
Phylogenetic Nonindependence and Meta-analysis

*Marc J. Lajeunesse, Michael S. Rosenberg,
and Michael D. Jennions*

AN IMPORTANT STATISTICAL ASSUMPTION of meta-analysis is that effect sizes are independent (Landman and Dawes 1982, Hedges and Olkin 1985, Gleser and Olkin 1994). This statistical independence means that the collection of effect sizes pooled in a meta-analysis does not have a correlated structure, and that each effect size (or sample) represents an independent piece of information. There are several reasons why a data set might have a correlated structure—for example, when multiple effect sizes are extracted from a single experiment or from different time points throughout a study. These forms of nonindependence are reviewed in Chapter 16. Here we focus on the statistical and conceptual issues arising from the nonindependence that emerges when meta-analysis is based on research from multiple species, which is nearly always the case in ecology and evolutionary biology.

Pooling research from multiple species can be a problem for ecological and evolutionary meta-analysis because species form a nested hierarchy of phylogenetic relationships. This shared phylogenetic history can introduce a correlated structure to effect size data because studies on closely related species may yield similar outcomes, and therefore similar estimates of effect sizes. This similarity is the product of shared (i.e., phylogenetically conserved) morphological, physiological, or behavioral characteristics (Harvey and Pagel 1991). Given that an explicit goal of many ecological and evolutionary meta-analyses is to include research from a diversity of taxa and to generalize across a broad range of species, there is the potential for bias should the literature synthesized be affected by the shared ancestry of taxa (Møller and Thornhill 1998; Verdú and Traveset 2004, 2005; Adams 2008; Lajeunesse 2009).

In this chapter, we describe statistical methods to account for phylogenetic nonindependence of species when pooling and testing for homogeneity of effect sizes. We also describe a method that compares the results of a traditional and a phylogenetically independent meta-analysis to evaluate which approach was more effective at explaining variation in research outcomes. Using these methods, we provide a worked example of a meta-analysis on trade-offs among plant antiherbivore defenses. Finally, we end the chapter with a discussion on approaches for collating phylogenetic information for meta-analysis.

The “apples and oranges problem”: It’s literal for ecology

The “apples and oranges problem” has a long history in meta-analysis and is a criticism of mixing studies with different conceptual and operational definitions; or more generally, a criticism

of mixing studies belonging to different groups (e.g., different experimental designs, different measures of outcome; Wolf 1986, Lynn 1989). The analogy here identifies that apples differ from oranges, and thus aggregating them is not appropriate or meaningful. The “apples and oranges problem” is not a valid criticism because the rationale for aggregating particular studies is relative to the focus of the review; if the focus is on apples then including oranges is inappropriate, but if the focus is on fruit, then grouping apples and oranges is meaningful. This distinction is important because it recognizes that, conceptually, objects form a hierarchy; apples and oranges are fruit because they share similar characteristics (e.g., in an informal sense they are edible, sweet, and have seeds). This concept of hierarchy should also resonate for all ecologists and evolutionary biologists because all species form a nested hierarchy of phylogenetic relationships. Apples and oranges have similar characteristics because of their shared ancestry among flowering plants (angiosperms). Considering the phylogenetic history of taxa is thus important for ecological and evolutionary meta-analyses because quantitative reviews often seek to generalize as broadly as possible by pooling or “mixing” all studies on a given conceptual topic.

But how should meta-analysts approach the problem of mixing research from different taxa? We believe they should first collate all the studies relevant to the conceptual topic under study—irrespective of whether they are based on birds, beetles, or nematodes. With these data, reviewers are then able to assess publication or research biases of taxa (e.g., if there is significant overrepresentation of beetle research over that of other taxa; see Chapter 14), conduct sensitivity analyses to determine whether the inclusion of certain taxa can bias results, and determine whether the phylogenetic history of these taxa is an issue for meta-analysis. An important benefit of collating all studies is the increased sample size of the review—large samples sizes decrease the likelihood of review type I and type II errors (Lajeunesse and Forbes 2003; Chapter 22) and allow for greater representation among moderator groupings essential for hypothesis testing.

Why is phylogenetic history an issue for meta-analysis?

Similarity of traits due to shared ancestry is a problem for meta-analysis because it violates two assumptions. First, that data are drawn from independent samples; this assumption is violated because effect sizes derived from closely related species may be similar in magnitude or direction due to their shared evolutionary history. Second, that effect sizes are sampled from a population that has a normal distribution with an expected variance. However, when effect sizes have a phylogenetic (correlated) structure, different variances are expected because lineages within phylogenies may have evolved at different rates (Harvey and Pagel 1991). Violating these assumptions increases the likelihood of a type I error (false positive) (Harvey and Pagel 1991). Below we review both assumptions in detail because they form the basis for the statistical methods we outline in this chapter.

First statistical violation: Lack of independence

Shared ancestry among species violates the assumption of nonindependence in meta-analysis because it can result in effect size data with a correlated structure—that is, some groupings of data points are more likely to have values that exceed (or are less than) the average value due to a shared error structure. Here these shared correlations are due to the nonindependent origins of traits and characteristics of related taxa. This phylogenetic conservatism (sometimes described as phylogenetic inertia; Wilson 1975) can emerge in effect size data in many ways. For example, effect size data quantifying the magnitude of difference or the correlation between characteristics (Table 17.1) can be nonindependent if the characters themselves are phylogenetically

TABLE 17.1. Variance estimates $\sigma^2(\delta)$ of effect size metrics (δ) that can be used with phylogenetically independent meta-analysis. Effect sizes and their variances are used to define the variance-covariance matrix (\mathbf{V}). This matrix is then applied to GLS models that estimate the phylogenetically independent weighted mean of a collection of studies. Here, effect sizes are estimated from a study that has a sample size (n), and control (C) and treatment (T) means (\bar{X}). Finally, the log odds ratio is based on count data of members belonging to groups A or B. Examples of these effect sizes are found in Van Zandt and Mopper (1998) for Hedges' d and $\ln R$, Koricheva et al. (2004) for correlation coefficients, and Beirinckx et al. (2006) for $\ln OR$.

effect size metric	effect size	effect size variance
Hedges' d^\dagger	$d = \frac{\bar{X}_T - \bar{X}_C}{s} \left(1 - \frac{3}{4(n_T + n_C) - 9} \right)$	$\frac{n_T + n_C}{n_T n_C} + \frac{d^2}{2(n_T + n_C)}$
log response ratio ($\ln R$)	$\ln R = \ln \left(\frac{\bar{X}_T}{\bar{X}_C} \right)$	$\frac{s_T^2}{n_T \bar{X}_T^2} + \frac{s_C^2}{n_C \bar{X}_C^2}$
Pearson's product-moment correlation coefficient (r)	$r = r$	$\frac{(1 - r^2)^2}{n - 2}$
Fisher's z transformation of r	$z = \frac{1}{2} \ln \left(\frac{1 + r}{1 - r} \right)$	$\frac{1}{n - 3}$
log odds ratio ($\ln OR$)	$\ln OR = \ln \left(\frac{n_{A_C} n_{B_T}}{n_{A_T} n_{B_C}} \right)$	$\frac{1}{n_{A_C}} + \frac{1}{n_{A_T}} + \frac{1}{n_{B_C}} + \frac{1}{n_{B_T}}$

[†]Hedges' d uses the pooled standard deviation (s) between the control and treatment; this is calculated as $s = ((n_T - 1)s_T^2 + (n_C - 1)s_C^2) / [n_T + n_C - 2]^{1/2}$.

conserved. Body size is an example of a trait that is often used in experimentation as a surrogate for fitness; but for many animals, closely related species or whole lineages may share similar sizes. Another way effect size data could be phylogenetically correlated is when the effect sizes themselves are phylogenetically conserved. An example is when the magnitude of difference or correlations between traits in body size occurs among animals that are sexually dimorphic.

Phylogenetic conservatism can generate effect size data with phylogenetic structure where studies based on related species yield very similar experimental outcomes (e.g., effect sizes). Typically, to statistically account for this source of dependence, researchers use Felsenstein's (1985) phylogenetically independent contrasts (PIC), which is a statistical method in comparative biology used for cross-species generalizations across traits. It transforms trait data into a set of contrasts (pairwise differences between trait values) with zero correlations. Here, although the absolute values of these data may be conserved, the pairwise differences between traits are not (Harvey and Pagel 1994). These contrasts are then analyzed using linear regression to evaluate whether traits share a correlated evolutionary history (for further details, see Martins and Garland 1991). Unfortunately, PICs are not practical for meta-analysis data because each effect size has a different variance and this violates the significance testing and variance estimating procedures of traditional regression models (Hedges 1994). This difference in variances among effect sizes is due to studies having different sample sizes (Hedges and Olkin 1985). Fortunately, PICs are special case of generalized least squares (GLS) models that can include phylogenetic information (Garland and Ives 2000, Rohlf 2001). Here a GLS model can be modified to fit the specific significance testing and variance estimation procedures for meta-analysis.

Second statistical violation: Heterogeneity of variances

Effect sizes with a phylogenetic structure violate the second statistical assumption that data are sampled from a normal distribution with an expected variance (Hedges and Olkin 1985). Variance here is defined as the rate of evolutionary change, and effect size data based on different taxa could have different expected variances because they belong to lineages evolving at different rates with varying divergence times (Harvey and Pagel 1991). Typically, comparative analyses account for this form of heterogeneity in variances by making assumptions on how lineages evolve (Felsenstein 1985, Martins and Garland 1991, Pagel 1997). For example, PICs assume that evolution proceeds as a Brownian motion (BM) process (e.g., random drift), where the expected variance of change is constant throughout the phylogeny; fitting this model of evolution is achieved by standardizing all the contrasts with the square root of the sum of all phylogenetic branch lengths (Martins and Garland 1991). In terms of GLS modeling, fitting a BM model of evolution is straightforward. Simply assume that the phylogenetic correlations used to define the phylogenetic structure of effect sizes are linearly related with the relative time since the divergence of each taxa. The following sections outline in more detail how to fit this evolutionary model into meta-analysis. Note also that although many of the statistics below are described as being “phylogenetically independent,” this is only the case if the assumption of Brownian motion evolution holds (e.g., the phylogenetic branch lengths are proportional to the expected evolutionary change). Several diagnostics have been developed to estimate whether data fit this evolutionary model (e.g., Pagel’s lambda; Pagel 1997), and these should be used with many of the statistics presented below to assess violations of this evolutionary assumption (Lajeunesse 2009).

STATISTICAL METHODS FOR PHYLOGENETICALLY INDEPENDENT META-ANALYSIS

Several comparative phylogenetic methods have been developed to account for phylogenetic nonindependence of species (Felsenstein 1985, Cheverud et al. 1985, Maddison 1990, Grafen 1989, Martins and Garland 1991, Pagel 1997). Felsenstein’s (1985) phylogenetically independent contrasts (PIC) remains the most widespread; however, alternative methods applying the generalized least squares (GLS) theory are increasing in popularity (Pagel 1994). In this chapter, we apply this GLS framework to account for phylogenetic history because meta-analysis can also be modeled under the GLS family of statistics (for further details, see Lajeunesse 2009). The advantage of using a combined statistical framework is the full flexibility to apply different statistical models to effect size data. The following methods can also be applied to any effect size metric (Table 17.1), but require that the variance of each effect size and a hypothesized phylogeny (how each effect size is correlated) are known.

Pooling effect sizes with phylogenetic correlations

Using matrix notation, we can pool K number of effect sizes (δ) into a weighted-mean (pooled) effect size ($\bar{\mu}_+$) with this GLS regression equation:

$$\bar{\mu}_+ = (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} \mathbf{Y}. \quad (17.1)$$

The 95% confidence intervals (CI) for $\bar{\mu}_+$ are calculated as follows:

$$95\% \text{ CI } [\bar{\mu}_+ \pm 1.96 \sqrt{(\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1}}]. \quad (17.2)$$

The components of Equations 17.1 and 17.2 are as follows. The superscripts T and $^{-1}$ indicate the transposition and inverse of a matrix, respectively. The effect sizes are contained in the $K \times 1$ column vector \mathbf{Y} . The averaging behavior of Equation 17.1 (i.e., pooling all effect sizes) is defined by \mathbf{X} , which is a $K \times 1$ column vector of ones and is known in GLS terms as the design matrix (Groß 2003). The most important feature of Equation 17.1 is the weighting matrix (\mathbf{W}), which is the inverse of the variance-covariance matrix (\mathbf{V}), such that $\mathbf{W} = \mathbf{V}^{-1}$. In traditional meta-analysis, \mathbf{V} is defined as a $K \times K$ matrix containing the effect size variances $\sigma^2(\delta)$ on its main diagonal, and zeros in all the remaining (off-diagonal) elements. When \mathbf{V} is defined this way, and applied to Equation 17.1, it will yield the same pooled effect size as the one generated by the traditional weighting method of meta-analysis described in Chapter 9.

To calculate a “phylogenetically independent” mean effect size, we need to integrate information in \mathbf{V} on how effect sizes vary together based on their correlated phylogenetic history. This is achieved by defining all the off-diagonal elements of \mathbf{V} with the phylogenetic covariance for each pair of effect sizes (for further details on the statistical background of this approach, see Hedges and Olkin 1985; Becker 1992; Gleser and Olkin 1994; Marín-Martínez and Sánchez-Meca 1999; Cheung and Chan 2004; and appendix of Lajeunesse 2009), such that the $i = 1, \dots, K$ rows and $j = 1, \dots, K$ columns of \mathbf{W} become

$$\mathbf{V}_{i,j} = \begin{cases} \sigma^2(\delta_i) & \text{when } i = j \text{ (on diagonal),} \\ \text{cov}(\delta_i, \delta_j) & \text{when } i \neq j \text{ (off diagonal),} \end{cases} \quad (17.3)$$

where each covariance is calculated as

$$\text{cov}(\delta_i, \delta_j) = P_{i,j} \sqrt{\sigma^2(\delta_i)} \sqrt{\sigma^2(\delta_j)}. \quad (17.4)$$

Here, information on shared ancestry between each i th and j th effect size is found in the phylogenetic correlation matrix \mathbf{P} . The elements of \mathbf{P} are defined by the shared internode branch-length distance between species on a phylogenetic tree. Figure 17.1 describes how this branch-length information is converted into a phylogenetic correlation, and Figure 17.2 illustrates how \mathbf{P} is then integrated into \mathbf{V} using Equation 17.4. This approach to defining \mathbf{P} forms the basis for many comparative methods based on the GLS framework (Rohlf 2001). For clarity, we now refer to $\bar{\mu}_+$ as the traditionally weighted pooled effect size, and $\bar{\rho}_+$ as the pooled effect size weighted by variance *and* phylogenetic correlations.

Finally, for all the phylogenetically independent analyses outlined in this chapter, only ultrametric trees (i.e., where all the tips of the phylogeny are aligned or contemporaneous; Figure 17.1) should be used to define \mathbf{P} . This is because Equation 17.4 assumes that the elements of \mathbf{P} are statistical correlations (i.e., with values ranging from zero to one) and not raw branch-length distances (Hedges and Olkin 1985). More precisely, \mathbf{P} should only contain elements on the main diagonal (i.e., species that are 100% correlated with themselves), and no off-diagonals should exceed one. When the tree is not ultrametric, such that each species has a different BL_{\max} (Fig. 17.2), then it is possible to generate a \mathbf{W} matrix that violates the statistical assumptions of GLS modeling (e.g., \mathbf{W} must be symmetric and positive definite; see Groß 2003). Using an ultrametric tree also meets the evolutionary assumption that phylogenetic branch lengths are proportional to the expected evolutionary change (see Rohlf 2001).

Phylogenetically independent homogeneity tests

Homogeneity statistics (Q) in meta-analysis test whether a collection of effect sizes share a common effect (i.e., $\delta_1 = \delta_2 = \dots = \delta_k$). A more thorough explanation and interpretation of Q is found in Chapter 9. The following GLS equation allows testing for homogeneity of effect sizes while accounting for phylogenetic correlations:

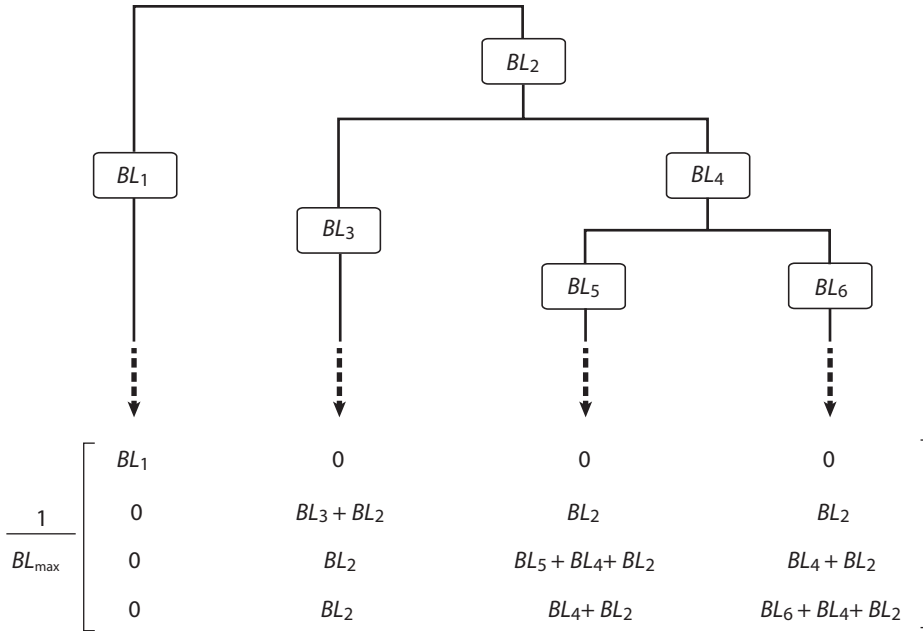


Figure 17.1. The phylogenetic correlation matrix **P** and how it is defined by the phylogenetic relationships of four species. Here, the internode branch lengths (*BL*) of the phylogeny are used to calculate the strength of correlations between species, and BL^{max} is the sum of all the internode branch-length distances from the root to tip of the phylogeny. The values within the matrix are the sum of the branch lengths of shared phylogenetic history between two taxa, divided by the depth of the tree (the maximum sum of branch lengths from the root to the tips). Thus, entry *ij* within the matrix represents the proportion of evolutionary history reflected in the tree that is shared by species *i* and *j*. By definition, the amount of history shared by a species with itself is 100%, and since the tree is ultrametric, the values on the diagonal will equal 1.

$$Q = \mathbf{Y}^T \mathbf{W} \mathbf{Y} - \bar{\rho}_+^T (\mathbf{X}^T \mathbf{W} \mathbf{X}) \bar{\rho}_+, \tag{17.5}$$

where again **W** is the inverse of **V** described in Equation 17.3. The significance of *Q* is tested against the critical value of a chi-square distribution (χ^2) with $K - 1$ degrees of freedom, such that if $Q \leq \chi^2_{K-1}$ then $\delta_1 = \delta_2 = \dots = \delta_K$. Microsoft Excel provides a function that calculates the *P*-value of this *Q*-test: =CHIDST(*Q*,*K*-1).

Polytomies in phylogenetic trees: A bias for homogeneity tests

Polytomies may influence type I error rates of homogeneity tests because they influence the degrees of freedom of analyses (Purvis and Garland 1993). The degrees of freedom are a measure of the number of independent pieces of information ($K - 1$ for a completely resolved tree); polytomies in phylogenetic trees cancel some of this independence because multiple lineages are specified with shared divergences. Polytomies are tree nodes that have more than two immediate descendants, and Madisson (1989) divides these into two types, one biological and the other statistical. The first type is referred to as a “hard” polytomy and depicts a true biological event where a group of sister taxa have diverged (speciated) simultaneously. Hard polytomies are most likely to arise due to explosive (rapid) speciation events, simultaneous fragmentation of populations, or introgressive hybridization and recombination events (Meinick and Hoelzer

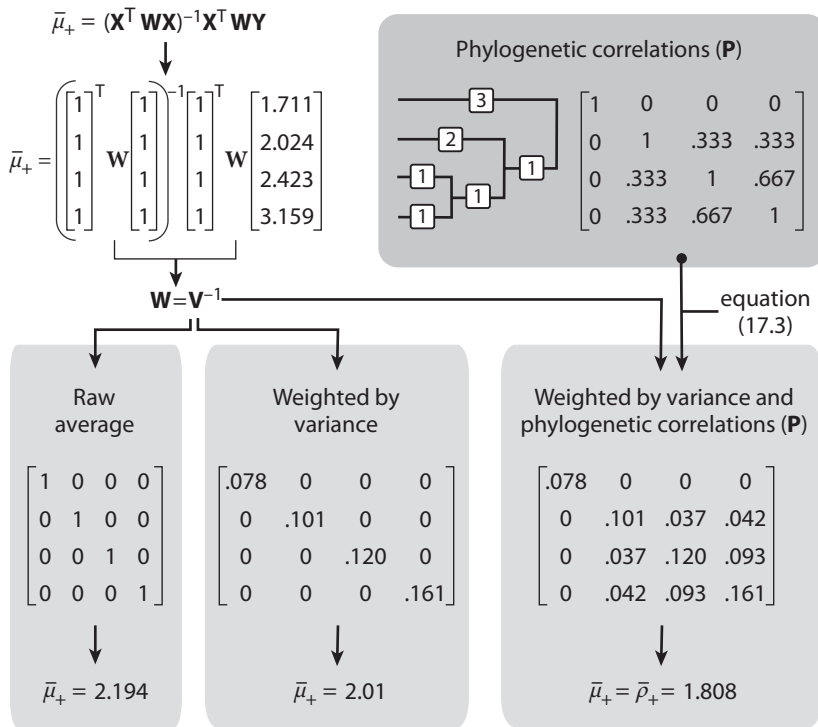


Figure 17.2. An illustration of applying three models of covariance matrices \mathbf{W} (light gray) to the GLS regression Equation 17.1 to generate a raw, weighted by variance (traditional meta-analysis or $\bar{\mu}_+$), and phylogenetically independent mean effect size ($\bar{\rho}_+$). Here $\mathbf{X} = [1 \ 1 \ 1 \ 1]^T$ is the design matrix, and $\mathbf{Y} = [1.711 \ 2.024 \ 2.423 \ 3.159]^T$ is the column vector of effect sizes. The phylogeny depicts how effect sizes are related (dark gray block) (effect sizes are aligned with the tips of the phylogeny), and \mathbf{P} is the phylogenetic correlation matrix of the shared internode branch-length distances from this phylogeny. These distances are in small font on the branches of the phylogeny.

1994). Walsh et al. (1999) describe an example of a hard polytomy, but more generally they appear to be uncommon. Polytomies more often occur because of statistical uncertainty when resolving relationships among taxa; these are known as “soft” polytomies. Statistical approaches to building phylogenetic trees (e.g., maximum likelihood) will generate soft polytomies when they are unable to statistically break up true speciation or bifurcation events among sister taxa.

The correct degrees of freedom for Q -tests should reflect this statistical imprecision in the phylogenetic tree. One approach is to conservatively bind the degrees of freedom to minimize type I errors. Purvis and Garland (1993) suggest a conservative df to be $K - p$. Here, p is the total number of polytomies in the phylogeny. Should the significance of Q be lost after correcting the degrees of freedom for polytomies, then all conclusions drawn from homogeneity tests should be considered as inconclusive given the current resolution of the phylogeny. Alternatively, Purvis and Garland (1993) suggest arbitrarily resolving each polytomy until the tree is fully bifurcated (no more polytomies), and then repeating this method to create a collection of phylogenies with alternative solutions. The results of applying each tree solution are then averaged to provide an overall synthesis. Unfortunately, GLS modeling summarizes phylogenetic

relationships using the \mathbf{P} correlation matrix. Here, arbitrary resolution of polytomies may not significantly change the shared branch-length distances between species because polytomies are resolved by setting the phylogenetic branch-length distance between intervening nodes as a very small distance (Rohlf 2001, Diniz-Filho 2001). Thus, GLS analyses may be insensitive to these resolutions if there are too few polytomies to resolve. However, when there are many polytomies, it is important to generate a broad sampling of tree solutions to avoid this issue. This approach is also useful when the internode branch-length distances are themselves arbitrarily set. This occurs when information on the topology of the phylogeny (i.e., the nodes on tree) is available, but intervening branch-length distances are not. Purvis and Garland (1993) outline multiple procedures for arbitrarily setting internode branch-length distances. Alternatively, missing internode branch-length distances can be simulated with different models of evolution (e.g., Brownian motion).

Integrating multiple studies of a single species

Including multiple studies of the same species requires only a slight modification to phylogenetically independent meta-analysis. These studies are important because they serve as species-level replicates, and their inclusion in meta-analyses can improve variance estimates. The challenge is to modify the GLS model such that these replicates retain their within-species study-level independence, while still accounting for the phylogenetic nonindependence that occurs when effect sizes are combined with other taxa. Figure 17.3 shows an example of how to modify the \mathbf{P} phylogenetic correlation for three species where data from one species are derived from two studies. Note that using the \mathbf{P} matrix in Figure 17.3 requires that the \mathbf{Y} column vector contains all the effect sizes of studies a , b , c_1 and c_2 (i.e., here the column vector rank becomes the number of effect sizes, not the number of species). This also applies for the design matrix \mathbf{X} with a column rank of K number of effect sizes. The degrees of freedom for this analysis are K number of taxa. Finally, this approach to integrating multiple studies is identical to pooling all the effect sizes and their variances of a single species before a phylogenetically independent meta-analysis (e.g., using Equation 17.1 and 17.2; Gurevitch and Hedges 1999); then, each species only has a single effect size estimate and variance for the final meta-analysis. These pooled effect sizes and variances are based on either a fixed- or random-effects meta-analysis.

Finally, we do not advocate treating these replicate studies as polytomies on the stem of a phylogeny for a given species. When treated this way, the whole group of replications will be

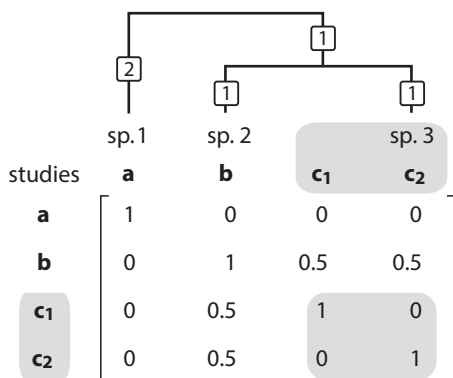


Figure 17.3. Example of a phylogenetic correlation matrix \mathbf{P} , including multiple studies (effect sizes) from a single species. Here, species 1 and 2 only have one effect size (**a** and **b**, respectively), whereas species 3 has two (**c₁** and **c₂**). The gray block emphasizes the correlation submatrix of species 3, which is parameterized to account for the between-study independence of studies **c** and **d**. These two studies share the same relative phylogenetic correlation with species 2. Finally, the small font numbers on the branches of the phylogeny indicate their length.

significantly underweighted in the final meta-analysis. This underweighting occurs because polytomies are often set to have a very short branch-length distance, and this small branch length will yield a very strong phylogenetic correlation ($P_{i,j} \rightarrow 1$). Here, the individual variation for each replication is lost because these strong phylogenetic correlations penalize the weighting of each effect size in the final analysis. Thus, in effect, the meta-analysis would generate the same result as if these replications were assumed to have equal variances.

Testing for a phylogenetic signal among effect sizes

Any meta-analysis for which the phylogenetic history is considered a source of bias should evaluate the degree to which effect size data show a phylogenetic signal. Detecting a phylogenetic signal is a test for phylogenetic conservatism. This test quantifies the degree to which related species tend to be more similar than distantly related species. When data show little phylogenetic conservatism, they are considered evolutionarily labile, and analysis of these data is assumed to be the least biased by phylogenetic nonindependence. Unfortunately, the application of information regarding phylogenetic signals is contentious, and it has been advised that if a phylogenetic signal is not detected, then researchers can skip phylogenetically controlled analyses (Westoby et al. 1995a, 1995b, 1995c, Ricklefs and Starck 1996, Björklund 1997). However, we believe testing for a signal should be used to test hypotheses rather than as a guide for directing (potentially informative) analyses; remember, the absence of evidence is not evidence of absence. The potential for error in these statistics is due to the continued statistical difficulty in detecting phylogenetic signals from phylogenies with few species (Martins 2000, Freckleton et al. 2002).

Several statistical tests for phylogenetic signals have already been formalized in the literature; these are Pagel's (1994, 1997) λ , and Blomberg et al.'s (2003) K (not to be confused with the K used as the review sample size of meta-analysis). These tests can be applied to effect sizes (Lajeunesse 2009), and are useful because they include calculations of 95% confidence intervals (CI) around λ or K . The interpretation of these statistics is straightforward; should λ or K be nonzero (i.e., the 95% CI does not overlap with zero), then closely related species tend to share similar effect sizes. This information can be used to assess the relative degree of phylogenetic bias in a meta-analysis, and to help evaluate whether "phylogenetically independent" analyses meet the important assumption of Brownian motion evolution.

Testing the fit of phylogenetic correlations

An alternative to evaluating a phylogenetic signal is to evaluate which statistical model, the traditional ($\bar{\mu}_+$) or a phylogenetically independent meta-analysis ($\bar{\rho}_+$), is better at explaining variation among effect size. What is needed to test model fit is the error sums of squares (SSE) of each meta-analysis; this is simply the homogeneity statistic Q (Equation 17.5; see also Hedges and Olkin 1985). With the homogeneity statistic forming the basis for model selection, Akaike's information criterion (AIC) is then used to compare each model. AIC scores are useful for model selection because they penalize models that (a) poorly fit effect sizes, and (b) use additional parameters to explain variation. AIC scores can be calculated as follows:

$$\text{AIC} = 2m - K \left[\ln \left(2\pi \frac{Q}{K} \right) + 1 \right] \quad (17.6)$$

where m is the number of parameters in the meta-analysis. Here, traditional meta-analysis will have an $m = 1$, whereas the phylogenetically independent meta-analysis will have $m = 2$. We penalize the phylogenetically independent meta-analysis because it uses an additional parameter

(phylogenetic correlations) to explain variation among effect sizes. Finally, the model that best fits the effect size data is the one with the lowest AIC score; for example, if $AIC_{\bar{\mu}_+} < AIC_{\bar{\rho}_+}$, then the traditional meta-analysis is the better fit. It should finally be noted that this approach to estimating AIC scores with Equation 17.6 may only be appropriate for comparing evolutionary models with equal K . Lajeunesse (2009) describes an alternative to calculating AIC using likelihood functions that may be more justifiable for complex model comparisons.

Alternative statistical approaches

Phylogenetically independent meta-analysis is a rapidly growing field of statistics and several alternative approaches to the one presented in this chapter have been proposed. One statistical approach was the use of phylogenetically informed simulations to evaluate the validity of meta-analysis (Verdú and Traveset 2004, 2005). Here, simulations are used to estimate phylogenetically independent critical values for the statistical tests of a weighted regression model (e.g., F -test; Garland et al. 1993). A limitation of this approach is that analyses use the significance tests (F -tests) and variance estimating procedures of weighted regression analyses. This is because effect size data do not meet the assumptions of conventional regression or ANOVA style analyses, given that individual effect sizes will have different variances due to their inverse relationship with sample size (Hedges and Olkin 1985, Hedges 1992). This difference in variance is the reason why the tools of statistical inference for meta-analysis differ from these conventional statistics and rely heavily on chi-square tests. Another limitation is that although this approach is only useful to evaluate bias in meta-analysis, it does not provide a solution to integrate phylogenetic information when pooling results or when testing for homogeneity.

Adams (2008) was the first to propose a method to integrate phylogenetic information when pooling effect size data. Here the effect size data (vector \mathbf{Y}) 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{X} and \mathbf{Y} with the inverse square root of the phylogenetic correlation matrix \mathbf{P} (Garland and Ives 2000; Groß 2003). These transformed vectors are then integrated in a traditional weighted regression model as follows:

$$\bar{\mu}_+^{Adams} = [(\mathbf{P}^{-1/2}\mathbf{X})^T \mathbf{W}(\mathbf{P}^{-1/2}\mathbf{X})]^{-1}(\mathbf{P}^{-1/2}\mathbf{X})^T \mathbf{W}(\mathbf{P}^{-1/2}\mathbf{Y}), \quad (17.7)$$

where \mathbf{W} contains the variances of each effect size on its main diagonal. However, Lajeunesse (2009) found that this approach can introduce a small bias when pooling effect sizes because the phylogenetic transformation has an effect of converting effect sizes into evolutionary units. This decreases the efficiency of the weights used to penalize studies with a large sampling error (Lajeunesse 2009). A more serious issue of this approach is that it applies traditional significance tests of comparative methods or weighted regression models to analyze effect size data. For example, Adams (2008) uses the t -test of a weighted regression to evaluate the significance of the pooled (phylogenetically independent) effect size. However, effect size data are not the same as the quantitative trait data typically analyzed with comparative methods, because each effect size has a different variance, and this violates the assumption of homogeneous variances of traditional significance tests. Using the incorrect statistical tests can significantly bias the outcome and interpretation of results.

Hadfield and Nakagawa (2010) also put forward a general mixed model framework for phylogenetic comparative methods and meta-analysis. This framework is by far the most flexible and sophisticated of all the approaches described in this chapter, and has the advantages of assuming nonnormal distributions and implementing meta-regression or multivariate models (see

Chapters 11 and 16). We predict that this framework and similar mixed modeling approaches will become essential to testing elaborate statistical models with phylogenetic meta-analysis.

Applying the GLS framework for other sources of nonindependence

Correlations used to estimate distances among effect sizes need not be limited to phylogenetic information; experimental, genetic, temporal, and spatial/geographic correlations could also be applied to the GLS framework presented in this chapter. One application could be to control for the overrepresentation of single studies due to extracting multiple effect sizes from a single experimental design (Rosenthal 1991). Here effect sizes are not independent because many of the experimental comparisons are based on multiple traits measured from the same individuals. Should a study report information on how these traits are correlated (i.e., Pearson product moment correlations), then a within-study pooled effect size (controlling for nonindependence among treatment effects) can be estimated using Equation 17.3 with a **P** matrix containing the correlations among traits (see also Gleser and Olkin 1994).

A similar approach can be applied to calculate a within-study pooled effect size from studies based on repeated measure designs. For example, the duration of measurement intervals in a time series analysis will dictate the independence of data, assuming that data measured at short intervals are more likely to be similar. The duration among intervals (as a linear distance) can be treated as correlations in the **P** matrix, where each effect size (drawn from each repeated measure) can be pooled based on the assumed temporal correlations. Simpler correlation structures could also be assumed, such as when measurements are only dependent on the previous measurement (autocorrelation). This avoids treating each repeated measure as a separate effect size or having to apply strict selection criteria that could lose information (e.g., extracting only the final endpoint as the study's effect size; Chapter 16).

AN ILLUSTRATIVE EXAMPLE: TRADE-OFFS IN ANTIHERBIVORE DEFENSES

Plants that defend themselves from herbivores using multiple (chemical or structural) defenses are predicted to show a negative association among these defenses (Bergelson and Purrington 1996). Plant defense theory predicts this negative association if defenses come at a cost to growth and reproduction so that, given limitations to resource investment, a trade-off in resource allocation is expected among different defenses. However, a meta-analysis of 31 independent studies testing this theory could not detect a trade-off in multiple defenses found among grasses, herbaceous plants, and trees (Koricheva et al. 2004). Although this meta-analysis provided a powerful framework to test this theory by synthesizing and exploring heterogeneity of results from multiple studies, while giving less weight to studies with poor sample precision, it did not provide an evolutionarily robust test of these findings. What was missing was an analysis that integrated the shared phylogenetic history of plants; this would account for any bias due to phylogenetic conservatism of antiherbivore defenses among related plant species (Agrawal 2007). Below, we compare models with and without phylogenetic information to evaluate the degree of phylogenetic bias in the findings of Koricheva et al.'s (2004) meta-analysis.

Methods

A hypothesized phylogenetic tree (Fig. 17.4) of the 22 plant species included in the original meta-analysis was generated using Phylomatic (Webb and Donoghue 2004). Phylomatic is a

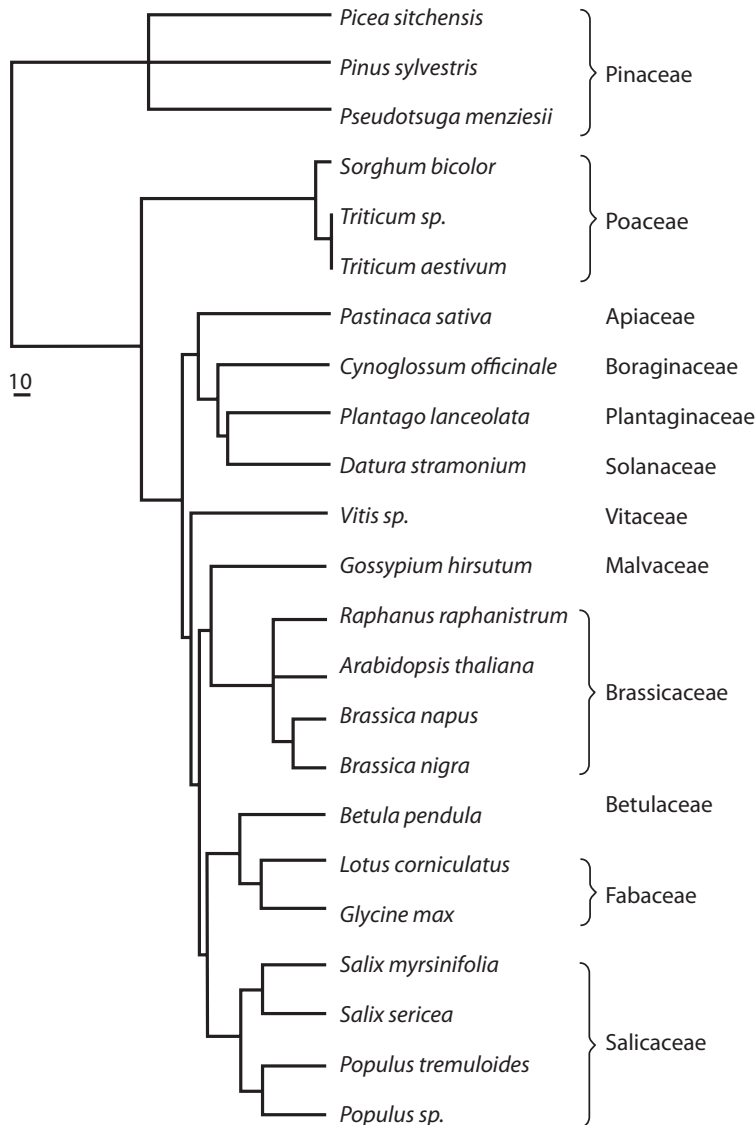


Figure 17.4. The hypothesized phylogenetic relationship of 23 plant species originally synthesized in a meta-analysis on tradeoffs in antiherbivore plant defenses; see Koricheva et al. (2004). See Table A17.1 for the NEWICK version of this tree. The plant families of these species are found to the far right. The small bar at the far left indicates the scale of branch-length distance within the phylogenetic tree (10 = 10 million years).

useful online program that estimates plant phylogenies from a modified (and updated) super-tree containing many of the major plant groups (see www.phylodiversity.net/phyloomatic). Our phylogeny also includes information on the temporal ordering of nodes. This information was based on estimates of the age of major divergence events among plant groups by Wikström et al. (2001), and was integrated in our tree using BLADJ (also on www.phylodiversity.net). Unfortunately, Phylomatic is only useful for connecting broad phylogenetic relationships, so

our tree is not entirely resolved (i.e., not completely bifurcated) among closely related species. For example, species belonging to Brassicaceae and Pinaceae families are described using polytomies.

The branch-length data of this tree are used to define the elements of the phylogenetic correlation matrix \mathbf{P} (Fig. 17.2), and \mathbf{P} is then applied to Equation 17.4 to estimate the covariance matrix \mathbf{V} for phylogenetically independent meta-analysis. However, in Koricheva et al. (2004), meta-analysis results are pooled among different categories of plant defenses such that each category contains only a subset of the species found in Figure 17.1. To resolve this issue, we used a subset tree for each category containing only the relevant species for which effect size data were available (Appendix Table 17.1). These subset trees (used to estimate \mathbf{P}) conserve all the phylogenetic branch-length information found in Figure 17.1.

One final note is that the original meta-analysis had more studies (experiments) than species. This is due to multiple studies testing different categories of defenses for a single species. To simplify analyses, we pooled the effect sizes of these replicate studies prior to our phylogenetically independent meta-analysis (Appendix Table 17.1). Thus, only one representative (pooled) effect size was used for each species within a given category of antiherbivore defenses. These pooled within-species effect sizes and their variances were estimated using traditional (fixed-effects) meta-analysis (Chapter 9). Finally, all the results reported in Koricheva et al. (2004) assume a random-effects model for pooling effect sizes across species. Here, we also apply a random-effects model, but for simplicity we assume that the between-study variance (τ) is not phylogenetically correlated among taxa. See Chapter 9 for details on how τ is estimated and Lajeunesse (2009) for information on how to calculate a phylogenetically independent version of τ . Thus before our phylogenetically independent analyses, each effect size variance was modified as follows to fit a random-effects model: $\hat{\sigma}^2(\delta_i) = \sigma^2(\delta_i) + \tau$ (details in Chapter 9).

Results and discussion

Although fewer effect sizes were used for each category of plant defenses (here K equaled the number of species in the analysis), our traditional meta-analysis yields the same conclusions reported by Koricheva et al. (2004) (Table 17.2). Studies contrasting individual defense compounds within a group of chemical defenses tended to show positive correlations among these compounds; studies contrasting constitutive (always present) versus induced plant defenses were more likely to find a negative correlation. However, phylogenetically independent meta-analysis revealed that the latter result (contrasting constitutive vs. induced defenses) was significantly influenced by the phylogenetic history of plants (Table 17.2). Here, the AIC scores indicated that the model including phylogenetic correlations (which also found no overall trade-off between these categories of defenses) was more useful in explaining variation among effect sizes. This effect of including phylogenetic correlations in the traditional meta-analysis may have occurred because it contained two species pairs that were very closely related: *Triticum* sp. and *T. aestivum*; and *Brassica napus* and *B. nigra* (Fig. 17.1). The traditional meta-analysis weighted these pairs equally in the overall analyses, even though the data derived from these taxa were not evolutionary independent.

DISCUSSION AND CONCLUSIONS

Integrating phylogenetic information into meta-analysis is challenging. Meta-analysts will have to collate data from published studies *and* gather information to generate a phylogeny for the taxa included in their meta-analysis. Below, we discuss issues that emerge when assembling

TABLE 17.2. Traditional and phylogenetically independent meta-analysis (MA) of studies testing for trade-offs in antiherbivore defenses among 21 plant species (see Figure 17.4). These studies are divided among different categories of plant defenses, and results are based on a random-effects model for pooling effect sizes. Lower (L) and Upper (U) bounds of 95% confidence intervals (CI) are also provided for the traditional ($\bar{\mu}_+$) and phylogenetically independent ($\bar{\rho}_+$) pooled effect size. Pooled effect sizes in bold indicate significant nonzero effects, whereas bolded AIC scores indicate the statistical model that best fit these pooled results. See Table 1 of Koricheva et al. (2004) to compare these results with the original findings.

Category of defenses among plant species	<i>K</i>	Traditional MA				Phylogenetically independent MA			
		$\bar{\mu}_+$	LCI	UCI	AIC	$\bar{\rho}_+$	LCI	UCI	AIC
Chemical defenses	13	-0.012	-0.392	0.368	46.0	0.033	-0.129	0.585	54.7
Individual compounds within a group	7	0.707	0.395	1.019	21.4	0.804	0.463	1.145	25.0
Groups of compounds within a class	4	-0.119	-0.325	0.086	12.0	-0.109	-0.448	0.230	14.1
Classes of compounds	4	-0.589	-1.710	0.532	11.0	-0.562	-1.652	0.528	13.4
Mechanical vs. chemical	2	0.158	-0.323	0.640	6.3	0.158	-0.323	0.640	8.3
Constitutive vs. induced	11	-0.499	-0.639	-0.359	29.1	-0.445	-1.243	0.353	16.8

phylogenetic information. We finally end with a discussion on how to test evolutionary hypotheses using the methods outlined in this chapter.

Phylogenetically independent meta-analysis requires a hypothesis about the evolutionary relationships of taxa. This information on phylogenetic history and divergence events is explicitly used to define the correlations among effect sizes with the intention of penalizing effect sizes from closely related taxa. However, a lack of information can influence the validity of these correlations. In our example, we benefited from an online resource (PhyloMatic) that generated a phylogenetic tree using a working hypothesis on the evolutionary history of all plants. Resources like this do not yet exist for other groups of taxa, and reviewers will have to mine published information or make use of the available genetic data to statistically construct phylogenies. However, each approach has its limitations. Cobbling a tree together using published phylogenies will provide a topology that will likely connect most of the taxa reviewed, but branch-length information on relative divergence times will be missing. Using Linnaean rankings would be the simplest way to sketch the topology of a tree. Constructing a tree using genetic information will provide these divergence times, but will be limited to species with genetic information. For example, a brief survey of GenBank revealed that although genetic data were available for 95% of the species found in Figure 17.4, these data were unevenly parsed among four genes useful for phylogenetic construction. A significant technical understanding of phylogenetic construction would be needed to make use of all these data, but even simplifying construction to a single gene would result in a subsampling of effect size data (Lajeunesse 2009). The difficulty of managing molecular data is further exacerbated when meta-analyses review a collection of species spread across broad taxonomic classes (e.g., a mix of plant and animals with different genes used for phylogenetic analyses).

Given these limitations, we suggest using a composite of these approaches to generate a phylogenetic tree that (a) is inclusive in that it contains many of the species under review (even if some species are specified as polytomies), and (b) contains some information about the relative timing of divergence events. Remember that this tree is a hypothesis on the phylogenetic

relationships among taxa, and that in comparative analysis, even trees specified with coarse (albeit correct) phylogenetic information can improve the inference of analyses (Freckleton et al. 2002). Of course, the closer the tree is to designating the “true” phylogenetic relationships, the better it will serve as a hypothesis for explaining variation in research outcomes. If there are, however, multiple solutions to designating phylogenetic relationships, then averaging meta-analytical results across these hypotheses is one approach to reconcile these differences.

Finally, the above approach to phylogenetically independent meta-analysis can be used to test different models of evolution. This approach extends the scope of meta-analysis by emphasizing a conceptual (evolutionary) interpretation of analyses rather than treating phylogenetic correlations as a statistical nuisance that needs correction (Westoby et al. 1995b, 1995c). For example, in our phylogenetically independent meta-analysis we assume only the purely neutral BM (Brownian motion) model of evolution. BM evolution assumes that drift is the major force acting to erase the strength of phylogenetic correlations (i.e., the strength of a phylogenetic correlation is linearly related to the time since divergence; Felsenstein 1985, Martins 1994). Figure 17.1 shows how a more recent divergence between two taxa will yield a stronger phylogenetic correlation than one derived from an ancient divergence. However, hypothesizing alternatives to BM can offer a powerful approach to providing more biologically realistic explanations for research outcomes. For example, it is also possible to assume that stabilizing selection is a force that influences the strength and maintenance of phylogenetic correlations; this model is known as an Ornstein-Uhlenbeck process (OU). Here selection is treated as a separate evolutionary parameter (in addition to drift) that acts to erase the strength of phylogenetic correlations (Hansen 1997). Hypothesizing OU models with meta-analyses can be useful when contrasted with results from a BM model, because it can help distinguish between neutral versus adaptive hypotheses in explaining variation among effect sizes (Butler and King 2004). Thus, by contrasting the fit of different evolutionary models, meta-analysis can become a powerful tool for exploring the historical processes responsible for the diversity and distribution of experimental outcomes across taxa (Lajeunesse 2009).

ACKNOWLEDGEMENTS

Work on this chapter by M.J.L. was supported by the University of South Florida and a National Science Foundation grant to the National Evolutionary Synthesis Center (EF-0423641).

APPENDIX 17

TABLE A17.1. The raw effect size data for each species grouped by each category of antiherbivore defenses. Effect sizes are the within-species pooled z -transformed Pearson’s correlation coefficients (\bar{Z}_r), and their variances $\sigma^2(\bar{Z}_r)$ are shown in parentheses. For further details of these data, see Koricheva et al. (2004). Abbreviations for each category are as follows: chemical defences (CD), individual compounds within a group (CDWG), groups of compounds within a class (CDWC), classes of compounds (CDCC), mechanical vs. chemical (MC), and constitutive vs. induced (CI).

Species *	Subgroups of CD					MC	CI
	CD	CDWG	CDWC	CDCC			
<i>Arabidopsis thaliana</i>	-	-	-	-		0.401 (0.007)	-
<i>Betula pendula</i>	-0.793 (0.143)	-	-	-0.793 (0.143)		-	-0.237 (0.071)

(continued)

TABLE A17.1. *Continued*

Species *	Subgroups of CD					CI
	CD	CDWG	CDWC	CDCC	MC	
<i>Brassica napus</i>	-	-	-	-	-	-0.698 (0.2)
<i>Brassica nigra</i>	-	-	-	-	-	-0.56 (0.008)
<i>Cynoglossum officinale</i>	-	-	-	-	-	-0.327 (0.071)
<i>Datura stramonium</i>	0.5 (0.01)	0.5 (0.01)	-	-	-	-
<i>Glycine max</i>	-	-	-	-	-	0.006 (0.2)
<i>Gossypium hirsutum</i>	-	-	-	-	-	-0.148 (0.143)
<i>Lotus corniculatus</i>	-1.673 (0.333)	-	-	-1.673 (0.333)	-	-
<i>Pastinaca sativa</i>	0.617 (0.02)	0.617 (0.02)	-	-	-	-
<i>Picea sitchensis</i>	1.306 (0.012)	1.306 (0.012)	-	-	-	-
<i>Pinus sylvestris</i>	-2.092 (0.083)	-	-	-	-	-
<i>Plantago lanceolata</i>	0.788 (0.012)	1.002 (0.035)	-	0.673 (0.019)	-	-
<i>Populus sp.</i>	-	-	-	-	-	-0.695 (0.111)
<i>Populus tremuloides</i>	0.441 (0.012)	0.971 (0.02)	-0.243 (0.026)	-	-	-
<i>Pseudotsuga menziesii</i>	-0.104 (0.143)	-	-0.104 (0.143)	-	-	-
<i>Raphanus raphanistrum</i>	0.161 (0.04)	-	0.161 (0.04)	-	-0.09 (0.01)	-
<i>Salix myrsinifolia</i>	-0.166 (0.03)	-0.045 (0.083)	-0.235 (0.048)	-	-	-0.789 (0.333)
<i>Salix sericea</i>	0.355 (0.044)	0.355 (0.044)	-	-	-	-
<i>Sorghum bicolor</i>	-0.829 (0.125)	-	-	-0.829 (0.125)	-	-
<i>Triticum aestivum</i>	-	-	-	-	-	-1.875 (1)
<i>Triticum sp.</i>	-	-	-	-	-	0.123 (0.5)
<i>Vitis sp.</i>	-	-	-	-	-	-0.321 (0.25)

* phylogenetic relationships in NEWICK format (as described in Fig. 17.4): ((Picea_sitchensis:154.761902, Pseudotsuga_menziesii:154.761917, Pinus_sylvestris:154.761917):170.238083, ((Sorghum_bicolour:12, (Triticum_sp:0.545457, Triticum_aestivum:0.545457):11.454545):149, ((Pastinaca_sativa:112.000003, (Plantago_lanceolata:87.666664, Datura_stramonium:87.666664):9.666672, Cynoglossum_officinale:97.333336):14.666667):12, (Vitis_sp:113.333336, ((Gossypium_hirsutum:95, (Raphanus_raphanistrum:33.166668, Arabidopsis_thaliana:33.166668, (Brassica_napus:16.583334, Brassica_nigra:16.583334):16.583334):61.833336):10.666664, ((Betula_pendula:88.5, (Lotus_corniculatus:56, Glycine_max:56):32.5):9.5, ((Salix_myrsinifolia:40.333328, Salix_sericea:40.333328):20.333334, (Populus_tremuloides:30.333332, Populus_sp:30.333332):30.333332):37.333336):7.666664):7.666672):10.666667):37):164):75;