

REVIEW AND SYNTHESES

Does phylogeny matter? Assessing the impact of phylogenetic information in ecological meta-analysis

Scott A. Chamberlain,^{1*} Stephen M. Hovick,¹ Christopher J. Dibble,¹ Nick L. Rasmussen,¹ Benjamin G. Van Allen,¹ Brian S. Maitner,¹ Jeffrey R. Ahern,¹ Lukas P. Bell-Dereske,¹ Christopher L. Roy,¹ Maria Meza-Lopez,¹ Juli Carrillo,¹ Evan Siemann,¹ Marc J. Lajeunesse² and Kenneth D. Whitney¹

¹Department of Ecology and Evolutionary Biology, Rice University, Houston, TX 77006, USA
²Department of Integrative Biology, University of South Florida, Tampa, FL 33620, USA

*Correspondence: E-mail: myrmecocystus@gmail.com

Abstract

Meta-analysis is increasingly used in ecology and evolutionary biology. Yet, in these fields this technique has an important limitation: phylogenetic non-independence exists among taxa, violating the statistical assumptions underlying traditional meta-analytic models. Recently, meta-analytical techniques incorporating phylogenetic information have been developed to address this issue. However, no syntheses have evaluated how often including phylogenetic information changes meta-analytic results. To address this gap, we built phylogenies for and re-analysed 30 published meta-analyses, comparing results for traditional vs. phylogenetic approaches and assessing which characteristics of phylogenies best explained changes in meta-analytic results and relative model fit. Accounting for phylogeny significantly changed estimates of the overall pooled effect size in 47% of datasets for fixed-effects analyses and 7% of datasets for random-effects analyses. Accounting for phylogeny also changed whether those effect sizes were significantly different from zero in 23 and 40% of our datasets (for fixed- and random-effects models, respectively). Across datasets, decreases in pooled effect size magnitudes after incorporating phylogenetic information were associated with larger phylogenies and those with stronger phylogenetic signal. We conclude that incorporating phylogenetic information in ecological meta-analyses is important, and we provide practical recommendations for doing so.

Keywords

Akaike Information Criterion, Colless' yule, fixed-effects, model fit, maximum likelihood, phylogenetic signal, phylogenetic non-independence, quantitative review, random-effects.

Ecology Letters (2012) 15: 627–636

INTRODUCTION

Meta-analysis is now an important tool in ecology and evolutionary biology where it is widely used to infer general patterns from primary studies. Meta-analyses synthesise data by calculating effect sizes, which measure the magnitude and direction of experimental outcomes in standardised units to facilitate among-study comparisons. Since the first quantitative meta-analysis in ecology and evolutionary biology (Gurevitch *et al.* 1992), hundreds of additional meta-analyses have been conducted (Fig. 1). Often, ecological and evolutionary meta-analyses summarise data from experiments involving different species, ranging from studies of a single taxonomic group (e.g. Insecta; Huberty & Denno 2004) to studies across multiple divergent groups (e.g. animals, plants and fungi; Persson *et al.* 2010).

A traditional (non-phylogenetic) meta-analysis that synthesises studies from different species can violate two statistical assumptions. First, samples are not independent because they share evolutionary history to varying degrees, and this shared history often leads to a correlated data structure. For example, one review of ecological traits measured in comparative studies found that nearly 90% of datasets had at least one trait with significant phylogenetic dependence (Freckleton *et al.* 2002). Thus, in most situations different species cannot be considered statistically independent. Second, samples are not drawn from a normally distributed population with a common variance because species come from lineages that have evolved at different rates (Lajeunesse 2009). Incorporating phylogenetic infor-

mation into ecological meta-analyses can ameliorate both of these problems (Adams 2008; Lajeunesse 2009). Furthermore, incorporating a phylogeny yields smaller variance estimates, reducing Type I error rates when parameter estimates are equal to zero and giving more powerful tests otherwise (Rohlf 2006). One potential challenge to the increased use of phylogenetic information in meta-analyses is that available phylogenies are often not fully resolved (leaving a number of soft polytomies). However, phylogenetic comparative methods appear to be relatively robust to some lack of resolution (Rohlf 2006; Stone 2011).

With increasing awareness of the dependence of species traits and ecological processes on phylogeny, meta-analyses now often use methods to account for phylogenetic history when effect sizes can be assigned to individual species (Fig. 1 inset). In perhaps the simplest form of a phylogenetic meta-analysis, taxonomic rank (e.g. genus or family) is included in analyses as a grouping variable to assess taxonomic differences in effect sizes (e.g. Marczak *et al.* 2007). If the effects of phylogenetic history play out only at coarse scales of taxonomic resolution, then complete phylogenies may not be necessary. Indeed, some traits that mediate ecological interactions are highly conserved at the family, subfamily or genus level; for example, among the subfamilies of leguminous plants, nitrogen fixation is nearly ubiquitous in the Mimosoideae and Papilionoideae, but rare in the Caesalpinioideae (de Faria *et al.* 1989). Another method to account for phylogenetic dependence uses pairwise distances (phylogenetic branch lengths) between pairs of species as a covariate

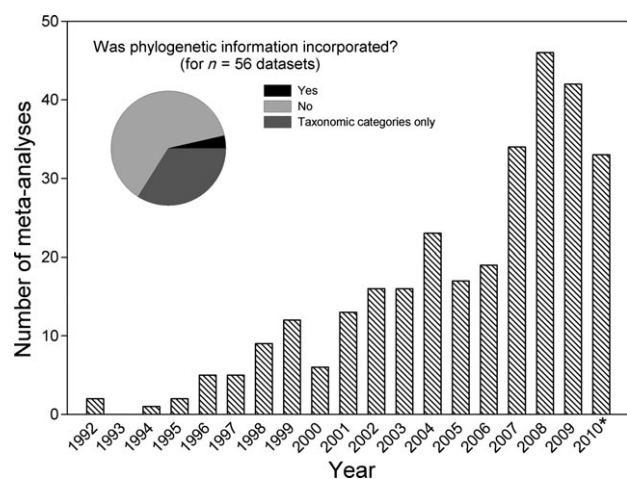


Figure 1 Bars show the number of published meta-analyses subject to potential phylogenetic non-independence (i.e. effect sizes were measured at the species level for at least three species; $n = 301$) from 1 January, 1992 to 25 October, 2010. For the subset of 56 meta-analyses that also made the dataset available and reported a measure of uncertainty (our full criteria for consideration in our re-analysis), the inset figure indicates the proportions that performed a phylogenetic meta-analysis (black; $n = 2$), assessed whether effect sizes differed among taxonomic categories using traditional meta-analysis (dark grey; $n = 19$) or conducted traditional meta-analyses only (light grey; $n = 35$). See Methods for details about the search criteria.

in the meta-analysis (e.g. Morales & Traveset 2009). A third approach accounts for phylogenetic dependence by transforming effect sizes prior to meta-analysis using phylogenetically independent contrasts (Abouheif & Fairbairn 1997; Dubois & Cezilly 2002). Finally, a recent development, proposed by Adams (2008), weights effect sizes by their relative sampling error (as in traditional meta-analysis) and then re-weights them using phylogenetic covariances. In a refinement to this method, Lajeunesse (2009) proposed simultaneously weighting effect sizes by relative sampling error and phylogenetic distances; this latter approach is quickly gaining use among ecologists (e.g. Carmona *et al.* 2011; DelBarco-Trillo 2011; Meunier *et al.* 2011; Munguia-Rosas *et al.* 2011).

Despite the increasing use of phylogenetic meta-analysis, there is no empirical assessment of how often, and under which circumstances, phylogenetic meta-analysis is important. Put simply, we do not know how often accounting for phylogenetic relatedness among taxa changes the outcome or interpretation of a meta-analysis. As incorporating phylogenetic information can take substantial effort, a broader understanding of the effects of incorporating phylogenetic history into ecological and evolutionary meta-analyses is timely.

Herein, we re-analyse datasets from previously published meta-analytic studies, comparing results of traditional and phylogenetic meta-analyses. In addition, we attempt to explain variation in the effect of phylogenetic information on meta-analytic outcomes by examining characteristics of phylogenies. We ask: (1) how does accounting for phylogenetic non-independence change results of individual meta-analyses? and (2) across datasets, what characteristics of phylogenies explain changes in effect size for phylogenetic vs. traditional meta-analyses? As a complement to our main questions, in Appendix A, we also ask (3) how does accounting for phylogenetic non-independence affect model fit of individual meta-analyses? and (4) across datasets, what characteristics of phylogenies explain variation in the relative fit of phylogenetic meta-analyses? Despite

the many compelling reasons to incorporate phylogenetic information into meta-analyses that involve multiple species, investigators often use model comparison criteria, such as Akaike's Information Criterion (AIC) to assess fit of phylogenetic vs. traditional meta-analytic models. We found a clear bias in relation to phylogeny size for one of the two methods currently used to quantify relative model fit (Q-based AIC), thus our findings have important implications for meta-analysts using such model comparisons (see Appendix A for details).

METHODS

Data selection criteria

To select datasets for our study, we conducted a comprehensive search for published meta-analyses using ISI Web of Science (<http://www.isiknowledge.com>). On 25 October 2010, we searched 'meta-analys*' or 'metaanalys*' within the ISI ecology and evolutionary biology subject areas, which yielded 937 journal articles published since 1992. From this set, we retained meta-analytic datasets that met three criteria. First, the effect sizes reported must have assessed a response at the level of individual taxa (i.e. species) for three or more taxa. In this way, we excluded datasets for which a phylogenetic meta-analysis would have been impossible (e.g. meta-analyses on a single focal species or on community-level responses, such as diversity or evenness). Second, effect size data must have been provided, either within the article itself or in an online archive. Third, some measure of uncertainty (e.g. variance) around the effect size estimate (or the data necessary to calculate it) must have been provided. Of the 56 meta-analytic datasets that met these three criteria, we randomly selected 30 for our analyses. Data and phylogenies for each dataset are provided in Appendix B.

It was not possible to fully re-create all analyses from the original datasets, which often included multiple meta-analyses per dataset, multiple grouping variables per meta-analysis and/or multiple effect sizes per species. From each study, we selected a single meta-analysis and a single grouping variable (if included in the original dataset), in both cases maximising the number of effect sizes (number of species or genera) to maximise statistical power. Grouping variables were utilised in many datasets to compare effect sizes among categories using a range of criteria, including habitat types, experimental methodology and functional categorisation (see Appendix C for details). Where more than one meta-analysis or grouping variable yielded the same sample size, we made selections at random. When a given meta-analysis reported multiple effect sizes for the same species, we pooled effect sizes for that species using a fixed-effect meta-analysis (Shadish & Haddock 1994).

Phylogeny reconstruction

We created phylogenetic trees with branch lengths for each dataset using a variety of methods. Plant-only phylogenies used the topology from the Davies *et al.* (2004) supertree (through the Phylomatic web service; Webb & Donoghue 2005) and node age estimates from Wikström *et al.* (2001). Topology and branch lengths for bird-only phylogenies were obtained from Hackett *et al.* (2008), with additional taxa added using the online tree of life (Maddison *et al.* 2007). For datasets including divergent animal taxa, we manually built phylogenies in MESQUITE v. 2.73 (Maddison & Maddison 2010) using information from multiple published phylogenies (see references in

Appendix Table D1); we then added branch lengths by ageing all possible internal nodes with TimeTree (Hedges *et al.* 2006). Node ages relied primarily on TimeTree's weighted average estimates and secondarily on TimeTree's Expert Results. For each phylogeny, we used the algorithm *bladj* in Phylocom (Webb *et al.* 2008) to interpolate ages for undated nodes. Branch lengths are presented in millions of years. When phylogenies could not be fully resolved, we retained polytomies rather than removing species with uncertain evolutionary relationships.

Finally, tree topology and branch lengths for the MacKenzie *et al.* (2003) fish dataset were obtained by building a molecular phylogeny using sequences (12s and 16s rRNA) obtained from GenBank (Appendix Table D2). Genes were aligned with Clustal multiple alignment using BioEdit v.7.0.9 (Hall 1999). Genes were aligned separately and then concatenated. We used MrBayes v.3.1.2 (Huelsenbeck *et al.* 2001; Ronquist & Huelsenbeck 2003) to build the phylogeny, with gamma-distributed rate variation across sites and a proportion of invariable sites (GTR model), with 1 000 000 generations. All 30 phylogenies are provided in Appendix D as plotted trees and in newick format and in Appendix B as individual text files.

Predictors of phylogenetic meta-analysis outcomes

In an attempt to explain variation in the effects of incorporating phylogeny across datasets, we analysed relationships between meta-analytic results (changes in effect size) and six predictor variables (phylogeny size, phylogenetic signal, phylogeny age, phylogenetic resolution and two metrics that quantify tree shape). As these analyses were intended to be exploratory, we included a broad selection of predictors that we thought might affect phylogenetic meta-analysis results. We quantified phylogeny size because larger phylogenies have more species and thus greater statistical power to detect phylogenetic effects (Freckleton *et al.* 2002; Rezende *et al.* 2007). We also quantified phylogeny age (age of the root node in millions of years, compiled from <http://www.timetree.org>), phylogenetic resolution (the proportion of dichotomous nodes in the phylogeny) and phylogenetic signal (Blomberg's K ; Blomberg *et al.* 2003) in the effect sizes. We expected that datasets with a strong phylogenetic signal would be most sensitive to phylogenetic meta-analytic methods. Values of K close to zero indicate that closely related species do not share similar trait (in this case, effect size) values, whereas values of K approaching and larger than one suggest that closely related species do share similar trait values (Fig. 2). Finally, we quantified two measures of tree shape from reconstructed phylogenies: Colless' yule, I_c (a measure of tree balance; Colless 1982) and γ (a measure of the distribution of internal nodes between the root and the tips; Davies *et al.* 2011). Smaller values of I_c indicate that a phylogeny is more balanced with speciation events spread equally across clades, whereas larger values of I_c indicate that a phylogeny is less balanced with speciation events occurring asymmetrically across the phylogeny (Fig. 2). Smaller values of γ suggest that speciation was concentrated early in a phylogeny, whereas larger values of γ indicate that speciation occurred relatively recently in time (Fig. 2).

The number of species, phylogenetic signal and phylogeny age were all \log_{10} transformed prior to analysis. Our continuous predictors were only moderately correlated with each other, if at all (mean $|r| = 0.41$). We omitted a number of additional predictors that were highly correlated with our final set ($r > 0.8$), including phylogenetic breadth (Σ branch lengths), mean phylogenetic distance and alternative

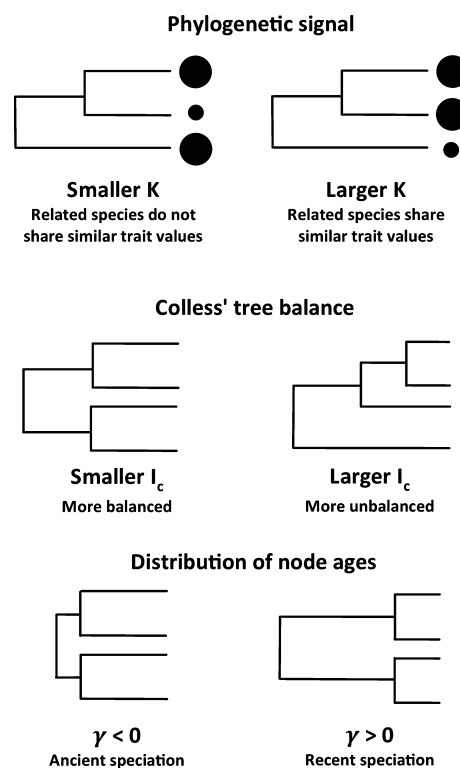


Figure 2 Characteristics of phylogenies (phylogenetic signal, tree balance and the distribution of node ages) that may influence the magnitude of differences in results between traditional and phylogenetic meta-analyses. See Methods for further description of these metrics.

measures of phylogeny shape (e.g. the beta-splitting index; Blum & François 2006). We did not use organismal group as a predictor because it was confounded with some of our other predictors (e.g. mean phylogeny age was greater for our set of plant-only phylogenies than it was for our set of bird-only phylogenies).

Data analysis

Comparison of traditional and phylogenetic meta-analyses within datasets: We performed traditional and phylogenetic meta-analyses for each of the 30 selected datasets, comparing the overall pooled effect sizes and their 95% confidence intervals (CI) from both methods. For datasets with grouping variables, we also assessed effect sizes and 95% CIs for each group. We present results for both fixed- and random-effects models because neither is the clear method of choice (i.e. both have caveats associated with their use). On one hand, multiple ecological datasets are unlikely to share one true underlying effect size, as assumed by fixed-effects models (Gurevitch *et al.* 2001). On the other hand, there are no established methods for estimating random effects when effect sizes are correlated (e.g. via shared evolutionary history). While fixed-effects analyses assume all replicates come from a single distribution and share a common variance, traditional random-effects analyses add an additional variance component (tau [τ]) to each replicate. As an estimate of between-replicate variance, τ represents additional variation in the dataset due to each replicate being drawn from a unique distribution. However, by adding the same value of τ to all replicates in a meta-analysis, existing phylogenetic random-effects models assume that the distributions underlying those replicates are

no more similar for closely related taxa than they are for distantly related taxa.

Both traditional and phylogenetic meta-analyses were performed with PhyloMeta v.1.2 (Lajeunesse 2009, 2011), with data extraction and collation automated using R v. 2.13.1 (R Development Core Team 2011; code for running PhyloMeta from R available at <http://schamberlain.github.com/2011/04/phylo-meta-from-r-update/>). Although we focus on the Lajeunesse (2009) method, we also compare it with another method used in the literature that of Adams (2008). Appendix E compares effect sizes obtained via these two methods and finds they do not significantly differ for the majority of our datasets.

What characteristics of phylogenies explain changes in effect size for phylogenetic relative to traditional meta-analyses? In some cases, traditional and phylogenetic methods estimated similar effect sizes, whereas in others the outcomes were quite different. We attempted to explain this variation across datasets by conducting a meta-analysis of meta-analyses; we refer to this as a meta-meta-analysis (MMA). We assessed differences in the overall effect size (δ) for each dataset using Hedges' d , which we calculated as:

$$d = \frac{|\delta_p| - |\delta_t|}{s} J,$$

where δ_p and δ_t are the effect sizes from phylogenetic and traditional meta-analyses, s is the pooled standard deviation and J corrects for bias due to small sample size (Hedges & Olkin 1985). We used the absolute value of effect sizes to calculate Hedges' d because our datasets varied with respect to the expected sign of an effect (e.g. plant biomass increases in response to mycorrhizal inoculation [Hoeksema *et al.* 2010], but herbivore performance declines in water-stressed plants [Huberty & Denno 2004]). Values of $d = 0$ indicate no difference in effect sizes between a traditional and phylogenetic meta-analysis. Positive values of d indicate that accounting for phylogeny increases the magnitude of an effect, making it more likely that δ would differ significantly from zero; negative values indicate that accounting for phylogeny decreases the magnitude of an effect, making it less likely that δ would differ significantly from zero. Hedges' d was calculated for both fixed- and random-effects meta-analyses.

We used fixed-effects weighted linear models (PROC GLM; SAS v. 9.1) to explain variation in d across datasets for the MMA. All models were weighted by the inverse of variance in d (Hedges & Olkin 1985). In most cases, the inclusion of all six predictor variables yielded an over-parameterised model, thus we adopted a variable selection and model averaging approach to better assess the importance of individual predictors. For both sets of MMA analyses (fixed- and random-effect models), we sequentially removed individual variables from the full model and repeated the analysis until an intercept-only model remained. As our criteria for variable elimination, we calculated Z -statistics which are more appropriate than F -statistics for inferences in meta-analysis where each effect size has its own variance (Hedges 1994); in each step we eliminated the predictor for which Z was closest to zero. The Z -statistic is calculated as $\beta_j / (SE_j / \sqrt{MS_E})$, where β_j and SE_j are the estimate and standard error for parameter j , and MS_E is the mean square error. We calculated the small-sample bias-corrected version of AIC and Akaike weights for all seven candidate models, ranking models by their Akaike weights (Burnham & Anderson 2002; Johnson &

Omland 2004). Akaike weights are interpreted as the probability of model i being the best model for the observed data given the set of models examined, where $\sum AW_i = 1$. We selected the first M models for which $\sum AW \geq 0.95$. This reduced set of candidate models was the basis for inferences regarding the importance of, and parameter estimates for, individual predictors.

We assessed the importance of an individual predictor using the sum of Akaike weights (re-normalised so that $\sum AW = 1$) for all models in which that term appeared. This sum is called the importance weight (Burnham & Anderson 2002), and we took a conservative approach by assessing the potential influence of parameters with an importance weight ≥ 0.25 . We calculated model-averaged parameter estimates and standard errors (SE), weighting single-model estimates by their re-normalised Akaike weights (Burnham & Anderson 2002; Johnson & Omland 2004). We estimated 95% CI around model-averaged parameter estimates as the parameter estimate ± 2 SE, and we consider a parameter to be significant if the 95% CI excludes zero (Burnham & Anderson 2002).

We checked residuals for normality, dropping outliers to yield a reduced dataset that met this statistical assumption for each of the individual MMAs that we intended to compare (e.g. for both fixed- and random-effect models). Individual MMA results were qualitatively unaffected by the exclusion of these outliers.

RESULTS

How does accounting for phylogenetic non-independence change results of individual meta-analyses?

We conducted traditional and phylogenetic meta-analyses using datasets derived from 30 published meta-analyses. These datasets varied in size from 8 to 287 species, varied in phylogeny age from 53 to 2622 mya and varied taxonomically from plant-only or animal-only datasets to datasets spanning multiple kingdoms. The questions addressed in these primary meta-analyses were diverse (see Appendix C for more details) from the effect of predator removals on breeding bird population sizes (Côté & Sutherland 1997) to the effect of experimental warming on litter decomposition rates of various plant species (Aerts 2006).

Accounting for phylogenetic relationships changed effect sizes to a much greater extent for fixed-effects analyses than for random-effects analyses, including both overall pooled effect sizes and effect sizes for individual groups (Figs 3 and 4). For fixed-effects models, overall pooled effect sizes differed significantly (95% CIs did not overlap) for phylogenetic vs. traditional meta-analyses in 47% of our datasets (14 of 30), and at least one effect size differed in 63% of them (19 of 30; Fig. 3). As expected, incorporating phylogenetic information into meta-analyses did not change effect sizes in a consistent direction: in six datasets effect sizes only increased, in five datasets they only decreased and eight datasets had a combination of increasing and decreasing effect sizes (Fig. 3). Incorporating phylogenetic information also changed whether an effect size was significantly different from zero in seven of our 30 datasets (23%), and there was no directional pattern to this change (Fig. 3).

For random-effects models, traditional and phylogenetic effect sizes differed in only 7% of datasets (one overall pooled effect size and one for a single level of the grouping variable; Fig. 4).

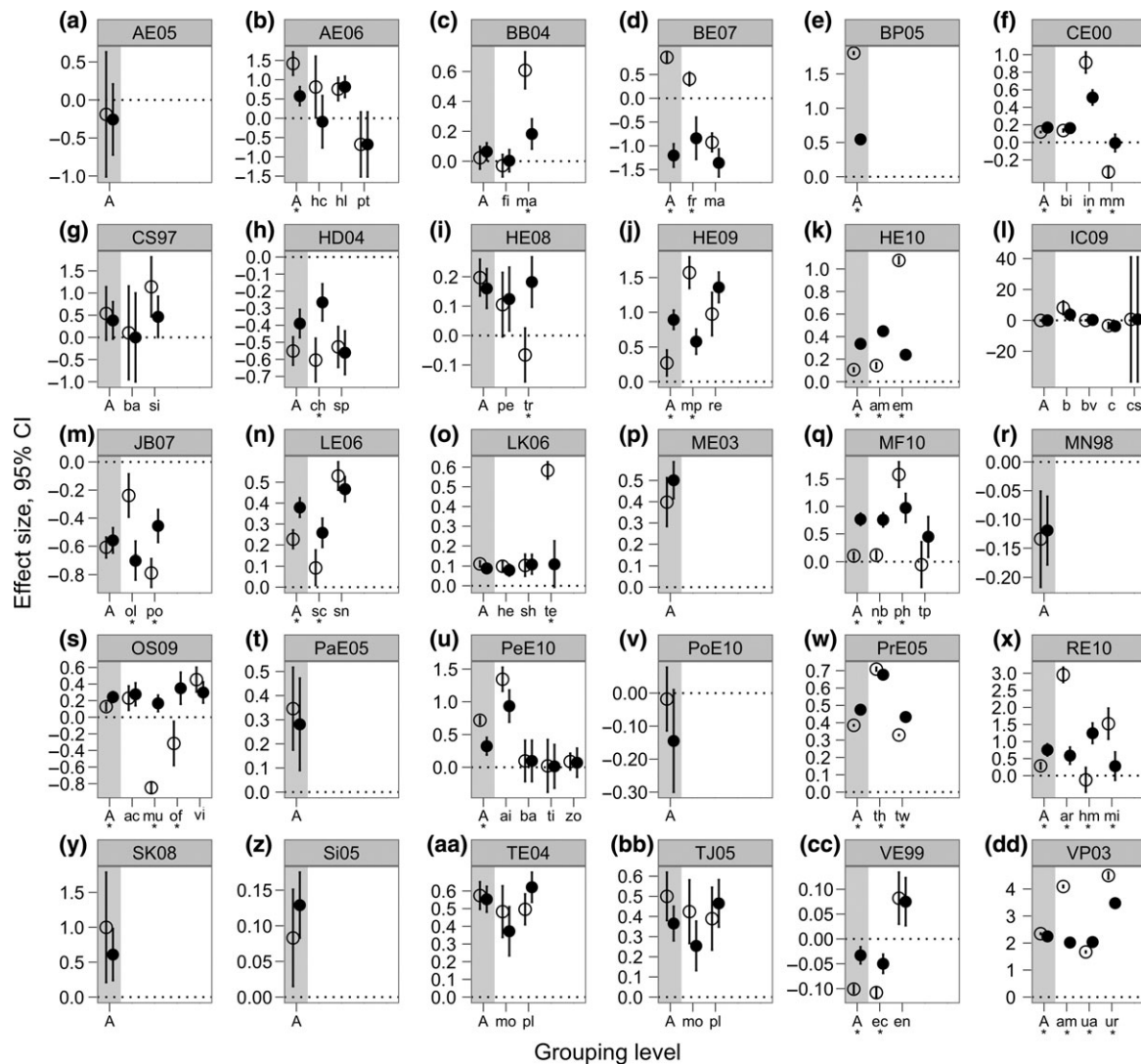


Figure 3 Traditional vs. phylogenetic fixed-effects meta-analysis results (mean pooled effect size and 95% CI) for individual datasets (a–dd). Datasets are sorted alphabetically; a key to the dataset identifier codes is given in Appendix Table D1. Open circles: phylogenetic meta-analysis. Filled circles: traditional meta-analysis. Overall pooled effect sizes are labelled by ‘A’ (indicating ‘all data’) and are shaded in grey. Two-letter codes indicate grouping levels within ‘A’ (see Appendix C for code definitions). Asterisks indicate datasets or groups for which the traditional and phylogenetic meta-analysis outcomes differed significantly.

However, incorporating phylogenetic information into random-effects models did change whether an effect size differed significantly from zero in 40% of datasets (12 of 30). In 10 of these 12 cases, effect sizes from traditional meta-analyses were significantly different from zero, but those from phylogenetic meta-analyses were not (Fig. 4).

Relative to traditional meta-analyses, incorporating phylogenetic information increased within-group heterogeneity (Q_w) in all 30 datasets. On average, this increase was nearly nine times the Q_w values for traditional meta-analyses (mean factor of increase = 8.7, median = 5.3, range = 1.4–77.6). As a result, for 63% of all datasets for which Q_w was non-significant in traditional meta-analysis, incorporating phylogeny lead to significant within-group heterogeneity (see Appendix C). This increase in heterogeneity resulted in larger CIs around effect size estimates for phylogenetic vs. traditional meta-analyses (Figs 3 and 4), often affecting whether those effect sizes differed significantly from zero.

Which characteristics of phylogenies explain changes in effect size for phylogenetic vs. traditional meta-analyses?

We quantified the degree to which including phylogenetic information changed meta-analysis effect sizes using Hedges’ d ($|\text{phylogenetic}| - |\text{traditional}|$ effect sizes) in a MMA. Our predictor variables explained more variation in d for fixed-effects models than they did for random-effects models (r^2 range: 0.37–0.47 and 0.15–0.37 for fixed- and random-effects models, respectively). However, because the magnitude of effect size change was lower for random- than fixed-effects models (see Fig. 5), there was also significantly less variation in random-effects models that could be explained by our predictors (Levene’s test: $F = 11.06$, $P = 0.002$).

For fixed-effects models, phylogenetic meta-analyses conducted using large phylogenies and those with strong phylogenetic signal had the largest decreases in effect size magnitude; in other words, the likelihood of changing one’s conclusions after accounting for

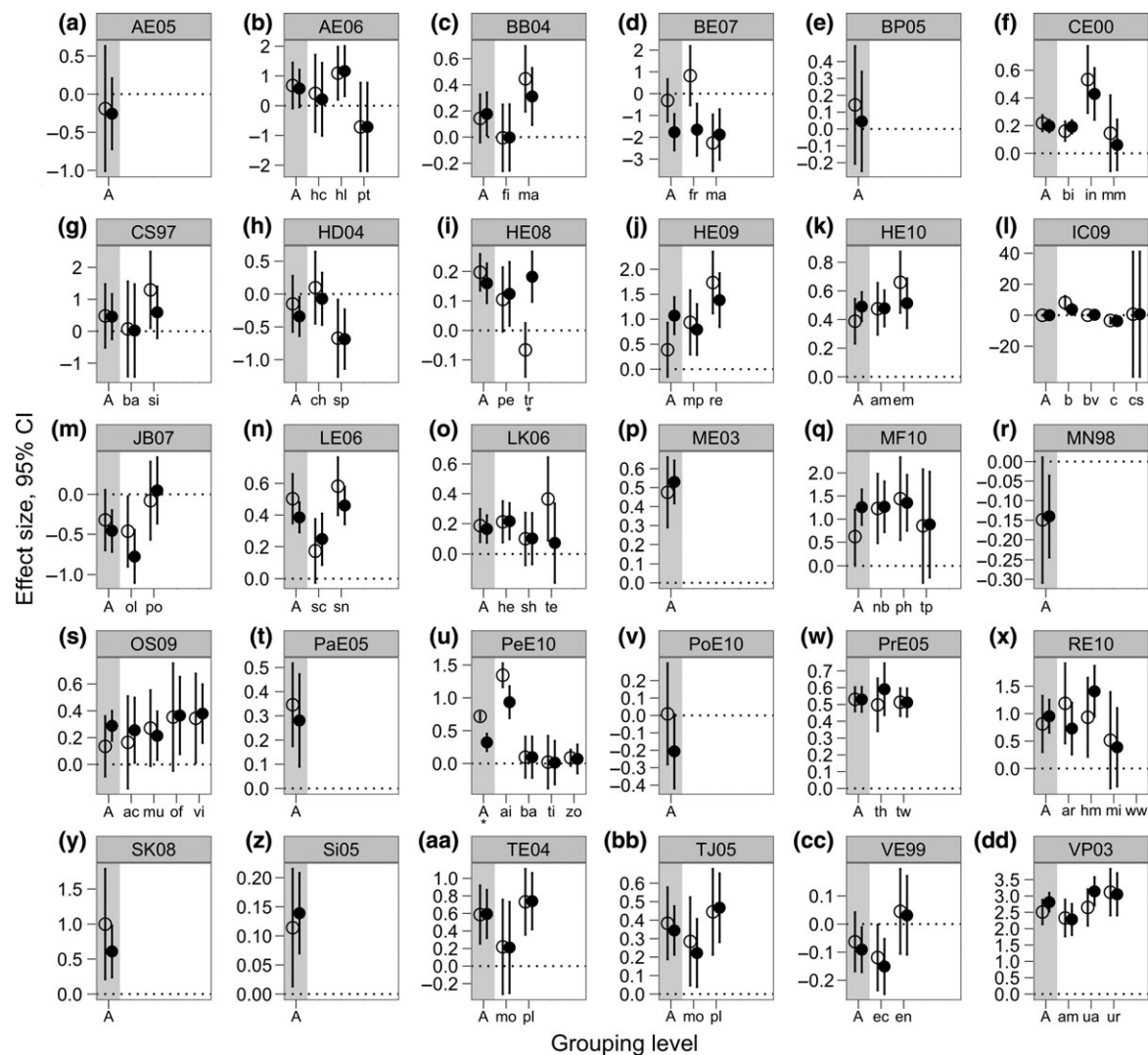


Figure 4 Traditional vs. phylogenetic random-effects meta-analysis results (mean pooled effect size and 95% CI) for individual datasets. See Fig. 3 caption for details.

phylogeny would be greatest using these types of datasets (Fig. 5; Table 1). The relationship between species number and effect size change was particularly strong; for seven of our ten datasets with at least 40 species, the magnitude of the overall pooled effect size from phylogenetic meta-analysis was significantly lower than that from traditional meta-analysis (Fig. 5; see Table A1 for phylogeny size data).

For random-effects models, phylogeny size was unrelated to variation in effect size change, although increased phylogenetic signal did lead to decreased effect size magnitudes after incorporating phylogenetic information (Fig. 5; Table 1). Accounting for phylogenetic information in random-effects analyses also led to decreased effect size magnitude in datasets with phylogenies in which the root node was more ancient (Fig. 5; Table 1).

Our metrics of tree shape were relatively unimportant for explaining variation in Hedges' d (all $IW < 0.25$), despite having model-averaged parameter estimates that were significantly different from zero (see Table 1). Thus, in response to incorporating phylogenetic information, effect size magnitudes from fixed-effects analyses declined weakly with increasingly unbalanced trees (large I_c), and those from random-effects analyses declined weakly in phylogenies with internal nodes that were

closer to the tips (large γ). Phylogenetic resolution did not explain variation in effect size change for either fixed- or random-effects analyses (i.e. it was absent as a predictor from all of our best-fit models).

DISCUSSION

By conducting traditional and phylogenetic meta-analyses across multiple datasets, and by assessing results from phylogenetic meta-analyses in relation to key characteristics of phylogenies, we provide the first empirical assessment of how this relatively new statistical method can affect meta-analytic inferences. Incorporating phylogeny often changed meta-analytic results, including quantitative changes to effect size estimates and whether those effect sizes were significantly different from zero. We found that the magnitude of effect size change following inclusion of phylogenetic data was strongly related to phylogenetic signal (as may be expected), phylogeny size (for fixed-effects models) and phylogeny age (for random-effects models). Our metrics of phylogeny shape (I_c and γ) also explained significant variation in effect size change, although neither metric was particularly important relative to our other predictors. Finally, we found that our

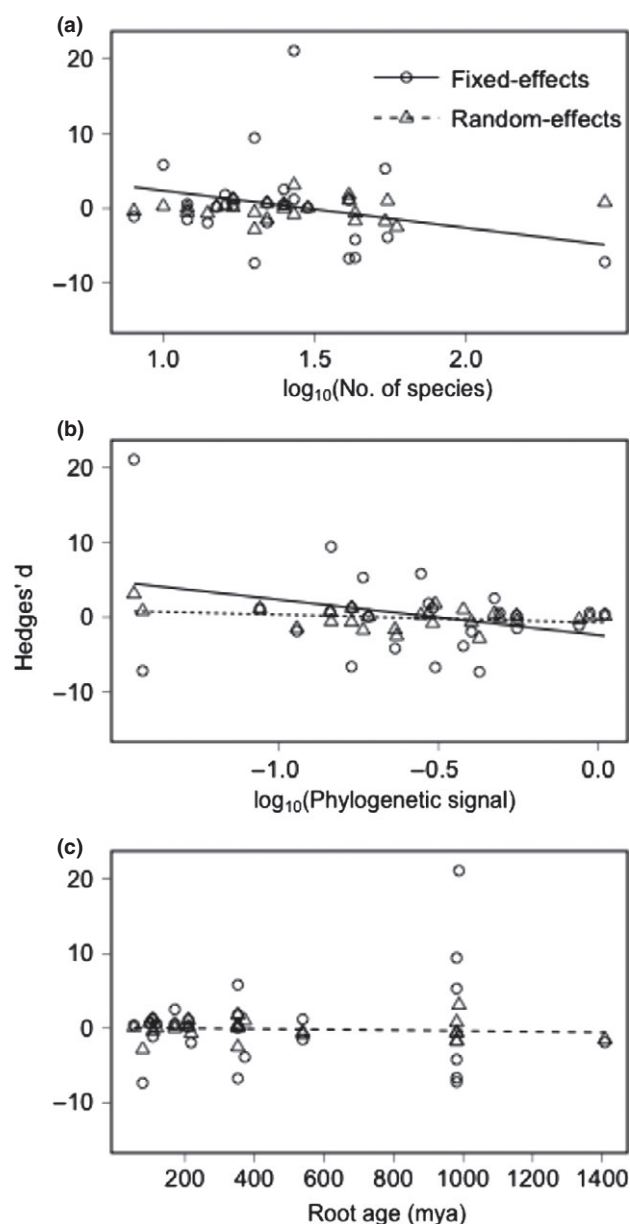


Figure 5 Explaining the magnitude of the difference between effect sizes from traditional and phylogenetic meta-analyses: Hedges' *d* in relation to (a) phylogeny size (number of species), (b) phylogenetic signal (the degree to which closer relatives have more similar trait values) and (c) phylogeny age (root age, mya). Results from fixed-effects models are depicted by circles (solid line shows best fit), and those from random-effects models are depicted by triangles (dashed line shows best fit). Note that these linear fits depict simple bivariate relationships without accounting for additional predictors that were included in our statistical models.

predictors explained little variation in effect size change for random-effects models relative to fixed-effects models; we discuss some implications of this distinction below.

How does incorporating phylogenies into meta-analysis affect overall pooled effect sizes?

For most individual datasets, incorporating phylogeny altered effect size estimates and whether those effect sizes were significantly different from zero (more so for fixed-effects than random-effects

Table 1 Model-averaging results for meta-meta-analysis of Hedges' *d*, which measures change in overall meta-analytic effect sizes from incorporating phylogenetic information

Model term	Imp. wt	Estimate (SE)	95% CI
Hedges' <i>d</i> (fixed-effects models)			
Intercept	1.00	8.04 (0.77)	(6.49, 9.58)
Number of species*	1.00	-6.34 (0.92)	(-8.18, -4.51)
Phylogenetic signal (K)*	0.43	-2.82 (0.37)	(-3.56, -2.07)
Tree balance (I_c)	<i>0.13</i>	<i>-0.55 (0.06)</i>	<i>(-0.68, -0.42)</i>
Distribution of node ages (γ)	—	—	—
Phylogeny age*	—	—	—
Hedges' <i>d</i> (random-effects models)			
Intercept	1.00	2.14 (1.33)	(-0.52, 4.80)
Number of species*	—	—	—
Phylogenetic signal (K)*	0.93	-2.01 (0.53)	(-3.08, -0.94)
Tree balance (I_c)	—	—	—
Distribution of node ages (γ)	<i>0.21</i>	<i>-0.15 (0.04)</i>	<i>(-0.23, -0.08)</i>
Phylogeny age*	0.76	-1.73 (0.19)	(-2.11, -1.35)

Shown here are the parameter importance weights, model-averaged parameter estimates (1 SE) and the 95% confidence interval (CI) around the model-averaged estimate. Parameter estimates significantly different from zero indicate that significant variation in the magnitude of effect size change relative to traditional meta-analysis is explained by a given predictor. Estimates greater than zero indicate an increase in the absolute value of effect sizes following incorporation of phylogeny; those less than zero indicate a decrease. Effects in bold are significant based on the 95% CI and considered important (Imp wt ≥ 0.25); those in italics are significant but not important (Imp wt < 0.25).

All analyses run with 26 datasets (PrE05, BP05, HE10 and PeE10 excluded).

* \log_{10} -transformed predictor variables.

models). The decision to use a phylogenetic vs. a traditional meta-analysis may therefore have crucial implications for the inferences and ultimate conclusions resulting from a meta-analytic investigation. Averaged across all of our datasets, incorporating phylogeny into traditional meta-analyses did not significantly alter effect sizes. The lack of such an average effect suggests there is no overall expected direction of effect size change when comparing phylogenetic with a traditional meta-analysis, a pattern also found when incorporating phylogenetic information into analyses based on other statistical techniques, e.g. regression (Rohlf 2006). We reiterate that there are clear and compelling statistical reasons to incorporate phylogenetic information into meta-analyses that synthesise information across multiple species. However, because incorporating phylogenetic information often affects meta-analytic inferences, and because this approach is being used with increasing regularity, it is critical to better understand the characteristics of trees and datasets most closely tied to changes in effect size.

When does conducting a phylogenetic meta-analysis result in large effect size changes?

Datasets with the largest phylogenies and the strongest phylogenetic signal showed decreases in overall effect size magnitude following the incorporation of phylogenetic information in fixed-effects analyses. The statistical assumption of independence among effect sizes is increasingly likely to be violated as closely related species are added to a given phylogeny. Of course, such a pattern will not arise if effect sizes are not phylogenetically conserved. However, the combination

of many related species and phylogenetic conservatism should both yield relatively large phylogenetic corrections, resulting in down-weighted phylogenetic effect sizes within groups of closely related species, as we observed. Interestingly, although the ability to detect significant phylogenetic signal increases with phylogeny size (Blomberg *et al.* 2003), across our datasets phylogeny size and phylogenetic signal were negatively correlated ($r = -0.54$, $P = 0.002$), suggesting that the effect of phylogeny size on meta-analytic outcomes was independent of phylogenetic signal.

For random-effects analyses, we found the same negative relationship between phylogenetic signal and phylogenetic effect size change as we did for fixed-effects analyses, although phylogeny size was no longer significant. Perhaps this reflects the positive correlation between the random-effects estimate (τ) and phylogeny size (Spearman's $\rho = 0.37$, $P = 0.063$, $n = 26$), which could have contributed to the relationship between phylogeny size and effect size change we observed from fixed-effects analyses. If this correlation is common across datasets, then one fortuitous outcome of using random-effects rather than fixed-effects analyses may be that underlying patterns in the data are more readily identified once the effect of phylogeny size is minimised.

For random-effects analyses, we also found that phylogenies for which the root node was more ancient had decreased effect-size magnitudes after incorporating phylogenetic information. Thus, parameter estimates for phylogeny age and signal were both negative, despite phylogenetic signal and phylogeny age being negatively correlated ($r = -0.61$, $P < 0.001$, $n = 30$). The correlation reflects a pattern whereby phylogenetic signal was strongest in phylogenies for which the root node was younger and which generally encompassed less phylogenetic breadth. The effects of phylogenetic signal and phylogeny age therefore appear to be independent; however, the underlying relationships between phylogeny age, phylogenetic breadth, phylogenetic signal and effect size change are likely to be complex.

We emphasise that although we have identified some intriguing relationships between meta-analysis outcomes and key characteristics of phylogenies, more work is needed. These meta-meta-analytic data are purely observational, and disentangling the independent effects of various phylogenetic characteristics on phylogenetic meta-analyses ultimately requires an experimental approach. This is particularly true for aspects of phylogenies, such as tree balance and the distribution of node ages, which were significantly related to Hedges' d despite being relatively unimportant across our datasets. Future simulation work will allow quantification of the relative importance of various phylogeny characteristics that are likely to be highly variable among meta-analysis datasets and also critically important for meta-analysis outcomes.

Random-effects vs. fixed-effects models for phylogenetic meta-analyses

In contrast to analyses based on fixed-effects models, our predictors explained relatively little variation in phylogenetic meta-analytic outcomes for random-effects models. In part, this reflects the fact that in our meta-meta-analyses, there was less variation in Hedges' d to be explained for random- vs. fixed-effects models. Random effects are incorporated into meta-analytic datasets as an increase in the within-study variance associated with each effect size (by the estimated between-study variance, τ). By adding this additional variance component, our results suggest that random-effects meta-analyses may have at least partially accounted for the increased variation

inherent in variance-covariance matrices from larger phylogenies. This suggests an unexpected potential benefit of using random-effects meta-analytic models within the phylogenetic context.

However, our findings also highlight a current statistical problem in meta-analysis: identifying the best way to calculate true random effects independent from incorporating phylogenetic information. The method we used (PhyloMeta v.1.2; Lajeunesse 2009, 2011) calculates τ from non-phylogenetically corrected data rather than first incorporating phylogenetic corrections, thereby assuming that estimates of τ are independent with respect to phylogeny. One consequence of this order of operations is that in some cases, the between study variance estimate (τ) may be inflated, not only accounting for random variation but also for variation that could otherwise be attributed to phylogenetic relationships. Ideally, a random-effects phylogenetic meta-analysis would incorporate phylogenetic information before calculating an estimate for τ . However, optimising the methodology for estimating τ in datasets with non-zero covariance (i.e. those accounting for pairwise phylogenetic distances) is a challenging issue facing statisticians (Riley *et al.* 2007; Jackson *et al.* 2010). Further developments in this field will greatly enhance our ability to conduct random-effects phylogenetic meta-analyses. In the meantime, we note that the assumption of independence in τ across species is one that may commonly be violated, and we recommend some degree of caution when interpreting random-effects phylogenetic meta-analyses using current methods.

CONCLUSION

Closely related species often share similar traits (Harvey & Purvis 1991) and occupy similar niches (cf. niche conservatism; Harvey & Pagel 1991). Despite these patterns, ecologists have rarely incorporated phylogenetic history into ecological meta-analyses either to account for non-independence due to shared ancestry or to test specific evolutionary hypotheses. Here, we have shown that incorporating phylogenies influences ecological meta-analysis outcomes, in many cases changing whether the observed effect size differs significantly from zero. We also show that the degree of difference between traditional and phylogenetic meta-analyses depends on key characteristics of phylogenies. Despite this potential complication, we strongly recommend incorporating phylogenetic information into ecological meta-analyses to account for species non-independence.

To conclude, we outline three recommendations for the use of phylogenetic meta-analyses in ecology and evolutionary biology:

- (1) *Use phylogenetic meta-analysis, but note that some response metrics are less likely to be affected by phylogenetic methods:* Incorporating phylogenetic relationships in meta-analysis addresses the non-independence of effect sizes from species with shared evolutionary history, thus solving a clear violation of statistical assumptions. However, phylogenetic corrections may have little effect on meta-analytic outcomes when effect sizes are not conserved and are therefore essentially independent. Conservation of effect sizes can be tested by determining if there is significant phylogenetic signal in the effect size.
- (2) *Include as many species as possible:* For a phylogenetic meta-analysis each data point represents an individual species, which can limit statistical power in cases where many effect sizes come from the same study species. Larger datasets ($c. > 20$ species) also permit greater statistical power to detect phylogenetic signal; a signif-

icant phylogenetic signal provides additional justification for conducting a phylogenetic meta-analysis. Although maximising sample size is always beneficial from the perspective of increasing statistical power, we suggest that in the context of phylogenetic meta-analyses, conducting a comprehensive data search is particularly critical. Thus, we caution against the use of search criteria that target only a few key journals or a limited number of publication years where phylogenetic meta-analyses are to be conducted.

- (3) *Be aware that phylogeny shape may influence meta-analytic outcomes.* As expected, phylogenetic signal and phylogeny size were the most important factors explaining how effect size magnitudes changed when incorporating phylogenetic information. Yet, despite being relatively unimportant in our analyses, both phylogeny balance (I_c) and the distribution of internal nodes between the root and the tips (γ) also influenced meta-analytic inferences. Planned simulation studies will allow us to better quantify the direct effects of phylogeny shape on meta-analytic outcomes. In the meantime, we recommend caution when conducting phylogenetic meta-analysis using highly unbalanced phylogenies and phylogenies with either very large or very small values of γ .

ACKNOWLEDGEMENTS

We thank Dean Adams, Carl Boettiger, Tom E.X. Miller, Shinichi Nakagawa, Craig Osenberg, Jennifer Rudgers, Jonathan Chase and two anonymous referees for thoughtful feedback on earlier versions of this manuscript. We also thank the authors of the 30 studies included in this analysis for providing their data online. We acknowledge support from NSF HRD-0450363, NSF DEB-0716868, the Ford Foundation Predoctoral Fellowship, NSF Graduate Research Fellowships and the Rice University Wray-Todd and Lodjeska Stockbridge Vaughn Fellowships.

AUTHORSHIP

All authors designed the study; SC, SH, CD, NR, BV, BM, JA, LBD, CR, MML, and JC collected meta-analysis datasets and compiled data; SC, SH, CD, NR, BV and BM reconstructed phylogenies; CD collected TimeTree data; SC and SH performed analyses, SC and SH wrote the manuscript; all authors contributed to revisions. SC and SH contributed equally.

REFERENCES

Abouheif, E. & Fairbairn, D.J. (1997). A comparative analysis of allometry for sexual size dimorphism: assessing Rensch's rule. *Am. Nat.*, 149, 540–562.

Adams, D.C. (2008). Phylogenetic meta-analysis. *Evolution*, 62, 567–572.

Aerts, R. (2006). The freezer defrosting: global warming and litter decomposition rates in cold biomes. *J. Ecol.*, 94, 713–724.

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

Blum, M.G.B. & François, O. (2006). Which random processes describe the tree of life? A large-scale study of phylogenetic tree imbalance. *Syst. Biol.*, 55, 685–691.

Burnham, K.P. & Anderson, D.R.. (2002). *Model Selection and Multimodel Inference: A Practical-Theoretic Approach*. Springer, Heidelberg.

Carmona, D., Lajeunesse, M.J. & Johnson, M.T.J. (2011). Plant traits that predict resistance to herbivores. *Funct. Ecol.*, 25, 358–367.

Colless, D.H. (1982). Review of: phylogenetics: the theory and practice of phylogenetic systematics. *Syst. Zool.*, 31, 100–104.

Côté, I.M. & Sutherland, W.J. (1997). The effectiveness of removing predators to protect bird populations. *Conserv. Biol.*, 11, 395–405.

Davies, K.F., Margules, C.R. & Lawrence, J.F. (2004). A synergistic effect puts rare, specialized species at greater risk of extinction. *Ecology*, 85, 265–271.

Davies, T.J., Allen, A.P., Borda-de-Água, L., Regetz, J. & Melián, C.J. (2011). Neutral biodiversity theory can explain the imbalance of phylogenetic trees but not the tempo of their diversification. *Evolution*, 65, 1841–1850.

DelBarco-Trillo, J. (2011). Adjustment of sperm allocation under high risk of sperm competition across taxa: a meta-analysis. *J. Evol. Biol.*, 24, 1706–1714.

Dubois, F. & Cezilly, F. (2002). Breeding success and mate retention in birds: a meta-analysis. *Behav. Ecol. Sociobiol.*, 52, 357–364.

de Faria, S.M., Lewis, G.P., Sprent, J.I. & Sutherland, J.M. (1989). Occurrence of nodulation in the Leguminosae. *New Phytol.*, 111, 607–619.

Freckleton, R.P., Harvey, P.H. & Pagel, M. (2002). Phylogenetic analysis and comparative data: a test and review of evidence. *Am. Nat.*, 160, 712–726.

Gurevitch, J., Morrow, L.L., Wallace, A. & Walsh, J.S. (1992). A meta-analysis of competition in field experiments. *Am. Nat.*, 140, 539–572.

Gurevitch, J., Curtis, P.S. & Jones, M.H.. (2001). Meta-analysis in ecology. *Advances in Ecological Research*, 32, 199–247.

Hackett, S.J., Kimball, R.T., Reddy, S., Bowie, R.C.K., Braun, E.L., Braun, M.J. *et al.* (2008). A phylogenomic study of birds reveals their evolutionary history. *Science*, 320, 1763–1768.

Hall, T.A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.*, 41, 95–98.

Harvey, P.H. & Pagel, M.D. (1991). *The Comparative Method in Evolutionary Biology*. Oxford University Press, Oxford.

Harvey, P.H. & Purvis, A. (1991). Comparative methods for explaining adaptations. *Nature*, 351, 619–624.

Hedges, L.V.. (1994). Fixed effects models. In: *The Handbook of Research Synthesis* (eds Cooper, H. & Hedges, L.V.). Russell Sage Foundation, New York, NY, pp. 285–299.

Hedges, L.V. & Olkin, I. (1985). *Statistical Methods for Meta-Analysis*. Academic Press, Orlando.

Hedges, S.B., Dudley, J. & Kumar, S. (2006). TimeTree: a public knowledge-base of divergence times among organisms. *Bioinformatics*, 22, 2971–2972.

Hoeksema, J.D., Chaudhary, V.B., Gehring, C.A., Johnson, N.C., Karst, J., Koide, R.T. *et al.* (2010). A meta-analysis of context-dependency in plant response to inoculation with mycorrhizal fungi. *Ecol. Lett.*, 13, 394–407.

Huberty, A.F. & Denno, R.F. (2004). Plant water stress and its consequences for herbivorous insects: a new synthesis. *Ecology*, 85, 1383–1398.

Huelsenbeck, J.P., Ronquist, F., Nielsen, R. & Bollback, J.P. (2001). Bayesian inference of phylogeny and its impact on evolutionary biology. *Science*, 294, 2310–2314.

Jackson, D., White, I.R. & Thompson, S.G. (2010). Extending DerSimonian and Laird's methodology to perform multivariate random effects meta-analyses. *Stat. Med.*, 29, 1282–1297.

Johnson, J. & Omland, K. (2004). Model selection in ecology and evolution. *Trends Ecol. Evol.*, 19, 101–108.

Lajeunesse, M.J. (2009). Meta-analysis and the comparative phylogenetic method. *Am. Nat.*, 174, 369–381.

Lajeunesse, M.J. (2011). phyloMeta: a program for phylogenetic comparative analyses with meta-analysis. *Bioinformatics*, 27, 2603–2604.

MacKenzie, B.R., Myers, R.A. & Bowen, K.G. (2003). Spawner-recruit relationships and fish stock carrying capacity in aquatic ecosystems. *Mar. Ecol. Prog. Ser.*, 248, 209–220.

Maddison, W.P. & Maddison, D.R.. (2010). *Mesquite: A Modular System for Evolutionary Analysis Version 2.73*. Available at: <http://mesquiteproject.org>. Last accessed 1 August 2011.

Maddison, D.R., Schulz, K.S. & Maddison, W.P. (2007). The tree of life web project. *Zootaxa*, 1668, 19–40.

Marczak, L.B., Thompson, R.M. & Richardson, J.S. (2007). Meta-analysis: trophic level, habitat, and productivity shape the food web effects of resource subsidies. *Ecology*, 88, 140–148.

Meunier, J., Pinto, S.F., Burri, R. & Roulin, A. (2011). Eumelanin-based coloration and fitness parameters in birds: a meta-analysis. *Behav. Ecol. Sociobiol.*, 65, 559–567.

- Morales, C.L. & Traveset, A. (2009). A meta-analysis of impacts of alien vs. native plants on pollinator visitation and reproductive success of co-flowering native plants. *Ecol. Lett.*, 12, 716–728.
- Munguía-Rosas, M.A., Ollerton, J., Parra-Tabla, V. & De-Nova, J.A. (2011). Meta-analysis of phenotypic selection on flowering phenology suggests that early flowering plants are favoured. *Ecol. Lett.*, 14, 511–521.
- Persson, J., Fink, P., Goto, A., Hood, J.M., Jonas, J. & Kato, S. (2010). To be or not to be what you eat: regulation of stoichiometric homeostasis among autotrophs and heterotrophs. *Oikos*, 119, 741–751.
- R Development Core Team (2011). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna.
- Rezende, E.L., Lavabre, J.E., Guimaraes, P.R., Jordano, P. & Bascompte, J. (2007). Non-random coextinctions in phylogenetically structured mutualistic networks. *Nature*, 448, 925–928.
- Riley, R.D., Abrams, K.R., Lambert, P.C., Sutton, A.J. & Thompson, J.R. (2007). An evaluation of bivariate random-effects meta-analysis for the joint synthesis of two correlated outcomes. *Stat. Med.*, 26, 78–97.
- Rohlf, F.J. (2006). A comment on phylogenetic correction. *Evolution*, 60, 1509–1515.
- Ronquist, F. & Huelsenbeck, J.P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, 19, 1572–1574.
- Shadish, W. & Haddock, C. (1994). Combining estimates of effect size. In: *The Handbook of Research Synthesis* (eds Cooper, H. & Hedges, L.V.). Russell Sage Foundation, New York, NY, pp. 261–281.
- Stone, E.A. (2011). Why the phylogenetic regression appears robust to tree misspecification. *Syst. Biol.*, 60, 245–260.
- Webb, C.O. & Donoghue, M.J. (2005). Phylomatic: tree assembly for applied phylogenetics. *Mol. Ecol. Notes*, 5, 181–183.
- Webb, C.O., Ackerly, D.D. & Kembel, S.W. (2008). Phylocom: software for the analysis of phylogenetic community structure and trait evolution. *Bioinformatics*, 24, 2098–2100.
- Wikström, N., Savolainen, V. & Chase, M.W. (2001). Evolution of the angiosperms: calibrating the family tree. *Proc. R. Soc. Lond. B Biol. Sci.*, 268, 2211–2220.

SUPPORTING INFORMATION

Additional Supporting Information may be downloaded via the online version of this article at Wiley Online Library (www.ecologyletters.com).

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organised for online delivery, but are not copy edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.

Editor, John Wiens

Manuscript received 14 November 2011

First decision made 27 December 2011

Manuscript accepted 13 March 2012