# Meta-Analysis and the Comparative Phylogenetic Method

**Marc J. Lajeunesse***

Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, New York 14853

*Submitted September 30, 2008; Accepted February 12, 2009; Electronically published July 28, 2009*
*Online enhancement:* appendix.

abstract: Meta-analysis has contributed substantially to shifting paradigms in ecology and has become the primary method for quantitatively synthesizing published research. However, an emerging challenge is the lack of a statistical protocol to synthesize studies and evaluate sources of bias while simultaneously accounting for phylogenetic nonindependence of taxa. Phylogenetic nonindependence arises from homology, the similarity of taxa due to shared ancestry, and treating related taxa as independent data violates assumptions of statistics. Given that an explicit goal of meta-analysis is to generalize research across a broad range of taxa, then phylogenetic nonindependence may threaten conclusions drawn from such reviews. Here I outline a statistical framework that integrates phylogenetic information into conventional meta-analysis when (*a*) taking a weighted average of effect sizes using fixed- and random-effects models and (*b*) testing for homogeneity of variances. I also outline how to test evolutionary hypotheses with meta-analysis by describing a method that evaluates phylogenetic conservatism and a model-selection framework that competes neutral and adaptive hypotheses to explain variation in meta-analytical data. Finally, I address several theoretical and practical issues relating to the application and availability of phylogenetic information for meta-analysis.

*Keywords:* Brownian motion, effect size, generalized least squares, Ornstein-Uhlenbeck process, phylogenetic conservatism, phylogenetic nonindependence.

## Introduction

Closely related taxa are more similar in morphology, physiology, behavior, and ecology than distantly related taxa (Harvey and Purvis 1991). This similarity resulting from shared phylogenetic history is a problem when analyzing data from a diversity of taxa because it violates two statistical assumptions. First, data are drawn from independent samples; phylogenetic history introduces a correlated structure to data because taxa form a nested hierarchy of phylogenetic relationships (Felsenstein 1985; Maddison 1990). Second, data are sampled from a population with a normal distribution with a common variance; sampling data with a phylogenetic structure can yield different variance structures because lineages within phylogenies may have evolved at different rates (Pagel 1992, 1999). Given that an explicit goal of recent meta-analyses on trade-offs (Koricheva et al. 2004), trophic cascades (Borer et al. 2005), and invasive biology (Parker et al. 2006) is to generalize and explain contingency in research across a broad range of taxa, then violating these assumptions may threaten the validity of conclusions drawn from such reviews.

Here I provide a general mathematical formulation of Adams's (2008) approach for integrating phylogenetic information into meta-analysis by unifying the statistics of meta-analysis and the comparative phylogenetic method. This unification is possible because both are special cases of the generalized least squares theory (Hedges and Olkin 1985; Cooper and Hedges 1994; Rohlf 2001). This more general formulation provides phylogenetically independent versions of all the most commonly used statistics in ecological meta-analysis, such as fixed- and random-effects models for pooling effect sizes and homogeneity tests (for a general text, see Hedges and Olkin 1985).

However, this general model for integrating phylogenetic information into meta-analysis also significantly extends the scope of quantitative reviews: now the evolutionary processes responsible for generating the diversity of experimental responses across taxa can be evaluated. For instance, phylogenetic conservatism or similarity due to shared ancestry is expected to obscure adaptive similarities resulting from convergent evolution (Schluter 2000). A phylogenetically independent meta-analysis has thus the conceptual advantage of distinguishing between evolutionary convergence of experimental responses among taxa and the alternative that responses are shared simply because of common ancestry.

To test evolutionary hypotheses with meta-analysis, I introduce diagnostics to evaluate phylogenetic conservatism and describe a model selection framework that contrasts neutral and adaptive hypotheses with meta-analytical

* Present address: National Evolutionary Synthesis Center, 2024 W. Main Street A200, Durham, North Carolina 27705; e-mail: marc.lajeunesse@nescent.org.

Am. Nat. 2009. Vol. 174, pp. 369–381. 
DOI: 10.1086/603628

data (Butler and King 2004). This evolutionary framework differs from the more common practice of using comparative analysis solely as a control for phylogenetic "pseudoreplication" (see Westoby et al. 1995*b*; Ricklefs and Starck 1996), because statistical models with and without phylogenetic information serve as competing hypotheses to explain variation in research outcomes. Reviewers need to think beyond simply phylogenetically correcting meta-analyses and should start explicitly recognizing that phylogenetic history may be an important explanatory variable for the diversity of experimental responses across taxa.

### *Previous Approaches to Phylogenetic Nonindependence*

Pooling studies from multiple taxa is meaningful only as long as the hierarchical relationship of taxa is recognized as a potential bias. One common approach to evaluate bias is to group studies by taxonomic rank (e.g., Linnean groups such as orders or families) and then treat these subsets as moderator variables in meta-analysis. Moderator variables are a necessary component of ecological meta-analysis because they serve as explanatory variables for heterogeneity in experimental outcomes (Hedges and Olkin 1985; Gurevitch and Hedges 1999). A similar method evaluates the sensitivity of meta-analysis to taxonomic bias by sequentially excluding groups of studies belonging to specific taxonomic ranks (e.g., Lajeunesse and Forbes 2002). Verdú and Traveset (2004, 2005) outline a more sophisticated approach to evaluating bias by contrasting the results from a conventional weighted regression model (a form of meta-analysis) to a range of results simulated using phylogenetic information (Garland et al. 1993). The function of these simulations is to provide phylogenetically independent critical values (e.g., *F*-tests) for hypothesis testing.

Although these approaches are useful to determine the presence of bias in meta-analysis, they do not explicitly use phylogenetic information to pool meta-analytical data; instead, phylogenetic history is used indirectly to assess the validity of traditional meta-analysis. Adams (2008) provided the first approach to directly integrate phylogenetic information into meta-analysis. Here the meta-analytical data are first converted into phylogenetically independent data using a generalized least squares (GLS) transformation method (see Garland and Ives 2000), and then a meta-analysis is performed using a second (weighted regression) GLS model. Adams's (2008) approach characterizes what is desired when pooling meta-analytical data; that is, (*a*) data are weighted by their relative sampling error (this weighting is fundamental to meta-analysis), and (*b*) data from closely related taxa are weighted less heavily in the overall analysis.

However, Adams's (2008) approach has limited application for meta-analysis because the statistical tools necessary to evaluate bias and test hypotheses (e.g., homogeneity tests) are missing. In addition, the GLS phylogenetic transformation method has an effect of converting meta-analytical data into evolutionary units. This change in units makes the comparison with pooled results from traditional meta-analysis impossible (Butler and King 2004) and also reduces the effectiveness of the weights used to penalize studies during meta-analysis (for further details, see appendix in the online edition of the *American Naturalist*). However, what is gleaned from Adams's phylogenetic meta-analysis is that the statistics of meta-analysis and the comparative method share a common GLS framework. Below, I outline an alternative mathematical formulation of Adams's (2008) method that unifies the weighting schemes of meta-analysis and the comparative method under a single GLS model. This approach offers all the familiar meta-analytical statistics necessary to perform a rigorous quantitative review (e.g., fixed- and random-effects models for pooling results, assessing heterogeneity, etc.).

### *Nonindependence and Expected Variance*

Synthesizing experiments based on taxa with a shared phylogenetic history violates the assumption of independence because data have a correlated structure: taxa form a nested hierarchy of phylogenetic relationships such that their traits and characteristics do not have an independent origin. In meta-analysis, the units of analysis are effect sizes—a statistical measure of the magnitude and direction of experimental outcomes—and examples can be envisioned where effect sizes are phylogenetically correlated. Consider an effect size that quantifies the magnitude of an experimental outcome using a control and treatment mean (i.e., Hedges's $d$ or $\ln \mathrm{RR}$; for examples, see Van Zandt and Mopper 1997). Here the means are derived from species traits, and these traits may be phylogenetically conserved. For example, body size is often used as a surrogate for fitness, and for many animals, body size is phylogenetically conserved such that closely related species or whole lineages tend to share similar sizes. The second bias can occur if the effect sizes themselves are phylogenetically conserved, such as phenotypic responses to multiple environments or among taxa with sexual dimorphism. It is known, for example, that the magnitude and direction of size dimorphism (females > males, or females < males) is dependent on the mating system of a lineage (e.g., polygamy or monogamy; Björklund 1997). These issues are further exacerbated should mating system itself be a homologous trait. These origins of phylogenetic correlations may not be independent but would generate similar bias for

meta-analysis; studies using related species for experimentation may yield similar study outcomes (effect sizes).

The second statistical assumption invalidated by data with a phylogenetic structure is that sampling occurs from a normally distributed population with an expected variance. Sampling from a phylogeny can generate data with different variances because lineages within a phylogeny may evolve at different rates or may have had more time to evolve (Harvey and Pagel 1991). Variance here is defined as the rate of evolutionary change within a phylogeny, and comparative analyses typically make assumptions about how lineages evolve to meet the statistical assumption of an expected variance (hereafter evolutionary variance). For instance, the widely used Felsenstein (1985) phylogenetically independent contrasts method assumes that evolution proceeds as a Brownian motion (BM) process (e.g., random drift) and uses information on phylogenetic branch lengths to calculate the expected evolutionary variance of change. This is because BM evolution predicts that long branches can accumulate more change and that evolutionary rates are the same throughout the phylogeny (and thus all taxa have the same expected variance of change; see Felsenstein 1985). Thus, to satisfy both assumptions of independence and common evolutionary variance, Felsenstein's approach transforms raw data into a set of contrasts that have zero (phylogenetic) covariance and standardizes these contrasts to have equal (evolutionary rate) variance (using the square root of the sum of all phylogenetic branch lengths as the expected variance of change).

Although BM forms the basis for nearly all phylogenetic comparative statistics (Cheverud et al. 1985; Felsenstein 1985; Maddison 1990; Martins and Garland 1991; Pagel 1997), it is a model that oversimplifies the process of evolution. For instance, BM assumes that character change is independent within each lineage and that the character variance among lineages will increase with time (Martins 1994). Yet selection is an important force in character change among taxa, and a more complete evolutionary model should account for this force. Hansen (1997) proposed an alternative stochastic model by means of an Ornstein-Uhlenbeck (OU) process, which assumes that natural selection also contributes to character change. Because OU explicitly models selection as a force for change within a phylogeny, then under some parameters of selection (i.e., stabilizing selection) it is expected that the evolutionary variance in traits will remain bounded and constant through time. This effect can meet the statistical assumption of homogeneity of evolutionary variances (Hansen 1997). Later, I describe how these different models of evolution serve as competing neutral (BM) and adaptive (OU) hypotheses for explaining evolutionary variation in research outcomes across taxa.

### *A Primer on Meta-Analysis and the Comparative Method*

One purpose of meta-analysis is to statistically weight study outcomes by their inverse sampling variance to control for within-study sampling error (see Lajeunesse and Forbes 2003). This downweighting is important because studies with large sampling variances or small sample sizes are sensitive to sampling error—perhaps under- or overestimating effect sizes. The comparative phylogenetic method, however, uses phylogenetic information during the regression of traits to control for shared ancestry of taxa (Felsenstein 1985) and to test explicit models of evolution (Pagel 1999). Again, a weighting scheme is used to penalize data; taxa stemming from short branches on phylogenies have their data downweighted because they may not represent independent pieces of information (e.g., not enough time for derived characteristics to change). To assist with the following sections, a roundup of the various terms and symbols used is shown in table 1.

What unites meta-analysis and the comparative phylogenetic method is that they are both special cases of the generalized theory of least squares (Adams 2008). Statistics based on ordinary least squares (OLS), such as regression and ANOVA, have several assumptions (for a general text, see Groß 2003), but the two of interest here are that effect sizes (the independent variable) share a common sampling variance (are homoscedastic) and are uncorrelated (statistically independent). A way of illustrating these assumptions using matrix notation is

$$\mathbf{E} \sim N(\mathbf{X}\bar{\delta}, \sigma^2\mathbf{I}), \tag{1}$$

where $\mathbf{E}$ is a $k \times 1$ column vector of $k$ number of effect sizes ($\delta$), which are assumed to be normally distributed ($N$) with an expected mean of $\mathbf{X}\bar{\delta}$ and sampling variance of $\sigma^2\mathbf{I}$. The expected mean ($\mathbf{X}\bar{\delta}$) of $\mathbf{E}$ designates the averaging behavior of effect sizes. How effect sizes are averaged is defined by the design matrix $\mathbf{X}$. Typically for meta-analysis, a pooled (average) effect size ($\bar{\delta}$) is generated by codifying $\mathbf{X}$ as a $k \times 1$ column vector of 1s. The sampling variance of effect sizes ($\sigma^2\mathbf{I}$) is known as the scalar variance-covariance matrix (i.e., $\mathrm{Var}[\mathbf{E}] = \sigma^2\mathbf{I}$), and this matrix defines how effect sizes are correlated. The identity matrix $\mathbf{I}$ indicates that the observed variances for each effect size are uncorrelated and share a common sampling variance ($\sigma^2$).

The method of generalized least squares is a statistical framework that directly addresses violations due to nonindependence and heteroscedasticity of data or, more precisely, instances where $\mathrm{Var}(\mathbf{E}) \neq \sigma^2\mathbf{I}$. These violations are explicitly modeled in a $k \times k$ covariance matrix ($\mathbf{\Sigma}$) such that the expected distribution of effect sizes is now

**Table 1:** A roundup of variables used in evolutionary meta-analysis

| Symbol | Definition |
|---|---|
| Effect size parameters: | |
| $\delta$ | Effect size (magnitude and direction of research outcome) |
| $\sigma^2(\delta)$ | Effect size variance |
| Meta-analysis: | |
| $k$ | Sample size of meta-analysis (number of effect sizes) |
| $\bar{\mu}$ | Raw unweighted pooled effect size |
| $\bar{\delta}$ | Weighted pooled effect size |
| $\mathbf{X}$ | The design matrix defined as $k \times 1$ modeled to take an average (column vector of 1s) or to hypothesize adaptive optima |
| $\mathbf{E}$ | Column vector of effect sizes ($\delta$) with a $k \times 1$ dimension |
| $Q$ | Homogeneity test of effect sizes |
| Comparative analysis: | |
| $\lambda$ | Degree of a phylogenetic signal |
| $b$ | Branch length on a phylogenetic tree |
| $\mathbf{P}$ | Phylogenetic correlation matrix of $k \times k$ dimensions containing all the shared $b$ between species within a phylogeny |
| BM | Brownian motion model of evolution for modeling the $b$ of $\mathbf{P}$ |
| OU | Ornstein-Uhlenbeck process used to model $\mathbf{P}$ |
| $\eta$ | Evolutionary parameter depicting the strength of drift in OU |
| $\beta$ | Evolutionary parameter depicting the strength of selection in OU |
| $\theta$ | Evolutionary optimum that designates the period under which selection has occurred |
| $m$ | Number of hypothesized evolutionary optima ($\theta$) |
| Evolutionary meta-analysis: | |
| $\boldsymbol{\Sigma}$ | Covariance matrix of $k \times k$ dimensions used to account for study sampling error and phylogenetic nonindependence |
| $\bar{\delta}^{\mathbf{P}}$ | Weighted and phylogenetically independent pooled effect size across $k$ studies |
| $Q_R^{\mathbf{P}}$ | Phylogenetically independent $\chi^2$ statistic testing $\bar{\delta}^{\mathbf{P}} \neq 0$ |
| $Q_H^{\mathbf{P}}$ | Phylogenetically independent homogeneity test for $\delta_1 = \delta_2 = \ldots = \delta_k$ |
| $\mathbf{W}$ | Column vector containing the percent weights of $m$ number of taxonomic classes |
| AIC | Akaike's Information Criterion used for selecting the best among competing evolutionary models |

Note: Bold type and nonitalicized symbols are column vectors or matrices, and transposed matrices are denoted with superscript T and inverse matrices with an exponent of $-1$.

$$\mathbf{E} \sim N(\mathbf{X}\bar{\delta}, \boldsymbol{\Sigma}). \quad (2)$$

For instance, traditional meta-analysis has a $\boldsymbol{\Sigma}$ matrix containing the sampling variances of each effect size $\sigma^2(\delta)$ on its main diagonal—modeling a weighted least squares regression where effect sizes with large variances are penalized during the pooling of effect sizes. This codification of $\boldsymbol{\Sigma}$ differs from the comparative method, which uses all off-diagonal elements (covariances) of $\boldsymbol{\Sigma}$ to account for the correlated evolution history of taxa, thereby giving less weight to taxa that are more closely related to other taxa when fitting a regression line through their data (see Pagel 1997, 1999). Thus, the elements of $\boldsymbol{\Sigma}$ can be formulated to serve the interests of both meta-analysis (weighting by sampling error) and the comparative method (weighting by relatedness).

## Statistical Framework

The following framework is divided into three sections. I first describe how to define the elements of the $\boldsymbol{\Sigma}$ covariance matrix to account for phylogenetic nonindependence and to study heteroscedasticity. I then apply this matrix to a GLS framework to calculate a phylogenetically independent meta-analysis of effect sizes. The final section extends this GLS framework to test neutral and adaptive evolutionary hypotheses with meta-analysis. An illustrative example of the use and interpretation of these methods is found in the appendix. Note that all the outlined methods can be applied to any metric of effect sizes that has a known variance. For examples of different effect sizes, see Van Zandt and Mopper (1997) for Hedges's $d$ and ln RR, Koricheva et al. (2004) for correlation coefficients, and Beirinckx et al. (2006) for log (OR).

### *Covariance Matrix and Bias*

Here I model both heteroscedasticity and phylogenetic correlations in a single covariance matrix ($\mathbf{\Sigma}$). Heteroscedasticity is first modeled on the main diagonal of $\mathbf{\Sigma}$, which contains the effect size variances $\sigma^2(\delta_i)$ for each $i$th effect size ($\delta_i$) of $k$ species. This modeling is the same as in traditional meta-analysis, where effect sizes with large sampling variances are weighted less heavily because they may represent inaccurate estimates of the "true" population effect size (Hedges and Olkin 1985). The second bias of phylogenetic nonindependence is modeled on all the off-diagonals of $\mathbf{\Sigma}$, which contain the between-study covariances (Cov). Here, the covariances measure how effect sizes vary together based on correlated phylogenetic history (as described in the $\mathbf{P}$ correlation matrix) and are calculated for each pair of effect sizes:

$$\text{Cov}(\delta_i, \delta_j) = \mathbf{P}_{i,j}\sqrt{\sigma^2(\delta_i)}\sqrt{\sigma^2(\delta_j)}. \quad (3)$$

The following is a more formal description of the symmetric covariance matrix $\mathbf{\Sigma}$ with elements in its $i = 1, \ldots, k$ rows and $j = 1, \ldots, k$ columns:

$$\mathbf{\Sigma}_{i,j} = \begin{bmatrix} \sigma^2(\delta_1) & \cdots & \text{Cov}(\delta_1, \delta_{j-1}) & \text{Cov}(\delta_1, \delta_j) \\ & \ddots & \vdots & \vdots \\ & & \sigma^2(\delta_{i-1}) & \text{Cov}(\delta_{i-1}, \delta_j) \\ & & & \sigma^2(\delta_i) \end{bmatrix}. \quad (4)$$

Taking the inverse of $\mathbf{\Sigma}$ yields a weight matrix (Groß 2003), but for simplicity I will continue to refer to $\mathbf{\Sigma}$ as the covariance matrix. Further details on the statistical background of equation (3) and $\mathbf{\Sigma}$ and how this approach relates to Adams's (2008) phylogenetic meta-analysis are found in the appendix.

### *Correlation Matrix and Phylogenetic History*

The $\mathbf{P}$ matrix contains the correlations among effect sizes due to the shared phylogenetic history of taxa. The strengths of these correlations are often defined as the phylogenetic branch length ($b$) distance between taxa—where, for example, the total branch lengths for each species are on the main diagonal, and the shared distance between species are all off-diagonals of the matrix. Defining the $\mathbf{P}$ matrix this way assumes the purely neutral BM model of evolution and is specifically coded as

$$\mathbf{P}^{\text{BM}}_{i,j} = \begin{cases} b_i^{\text{total}} & \text{if } i = j \\ b_{i,j}^{\text{shared}} & \text{if } i \neq j \end{cases}. \quad (5)$$

Here $b$ denotes the phylogenetic branch length, where the total distance from the root to tip is $b^{\text{total}}$ and $b^{\text{shared}}$ is the branch length that taxon $i$ shares with taxon $j$. Rohlf (2001) provides an example of the $\mathbf{P}^{\text{BM}}$ matrix. Comparative methods based on GLS often treat this distance matrix as the covariance matrix (e.g., $\mathbf{\Sigma} = \mathbf{P}^{\text{BM}}$; see Rohlf 2001). This matrix also forms the basis for Adams's (2008) approach to including phylogenetic information into meta-analysis.

Another useful correlation matrix described by Hansen and Martins (1996) and Hansen (1997) assumes an Ornstein-Uhlenbeck process (OU; see also Gardiner 1985). Unlike the BM model, which predicts a negative linear relationship between phylogenetic correlations and divergence time, OU assumes that closely related clades are exponentially more similar to one another than to more distantly related taxa (see fig. 1). The rate of this exponential change increases in lineages undergoing strong selection, with selection acting to erase ancestral or derived constraints (Hansen 1997). Under the OU model, $\mathbf{P}$ is defined as

$$\mathbf{P}^{\text{OU}}_{i,j} = \begin{cases} \left(1 - e^{-2\beta b^{\text{total}}}\right)/2\beta & \text{if } i = j \\ \left(e^{-2\beta(b^{\text{total}} - b_{i,j}^{\text{shared}})} - e^{-2\beta b^{\text{total}}}\right)/2\beta & \text{if } i \neq j \end{cases}. \quad (6)$$

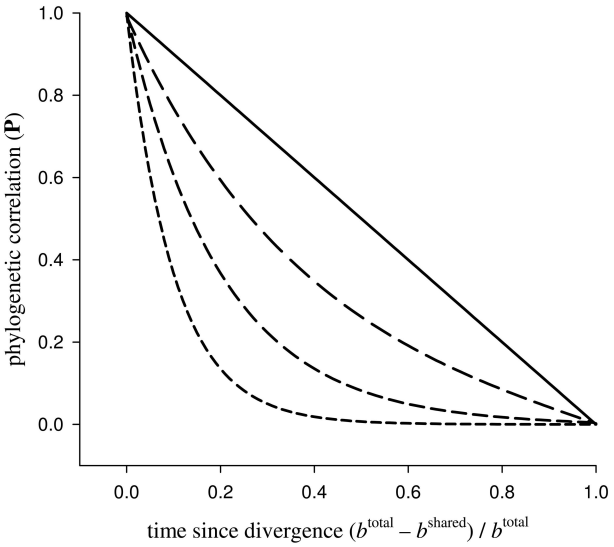



**Figure 1:** The predicted relationship between similarity among taxa (quantified as a phylogenetic correlation) and their relative time since divergence for Brownian motion (*solid line*) and Ornstein-Uhlenbeck (OU; *dashed lines*) models of evolution. The multiple dashed lines show how under an OU model the phylogenetic correlations decrease exponentially with time when lineages undergo weak to moderate to strong selection. The dashed line with the most pronounced curve indicates a model with the strongest selection. Because Brownian motion evolution is nested within OU, the solid line also depicts OU when selection is zero ($\beta = 0$).

Here selection ($\beta$) is used to model phylogenetic correlations and can range from 0 to infinity. An important feature of $\mathbf{P}^{OU}$ is that it converges to $\mathbf{P}^{BM}$ when the strength of selection nears 0 ($\beta \to 0$). Later I discuss how to estimate selection and how the nestedness of $\mathbf{P}^{BM}$ in $\mathbf{P}^{OU}$ allows the sequential hypothesis testing of neutral and adaptive evolutionary models (sensu Butler and King 2004).

There are two issues that should be considered before $\mathbf{P}^{BM}$ and $\mathbf{P}^{OU}$ are used to calculate effect size covariances. First, the covariance equation (3) assumes that all the diagonal correlations of $\mathbf{P}$ are equal and that all nondiagonals do not equal or exceed these diagonal correlations (Hedges and Olkin 1985). This assumption controls for computational issues during matrix inversion (e.g., the matrix must be positive definite; Groß 2003). Second, BM and OU assume a relationship with time (fig. 1). To satisfy these assumptions, only ultrametric trees (e.g., chronograms where tips of the phylogeny are aligned or contemporaneous) should be used to code the phylogenetic correlation matrix $\mathbf{P}$. The timescale of the ultrametric tree does not have to be absolute but should at least contain information on the relative divergence time between taxa. Thus, it is necessary to standardize the $\mathbf{P}$ matrix with

$$\mathbf{P}_{i,j} = \begin{cases} 1 & \text{if } i = j \\ 2b_{i,j}^{\text{shared}}(b_i^{\text{total}} + b_j^{\text{total}})^{-1} & \text{if } i \neq j \end{cases} \quad (7)$$

Equation (7) is equivalent to standardizing (dividing) the elements of the $\mathbf{P}$ matrix by $b^{\text{total}}$ (see Pagel 1994) but has an added control to standardize $\mathbf{P}$ to fit statistical assumptions even when the phylogenetic tree is not ultrametric. Finally, standardizing the $\mathbf{P}$ correlation matrix to these values has the practical effect of allowing for the direct comparison between the pooled effect size with and without phylogenetic correlations.

### *Phylogenetically Independent Mean Effect Sizes and Variances*

Now that the covariance matrix $\mathbf{\Sigma}$ is modeled to account for heteroscedacity and phylogenetic correlations (either $\mathbf{P}^{BM}$ or $\mathbf{P}^{OU}$), we can estimate a phylogenetically independent weighted mean effect size ($\bar{\delta}^{\mathbf{P}}$) with the standard GLS regression equation:

$$\bar{\delta}^{\mathbf{P}} = (\mathbf{X}^{\mathrm{T}}\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{\Sigma}^{-1}\mathbf{E}, \quad (8)$$

which has a variance of

$$\sigma^2(\bar{\delta}^{\mathbf{P}}) = (\mathbf{X}^{\mathrm{T}}\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}. \quad (9)$$

Here, the design matrix $\mathbf{X}$ is a $k \times 1$ column vector of 1s, and $\mathbf{E}$ is a $k \times 1$ column vector $k$ number of effect sizes ($\delta$). To evaluate whether $\bar{\delta}^{\mathbf{P}}$ is nonzero, 95% confidence intervals (CIs) should be used and are calculated as follows:

$$95\%\ \text{CI}\left[\bar{\delta}^{\mathbf{P}} - 1.96\sqrt{\sigma^2(\bar{\delta}^{\mathbf{P}})};\ \bar{\delta}^{\mathbf{P}} + 1.96\sqrt{\sigma^2(\bar{\delta}^{\mathbf{P}})}\right]. \quad (10)$$

An additional advantage of 95% CIs is that they provide information on the statistical power of meta-analysis: broad 95% CIs indicate a poor ability to detect a nonzero $\bar{\delta}^{\mathbf{P}}$ should it exist (see Nakagawa and Cuthill 2007). A more direct test for whether $\bar{\delta}^{\mathbf{P}} \neq 0$ uses the following regression test statistic:

$$Q_{\mathrm{R}}^{\mathbf{P}} = (\bar{\delta}^{\mathbf{P}})^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{\Sigma}^{-1}\mathbf{X}\bar{\delta}^{\mathbf{P}}. \quad (11)$$

This test is based on a $\chi^2$ distribution with 1 degree of freedom, such that if $Q_{\mathrm{R}}^{\mathbf{P}} > \chi^2_{\mathrm{df}=1}$, then there is not enough evidence to indicate that $\bar{\delta}^{\mathbf{P}}$ differs from 0 (Hedges 1992).

Reductions to the main diagonal and off-diagonal elements of $\mathbf{\Sigma}$ will also yield other important mean effect sizes (for additional details, see appendix). For instance, reducing all off-diagonal elements of $\mathbf{\Sigma}$ to 0, such that $\mathbf{\Sigma} = \text{diag}[\sigma^2(\delta_1), \ldots, \sigma^2(\delta_k)]$, will treat effect sizes as independent and thus generate the traditional weighted mean effect size ($\bar{\delta}$) when applied to equation (8). A second reduction of the main diagonal elements to 1s will result in $\mathbf{\Sigma} = \mathbf{I}$. Applying $\mathbf{I}$ to equation (8) will yield a simple arithmetic mean of effect sizes ($\bar{\mu}$).

### *Homogeneity Test for Phylogenetically Independent Effect Sizes*

Homogeneity statistics ($Q_{\mathrm{H}}$) determine whether within-study sampling error can explain the observed variation among a collection of effect sizes (Hedges and Olkin 1985). If sampling error is the primary source of variation (e.g., nonsignificant $Q_{\mathrm{H}}$), then $\delta_1 = \delta_2 = \ldots = \delta_k$ and effect sizes can be pooled; however, a significant $Q_{\mathrm{H}}$ indicates that further exploration of the data is needed to explain heterogeneity among effect sizes. The appendix describes how to use a moderator variable to explore variation and how to use a phylogenetically independent random-effects model when effect sizes continue to fail homogeneity tests. Calculation of a homogeneity test ($Q_{\mathrm{H}}^{\mathbf{P}}$) for phylogenetically independent effect sizes is as follows:

$$Q_{\mathrm{H}}^{\mathbf{P}} = \mathbf{E}^{\mathrm{T}}\mathbf{\Sigma}^{-1}\mathbf{E} - Q_{\mathrm{R}}^{\mathbf{P}}, \quad (12)$$

where if $Q_{\mathrm{H}}^{\mathbf{P}} \leq \chi^2_{\mathrm{df}=k-1}$ then the observed variation is due to sampling error.

### *A Diagnostic for Taxonomic Bias*

A common heuristic used to assess taxonomic bias when pooling effect sizes is to calculate the percentage of studies based on a specific taxonomic group. For example, 70% of $k$ studies were based on invertebrates. The caveat of this example is that other taxonomic groups such as vertebrates are underrepresented in the literature, and thus, the pooled effect size will favor the outcomes of invertebrate research. However, given that equation (8) has information on the weights due to sampling precision and phylogenetic correlations (found in weight matrix $\mathbf{\Sigma}^{-1}$), then partitioning these weights among taxonomic ranks will give more information on the actual contribution of these classes during the meta-analysis of effect sizes.

Estimating the overall contribution (quantified as the percentage of the overall weight on the pooled effect size) for $m$ number of taxonomic ranks (e.g., order or family) is as follows:

$$\mathbf{W} = \mathbf{X}^{\mathrm{T}}\mathbf{\Sigma}^{-1}\mathbf{Z}(\mathbf{Z}^{\mathrm{T}}\mathbf{\Sigma}^{-1}\mathbf{Z})^{-1}, \tag{13}$$

where $\mathbf{W}$ is the $m \times 1$ column vector containing the percentage weight of each rank (where the sum of its elements will equal 100%) and $\mathbf{Z}$ is a $k \times 1$ column vector containing 0.01. The $\mathbf{Z}$ matrix serves the purpose for calculating percentages. Finally, $\mathbf{X}$ is the design matrix with $m$ number of ranks, where each rank is a column containing 1 if the taxa belongs to that rank and 0 otherwise. For example, if a meta-analysis has three taxa and two belong to the same genus, then the $k \times m$ design matrix (here $3 \times 2$) is

$$\mathbf{X} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}^{\mathrm{T}}. \tag{14}$$

Much like when pooling effect sizes with equation (8), different percentages can be estimated based on the design of the covariance matrix $\mathbf{\Sigma}$. Assuming that $\mathbf{\Sigma} = \mathbf{I}$ in equation (13) will yield the weight of each taxonomic class if each taxon (effect size) is evenly weighted (e.g., not weighted by sample precision or phylogenetic correlations). Following the example above, the raw percentage weights for the two genera are $\mathbf{W}^{\mathbf{I}} = [66.7, 33.3]^{\mathrm{T}}$. Assuming that $\mathbf{\Sigma} = \mathrm{diag}[\sigma^2(\delta_1), \ldots, \sigma^2(\delta_k)]$, as in traditional meta-analysis, will yield the overall weighting of each rank relative to their sample precision, whereas assuming that $\mathbf{\Sigma} = \mathbf{\Sigma}^{\mathrm{BM}}$ will further adjust these weights based on their phylogenetic correlations. Interpreting the percentages in $\mathbf{W}$ is straightforward: a high percentage indicates a greater weight of the effect sizes in that taxonomic class when pooling effect sizes.

### *Detecting a Phylogenetic Signal among Effect Sizes*

Testing for a phylogenetic signal is good practice in comparative analysis (Freckleton et al. 2002) and should be part of any meta-analysis that includes information on phylogenetic history. Estimating a phylogenetic signal is a test of phylogenetic conservatism (also known as phylogenetic inertia) and determines the degree to which related taxa tend to be more similar than distantly related species. Knowing the degree to which data are phylogenetically conserved allows for a more informative interpretation of phylogenetically independent results (Björklund 1997): strong phylogenetic signal may indicate strong bias due to phylogenetic nonindependence (e.g., strong phylogenetic correlations in $\mathbf{P}$). When traits show little phylogenetic conservatism, they are considered evolutionarily labile and appear distributed randomly among the tips of a phylogeny.

Pagel (1994) formally defines phylogenetic conservatism as the degree to which data fit the BM model of evolution: that the phylogenetic correlations among effect sizes are linearly related with time of divergence between the taxa for which they are based (see fig. 1). As $\lambda$ nears 1 ($\lambda \rightarrow 1$), then effect sizes are distributed phylogenetically as expected by BM, whereas when $\lambda \rightarrow 0$, then data are randomly distributed and uncorrelated (e.g., no conservatism). Here, $\lambda$ is treated as a scaling factor that transforms the correlations among taxa from having no signal (no correlation) to having a full signal (full correlations as predicted by $\mathbf{P}^{\mathrm{BM}}$):

$$\mathbf{P}_{i,j}^{\mathrm{BM}(\lambda)} = \begin{cases} b_i^{\mathrm{total}} & \text{if } i = j, \\ \lambda b_{i,j}^{\mathrm{shared}} & \text{if } i \neq j. \end{cases} \tag{15}$$

Approaches to evaluate the contribution of $\lambda$ can use either a manual or a maximum likelihood (ML) approach that optimizes $\lambda$ to effect sizes (Pagel 1994, 1997). The manual approach adopts a "T-shirt" philosophy, where a range of small (no correlation or $\lambda = 0$), medium ($\lambda = 0.5$), and large (full correlation or $\lambda = 1$) values of $\lambda$ are plugged into $\mathbf{P}^{\mathrm{BM}}$. These transformed phylogenetic correlation matrices ($\mathbf{P}^{\mathrm{BM}(\lambda)}$) are then applied to meta-analysis (e.g., starting with eq. [8]). For instance, using $\mathbf{P}^{\mathrm{BM}(\lambda=0)}$ to calculate a pooled effect size will generate the traditional weighted (nonphylogenetically corrected) mean effect size ($\bar{\delta}$).

Alternatively, $\lambda$ can be optimized via ML. This method finds the best $\lambda$ that minimizes the residual sums of squares (SSE) of the model and thus can provide more information on whether the phylogenetic conservatism of effect sizes is closer 0 or 1. Under assumptions of normality among effect sizes ($\mathbf{E}$), the least squares likelihood ($L$) of $\lambda$ is

$$L[\lambda|\mathbf{E}] = \frac{e^{\mathrm{SSE}[-2\sigma^2(\bar{\mu}_{\mathrm{BM}(\lambda)})]^{-1}}}{\sqrt{[2\pi\sigma^2(\bar{\mu}_{\mathrm{BM}(\lambda)})]^k \det(\mathbf{P}^{\mathrm{BM}(\lambda)})}}, \quad (16)$$

where

$$\bar{\mu}_{\mathrm{BM}(\lambda)} = [\mathbf{X}^{\mathrm{T}}(\mathbf{P}^{\mathrm{BM}(\lambda)})^{-1}\mathbf{X}]^{-1}\mathbf{X}^{\mathrm{T}}(\mathbf{P}^{\mathrm{BM}(\lambda)})^{-1}\mathbf{E}, \quad (17)$$

$$\mathrm{SSE} = (\mathbf{E} - \mathbf{X}\,\bar{\mu}_{\mathrm{BM}(\lambda)})^{\mathrm{T}}(\mathbf{P}^{\mathrm{BM}(\lambda)})^{-1}(\mathbf{E} - \mathbf{X}\,\bar{\mu}_{\mathrm{BM}(\lambda)}), \quad (18)$$

and $\sigma^2(\bar{\mu}_{\mathrm{BM}(\lambda)}) = \mathrm{SSE}(k-1)^{-1}$. Finally, $\bar{\mu}_{\mathrm{BM}(\lambda)}$ is the pooled (unweighted) effect size with a variance of $\sigma^2(\bar{\mu}_{\mathrm{BM}(\lambda)})$ under the BM model. The ML estimate of $\lambda$ (hereafter $\hat{\lambda}$) is found by linear optimization, where, for instance, equations (16)–(18) are calculated for values of $\lambda$ ranging from 0.0001 to 1 and where the smallest value of $L$ (with the least amount of error) is chosen for $\hat{\lambda}$. Once $\hat{\lambda}$ is found, it is then applied to $\mathbf{P}^{\mathrm{BM}(\hat{\lambda})}$. Irrespective of whether the manual or ML approach is taken, all of the above effect size statistics (e.g., pooled effect size) will be fitted conditionally on $\lambda$ or $\hat{\lambda}$.

### *Estimating a Signal of Selection and Drift*

I previously outlined how to model the phylogenetic correlation matrix **P** using an Ornstein-Uhlenbeck process. I now extend this evolutionary model into a more formal framework to evaluate the contribution of selection ($\beta$) and drift ($\eta$) in explaining evolutionary variation in research outcomes. These evolutionary parameters are useful for meta-analysis because they will provide the basis for testing neutral and adaptive hypotheses. I avoid theory when possible and present only the equations necessary to calculate $\beta$; for further information on background and derivation, see work by Hansen (1997) and Butler and King (2004).

The Ornstein-Uhlenbeck process models the strength of phylogenetic correlations using selection (see fig. 1). It assumes that drift acts to push effect sizes away from an adaptive optimum (e.g., a peak in a fitness landscape) and that selection counteracts this drift from the optimum. In the absence of selection ($\beta \to 0$), the OU model collapses to BM. Estimating the contribution of selection in explaining variation among effect sizes is similar to estimating a phylogenetic signal ($\lambda$): this parameter can be fit via a manual or ML approach. For instance, a range of small, medium, and large values of selection (e.g., $\beta = 0.1, 1, 10$) are incorporated into $\mathbf{P}^{\mathrm{OU}}$ and then applied to meta-analysis (e.g., starting with eq. [3]). The ML estimate of selection ($\hat{\beta}$) is optimized to effect sizes using equations (16)–(18), but $\mathbf{P}^{\mathrm{OU}}$ is applied instead of $\mathbf{P}^{\mathrm{BM}}$. However, a nonlinear optimization method is required to estimate selection because $\beta$ does not fit linearly in equation (17). Statistical languages such as R provide nonlinear optimization libraries that are useful to solve this issue. Again, all of the effect size statistics (e.g., pooled effect size, homogeneity test) are fitted conditionally on selection.

Under the OU model, the intensity of drift ($\eta$) is the rate at which phylogenetic correlations are lost among taxa. This intensity can be illustrated in figure 1 as the full (BM) line converging to a null slope. Hansen (1997) estimates $\eta$ as being conditional on selection, and it is calculated directly from the residual sums of squares (SSE) of a GLS model with $\hat{\beta}$ as follows:

$$\eta = \sqrt{\frac{\mathrm{SSE}^{\mathrm{OU}(\hat{\beta})}}{k-m}}, \quad (19)$$

where $m$ is the column rank of **X**.

### *Moderator Variables as Adaptive Optima*

Moderator variables are important for evaluating bias and testing hypotheses with meta-analysis (see appendix; Cooper and Hedges 1994). They function as explanatory variables for variation among effect sizes. Here, I outline how to modify moderator variables to test whether they can also serve as adaptive explanatory variables. This is possible because when estimating phylogenetic correlations, the OU model assumes that effect sizes are maintained near an adaptive (fitness) optimum through selection (Hansen 1997). It is further assumed that selection ($\beta$) will act to release ancestral constraints on adaptation by pushing taxa to evolve toward a new primary (adaptive) optimum and the strength of $\beta$ will determine the rate at which taxa evolve from the ancestral to the primary (contemporaneous) optimum (fig. 1). Should moderator groupings had functioned as adaptive optima throughout the evolutionary history of taxa, then the (contemporaneous) effect sizes should also contain a signature of selection relative to the position of these optima. Thus, by hypothesizing (*a*) multiple optima using moderator variables and (*b*) their relative positioning in the phylogeny (e.g., whether they are primary or ancestral), it is possible to test whether these moderator variables have been adaptive for the taxa (e.g., have functioned as true optima). This is done evaluating whether a signature of selection can be recovered from effect sizes based on these hypothesized optima (Hansen 1997; Butler and King 2004; Hansen et al. 2008).

To generate an OU model with $m$ number of adaptive optima ($\theta$), the $k \times m$ design matrix (**X**) of equation (8) is modeled to weight effect sizes based on these optima. These weights are based on the time spent at each optimum, such that taxa evolving under a primary optimum will have their effect sizes weighted more heavily than taxa evolving under an ancestral optimum. Using this new **X**

matrix in equation (8) will yield weighted averages among moderator groups, where more weight will be given to effect sizes from taxa evolving under the primary optimum. Using ML (via eq. [16]), we can then estimate whether there is a signature of selection based on these moderator groups. For example, if hypothesizing a particular grouping resulted in a strong signature of selection (e.g., $\beta \to 0$), then these moderator groups serve as good adaptive explanations for the phylogenetic patterning of effect sizes.

There are several ways to model moderator variables as adaptive optima. However, the elements of $\mathbf{X}$ for all models can be generalized as

$$\mathbf{X}_{i,m} = e^{-\beta\theta_i^{\text{end}}} - e^{-\beta\theta_i^{\text{start}}}, \tag{20}$$

where for each $i$th taxon, $\theta^{\text{start}}$ and $\theta^{\text{end}}$ are the start and end of a period throughout the phylogeny occupied by this optimum. The phylogenetic branch length relative to $\theta^{\text{start}}$ and $\theta^{\text{end}}$ of each optimum (designated by $b^{\theta}$) will form the weighting scheme: effect sizes derived from taxa evolving for long periods under the primary optimum will be weighted more heavily than taxa evolving under the ancestral optimum. Similar to modifying the design matrix to conduct a one-way ANOVA, the sum of elements of each row vector of $\mathbf{X}$ must equal 1, and the global sum of all the elements of $\mathbf{X}$ will equal $k$ (Groß 2003).

Examples of coding the elements of $\mathbf{X}$ with two adaptive optima are found in figure 2, but more complex models can also be hypothesized. For instance, we can assume that the ancestral optimum is unknown and that there are two primary optima derived from this unknown. Here a third column in $\mathbf{X}$ will designate the unknown (ancestral) optimum, with elements equaling $e^{-\beta b_i^{\text{total}}}$, and the remaining columns will include the distances relative to the first and second primary optima to this unknown (e.g., $\mathbf{x}_{i,m}^{\theta} = [e^{-\beta b_i^{\text{total}}}, e^{-\beta b_i^{\text{total}}} - e^{-\beta b_i^{\theta}}, e^{-\beta b_i^{\theta}}]$). The appendix provides an example of hypothesizing multiple optima using meta-analysis.

### *Model Selection among Competing Evolutionary Hypotheses*

The above framework generates multiple evolutionary hypotheses to explain evolutionary variation among effect sizes—from simple neutral models to more elaborate explanations with multiple adaptive optima. These models have a design based on an Ornstein-Uhlenbeck process (Hansen 1997; Butler and King 2004) and serve as competing hypotheses to explain evolutionary variation among effect sizes. To determine the best fit of these models, Akaike's Information Criterion (AIC) values are compared, with the lowest AIC chosen as the best fit. This approach differs from the more common application of likelihood-ratio tests to contrast different evolutionary hypotheses in comparative analysis (see Pagel 1997). The AIC scores are more useful for evolutionary meta-analysis because of the potentially large number of evolutionary hypotheses that could be generated using OU models. The AIC scores also avoid problems relating to the sequential testing of multiple hypotheses within a nested design, such as making assumptions about which model will serve as the null hypothesis and the subsequent statistical nonindependence of comparisons resulting from shared null hypotheses (see Cohen 1994).

Each evolutionary model predicts a different way effect sizes can be phylogenetically correlated ($\mathbf{P}$), and these

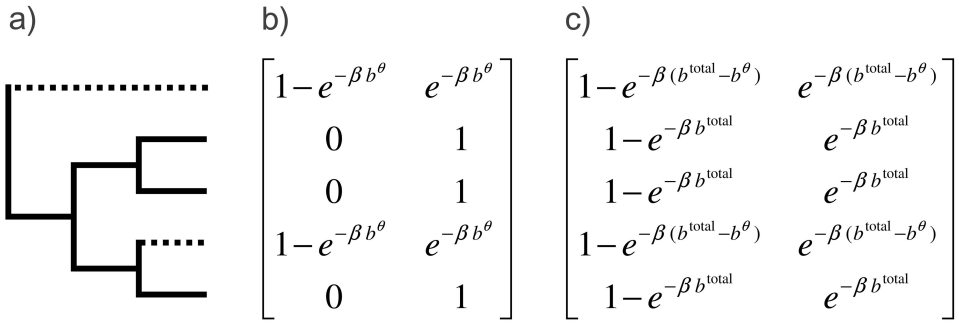



**Figure 2:** Two examples of coding the design matrix ($\mathbf{X}$) to test different evolutionary hypotheses on the influence of a moderator variable (modeled as adaptive optima) in explaining variation among effect sizes. The phylogeny (*a*) has two adaptive optima: taxa designed with a dashed line have the primary optimum, whereas the remaining taxa have the secondary (ancestral) optimum (*solid line*). The design matrix (*b*) is coded assuming that the primary optimum ($\theta_1$) is derived from the secondary optimum with an unknown (unrooted) origin ($\theta_2^{\text{start}} = \infty$). The second design matrix (*c*) assumes that the primary optimum evolved from an unknown but rooted secondary optimum ($\theta_2^{\text{start}} = b^{\text{total}}$). The elements of these matrices are simplified versions of equation (20).

models form a nested hierarchy of hypotheses illustrated as follows:

$$\mathbf{P}^{\mathbf{I}} \overset{\lambda}{\subset} \mathbf{P}^{\mathrm{BM}} \overset{\beta}{\subset} \mathbf{P}^{\mathrm{OU}} \overset{\beta(m-1)}{\subset} \mathbf{P}^{\mathrm{OU}(m-1)} \subset \cdots \overset{\beta(m)}{\subset} \mathbf{P}^{\mathrm{OU}(m)}, \tag{21}$$

where $\subset$ indicates that the left hypothesis is a subset of the one to the right and where the symbols above $\subset$ are evolutionary parameters ($\lambda$ and $\beta$) used to parameterize the fit of the proceeding model. An empirical example of the nestedness of different evolutionary models is found in the appendix. The relative fit of each model is evaluated with an AIC score

$$\mathrm{AIC} = 2m - 2\ln(L). \tag{22}$$

Here the likelihood estimate ($L$) of each GLS model (see eq. [16]) forms the basis for model selection. This approach penalizes models with high error (low fit) in describing the data and models that use multiple evolutionary parameters to describe data (e.g., $\mathbf{P}^{\mathrm{OU}(m)}$). For instance, should $\mathrm{AIC}(\bar{\delta}^{BM(\lambda=0)}) < \mathrm{AIC}(\bar{\delta}^{BM(\lambda=1)})$, then pooling effect sizes with a model without phylogenetic correlations was more effective in minimizing statistical error than a model assuming phylogenetic correlations under a BM process.

## Discussion

The framework outlined in this article serves to improve the statistical inference of meta-analysis based on research from multiple taxa. However, several theoretical and practical issues relating to the application of phylogenetic information need discussion. One issue that will affect the ability of meta-analysis to test evolutionary hypotheses is the availability and ease of analysis of information used to connect evolutionary relationships. For example, I could find useful molecular data for only half of the species found in a published meta-analysis (see appendix). This lack of information resulted in a significant subsampling of the studies from the original review. Clearly, publication bias is already known to affect the sampling of studies used in meta-analysis—where nonsignificant or marginally significant research is less likely to be published and thus less likely to be included in meta-analyses (e.g., a file drawer problem; see Arnqvist and Wooster 1995). What is less clear is whether excluding research from species lacking phylogenetic information can exacerbate this bias. If the reasons why species lacked phylogenetic information were random, then their omission from meta-analysis would not cause additional bias or at least not exacerbate publication bias; it would only erode statistical power because fewer effect sizes would be available for review (Rosenthal 1991). However, investigator bias in the genes or species used for phylogenetic construction (e.g., using only model species or single representative species for an entire genus; Hillis 1998; Wiens 2003) and bias in how this information is reported in publications or public databases (Leebens-Mack et al. 2006) will likely contribute to nonrandom gaps in phylogenetic information.

The taxonomic composition of effect size data can further bias the statistical inference of meta-analysis. The effect size data will likely be based on a collection of taxa with a nonrandom paraphyletic (e.g., a collection of insect orders) or polyphyletic structure (e.g., only taxa where there is published research). This composition can arise from natural differences in species richness and diversity of taxa or from taxon and publication bias in research—where taxa from model systems are more likely to be studied and show significant results because more information is known to control for experimental bias (Clark and May 2002; Cassey et al. 2004). In the example outlined in the appendix, many of species included in the meta-analysis were agricultural pests. A nonrandom (phylogenetic) sample may also yield a collection of distantly related taxa—this would result in small phylogenetic correlations and little observable difference between a phylogenetically controlled and traditional meta-analysis (e.g., $\boldsymbol{\Sigma}^{\mathrm{BM}(\lambda=1)} \approx \boldsymbol{\Sigma}^{\mathrm{BM}(\lambda=0)}$; Martins and Housworth 2002). Testing evolutionary hypotheses with a phylogenetically nonrandom sample will further bias the ability of statistical tests to detect evolutionary signals should they exist—magnifying problems associated with poor sample size (Freckleton et al. 2002) and data type (e.g., morphological vs. behavioral; see Blomberg et al. 2003).

To improve the sample size of evolutionary meta-analysis and to reduce subsampling biases, less restrictive criteria for assembling phylogenetic relationships can be applied (e.g., integrating information from published phylogenies or Linnean rankings). Using simulations, Freckleton et al. (2002) found that coarse phylogenetic information may still improve statistical inferences and the description of the data—despite evolutionary relationships having numerous polytomies or lacking information on relative divergence times (see Purvis and Garland 1993). A balance must be met between restricting analyses to taxa with precise phylogenetic information and expanding the scope of the review by including numerous taxa with coarse phylogenetic correlations. This is a problem similar to when effect size metrics that require multiple pieces of information to quantify research outcomes are used over less restrictive metrics. Here again, studies lacking information are excluded from reviews. Using simulations, however, Lajeunesse and Forbes (2003) found that meta-analyses based on few but high-precision data had improved error rates because more within-study information

was used to control for bias. Evolutionary meta-analysis will also likely benefit from this effect: the addition of phylogenetic information should improve statistical inference by decreasing Type I errors due to nonindependence (see Harvey and Purvis 1991; Purvis and Garland 1993).

Given these issues, when should reviewers integrate phylogenetic information into ecological meta-analysis? Reviews where the units of analysis are communities or ecosystems, for example, are the least likely to be biased by phylogenetic nonindependence—given that experimental outcomes are a composite of biotic and abiotic effects. The greatest potential for bias will occur when there is a single effect size per species and when there is a strong phylogenetic signal of experimental responses between these species. Unfortunately, there is the philosophy to wave off "phylogenetically controlled" analyses should data lack a phylogenetic signal (Westoby et al. 1995*a*, 1995*c*). This should be avoided because failing to detect a signal may be more of a statistical than a biological issue (see Martins 2000). For instance, there is continued difficulty in recovering phylogenetic signals from small phylogenies (Freckleton et al. 2002). It also remains unknown how the associated error due to incomplete taxon sampling and publication bias will further affect the ability to detect phylogenetic signals for meta-analytical data. To reflect these limitations, a philosophy that explicitly recognizes phylogenetic relationships should be used—even when effect sizes show no phylogenetic signal. Remember that testing for a phylogenetic signal is the same as evaluating the fit of a BM model to data (see Pagel 1997), and thus it is possible to integrate this test within a model selection framework (see example in appendix). This approach avoids making research decisions based solely on the significance testing of phylogenetic signals (e.g., not pursuing evolutionary analyses) and provides more information on which evolutionary model (with and without phylogenetic information) is better at describing variation among effects sizes (see Butler and King 2004).

Finally, the model selection framework outlined in this article uses moderator variables to test hypotheses on adaptive evolution. Moderator variables are modeled as adaptive optima (e.g., local maxima or peaks in an adaptive landscape), and these optima are hypothesized sources of variation among effect sizes. This optimality approach (based on OU process) avoids having to estimate ancestral states when hypothesizing the evolution and distribution of effect sizes (Webster and Purvis 2002). This is because the evolutionary processes modeling phylogenetic correlations and adaptive optima (e.g., pure drift and stabilizing selection) do not need explicit hypotheses on the optimal states of internodes found in phylogenetic trees. However, defining the states of these internodes, by using ancestral state reconstruction or heuristically hypothesizing plausible intermediate evolutionary transitions, would allow for evolutionary meta-analysis to test hypotheses on direction selection. Directional selection can be integrated in evolutionary meta-analysis by modifying the design matrix (see fig. 2) to include these intermediate transitional optima. However, only the most parsimonious directional model (with the fewest number of evolutionary transitions and intermediate optima) should be applied to the model selection framework. Otherwise, the directional model will always fair poorly during model selection because models with numerous optima have severely penalized AIC scores.

## Conclusions

Meta-analysis is a retrospective endeavor—and the lessons learned from synthesizing published research should serve as a stepping point for future experiments. Testing evolutionary hypotheses with meta-analysis can reveal phylogenetic signals in experimental responses across taxa and help distinguish whether these signals are due to neutral and adaptive processes. This information should feed back into a comparative approach to experimentation: where designs explicitly consider phylogenetic relationships of taxa (see Webb et al. 2002) and conceptualize experimentation through effect sizes (e.g., Nakagawa and Cuthill 2007; Strauss et al. 2008). This approach would allow both the primary researcher and the meta-analyst to reach the broadest generalization possible and perhaps yield causal explanations for the diversity of ecological and evolutionary patterns observed among taxa.

## Acknowledgments

I thank the National Center for Ecological Analysis and Synthesis (NCEAS) meta-analysis work group for thoughtful discussion on this topic, in particular, J. Gurrevitch; M. Jennions, for lending his study for analysis; J. Koricheva; K. Mengersen; and M. Rosenberg, for a generous solution to making these statistics applicable to all effect-size metrics. I also thank D. Adams, A. Agrawal, B. Danforth, N. Hairston Jr., R. Harrison, and an anonymous reviewer for valuable feedback on this manuscript. This research was supported by National Science Foundation grants to the National Evolutionary Synthesis Center (EF-0423641) and the NCEAS (DEB-0553768).

## Literature Cited

Adams, D. C. 2008. Phylogenetic meta-analysis. Evolution 62:567–572.

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

Beirinckx, K., H. Van Gossum, M. J. Lajeunesse, and M. R. Forbes. 2006. Sex biases in dispersal and philopatry: insights from a meta-

analysis based on capture-mark-recapture studies of damselflies. Oikos 113:539–547.

Björklund, M. 1997. Are "comparative methods" always necessary? Oikos 80:607–612.

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

Borer, E. T., E. W. Seabloom, J. B. Shurin, K. E. Anderson, C. A. Blanchette, B. Broitman, S. D. Cooper, and B. S. Halpern. 2005. What determines the strength of a trophic cascade? Ecology 86:528–537.

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

Cassey, P., J. G. Ewen, T. M. Blackburn, and A. P. Møller. 2004. A survey of publication bias within evolutionary ecology. Proceedings of the Royal Society B: Biological Sciences 271:S451–S514.

Cheverud, J. M., M. M. Dow, and W. Leutenegger. 1985. The quantitative assessment of phylogenetic constraints in comparative analyses: sexual dimorphism in body-weight among primates. Evolution 39:1335–1351.

Clark, J. A., and R. M. May. 2002. Taxonomic bias in conservation research. Science 297:191–192.

Cohen, J. 1994. The earth is round ($p < .05$). American Psychologist 49:997–1003.

Cooper, H., and L. V. Hedges, eds. 1994. Handbook of research synthesis. Russell Sage Foundation, New York.

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.

Gardiner, C. W. 1985. Handbook of stochastic methods for physics, chemistry and the natural sciences. Springer, Berlin.

Garland, T., Jr., 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.

Garland, T., Jr., A. W. Dickerman, C. M. Janis, and J. A. Jones. 1993. Phylogenetic analysis of covariance by computer simulation. Systematic Biology 42:265–292.

Groß, J. 2003. Linear regression. Springer, Berlin.

Gurevitch, J., and L. V. Hedges. 1999. Statistical issues in conducting ecological meta-analyses. Ecology 80:1142–1149.

Hansen, T. F. 1997. Stabilizing selection and the comparative analysis of adaptation. Evolution 51:1341–1351.

Hansen, T. F., and E. P. Martins. 1996. Translating between microevolutionary process and macroevolutionary patterns: the correlation structure of interspecific data. Evolution 50:1404–1417.

Hansen, T. F., J. Pienaar, and S. H. Orzack. 2008. A comparative method for studying adaptation to a randomly evolving environment. Evolution 62:1965–1977.

Harvey, P. H., and M. D. Pagel. 1991. The comparative method in evolutionary biology. Vol. 1. Oxford Series in Ecology and Evolution. Oxford University Press, Oxford.

Harvey, P. H., and A. Purvis. 1991. Comparative methods for explaining adaptation. Nature 351:619–624.

Hedges, L. V. 1992. Meta-analysis. Journal of Educational Statistics 17:279–296.

Hedges, L. V., and I. Olkin. 1985. Statistical methods for meta-analysis. Academic Press, Orlando, FL.

Hillis, D. M. 1998. Taxonomic sampling, phylogenetic accuracy, and investigator bias. Systematic Biology 47:3–8.

Koricheva, J., H. Nykänen, and E. Gianoli. 2004. Meta-analysis of trade-offs among plant antiherbivore defenses: are plants jacks-of-all-trades, masters of all? American Naturalist 163:E64–E75.

Lajeunesse, M. J., and M. R. Forbes. 2002. Host range and local parasite adaptation. Proceedings of the Royal Society B: Biological Sciences 269:703–710.

———. 2003. Variable reporting and quantitative reviews: a comparison of three meta-analytical techniques. Ecology Letters 6:448–454.

Leebens-Mack, J., T. Vision, E. Brenner, J. E. Bowers, S. Cannon, M. J. Clement, C. W. Cunningham, et al. 2006. Taking the first steps towards a standard for reporting on phylogenies: minimum information about a phylogenetic analysis (MIAPA). Omics 10:231–237.

Maddison, W. 1990. A method for testing the correlated evolution of two binary characters: are gains or losses concentrated on branches of a phylogenetic tree? Evolution 44:539–557.

Martins, E. P. 1994. Estimating the rate of phenotypic evolution from comparative data. American Naturalist 144:193–209.

———. 2000. Adaptation and the comparative method. Trends in Ecology & Evolution 15:296–299.

Martins, E. P., and T. Garland Jr. 1991. Phylogenetic analyses of the correlated evolution of continuous characters: a simulation study. Evolution 45:534–557.

Martins, E. P., and E. A. Housworth. 2002. Phylogeny shape and the phylogenetic comparative method. Systematic Biology 51:873–880.

Nakagawa, S., and I. C. Cuthill. 2007. Effect size, confidence interval and statistical significance: a practical guide for biologists. Biological Review 82:591–605.

Pagel, M. 1994. Detecting correlated evolution on phylogenies: a general method for the comparative analysis of discrete characters. Proceedings of the Royal Society B: Biological Sciences 255:37–45.

———. 1997. Inferring evolutionary processes from phylogenies. Zoologica Scripta 26:331–348.

Pagel, M. D. 1992. A method for the analysis of comparative data. Journal of Theoretical Biology 156:431–442.

———. 1999. Inferring the historical patterns of biological evolution. Nature 401:877–884.

Parker, J. D., D. E. Burkepile, and M. E. Hay. 2006. Opposing effects of native and exotic herbivores on plant invasions. Science 311: 1459–1461.

Purvis, A., and T. Garland Jr. 1993. Polytomies in comparative analyses of continuous characters. Systematic Biology 42:569–575.

Ricklefs, R. E., and J. M. Starck. 1996. The application of phylogenetically independent contrasts: a mixed progress report. Oikos 77:167–172.

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

Rosenthal, R. 1991. Meta-analytic procedures for social research. Sage, Newbury Park, CA.

Schluter, D. 2000. The ecology of adaptive radiations. Oxford University Press, Oxford.

Strauss, S. Y., J. A. Lau, T. W. Schoener, and P. Tiffin. 2008. Evolution in ecological field experiments: implications for effect size. Ecology Letters 11:199–207.

Van Zandt, P. A., and S. Mopper. 1998. A meta-analysis of adaptive deme formation in phytophagous insect populations. American Naturalist 152:595–604.

Verdú, M., and A. Traveset. 2004. Bridging meta-analysis and the

comparative method: a test of seed size effect on germination after frugivores' gut passage. Oecologia (Berlin) 138:414–418.

———. 2005. Early emergence enhances plant fitness: a phylogenetically controlled meta-analysis. Ecology 86:1385–1394.

Webb, C. O., D. D. Ackerly, M. A. McPeek, and M. J. Donoghue. 2002. Phylogenies and community ecology. Annual Review of Ecology and Systematics 33:475–505.

Webster, A. J., and A. Purvis. 2002. Testing the accuracy of methods for reconstructing ancestral states of continuous characters. Proceedings of the Royal Society B: Biological Sciences 269:143–149.

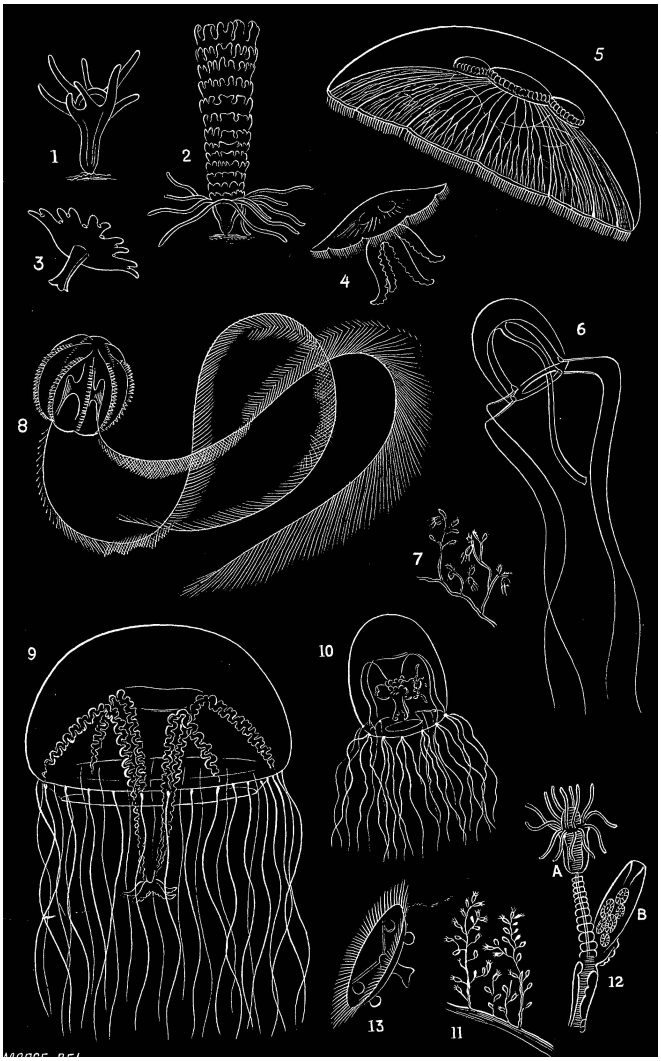
Westoby, M., M. R. Leishman, and J. M. Lord. 1995*a*. Further remarks on phylogenetic correction. Journal of Ecology 83:727–729.

———. 1995*b*. Issues of interpretation following phylogenetic correction. Journal of Ecology 83:892–893.

———. 1995*c*. On misinterpreting the "phylogenetic correction." Journal of Ecology 83:531–534.

Wiens, J. J. 2003. Missing data, incomplete taxa, and phylogenetic accuracy. Systematic Biology 52:528–538.

Associate Editor: Stephen B. Heard
Editor: Michael C. Whitlock




1) Early condition of *Aurelia* (after Agassiz); 2) older condition of same, showing individuals about to separate (after Agassiz); 3) *Aurelia*, a short time after freeing itself; 4) advanced stage; 5) *Aurelia* in adult condition (after Agassiz); 6) *Coryne mirabilis* Ag.; 7) hydroid community of *Coryne* magnified, showing Jelly-fish buds about to separate (after Clark); 8) *Pleurobrachia* (after A. Agassiz); 9) *Tima formosa* Ag.; 10) *Lizzia grata* Ag., showing young ones budding from stomach; 11) hydroid community of *Eucope diaphana* Ag. (after A. Agassiz); 12) showing one twig of Eucope with fixed individual A, and reproductive capsule B, containing a number of young Jelly-fishes (after A. Agassiz); 13) *Eucope*, in adult condition, magnified, from "Something about Jelly-Fishes" by Edward S. Morse (*American Naturalist*, 1867, 1:244–253).

 DOI: 10.1086/603628

# Appendix from M. J. Lajeunesse, "Meta-Analysis and the Comparative Phylogenetic Method"

## (Am. Nat., vol. 174, no. 3, p. 369)

## The Covariance Matrix of Phylogenetically Independent Meta-Analysis

I begin with a formal description on how the correlation and covariance between pairs of effect sizes are statistically related and then outline how this relationship is used to define the elements of the covariance matrix $\mathbf{\Sigma}$ for phylogenetically independent meta-analysis. This approach to estimating $\mathbf{\Sigma}$ is used by all meta-analytical generalized least squares (GLS) methods that account for the statistical dependence that arises when a collection of effect sizes share a correlated structure (Hedges and Olkin 1985; Becker 1992; Glesser and Olkin 1994; Marín-Martínez and Sánchez-Meca 1999; Cheung and Chan 2004). In the following section, I also show that a slightly modified version of Adams's (2008) phylogenetic meta-analysis will yield the same pooled effect size as the method presented below.

A correlation is formally defined as a scaled version of the covariance (Hedges and Olkin 1985); for example, the correlation between effect sizes $\delta_a$ and $\delta_b$ is defined as

$$\text{cor}(\delta_a, \delta_b) = \frac{\text{Cov}(\delta_a, \delta_b)}{\text{SD}(\delta_a)\text{SD}(\delta_b)}. \tag{A1}$$

Here, SD is the standard deviation of $\delta$. Assuming that information on $\text{cor}(\delta_a, \delta_b)$ is already known, as well as the SD for each effect size, then it is possible to rearrange equation (A1) to yield the covariance between $\delta_a$ and $\delta_b$:

$$\text{Cov}(\delta_a, \delta_b) = \text{cor}(\delta_a, \delta_b)\text{SD}(\delta_a)\text{SD}(\delta_b). \tag{A2}$$

The covariance equation (A2) has two important properties. First, when two effect sizes are fully independent (e.g., $\text{cor}[\delta_a, \delta_b] = 0$), then their covariance will equal 0. Second, equation (A2) reduces to the variance $\sigma^2(\delta)$ of an effect size when it is correlated with itself (e.g., $\text{cor}[\delta_a, \delta_a] = 1$). For example, the covariance of $\delta_a$ is

$$\text{Cov}(\delta_a, \delta_a) = 1 \times \text{SD}(\delta_a)\text{SD}(\delta_a) = [\text{SD}(\delta_a)]^2 = \sigma^2(\delta_a). \tag{A3}$$

These properties of equation (A2) allow for the proper codification of the elements of $\mathbf{\Sigma}$ to simultaneously weight effect sizes by sample precision and phylogenetic correlations. More precisely, for each pair of $k$ number of phylogenetically dependent effect sizes, the $i = 1, \ldots, k$ rows and $j = 1, \ldots, k$ columns of $\mathbf{\Sigma}$ become

$$\mathbf{\Sigma}_{i,j} = \begin{cases} \sigma^2(\delta_i) & \text{if } i = j \\ \mathbf{P}_{i,j}\sqrt{\sigma^2(\delta_i)}\sqrt{\sigma^2(\delta_j)} & \text{if } i \neq j \end{cases}. \tag{A4}$$

Here, the off-diagonals of $\mathbf{\Sigma}$ (when $i \neq j$) use the correlations found in the phylogenetic matrix $\mathbf{P}$ defined by equations (5) and (6) while converting the known sampling variances of each effect size into standard deviations. These diagonal covariances serve to weight effect sizes by their relative phylogenetic correlation, where effect sizes derived from closely related species are downweighted in the overall meta-analysis. The main diagonal of $\mathbf{\Sigma}$ simply becomes equation (A3) because effect sizes are 100% phylogenetically correlated with themselves. This main diagonal serves to downweight effect sizes with large sampling variances because these effect size estimates may be more sensitive to sampling error.

Applying this covariance matrix $\mathbf{\Sigma}$ to the GLS equation (8) will yield a phylogenetic-independent and

weighted pooled effect size. For a more straightforward description of how $\boldsymbol{\Sigma}$ is used in GLS models and how it satisfies the conditions for phylogenetic meta-analysis, it is necessary to define $\boldsymbol{\Sigma}$ using matrix notation as follows:

$$\boldsymbol{\Sigma} = \mathbf{DPD}. \tag{A5}$$

Here, $\mathbf{D}$ is a $k \times k$ matrix containing the SD of each effect size on the main diagonal and 0s in all off-diagonals (e.g., $\mathbf{D} = \text{diag}[\text{SD}(\delta_1), \ldots, \text{SD}(\delta_k)]$), and $\mathbf{P}$ again is the phylogenetic matrix defined by equations (5) and (6). Equation (A5) is the same covariance matrix described in equation (A4), and this matrix version of $\boldsymbol{\Sigma}$ is also found in Hedges and Olkin (1985). Two important statistical conditions of the covariance matrix for GLS modeling are that $\boldsymbol{\Sigma}$ is symmetric and positive definite (Groß 2003). Equations (A4) and (A5) satisfy these conditions as long as $\mathbf{P}$ is treated as a statistical correlation matrix with the following constraints: (*a*) taxa can be 100% correlated only with themselves, and (*b*) all off-diagonals contain the relative phylogenetic correlations among taxa ranging from 0 to near 1 but not equaling 1. Equation (4) serves to maintain these conditions during meta-analysis.

A few simplifications of equation (A5) can also show how $\boldsymbol{\Sigma}$ retains the weighting properties of both traditional meta-analysis and comparative analyses based on Brownian motion (BM) evolution. Assuming that effect sizes are not phylogenetically correlated, then $\mathbf{P} = \mathbf{I}$, and the covariance matrix simplifies to $\boldsymbol{\Sigma} = \mathbf{DID} = \mathbf{DD} = \text{diag}[\sigma^2(\delta_1), \ldots, \sigma^2(\delta_k)]$. Here $\mathbf{I}$ is an identity matrix containing only 1s on the main diagonal and 0s in all off-diagonals. This covariance matrix containing only the sampling variances of each effect size on its main diagonal is the same covariance matrix used in traditional meta-analysis (Hedges and Olkin 1985). Pagel (1997) and Rohlf (2001) define the covariance matrix as fitting the assumptions of BM evolution when the raw branch lengths of the phylogenetic tree are explicitly used to define $\boldsymbol{\Sigma}$. That is, $\boldsymbol{\Sigma} = \mathbf{P}$ when $\mathbf{P}$ is defined with equations (5) or (15) with $\lambda = 1$. Simplifying equation (A4), by removing the meta-analytical component of $\boldsymbol{\Sigma}$ (e.g., $\mathbf{D} = \mathbf{I}$), also yields the same covariance matrix used in comparative analyses assuming BM ($\boldsymbol{\Sigma} = \mathbf{IPI} = \mathbf{P}$). Figure A1 further illustrates how more complex evolutionary models, such as the Ornstein-Uhlenbeck (OU) process, have a nested design that can also be reduced to BM. Further, assuming that there are no phylogenetic correlations ($\mathbf{P} = \mathbf{I}$) and no differences in sampling variances among effect sizes ($\mathbf{D} = \mathbf{I}$) results in $\boldsymbol{\Sigma} = \mathbf{DPD} = \mathbf{III} = \mathbf{I}$. This reduction to the covariance matrix yields the simple arithmetic mean of effect sizes when applied to the GLS equation (8).

## A Comparison with Adams's (2008) Phylogenetic Meta-Analysis

Here, I compare the evolutionary meta-analysis outlined in this article with Adams's (2008) phylogenetic meta-analysis. I reanalyze Adams's (2008) meta-analytical data on latitudinal patterns on body size in mammals to determine whether $\boldsymbol{\Sigma}$, as specified in equation (A5), provides a pooled effect size equivalent to that of the Adams method for a fixed-effects model assuming BM evolution. However, the pooled effect size from Adams's paper is incorrect because the effect size data (contained in Adams's $\mathbf{E}$ vector) are misaligned with the phylogenetic correlation matrix (Adams refers to this matrix as $\boldsymbol{\Sigma}$, but I will continue to use $\mathbf{P}$). Here, I provide the correct (aligned) analysis of Adams's phylogenetic meta-analysis. To reanalyze these effect size data, I first extrapolated $\mathbf{P}$ from the phylogeny found in figure 2 of Adams's paper (see box A1 for this extrapolated phylogeny) and then aligned the species effect size data with $\mathbf{P}$. This matrix and aligned effect size data were then analyzed with the R code available online for Adams's (2008) method (see Dean Adams's Web site: http://www.public.iastate.edu/~dcadams/). Pooling effect sizes using Adams's approach resulted in $\bar{\bar{E}}_{p-m-a} = -0.067$, whereas assuming the same evolutionary model (BM) but using evolutionary meta-analysis yielded a pooled effect size of $\bar{\delta}^{\mathbf{P}} = -0.049$.

This difference between $\bar{\bar{E}}_{p-m-a}$ and $\bar{\delta}^{\mathbf{P}}$ is due to the variances in Adams's approach not having the same units as the effect sizes. A brief description of the phylogenetic GLS transformation method used by Adams is necessary to describe why this unit problem will introduce bias into meta-analysis. In Adams's approach, the effect size data (Adams's column vector $\mathbf{E}$) and the design matrix (vector $\mathbf{X}$) are first transformed to have zero phylogenetic correlations before analysis using a traditional weighted regression. This transformation is achieved by multiplying $\mathbf{E}$ and $\mathbf{X}$ with the inverse square root of the phylogenetic correlation matrix $\mathbf{P}$ (see Garland and Ives 2000; Groß 2003). These transformed vectors ($\mathbf{P}^{-1/2}\mathbf{E}$ and $\mathbf{P}^{-1/2}\mathbf{X}$) are then integrated in a traditional weighted regression model as follows:

$$\bar{\bar{E}}_{p-m-a} = [(\mathbf{P}^{-1/2}\mathbf{X})^{\mathrm{T}}\mathbf{W}(\mathbf{P}^{-1/2}\mathbf{X})]^{-1}(\mathbf{P}^{-1/2}\mathbf{X})^{\mathrm{T}}\mathbf{W}(\mathbf{P}^{-1/2}\mathbf{E}). \quad \text{(A6)}$$

Meta-analysis (via a weighted regression) uses the sampling variance of each effect size as its weight, where, for example, effect sizes with large variances get downweighted when pooling results. These sampling variances are the diagonals of the **W** weighting matrix. For most effect size metrics, these variances are actually approximations based on large-sample theory (Hedges and Olkin 1985). This means that the sample sizes of studies heavily determine the variances. Because these variances are approximations, their efficiency to act as weights assumes that the variances approximate the distribution of the effect sizes. However, when $\mathbf{P}^{-1/2}\mathbf{E}$ is used instead of **E** in a weighted regression based on **W**, this assumption is violated because variances in **W** (which are meant to approximate the distribution of **E**) are now used to weight transformed effect sizes with a different distribution: $\mathbf{P}^{-1/2}\mathbf{E}$. Thus, for Adams's method to provide unbiased (phylogenetically correct) pooled effect sizes, it would have to also adjust **W** to the same evolutionary units as $\mathbf{P}^{-1/2}\mathbf{E}$.

A much simpler solution to the Adams problem is to first use the GLS transformation method to convert **E** and **X** via **W**, followed by a phylogenetic least square regression (PGLS) using **P**. This PGLS analysis, as described in equation (4) in Adams's paper, would apply **E** and **X** transformed via **W** as follows:

$$\bar{\bar{E}}_{p-m-a} = [(\mathbf{W}^{-1/2}\mathbf{X})^{\mathrm{T}}\mathbf{P}(\mathbf{W}^{-1/2}\mathbf{X})]^{-1}(\mathbf{W}^{-1/2}\mathbf{X})^{\mathrm{T}}\mathbf{P}(\mathbf{W}^{-1/2}\mathbf{E}). \quad \text{(A7)}$$

Under this protocol, the required connection between the variances and effect sizes is retained, and what remain are the weightings based on phylogenetic correlations (**P**) to adjust these transformed effect sizes. Accounting for this source of bias in Adams's original meta-analysis results in the same pooled effect size ($\bar{\bar{E}}_{p-m-a} = -0.049$) as in the approach outlined in this article ($\bar{\delta}^{\mathbf{P}} = -0.049$).

## Homogeneity Tests, Moderator Variables, and the Random-Effects Model

Moderator variables are the primary method to test hypotheses in ecological meta-analysis (Gurevitch and Hedges 1999). Here, behavioral or ecological characteristics (e.g., mating system, trophic rank, or geographic region) are used as grouping variables for effect sizes to test hypotheses in explaining heterogeneity among effect sizes. Moderator variables, or quantitative predictors and covariates, are integrated into meta-analysis by modifying the design matrix **X** of GLS models. This approach is akin to one-way ANOVA or regression (Groß 2003). Here, I describe only how to include moderator variables in the design matrix; a description on how to test the significance of moderator groupings using within and between homogeneity statistics can be found in Hedges and Olkin (1985).

Moderator groups are integrated into meta-analysis by adding $m$ number of columns to **X** that code whether effect sizes belong to a particular group. For example, the following design matrix defines how five effect sizes are divided among two moderator groups ($m = 2$):

$$\mathbf{X} = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}^{\mathrm{T}}. \quad \text{(A8)}$$

Applying this design matrix when pooling effect sizes using the regression equation (8) will yield $\bar{\delta}^{\mathbf{P}}$ as a vector containing two pooled effect sizes for each moderator grouping—these are weighted by variance and controlled for phylogenetic nonindependence. It is also possible to integrate a covariate $c$ (e.g., a continuous variable such as migration rates or body size) by treating $c$ as an additional column in **X**:

$$\mathbf{X} = \begin{bmatrix} 1 & \dots & 1 \\ c_1 & \dots & c_k \end{bmatrix}^{\mathrm{T}}. \quad \text{(A9)}$$

This design matrix will yield $\bar{\delta}^{\mathbf{P}}$, a vector containing both the intercept and the regression coefficient—again, these statistics are weighted by variance and controlled for phylogenetic nonindependence. In the medical sciences, integrating a covariate in a GLS model is referred to as meta-regression (Thompson and Higgins 2002). Using matrices (A8) and (A9) in equation (12) will test whether these groups (or the covariate) are useful to explain variation among effect sizes.

**Box A1: Newick Tree**
**Mammal phylogeny extrapolated from figure 2 of Adams (2008)**

((((Alces_alces:357,(Puma_concolor:271,(Procyon_lotor:128,(Martes_martes:99,(Mustela_erminea:38,Mustela_frenata:38,Mustela_nivalis:38):61):29):143):86):1,(Anoura_cultrata:256,(Myotis_californicus:53,Myotis_daubentoni:53):203):102):12,(Blarina_brevicauda:338,Scapanus_latimanus:338):32):30,(Homo_sapiens:381,(Lepus_americanus:369,(((((Chaetodipus_penicillatus:135,(((Dipodomys_californicus:23,Dipodomys_phillipsii:23,(Dipodomys_microps:17,(Dipodomys_agilis:13,Dipodomys_venustus:13):4):6,(Dipodomys_heermanni:12,Dipodomys_panamintinus:12):11):4,(Dipodomys_compactus:24,Dipodomys_ordii:24):3,(Dipodomys_merriami:9,Dipodomys_nitratoides:9):18):1,(Dipodomys_deserti:26,(Dipodomys_nelsoni:11,Dipodomys_spectabilis:11):15):2):107):7,(Heteromys_gaumeri:101,Thomomys_bottae:101):41):182,(Rattus_rattus:198,(Microtus_montebelli:41,Microtus_pennsylvanicus:41):157,(Sigmodon_hispidus:160,(Neotoma_cinerea:108,(Peromyscus_eremicus:47,Peromyscus_maniculatus:47):61):52):38):126):19,(Tamiasciurus_douglasii:16,Tamiasciurus_hudsonicus:16):327):26):12):19);

When moderator groups are not useful for explaining heterogeneity but an unbiased pooled effect size is still needed to assess the overall magnitude of effect, then a random-effects model may be more appropriate for meta-analysis. The random-effects model assumes that effect sizes are random constants, each with its own variance (Hedges 1992; Cooper and Hedges 1994). This differs from the above approach to calculating $\bar{\delta}^{\mathbf{P}}$ because the underlying assumption of the regression equation (8) is that effect sizes have a common variance. This is referred to as the fixed-effects model (Hedges and Olkin 1985). Thus, assuming a random-effects model may be more appropriate when there is significant heterogeneity among effect sizes; in fact, it has been argued that the random-effects model should be the only approach to summarizing research in ecology (see Gurevitch and Hedges 1999).

To pool effect sizes with a random-effects model (also see Hedges 1992), an estimate of the between-effect size variance ($\tau$) is needed:

$$\tau = \begin{cases} \dfrac{Q_\tau - \mathrm{df}}{\mathrm{tr}(\boldsymbol{\Sigma}^{-1}) - \mathrm{tr}[(\mathbf{X}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}(\boldsymbol{\Sigma}\boldsymbol{\Sigma})^{-1}\mathbf{X}]} & \text{if } Q_\tau > \mathrm{df} \\ 0 & \text{if } Q_\tau \le \mathrm{df} \end{cases}, \tag{A10}$$

where tr is the trace of a matrix, $Q_\tau = \mathbf{E}^{\mathrm{T}}[\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\mathbf{X}(\mathbf{X}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}]\mathbf{E}$, $\boldsymbol{\Sigma} = \mathrm{diag}[\sigma^2(\delta_1), \ldots, \sigma^2(\delta_k)]$, and $m$ is the column rank of $\mathbf{X}$. Here, the between-effect size variance ($\tau$) is set to 0 when $Q_\tau$ is smaller than the degrees of freedom of the meta-analysis ($\mathrm{df} = k - m$), because by definition, the between-study variance cannot be negative (Hedges 1992). Note that $Q_\tau$ also functions as a test for whether $\tau$ is nonzero, where, for example, if $Q_\tau \le \chi^2_{\mathrm{df}=k-1}$, then there is little evidence to indicate that $\tau \neq 0$. Integrating the between-study variance is straightforward and involves adding $\tau$ to the variances of all effect sizes:

$$\hat{\sigma}^2(\delta_i) = \sigma^2(\delta_i) + \tau. \tag{A11}$$

These new variances are then used to calculate the variance-covariance matrix $\boldsymbol{\Sigma}$ with phylogenetic correlations. Finally, it is important to note that under the random-effects model, diagnostics for publication bias, such as the funnel plot (see Cooper and Hedges 1994), are ineffective because they assume a fixed-effect model.

## An Illustrative Example

*A Meta-Analysis with Multiple Evolutionary Hypotheses*

Here I revisit the Torres-Vila and Jennions (2005) meta-analysis on whether females have greater reproductive output if they mate with virgin males. This study is a good example of where an evolutionary meta-analysis could have been useful to integrate phylogenetic information and test hypotheses because (*a*) the authors found a family-level effect when testing for taxonomic bias and (*b*) they found that this effect related to the mating history of the taxa (e.g., among polyandrous and monandrous lepidopterans). I first begin by reanalyzing their data by pooling effect size and heterogeneity tests, followed by an evolutionary test of their hypothesis that male mating history will have a stronger effect in polyandrous species than in monandrous species.

*Methods*

A necessary aspect of meta-analysis is reporting the inclusion/exclusion criteria used to collate the group of studies for synthesis (Cooper and Hedges 1994). This helps identify sources of bias that can affect the representation of studies included in meta-analysis. This philosophy of transparency should also cross over to the approaches and methods used to collate data for phylogenetic tree construction. Below is a sketch of my inclusion/exclusion criteria for building a tree using genetic sequence data from GenBank. My approach is likely the most exclusive way of building a tree because the resulting meta-analysis will be based solely on taxa with available genetic data. However, for the purposes of this article, my approach is useful to (*a*) draw attention to taxonomic and genetic biases in public databases, and (*b*) to generate a phylogeny with relative branch lengths, which is important for testing evolutionary hypotheses with meta-analysis.

Partial sequence data useful for phylogenetic construction were available for 84% of 25 lepidopteran species analyzed in the original meta-analysis ($N = 25$; see table A1). These data were distributed across eight genes (COI, COII, 12S rRNA, 16S rRNA, 28S rRNA, Cyt B, NADH, EF-1 alpha), but no one gene was available for all species. A more involved analysis could make use of these data; however, for simplicity, I limited my analysis to the mtDNA COII gene (albeit only available for 53% of the 25 species). This gene best resolved the topology of taxa following known published relationships (Nylin et al. 2001). Further, COII data were not available for *Zeiraphera canadiensis*, but information existed for another species within the same genus (*Zeiraphera diniana*). I thus substituted information between these species, given that there were no other taxa from that genus included in the original meta-analysis. These 14 species belong to the suborder Ditrysia, and an additional two species from a sister suborder Incurvarioidea were selected as outgroups: *Prodoxus gypsicolor* (AF150920) and *Greya variabilis* (AF150909). All sequences were aligned with ClustalW (Larkin et al. 2007) and then visually inspected for consistency. This analysis found that GenBank data from *Diatraea considerate* did not align well (< 20% alignment) with other sequences, and thus this species was excluded from phylogenetic construction.

My final data set included 13 species with meta-analytical data plus two additional species serving as outgroups for phylogenetic construction. A Modeltest analysis (ver. 3.5; Posada and Crandall 1998) using Akaike's Information Criterion selection criteria determined that GTR+I+G was the best nucleotide substitution model for these data. This model was then applied to a maximum likelihood analysis using a heuristic search with tree bisection-reconnection (TBR) branch swapping and a stepwise addition starting tree with the ASIS stepwise addition option. This ML analysis was performed with PAUP* 4.0b10 (Swofford 2003) with a molecular clock assumption. Although these analyses were not calibrated against real time, the generated phylogeny still describes the relative temporal orderings of nodes (and relative intervening branch lengths), which again are necessary to test evolutionary hypotheses.

Using this phylogenetic hypothesis, I then performed an evolutionary meta-analysis to pool effect sizes across taxa using the following models: traditional (weighted) meta-analysis (labeled as N for normal); two BM models where the first was transformed with the ML estimate of phylogenetic conservatism ($\lambda = \text{ML}$) and the second assumed full phylogenetic correlations among taxa ($\lambda = 1$); and, finally, an OU model based on the ML estimate of selection ($\beta = \text{ML}$). I then repeated these analyses using monandry and polyandry as a moderator variable. I further treated these moderator variables as hypothesized adaptive optima using two multioptima OU models. The first model assumed that monandry was the primary optima (for which selection will act to maintain) and polyandrous taxa have the derived optima. I report only the results for this direction in the evolution among optima because opposite ordering (e.g., monandry as ancestral) would yield similar results. My second and most complex model assumed that both monandry and polyandry are separate primary optima derived from a third unknown adaptive optima. Hypothesizing this model is necessary given that the origin of the mating system is unknown. All of these evolutionary hypotheses are found in figure 4. Finally, for all GLS models I calculated homogeneity tests and AIC scores; these statistics will assess the fit of these competing phylogenetic models.

*Results*

As in the original meta-analysis by Torres-Vila and Jennions (2005), females mated to virgin males had greater reproductive output than females mated with nonvirgins among 13 lepidopteran species (table A2). A homogeneity test revealed significant variation among these 13 taxa under a fixed-effects model (table A2), but this variation was largely removed by parsing taxa into monandrous ($k = 5$) or polyandrous ($k = 8$) mating systems (table A2). For simplicity, I will thus continue to assume a fixed-effects model. In addition, females of

polyandrous taxa tended to have greater reproductive output than monandrous females when mated to virgin males (95% CIs of the two groups did not overlap; table A2). This latter result was marginal only in the original meta-analysis and was dependent on the further parsing of taxa by Rhopalocera (moths) and Heterocera (butterflies) taxonomic ranks (see Torres-Vila and Jennions 2005).

Accounting for the evolutionary history of Lepidoptera did not significantly change the overall meta-analysis: female reproductive output is strongly affected by male mating history (fig. 4; table A2). The GLS model with the best AIC score was the traditional (nonphylogenetically corrected) meta-analysis, whereas the meta-analysis with full phylogenetic correlations (BM with $\lambda = 1$) was the least effective in minimizing error when pooling effect sizes. This was expected given that the ML estimate of phylogenetic conservatism of effect sizes ($\lambda$) was near 0 (table A2) and the ML estimate of selection ($\beta$) from the OU model was very high. Both evolutionary parameters here removed all the phylogenetic correlations from meta-analysis, effectively modeling the same covariance matrix as traditional meta-analysis (e.g., $\mathbf{\Sigma}^{\mathrm{OU}(\beta=\mathrm{ML})} \cong \mathbf{\Sigma}^{\mathrm{BM}(\lambda=\mathrm{ML})} \cong \mathrm{diag}[\sigma^2(\delta_1), \ldots, \sigma^2(\delta_k)]$; see Diniz-Filho 2001). These evolutionary models were further penalized for modeling traditional meta-analysis with evolutionary parameters (see AIC scores in table A2).

A GLS model that included mating system as a moderator grouping had a higher AIC score than a model lacking these groupings (table A2). This effect was also indirectly observed with homogeneity tests: parsing species by mating system significantly removed all within-study heterogeneity among effect sizes (table A2). Thus, grouping species under monandrous and polyandrous mating system explained much of the variation in effect sizes across the 13 studies. However, as in the pooled analysis across all species, the evolutionary model with full phylogenetic correlations (BM where $\lambda = 1$) was least effective in explaining the patterning of effect sizes.

Modeling mating systems as evolutionary optima also did not improve the fit of GLS models and indicated again that selection was strong—yielding an evolutionary meta-analysis that was equivalent to traditional meta-analysis (table A2). These models with hypothesized adaptive optima had poor AIC scores because they explained the same amount of information as traditional meta-analysis (with a moderator variable) but with additional (ineffective) evolutionary parameters. Models with strong selection indicate again that mating system serves as an important explanatory variable for female reproductive output—should the ML estimate of selection have been weak, then this effect would have indicated that treating each mating system as an evolutionary optimum did not serve as a good explanation for the patterning of effect sizes.

Should modeling meta-analysis with phylogenetic information have resulted in a better fit to the data, then analyses would have had only a small marginal effect of the weighting of taxa belonging to either Rhopalocera (moths) and Heterocera (butterflies) taxonomic ranks (table A3). Perhaps this indicates that the significant effect among mating systems when grouping taxa among moths and butterflies in the original published meta-analysis was not a product of having accounted for the shared ancestry among these classes. The significant effect may have been an epiphenomenon of some other unknown moderating characteristic that is closely aligned with these two taxonomic classes. Models with phylogenetic information would have also given more weight to two species from the grass moth family (Crambidae): *Chilo partellus* and *Ostrinia nubilalis* (table A3). These species were weighted more heavily in all analyses because they have the greatest shared mean branch length to all other species (see phylogeny in fig. 4) and are thus the least phylogenetically correlated with other lepidopterans.

*Discussion*

I found that only partial sequence data was available for 53% of the species included in the original meta-analysis—albeit my inclusion criteria were narrow and limited to one gene. This subsampling of the original meta-analytical data resulted in a significant mating system effect that was detected in the original study only after correcting for taxonomic ranks and likely decreased the power to detect a phylogenetic signal given the small sample size (table A2). This bias also likely resulted in competing evolutionary models being less effective in fitting effect size data than compared to a model lacking phylogenetic information.

Dissecting the composition of effect sizes can provide further information for explaining these results with evolutionary meta-analysis. For instance, the effect sizes in Torres-Vila and Jennions's (2005) meta-analysis estimate the difference in lifetime fecundity between females mated with virgin and nonvirgin males. Among taxa where raw data were available, the lifetime fecundity of females mated with both virgin and experienced males had strong (nonzero) phylogenetic signals: fecundity with virgins ($\hat{\lambda} = 0.71$) and with experienced males ($\hat{\lambda} = 0.8$). These strong signals match empirical data showing that fecundity is often highly constrained by body

size (Honěk 1993) and that size is phylogenetically conserved among Lepidoptera. In addition, there is also a shared correlated evolutionary history between testis size and body size among Lepidoptera (Gage 1994). Presumably, this degree of trait conservatism would also be conserved experimentally across taxa, but the raw difference in fecundity among females mated with virgin or experienced males was not phylogenetically conserved ($\hat{\lambda} = 0.0$).

The results of evolutionary meta-analysis suggest that the mating difference in reproductive output of females is evolutionary labile and diverges rapidly and independently from constraints imposed by evolutionary history. This is expected, perhaps, given that mating with virgin and nonvirgin males has a direct effect on fitness and that the evolutionary meta-analysis found strong selection for increased fecundity when mating with virgin males (table A2). This selection would erase the contribution of phylogenetic correlations due to shared ancestry. However, there is some evidence to indicate that mating system can serve a constraint given the significant effect of parsing studies as monondrous and polyandrous groups (table A2)—however, whether differences in mating system serve as an evolutionary constraint remains unclear (table A2).

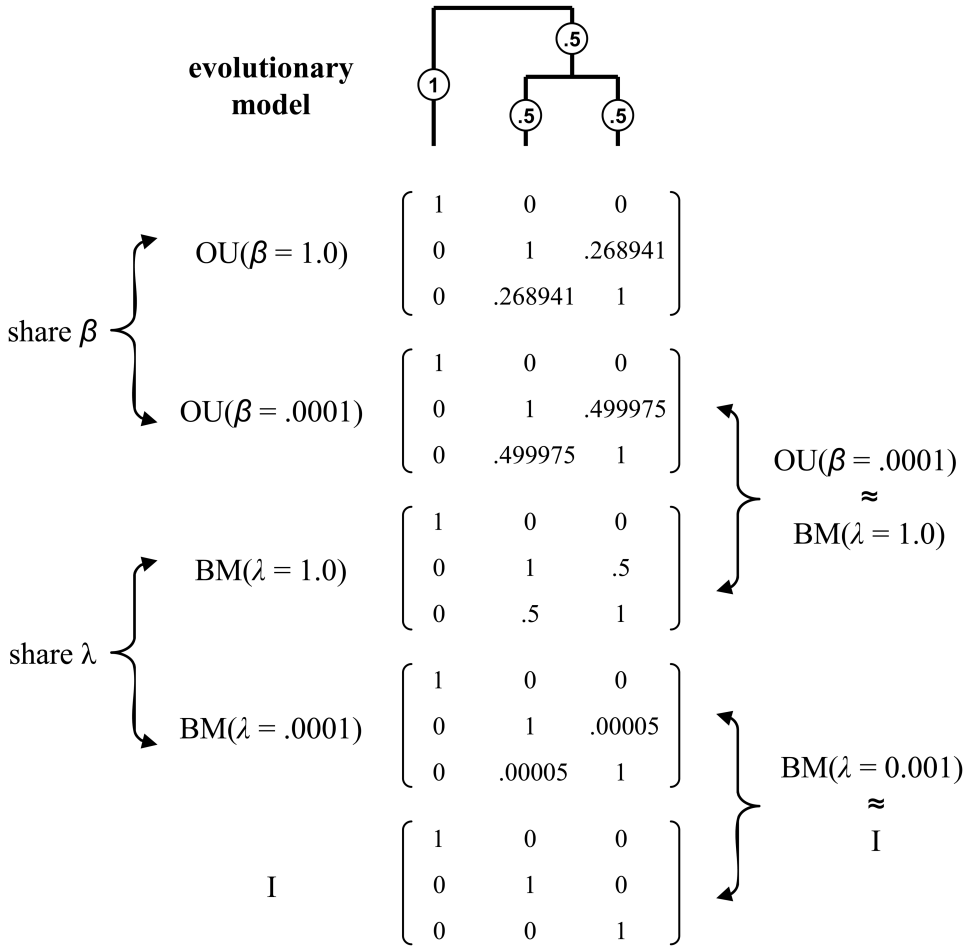



**Figure A1:** An illustration of the nestedness of evolutionary models used to define the phylogenetic correlation matrix (**P**) for meta-analysis. Above the five matrices is the hypothesized phylogenetic relationship of three taxa. The raw branch lengths of this tree (indicated on each branch) are used to define the elements of **P**. The five **P** matrices are ordered from the most complex model of evolution (*top*) to the simplest model assuming no phylogenetic correlations (**I**). The equations for calculating $\mathbf{P}^{OU}$ (OU = Ornstein-Uhlenbeck) is found in equation (6) and $\mathbf{P}^{BM}$ (BM = Brownian motion) in equations (5) and (15). For example, assuming an OU model with negligible selection ($\beta = 0.0001$) converges to the same matrix generated when assuming a BM model with a full phylogenic signal ($\lambda = 1$).

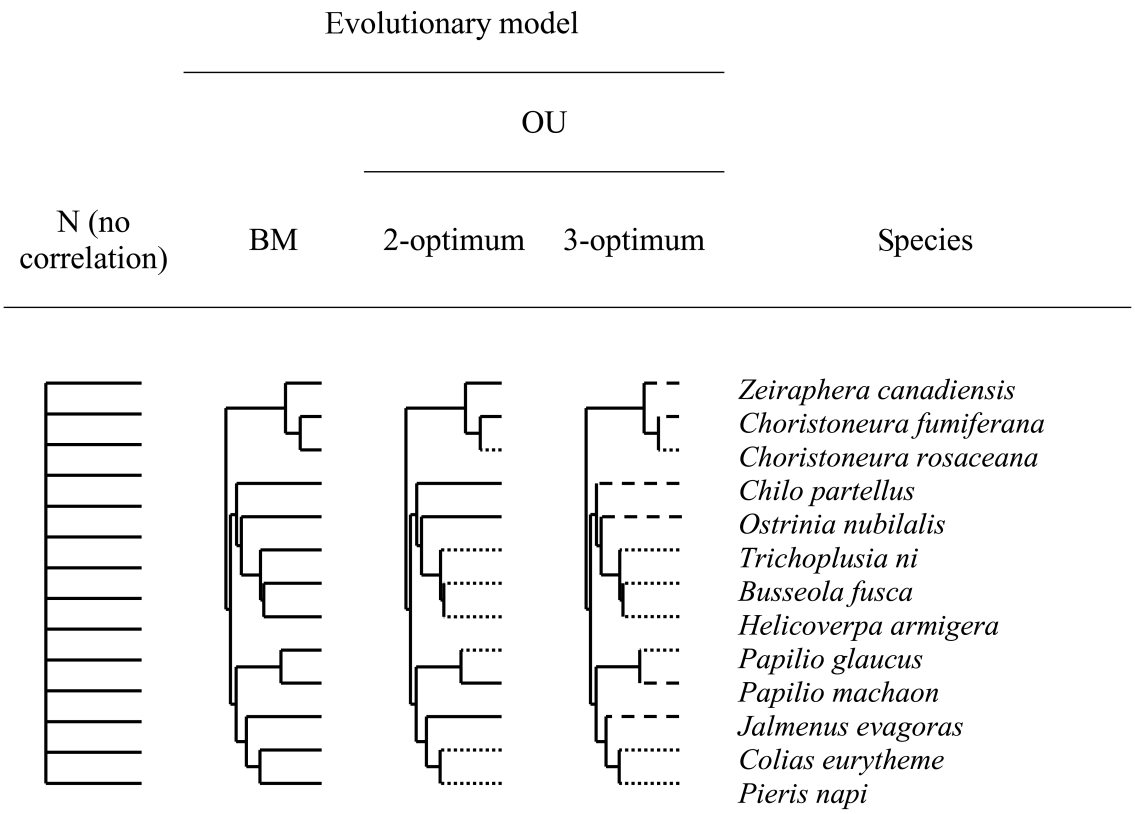



**Figure A2:** Alternative phylogenetic models for effect size evolution: having no phylogenetic correlations (N), having phylogenetic correlations based on Brownian motion (BM), and using an Ornstein-Uhlenbeck process (OU) having two and three hypothesized adaptive optima. The dotted and dashed lines indicate the start and end of the primary optimum for a particular mating system, and the solid lines indicate the ancestral optimum. In the two-optimum model, the monandrous species have evolved under the primary optimum, whereas polyandrous species have the ancestral optimum; this model is equivalent to a model that hypothesizes the opposite, where monandrous species have the ancestral state. Finally, the three-optimum model hypothesized that both monandrous and polyandrous species (*dashed lines*) are separate primary optima that are derived from shared but unknown ancestral optimum. Note that an OU model with one optimum is a BM model (BM ~ OU; seefig. A1).

**Table A1**

Lepidopteran species synthesized in Torres-Vila and Jennions's (2005) meta-analysis

| Species | Family | Type | Mating system | Hedges's $d$ | $\sigma^2(d)$ | COII accession no. |
|---|---|---|---|---|---|---|
| *Busseola fusca* | Noctuidae | H | P | .469 | .082 | AY320474 |
| *Chilo partellus* | Crambidae | H | M | −.041 | .052 | AY320482 |
| *Choristoneura fumiferana* | Tortricidae | H | M | .137 | .048 | L19098 |
| *Choristoneura rosaceana* | Tortricidae | H | P | 1.028 | .037 | L19099 |
| *Colias eurytheme* | Pieridae | R | P | 1.013 | .205 | AF044024 |
| *Helicoverpa armigera* | Noctuidae | H | P | .304 | .128 | DQ059302 |
| *Jalmenus evagoras* | Lycaenidae | R | M | .366 | .071 | DQ456502 |
| *Ostrinia nubilalis* | Crambidae | H | M | .359 | .017 | AF321880 |
| *Papilio glaucus* | Papilionidae | R | P | .232 | .044 | EF126474 |
| *Papilio machaon* | Papilionidae | R | M | .251 | .155 | AY457593 |
| *Pieris napi* | Pieridae | R | P | 1.169 | .255 | AF170861 |
| *Trichoplusia ni* | Noctuidae | H | P | .263 | .075 | AB158623 |
| *Zeiraphera canadiensis* | Tortricidae | H | M | .016 | .067 | DQ241506 |

**Note:** These species are grouped as moths (Heterocera) or butterflies and skippers (Rhopalocera) and also by mating system (polyandrous or monandrous). Also presented are the original effect size data ($\delta$ = Hedges's $d$) and variances $\sigma^2(d)$ from Torres-Vila and Jennions (2005) and the GenBank accession numbers for phylogenetic construction.

**Table A2**

Results from integrating phylogenetic correlations into Torres-Vila and Jennions (2005) meta-analysis on the mating success of females mated with virgin or nonvirgin males

| | | Pooled effect size | | | Homogeneity test | | | Evolutionary parameters | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| GLS evolutionary model | $k$ | $\bar{\delta}$ | UCI | LCI | $Q_H$ | df | $p$ | $\lambda$ | $\beta$ | $\eta$ | AIC |
| All studies: | | | | | | | | | | | |
| N | 13 | .374 | .243 | .506 | 23.3 | 12 | <.001 | .0000 | . . . | . . . | 13.58 |
| BM ($\lambda$ = ML) | 13 | .374 | .243 | .506 | 23.3 | 12 | <.001 | .0001 | . . . | . . . | 15.58 |
| BM ~ OU ($\beta$ = 0, $\lambda$ = 1) | 13 | .398 | .228 | .568 | 63.4 | 12 | <.001 | 1.0000 | .0 | .53 | 25.99 |
| OU ($\beta$ = ML) | 13 | .374 | .243 | .506 | 23.3 | 12 | <.001 | . . . | 74.0 | 4.59 | 17.58 |
| By mating system: | | | | | | | | | | | |
| Monandrous (N) | 7 | .225 | .056 | .394 | 3.5 | 6 | .744 | . . . | . . . | . . . | 10.42 |
| Polyandrous (N) | 6 | .604 | .395 | .814 | 12.1 | 5 | .033 | | | | |
| Monandrous (BM ~ OU, $\beta$ = 0, $\lambda$ = 1) | 7 | .171 | –.015 | .356 | 3.6 | 6 | .729 | 1.0000 | .0 | .37 | 18.97 |
| Polyandrous (BM ~ OU, $\beta$ = 0, $\lambda$ = 1) | 6 | .742 | .413 | .840 | 11.7 | 5 | .039 | | | | |
| 2-optimum OU ($\beta$ = ML): | | | | | | | | | | | |
| Monandrous | 7 | .223 | .049 | .397 | 3.5 | 6 | .744 | . . . | 100.0[a] | 4.24 | 16.39 |
| Polyandrous | 6 | .605 | .395 | .815 | 12.1 | 5 | .033 | | | | |
| 3-optimum OU ($\beta$ = ML): | | | | | | | | | | | |
| Monandrous | 7 | .206 | .022 | .385 | 3.4 | 6 | .757 | . . . | 77.7 | 3.63 | 18.14 |
| Polyandrous | 6 | .605 | .395 | .814 | 12.1 | 5 | .033 | | | | |
| Unknown ancestor | . . . | 6.709 | . . . | . . . | . . . | . . . | . . . | | | | |

**Note:** Here $k$ is the review sample size, and parentheticals for generalized least squares (GLS) models are the values of evolutionary parameters: phylogenetic signal, selection, and drift. ML indicates that these parameters were optimized via maximum likelihood. See figure 4 for definitions and details of the different evolutionary hypotheses tested.

[a] Here the ML algorithm failed to estimate selection and indicates the maximum range of values explored during optimization.

**Table A3**
The overall weights (**W**) of different taxonomic ranks on the mean effect size pooled across 13 studies (see box A1)

| | | % taxonomic weight (**W**) on pooled effect sizes ($k = 13$) | | |
|---|---|---|---|---|
| Taxonomic rank | $k$ | $\bar{\mu}$ | $\bar{\delta}$ | $\bar{\delta}^{P(BM)}$ |
| Division: | | | | |
| Heterocera | 8 | 61.6 | 76.5 | 77.2 |
| Rhopalocera | 5 | 38.4 | 23.5 | 22.8 |
| Family: | | | | |
| Crambidae | 2 | 15.4 | 35.2 | 48.6 |
| Lycaenidae | 1 | 7.6 | 6.3 | 7.5 |
| Noctuidae | 3 | 23.1 | 13.0 | 7.4 |
| Papilionidae | 2 | 15.4 | 13.2 | 13.9 |
| Pieridae | 2 | 15.4 | 4.0 | 1.4 |
| Tortricidae | 3 | 23.1 | 28.3 | 21.2 |

**Note:** These percentages are derived from a raw average ($\bar{\mu}$) of effect sizes (where each study has equal weight), meta-analysis ($\bar{\delta}$; weighted by sample precision), and phylogenetically corrected meta-analysis ($\bar{\delta}^{P(BM)}$; weighted by sample precision and phylogenetic correlations based on a BM model of evolution assuming $\lambda = 1$). BM = Brownian motion.

## Literature Cited in Appendix


Becker, B. J. 1992. Using results from replicated studies to estimate linear models. Journal of Educational Statistics 17:341–362.

Cheung, S. F., and D. K.-S. Chan. 2004. Dependent effect sizes in meta-analysis: incorporating the degree of interdependence. Journal of Applied Psychology 89:780–791.

Diniz-Filho, J. A. F. 2001. Phylogenetic autocorrelation under distinct evolutionary processes. Evolution 55: 1104–1109.

Gage, M. J. G. 1994. Associations between body size, mating pattern, testis size and sperm lengths across butterflies. Proceedings of the Royal Society B: Biological Sciences 258:247–254.

Glesser, L. J., and I. Olkin. 1994. Stochastically dependent effect sizes. Pages 339–355 *in* H. Cooper and L. V. Hedges, eds. Handbook of research synthesis. Russell Sage Foundation, New York.

Honěk, A. 1993. Intraspecific variation in body size and fecundity in insects: a general relationship. Oikos 66: 483–492.

Larkin, M. A., G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, et al. 2007. ClustalW and ClustalX version 2. Bioinformatics 23:2947–2948.

Marín-Martínez, F., and J. Sánchez-Meca. 1999. Averaging dependent effect sizes in meta-analysis: a cautionary note about procedures. Spanish Journal of Psychology 2:32–38.

Nylin, S., K. Nyblom, F. Ronquist, N. Janz, J. Belicek, and M. Källersjö. 2001. Phylogeny of *Polygonia*, *Nymphalis* and related butterflies (Lepidoptera: Nymphalidae): a total-evidence analysis. Zoological Journal of the Linnean Society 132:441–468.

Posada, D., and K. A. Crandall. 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics 14: 817–818.

Swofford, D. L. 2003. PAUP*: phylogenetic analysis using parsimony (*and other methods). Version 4.0b10. Sinauer, Sunderland, MA.

Thompson, S. G., and J. P. T. Higgins. 2002. How should meta-regression analyses be undertaken and interpreted? Statistics in Medicine 21:1559–1573.

Torres-Vila, L. M., and M. D. Jennions. 2005. Male mating history and female fecundity in the Lepidoptera: do male virgins make better partners? Behavioral Ecology and Sociobiology 57:318–326.