A Standardized Effect Size for Evaluating the Strength of Phylogenetic Signal, and Why Lambda is not Appropriate

Keywords: phylogenetic signal, effect size, Pagel's lambda

Short Title: An Effect Size for Phylogenetic Signal

Abstract

{conclusion holds: interpreting the regression is not appreciably different (in terms of slopes and f values)}

Introduction

Investigating macroevolutionary patterns of trait variation requires a phylogenetic perspective, because the shared ancestry among species generates statistical non-independence (Felsenstein 1985; Harvey and Pagel 1991). Accounting for this evolutionary non-independence is the purview of *phylogenetic comparative methods* (PCMs); a suite of analytical tools that condition the data on the phylogeny through the course of statisical evaluations of phenotypic trends (e.g., Grafen 1989; Garland and Ives 2000; Rohlf 2001; Butler and King 2004). The past several decades have witnessed a rapid expansion in the development of PCMs to address an ever-growing set of macroevolutionary hypotheses (Martins and Hansen 1997; O'Meara et al. 2006; Revell and Harmon 2008; Beaulieu et al. 2012; Adams 2014b,a; Adams and Collyer 2018). These methods are predicated on the notion that phylogenetic signal – the tendancy for closely related species to display similar trait values – is present in cross-species datasets (Felsenstein 1985; Pagel 1999; Blomberg et al. 2003). 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 covaration among related taxa (see Felsenstein 1985; Blomberg et al. 2003; Revell et al. 2008).

Several analytical tools have been developed to quantify phylogenetic signal in phenotypic datasets, including measures of serial independence (**C**: Abouheif 1999), autocorrelation estimates (I: Gittleman and Kot 1990), statistical ratios of trait variation relative to what is expected given the phylogeny ($Kappa$: Blomberg et al. 2003; Adams 2014a), and scaling parameters used in maximum likelihood fitting of the data to the phylogeny (λ: Pagel 1999), among others (e.g., Klingenberg and Gidaszewski 2010). The statistical properties of these methods – namely type I error rates and power – have also been investigated to determine when phylogenetic signal can be detected and under what conditions (e.g., Munkemuller et al. 2012; Pavoine and Ricotta 2012; Diniz-Filho et al. 2012; Adams 2014a; Molina-Venegas and Rodriguez 2017; see also Revell et al. 2008; Revell 2010). One of the most widely used methods for characterizing phylogenetic signal in macroevolutionary studies is Pagel's λ (Pagel 1999). Here, maximum likelihood is used to fit the data to the phylogeny under a Brownian motion model of evolution. A parameter (λ) is included, which transforms the lengths of the internal branches of the phylogeny to improve the fit (Pagel 1999; Freckleton et al. 2002). Pagel's λ ranges from $0 \rightarrow 1$, with larger values signifying a greater dependence of observed trait variation on the phylogeny. Pagel's λ also has the appeal that it may be included in phylogenetic regression (PGLS) to account for the degree of phylogenetic signal in comparative analyses (see Freckleton et al. 2002).

Evolutionary biologists commonly seek to describe the relative strength of phylogenetic signal in phenotypic traits, to determine the extent to which shared evolutionary history has influenced trait covariation among taxa. This is often accomplished by interpreting empirical estimates of λ; with smaller values signifying 'weak' phylogenetic signal, while larger values are interpreted as 'strong' phylogenetic signal (e.g., De Meester et al. 2019; Pintanel et al. 2019; Su et al. 2019). Other approaches for interpreting λ are more statistical. For instance, some have evaluated whether the observed λ differs from some expected value through the use of confidence intervals (Vandelook et al. 2019) or by performing likelihood ratio tests that compare the observed model fit to that obtained when $\lambda = 0$ or $\lambda = 1$ (Freckleton et al. 2002; Cooper et al. 2010; Bose et al. 2019). Additionally, qualitative comparisons of λ estimates obtained from multiple phenotypic traits have been used to infer whether the strength of phylogenetic signal is greater in one trait as compared to another (e.g., Liu et al. 2019; Bai et al. 2019). Indeed, statements regarding the strength of phylogenetic signal based on λ are rather common in the evolutionary literature. We conducted a literature survey in Google.scholar and found that of the 204 papers published in 2019 that estimated and reported Pagel's λ, 40% interpreted the strength of phylogenetic signal for at least one phenotypic trait. Additionally, nearly 30% of the 421 λ estimates between 0.25 and 0.75 were assigned a strength of signal, and because nearly half of the 1,572 values reported were near the limits of the parameter (Figure 1), this percentage is even higher, as biological interpretation of phylogenetic signal at the limits of λ are known.

[insert Figure 1 here]

It seems intuitive to interpret the strength of phylogenetic signal based on the value of λ, as λ is a parameter on a bounded scale ($0 \rightarrow 1$) for which interpretation of its extremal points are understood. Specifically, $\lambda = 0$ represents no phylogenetic signal, while $\lambda = 1$ is phylogenetic signal as expected under Brownian motion. However, equating values of λ directly to the strength of phylogenetic signal presumes two important statistical properties that have not been fully explored. First, it presumes that values of λ can be precisely estimated, as biological inferences regarding the strength of phylogenetic signal depend on high accuracy in its estimation. Therefore, understanding the precision in estimating λ is paramount. One study (Boettiger et al. 2012) found that estimates of Pagel's λ displayed less variation (i.e., greater precision) when data were simulated on a large phylogeny ($N = 281$) as compared to a small one ($N = 13$). From this observation it was concluded that insufficient data (i.e., the number of species) was the underlying cause of the increased variation across parameter estimates (Boettiger et al. 2012). Indeed, such a pattern is common with statistical estimators, as summary statistics and parameters are often more precise at greater sample sizes

(Cohen 1988). However, this conclusion also assumes that the precision of λ remains constant across its range ($\lambda = 0 \rightarrow 1$); an assumption that to date, has not been verified. Thus, despite widespread use of Pagel's (1999) λ in macroevolutionary studies, at present, we still lack a general understanding of the precision with which λ can estimate levels of phylogenetic signal in phenotypic datasets.

Second, while estimates of λ are within a bounded scale ($0 \rightarrow 1$), this does not *de-facto* imply that the estimated values of this parameter correspond to the actual strength of the underlying input signal in the data. For this to be the case, λ must be a statistical effect size. Effect sizes are a measure the magnitude of a statistical effect in data, represented on a common scale (Glass 1976; Cohen 1988). Effect sizes have widespread use in many areas of the quantiative sciences, as they represent measures that may be readily summarized across datasets as in meta-analysis (Glass 1976; Hedges and Olkin 1985; Arnqvist and Wooster 1995), or compared among datasets (e.g., Adams and Collyer 2016, 2019a). Unfortunatley, not all model parameters and test statistics are effect sizes, and thus many summary measures must first be converted to standardized units (i.e., an effect size) for meaningful comparison (see Rosenthal 1994). As a consequence, it follows that only if λ is a statistical effect size can comparisons of estimates across datasets be interpretable. For the case of λ, this has not yet been explored.

In this study, we evaluate the precision of Pagel's λ in estimating known levels of phylogenetic signal in phenotypic data. We use computer simulations with differing numbers of species, differently shaped phylogenies, and differing input levels of phylogenetic signal, to explore the degree to which λ correctly identifies known levels of phylogenetic signal, and under what circumstances. We find that while PGLS parameters (e.g., β) are accurately estimated with the inclusion of phylogenetic signal, estimates of λ are not. We also find that estimates of λ vary widely for a given input value of phylogenetic signal, and that the precision in estimating λ is not constant across the range of input signal, with decreased precision when phylogenetic signal is of intermediate strength. Additionally, the same λ_{est} may be obtained from datasets containing vastly different input levels of phylogenetic signal. Thus, λ is not a reliable estimate of the strength of phylogenetic signal in phenotypic data. We subsequently derive a standardized effect size for measuring the strength of phylogenetic signal in phenotypic datasets, and apply the concept to two common measures of phylogenetic signal: λ and $Kappa$. Through simulations across a wide range of conditions, we find that the precision of effect sizes based on λ (Z_λ) are less reliable than that those based on $Kappa$ (Z_K), implying that Z_K is a more robust effect size measure. Additionally, we propose a two-sample test statistic that may be used to compare the strength of phylogenetic signal among datasets, and provide an empirical example to demonstrate its use. We conclude that estimates of phylogenetic signal using Pagel's λ are often inaccurate, and thus interpreting strength of phylogenetic signal in phenotypic datasets based on this measure is compromised. By contrast, effect sizes obtained from $Kappa$ hold promise for characterizing phylogenetic signal, and for comparing the strength of phylogenetic signal across datasets.

Methods and Results

The Precision of λ is Variable

We conducted a series of computer simulations to evaluate the precision of Pagel's λ. Our primary simulations were based on pure-birth phylogenies; however, we also evaluated patterns on both balanced and pectinate trees to determine whether tree shape affected our findings (see Supporting Information). First we generated 50 pure-birth phylogenies at each of six different tree sizes, ranging from 32 to 1024 taxa ($n = 2^5 - 2^{10}$). Next, we rescaled the simulated phylogenies by multiplying the internal branches by λ_{in}, using 21 intervals of 0.05 units across its range ($\lambda_{in} = 0.0 \rightarrow 1.0$), resulting in 1050 scaled phylogenies at each level of species richness (n). Continuous traits were then simulated on each phylogeny under a Brownian motion model of evolution to obtain datasets with differing levels of phylogenetic signal, that ranged from no phylogenetic signal (when $\lambda_{in} = 0$), to phylogenetic signal corresponding reflecting Brownian motion (when $\lambda_{in} = 1$). For each dataset we then estimated phylogenetic signal (λ_{est}), and calculated the precision of λ using the variance (σ^2_λ) across datasets at each input level of phylogenetic signal and level of species richness.

We also evaluated the precision of λ when estimated in PGLS regression and ANOVA (i.e., $Y \sim X$). Here, an independent variable X was simulated on each phylogeny under a Brownian motion model of evolution (for PGLS regression). For phylogenetic ANOVA, random groups (X) were obtained by simulating a discrete (binary) character on each phylogeny. Next, the dependent variable was simulated in such a manner as to contain a known relationship with X plus random error containing phylogenetic signal. This was accomplished as: $Y = \beta X + \epsilon$. Here, the association between Y and X was modeled using a range of values: $\beta = (0.0, 0.25, 0.5, 0.75, 1.0)$, and the residual error was modeled to contain phylogenetic signal simulated under a Brownian motion model of evolution: $\epsilon = \mathcal{N}(\mu = 0, \sigma = \mathbf{C})$: (see Revell 2010 for a similar simulation design). The fit of the phylogenetic regression was estimated using maximum likelihood, and parameter estimates (β_{est} and λ_{est}) were obtained. Precision estimates (σ^2_λ) at each input level of phylogenetic signal and level of species richness were then observed.

All analyses were performed in R v3.6.0 (R Core Team 2019) using the packages `geiger` (Harmon et al. 2008), `caper` (Orme et al. 2013), `phytools` (Revell 2012), and `geomorph` (Adams and Otárola-Castillo 2013; Adams et al. 2020). R-scripts are found in the Supporting Information.

Results. We found that the precision of λ_{est} varied widely across simulation conditions. Predictably, precision improved as the number of species increased (Figure 2). This confirmed earlier findings of Boettiger et al. (2012), and adhered to parametric statistical theory. However, in many cases the set of λ_{est} spanned nearly the entire range of possible values (e.g., $n = 32$; $\lambda_{in} = 0.5$: $\lambda_{est} = 0.0 \rightarrow 0.985$), revealing that estimates of λ were not a reliable indicator of input phylogenetic signal. Importantly, the precision of λ_{est} was not uniform across all levels of phylogenetic signal, with the worst precision at intermediate levels of signal ($\lambda_{in} \approx 0.5$), and improved precision as input levels approached the extremes of its range (i.e., $\lambda_{in} \rightarrow 0$ & $\lambda_{in} \rightarrow 1$). Thus, estimates of λ were least reflective of the true input signal at intermediate values. Additionally, even at large levels of species richness, we found that the range of λ_{est} still encompassed a substantial portion of possible values (e.g., $n = 512$; $\lambda_{in} = 0.5$: $\lambda_{est} = 0.32 \rightarrow 0.68$). Likewise, the same λ_{est} could be obtained from datasets containing vastly different input levels of phylogenetic signal (e.g., $n = 512$; $\lambda_{est} = 0.5$; $\lambda_{in} = 0.25 \rightarrow 0.65$). Results were similar when λ was co-estimated with regression parameters in PGLS regression (Figure 3). Here, regression parameters (β) were accurately estimated, confirming earlier findings of Revell 2010 (2010) (see Supporting Information). However, estimates of phylogenetic signal were not, and the spread of λ_{est} was even broader than that observed when λ was estimated for only the dependent variable. Taken together, these findings reveal that λ_{est} does not precisely characterize observed levels of phylogenetic signal in phenotypic datasets, and that biological interpretations of the strength of phylogenetic signal based on λ may be highly inaccurate.

[insert Figure 2 here]

[insert Figure 3 here]

A Standardized Effect Size for Phylogenetic Signal

The results above demonstate that λ is not a reliable estimate of the phylogenetic signal in phenotypic data. As such, biological interpretations of the strength of phylogenetic signal, and comparisons of the magnitude of such effects across datasets, are severely compromised when based on this parameter. As an alternative, we propose that summary estimates of phylogenetic signal be converted to effect sizes for interpretation and comparison. Statistically, a standardized effect size may be found as:

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

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 (Glass 1976; Cohen 1988; Rosenthal 1994). Z_θ expresses the magnitude of the effect in θ_{obs} by transforming the original test statistic to a standard normal deviate (Glass 1976; Kelley and Preacher 2012). 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 (Collyer et al. 2015; Adams and Collyer 2016, 2019a) have shown that $E(\theta)$ and σ_θ may also be obtained from an empirical sampling distribution of θ obtained from permutation procedures.

Adams and Collyer (2019b) recently suggested that the strength of phylogenetic signal could be represented as an effect size, based on the *Kappa* statistic and its empirical sampling distribution from permutation. Here we formalize that suggestion, and find an effect size as:

$$Z_K = \frac{K_{obs} - \hat{\mu}_K}{\hat{\sigma}_K} \tag{2}$$

where K_{obs} is the observed phylogenetic signal, and $\hat{\mu}_K$ and $\hat{\sigma}_K$ are the mean and standard deviation of the empirical sampling distribution of *Kappa* obtained via permutation. Similarly, an effect size based on λ could be envisioned as:

$$Z_\lambda = \frac{\lambda_{obs} - 0}{\hat{\sigma}_\lambda}. \tag{3}$$

In this case, λ_{obs} and $\hat{\sigma}_\lambda$ are empirically derived using maximum likelihood. Note also that under the null hypothesis, $E(\lambda) = 0$, a no phylogenetic signal is expected under this condition (Freckleton et al. 2002).

To evaluate the utility of Z_K and Z_λ we calculated both effect sizes for the simulated datasets generated above, and summarized the precision of each using its variance ($\sigma^2_{Z_K}$ and $\sigma^2_{Z_\lambda}$). Results are found in Figure 4. Here two things are evident. First, estimates of Z_K track the input phylogenetic signal in a more linear fashion than do estimates of Z_λ. Second, the precision of Z_K is considerably more stable as compared with Z_λ, as coefficients of variation for the set of $\sigma^2_{Z_K}$ across input levels of phylogenetic signal were an order of magnitude smaller for than was observed for $\sigma^2_{Z_\lambda}$ (Figure 4). This implied that estimates of the strength of phylogenetic signal were more reliable and robust when using Z_K as compared with Z_λ.

[insert Figure 4 here]

Statistical Comparisons of Phylogenetic Signal

Once the magnitude of phylogenetic signal is characterized using Z_K, it may be of interest to compare such measures across datasets. This is useful, for instance, to determine whether the strength of phylogenetic signal is greater in one phenotypic trait as compared with another. As with other effect sizes derived from permutation distributions (e.g., Adams and Collyer 2016, 2019a), a two-sample test statistic may be found as:

$$\hat{Z}_{12} = \frac{|(K_1 - \hat{\mu}_{K_1}) - (K_2 - \hat{\mu}_{K_2})|}{\sqrt{\hat{\sigma}^2_{K_1} + \hat{\sigma}^2_{K_2}}} \tag{4}$$

where K_1, K_2, $\hat{\mu}_{K_1}$, $\hat{\mu}_{K_2}$, $\hat{\sigma}_{K_1}$, and $\hat{\sigma}_{K_2}$ are as defined above for equation 2. Estimates of significance of $\hat{Z}_{12}$ may be obtained from a standard normal distribution. As with other two-sample tests, $\hat{Z}_{12}$ is typically considered a two-tailed test, however directional (one-tailed) tests may be specified should the empirical situation require it (see Adams and Collyer 2016, 2019a).

Empirical Example

Conclusions and Implications

1: summary paragraph

2: expand on Lambda.. lambda innacurate, not precise, level of precision varies with input physig (worse in mid-range). NEW RESULT. We are first to show this. NOTE: pattern is obvious with reflection. Since it is a 'bounded' parameter estimation should be best at the extremes… (state this?).. hmm.

Patterns worse with PGLS, though beta still estimated properly. Conclusion, lambda not overly useful.

3: By contrast, effect size Z-K useful, equally precise across range of values. Can be used to characterize the strength of physignal, and because robust to input levels, etc. may be used to compare across datasets.

Somewhere, recognize that this is somewhat 'backwards' from prior recommendations where Kappa had somewhat lower performance in terms of type I and type II error (which?? I forget). However, recall that those studies did not examine the precision of the estimates. Nor was Z-k included, because it was not yet invented. So Use of Z-k should make good sense here.

Closing paragraph.

More discussion paragraphs

References

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

Adams, D. C. 2014a. A generalized Kappa statistic for estimating phylogenetic signal from shape and other high-dimensional dultivariate data. Systematic Biology 63:685–697.

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

Adams, D. C., and M. L. Collyer. 2019a. Comparing the strength of modular signal, and evaluating alternative modular hypotheses, using covariance ratio effect sizes with morphometric data. Evolution 73:2352–2367.

Adams, D. C., and M. L. Collyer. 2016. On the comparison of the strength of morphological integration across morphometric datasets. Evolution 70:2623–2631.

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

Adams, D. C., and M. L. Collyer. 2019b. Phylogenetic comparative methods and the evolution of multivariate phenotypes. Annual Review of Ecology, Evolution, and Systematics 50:405–425.

Adams, D. C., M. L. Collyer, and A. Kaliontzopoulou. 2020. Geomorph: Software for geometric morphometric analyses. R package version 3.2.1.

Adams, D. C., and E. Otárola-Castillo. 2013. Geomorph: An r package for the collection and analysis of geometric morphometric shape data. Methods in Ecology and Evolution 4:393–399.

Arnqvist, G., and D. Wooster. 1995. Meta-analysis: Synthesizing research findings in ecology and evolution. Trends in Ecology and Evolution 10:236–240.

Bai, K., S. Lv, S. Ning, D. Zeng, Y. Guo, and B. Wang. 2019. Leaf nutrient concentrations associated with phylogeny, leaf habit and soil chemistry in tropical karst seasonal rainforest tree species. Plant and Soil 434:305–326.

Beaulieu, J. M., D. C. Jhwueng, C. Boettiger, and B. C. O'Meara. 2012. Modeling stabilizing selection: Expanding the ornstein-uhlenbeck model of adaptive evolution. Evolution 66:2369–2383.

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

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

Bose, R., B. R. Ramesh, R. Pélissier, and F. Munoz. 2019. Phylogenetic diversity in the western ghats biodiversity hotspot reflects environmental filtering and past niche diversification of trees. Journal of Biogeography 46:145–157.

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

Cohen, J. 1988. Statistical power analysis for the behavioral sciences. Routledge.

Collyer, M. L., D. J. Sekora, and D. C. Adams. 2015. A method for analysis of phenotypic change for phenotypes described by high-dimensional data. Heredity 115:357–365.

Cooper, N., W. Jetz, and R. P. Freckleton. 2010. Phylogenetic comparative approaches for studying niche conservatism. Journal of Evolutionary Biology 23:2529–2539.

De Meester, G., K. Huyghe, and R. Van Damme. 2019. Brain size, ecology and sociality: A reptilian perspective. Biological Journal of the Linnean Society 126:381–391.

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

Felsenstein, J. 1985. Phylogenies and the comparative method. American Naturalist 125:1–15.

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

Garland, T. J., and A. R. Ives. 2000. Using the past to predict the present: Confidence intervals for regression equations in phylogenetic comparative methods. American Naturalist 155:346–364.

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

Glass, G. V. 1976. Primary, secondary, and meta-analysis of research. Educational Researcher 5:3–8.

Grafen, A. 1989. The phylogenetic regression. Philosophical Transactions of the Royal Society of London B, Biological Sciences 326:119–157.

Harmon, L. J., J. T. Weir, C. D. Brock, R. E. Glor, and W. Challenger. 2008. GEIGER: Investigating evolutionary radiations. Bioinformatics 24:129–131.

Harvey, P. H., and M. D. Pagel. 1991. The comparative method in evolutionary biology. Oxford University Press, Oxford.

Hedges, L. V., and I. Olkin. 1985. Statistical methods for meta-analysis. Elsevier.

Kelley, K., and K. J. Preacher. 2012. On effect size. Psychological Methods 17:137–152.

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

Liu, H., C. P. Osborne, D. Yin, R. P. Freckleton, G. Jiang, and M. Liu. 2019. Phylogeny and ecological processes influence grass coexistence at different spatial scales within the steppe biome. Oecologia 191:25–38.

Martins, E. P., and T. F. Hansen. 1997. Phylogenies and the comparative method: A general approach to incorporating phylogenetic information into the analysis of interspecific data. American Naturalist 149:646–667.

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

Munkemuller, T., S. Lavergne, B. Bzeznik, S. Dray, T. Jombart, K. Schiffers, and W. Thuiller. 2012. How to measure and test phylogenetic signal. Methods in Ecology and Evolution 3:743–756.

O'Meara, B. C., C. Ane, M. J. Sanderson, and P. C. Wainwright. 2006. Testing for different rates of continuous trait evolution using likelihood. Evolution 60:922–933.

Orme, D., R. P. Freckleton, G. H. Thomas, T. Petzoldt, S. A. Fritz, and N. Isaac. 2013. CAPER: Comparative analyses of phylogenetics and evolution in r. Methods in Ecology and Evolution 3:145–151.

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

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

Pintanel, P., M. Tejedo, S. R. Ron, G. A. Llorente, and A. Merino-Viteri. 2019. Elevational and microclimatic drivers of thermal tolerance in andean pristimantis frogs. Journal of Biogeography 46:1664–1675.

R Core Team. 2019. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

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

Revell, L. J. 2012. Phytools: An r package for phylogenetic comparative biology (and other things). Methods in Ecology and Evolution 3:217–223.

Revell, L. J., and L. J. Harmon. 2008. Testing quantitative genetic hypotheses about the evolutionary rate matrix for continuous characters. Evolutionary Ecology Research 10:311–331.

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

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

Rosenthal, R. 1994. The handbook of research synthesis. Pp. 231–244 *in* L. V. Cooper H Hedges, ed. Russell Sage Foundation.

Su, G., S. Villéger, and S. Brosse. 2019. Morphological diversity of freshwater fishes differs between realms, but morphologically extreme species are widespread. Global ecology and biogeography 28:211–221.

Vandelook, F., S. Janssens, P. Gijbels, E. Fischer, W. Van den Ende, O. Honnay, and S. Abrahamczyk. 2019. Nectar traits differ between pollination syndromes in balsaminaceae. Annals of Botany 124:269–279.

Figure Legends

Figure 1. Frequency distribution of λ estimates published in 2019. The majority of these values were close to 0 or 1, and from phylogenies with fewer than 200 taxa.

Figure 2. Precision of Pagel's λ across known levels of input phylogenetic signal (λ_{in}) on phylogenies of various sizes. As phylogenies increase in size, variation in λ_{in} decreases; however the precision is not constant across the range of input levels ($\lambda_{in} : 0 \rightarrow 1$), and is highest at intermediate levels of phylogenetic signal.

Figure 3. Precision of Pagel's λ when incorporated in phylogenetic regression ($Y \sim X$), across known levels of input phylogenetic signal (λ_{in}) on phylogenies of various sizes. As phylogenies increase in size, variation in λ_{in} decreases; however the precision is not constant across the range of input levels ($\lambda_{in} : 0 \rightarrow 1$), and is highest at intermediate levels of phylogenetic signal.

Figure 4. Variation in estimates of phylogenetic signal across input levels of phylogenetic signal. (A) Estimates of Pagel's λ for data simulated on phylogenies with 128 taxa ($n = 128$), (B) Estimates of Z_K for data simulated on phylogenies with 128 taxa ($n = 128$), (C) Variance in the variation of λ_{est} across input levels of phylogenetic signal, estimated on phylogenies containing differing numbers of species. (D) Variance in the variation of Z_K across input levels of phylogenetic signal, estimated on phylogenies containing differing numbers of species.

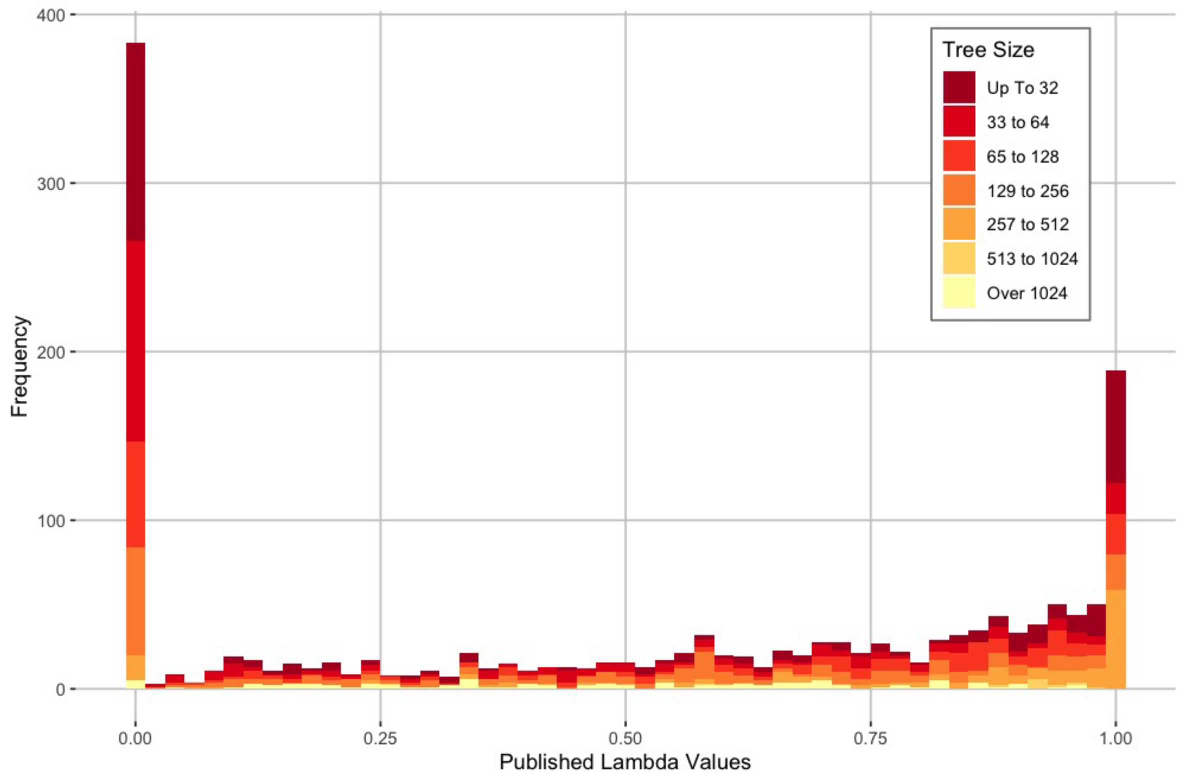

Figure 1. Frequency distribution of λ estimates published in 2019. The majority of these values were close to 0 or 1, and from phylogenies with fewer than 200 taxa.

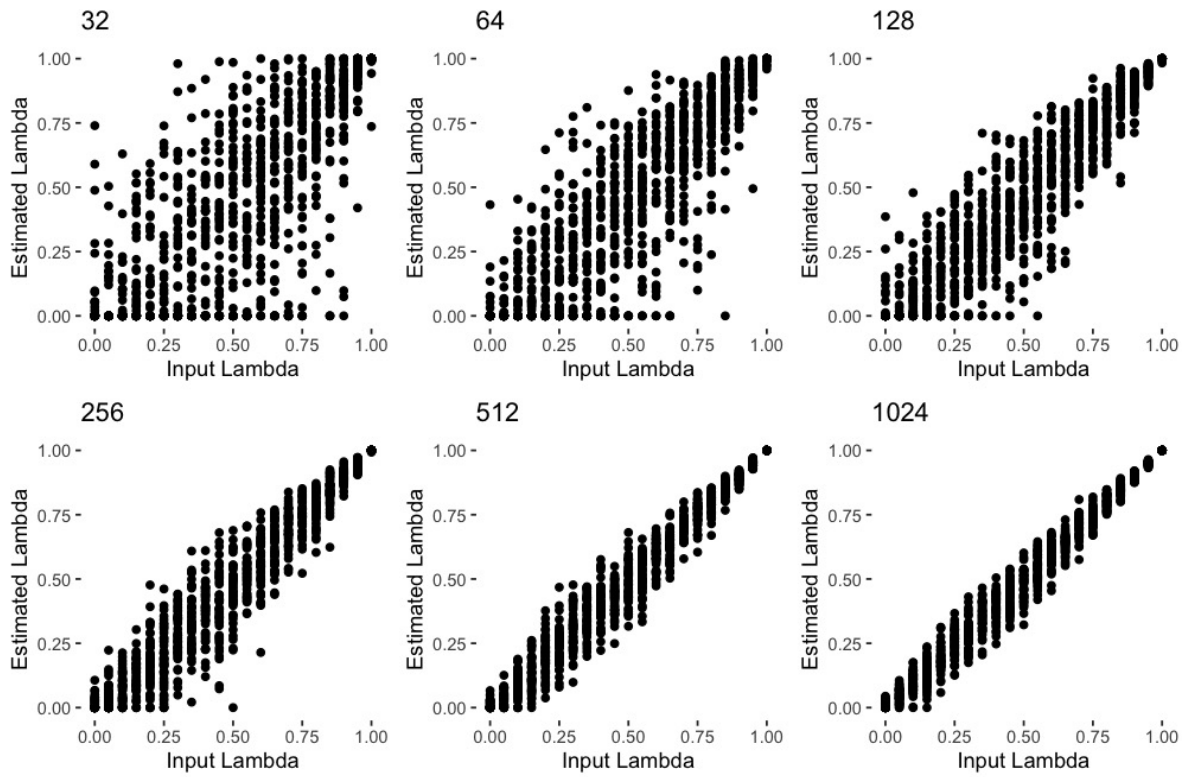

Figure 2. Precision of Pagel's λ across known levels of input phylogenetic signal (λ_{in}) on phylogenies of various sizes. As phylogenies increase in size, variation in λ_{in} decreases; however the precision is not constant across the range of input levels ($\lambda_{in} : 0 \rightarrow 1$), and is highest at intermediate levels of phylogenetic signal.

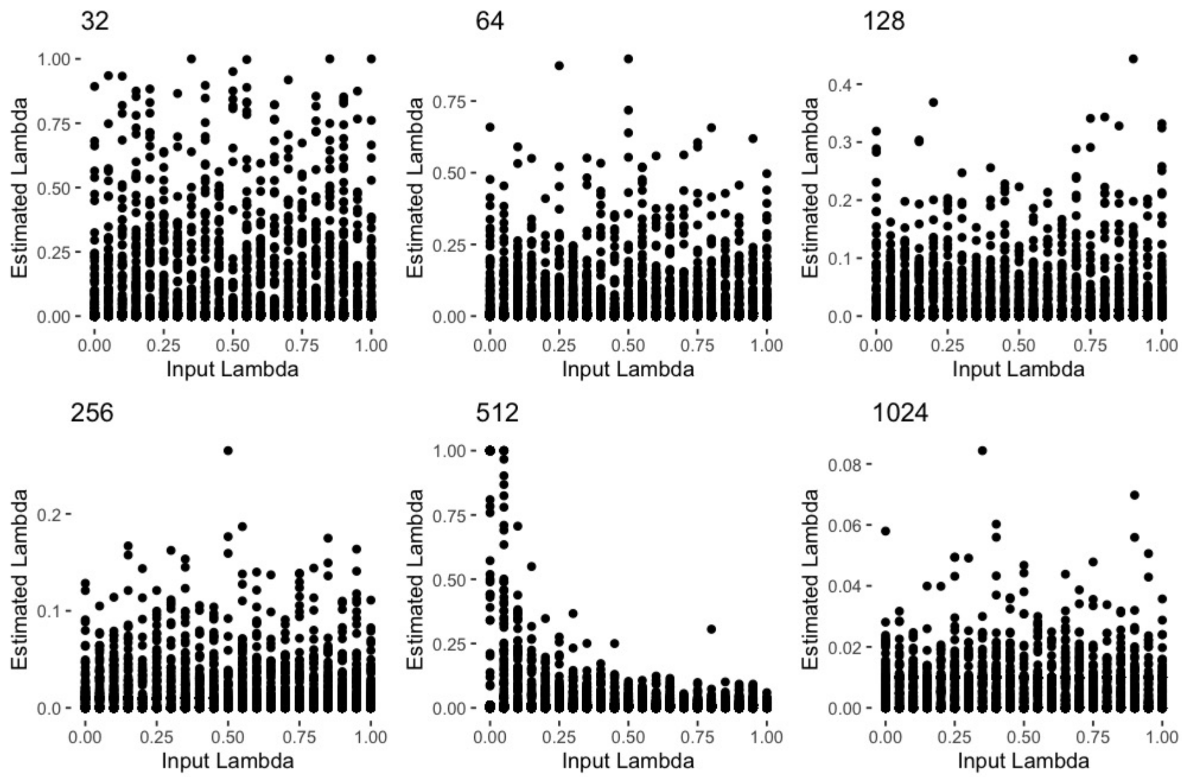

Figure 3. Precision of Pagel's λ when incorporated in phylogenetic regression ($Y \sim X$), across known levels of input phylogenetic signal (λ_{in}) on phylogenies of various sizes. As phylogenies increase in size, variation in λ_{in} decreases; however the precision is not constant across the range of input levels ($\lambda_{in} : 0 \rightarrow 1$), and is highest at intermediate levels of phylogenetic signal.

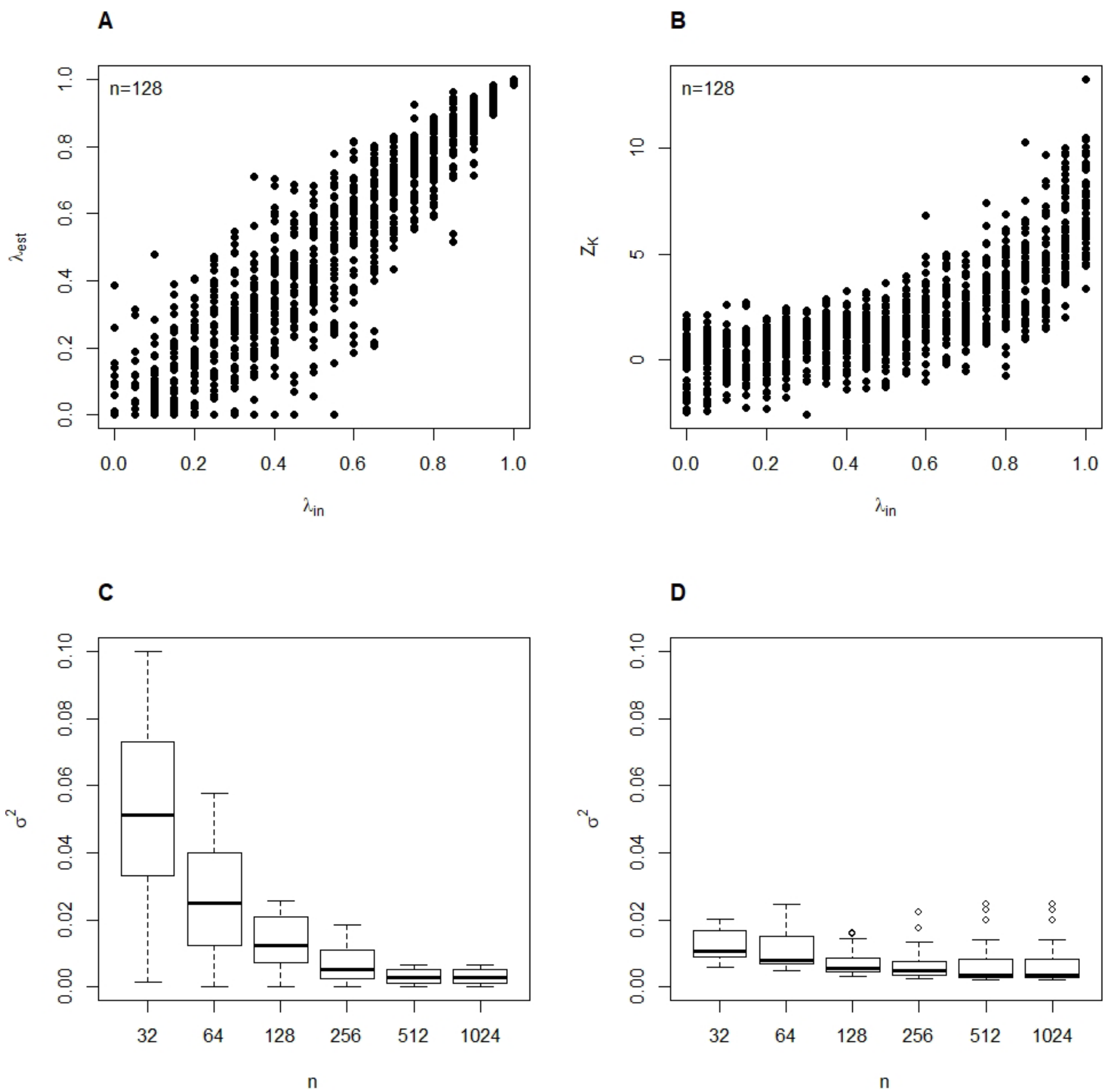

Figure 4. Variation in estimates of phylogenetic signal across input levels of phylogenetic signal. (A) Estimates of Pagel's λ for data simulated on phylogenies with 128 taxa ($n = 128$), (B) Estimates of Z_K for data simulated on phylogenies with 128 taxa ($n = 128$), (C) Variance in the variation of λ_{est} across input levels of phylogenetic signal, estimated on phylogenies containing differing numbers of species. (D) Variance in the variation of Z_K across input levels of phylogenetic signal, estimated on phylogenies containing differing numbers of species.