

phyloMeta: a program for phylogenetic comparative analyses with meta-analysis

Marc J. Lajeunesse^{1,2}¹Department of Integrative Biology, University of South Florida, Tampa, FL 33620 and ²National Evolutionary Synthesis Center, Durham, NC 27705-4667, USA

Associate Editor: David Posada

ABSTRACT

Summary: *phyloMeta* is an easy to use console program for integrating phylogenetic information into meta-analysis. It is designed to help ecologists, evolutionary biologists and conservation biologists analyze effect size data extracted from published studies in a comparative phylogenetic context. This software estimates phylogenetic versions of all the traditional meta-analytical statistics used for: pooling effect sizes with weighted regressions; evaluating the homogeneity of these effect sizes; performing moderator tests akin to ANOVA style analyses; and analyzing data with fixed- and random-effects models. *phyloMeta* is developed in C/C++ and can be used via command line in MS Windows environments.

Availability: *phyloMeta* can be obtained freely as an executable on the web at <http://lajeunesse.myweb.usf.edu/publications>

Contact: lajeunesse@usf.edu

Received on May 25, 2011; revised on July 18, 2011; accepted on July 19, 2011

1 INTRODUCTION

The statistics of meta-analysis consist of quantifying experimental outcomes with effect sizes, assigning a weight to each effect size that is inversely proportional to its variance, and then averaging these weighted effect sizes to provide an overall synthesis of multiple published studies (Hedges and Olkin, 1985). Ecologists and evolutionary biologists rely heavily on these statistics to provide research summaries and to test hypotheses on moderators of experimental outcomes—often pooling research from different taxa to achieve these goals (Lajeunesse, 2010).

However, pooling effect size data from multiple taxa can potentially bias the outcome of meta-analysis because of the non-independence of effect sizes due to the shared phylogenetic history of taxa (Lajeunesse, 2009). *phyloMeta* fills a gap in the statistical tools needed to assess and account for this form of non-independence when performing a meta-analysis on ecological research. This software reports all the familiar statistics of meta-analysis (e.g. pooled effect sizes paired with 95% confidence intervals and homogeneity tests) along with phylogenetic versions of these statistics (as reported in Lajeunesse, 2009; Lajeunesse *et al.*, 2011). My intention with this software is to overcome the current limitations of widespread meta-analytical software such as metaWin (Rosenberg *et al.*, 2000) and the emerging metafor package in R (Viechtbauer, 2010). These tools are unable to integrate the covariances necessary to account for phylogenetic correlations among effect sizes from related taxa.

2 STATISTICAL METHODS

The statistical methods reported by *phyloMeta* are based on the generalized least squares (GLS) approach to meta-analysis described by Lajeunesse (2009) and Lajeunesse *et al.* (2011). These include: weighted regressions to pool effect sizes paired with confidence intervals; homogeneity statistics (*Q*-tests) used for evaluating whether effect size should be pooled without bias; conservative adjustments of homogeneity tests should the phylogeny contain polytomies; analyses for both fixed- and random-effects models; ANOVA style moderator analyses, based on partitioning within- and between-group homogeneity tests; calculation of model selection criteria (AIC scores) to evaluate the relative fit of competing regression models. These methods are modifications of the regression models of Hedges and Olkin (1985) and Hedges (1992).

Briefly for example, in a fixed-effects meta-analysis of *k* number of effect sizes, the following GLS regression equation in matrix notation is used to pool all these *k* effects into an overall average effect $\bar{\delta}$:

$$\bar{\delta} = (\mathbf{X}^T \mathbf{W}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^{-1} \mathbf{E}, \quad (1)$$

where \mathbf{X} is a $k \times 1$ column vector of ones, \mathbf{E} is a $k \times 1$ column vector of the effect sizes, and \mathbf{W} is the variance–covariance matrix that defines how effect sizes are to be weighted and how they relate to one another, and is defined as:

$$\mathbf{W} = \mathbf{D} \mathbf{P} \mathbf{D}. \quad (2)$$

This variance–covariance matrix contains the weights due to sampling error as required by meta-analysis (these are defined in \mathbf{D} , a $k \times k$ diagonal matrix of the SDs of each effect size) and the phylogenetic correlations. The phylogenetic correlations \mathbf{P} is a $k \times k$ diagonal matrix containing the shared branch length distances of each taxa on a phylogenetic tree. See Rohlf (2001) for an example of this correlation matrix. In a traditional meta-analysis, \mathbf{P} is simply defined as an Identity matrix—indicating that effect sizes are independent (Lajeunesse, 2009). Currently, *phyloMeta* (v. 1.3 as of 05/12/2011) assumes, as in Felsenstein's (1985) phylogenetically independent contrasts, that branch lengths are proportional to the expected change (e.g. a Brownian motion model of evolution). Future releases will include diagnostics of this assumption [as described in Lajeunesse (2009)].

3 IMPLEMENTATION

The software requires no installation and little input from the user other than a text file containing the meta-analytical data and second

text file of the hypothesized phylogeny in NEWICK format. This NEWICK tree is converted into a correlation matrix (**P**) that is integrated into the GLS weighted regression models (Lajeunesse, 2009). Users are prompted for the names of each file on each run, and all analyses (e.g. *Q*-tests, fixed- and random-effects models) are executed automatically. As a console program, *phyloMeta* also accepts these two file names as command line arguments; this facilitates the calling of *phyloMeta* in R for rapid visualization of results or for simulation analyses.

After each run, *phyloMeta* saves all the results in a single text file. Finally, *phyloMeta* can analyze data based on any effect size metric (e.g. Hedges' *d*, log response ratio, correlations) as long as these data are paired with valid variance estimates. By definition, variances are required for meta-analysis and are used as weights in a weighted regression (Hedges, 1992).

I currently compile *phyloMeta* with an early version of MS VC++ (6.0) to avoid use of the Windows .Net framework. This allows for a broad portability among early and later releases of Microsoft Windows (i.e. NT, XP, Vista, Windows 7) without any additional installation from the user.

4 PUBLISHED APPLICATIONS

Several published meta-analyses have already used a beta version of *phyloMeta* for pooling and testing hypotheses with effect size data in a phylogenetic context. I highlight below a few of these studies to emphasize the potential broad applications of *phyloMeta* for ecological synthesis. These studies also serve as illustrative examples on how to use and report the output of *phyloMeta*, and provide important information on how to acquire or construct phylogenetic trees for meta-analysis.

For example, Carmona *et al.* (2011) pooled correlations among 40 plant species to test the importance of secondary metabolites over other plant characteristics (e.g. life history traits) as predictors of anti-herbivore defense. They found no significant differences between the traditional and phylogenetically independent meta-analyses, and chose to report only the latter. Their phylogenetic tree of plants was estimated using phylomatic (Webb and Donoghue, 2004) with internal branch length distances based on divergence times found on the timetree website (Hedges and Kumar, 2009). Meunier *et al.* (2011) also pooled correlations with *phyloMeta*, but here focused on testing hypotheses of the adaptive function of melanin-based coloration in 26 bird species. In this meta-analysis, they used a molecular phylogeny estimated from gene sequence data publically available online at GenBank. In another meta-analysis, results from traditional and phylogenetically independent meta-analysis were compared among studies based on 87 plant species to evaluate correlations between fitness and flowering synchrony (Munguía-Rosas *et al.*, 2011). Finally, Ord *et al.* (2011) broadly surveyed published research on amphibians, birds, fish, mammals and insects to test how animals respond to encounters with conspecific or heterospecific species in their environments. For their meta-analysis, they used a composite phylogenetic tree with

a topology derived from public web databases (e.g. Encyclopedia of Life, *eol.org*) and published phylogenies.

5 CONCLUSION AND PROSPECTUS

phyloMeta is a stand-alone software for assessing and accounting of phylogenetic non-independence of meta-analytical data. I am continuously developing this software and custom versions can be made upon request for specific research requirements.

Further developments of the software will include optimizations via maximum likelihood for different models of evolution [e.g. Ornstein–Uhlenbeck processes; Lajeunesse (2009)], and integration of diagnostics useful for evaluating phylogenetic bias and publication bias in effect size data.

ACKNOWLEDGEMENTS

I thank two anonymous reviewers for comments, and S. Chamberlain, S. Hovick and the Rice University EEB Department for helpful feedback on an early version of *phyloMeta*.

Funding: College of Arts and Sciences at the University of South Florida; National Science Foundation grant to the National Evolutionary Synthesis Center (EF-0423641).

Conflicts of Interest: none declared.

REFERENCES

- Carmona, D. *et al.* (2011) Plant traits that predict resistance to herbivores. *Funct. Ecol.*, **25**, 358–367.
- Felsenstein, J. (1985) Phylogenies and the comparative method. *Am. Nat.*, **125**, 1–15.
- Hedges, L.V. (1992) Meta-analysis. *J. Educ. Stat.*, **17**, 279–296.
- Hedges, S.B. and Kumar, S. (2009) *The Timetree of Life*. Oxford University Press, New York.
- Hedges, L.V. and Olkin, I. (1985) *Statistical Methods for Meta-analysis*. Academic Press, NY.
- Lajeunesse, M.J. (2009) Meta-analysis and the comparative phylogenetic method. *Am. Nat.*, **174**, 369–381.
- Lajeunesse, M.J. (2010) Achieving synthesis with meta-analysis by combining and comparing all available studies. *Ecology*, **91**, 2561–2564.
- Lajeunesse, M.J. *et al.* (2011) Phylogenetically-independent meta-analysis. In Koricheva, J. *et al.* (eds) *Handbook of Meta-Analysis in Ecology and Evolution*. Princeton University Press, Princeton, New Jersey, USA (in press).
- Meunier, J. *et al.* (2011) Eumelanin-based coloration and fitness parameters in birds: a meta-analysis. *Behav. Ecol. Sociobiol.*, **65**, 559–567.
- Munguía-Rosas, M.A. *et al.* (2011) Meta-analysis of phenotypic selection on flowering phenology suggests that early flowering plants are favoured. *Ecol. Lett.*, **14**, 511–521.
- Ord, T.J. *et al.* (2011) Contrasting theory with the empirical data of species recognition. *Evolution* (in press).
- Rohlf, F.J. (2001) Comparative methods for the analysis of continuous variables: geometric interpretations. *Evolution*, **55**, 2143–2160.
- Rosenberg, M.S. *et al.* (2000) *MetaWin: Statistical Software for Meta-Analysis*. Sinauer, Sunderland, MA.
- Viechtbauer, W. (2010) Conducting meta-analyses in R with the metafor package. *J. Stat. Software*, **36**, 1–48.
- Webb, C.O. and Donoghue, M.J. (2004) Phylomatic: tree assembly for applied phylogenetics. *Mol. Ecol. Notes*, **5**, 181–183.