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, which ranges from $0 \rightarrow 1$, is then treated as a measure of phylogenetic signal; with values close to 0 signifying little phylogenetic signal, and values close to 1 indicating that trait variation accumulates according to Brownian motion. 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 $\lambda \approx 0$ signifying 'weak' phylogenetic signal, and $\lambda \approx 1$ being 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, through the use of confidence intervals (Vandelook et al. 2019) or 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). Likewise, qualitative comparisons of λ across multiple phenotypic traits have also 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, it seems intuitive to interpret the strength of phylogenetic signal in this manner, as λ is a parameter on a bounded scale ($0 \rightarrow 1$) for which interpretation of its extremal points are well known ($\lambda = 0$ is no phylogenetic signal, while $\lambda = 1$ corresponds to 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, and that parameter estimates of λ are representative of the underlying phylogenetic signal. However, for such applications, understanding the precision in estimating levels of phylogenetic signal is paramount, because biological inferences regarding the strength of phylogenetic signal in phenotypic datasets requires high precision of its estimation. 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 with greater sample sizes (Cohen 1988). However, this conclusion also assumes that the precision in estimating 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 λ is observed on 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 (e.g., meta-analysis: Hedges and Olkin 1985; Glass 1976; Arnqvist and Wooster 1995), or compared across datasets (e.g., Adams and Collyer 2016, 2019a). Unfortunatley, not all statistical 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 effect size for measuring the strength of phylogenetic signal. 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. 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

-Lambda is not precise: holds in estimates, and PGLSregression -Precision varies with N (duh) but also WITH INPUT! -same input on different N, different lambda.est -same lambda.est from different input -Boettiger

example: add small tips, greatly change lambda... -Conclusion: Lambda does not quantify strength of phylogenetic signal as is often assumed

-Empirical examples: how interpretation can get folks ito trouble.

-A proposed Effect size for the strength of physig -Derivation of effect size (expand on Adams and Collyer 2019 suggestion that an effect size for phylogenetic signal is needed. The 'gold standard' of these is a standaredized effect size: (following the meta-analytic literature beginning with Glass 1976). In other words, a Z-score. So, how does Z work when based on lambda? On kappa? - Z-score. could be Lambda or Kappa. Show it is Kappa -Comparing the strength of physig - two sample Z-score

-Conclusions and Implications

λ is not Precise

Lambda does not measure phylogenetic signal

An effect size for Physig

Materials and Methods

Simulations

We conducted a series of computer simulations to evaluate the ability of Pagel's λ to estimate known levels of phylogenetic signal. 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). Our first set of simulations evaluated the extent to which values of λ estimated from the data corresponded with actual input levels of phylogenetic signal. For these simulations 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. We used a set of values that encompassed the entire range of λ: ($\lambda_{in} = 0.0 \rightarrow 1.0$; in 21 intervals of 0.05 units), 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 known and differing levels of phylogenetic signal. These varied from datasets with no phylogenetic signal (when $\lambda_{in} = 0$) to datsets with phylogenetic signal corresponding to what was expected under Brownian motion (when $\lambda_{in} = 1$).

Using the simulated phylogenies we estimated the phylogenetic signal in each dataset using Pagel's λ. Next, for the 1050 datasets at each level of species richness (n), we assessed the relationship between λ_{in} and λ_{est} using regression. Additionally, the precision of λ was approximated by observing the variation of λ_{est} obtained at each input level of phylogenetic signal (λ_{in}). Finally, for comparison we characterized the strength of phylogenetic signal in each dataset using a standardized effect size (Z_K: sensu Adams and Collyer 2016, 2019a) based on $Kappa$. As suggested by Adams and Collyer (2019b), an effect size for phylogenetic signal may be estimated as: $Z_K = \frac{K_{obs} - \mu_K}{\sigma_K}$, where K_{obs} was the observed $Kappa$, and μ_K and σ_K were the mean and standard deviation of the empirical sampling distribution of values obtained from the permutation distribution. Z_K describes the strength of phylogenetic signal as a standard deviate from its sampling distribution, and thus directly measures the strength of signal in a manner that is comparable across datasets. Variation in the set of Z_K at each input level of phylogenetic signal was then calculated as an estimate of precision in Z_K. However, because Z_K differs in scale from λ, we used a linear normalization to standardize Z_K to a uniform distribution ($0 \rightarrow 1$), and estimated the precision of Z_K from the normalized values.

Our second set of simulations evaluated the extent to which values of λ estimated in PGLS regression and ANOVA (i.e., $Y \sim X$) corresponded with actual input levels of phylogenetic signal in the response variable. As before we generated 50 pure-birth phylogenies at each of six levels of species richness ($n = 2^5 - 2^{10}$), and rescaled each with a set of input λ values ($\lambda_{in} = 0.0 - 1.0$; in 21 intervals of 0.05 units). Next, 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). For each dataset, the fit of the phylogenetic regression was estimated using maximum likelihood, and parameter estimates (β_{est} and λ_{est}) were obtained and evaluated as above. 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.

Literature Survey

To determine how Pagel's λ is commonly utilized in empirical studies, we conducted a literature survey. From Google.scholar we obtained a list of all papers published in 2019 that used λ; resulting in 341 studies. For each study, we extracted all λ_{est}, the size of the phylogeny (n), and noted whether authors reported confidence intervals or performed significance tests assessing difference of λ_{est} from either zero or one. We also noted whether biological interpretations based on λ_{est} were made, and for studies that reported more than one λ_{est}, we also noted whether these were compared in some manner, and whether such comparisons were accompanied with statistical tests between λ_{est}.

Results

Simulations

Our first set of simulations revealed several patterns. First, the relationship between λ_{est} and λ_{in} across the range input levels of phylogenetic signal was $\beta \sim 1.0$; indicating that the average λ_{est} across datasets for a given simulation condition correctly reflected the input levels of phylogenetic signal. However, as shown in Figure 1, the precision of those estimates varied widely, and in several interesting ways. Predictably, precision was worse at low levels of species richness, where in many cases the set of λ_{est} spanned nearly the entire range of possible values (e.g., $n = 32$; $\lambda_{in} = 0.5$: range of $\lambda_{est} = 0.0 \rightarrow 0.985$). Second, as species richness increased, variation across estimates of λ decreased (Figure 1). This confirmed the pattern identified by Boettiger et al. (2012) on a small ($n = 13$) versus a large ($n = 281$) phylogeny, and demonstrated that the trend of increasing precision with higher species richness was general; adhering to parametric statistical theory. Importantly however, our broader set of simulations revealed that the precision of λ_{est} was not uniform across all levels of phylogenetic signal. Specifically, we found that variation in λ_{est} was highest at intermediate levels of phylogenetic signal, and decreased at both lower and higher levels of input signal (Figure 1). This implied that the precision in estimating λ was worse at intermediate values, and improved as the levels of phylogenetic signal were closer to $\lambda_{in} = 0$ (no phylogenetic signal) or $\lambda_{in} = 1$ (Brownian motion). Notably, even at large levels of species richness, the range of λ_{est} still encompassed a substantial portion of possible values (e.g., $n = 512$; $\lambda_{in} = 0.5$: range of $\lambda_{est} = 0.32 \rightarrow 0.68$). Likewise, the same λ_{est} could be obtained from datasets containing vastly different input levels of phylogenetic signal (e.g., $n = 512$; $\lambda_{est} = 0.5$; range of $\lambda_{in} = 0.25 \rightarrow 0.65$). 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 1 here]

By contrast, when characterizing phylogenetic signal with the standardized effect size (Z_K) of $Kappa$, we found that the precision of Z_K was more stable, as variation across datasets was far more consistent across the range of input values. For example, when $n = 128$ the precision of λ_{est} (Figure 2A) varied considerably more across input levels of phylogenetic signal than did the precision of Z_K for the same datasets (Figure 2B). Further, for a given n, the variance in precision estimates of Z_K was considerably smaller than the variance in precision estimates of λ_{est} (Fig. 2C,D); implying that estimates of phylogenetic signal were more reliable and robust when using Z_K (for additional results see Supporting Information). Finally, it should also be recognized that because Z_K is a standardized effect size, the strength of phylogenetic signal is more readily interpretable when using this measure, as Z_K expresses the strength of phylogenetic signal in standard deviation units relative to the mean (see Adams and Collyer 2019b). This further implies that comparisons of the strength of phylogenetic signal among phenotypic traits are possible, and may be accomplished statistically via a two-sample test that formally compares Z_K across datasets (for comparisons of multivariate effect sizes see: Adams and Collyer 2016, 2019a).

[insert Figure 2 here]

When λ was incorporated in PGLS regression and ANOVA (i.e., $Y \sim X$), we found much the same pattern as in our earlier simulations. Namely, the precision of λ_{est} covaried with species richness; where greater precision was obtained at higher levels of species richness (Figure 3). Likewise, the precision in estimating λ was worse at intermediate values (Figure 3), and improved as the levels of phylogenetic signal were closer to $\lambda_{in} = 0$ (no phylogenetic signal) or $\lambda_{in} = 1$ (Brownian motion). Generally, the spread of estimates was slightly broader when λ was co-estimated with regression parameters in PGLS regression (Figure 3), as compared to when λ was estimated for only the dependent variable (Figure 1). Finally, we found that regression parameters (β) were accurately estimated when λ was included during phylogenetic regression; a result which confirmed earlier findings of Revell (2010) (see Supporting Information).

[insert Figure 3 here]

Literature Survey

We found 182 manuscripts from 2019 that estimated and reported Pagel's lambda values using PGLS methods. These papers averaged 8.527 lambda values, ranging from a single lambda estimate up to 71 estimated lambdas. Almost exactly half of the published lambda estimates were either below 0.05 (25.32%) or above 0.9 (24.74%; Figure 3). 73.32% of the published lambdas were estimated using phylogenies with fewer than 200 tips, and 348 lambda estimates (8.57% of all published estimates) came from phylogenies with fewer than 30 tips.

[insert Figure 4 here]

Many of the reviewed manuscripts liberally interpreted the magnitude of the estimate lambda, using phrases such as "strong" or "weak" phylogentic signal when statistically, all that was clear was a difference between the estimated lambda and 0 or 1 respectively. We estimated that about 20.49% of the manuscripts revealed some sort of biological interpretation of the magnitude of estimated phylogenetic signal that overreached the statistical findings. We also identified seven manuscripts as having inappropriately interpreted differences in lambda values, indicating that some traits had stronger or weaker signal than other traits without the appropriate statistical tests.

As is evidenced by macroevolutionary papers published in 2019 papers, Pagel's lambda estimation methods are often misused and over-interpretted. Despite the urging of Boettiger and colleagues to publish confidence intervals with all lambda parameter estimates, only 18% of papers published in 2019 do so.

Results

Discussion

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.

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.

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. 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 2. 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 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. Frequency of estimated lambda values published in manuscripts in 2019. The majority of these values were close to 0 or 1, and from phylogenies with fewer than 200 taxa.

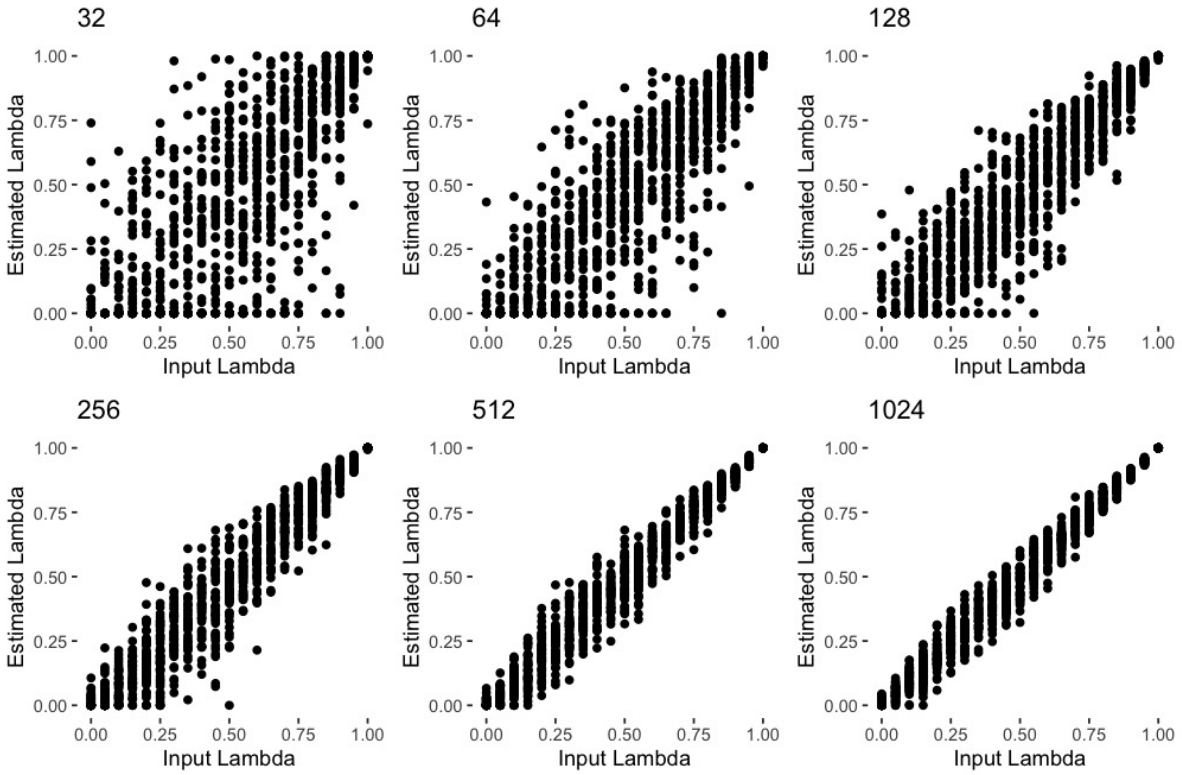

Figure 1. Precision of Pagel's λ across known levels of input phylogenetic signal (λ_{in}) on phylogenies of various sizes. As phylogenies increase in size, variation in λ_{in} decreases; however the precision is not constant across the range of input levels ($\lambda_{in} : 0 \rightarrow 1$), and is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

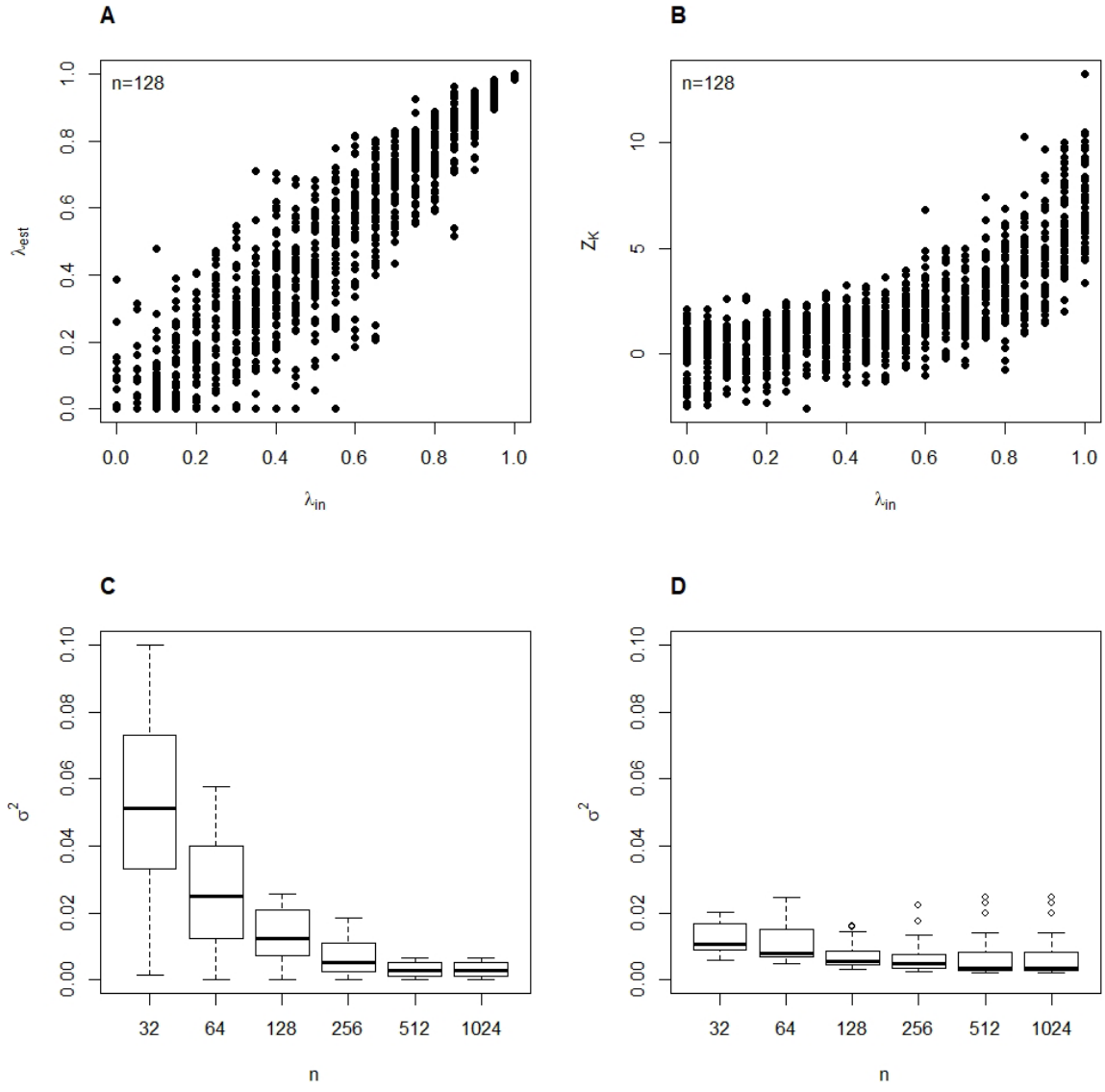

Figure 2. Variation in estimates of phylogenetic signal across input levels of phylogenetic signal. (A) Estimates of Pagel's λ for data simulated on phylogenies with 128 taxa ($n = 128$), (B) Estimates of Z_K for data simulated on phylogenies with 128 taxa ($n = 128$), (C) Variance in the variation of 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.

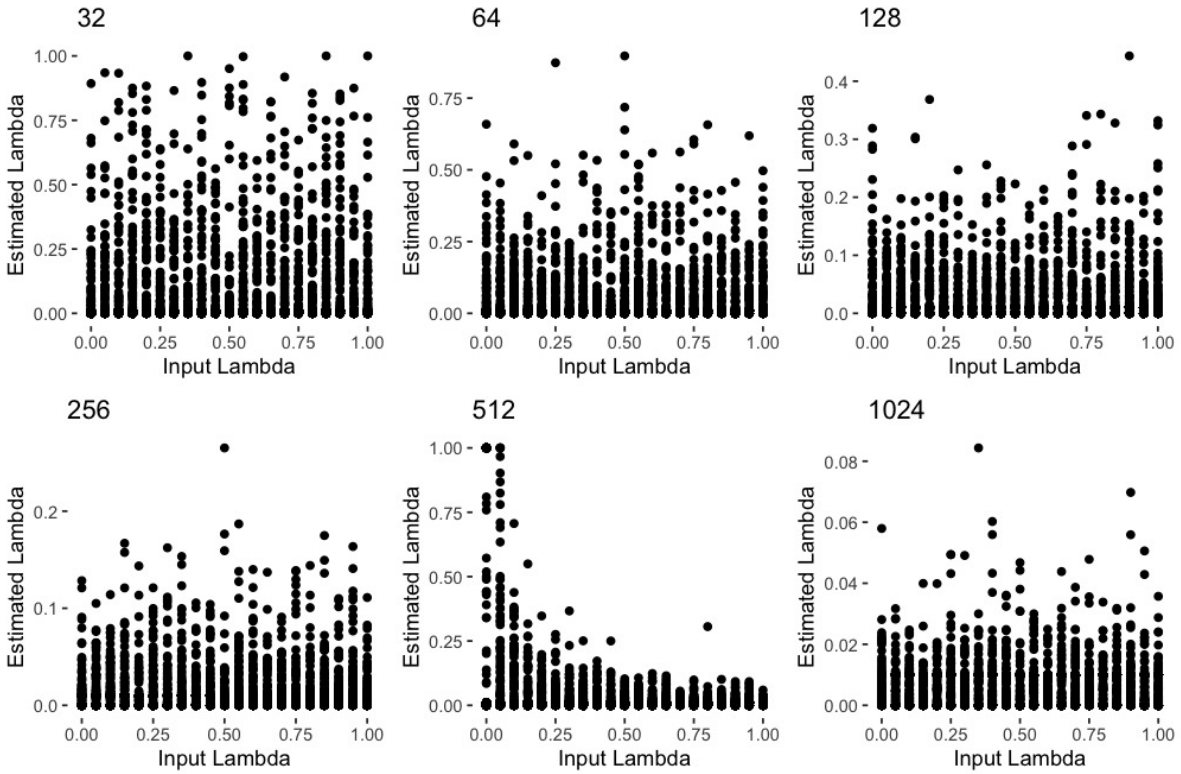

Figure 3. Precision of Pagel's λ when incorporated in phylogenetic regression ($|Y \sim X$), across known levels of input phylogenetic signal (λ_{in}) on phylogenies of various sizes. As phylogenies increase in size, variation in λ_{in} decreases; however the precision is not constant across the range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

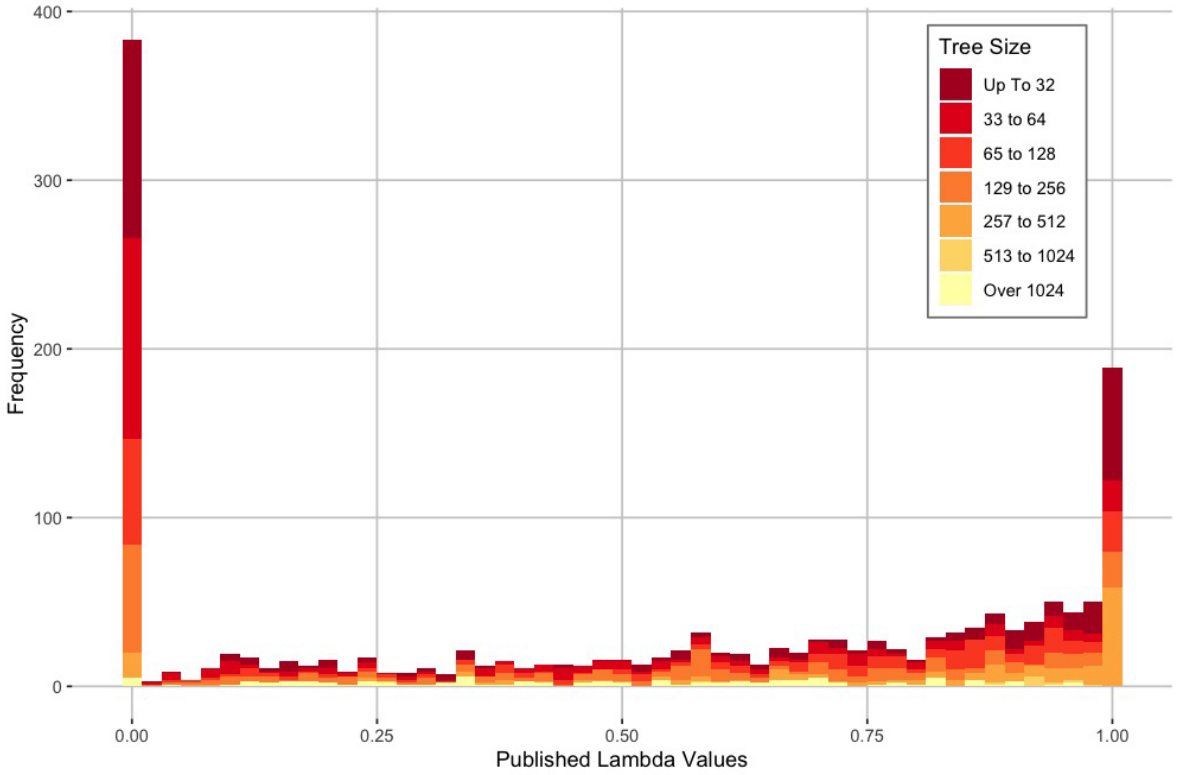

Figure 4.Frequency of estimated lambda values published in manuscripts in 2019. The majority of these values were close to 0 or 1, and from phylogenies with fewer than 200 taxa.