative formulations could be envisioned: $Z_{\lambda} = d\sqrt{\chi_{\hat{\lambda}}^2}$, where d is a binary value $(-1, 1)$ to indicate a direction based on whether $\hat{\lambda}$ is below or above a critical value of λ , for a quantile from a χ^2 distribution at a probability of 0.5. However, preliminary investigations found that mapping $\chi_{\hat{\lambda}}^2$ values in this manner did not produce effect sizes that were symmetrical about $Z = 0$, as the mapping was not linear and the log-likelihood profiles can be rather flat for small trees.

Derivation of two-sample test statistic. To compare the strength of phylogenetic signal across datasets, a two-sample test statistic may be calculated as:

$$\hat{Z}_{12} = \frac{|(\kappa_1 - \hat{\mu}_{\kappa_1}) - (\kappa_2 - \hat{\mu}_{\kappa_2})|}{\sqrt{\hat{\sigma}_{\kappa_1}^2 + \hat{\sigma}_{\kappa_2}^2}} = \frac{|Z_{\kappa_1} - Z_{\kappa_2}|}{\sqrt{2}} \quad (4)$$

where κ_1 , κ_2 , $\hat{\mu}_{\kappa_1}$, $\hat{\mu}_{\kappa_2}$, $\hat{\sigma}_{\kappa_1}$, and $\hat{\sigma}_{\kappa_2}$ are as defined above for equation 2. The right side of the equation illustrates that if Z_κ has already been calculated for two sampling distributions as in equation 2, the sampling distributions have unit variance for each of the Z_κ statistics. Estimates of significance of \hat{Z}_{12} may be obtained from a standard normal distribution. Typically, \hat{Z}_{12} is considered a two-tailed test, however directional (one-tailed) tests may be specified should the empirical situation require it (39, 41).

Simulations. Simulations were conducted by generating pure-birth phylogenies at each of six different tree sizes ($n = 2^5, 2^6, \dots, 2^{10}$), and with differing levels of phylogenetic signal ($\lambda = 0.0, 0.5, \dots, 1.0$). We generated 100 random trees for each intersection of tree size and λ . For each λ within each tree size, continuous traits were then simulated on each phylogeny under a BM model of evolution. For each set of 100 trees we measured the mean values of $\hat{\lambda}$ and κ , their standard deviation, and calculated the Shapiro-Wilk W statistic as a departure from normality (symmetry). For the latter, a value of 1.0 indicates normally distributed values, while departures from 1.0 indicate skewness. Simulations were then repeated for both balanced and pectinate trees, which yielded qualitatively similar results (see Supporting Information). Trees containing polytomies, and an evaluation of $\hat{\lambda}$ from models of linear regression and phylogenetic ANOVA, were also investigated, and results were qualitatively similar to those reported above (see Supporting Information).

Empirical Data. Surface area to volume ratios (SA:V) and relative body width ($\frac{BW}{SVL}$) measures were obtained from individuals of 305 species (2,781 and 3,371 respectively), from which species means were obtained (33, 34). A time-dated molecular phylogeny for the group (43) was then pruned to match the species in the phenotypic dataset. The phylogenetic signal in each trait was then characterized using κ , which was converted to its effect size (Z_κ) using **geomorph** 3.3.1 (44, 45).

Acknowledgements

This work was supported in part by NSF grant DBI-1902511 (to D.C.A.) and DBI-1902694 (to M.L.C.).

References

1. J. Felsenstein, Phylogenies and the comparative method. *American Naturalist* **125**, 1–15 (1985).
2. P. H. Harvey, M. D. Pagel, *The comparative method in evolutionary biology* (Oxford University Press, 1991).
3. A. Grafen, The phylogenetic regression. *Philosophical Transactions of the Royal Society of London B, Biological Sciences* **326**, 119–157 (1989).
4. T. J. Garland, A. R. Ives, Using the past to predict the present: Confidence intervals for regression equations in phylogenetic comparative methods. *American Naturalist* **155**, 346–364 (2000).
5. F. J. Rohlf, Comparative methods for the analysis of continuous variables: Geometric interpretations. *Evolution* **55**, 2143–2160 (2001).
6. M. A. Butler, A. A. King, Phylogenetic comparative analysis: A modeling approach for adaptive evolution. *American Naturalist* **164**, 683–695 (2004).
7. E. P. Martins, T. F. Hansen, Phylogenies and the comparative method: A general approach to incorporating phylogenetic information into the analysis of interspecific data. *American Naturalist* **149**, 646–667 (1997).
8. B. C. O’Meara, C. Ane, M. J. Sanderson, P. C. Wainwright, Testing for different rates of continuous trait evolution using likelihood. *Evolution* **60**, 922–933 (2006).
9. L. J. Revell, L. J. Harmon, Testing quantitative genetic hypotheses about the evolutionary rate matrix for continuous characters. *Evolutionary Ecology Research* **10**, 311–331 (2008).
10. J. M. Beaulieu, D. C. Jhvueng, C. Boettiger, B. C. O’Meara, Modeling stabilizing selection: Expanding the ornstein-uhlenbeck model of adaptive evolution. *Evolution* **66**, 2369–2383 (2012).
11. D. C. Adams, A method for assessing phylogenetic least squares models for shape and other high-dimensional multivariate data. *Evolution* **68**, 2675–2688 (2014).
12. D. C. Adams, M. L. Collyer, Phylogenetic anova: Group-clade aggregation, biological challenges, and a refined permutation procedure. *Evolution* **72**, 1204–1215 (2018).
13. M. D. Pagel, Inferring the historical patterns of biological evolution. *Nature* **401**, 877–884 (1999).
14. S. P. Blomberg, T. Garland, A. R. Ives, Testing for phylogenetic signal in comparative data: Behavioral

traits are more labile. *Evolution* **57**, 717–745 (2003).

15. L. J. Revell, L. J. Harmon, D. C. Collar, Phylogenetic signal, evolutionary process, and rate. *Systematic Biology* **57**, 591–601 (2008).

16. E. Abouheif, A method for testing the assumption of phylogenetic independence in comparative data. *Evolutionary Ecology Research* **1**, 895–909 (1999).

17. J. L. Gittleman, M. Kot, Adaptation: Statistics and a null model for estimating phylogenetic effects. *Systematic Zoology* **39**, 227–241 (1990).

18. D. C. Adams, A generalized Kappa statistic for estimating phylogenetic signal from shape and other high-dimensional multivariate data. *Systematic Biology* **63**, 685–697 (2014).

19. C. P. Klingenberg, N. A. Gidaszewski, Testing and quantifying phylogenetic signals and homoplasy in morphometric data. *Systematic biology* **59**, 245–261 (2010).

20. T. Münkemüller *et al.*, How to measure and test phylogenetic signal. *Methods in Ecology and Evolution* **3**, 743–756 (2012).

21. S. Pavoine, C. Ricotta, Testing for phylogenetic signal in biological traits: The ubiquity of cross-product statistics. *Evolution: International Journal of Organic Evolution* **67**, 828–840 (2012).

22. J. A. F. Diniz-Filho, T. Santos, T. F. Rangel, L. M. Bini, A comparison of metrics for estimating phylogenetic signal under alternative evolutionary models. *Genetics and Molecular Biology* **35**, 673–679 (2012).

23. R. Molina-Venegas, M. A. Rodríguez, Revisiting phylogenetic signal; strong or negligible impacts of polytomies and branch length information? *BMC evolutionary biology* **17**, 53 (2017).

24. L. J. Revell, Phylogenetic signal and linear regression on species data. *Methods in Ecology and Evolution* **1**, 319–329 (2010).

25. C. Boettiger, G. Coop, P. Ralph, Is your phylogeny informative? Measuring the power of comparative methods. *Evolution* **67**, 2240–2251 (2012).

26. R. P. Freckleton, P. H. Harvey, M. Pagel, Phylogenetic analysis and comparative data: A test and review of evidence. *American Naturalist* **160**, 712–726 (2002).

27. N. Cooper, W. Jetz, R. P. Freckleton, Phylogenetic comparative approaches for studying niche conservatism.

Journal of Evolutionary Biology **23**, 2529–2539 (2010).

28. R. Bose, B. R. Ramesh, R. Péliissier, F. Munoz, Phylogenetic diversity in the western ghats biodiversity hotspot reflects environmental filtering and past niche diversification of trees. *Journal of Biogeography* **46**, 145–157 (2019).
29. F. Vandeloocket *al.*, Nectar traits differ between pollination syndromes in balsaminaceae. *Annals of Botany* **124**, 269–279 (2019).
30. G. De Meester, K. Huyghe, R. Van Damme, Brain size, ecology and sociality: A reptilian perspective. *Biological Journal of the Linnean Society* **126**, 381–391 (2019).
31. P. Pintanel, M. Tejedo, S. R. Ron, G. A. Llorente, A. Merino-Viteri, Elevational and microclimatic drivers of thermal tolerance in andean pristimantis frogs. *Journal of Biogeography* **46**, 1664–1675 (2019).
32. G. Su, S. Villéger, S. Brosse, Morphological diversity of freshwater fishes differs between realms, but morphologically extreme species are widespread. *Global ecology and biogeography* **28**, 211–221 (2019).
33. E. K. Baken, D. C. Adams, Macroevolution of arboreality in salamanders. *Ecology and Evolution* **9**, 7005–7016 (2019).
34. E. K. Baken, L. E. Mellenthin, D. C. Adams, Macroevolution of desiccation-related morphology in plethodontid salamanders as inferred from a novel surface area to volume ratio estimation approach. *Evolution* **74**, 476–486 (2020).
35. G. V. Glass, Primary, secondary, and meta-analysis of research. *Educational Researcher* **5**, 3–8 (1976).
36. J. Cohen, *Statistical power analysis for the behavioral sciences* (Routledge, 1988).
37. R. Rosenthal, “The handbook of research synthesis” in L. V. Cooper H Hedges, Ed. (Russell Sage Foundation, 1994), pp. 231–244.
38. M. L. Collyer, D. J. Sekora, D. C. Adams, A method for analysis of phenotypic change for phenotypes described by high-dimensional data. *Heredity* **115**, 357–365 (2015).
39. D. C. Adams, M. L. Collyer, On the comparison of the strength of morphological integration across morphometric datasets. *Evolution* **70**, 2623–2631 (2016).
40. M. L. Collyer, D. C. Adams, RRPP: An r package for fitting linear models to high-dimensional data using residual randomization. *Methods in Ecology and Evolution* **9**, 1772–1779 (2018).

- 285 41. D. C. Adams, M. L. Collyer, Comparing the strength of modular signal, and evaluating alternative
286 modular hypotheses, using covariance ratio effect sizes with morphometric data. *Evolution* **73**, 2352–2367
287 (2019).
- 288 42. D. C. Adams, M. L. Collyer, Phylogenetic comparative methods and the evolution of multivariate
289 phenotypes. *Annual Review of Ecology, Evolution, and Systematics* **50**, 405–425 (2019).
- 290 43. R. M. Bonett, A. L. Blair, Evidence for complex life cycle constraints on salamander body form
291 diversification. *Proceedings of the National Academy of Sciences, U.S.A.* **114**, 9936–9941 (2017).
- 292 44. D. C. Adams, E. Otárola-Castillo, Geomorph: An r package for the collection and analysis of geometric
293 morphometric shape data. *Methods in Ecology and Evolution* **4**, 393–399 (2013).
- 294 45. D. C. Adams, M. L. Collyer, A. Kaliontzopoulou, Geomorph: Software for geometric morphometric
295 analyses. R package version 3.3.1. (2020).

Real figures will have 6 panels for lambda and kappa and probably only a subset for z. Simulations need to be re-run.

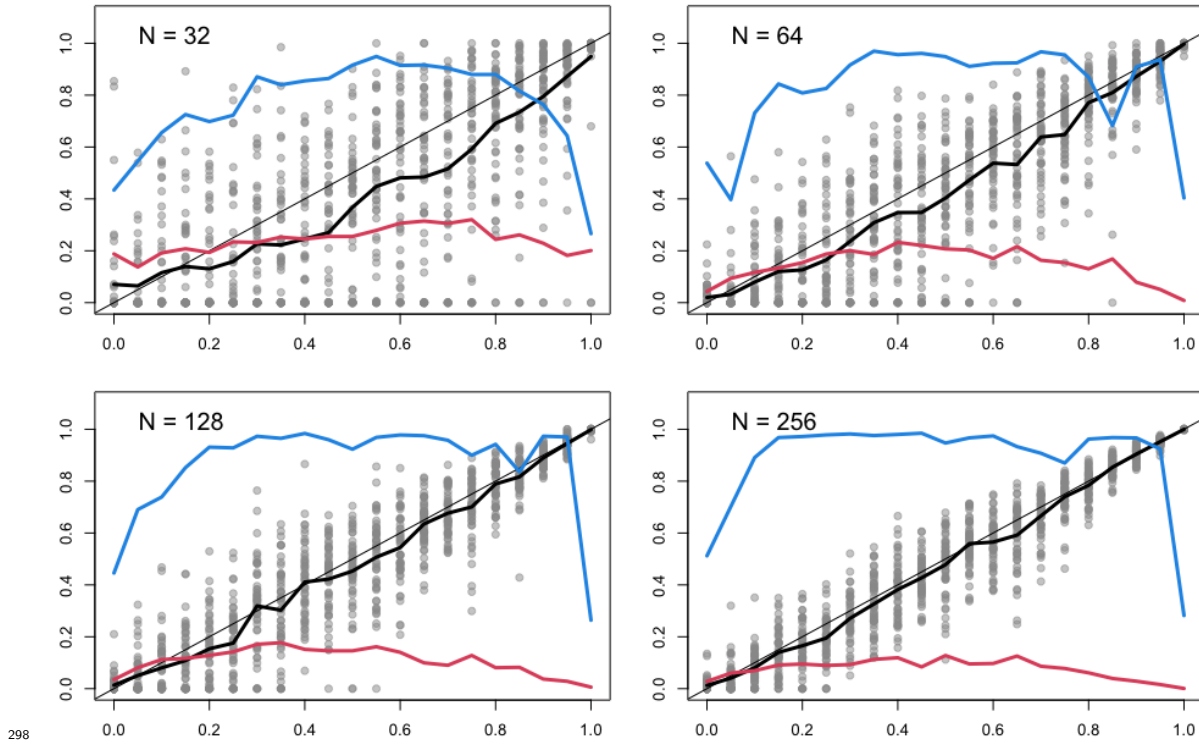
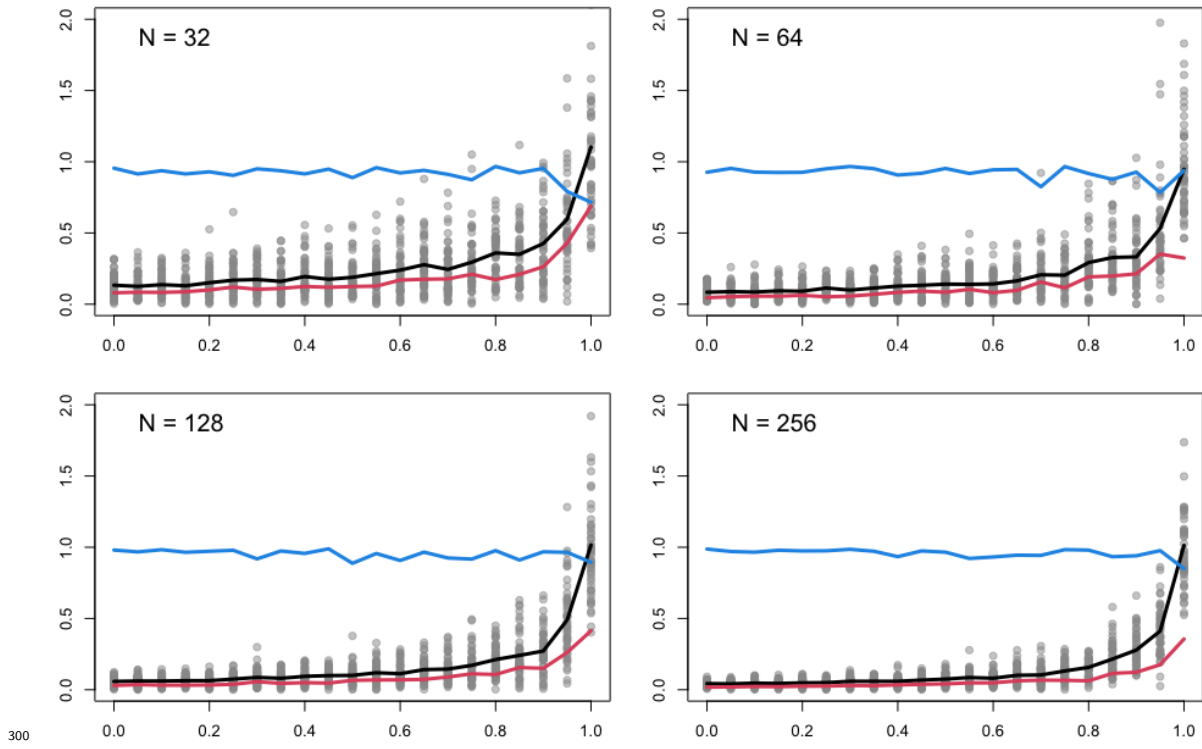
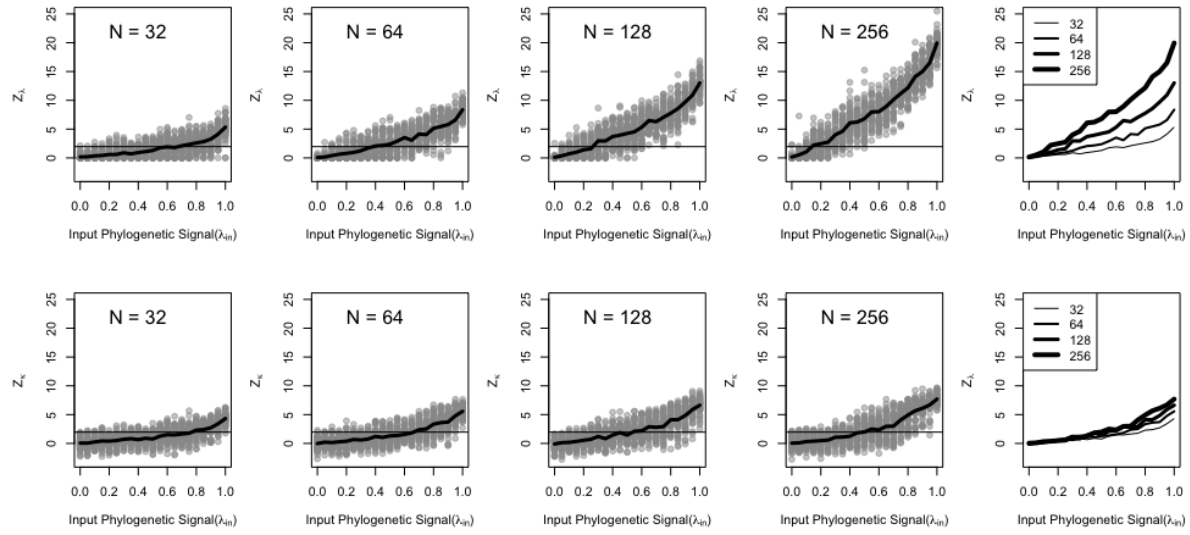


Figure 1. Temporary Fig. 1 showing lambda patterns



301 **Figure 2.** Temporary Fig. 2 showing kappa patterns



302

303 **Figure 3.** Temporary Fig. 3 showing z patterns

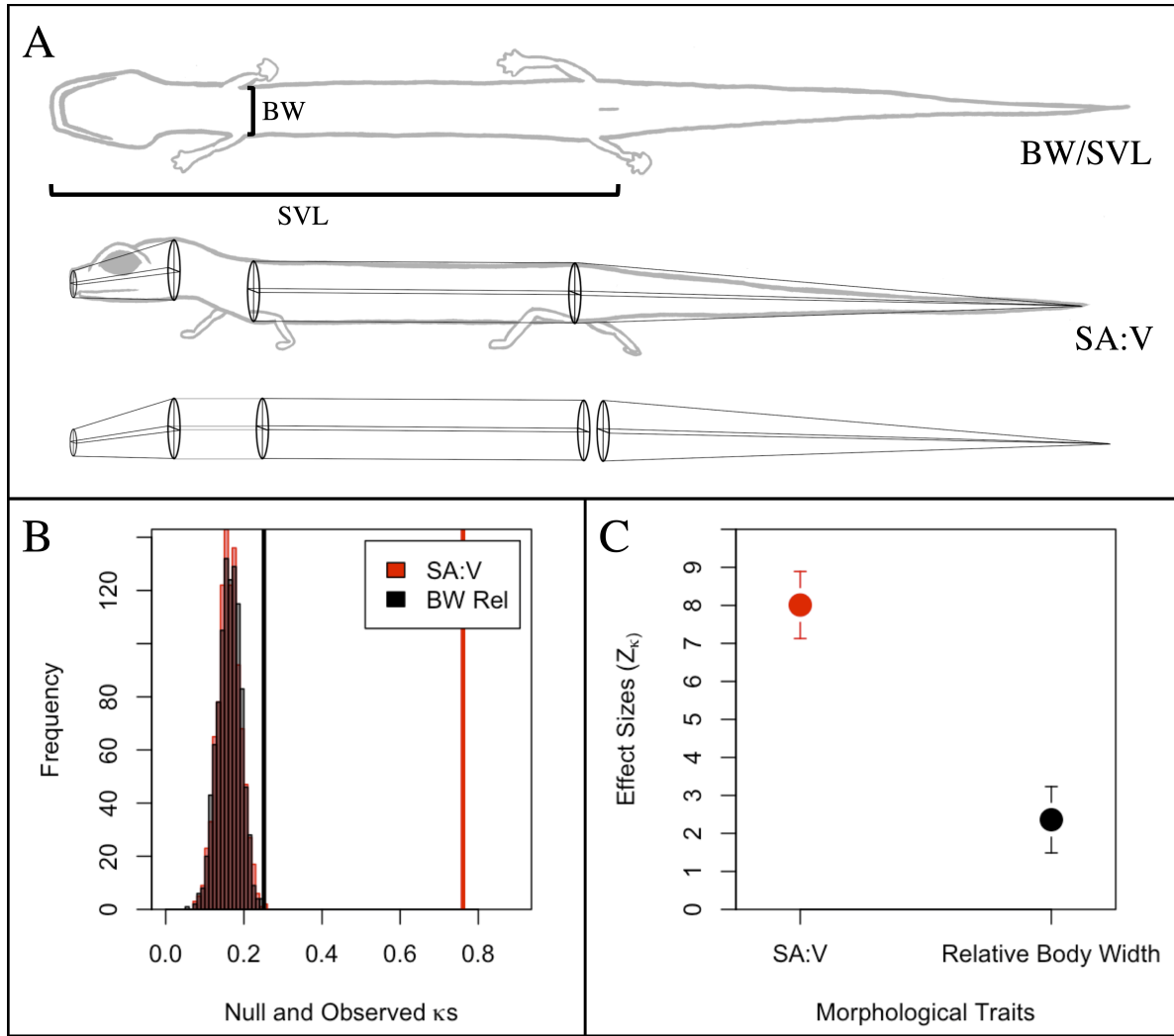


Figure 4. (A) Linear measures for relative body size, and regions of the body used to estimate surface area to volume (SA:V) ratios. (B) Permutation distributions of phylogenetic signal for SA:V and $\frac{BW}{SVL}$, with observed values shown as vertical bars. (C) Effect sizes (Z_{κ}) for SA:V and $\frac{BW}{SVL}$, with their 95% confidence intervals (CI not standardized by $\sqrt{(n)}$).