# Multi-Response Phylogenetic Mixed Models: Concepts and Application

Ben Halliwell[1,2,3], Luke A. Yates[1,2], and Barbara R. Holland [1,2]

[1]School of Natural Sciences, University of Tasmania, Hobart, Tasmania, Australia

[2]ARC Centre of Excellence for Plant Success in Nature and Agriculture, University of Tasmania, Sandy Bay, Tasmania, Australia

[3]Corresponding author. Email: benjamin.halliwell@utas.edu.au

Open Research: The data set utilised for this research is "AusTraits.v4.2.0" available at https://zenodo.org/records/8353840
Online tutorial is available at https://github.com/Benjamin-Halliwell/MR-PMM. Upon acceptance, all novel code will be archived with a DOI on zenodo.

**Key words:** evolutionary ecology, trait evolution, phylogenetic comparative methods, multivariate statistics, variance partitioning, linear mixed models

## Abstract

The scale and resolution of trait databases and molecular phylogenies is increasing rapidly. These resources permit many open questions in comparative biology to be addressed with the right statistical tools. Multi-response (MR) phylogenetic mixed models (PMM) offer great potential for multivariate analyses of trait evolution. While flexible and powerful, these methods are not often employed by researchers in ecology and evolution, reflecting a specialised and technical literature that creates barriers to usage for many biologists. Here we present a practical and accessible guide to MR-PMM. We begin with a review of single-response (SR) models, the use of covariance matrices to model correlative structure, and extensions to multilevel meta-analyses. We discuss the limitations of SR-PMM, such as phylogenetic generalised least squares, for characterizing patterns of trait covariation, and emphasise MR models as a preferable approach to analyses involving multiple species traits. We discuss multivariate models of continuous trait evolution, extensions to non-Gaussian response traits, and techniques for causal inference using precision matrices. Using simulated data and visual examples, we demonstrate the motivation for partitioning trait covariances into phylogenetic and non-phylogenetic components in MR-PMM. We discuss interpretation, prediction, and model validation, including methods based on leave-one-out cross-validation. We then apply this approach to a data set of leaf traits in Eucalyptus to show how MR-PMM can provide a more nuanced understanding of trait correlations compared to traditional SR approaches. Finally, we discuss the emerging synthesis of comparative techniques made possible by MR-PMM, highlight strengths and weaknesses, and offer practical recommendations to analysts. To complement this material, we provide extensive online tutorials including side-by-side model implementations in two popular R packages, `MCMCglmm` and `brms`.

# Contents


**1 Introduction** **6**

**2 Models** **11**

2.1 Mixed Models . . . . . 11

2.2 Phylogenetic Mixed Models . . . . . 13

2.2.1 Phylogenetic meta-analysis . . . . . 15

2.2.2 Phylogenetic signal . . . . . 18

2.2.3 Models of evolution . . . . . 19

2.2.4 Limitations of single response models . . . . . 20

2.3 Multi-Response Phylogenetic Mixed Models . . . . . 22

2.3.1 Multi-response phylogenetic meta-analyses . . . . . 24

2.3.2 Multivariate models of evolution . . . . . 27

2.3.3 Non-Gaussian response traits . . . . . 29

2.3.4 Considerations for fixed effects . . . . . 30

2.3.5 Assessing conditional dependencies . . . . . 31

2.4 Summary . . . . . 33

**3 Interpretation** **34**

3.1 Phylogenetic and non-phylogenetic (co)variances . . . . . 34

3.2 Pattern versus process . . . . . 35

3.3 Visualisations . . . . . 37

3.4 Summary . . . . . 39

**4 Prediction** **40**

4.1 Predicting to new or missing observations . . . . . 40

4.2 Model validation using posterior predictive checks . . . . . 42

4.3 Predictive assessment using cross validation . . . . . 43

**5 Example Analysis - Leaf Traits in Eucalyptus** **44**

5.1 Methods . . . . . 44

5.2 Results and Discussion . . . . . 45

**6 Discussion** **48**

6.1 Summary . . . . . 48

6.2 Strengths and Weaknesses . . . . . 49

6.3 Recommendations and Future directions . . . . . 53

6.4 Conclusions . . . . . 56

**7 Tutorial** **57**

**8 Acknowledgements** **57**

# 1 Introduction

The motivation for developing multivariate phylogenetic comparative methods is now broadly appreciated (Adams and Collyer 2018; Uyeda et al. 2015, 2018; Garamszegi 2014). Modern approaches aim to move beyond phylogenetic regression and variance partitioning of individual Gaussian variables (e.g., Lynch 1991; Pagel 1999; Blomberg et al. 2003), to methods capable of evaluating the strength, direction, and conservatism of eco-evolutionary relationships within networks of continuous and discrete variables (Westoby et al., 2023; Haba and Kutsukake, 2019; Brommer et al., 2019; Hadfield, 2010). These multivariate techniques are applicable to a broad range of species traits, from morphology, physiology, and behavior, to environmental tolerance limits and niche characteristics. Furthermore, as the tide of 21st-century data collection erodes historical constraints on the scale and complexity of analyses, opportunities to apply these methods are increasing: Advances in climate modeling, remote sensing technology, and collaborative data projects are accelerating the availability and resolution of suitable datasets (Green et al., 2022; Herberstein et al., 2022; Falster et al., 2021; Kattge et al., 2020); the genomic revolution continues to provide more accurate and complete phylogenies (Young and Gillung, 2020; Laumer et al., 2019; Smith and Brown, 2018; Yeates et al., 2016); and the computational resources necessary to pursue fully Bayesian analyses are now accessible to most researchers. Despite this coalescence of opportunities, challenges in implementing modern methods create barriers to usage for many biologists.

A robust literature developing statistical models of multivariate trait evolution has existed for some time. However, new techniques are often slow to absorb into research practice, especially where they demand a technical or conceptual leap from users. Development of

the generalised least-squares framework, helped to bring phylogenetic regression and correlative analyses of trait evolution into the mainstream (Rohlf, 2001; Martins and Hansen, 1997; Lynch, 1991; Grafen, 1989; Felsenstein, 1985). Subsequent extensions to phylogenetic generalised linear mixed models (PMM) provided additional benefits, including support for non-Gaussian traits, flexibility in the specification of hierarchical group effects, and a growing familiarity with mixed models among researchers in ecology and evolution (Ives and Helmus, 2011; Ives and Garland Jr, 2010; Bolker et al., 2009; Housworth et al., 2004). Currently, PMM also supports multi-response (MR) analyses, facilitating the estimation (and decomposition) of covariances between discrete and continuous species traits (Hadfield, 2010; Hadfield and Nakagawa, 2010).

The development of user-friendly software for fitting MR-PMM (Bürkner, 2017; Hadfield, 2010) has prompted researchers from diverse scientific fields to implement the method for sophisticated comparative analyses. For example, MR-PMM has been used in anthropology to examine evolutionary relationships between climate and cranial form among neo-lithic humans (Katz et al., 2016), and the influence of phylogenetic history on marriage patterns among modern human populations (Minocher et al., 2019); in animal behaviour, to study the evolution of multivariate behavioral repertoires in Drosophilidae (Hernández et al., 2021), and the evolution of group size and social complexity in birds (Downing et al., 2020); in epidemiology and disease ecology, to understand relationships between growth rate, transmission mode, and virulence in pathogens (Leggett et al., 2017), the role of sexual dimorphism in susceptibility to viral infection (Roberts and Longdon, 2023), and covariance in macro- and micro-parasite species richness among host species (Gutiérrez et al., 2019); and, of course, in evolutionary ecology and eco-physiology, to understand the multivariate evolution of species

functional traits. Examples of the latter include coevolution among milk macro-nutrient concentrations in mammals (Blomquist, 2019), interactions between climate and functional traits on plant vital rates and life history evolution (Kelly et al., 2021), the evolutionary coordination of plant hydraulic traits (Sanchez-Martinez et al., 2020), and covariance between species traits and flammability in fire-prone grasslands (Simpson et al., 2016).

Many early applications of MR-PMM to comparative biology involved critical tests of theory made possible by the estimation of phylogenetic covariances (Section 3.1). For example, the effects of sexual selection on brain and body size evolution (García-Peña et al., 2013), whether character displacement is driven by trait divergence in allopatry or sympatry (Tobias et al., 2014), and whether low rates of extra-pair paternity, long lifespans, and tolerance of harsh environments represent causes or consequences of transitions to cooperative breeding (Cornwallis et al., 2010; Downing et al., 2015; Cornwallis et al., 2017). More recently, MR-PMM has been highlighted as a powerful framework for eco-evolutionary studies incorporating spatio-temporal random effects (Gomes et al., 2023), trait-based models of community assembly (Gallinat and Pearse, 2021), and analyses of function-valued traits (Pottier et al., 2023), as well as for disentangling opposing correlations between response variables that manifest at different levels within hierarchical datasets (Westoby et al. 2023, also see Downs and Dochtermann 2014). These strengths of MR-PMM echo a growing consensus that multivariate statistical approaches are often valuable even when associations between response variables are not of specific interest; fitting multiple correlated response variables improves inferential and predictive accuracy due to borrowing of strength (see Section 4, also see Riley et al. 2017; Pottier et al. 2023).

Despite these benefits, and the example of pioneering researchers, MR-PMM remains un-

derutilized, even for datasets to which it would be well suited, such as recent large-scale comparative analyses of species traits (e.g., Cássia-Silva et al. 2020; Grossnickle 2020; Bruelheide et al. 2018; Díaz et al. 2016). A sobering list of publications highlighting widespread misconceptions about phylogenetic comparative methods, erroneous statistical practices, and the potential for spurious inference, confirm the need for more translational research (Uyeda et al., 2018; Cooper et al., 2016; Uyeda et al., 2015; Revell, 2010; Freckleton, 2009; Revell et al., 2008). In particular, practical guidance for biologists wanting to apply multivariate methods that provide a more meaningful decomposition of trait relationships is needed to advance the field beyond the limits of univariate approaches (Westoby et al., 2023).

This review aims to bridge the gap between theory and practice in mixed model analyses of trait evolution. We focus on Bayesian implementations of MR-PMM, because this model class provides a highly flexible, multivariate framework with general utility for comparative biology, although maximum likelihood implementations do exist for certain special cases (Mitov et al., 2020; Butler et al., 2017). The capacity for the MR version of PMM to provide more nuanced analyses of trait (co)evolution than the SR version has long been recognised (Housworth et al., 2004; Hadfield and Nakagawa, 2010). However, despite a long history of use among quantitative geneticists (Meyer, 1991), we believe a lack of familiarity, along with practical barriers to implementation, continues to prevent common usage of MR-PMM among ecologists and evolutionary biologists. Thus, we begin with a broad conceptual introduction to mixed models, demonstrating how covariance matrices are used to model correlated error structures, such as those present in phylogenetic effects. We define the basic SR-PMM, and discuss its equivalence with phylogenetic generalised least squares (PGLS), the different parametrizations used to quantify phylogenetic signal, extensions to multilevel

meta-analyses, and implementation of alternative models of evolution (i.e., other than pure Brownian Motion). After outlining the limitations of SR models for characterising patterns of coevolution between traits, we shift focus to MR-PMM. In particular, we highlight the different components of trait covariance (phylogenetic and non-phylogenetic) that arise in MR-PMM, how these error structures are specified using covariance matrices, and how parametrizations of these matrices are used to estimate phylogenetic and non-phylogenetic correlations between species traits. We discuss multivariate models of continuous trait evolution, extensions to non-Gaussian response traits, and techniques for causal inference using precision matrices and graphical models. In section 3 we discuss interpretation and explore data simulated from MR-PMM to clarify the distinction between correlations operating on the phylogenetic and non-phylogenetic level. In Section 4 we cover predictive assessment, including methods based on posterior predictive distributions and leave-one-out (LOO) cross-validation procedures. In section 5, we present an example analysis on leaf economic traits in Eucalyptus from the AusTraits database (Falster et al., 2021). Finally, in section we discuss the strengths and weaknesses of MR-PMM, highlight emerging syntheses facilitated by the method, and provide recommendations to analysts.

To support our exposition, we provide extensive online tutorials, including annotated code and mathematical appendices. These tutorials demonstrate how to simulate multivariate trait data containing phylogenetic signal, then specify, validate, and interpret corresponding MR-PMM in two popular R packages, `MCMCglmm` (Hadfield, 2010) and `brms` (Bürkner, 2017). They also present a detailed example analysis using real data on leaf traits of Eucalyptus, covering common tasks and challenges in the MR-PMM workflow.

# 2 Models

## 2.1 Mixed Models

Biologists are familiar with structured sampling designs: these apply when each observation from a survey or experiment is a member of one or more nested groups. Acknowledging structure in our data is important, because observations within a group will often be more similar than can be explained by available predictors. For example, observations of species traits collected from the same transect may be more similar to each other than those collected from different transects, due to spatial heterogeneity in unobserved causal factors. This covariance between observations—the tendency for members of a group to have similar trait values—may exist at various scales, producing a hierarchy of group-level effects. For example, observations (level 1) from transects (level 2) could be nested within sites (level 3). Hierarchical (or mixed) models aim to capture this structural hierarchy to more accurately model variation in the data.

In the simplest case, the hierarchical structure is limited to a single random (or group-level) intercept. For the popular `lme4` package in R, the syntax used to express this model would be:

$$y \sim 1 + X + (1|\text{group}), \tag{1}$$

where $y$ is the observed response, 1 specifies a population-level intercept (often omitted from the model formula as fit by default), the measured covariate $X$ specifies a fixed effect, and $(1|\text{group})$ specifies a random intercept at the group level, e.g., transect. In other words, we

model $y$ as a linear combination of effects, including a random effect that accounts for the fact that observations belong to distinct groups.

To express a more general form of this model and make precise the notion of a random effect, it is helpful to adopt formal mathematical notation. In particular, we need to be explicit about the choice of statistical distributions governing each random variable,

$$
\begin{aligned}
\boldsymbol{y} &= \boldsymbol{\mu} + \boldsymbol{b} + \boldsymbol{e} \\
\boldsymbol{\mu} &= \beta_0 1 + \beta_1 X_1 + ... + \beta_k X_k \\
\boldsymbol{b} &\sim \mathcal{N}(0, \sigma_b^2 C) \\
\boldsymbol{e} &\sim \mathcal{N}(0, \sigma_e^2 I),
\end{aligned} \tag{2}
$$

where, $\boldsymbol{\mu}$ is a vector of fixed effects including a population-level intercept and fixed effect coefficients associated with predictors $X_i$, $i = 1, .., k$, $\boldsymbol{b}$ a vector of group-level random effects, and $\boldsymbol{e}$ a vector of independent residual errors. The notation $\sim \mathcal{N}(\cdots)$ means "drawn from a (multivariate) normal distribution". For the random effects, an $\mathrm{N_{obs}} \times \mathrm{N_{obs}}$ covariance matrix $C$ encodes the hierarchical structure among observations, and the identity matrix $I$ (a matrix with 1's on the diagonal and 0's in all off-diagonals) encodes residual error for each observation (Figure 1A). Finally, the parameters $\sigma_b^2$ and $\sigma_e^2$ denote the magnitude of the structured (group-level) and unstructured (residual) variance components, respectively.

To illustrate, again suppose we are analysing data from species sampled in different transects, and we want to use random intercepts to model the covariance between observations within each transect. For Gaussian distributed errors, the two variance components can simply be

added to form the total (co)variance:

$$\Sigma = \sigma_b^2 C + \sigma_e^2 I. \tag{3}$$

This *marginal* model is obtained by averaging over the distribution of the random effects, which means treating $\boldsymbol{b} + \boldsymbol{e}$ as the (marginalised) residuals around a mean $\boldsymbol{\mu}$,

$$\boldsymbol{b} + \boldsymbol{e} \sim \mathcal{N}(0, \Sigma). \tag{4}$$

In this view, the hierarchical structure of our transect sampling design can be represented as a tree comprising a polytomous clade for each transect (Figure 1A). The corresponding covariance matrix $\Sigma$ assumes a block-diagonal form for which the block sizes are determined by the number of observations within each transect, and the relative magnitudes of $\sigma_b^2$ and $\sigma_e^2$ correspond to relative lengths of the internal and external branches in the tree (also see Ives 2018). In many instances, $C$ is re-scaled to a correlation matrix (i.e., a matrix with 1's on the diagonal) which places it on the same scale as $I$ and allows the relative proportion of each variance component to be calculated from $\sigma_b^2$ and $\sigma_e^2$. This diagrammatic representation of observation-level covariance makes clear how matrices are used to model sources of structural dependence in the data, including those arising from phylogenetic history (Figure 1B).

## 2.2 Phylogenetic Mixed Models

In analyses of cross-species data, dependence arises from phylogenetic signal: the tendency for closely related species to resemble each other phenotypically due to the hierarchical evolutionary history of life (Pagel, 1999; Blomberg et al., 2003). This phylogenetic structure can be handled in exactly the same way as a nested sampling design, by defining a covariance

matrix that encodes our expectation of dependence in the data (Figure 1B). The matrix shown in Figure 1B assumes a Brownian Motion model of evolution, in which conserved covariances among species trait values reflect the proposed phylogenetic structure (Felsenstein, 1985). While it is possible to encode phylogenetic covariances assuming other models of evolution, Brownian Motion is the default assumption.

Brownian motion (BM) represents a stochastic random walk—trait values increasing or decreasing at a constant rate every time step—that branches at each node in the phylogeny. This process has the property that trait variance increases linearly with time, meaning that shared branch lengths are proportional to the expected (co)variances of trait values among species. Thus, for BM, there is a direct and simple translation from phylogeny to covariance matrix (Ives, 2018). In PMM, phylogenetic structure enters the model as the covariance $C$ of the random effects $\boldsymbol{b}$, where $C$ is scaled by the phylogenetic variance (or, evolutionary rate) $\sigma_b^2$ (3). The phylogeny is supplied by the user and is usually considered known without error (although in a Bayesian setting, phylogenetic uncertainty is easily incorporated by combining posterior samples across a set of candidate topologies). The rate parameter $\sigma_b^2$ is unknown and to be estimated from the data.

For $\mathrm{N}_{\mathrm{species}}$ in a phylogeny, $C$ is an $\mathrm{N}_{\mathrm{species}} \times \mathrm{N}_{\mathrm{species}}$ matrix containing all pairwise similarity scores between species. Figure 1B shows a phylogeny, and associated $C$ matrix, for 6 species. Note that $C$ is symmetric, with values on the diagonal representing the sum of branch lengths for each path from root to tip. For an ultrametric tree (tips appear right-adjusted) like that presented in Figure 1B, all diagonal elements $C_{ii}$ will be equal because all species were sampled in the present day. In contrast, for a pair of species $i$ and $j$, the expected phylogenetic covariance is given by the off-diagonal element $C_{ij}$, the sum of branch lengths

from the root node to the most recent common ancestor of species $i$ and $j$ (see Rohlf 2001 for details). In other words, the sum of shared branch lengths for species $i$ and $j$.

To summarise, PMM is a special case of a linear mixed model (2) that incorporates phylogenetic relatedness via a scaled covariance matrix, $C$. Species trait values $\boldsymbol{y}$ are modelled as the sum of fixed effects $\boldsymbol{\mu}$, a phylogenetic random effect $\boldsymbol{b}$, and non-phylogenetic (residual) error $\boldsymbol{e}$ (see Section 2.3.3 for non-Gaussian extensions). In the simplest case, the correlation structure of the random effects $C$ assumes a Brownian motion model of evolution.

### 2.2.1 Phylogenetic meta-analysis

Phylogenetic comparative analyses are often performed on data compiled from multiple studies. A common case is where observations represent species mean trait values (or similar statistics) with associated sampling variances. This casts PMM in a meta-analytic framework where species trait values are the study-level effect sizes. Thus, assuming a single effect size per study, we can specify a more general form of (2) as,

$$\begin{aligned} \boldsymbol{y} &= \boldsymbol{\mu} + \boldsymbol{b} + \boldsymbol{e} + \boldsymbol{m} \\ \boldsymbol{m} &\sim \mathcal{N}(0, V), \end{aligned} \tag{5}$$

where $\boldsymbol{b}$ and $\boldsymbol{e}$ are distributed as in (2), $\boldsymbol{m}$ is a vector of sampling (measurement) errors and $V$ is an $\mathrm{N_{obs}} \times \mathrm{N_{obs}}$ diagonal matrix of sampling variances. Though we focus on species mean trait values here, the models presented generalize to other standard meta-analytic effect size statistics (e.g., differences between the means of two groups, or the association between two variables), making them suitable for any meta-analysis including data on multiple species

(Pottier et al., 2023; Nakagawa and Santos, 2012).

Meta-analyses offer several benefits over traditional comparative analyses, including the ability to partition effect variation among multiple sources (e.g., study, species, and observation level), account for variation in sampling effort, and test for publication bias (Pottier et al., 2023; Nakagawa and Santos, 2012). Indeed, accounting for multilevel structure (i.e., nested levels of dependence, such as trait measurements on different species, or populations, from different studies) may be crucial for valid inference of phylogenetic (co)variances, which are likely to be confounded with other sources of between-species variation (Cinar et al., 2022; Garamszegi and Møller, 2017). Despite this, the synthesis of phylogenetic comparative- and multilevel meta-analyses has been slow to emerge. For example, failing to account for phylogenetic dependence remains a common shortcoming of biological meta-analyses: out of 322 meta-analyses in plant ecology published between 1996–2013 where it was deemed necessary to account for phylogeny, only 12.2% of studies did so (Koricheva and Gurevitch 2014; also see Chamberlain et al. 2012). Conversely, studies employing phylogenetic comparative analyses rarely adopt a meta-analytic approach, often ignoring (partially or completely) the multilevel structure inherent in the data.

For example, the model in (5) is appropriate when analyses are restricted to studies that contribute a single effect size, (e.g., the mean of a single trait across multiple individuals of a species, or a raw trait measurement of a single individual from a single species). However, in order to include data from as many studies as possible, models must accommodate more complex structural dependence among records (Nakagawa and Santos, 2012). Indeed, a limiting assumption of (5) is that all within-study variance derives from sampling variance. In practice, published studies often report multiple effect sizes (e.g., observations on

multiple traits for multiple species), introducing a multi-level structure and the potential for dependence among observations within studies (i.e., due to differences in measurement technique, systematic instrument errors, etc.). In some studies, multiple values of a trait for a given species are measured under different conditions (e.g., different sites, populations, life stages), introducing an additional source of within-species variability if all data are to be included. Phylogenetic multi-level meta-analyses offer a general framework for dealing with these complex effect structures (Cinar et al., 2022; Nakagawa and Santos, 2012). For example, using a multi-level model, we could extend (5) by including a random effect at the study level $\boldsymbol{u}$, which captures variability between studies, and an observation-level random effect $\boldsymbol{\epsilon}$, which captures variability within studies (i.e., the residual error term),

$$
\begin{aligned}
y_i &= \mu + b_{k[i]} + e_{k[i]} + u_{j[i]} + \epsilon_i + m_i \\
\boldsymbol{u} &\sim \mathcal{N}(0, \sigma_u^2 I_{\mathrm{N_{study}}}) \\
\boldsymbol{\epsilon} &\sim \mathcal{N}(0, \sigma_\epsilon^2 I_{\mathrm{N_{obs}}}),
\end{aligned} \tag{6}
$$

where $k[i]$ and $j[i]$ are the species and study indices, respectively, for the ith effect size. For species $k = 1, ..., \mathrm{N_{species}}$ and studies $j = 1, ..., \mathrm{N_{study}}$, $b_k$, $e_k$, and $u_j$ are the corresponding components of the phylogenetic, non-phylogenetic, and study effects. For effect sizes $i = 1, ..., \mathrm{N_{obs}}$, $\epsilon_i$ and $m_i$ are the within-study (residual) effect and sampling error, respectively. Finally, $\sigma_u^2$ and $\sigma_\epsilon^2$ are the study-level and residual variances, and $I_{\mathrm{N_{study}}}$ is an $\mathrm{N_{study}} \times \mathrm{N_{study}}$ identity matrix. In practice, observation-specific sampling variances $v_i^2$ (e.g. squared standard errors associated with species mean trait values) are often modeled as additive components of the residual variance (i.e., $\epsilon_i' \sim \mathcal{N}(0, \sigma_\epsilon^2 + v_i^2)$), for example, using the `se()` and `mev()` arguments in `brms` and `MCMCglmm`, respectively (see tutorial; also see Figure 4 in Nakagawa et al. (2017) for a helpful visualization of multi-level models).

The model in (6) could of course be extended to include additional covariance structures between observations, such as spatial effects (e.g. Markovski et al. 2023; also see Freckleton and Jetz 2009). This flexibility makes phylogenetic multi-level meta-analyses well-suited for large-scale comparative studies. In particular, because data compiled from numerous sources (e.g., literature reviews, public data repositories, etc) are inherently hierarchical, introducing sources of dependence that must be modeled to mitigate the invalid inferences that arise from overly simplistic models (Cinar et al., 2022).

### 2.2.2 Phylogenetic signal

A common objective of phylogenetic comparative analyses is to quantify phylogenetic signal in species traits. Numerous metrics have been developed for this purpose, each with different assumptions and applications (Münkemüller et al., 2012). In PMM (2), phylogenetic heritability $h^2$ describes the proportion of variance in $\boldsymbol{y}$ attributable to phylogeny and is expressed in terms of the estimated variance components as

$$h^2 = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_e^2}. \tag{7}$$

Importantly, when fixed or relevant random effects (e.g. (6)) are added to the model, the variance explained by these components must also be included in the denominator of the calculation for $h^2$ (de Villemereuil et al., 2016). That is, $h^2 = \frac{\sigma_b^2}{\sigma_{\text{total}}^2}$.

### 2.2.3 Models of evolution

Values of $h^2 < 1$ imply a component of variation in $\boldsymbol{y}$ that cannot be explained by Brownian motion on the tree. In the original presentation of PGLS, a scalar multiple of the phylogenetic correlation matrix $C$ is taken as the total covariance matrix (Grafen, 1989), which assumes traits have evolved via Brownian motion with no additional error. This assumption is equivalent to fixing $\sigma_e^2 = 0$, or ignoring red branch lengths in Figure 1B. With subsequent developments, $C$, or equivalently the branch lengths of the corresponding phylogenetic tree, is transformed as a function of one or more additional parameters $\boldsymbol{\theta}$, allowing different evolutionary models to be fit to the data (Symonds and Blomberg, 2014; Harmon et al., 2008). These parameters characterise branch-length transformations that produce a transformed phylogenetic correlation matrix $C(\boldsymbol{\theta})$, which is multiplied by a scalar parameter $\sigma^2$ to define the total covariance,

$$\Sigma = \sigma^2 C(\boldsymbol{\theta}). \tag{8}$$

A common instance is when $C(\theta)$ is a weighted sum of $C$ and $I$ characterised by a single parameter, $\lambda$ (Pagel, 1999),

$$C(\lambda) = \lambda C + (1 - \lambda) I. \tag{9}$$

In this formulation, the additive terms express the relative proportions, scaled by $\lambda$, of the total variance attributable to phylogenetic and non-phylogenetic error. From this we can see that $\lambda$ and $h^2$ are in fact mathematically equivalent, as $\sigma^2 = \sigma_b^2 + \sigma_e^2$ and $\lambda = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_e^2} = h^2$ (also see Housworth et al. 2004; Cinar et al. 2022). These equivalencies help clarify the relationship between PMM and other popular methods: PMM is equivalent to PIC when $\Sigma = \sigma_b^2 C$ (i.e.,

$\sigma_e^2 = 0$) and to ordinary least squares when $\Sigma = \sigma_e^2 I$ (i.e., $\sigma_b^2 = 0$) (Blomberg et al., 2012). PGLS with $\lambda$ estimated is equivalent to a marginal PMM (3). The phylogenetic threshold model of Felsenstein (2005) is equivalent to PMM with a binary response trait (Hadfield 2015; see section 2.3.3).

In addition to (9), many alternative models for the evolution of continuous traits have been proposed, including those with directional trends, stabilising selection, or changes in evolutionary rates over time (Harmon et al., 2010; Butler and King, 2004; Blomberg et al., 2003; Pagel, 1999; Hansen, 1997). Importantly, several popular models are directly implemented in PMM (e.g., $\delta$, $\kappa$, early burst (EB), Ornstein-Uhlenbeck (OU), accelerating-decelerating (ACDC)) by transformations to $C$, as in (8). This extends (2) to a range of evolutionary models besides Brownian motion (see Tung Ho and Ané 2014; Blomberg et al. 2003), for which model selection techniques can be used to discriminate among alternative evolutionary hypotheses.

### 2.2.4 Limitations of single response models

PMM provides an elegant solution to the statistical problem of phylogenetic dependence in cross-species data. However, univariate implementations (2) have important limitations that arise from treating one species trait, often arbitrarily, as the response variable and other traits as predictor variables. In particular, the validity of inferences is compromised when predictor traits themselves contain phylogenetic signal due to confounding of the phylogenetic signal in the predictor with that in the residuals (Westoby et al., 2023; Warton, 2022). Furthermore, the asymmetric preferencing of one trait as the response variable in regression analyses

implies (though does not strictly assume) a causal relationship, where evolutionary changes in the response trait are driven by changes in predictor traits. More often, species traits are correlated due to a common causal mechanism (e.g., multi-trait selection, or pleiotropy), or co-vary due to reciprocal causation, meaning interpretations of unidirectional causality are often not biologically defensible (Westoby et al., 2023).

Suppose we observe a phenotypic correlation between two traits, $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$, across a number of species and that both traits display phylogenetic signal; a common occurrence in empirical data sets (Adams and Collyer, 2019; Blomberg et al., 2003; Freckleton et al., 2002). Part of the covariance between $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$ across species is due to a phylogenetically conserved relationship between these traits (e.g., deep functional integration, or phylogenetic niche conservatism), with the remainder due to causes that are independent of phylogeny (e.g., phenotypic plasticity, local adaptation, or selection imposed by fluctuating environmental conditions). If we model these data using a single-response model (2), such as PGLS, treating $\boldsymbol{y}_2$ as a fixed predictor of $\boldsymbol{y}_1$ reduces this composite of correlations operating at the phylogenetic and non-phylogenetic levels to a single slope parameter. Not only are the separate (co)variance components not recoverable from the model, but phylogenetic signal in the predictor trait $\boldsymbol{y}_2$ is confounded with (phylogenetic components of) the residual covariance structure (Warton 2022; supporting information in Westoby et al. 2023; also see Wilson 2008 and Marques et al. 2022 for analogous issues in quantitative genetics and spatial statistics, respectively).

In other words, SR models cannot effectively separate covariance between traits that is phylogenetically structured (conserved correlation between $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$) from covariance between traits that is independent of the phylogenetic structure (non-phylogenetic correlation be-

tween $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$). One consequence of this confound is that estimated slope parameters will be misleading when phylogenetic and non-phylogenetic correlations have different signs, or vary considerably in magnitude. In a worst-case scenario, these two correlation components may cancel each other out when reduced to a single (scalar) slope estimate, leading to the erroneous conclusion that the traits in question are not meaningfully related (Westoby et al., 2023).

An alternative approach, which re-frames this analysis as a multivariate statistical problem, is to treat both traits $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$ as response variables. That is, move all traits to the left-hand side of the model equation (10). This shift to a multivariate framework has several distinct benefits: 1) it avoids often arbitrary decisions about which trait should be considered the response variable, 2) it allows phylogenetic signal to be modelled in all traits simultaneously, 3) it permits a partitioning of trait correlations into phylogenetic and non-phylogenetic components, 4) it exploits correlations between response traits to improve the precision of parameter estimates. We argue this approach is often more appropriate and more meaningful for comparative analyses including multiple species traits (also see Section 3).

## 2.3 Multi-Response Phylogenetic Mixed Models

In MR-PMM, all species traits are modelled jointly as response variables, which means both the phylogenetic random effects $\mathbf{b}$ and the non-phylogenetic errors $\mathbf{e}$ must also be modelled jointly. For Gaussian response traits, this is achieved directly using multivariate normal distributions (but see Section 2.3.3 for non-Gaussian response types). In a natural extension of the single response model (2), each trait appears in a separate row, which for two traits

can be written as

$$
\begin{aligned}
\begin{pmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \end{pmatrix} &= \begin{pmatrix} \boldsymbol{\mu}_1 + \mathbf{b}_1 + \mathbf{e}_1 \\ \boldsymbol{\mu}_2 + \mathbf{b}_2 + \mathbf{e}_2 \end{pmatrix} \\
(\mathbf{b}_1, \mathbf{b}_2)^{\mathrm{T}} &\sim \mathcal{N}(0, \Sigma^{\mathrm{b}} \otimes C) \\
(\mathbf{e}_1, \mathbf{e}_2)^{\mathrm{T}} &\sim \mathcal{N}(0, \Sigma^{\mathrm{e}} \otimes I),
\end{aligned} \tag{10}
$$

where between-trait covariances across both phylogenetic and non-phylogenetic components are modelled with the trait covariance matrices

$$
\Sigma^{\mathrm{b}} = \begin{pmatrix} \Sigma^{\mathrm{b}}_{11} & \Sigma^{\mathrm{b}}_{12} \\ \Sigma^{\mathrm{b}}_{21} & \Sigma^{\mathrm{b}}_{22} \end{pmatrix} \qquad \text{and} \qquad \Sigma^{\mathrm{e}} = \begin{pmatrix} \Sigma^{\mathrm{e}}_{11} & \Sigma^{\mathrm{e}}_{12} \\ \Sigma^{\mathrm{e}}_{21} & \Sigma^{\mathrm{e}}_{22} \end{pmatrix}, \tag{11}
$$

where the matrix elements $\Sigma^{\mathrm{e}}_{ij}$ and $\Sigma^{\mathrm{b}}_{ij}$ are scalar covariance components between traits $i$ and $j$. The Kronecker product $\otimes$ is a type of matrix multiplication used here, for example, to combine the $\mathrm{N}_{\mathrm{trait}} \times \mathrm{N}_{\mathrm{trait}}$ trait covariance matrix $\Sigma^{\mathrm{b}}_{ij}$ with the $\mathrm{N}_{\mathrm{species}} \times \mathrm{N}_{\mathrm{species}}$ phylogenetic covariance matrix $C$ into an $\mathrm{N}_{\mathrm{species}}\mathrm{N}_{\mathrm{trait}} \times \mathrm{N}_{\mathrm{species}}\mathrm{N}_{\mathrm{trait}}$ covariance matrix $\Sigma^{\mathrm{b}} \otimes C$. This larger matrix matches the dimensions of the multi-trait response vector and correctly encodes the parameterised correlation structure across species and traits. Thus, the Kronecker product for the phylogenetic component can be written as

$$
\Sigma^{\mathrm{b}} \otimes C = \begin{pmatrix} \Sigma^{\mathrm{b}}_{11} C & \Sigma^{\mathrm{b}}_{12} C \\ \Sigma^{\mathrm{b}}_{21} C & \Sigma^{\mathrm{b}}_{22} C \end{pmatrix}, \tag{12}
$$

which has the effect of introducing phylogenetic relatedness, via $C$, into the covariance structure of each trait and each pairwise trait relationship (see tutorial for fully worked examples). Like the SR case (3), the total covariance matrix of the marginal model is the

sum of the phylogenetic and non-phylogenetic components,

$$\Sigma = \Sigma^{\mathrm{b}} \otimes C + \Sigma^{\mathrm{e}} \otimes I. \tag{13}$$

Despite the large dimension of $\Sigma$, the number of estimated (co)variance parameters in MR-PMM is only $\mathrm{N}_{\mathrm{trait}}(\mathrm{N}_{\mathrm{trait}} + 1)$ (e.g., 6 parameters for the two-trait case). These parameters are the standard deviations $\sigma$ and the correlations $\rho$ used to characterise the matrix elements of both $\Sigma^{\mathrm{b}}$ and $\Sigma^{\mathrm{e}}$, expressed as

$$\Sigma^{\mathrm{b}}_{ij} = \sigma^{\mathrm{b}}_i \sigma^{\mathrm{b}}_j \rho^{\mathrm{b}}_{ij} \qquad \text{and} \qquad \Sigma^{\mathrm{e}}_{ij} = \sigma^{\mathrm{e}}_i \sigma^{\mathrm{e}}_j \rho^{\mathrm{e}}_{ij}, \tag{14}$$

for $i, j = 1, .., \mathrm{N}_{\mathrm{trait}}$.

A key distinction between SR- and MR-PMM is the decomposition of trait (co)variances. In the MR case, the Kronecker product (12) provides a means to include both phylogenetic and non-phylogenetic structures in the modelled (co)variance term for each trait and trait pair. Specifically, the estimated quantity $\widehat{\Sigma}^{\mathrm{b}}_{ij} \neq 0$ provides evidence for a component of correlation between traits $i$ and $j$ that is conserved over evolutionary time (i.e., a portion of trait covariance that is associated with phylogeny), while $\widehat{\Sigma}^{\mathrm{e}}_{ij} \neq 0$ provides evidence for a component of correlation between these traits that is independent of phylogeny (see Figure 3 for an illustration).

### 2.3.1 Multi-response phylogenetic meta-analyses

Multi-response (or multivariate) meta-analyses have become increasingly common in clinical medical research over the past decade (Jackson et al., 2011; Riley et al., 2017), and are currently in focus for researchers in ecology and evolution (Cinar et al., 2022; Pottier et al.,

2023). Two general advantages of multi-response analyses (15) over separate univariate analyses (5) emerge from considering correlations between response variables: 1) results for each response variable make use of data on all other response variables; 2) all studies that report at least one response variable of interest for at least one species can be included within a single analysis. Both points relate to the information gained by "borrowing strength" from correlated response variables, which facilitates the imputation of missing response values, and reduces the uncertainty of parameter estimates (Riley et al., 2017; Jackson et al., 2011). In the context of phylogenetic models, prediction is improved by leveraging the estimated covariance between traits acting on both the phylogenetic and non-phylogenetic levels (see Section 4). Meta-analytic models may also provide more robust inferences on trait correlations (and partial correlations; see Section 2.3.5) because they reduce the potential for spurious relationships arising from un-modelled dependencies in the data (Cinar et al., 2022; Garamszegi and Møller, 2017).

Following (6), we can define a multi-response meta-analytical model as,

$$\begin{pmatrix} y_{1,i} \\ y_{2,i} \\ \vdots \\ y_{\mathrm{N_{trait}},i} \end{pmatrix} = \begin{pmatrix} \mu_1 + b_{1,k[i]} + e_{1,k[i]} + u_{1,j[i]} + \epsilon_{1,i} + m_{1,i} \\ \mu_2 + b_{2,k[i]} + e_{2,k[i]} + u_{2,j[i]} + \epsilon_{2,i} + m_{2,i} \\ \vdots \\ \mu_{\mathrm{N_{trait}}} + b_{\mathrm{N_{trait}},k[i]} + e_{\mathrm{N_{trait}},k[i]} + u_{\mathrm{N_{trait}},j[i]} + \epsilon_{\mathrm{N_{trait}},i} + m_{\mathrm{N_{trait}},i} \end{pmatrix}, \tag{15}$$

for which the random variables are sampled as

$$
\begin{aligned}
(\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_{\mathrm{N_{trait}}})^{\mathrm{T}} &\sim \mathcal{N}(0, \Sigma^{\mathrm{b}} \otimes C) \\
(\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_{\mathrm{N_{trait}}})^{\mathrm{T}} &\sim \mathcal{N}(0, \Sigma^{\mathrm{e}} \otimes I_{\mathrm{N_{obs}}}) \\
(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_{\mathrm{N_{trait}}})^{\mathrm{T}} &\sim \mathcal{N}(0, \sigma_u^2 I_{\mathrm{N_{study}}} \otimes I_{\mathrm{N_{trait}}}) \\
(\boldsymbol{\epsilon}_1, \boldsymbol{\epsilon}_2, \ldots, \boldsymbol{\epsilon}_{\mathrm{N_{trait}}})^{\mathrm{T}} &\sim \mathcal{N}(0, \sigma_\epsilon^2 I_{\mathrm{N_{obs}}} \otimes I_{\mathrm{N_{trait}}}) \\
(\mathbf{m}_1, \mathbf{m}_2, \ldots, \mathbf{m}_{\mathrm{N_{trait}}})^{\mathrm{T}} &\sim \mathcal{N}(0, V \otimes I_{\mathrm{N_{trait}}}),
\end{aligned}
$$

where $y_{1,i}$ is the $i$th effect size for trait 1, and $k[i]$ and $j[i]$ are the species and study indices, respectively.

Use of the meta-analytic form of MR-PMM should be considered whenever data are derived from multiple studies, and analyses involve multiple traits of multiple species. Meaningful reductions in parameter uncertainty compared to SR models may require that traits are at least moderately correlated (Jackson et al., 2011; Riley et al., 2017). However, the latter is common in comparative biology, with studies often investigating trade-offs, co-selection, or coordination among species traits (see *Introduction*, also example analysis in Section 5). Recent work also highlights the utility of this approach for addressing hypotheses relating to function-valued traits such as performance curves, growth trajectories, or reaction norms (Pottier et al., 2023). Parameters that describe these functions (e.g., intercept, slope, or maxima of a fitted curve) are often correlated (lending strength), and these correlations may themselves be biologically meaningful (e.g., Pettersen et al. 2023; Kontopoulos et al. 2020).

### 2.3.2 Multivariate models of evolution

Over recent years, there have been considerable advances in software implementation and efficient computation of multivariate evolutionary models (Bartoszek et al., 2023; Blomberg et al., 2020; Mitov et al., 2020; Clavel et al., 2015; Bartoszek et al., 2012). As models of evolution are stochastic processes, they are typically expressed as stochastic differential equations. Solutions (i.e., the distribution of trait values) are obtained by integrating these equations over the phylogeny. For many multivariate models (e.g., multivariate BM, OU, EB and ACDC), the trait distributions are Gaussian and therefore characterised by their means and covariances. In the multivariate BM case, the solutions are easily obtained since the phylogenetic covariance is proportional to $C$ (12). The addition of a non-phylogenetic variance (i.e., a multivariate $\lambda$ model) can be obtained by simply re-scaling this matrix (13). For more complex models of evolution, the covariance matrix cannot generally be expressed as a simple branch-length transformation, which appears to limit the scope of MR-PMM in accommodating alternative evolutionary models. However, in certain cases, analytic expressions for the total covariance have been derived. To illustrate we consider the most general multivariate OU model,

$$d\boldsymbol{Y}(t) = A(\boldsymbol{Y}(t) - \boldsymbol{\theta})dt + \Lambda d\boldsymbol{W}(t), \tag{16}$$

where, $\boldsymbol{Y}$ is an $\mathrm{N}_{\mathrm{species}}\mathrm{N}_{\mathrm{trait}}$-vector of trait values, $A$ is a $\mathrm{N}_{\mathrm{trait}} \times \mathrm{N}_{\mathrm{trait}}$ strength of selection (or rate of adaptation) matrix, $\boldsymbol{\theta}$ is an $\mathrm{N}_{\mathrm{species}}\mathrm{N}_{\mathrm{trait}}$-vector of trait optima values, and $\boldsymbol{W}$ is an $\mathrm{N}_{\mathrm{species}}\mathrm{N}_{\mathrm{trait}}$-dimensional Brownian process with diffusion matrix $\Lambda$. The $\mathrm{N}_{\mathrm{trait}} \times \mathrm{N}_{\mathrm{trait}}$ trait covariance matrix between two species i and j at the tips is given by (Bartoszek et al.

2012, eq. B.3; also see Clavel et al. 2015),

$$Cov[Y_i, Y_j] = e^{-A(C_{ii}-C_{ij})} \left( \int_0^{C_{ij}} e^{-A\nu} \Lambda\Lambda^T e^{-A^T\nu} \, d\nu \right) e^{-A^T(C_{jj}-C_{ij})}. \tag{17}$$

As intimidating as it appears, (17) can be evaluated directly provided the matrix $A$ admits an eigen decomposition (eq. A.1 in Bartoszek et al. 2012) which is likely to be satisfied under most circumstances (Mitov et al., 2020). In the limiting case that $A$ goes to zero, (17) reduces to the standard multivariate BM where $\Sigma^{\mathrm{b}} = \Lambda\Lambda^T$. Many common OU variants are characterised by strong constraints on the form of $A$ which further simplify the evaluation of (17). For example, when attraction to optima is assumed to be equal and independent for all traits, i.e., when $A$ is a multiple of the identity matrix (Goolsby et al., 2017; Tung Ho and Ané, 2014). This result in principle extends MR-PMM to a broad class of multivariate Gaussian evolutionary models (e.g., for multivariate OU, it may be sufficient to use (17) in place of $\Sigma^{\mathrm{b}} \otimes C$ in (10)). However, options to implement models other than BM are yet to be integrated into popular software packages such as `MCMCglmm` and `brms`, despite being well developed in other modeling contexts (Bartoszek et al., 2023; Mitov et al., 2020; Clavel et al., 2015). A valuable direction for future work would therefore be to explore the repertoire of evolutionary models compatible with MR-PMM and integrate them into standard software. Progress in this direction would yield a powerful modeling framework, capable of synthesizing model selection among alternate multivariate evolutionary models with the flexible multi-level meta-analytic framework of MR-PMM.

### 2.3.3 Non-Gaussian response traits

One advantage of MR-PMM compared to other methods (e.g., Goolsby et al. 2017; Clavel et al. 2015; Bartoszek et al. 2012) is the capacity to include both continuous and discrete response traits. To simultaneously include different trait types, a latent-variable formulation is used. This models between-trait covariance in the usual way, via a multivariate Gaussian distribution, while at the same time permitting trait-specific probability distributions and link functions (Hadfield, 2010). For traits $i = 1, .., \mathrm{N}_{\mathrm{trait}}$, the latent-variable formulation is written

$$\begin{pmatrix} \boldsymbol{l}_1 \\ \boldsymbol{l}_2 \\ \vdots \\ \boldsymbol{l}_{\mathrm{N}_{\mathrm{trait}}} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\mu}_1 + \mathbf{b}_1 + \mathbf{e}_1 \\ \boldsymbol{\mu}_2 + \mathbf{b}_2 + \mathbf{e}_2 \\ \vdots \\ \boldsymbol{\mu}_{\mathrm{N}_{\mathrm{trait}}} + \mathbf{b}_{\mathrm{N}_{\mathrm{trait}}} + \mathbf{e}_{\mathrm{N}_{\mathrm{trait}}} \end{pmatrix}, \tag{18}$$

where $\boldsymbol{l}_i$ are the latent variables and $\mathbf{b}$ and $\mathbf{e}$ are drawn from multivariate normal distributions characterised by the phylogenetic ($\Sigma^{\mathrm{b}}$) and non-phylogenetic ($\Sigma^{\mathrm{e}}$) trait-correlation structures as in (10). A model for the observations $\boldsymbol{y}_i$ is given in terms of a trait-specific probability distribution $f_i$ and link function $\eta_i$

$$\boldsymbol{y}_i \sim f_i(\eta_i^{-1}(\boldsymbol{l}_i), \boldsymbol{\phi}_i), \tag{19}$$

where $\boldsymbol{\phi}_i$ includes any distribution-specific parameters (if required). For example, count data might be modelled using a negative binomial distribution with log link function and overdispersion $\phi$. For the formulation (19), the joint likelihood of the trait observations and

the latent variables is

$$p(\boldsymbol{y}, \boldsymbol{l}) = \prod_{i=1}^{\mathrm{N_{trait}}} p(\boldsymbol{y}_i \mid \boldsymbol{l}_i, \phi_i)\, p(\boldsymbol{l} \mid \Sigma^{\mathrm{b}}, \Sigma^{\mathrm{e}}). \tag{20}$$

Depending on the choice of distribution $f_i$, additional constraints must be imposed to ensure parameter identifiability. For example, using a probit threshold model for binary data (Hadfield, 2015), the mean $\mu_\varepsilon$ and variance $\sigma^2_\varepsilon$ of the associated normal distribution must be fixed (typically to define a cumulative standard normal distribution) as well as the corresponding non-phylogenetic variance component $\Sigma^{\mathrm{e}}_{ii}$ (typically fixed to one). It should be noted that these conventional, yet arbitrary, constraint choices re-scale the other parameters estimates accordingly. If $f_i$ is Gaussian, then $\phi_i = \sigma_i$ must be fixed to make $\Sigma^{\mathrm{e}}_{ii}$ identifiable.

Variance partitioning to define an analogue of $h^2$ (7) is possible for latent-variable models, but is more nuanced because there is a distinction between latent- and observation-scale partitioning where distribution-specific variances must be taken into account (Nakagawa et al., 2017). For example, using the above-stated constraints (i.e., $\sigma^2_e = \sigma^2_\varepsilon = 1$) in a probit model for a binary trait, the proportion of variance attributable to phylogeny on the scale of the threshold variable is

$$h^2_{\mathrm{probit}} = \frac{\sigma^2_b}{\sigma^2_b + \sigma^2_e + \sigma^2_\varepsilon} = \frac{\sigma^2_b}{\sigma^2_b + 2}. \tag{21}$$

### 2.3.4 Considerations for fixed effects

There are two main reasons for including fixed effects in a MR-PMM. The first is to account for sampling biases, such as differences in data-collection methods or laboratory protocols.

For example, in the field of plant hydraulics, physiological traits relating to xylem vulnerability are often measured using different techniques that each have distinct assumptions and sources of error (Jansen et al., 2015). In such situations, it is advisable to account for measurement methodology with a fixed effect covariate. The second reason is to explore relationships between responses conditional on covariates, typically to evaluate hypotheses about the mechanistic basis of trait correlations. For example, Sanchez-Martinez et al. (2020) used MR-PMM to partition correlations between a range of plant hydraulic traits into phylogenetic and non-phylogenetic contributions, then re-fit models with relevant climate covariates included as fixed effects. Whether or not trait correlations remained after accounting for climate effects was used to evaluate hypotheses about trade-offs and integration between these traits in the evolution of plant hydraulic systems (see also Kruuk and Hadfield 2007 for related examples from quantitative genetics). Equivalent procedures are used to control for environmental variation in multivariate species co-occurrence models (Ovaskainen et al., 2017; Warton et al., 2015). However, a preferable approach will often be to treat covariates as response traits and assess conditional dependencies.

#### 2.3.5 Assessing conditional dependencies

The practice of assessing conditional relationships between traits (i.e., after controlling for relevant covariates) is well-established for univariate models. For example, it is common for researchers to use multiple regression to estimate partial regression coefficients for each fixed effect predictor. However, in MR-PMM, it is often preferable to estimate conditional relationships from a joint model, where both focal traits and covariates are modelled as response variables and feature on the left-hand side of the model equation. This approach

is necessary because traits commonly treated as covariates, such as those characterising the climatic or environmental niche of species, often display strong phylogenetic signal (Liu et al., 2020; Kubota et al., 2017; Peixoto et al., 2017; Hof et al., 2010; Kozak and Wiens, 2010; Smith and Beaulieu, 2009; Hawkins et al., 2007). Treating covariates as response variables both avoids potential confounds between fixed and phylogenetic effects (see section 2.2.4) and allows conditional dependencies between traits to be estimated on the phylogenetic and non-phylogenetic levels, providing more nuanced interpretations. Conditional dependencies between response traits in MR-PMM are obtained by computing the matrix inverse of the estimated trait covariance matrices, known as precision matrices,

$$\Omega^{\mathrm{b}} = (\Sigma^{\mathrm{b}})^{-1} \qquad \text{and} \qquad \Omega^{\mathrm{e}} = (\Sigma^{\mathrm{e}})^{-1}. \tag{22}$$

The use of precision matrices to characterise conditional dependencies casts MR-PMM in the framework of Gaussian graphical models (Popovic et al., 2019; Epskamp et al., 2018; Yuan and Lin, 2007), which have great potential for clarifying assumptions of phylogenetic comparative methods (Uyeda et al., 2018). Elements of the precision matrices relate directly to the existence of edges in a graphical network (Figure 2). For example, suppose we obtain data on three co-varying species traits $x$, $y$ and $z$. The absence of an edge between $x$ and $y$ after accounting for $z$ signifies conditional independence: that $x$ and $y$ are independent, given $z$. Precision matrices are sparser than their corresponding covariance matrices (i.e., contain more off-diagonal elements that are close to 0), which focuses our inference on candidate causal relationships. Outside of controlled experiments, it remains possible that edges in a graph could be explained by missing variables. Nonetheless, precision matrices are a powerful tool for generating candidate causal hypotheses from observational data (Popovic et al., 2019; Pearl, 1995). Furthermore, MR-PMM can be parameterised directly in terms

of the precision elements, which naturally reduces model complexity and allows priors to be placed on these target parameters; e.g., for the purpose of regularization (see tutorial).

## 2.4 Summary

Many species traits intended for use as predictors or covariates in SR-PMM (2) will themselves contain phylogenetic signal. If the evolutionary processes generating data we observe are characterised by reciprocal co-evolution (or joint distribution) of two or more traits, then the prediction-focused SR model may be badly misspecified; it not only fails to adequately partition covariances but can also fail to identify the presence of significant correlative effects. By partitioning trait correlations into phylogenetic and non-phylogenetic contributions, MR-PMM (10) allows us to decompose cross-species trait correlations into a more meaningful set of contributing factors. This framework permits multilevel meta-analyses, which account for complex structural dependencies in the data. It also permits the estimation of conditional dependencies between response traits (on both the phylogenetic and non-phylogenetic levels), which describe how pairs of traits are related after controlling for covariances with other response traits. As partitioning trait variation and identifying functional trait relationships are principal objectives of the eco-evolutionary research program, MR-PMM will often represent a better approximating model and thus offer more valuable inferences. Considering these benefits, there is an emerging consensus that MR-PMM is underutilised in ecological and evolutionary research (Westoby et al., 2023; Pottier et al., 2023).

# 3 Interpretation

The biological interpretation of PMM (in particular PGLS (2)), and its utility for addressing specific ecological and evolutionary hypotheses, has been a subject of lively debate (Freckleton et al., 2002; Westoby et al., 1995; Björklund, 1997; Harvey et al., 1995). Reflecting on this constructive body of work, we believe MR-PMM emerges as a flexible, pluralistic framework for analyses of multiple species traits. However, consensus on interpretation is critical to realise the full, operational potential of the method.

## 3.1 Phylogenetic and non-phylogenetic (co)variances

Phylogenetic (co)variances estimate the conserved component of species phenotypes, given a model of evolution and a hypothesis of the phylogenetic relationship between taxa. The extent to which these conserved effects result from neutral Brownian motion, or reflect constraints and adaptive responses to selection, will vary for different traits, between clades, across phylogenetic scales, and is likely to fluctuate throughout the course of evolutionary history (Revell et al., 2008; Losos, 2008, 2011). As Housworth et al. (2004), point out "[PMM] envisions phenotypic evolution as being the result of a complex of forces including some that are retained over long periods of time, forming patterns in trait variation that reflect the underlying phylogenetic structure, and others that act more quickly, in bursts of change that are lost easily at new speciation events.". Expressed another way, the phylogenetic component of trait (co)variance in PMM will capture all variation attributable to the phylogenetic relationships among taxa. Examples include variation arising from genetic differences between clades that have accumulated over evolutionary history, as well as

non-genetic (e.g., spatial, environmental or developmental) effects that are phylogenetically structured for one reason or another.

The non-phylogenetic component will capture variation not attributable to phylogenetic structure, including effects from processes operating exclusively in the present day, as well as effects from historical processes that leave no phylogenetic signature in the data (e.g., because all species are equally affected, or because effects are random with respect to phylogeny). Examples include measurement error, phenotypic plasticity, and rapid evolutionary change following divergence from sister taxa. The latter point is particularly interesting as, for the MR case, this means that $\Sigma_{ij}^{\text{ind}}$ may partly reflect nascent phylogenetic covariance between traits $i$ and $j$; i.e., over subsequent divergences through evolutionary time, such effects may come to reflect the future phylogenetic structure, and thus be attributed instead to $\Sigma_{ij}^{\text{phy}}$. This reinforces a key conceptual point that the distinction between phylogenetic and non-phylogenetic effects in PMM is as much about time as it is about mechanism (Housworth et al., 2004). Indeed, the term 'conservative trait correlation' may be preferable to 'phylogenetic trait correlation', as the former descriptor is mechanism-agnostic, accommodating the possibility that historical selective agents persist into the present day, while the latter implies trait correlation resulting from strictly historical causes (Westoby et al. 2023).

## 3.2 Pattern versus process

There are inherent limitations to the inferences one can make on evolutionary processes from data observed in the present day. Indeed, many have argued that the interpretation of phylogenetic comparative analyses should be limited to patterns of variation, rather than the

explicit processes that generated them (Revell et al., 2008; Losos, 2008, 2011; Ives, 2018). Others have suggested that integration of fossil evidence, path analyses, ancestral state reconstruction, and simulation studies may extend our epistemic reach to hypothesis tests of the processes generating observable variation (Uyeda et al., 2018; Quental and Marshall, 2010; Slater et al., 2012; Uyeda and Harmon, 2014).

In MR-PMM, the covariance of each trait pair (and its decomposition into phylogenetic and non-phylogenetic components) is the focus of analysis rather than the prediction of one trait as a function of the others. This approach is clearly preferable when there are no *a priori* hypotheses about (unidirectional) causal relationships between traits (Warton et al., 2006). However, because phylogenetic and non-phylogenetic correlations can arise from multiple distinct causal processes, it also argues for caution around mechanistic interpretations of trait associations. Two traits may be phylogenetically correlated due to functional integration if, for example, conserved genes underlying a set of traits show linkage or pleiotropy that constrains the evolutionary potential for certain trait combinations, or because traits form part of a coordinated life history strategy or functional trait spectrum and clades display niche conservatism (Westoby et al., 2023; Wiens et al., 2010). Alternatively, phylogenetic correlations may also reflect indirect relationships between traits, owing to similar selective pressures acting on multiple traits independently throughout evolutionary history, such as independent responses to shared environmental effects in clades displaying niche conservatism. Furthermore, these hypotheses should not be considered mutually exclusive, as numerous processes across time may contribute to observed patterns of phylogenetic (co)variance in the present (Losos, 2011; Revell et al., 2008). Deriving conditional dependencies can help evaluate evidence for alternative causal hypotheses in MR-PMM (see Section 2.3.5). Ulti-

mately, however, it is rarely the pleasure of a comparative biologist to declare causation. Rather, to uncover meaningful patterns of variation and generate testable hypotheses for future comparative and experimental research.

## 3.3 Visualisations

To compare and contrast phylogenetic and non-phylogenetic variance components, and clarify their biological interpretation, we simulated data under MR-PMM for different values of phylogenetic and non-phylogenetic (co)variance between two Gaussian traits $x$ and $y$. Figure 3 shows scatter plots of simulated data together with the tree from which phylogenetic random effects were generated. The ellipses within the scatter plot provide a visual representation of the simulation conditions. A heat map of trait data plotted against the phylogeny highlights the distinct signature each source of structural covariance leaves in the data.

In figure 3A, $x$ and $y$ were simulated with no phylogenetic signal ($\sigma^b = 0$), but a strong positive correlation operating on the non-phylogenetic level ($\rho^b = 0, \rho^e = 0.9$). Clades overlap completely in the scatter plot, together forming a distinguishable positive ellipse. The heatmap shows bands of colour across $x$ and $y$ that appear randomised with respect to the phylogeny (i.e., for any given species, low values of $x$ tend to be observed together with low values of $y$, but the structure of the phylogeny is not reflected in the distribution of $x$ and $y$ values).

Figure 3B shows the opposing situation, $x$ and $y$ are strongly positively correlated but now this covariance manifests entirely with respect to the phylogenetic structure ($\rho^b = 0.9, \rho^e = 0$) and there is no residual variation in either trait ($\sigma^e = 0$). The scatter plot in Figure 3B

shows clearly distinguishable clades, with a tendency for both within- and between-clade correlation. The extent of between-clade correlation (the tendency for clades to arrange along a positive slope) depends on the topology of the tree, with deep splits between major clades promoting separation along the major axis of co-variation. This phenomenon is apparent across clades in 3B, but also within the green clade: a particularly early split between subclades is reflected in two groups of green points drawn out along a positive slope between $x$ and $y$. The heatmap in figure 3B shows phylogenetic structure corresponding to the different bivariate means of each clade and subclade. This structure weakens across clades from green, to orange to blue as the topology becomes more deeply nested, i.e., as subclades become less clearly separated in evolutionary time.

In figure 3C, $x$ and $y$ were simulated with equal phylogenetic and non-phylogenetic variances, and a strong positive correlation operating on both levels. This scenario, where phylogenetic and non-phylogenetic correlations show the same sign and similar magnitude, is likely to be common for many pairs of species traits. For example, where selection pressures driving trait correlations over evolutionary time persist into the present day, and adaptation at the species level reflects the same functional relationships between traits. The data in 3C are more variable ($\sigma^b + \sigma^e = 2$) but show similarities to the patterns in both 3A and 3B. For example, clades are distinguishable in the scatterplot but together conform more closely to the non-phylogenetic bivariate density imposed by $\rho^e$.

In Figure 3D, both traits contain phylogenetic and non-phylogenetic variance, but the correlation between traits is only present on the phylogenetic level. This shows how easily conserved correlations may be obscured when random sources of variation contribute substantially to the total variance in species traits. Notably, figure 3D represents a set of

conditions for which SR-PMM, such as PGLS, will typically fail to detect a significant association between $x$ and $y$ (Westoby et al., 2023). As conserved covariances may result from deep functional integration between traits, and can even be the product of selection pressures that continue into the present day, these phylogenetic effects will often be relevant for understanding ongoing ecological and evolutionary associations between traits (Westoby et al., 2023).

Visualizations of simulated data are powerful heuristics for biologists, as they provide tangible representations of the abstract covariance structures we aim to partition with MR-PMM. For more technical coverage of simulating multivariate data containing phylogenetic and non-phylogenetic covariance structure, we include detailed examples and R code in the tutorial.

## 3.4 Summary

We see the nuanced interpretive framework of MR-PMM as having three strong usage cases for biologists: 1) to provide a more meaningful decomposition of trait (co)variances, including phylogenetic and non-phylogenetic components; 2) to test for conditional dependencies, consistent with causal relationships between species traits, in a rigorous phylogenetic framework; 3) to generate new hypotheses about the ecological and evolutionary drivers of trait variation. With an emerging synthesis of evolutionary, ecological and genomic approaches to comparative analyses (James et al., 2021; Smith et al., 2020), we concur with recent work recognizing the considerable potential of MR-PMM to advance our understanding of trait evolution, niche conservatism, and community assembly at numerous scales (Pottier et al. 2023; Westoby et al. 2023; Gallinat and Pearse 2021, also see Abrego et al. 2020).

# 4 Prediction

While the estimation of trait covariance, rather than trait prediction *per se*, is the focus of MR-PMM, the inclusion of traits as multiple responses rather than multiple predictors does not in fact limit our ability to make predictions from the model. On the contrary, the estimated covariance structure can be fully exploited for predictive goals through the use of conditional predictive distributions (described below). This approach actually improves on ordinary multiple regression for prediction as it makes use of the phylogenetic structure in the predictors. The predictive distribution of a fitted model is also useful for model validation, such as posterior-predictive checks, or model comparison using predictive assessment such as cross-validation. We now review conceptual aspects of predictive distributions in MR-PMM, and refer the reader to the accompanying tutorial for worked examples of how these predictive methods can be implemented in R.

## 4.1 Predicting to new or missing observations

The assumed phylogenetic relationships between taxa and estimated trait correlations from a MR-PMM can be used to predict to new taxa and traits given only partially observed data. To generate this predictive model, the new responses $\boldsymbol{y}_{\text{new}}$ (those to be predicted) are appended to the observed responses $\boldsymbol{y}_{\text{obs}}$ to define a joint multivariate normal distribution (compare with (2) and (10)),

$$\begin{pmatrix} \boldsymbol{y}_{\text{obs}} \\ \boldsymbol{y}_{\text{new}} \end{pmatrix} \sim \mathcal{N}\left( \begin{pmatrix} \widehat{\boldsymbol{\mu}}_{\text{obs}} \\ \widehat{\boldsymbol{\mu}}_{\text{new}} \end{pmatrix}, \begin{pmatrix} \widehat{\Sigma}_{\text{obs, obs}} & \widehat{\Sigma}_{\text{obs, new}} \\ \widehat{\Sigma}_{\text{new, obs}} & \widehat{\Sigma}_{\text{new, new}} \end{pmatrix} \right), \tag{23}$$

where the set of observed values includes both those used to fit the original PMM and any new observations that have been made. The components $\widehat{\boldsymbol{\mu}}$ and $\widehat{\Sigma}$ depend only on the estimated parameters of the fitted PMM and, if available, an extended phylogenetic covariance matrix that includes any new species. Here we restrict our attention to the Gaussian case, although the technique can be extended to the latent-variable formulation presented in Section 2.3.3; see also Ovaskainen and Abrego (2020, Chapter 7) for a detailed exposition in the context of joint species distributions.

It is a standard result from the theory of multivariate normal distributions (e.g., Tong, 2012) that the predictive distribution of the new responses $\boldsymbol{y}_{\text{new}}$ conditional on the observed data $\boldsymbol{y}_{\text{obs}}$ is

$$\boldsymbol{y}_{\text{new}} | \boldsymbol{y}_{\text{obs}} \sim N(\widehat{\boldsymbol{\mu}}'_{\text{new}}, \Sigma'_{\text{new}}), \tag{24}$$

where

$$\widehat{\boldsymbol{\mu}}'_{\text{new}} = \widehat{\boldsymbol{\mu}}_{\text{new}} + \Sigma_{\text{new, obs}} \Sigma^{-1}_{\text{obs, obs}} (\boldsymbol{y}_{\text{obs}} - \hat{\boldsymbol{\mu}}_{\text{obs}}) \tag{25}$$

$$\widehat{\Sigma}'_{\text{new}} = \widehat{\Sigma}_{\text{new, new}} - \widehat{\Sigma}_{\text{new, obs}} \widehat{\Sigma}^{-1}_{\text{obs, obs}} \widehat{\Sigma}_{\text{obs, new}}. \tag{26}$$

The observed responses $\boldsymbol{y}_{\text{obs}}$ enter the conditional model (25) as linear predictors of the new responses which adjusts the prediction from the unconditional mean $\widehat{\boldsymbol{\mu}}_{\text{new}}$ and reduces the predictive variance (26).

Conditional predictive distributions are useful not only in the context of prediction to new data, but also for the imputation of missing data and for cross-validation approaches to predictive assessment (see Section 4.3). We illustrate the use of conditional prediction for predictive assessment in an example analysis using real data in Section 5.

## 4.2 Model validation using posterior predictive checks

In Bayesian settings, posterior predictive checks use simulated data from the fitted model to test both the adequacy of the fit to the data and the plausibility of the model predictions (Gelman et al., 2013). These checks are typically visual plots that rely on qualitative assessments (Gabry et al., 2019). To test for adequacy of fit, one option is to superpose the observed data onto a plot of the distribution of the simulated data. For a PMM, this type of check could be performed using a separate plot for each trait type with the observed data point for each species plotted on top of a 5 point summary of the predicted distribution (see Figure 4 for an example). For assessment of model plausibility, the current state of knowledge should be used to evaluate model predictions within ecologically plausible but perhaps unobserved ranges of the included covariates. For example, do predictions for $\boldsymbol{y}_1$ make sense across the entire range of plausible $\boldsymbol{y}_2$ values for a given set of taxa and/or region of interest?

The distinction between marginal and conditional models with respect to the phylogenetic random effects in PMM leads to two corresponding types of predictive distributions. The *marginal predictive distribution* concerns population-level prediction and the *conditional predictive distribution* concerns prediction to specific taxa conditional on the taxon-specific random effect estimate (Figure 4).

## 4.3 Predictive assessment using cross validation

Cross validation is the use of data splitting to estimate the predictive performance of one or more statistical models, usually for the purpose of model comparison and selection (Yates et al., 2022). Model selection is used when discrete decisions must be made about model structure. For example, whether or not to include various fixed effects (i.e., variable selection) or the choice of probability distribution (e.g., Poisson or negative binomial for count data). Predictive assessment tools such as cross validation are also useful to quantify or simply visualise how well a model can predict to new data (e.g., to new taxon-trait pairs in a PMM) which is distinct from the assessment of model adequacy which concerns prediction of data to which a model was fit.

Cross validation works by fitting each model to a subset of the available data and then assessing the models' predictive capacities on the remaining portion of the data. The splitting procedure is systematically iterated to select different test data and the overall predictive performance is summarised as a cross validation *score* (Arlot and Celisse, 2010). When the measure of predictive performance is the log likelihood of the test data then the predictive assessment is said to be *information theoretic*. Information criteria such as Akaike's Information Criteria (AIC, Akaike, 1973), or for Bayesian analyses the widely applicable information criteria (WAIC, Watanabe, 2010), approximate predictive log likelihood without data splitting by adding a bias correction to the log likelihood of the full data which requires only a single fit for each model. Information criteria are therefore faster to compute than cross-validation scores, however the latter are often preferred as they are less sensitive to violations of model assumptions and are readily combined with techniques to mitigate over-

fitting (Yates et al., 2021). For Bayesian MR-PMM estimated using Monte Carlo sampling, model fitting is often too slow to permit the use of ordinary cross validation, however recently developed approximate methods provide a rapidly computed and accurate alternative (Vehtari et al., 2017; Bürkner et al., 2021).

# 5 Example Analysis - Leaf Traits in Eucalyptus

To demonstrate potential applications of MR-PMM, we present an example analysis using data on leaf traits of Eucalyptus species from the AusTraits data set (Falster et al., 2021). Our intentions for this analysis were to evaluate evidence for: 1) phylogenetic signal in each trait; and 2) correlations between traits on the phylogenetic and non-phylogenetic level. For simplicity, we present an analysis based on species mean trait values as in (10). Full details of the analysis including model specification, diagnostics, validation, and inference (as well as extensions to multilevel meta-analyses) are available in the tutorial.

## 5.1 Methods

We derived the phylogenetic correlation matrix $C$ from the maximum likelihood Eucalypt phylogeny "ML1" presented in Thornhill (2019). We focused on three target leaf traits: leaf mass per unit area (LMA); nitrogen content per dry mass of leaf tissue ($\mathbf{N}_{\text{mass}}$); and the ratio of carbon isotopes 12 and 13 in leaf tissue ($\delta^{13}\mathbf{C}$). In `MCMCglmm`, missing trait values are imputed by default during model fitting. For analyses using `brms`, we chose to prune the tree to complete cases for simplicity, acknowledging work that suggests phylogenetic imputation

may be a preferable approach to missing data in many circumstances (Debastiani et al. 2021; Penone et al. 2014). We calculated weighted means and standard errors for each trait by pooling all observations at the species level. Where transformations to trait data were necessary, these were applied to individual observations, before means and standard errors were calculated, to avoid subsequent approximations of the standard error (Garamszegi 2014, Chapter 7). Species for which the standard error of a trait could not be calculated because only a single observation was recorded were assigned a standard error equal to the 90th percentile of the standard error estimates for that trait across species. This represents a conservative approach to the assignment of unknown sampling variance in a meta-analytical context (Weir et al., 2018; Ives et al., 2007).

For model validation, we performed marginal and conditional posterior predictive checks. The proportion of data included in the nominal predictive intervals verified the capacity of the model to generate plausible data (Figure 4), although the proportion was lower for mass and $\delta^{13}\mathbf{C}$ than for LMA (see tutorial). The shortfall in predictive capacity for these traits may indicate the model is missing relevant covariates or that the chosen probability distributions are less well-suited to these two traits. For predictive assessment using approximate LOO-CV (Bürkner et al., 2021) see the tutorial.

## 5.2 Results and Discussion

Each trait showed strong phylogenetic signal (posterior mean and 95% credible interval for $\hat{h}^2$: LMA = 0.83 (0.75, 0.88), $\mathbf{N}_{\text{mass}}$ = 0.77 (0.63, 0.87), $\delta^{13}\mathbf{C}$ = 0.73 (0.58, 0.84)) and hence the tendency for similar values among closely related species (Figure 6). The traits

selected for this analysis are tightly linked to resource-use strategies and leaf economics (Reich, 2014; Wright et al., 2004). For example, Prieto et al. (2018) found that plant species in a Mediterranean woodland showing more resource acquisitive strategies (low LMA and high $\mathbf{N}_{\text{mass}}$), were associated with lower water use efficiency (low $\delta^{13}\mathbf{C}$). This is consistent with predictions from leaf economic theory (Wright et al., 2004), that LMA, $\delta^{13}\mathbf{C}$ and $\mathbf{N}_{\text{mass}}$ represent components of a coordinated life history strategy which trade-off predictably as different regions of niche space are explored during species diversification. These co-evolutionary relationships are clearly supported by the fitted model, which reports a positive phylogenetic correlation between LMA and $\delta^{13}\mathbf{C}$, and negative phylogenetic correlations between these traits and $\mathbf{N}_{\text{mass}}$. Thus, in line with theory and empirical observations from other plant groups, these traits appear to co-vary predictably over evolutionary time in Eucalyptus.

A notable result is that for $\mathbf{N}_{\text{mass}}$ and $\delta^{13}\mathbf{C}$, the non-phylogenetic correlation is positive while the phylogenetic correlation is negative (Figure 5). This situation highlights an important strength of MR-PMM; the capacity to identify opposing correlations operating at the phylogenetic and non-phylogenetic level. Indeed, unlike the consistent correlations identified for other trait pairs, contrasting correlations may obscure meaningful relationships in cross-species data (Figure 6; Figure 3D; see tutorial). Given this result, the subsequent objective is to identify plausible biological mechanisms that might explain these opposing relationships. Carbon fractionation within plant tissues is approximately proportional to the ratio of intercellular to ambient $CO_2$ mole fractions (ci:ca, see Farquhar et al. 1982). This inequality is strongly influenced by stomatal conductance, because stomatal aperture determines the rate of diffusion of $CO_2$ into leaves (Farquhar and Sharkey, 1982). As a result,

leaf $\delta^{13}\mathbf{C}$ is generally considered a proxy for water use efficiency, with higher values found in species adapted to low soil moisture availability and showing greater stomatal regulation (Cernusak et al., 2013; Diefendorf et al., 2010; Kohn, 2010). This is consistent with the negative phylogenetic correlation observed between $\mathbf{N}_{\text{mass}}$ and $\delta^{13}\mathbf{C}$, i.e., as water use efficiency is necessarily traded-off against growth rate and resource acquisitiveness in the divergence of life history strategies. However, leaf nitrogen content can also influence ci:ca, and hence carbon discrimination, through effects on photosynthetic capacity. Specifically, as variation in $\mathbf{N}_{\text{mass}}$ reflects variation in the carboxylating enzyme Rubisco (Evans and Clarke, 2019), greater $\mathbf{N}_{\text{mass}}$ will be associated with increased demand for $CO_2$ in the mesophyll, relaxing isotope discrimination and increasing the $\delta^{13}\mathbf{C}$ of leaf tissue (Cernusak et al., 2013). This is consistent with the positive non-phylogenetic correlation observed, i.e. that after accounting for the apparent trade-off between these traits over evolutionary time, species with greater $\mathbf{N}_{\text{mass}}$ tend to have greater $\delta^{13}\mathbf{C}$ due to the influence of photosynthetic capacity on carbon isotope discrimination.

One important consideration for this result is that variation in leaf nitrogen has different eco-physiological interpretations depending on whether it is expressed on a per mass or per area basis (Wright et al., 2005; Dong et al., 2020). Specifically, carboxylation capacity is expected to be more strongly related to $\mathbf{N}_{\text{area}}$ than $\mathbf{N}_{\text{mass}}$ (Evans and Clarke, 2019). Interestingly, while $\mathbf{N}_{\text{mass}}$ and $\mathbf{N}_{\text{area}}$ were not correlated in this dataset (Pearson correlation coefficient: $r = 0.09$, $P = 0.11$), $\mathbf{N}_{\text{area}}$ was positively correlated with the residuals of $\mathbf{N}_{\text{mass}}$ from the fitted model ($r = 0.29$, $P < 0.001$). This suggests that part of the positive non-phylogenetic correlation observed between $\mathbf{N}_{\text{mass}}$ and $\delta^{13}\mathbf{C}$ (Figure 5) represents variation in carboxylation capacity, corroborating our initial interpretation of results.

# 6 Discussion

## 6.1 Summary

Phylogenetic mixed models are a familiar, flexible, and scalable framework for comparative analyses of species traits. Phylogenies are used to derive an expectation of similarity among species trait values, which enters the model as a covariance matrix. Different evolutionary models are fit via transformations of this covariance matrix. For a single response trait, modelling phylogenetic structure means the proportion of variance attributable to phylogeny (i.e., phylogenetic signal) can be estimated. However, single-response models with fixed effects may confound an ensemble of correlations if predictor and response variables both show phylogenetic signal. MR-PMM provides a solution via the explicit decomposition of trait covariances. This allows correlations between response traits to be partitioned into phylogenetic and non-phylogenetic contributions, providing a more nuanced understanding of trait relationships. We review these models, with a focus on statistical concepts, model specification, parameter interpretation, and practical implementation. We discuss recent advances in multivariate evolutionary models and propose a synthesis of these methods with the meta-analytic framework of MR-PMM as a valuable direction for future work. Finally, we provide an extensive tutorial featuring worked examples and supplementary code to help users get started, as well as additional mathematical details for those seeking a more technical understanding.

## 6.2 Strengths and Weaknesses

One limitation of MR-PMM, as currently implemented in the popular R packages `MCMCglmm` and `brms`, is the restriction to a multivariate BM model of evolution. By construction, MR-PMM includes additive residual (co)variances, which facilitate a data-driven scaling between phylogenetic and non-phylogenetic effects. This model is surprisingly flexible and will provide an adequate approximation for the evolutionary diversification of many species traits at many phylogenetic scales. However, despite its statistical elegance, BM is clearly a gross oversimplification of the varied and dynamic processes generating phenotypic diversity over macro-evolutionary time (Uyeda et al., 2018; Losos, 2011; Freckleton and Harvey, 2006). In particular, BM does not attempt to model the fundamental processes of adaptive evolution (e.g. stabilizing and directional selection). At first glance, this failure of BM to explicitly model adaptive processes appears to be a fatal weakness. However, phylogenetic signal consistent with BM is, in fact, also consistent with adaptation via niche conservatism; these processes can produce indistinguishable phylogenetic patterns in ecological data (reviewed in Westoby et al. 2023; Wiens et al. 2010). Brownian phylogenetic (co)variance is therefore relevant for understanding divergence in ecological strategy between clades, a major contributor to the functional diversity of ecosystems.

Furthermore, while MR-PMM offers support for more sophisticated multivariate evolutionary models (see section 2.3.2), BM is easily the most tractable, a virtue so often repeated it risks being overlooked (Blomberg et al., 2020). In particular, modern comparative analyses including many thousands of taxa pose genuine computational challenges to more complex, parameter-rich evolutionary models. Of course, such challenges do not obviate the need

to account for structural dependencies in comparative data, rather they call for a more tractable model. Nonetheless, large-scale studies, particularly in the field of plant functional ecology, continue to be published despite no effort to account for phylogeny (e.g., Bruelheide et al. 2018; Díaz et al. 2016; also see Koricheva and Gurevitch 2014). At the very least, phylogenetic models assuming BM evolution should feature in the analysis of such datasets, and MR-PMM offers a robust, flexible and informative framework to do so (Westoby et al., 2023). With the inexorable expansion of biological trait databases (Falster et al., 2021; Kattge et al., 2020), even models assuming BM become computationally burdensome as the number of species and traits included in analyses has grown. Future work could address this by developing approximate inference methods for MR-PMM, such as mean field variational Bayes algorithms (Ormerod and Wand 2010, though the accuracy of covariance components for non-Gaussian response traits remains a challenge, see Hughes et al. 2023). However, a particularly powerful synthesis would emerge from integrating recent advances in the computational efficiency of non-Brownian multivariate evolutionary models (Mitov et al., 2020) into the generalizable, multi-level framework of MR-PMM.

Looking beyond PMM, there exists a broad class of processes expressible in terms of stochastic partial differential equations that characterise more complex and dynamic multivariate evolutionary models (Blomberg et al., 2020; Bartoszek et al., 2012; Butler and King, 2004). These include fully parameterised multivariate OU models (i.e., unstructured $\boldsymbol{A}$ matrices), as well as multiple optima and shift models (see below), examples of which are implemented in `mvSLOUCH` (Bartoszek et al., 2012), `mvMORPH` (Clavel et al. 2015) and `Rphylopars` (Goolsby et al., 2017), with some models also estimable via a fast linear-time maximum likelihood algorithm in `PCMbase` (Mitov et al., 2020). Like MR-PMM, these methods accommodate

missing data and within-species variation but offer more detailed and (potentially more) realistic evolutionary models. On the other hand, model selection approaches to identify the correct structure (e.g, for $\boldsymbol{A}$) are non-trivial, and may require very large comparative datasets (Bartoszek et al., 2023). Furthermore, these methods are restricted to continuous traits, and do not permit a decomposition of trait covariances into phylogenetic and non-phylogenetic components, which is a desirable and informative simplification for many biological questions (Westoby et al., 2023).

An important limitation of (10), is that it assumes the chosen evolutionary model is both constant through time and homogeneous across the tree, making it vulnerable to misspecification under certain evolutionary scenarios (Uyeda et al., 2018; Clavel et al., 2015). Some alternative methods allow for shifts, meaning parameters of the evolutionary model (e.g. $\sigma_b^2$), even the model itself (BM, OU, etc), are allowed to vary between clades and across time (Pagel et al., 2022; Mitov et al., 2020, 2019; Clavel et al., 2015), including (when furnished with fossil data) changes in the trait-level covariance matrix itself through time (i.e., $\Sigma = \Sigma(\theta, t)$) (Blomberg et al., 2020). Currently, it is not clear how these more dynamic models of trait evolution might be expressed in the PMM framework. Thus, while MR-PMM is well suited to ancestral state reconstruction under the assumption of a constant BM model (Section 5), other techniques will be more appropriate when trying to infer the exact timing or phylogenetic position of rare evolutionary events or shifts in the underlying processes generating trait variation (Pagel et al. 2022; Uyeda and Harmon 2014; Slater 2013).

Compared with other multivariate phylogenetic comparative methods (see above), we highlight two general strengths of the mixed model approach. First, is the use of link functions to include (Gaussian) random effects in the modelled mean of traits that are themselves

not Gaussian distributed (Nakagawa and Schielzeth, 2010; Hadfield, 2010). This capability is exploited in MR-PMM to estimate both phylogenetic and non-phylogenetic correlations between Gaussian and non-Gaussian response variables, where covariances are modelled on the link (or latent) scale (see Section 2.3.3). Many evolutionary hypotheses invoke causal relationships between continuous and discrete traits. Currently, however, MR-PMM is one of very few methods available for estimating correlations between continuous and discrete traits in a fully phylogenetic framework (see also Haba and Kutsukake 2019), enabling several longstanding evolutionary hypotheses to be addressed in recent years (e.g., Downing et al. 2020; Cornwallis et al. 2017).

A second strength of MR-PMM, is the capacity to specify multi-level meta-analytical models that appropriately capture hierarchical study-level effects (Cinar et al., 2022; Nakagawa and Santos, 2012; Hadfield and Nakagawa, 2010). Integrating data from multiple studies to draw general conclusions is a common practice in comparative biology. It enables researchers to increase the power and complexity of analyses and is simply necessary to make full use of growing public data repositories. However, failure to account for multi-level data remains common, risking poor statistical performance from models and spurious inferences from analyses (Cinar et al., 2022). The capacity to account for multi-level structure is therefore crucial for methods to effectively scale with modern data compilations (e.g. Falster et al. 2021; Kattge et al. 2020).

Despite many advantages, one obstacle to broader uptake of MR-PMM is that it represents a relatively complex and technical approach compared to univariate alternatives, such as PGLS. In particular, special care must be taken during model specification to ensure random effects are structured as intended (particularly for multilevel meta-analyses), dur-

ing model validation and predictive assessment, as well as in the interpretation of results. Nonetheless, we believe efforts to engage with this sophisticated modeling framework are richly rewarded by more tenable assumptions, sharper inferences, and ultimately more biologically meaningful analyses. Our work provides a current and comprehensive review of MR-PMM, covering all major considerations and providing practical guidelines for analysts. Nonetheless, the following texts on mixed models in biology (Nakagawa and Schielzeth, 2010; Wilson et al., 2010; Bolker et al., 2009), and specifically multi-response mixed models (Brommer et al., 2019; de Villemereuil, 2018; Dingemanse and Dochtermann, 2013; Hadfield, 2010; Housworth et al., 2004), are valuable reading for practitioners looking to deepen their understanding of these methods.

## 6.3 Recommendations and Future directions

A common pitfall in phylogenetic comparative analyses is the failure to account for within-species variation. This problem arises when omitted variables influence phenotypes at the individual or population level, adding error and potential confounds to estimates of inter-specific (co)variance components (Ives et al., 2007). Examples include local environmental or habitat variation, species community composition, or population genetic structure. In practice, data on all relevant covariates are rarely available and data pooling procedures aimed at marginalizing within-species variation are adopted instead. For example, it is common to use species mean trait values pooled from multiple observations or studies as the response vector, and model variation introduced by omitted variables as sampling error around these species mean values (Section 5; also see the tutorial). Assuming sufficient replication, an alternative approach (see Section 2.3.1) is to model species traits at the

lowest possible level (e.g., observations on individual organisms), using multilevel meta-analyses (Cinar et al. 2022 also see Felsenstein 2008).

This approach offers several benefits besides the inclusion of observation-specific covariates. For example, effective separation of within- and between-population variances (Garamszegi and Møller, 2017), phylogenetic and spatial effects (e.g., where observations on individuals or populations are spread across spatially structured environmental gradients) (Gomes et al., 2023; Freckleton and Jetz, 2009), and the estimation of within-species residual (co)variances (i.e., patterns of species-specific phenotypic plasticity and trait co-variation) (Goolsby et al., 2017). As species differ not only in mean trait values but also in the magnitude and structure of trait variation, partitioning variances both within and between species may be necessary to identify drivers of selection. For example, partitioning variance in anti-predator behaviour of 254 bird species into within-population, between-population, and between-species components supported the view that within- and between-population variances were being driven by different selective forces (Garamszegi and Møller, 2017). Similarly, a global meta-analysis showed that within-species variances account for a considerable portion of community-level variances in plant functional traits and that within-species variance tends to be greater in whole-plant (e.g. plant height), rather than organ-level (e.g. leaf size), traits (Siefert et al., 2015). Looking forward, integration of within- or between-population genetic relatedness (i.e., pedigree) into such studies would represent a powerful synthesis of quantitative genetic, ecological and evolutionary approaches. One can envisage MR-PMM models that combine phylogenetic and pedigree information into a unified framework, permitting a decomposition of species trait (co)variances into additive genetic, phylogenetic, spatial, environmental, and residual components, all while accounting for the multi-level structure of observations.

With the current pace of genomic and comparative data collection, opportunities to fit these nested, multilevel covariance structures will begin to emerge.

One clear direction for future research is a more comprehensive assessment of statistical power and sensitivity in MR-PMM. In particular, it would be instructive to have clearer expectations around parameter uncertainty with respect to the number of response traits, the mixture of non-Gaussian distributions among response traits, the relative magnitude of phylogenetic and non-phylogenetic (co)variances, the magnitude of sampling error, phylogenetic uncertainty and sample sizes. In one of the first presentations of MR-PMM, Housworth et al. (2004) provide a basic power analyses for phylogenetic covariance between two Gaussian traits, indicating unpredictable performance for small sample sizes (i.e., <100 taxa). However, with subsequent implementation of more robust algorithms for sampling parameter values in MR-PMM (Hadfield 2010), these results may no longer be relevant. In general, MR-PMM is well suited to datasets with many species $n$ relative to the number of traits $p$ (i.e., high $n : p$ ratios). Alternative methods, or rank-reduced versions, will be more appropriate for higher dimensional data sets with low $n : p$ ratios, such as those generated in studies of geometric morphometrics (Clavel and Morlon 2020; Adams and Collyer 2019, 2018; Collyer et al. 2015; Adams 2014, also see Runcie et al. 2021 for methods relevant to genomic prediction). However, a formal examination of power and sensitivity, provided by extensive simulation studies, is necessary to identify the inherent limitations of MR-PMM and provide firmer guidelines for analysts.

## 6.4 Conclusions

1. Species phenotypes are the product of a complex network of causal processes that promote co-selection of traits, over both short and long time scales. For some traits, (co)variation is conserved over evolutionary time, creating patterns in cross-species data that reflect phylogenetic relationships (Section 2).

2. Distinguishing the influence of these conserved effects from those that are decoupled from phylogenetic history is a fundamental objective of evolutionary ecology, because it is the balance of these forces that defines whether constraints, trade-offs and coordinated strategies, should be understood in terms of deep evolutionary integration or labile responses to prevailing conditions (Section 3).

3. Traditional approaches to cross-species analysis, such as PGLS, are not parameterised to distinguish between these components of trait correlation (Section 2), leading to potential confounds and inferential limitations. In contrast, MR-PMM, in which correlations between traits can be partitioned into phylogenetic and non-phylogenetic components, offer a more plausible and informative model of trait evolution under many circumstances (Section 3, 5).

4. The capacity to estimate conditional dependencies between traits (Section 2.3.5), accommodate different error distributions (Section 2.3.3), random effect structures (Section 2.3.1) and models of trait evolution (Section 2.3.2), and employ conditional predictions for imputation and model validation (Section 4), makes MR-PMM a robust and unifying approach to many open questions in comparative biology.

5. In particular, MR-PMM is well-equipped to advance an emerging synthesis between

genetic, ecological, and evolutionary approaches to the study of trait variation. Data sets will soon exist that allow the partitioning of within- and between-species trait variances into additive genetic, phylogenetic, spatial, environmental, and residual components.

6. We expect more researchers in ecology and evolution will adopt these methods as barriers to implementation for non-specialists are broken down. To this end, we provide extensive tutorial materials, including annotated code and applied examples.

# 7 Tutorial

Tutorial material relevant to this article can be found here.

# 8 Acknowledgements

We thank Ian Wright, Mark Westoby, David Warton, and Simone Blomberg for helpful discussions during preparation of the manuscript. We thank Shinichi Nakagawa for editorial suggestions that improved the manuscript. We thank Rachael Gallagher for consultation on the AusTraits database. This work was funded by The Australian Research Council Centre of Excellence for Plant Success in Nature and Agriculture (CE200100015).

# References


Abrego, N. et al. (2020). Joint species distribution modelling with applications in r.

Adams, D. C. (2014). A method for assessing phylogenetic least squares models for shape and other high-dimensional multivariate data. *Evolution 68*(9), 2675–2688.

Adams, D. C. and M. L. Collyer (2018). Multivariate phylogenetic comparative methods: evaluations, comparisons, and recommendations. *Systematic biology 67*(1), 14–31.

Adams, D. C. and M. L. Collyer (2019). Phylogenetic comparative methods and the evolution of multivariate phenotypes. *Annual Review of Ecology, Evolution, and Systematics 50*, 405–425.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov and F. Csaki (Ed.), *Second International Symposium on Information Theory (Tsahkadsor, 1971)*, pp. 267–281. Academiai Kiado, Budapest.

Arlot, S. and A. Celisse (2010). A survey of cross-validation procedures for model selection. *Statistics Surveys 4*, 40–79.

Bartoszek, K., J. Fuentes-González, V. Mitov, J. Pienaar, M. Piwczyński, R. Puchałka, K. Spalik, and K. L. Voje (2023). Model selection performance in phylogenetic comparative methods under multivariate ornstein–uhlenbeck models of trait evolution. *Systematic Biology 72*(2), 275–293.

Bartoszek, K., J. Pienaar, P. Mostad, S. Andersson, and T. F. Hansen (2012). A phylogenetic comparative method for studying multivariate adaptation. *Journal of Theoretical Biology 314*, 204–215.

Björklund, M. (1997). Are 'comparative methods' always necessary? *Oikos 80*(3), 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*(4), 717–745.

Blomberg, S. P., J. G. Lefevre, J. A. Wells, and M. Waterhouse (2012, January). Independent contrasts and PGLS regression estimators are equivalent. *Systematic Biology 61*(3), 382–391.

Blomberg, S. P., S. I. Rathnayake, and C. M. Moreau (2020). Beyond brownian motion and the ornstein-uhlenbeck process: stochastic diffusion models for the evolution of quantitative characters. *The American Naturalist 195*(2), 145–165.

Blomquist, G. E. (2019). Adaptation, phylogeny, and covariance in milk macronutrient composition. *PeerJ 7*, e8085.

Bolker, B. M., M. E. Brooks, C. J. Clark, S. W. Geange, J. R. Poulsen, M. H. H. Stevens, and J.-S. S. White (2009). Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in ecology & evolution 24*(3), 127–135.

Brommer, J. E., B. Class, and G. Covarrubias-Pazaran (2019). Multivariate mixed models in ecology and evolutionary biology: inference and implementation in r. *EcoEvoRxiv Prepr 10*.

Bruelheide, H., J. Dengler, O. Purschke, J. Lenoir, B. Jiménez-Alfaro, S. M. Hennekens, Z. Botta-Dukát, M. Chytrỳ, R. Field, F. Jansen, et al. (2018). Global trait–environment relationships of plant communities. *Nature ecology & evolution 2*(12), 1906–1917.

Bürkner, P.-C. (2017). brms: An r package for bayesian multilevel models using stan. *Journal of statistical software 80*, 1–28.

Bürkner, P.-C., J. Gabry, and A. Vehtari (2021, 11). Efficient leave-one-out cross-validation for Bayesian non-factorized normal and Student-t models. *Computational Statistics 36*(2), 1243–1261.

Butler, D., B. Cullis, A. Gilmour, B. Gogel, and R. Thompson (2017). Asreml-r reference manual version 4. *VSN International Ltd, Hemel Hempstead, HP1 1ES, UK*.

Butler, M. A. and A. A. King (2004). Phylogenetic comparative analysis: a modeling approach for adaptive evolution. *The american naturalist 164*(6), 683–695.

Cássia-Silva, C., C. G. Freitas, L. P. Lemes, G. B. Paterno, P. A. Dias, C. D. Bacon, and R. G. Collevatti (2020). Higher evolutionary rates in life-history traits in insular than in mainland palms. *Scientific reports 10*(1), 1–10.

Cernusak, L. A., N. Ubierna, K. Winter, J. A. Holtum, J. D. Marshall, and G. D. Farquhar (2013). Environmental and physiological determinants of carbon isotope discrimination in terrestrial plants. *New Phytologist 200*(4), 950–965.

Chamberlain, S. A., S. M. Hovick, C. J. Dibble, N. L. Rasmussen, B. G. Van Allen, B. S. Maitner, J. R. Ahern, L. P. Bell-Dereske, C. L. Roy, M. Meza-Lopez, et al. (2012). Does phylogeny matter? assessing the impact of phylogenetic information in ecological meta-analysis. *Ecology Letters 15*(6), 627–636.

Cinar, O., S. Nakagawa, and W. Viechtbauer (2022). Phylogenetic multilevel meta-analysis: A simulation study on the importance of modelling the phylogeny. *Methods in Ecology and Evolution 13*(2), 383–395.

Clavel, J., G. Escarguel, and G. Merceron (2015). mvmorph: an r package for fitting multivariate evolutionary models to morphometric data. *Methods in Ecology and Evolution 6*(11), 1311–1319.

Clavel, J. and H. Morlon (2020). Reliable phylogenetic regressions for multivariate comparative data: illustration with the manova and application to the effect of diet on mandible morphology in phyllostomid bats. *Systematic biology 69*(5), 927–943.

Collyer, M. L., D. J. Sekora, and D. C. Adams (2015). A method for analysis of phenotypic change for phenotypes described by high-dimensional data. *Heredity 115*(4), 357–365.

Cooper, N., G. H. Thomas, and R. G. FitzJohn (2016). Shedding light on the 'dark side'of phylogenetic comparative methods. *Methods in Ecology and Evolution 7*(6), 693–699.

Cornwallis, C. K., C. A. Botero, D. R. Rubenstein, P. A. Downing, S. A. West, and A. S. Griffin (2017). Cooperation facilitates the colonization of harsh environments. *Nature Ecology & Evolution 1*(3), 1–10.

Cornwallis, C. K., S. A. West, K. E. Davis, and A. S. Griffin (2010). Promiscuity and the evolutionary transition to complex societies. *Nature 466*(7309), 969–972.

de Villemereuil, P. (2018). Quantitative genetic methods depending on the nature of the phenotypic trait. *Annals of the New York Academy of Sciences 1422*(1), 29–47.

de Villemereuil, P., H. Schielzeth, S. Nakagawa, and M. Morrissey (2016, 11). General Methods for Evolutionary Quantitative Genetic Inference from Generalized Mixed Models. *Genetics 204*(3), 1281–1294.

Debastiani, V. J., V. A. Bastazini, and V. D. Pillar (2021). Using phylogenetic information to

impute missing functional trait values in ecological databases. *Ecological Informatics 63*, 101315.

Díaz, S., J. Kattge, J. H. Cornelissen, I. J. Wright, S. Lavorel, S. Dray, B. Reu, M. Kleyer, C. Wirth, I. Colin Prentice, et al. (2016). The global spectrum of plant form and function. *Nature 529*(7585), 167–171.

Diefendorf, A. F., K. E. Mueller, S. L. Wing, P. L. Koch, and K. H. Freeman (2010). Global patterns in leaf 13c discrimination and implications for studies of past and future climate. *Proceedings of the National Academy of Sciences 107*(13), 5738–5743.

Dingemanse, N. J. and N. A. Dochtermann (2013). Quantifying individual variation in behaviour: mixed-effect modelling approaches. *Journal of Animal Ecology 82*(1), 39–54.

Dong, N., I. C. Prentice, I. J. Wright, B. J. Evans, H. F. Togashi, S. Caddy-Retalic, F. A. McInerney, B. Sparrow, E. Leitch, and A. J. Lowe (2020). Components of leaf-trait variation along environmental gradients. *New Phytologist 228*(1), 82–94.

Downing, P. A., C. K. Cornwallis, and A. S. Griffin (2015). Sex, long life and the evolutionary transition to cooperative breeding in birds. *Proceedings of the Royal Society B: Biological Sciences 282*(1816), 20151663.

Downing, P. A., A. S. Griffin, and C. K. Cornwallis (2020). Group formation and the evolutionary pathway to complex sociality in birds. *Nature Ecology & Evolution 4*(3), 479–486.

Downs, C. and N. Dochtermann (2014). Testing hypotheses in ecoimmunology using mixed models: disentangling hierarchical correlations.

Epskamp, S., L. J. Waldorp, R. Mõttus, and D. Borsboom (2018). The gaussian graphical model in cross-sectional and time-series data. *Multivariate behavioral research 53*(4), 453–480.

Evans, J. R. and V. C. Clarke (2019). The nitrogen cost of photosynthesis. *Journal of Experimental Botany 70*(1), 7–15.

Falster, D., R. Gallagher, E. H. Wenk, I. J. Wright, D. Indiarto, S. C. Andrew, C. Baxter, J. Lawson, S. Allen, A. Fuchs, et al. (2021). Austraits, a curated plant trait database for the australian flora. *Scientific Data 8*(1), 1–20.

Farquhar, G. D., M. H. O'Leary, and J. A. Berry (1982). On the relationship between carbon isotope discrimination and the intercellular carbon dioxide concentration in leaves. *Functional Plant Biology 9*(2), 121–137.

Farquhar, G. D. and T. D. Sharkey (1982). Stomatal conductance and photosynthesis. *Annual review of plant physiology 33*(1), 317–345.

Felsenstein, J. (1985). Phylogenies and the comparative method. *The American Naturalist 125*(1), 1–15.

Felsenstein, J. (2005). Using the quantitative genetic threshold model for inferences between and within species. *Philosophical Transactions of the Royal Society B: Biological Sciences 360*(1459), 1427–1434.

Felsenstein, J. (2008). Comparative methods with sampling error and within-species variation: contrasts revisited and revised. *The American Naturalist 171*(6), 713–725.

Freckleton, R. (2009). The seven deadly sins of comparative analysis. *Journal of evolutionary biology 22*(7), 1367–1375.

Freckleton, R. P. and P. H. Harvey (2006). Detecting non-brownian trait evolution in adaptive radiations. *PLoS biology 4*(11), e373.

Freckleton, R. P., P. H. Harvey, and M. Pagel (2002). Phylogenetic analysis and comparative data: a test and review of evidence. *The American Naturalist 160*(6), 712–726.

Freckleton, R. P. and W. Jetz (2009). Space versus phylogeny: disentangling phylogenetic and spatial signals in comparative data. *Proceedings of the Royal Society B: Biological Sciences 276*(1654), 21–30.

Gabry, J., D. Simpson, A. Vehtari, M. Betancourt, and A. Gelman (2019, January). Visualization in bayesian workflow. *Journal of the Royal Statistical Society: Series A (Statistics in Society) 182*(2), 389–402.

Gallinat, A. S. and W. D. Pearse (2021). Phylogenetic generalized linear mixed modeling presents novel opportunities for eco-evolutionary synthesis. *Oikos 130*(5), 669–679.

Garamszegi, L. Z. (2014). *Modern phylogenetic comparative methods and their application in evolutionary biology: Concepts and practice*. Springer.

Garamszegi, L. Z. and A. P. Møller (2017). Partitioning within-species variance in behaviour to within-and between-population components for understanding evolution. *Ecology Letters 20*(5), 599–608.

García-Peña, G., D. Sol, A. Iwaniuk, and T. Székely (2013). Sexual selection on brain size in shorebirds (c haradriiformes). *Journal of Evolutionary Biology 26*(4), 878–888.

Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin (2013, 11). *Bayesian Data Analysis* (Third ed.). Chapman and Hall/CRC.

Gomes, L. d. M., G. S. Garcia, C. A. Cordeiro, N. A. Gouveia, C. E. Ferreira, M. G. Bender, G. O. Longo, J. P. Quimbayo, and D. F. Gherardi (2023). Complex phylogenetic origin and geographic isolation drive reef fishes response to environmental variability in oceanic islands of the southwestern atlantic. *Ecography*, e06559.

Goolsby, E. W., J. Bruggeman, and C. Ané (2017). Rphylopars: fast multivariate phylogenetic comparative methods for missing data and within-species variation. *Methods in Ecology and Evolution 8*(1), 22–27.

Grafen, A. (1989). The phylogenetic regression. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences 326*(1233), 119–157.

Green, S. J., C. B. Brookson, N. A. Hardy, and L. B. Crowder (2022). Trait-based approaches to global change ecology: moving from description to prediction. *Proceedings of the Royal Society B 289*(1971), 20220071.

Grossnickle, D. M. (2020). Feeding ecology has a stronger evolutionary influence on functional morphology than on body mass in mammals. *Evolution 74*(3), 610–628.

Gutiérrez, J. S., T. Piersma, and D. W. Thieltges (2019). Micro-and macroparasite species richness in birds: The role of host life history and ecology. *Journal of Animal Ecology 88*(8), 1226–1239.

Haba, Y. and N. Kutsukake (2019). A multivariate phylogenetic comparative method incorporating a flexible function between discrete and continuous traits. *Evolutionary Ecology 33*, 751–768.

Hadfield, J. and S. Nakagawa (2010). General quantitative genetic methods for comparative biology: phylogenies, taxonomies and multi-trait models for continuous and categorical characters. *Journal of evolutionary biology 23*(3), 494–508.

Hadfield, J. D. (2010). Mcmc methods for multi-response generalized linear mixed models: the mcmcglmm r package. *Journal of statistical software 33*, 1–22.

Hadfield, J. D. (2015). Increasing the efficiency of mcmc for hierarchical phylogenetic models of categorical traits using reduced mixed models. *Methods in Ecology and Evolution 6*(6), 706–714.

Hansen, T. F. (1997). Stabilizing selection and the comparative analysis of adaptation. *Evolution 51*(5), 1341–1351.

Harmon, L. J., J. B. Losos, T. Jonathan Davies, R. G. Gillespie, J. L. Gittleman, W. Bryan Jennings, K. H. Kozak, M. A. McPeek, F. Moreno-Roark, T. J. Near, et al. (2010). Early bursts of body size and shape evolution are rare in comparative data. *Evolution 64*(8), 2385–2396.

Harmon, L. J., J. T. Weir, C. D. Brock, R. E. Glor, and W. Challenger (2008). Geiger: investigating evolutionary radiations. *Bioinformatics 24*(1), 129–131.

Harvey, P. H., A. F. Read, and S. Nee (1995). Why ecologists need to be phylogenetically challenged. *Journal of Ecology 83*(3), 535–536.

Hawkins, B. A., J. A. F. Diniz-Filho, C. A. Jaramillo, and S. A. Soeller (2007). Climate, niche conservatism, and the global bird diversity gradient. *the american naturalist 170*(S2), S16–S27.

Herberstein, M. E., D. J. McLean, E. Lowe, J. O. Wolff, M. K. Khan, K. Smith, A. P. Allen, M. Bulbert, B. A. Buzatto, M. D. Eldridge, et al. (2022). Animaltraits-a curated animal trait database for body mass, metabolic rate and brain size. *Scientific Data 9*(1), 265.

Hernández, D. G., C. Rivera, J. Cande, B. Zhou, D. L. Stern, and G. J. Berman (2021). A framework for studying behavioral evolution by reconstructing ancestral repertoires. *Elife 10*, e61806.

Hof, C., C. Rahbek, and M. B. Araújo (2010). Phylogenetic signals in the climatic niches of the world's amphibians. *Ecography 33*(2), 242–250.

Housworth, E. A., E. P. Martins, and M. Lynch (2004). The phylogenetic mixed model. *The American Naturalist 163*(1), 84–96.

Hughes, D. M., M. García-Fiñana, and M. P. Wand (2023). Fast approximate inference for multivariate longitudinal data. *Biostatistics 24*(1), 177–192.

Ives, A. R. (2018). Mixed and phylogenetic models: a conceptual introduction to correlated data. *See https://leanpub.com/correlateddata*.

Ives, A. R. and T. Garland Jr (2010). Phylogenetic logistic regression for binary dependent variables. *Systematic biology 59*(1), 9–26.

Ives, A. R. and M. R. Helmus (2011). Generalized linear mixed models for phylogenetic analyses of community structure. *Ecological Monographs 81*(3), 511–525.

Ives, A. R., P. E. Midford, and T. Garland Jr (2007). Within-species variation and measurement error in phylogenetic comparative methods. *Systematic biology 56*(2), 252–270.

Jackson, D., R. Riley, and I. R. White (2011). Multivariate meta-analysis: potential and promise. *Statistics in medicine 30*(20), 2481–2498.

James, M. E., M. J. Wilkinson, D. M. Bernal, H. Liu, H. L. North, J. Engelstädter, and D. Ortiz-Barrientos (2021). Phenotypic and genotypic parallel evolution in parapatric ecotypes of senecio. *Evolution 75*(12), 3115–3131.

Jansen, S., B. Schuldt, and B. Choat (2015). Current controversies and challenges in applying plant hydraulic techniques. *New Phytologist 205*(3), 961–964.

Kattge, J., G. Bönisch, S. Díaz, S. Lavorel, I. C. Prentice, P. Leadley, S. Tautenhahn, G. D. Werner, T. Aakala, M. Abedi, et al. (2020). Try plant trait database–enhanced coverage and open access. *Global change biology 26*(1), 119–188.

Katz, D. C., M. N. Grote, and T. D. Weaver (2016). A mixed model for the relationship between climate and human cranial form. *American Journal of Physical Anthropology 160*(4), 593–603.

Kelly, R., K. Healy, M. Anand, M. E. Baudraz, M. Bahn, B. E. Cerabolini, J. H. Cornelissen, J. M. Dwyer, A. L. Jackson, J. Kattge, et al. (2021). Climatic and evolutionary contexts are required to infer plant life history strategies from functional traits at a global scale. *Ecology Letters 24*(5), 970–983.

Kohn, M. J. (2010). Carbon isotope compositions of terrestrial c3 plants as indicators of (paleo) ecology and (paleo) climate. *Proceedings of the National Academy of Sciences 107*(46), 19691–19695.

Kontopoulos, D.-G., E. van Sebille, M. Lange, G. Yvon-Durocher, T. G. Barraclough, and

S. Pawar (2020). Phytoplankton thermal responses adapt in the absence of hard thermodynamic constraints. *Evolution 74*(4), 775–790.

Koricheva, J. and J. Gurevitch (2014). Uses and misuses of meta-analysis in plant ecology. *Journal of Ecology 102*(4), 828–844.

Kozak, K. H. and J. J. Wiens (2010). Accelerated rates of climatic-niche evolution underlie rapid species diversification. *Ecology letters 13*(11), 1378–1389.

Kruuk, L. E. and J. D. Hadfield (2007). How to separate genetic and environmental causes of similarity between relatives. *Journal of evolutionary biology 20*(5), 1890–1903.

Kubota, Y., B. Kusumoto, T. Shiono, and T. Tanaka (2017). Phylogenetic properties of tertiary relict flora in the east asian continental islands: imprint of climatic niche conservatism and in situ diversification. *Ecography 40*(3), 436–447.

Laumer, C. E., R. Fernández, S. Lemer, D. Combosch, K. M. Kocot, A. Riesgo, S. C. Andrade, W. Sterrer, M. V. Sørensen, and G. Giribet (2019). Revisiting metazoan phylogeny with genomic sampling of all phyla. *Proceedings of the royal society B 286*(1906), 20190831.

Leggett, H. C., C. K. Cornwallis, A. Buckling, and S. A. West (2017). Growth rate, transmission mode and virulence in human pathogens. *Philosophical Transactions of the Royal Society B: Biological Sciences 372*(1719), 20160094.

Liu, H., Q. Ye, and J. J. Wiens (2020). Climatic-niche evolution follows similar rules in plants and animals. *Nature Ecology & Evolution 4*(5), 753–763.

Losos, J. B. (2008). Phylogenetic niche conservatism, phylogenetic signal and the relationship between phylogenetic relatedness and ecological similarity among species. *Ecology letters 11*(10), 995–1003.

Losos, J. B. (2011). Seeing the forest for the trees: The limitations of phylogenies in comparative biology: (american society of naturalists address). *The American Naturalist 177*(6), 709–727.

Lynch, M. (1991). Methods for the analysis of comparative data in evolutionary biology. *Evolution 45*(5), 1065–1080.

Markovski, J., C. T. Callaghan, W. K. Cornwell, and S. Nakagawa (2023). A global analysis reveals the dynamic relationship between sexual selection and population abundance in space and time.

Marques, I., T. Kneib, and N. Klein (2022). Mitigating spatial confounding by explicitly correlating gaussian random fields. *Environmetrics 33*(5), e2727.

Martins, E. P. and T. F. Hansen (1997). Phylogenies and the comparative method: a general approach to incorporating phylogenetic information into the analysis of interspecific data. *The American Naturalist 149*(4), 646–667.

Meyer, K. (1991). Estimating variances and covariances for multivariate animal models by restricted maximum likelihood. *Genetics Selection Evolution 23*(1), 67–83.

Minocher, R., P. Duda, and A. V. Jaeggi (2019). Explaining marriage patterns in a globally representative sample through socio-ecology and population history: A bayesian phylogenetic analysis using a new supertree. *Evolution and human behavior 40*(2), 176–187.

Mitov, V., K. Bartoszek, G. Asimomitis, and T. Stadler (2020). Fast likelihood calculation for multivariate gaussian phylogenetic models with shifts. *Theoretical Population Biology 131*, 66–78.

Mitov, V., K. Bartoszek, and T. Stadler (2019). Automatic generation of evolutionary hypotheses using mixed gaussian phylogenetic models. *Proceedings of the National Academy of Sciences 116*(34), 16921–16926.

Münkemüller, T., S. Lavergne, B. Bzeznik, S. Dray, T. Jombart, K. Schiffers, and W. Thuiller (2012). How to measure and test phylogenetic signal. *Methods in Ecology and Evolution 3*(4), 743–756.

Nakagawa, S., P. C. D. Johnson, and H. Schielzeth (2017, 9). The coefficient of determination $r^2$ and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. *Journal of The Royal Society Interface 14*, 20170213.

Nakagawa, S., D. W. Noble, A. M. Senior, and M. Lagisz (2017). Meta-evaluation of meta-analysis: ten appraisal questions for biologists. *BMC biology 15*(1), 1–14.

Nakagawa, S. and E. S. Santos (2012). Methodological issues and advances in biological meta-analysis. *Evolutionary Ecology 26*, 1253–1274.

Nakagawa, S. and H. Schielzeth (2010). Repeatability for gaussian and non-gaussian data: a practical guide for biologists. *Biological Reviews 85*(4), 935–956.

Ormerod, J. T. and M. P. Wand (2010). Explaining variational approximations. *The American Statistician*, 140–153.

Ovaskainen, O. and N. Abrego (2020). *Joint Species Distribution Modelling: With Applications in R*. Ecology, Biodiversity and Conservation. Cambridge University Press.

Ovaskainen, O., G. Tikhonov, A. Norberg, F. Guillaume Blanchet, L. Duan, D. Dunson, T. Roslin, and N. Abrego (2017). How to make more out of community data? a conceptual framework and its implementation as models and software. *Ecology letters 20*(5), 561–576.

Pagel, M. (1999). Inferring the historical patterns of biological evolution. *Nature 401*(6756), 877–884.

Pagel, M., C. O'Donovan, and A. Meade (2022). General statistical model shows that macroevolutionary patterns and processes are consistent with darwinian gradualism. *Nature communications 13*(1), 1–12.

Pearl, J. (1995). Causal diagrams for empirical research. *Biometrika 82*(4), 669–688.

Peixoto, F. P., F. Villalobos, and M. V. Cianciaruso (2017). Phylogenetic conservatism of climatic niche in bats. *Global Ecology and Biogeography 26*(9), 1055–1065.

Penone, C., A. D. Davidson, K. T. Shoemaker, M. Di Marco, C. Rondinini, T. M. Brooks, B. E. Young, C. H. Graham, and G. C. Costa (2014). Imputation of missing data in life-history trait datasets: Which approach performs the best? *Methods in Ecology and Evolution 5*(9), 961–970.

Pettersen, A. K., N. Feiner, D. W. Noble, G. M. While, T. Uller, and C. K. Cornwallis (2023). Maternal behavioral thermoregulation facilitated evolutionary transitions from egg laying to live birth. *Evolution Letters*, qrad031.

Popovic, G. C., D. I. Warton, F. J. Thomson, F. K. Hui, and A. T. Moles (2019). Untangling

direct species associations from indirect mediator species effects with graphical models. *Methods in Ecology and Evolution 10*(9), 1571–1583.

Pottier, P., D. W. Noble, F. Seebacher, N. C. Wu, M. Lagisz, L. Schwanz, S. M. Drobniak, and S. Nakagawa (2023). New horizons for comparative studies and meta-analyses.

Prieto, I., J. I. Querejeta, J. Segrestin, F. Volaire, and C. Roumet (2018). Leaf carbon and oxygen isotopes are coordinated with the leaf economics spectrum in mediterranean rangeland species. *Functional Ecology 32*(3), 612–625.

Quental, T. B. and C. R. Marshall (2010). Diversity dynamics: molecular phylogenies need the fossil record. *Trends in ecology & evolution 25*(8), 434–441.

Reich, P. B. (2014). The world-wide 'fast–slow' plant economics spectrum: a traits manifesto. *Journal of ecology 102*(2), 275–301.

Revell, L. J. (2010). Phylogenetic signal and linear regression on species data. *Methods in Ecology and Evolution 1*(4), 319–329.

Revell, L. J., L. J. Harmon, and D. C. Collar (2008). Phylogenetic signal, evolutionary process, and rate. *Systematic biology 57*(4), 591–601.

Riley, R. D., D. Jackson, G. Salanti, D. L. Burke, M. Price, J. Kirkham, and I. R. White (2017). Multivariate and network meta-analysis of multiple outcomes and multiple treatments: rationale, concepts, and examples. *bmj 358*.

Roberts, K. E. and B. Longdon (2023). Heterogeneities in infection outcomes across species: sex and tissue differences in virus susceptibility. *Peer community journal 3*.

Rohlf, F. J. (2001). Comparative methods for the analysis of continuous variables: geometric interpretations. *Evolution 55*(11), 2143–2160.

Runcie, D. E., J. Qu, H. Cheng, and L. Crawford (2021). Megalmm: mega-scale linear mixed models for genomic predictions with thousands of traits. *Genome biology 22*(1), 1–25.

Sanchez-Martinez, P., J. Martínez-Vilalta, K. G. Dexter, R. A. Segovia, and M. Mencuccini (2020). Adaptation and coordinated evolution of plant hydraulic traits. *Ecology Letters 23*(11), 1599–1610.

Siefert, A., C. Violle, L. Chalmandrier, C. H. Albert, A. Taudiere, A. Fajardo, L. W. Aarssen, C. Baraloto, M. B. Carlucci, M. V. Cianciaruso, et al. (2015). A global meta-analysis of the relative extent of intraspecific trait variation in plant communities. *Ecology letters 18*(12), 1406–1419.

Simpson, K. J., B. S. Ripley, P.-A. Christin, C. M. Belcher, C. E. Lehmann, G. H. Thomas, and C. P. Osborne (2016). Determinants of flammability in savanna grass species. *Journal of Ecology 104*(1), 138–148.

Slater, G. J. (2013). Phylogenetic evidence for a shift in the mode of mammalian body size evolution at the c retaceous-p alaeogene boundary. *Methods in Ecology and Evolution 4*(8), 734–744.

Slater, G. J., L. J. Harmon, and M. E. Alfaro (2012). Integrating fossils with molecular phylogenies improves inference of trait evolution. *Evolution: International Journal of Organic Evolution 66*(12), 3931–3944.

Smith, S. A. and J. M. Beaulieu (2009). Life history influences rates of climatic niche evolu-

tion in flowering plants. *Proceedings of the Royal Society B: Biological Sciences 276*(1677), 4345–4352.

Smith, S. A. and J. W. Brown (2018). Constructing a broadly inclusive seed plant phylogeny. *American journal of botany 105*(3), 302–314.

Smith, S. D., M. W. Pennell, C. W. Dunn, and S. V. Edwards (2020). Phylogenetics is the new genetics (for most of biodiversity). *Trends in Ecology & Evolution 35*(5), 415–425.

Symonds, M. R. and S. P. Blomberg (2014). A primer on phylogenetic generalised least squares. In *Modern phylogenetic comparative methods and their application in evolutionary biology*, pp. 105–130. Springer.

Thornhill, A. H., M. D. Crisp, C. Külheim, K. E. Lam, L. A. Nelson, D. K. Yeates, and J. T. Miller (2019). A dated molecular perspective of eucalypt taxonomy, evolution and diversification. *Australian Systematic Botany 32*(1), 29–48.

Tobias, J. A., C. K. Cornwallis, E. P. Derryberry, S. Claramunt, R. T. Brumfield, and N. Seddon (2014). Species coexistence and the dynamics of phenotypic evolution in adaptive radiation. *Nature 506*(7488), 359–363.

Tong, Y. (2012). *The Multivariate Normal Distribution*. Springer Series in Statistics. Springer New York.

Tung Ho, L. s. and C. Ané (2014). A linear-time algorithm for gaussian and non-gaussian trait evolution models. *Systematic biology 63*(3), 397–408.

Uyeda, J. C., D. S. Caetano, and M. W. Pennell (2015). Comparative analysis of principal components can be misleading. *Systematic Biology 64*(4), 677–689.

Uyeda, J. C. and L. J. Harmon (2014). A novel bayesian method for inferring and interpreting the dynamics of adaptive landscapes from phylogenetic comparative data. *Systematic biology 63*(6), 902–918.

Uyeda, J. C., R. Zenil-Ferguson, and M. W. Pennell (2018). Rethinking phylogenetic comparative methods. *Systematic Biology 67*(6), 1091–1109.

Vehtari, A., A. Gelman, and J. Gabry (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing 27*(5), 1413–1432.

Warton, D. I. (2022). *Eco-Stats: Data Analysis in Ecology: From t-tests to Multivariate Abundances*. Springer Nature.

Warton, D. I., F. G. Blanchet, R. B. O'Hara, O. Ovaskainen, S. Taskinen, S. C. Walker, and F. K. Hui (2015). So many variables: joint modeling in community ecology. *Trends in ecology & evolution 30*(12), 766–779.

Warton, D. I., I. J. Wright, D. S. Falster, and M. Westoby (2006). Bivariate line-fitting methods for allometry. *Biological reviews 81*(2), 259–291.

Watanabe, S. (2010, dec). Asymptotic equivalence of bayes cross validation and widely applicable information criterion in singular learning theory. *J. Mach. Learn. Res. 11*, 3571–3594.

Weir, C. J., I. Butcher, V. Assi, S. C. Lewis, G. D. Murray, P. Langhorne, and M. C. Brady (2018). Dealing with missing standard deviation and mean values in meta-analysis of continuous outcomes: a systematic review. *BMC medical research methodology 18*(1), 1–14.

Westoby, M., M. R. Leishman, and J. M. Lord (1995). On misinterpreting the 'phylogenetic correction'. *Journal of Ecology 83*(3), 531–534.

Westoby, M., L. Yates, B. Holland, and B. Halliwell (2023). Phylogenetically conservative trait correlation: quantification and interpretation. *Journal of Ecology*.

Wiens, J. J., D. D. Ackerly, A. P. Allen, B. L. Anacker, L. B. Buckley, H. V. Cornell, E. I. Damschen, T. Jonathan Davies, J.-A. Grytnes, S. P. Harrison, et al. (2010). Niche conservatism as an emerging principle in ecology and conservation biology. *Ecology letters 13*(10), 1310–1324.

Wilson, A. (2008). Why h2 does not always equal va/vp? *Journal of evolutionary biology 21*(3), 647–650.

Wilson, A. J., D. Reale, M. N. Clements, M. M. Morrissey, E. Postma, C. A. Walling, L. E. Kruuk, and D. H. Nussey (2010). An ecologist's guide to the animal model. *Journal of animal ecology 79*(1), 13–26.

Wright, I. J., P. B. Reich, J. H. Cornelissen, D. S. Falster, P. K. Groom, K. Hikosaka, W. Lee, C. H. Lusk, Ü. Niinemets, J. Oleksyn, et al. (2005). Modulation of leaf economic traits and trait relationships by climate. *Global Ecology and Biogeography 14*(5), 411–421.

Wright, I. J., P. B. Reich, M. Westoby, D. D. Ackerly, Z. Baruch, F. Bongers, J. Cavender-Bares, T. Chapin, J. H. Cornelissen, M. Diemer, et al. (2004). The worldwide leaf economics spectrum. *Nature 428*(6985), 821–827.

Yates, L. A., Z. Aandahl, S. A. Richards, and B. W. Brook (2022). Cross validation for model selection: a review with examples from ecology. *Ecological Monographs*, e1557. e1557 ECM22-0048.R1.

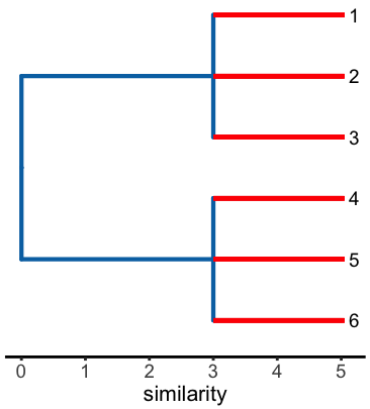
Yates, L. A., S. A. Richards, and B. W. Brook (2021). Parsimonious model selection using information theory: a modified selection rule. *Ecology 102*(10), e03475.

Yeates, D. K., K. Meusemann, M. Trautwein, B. Wiegmann, and A. Zwick (2016). Power, resolution and bias: recent advances in insect phylogeny driven by the genomic revolution. *Current Opinion in Insect Science 13*, 16–23.

Young, A. D. and J. P. Gillung (2020). Phylogenomics—principles, opportunities and pitfalls of big-data phylogenetics. *Systematic Entomology 45*(2), 225–247.

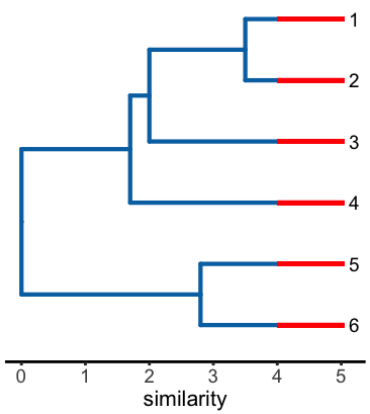
Yuan, M. and Y. Lin (2007). Model selection and estimation in the gaussian graphical model. *Biometrika 94*(1), 19–35.

**A**




$$\rightarrow \Sigma = \sigma_{\mathbf{b}}^{\mathbf{2}} C + \sigma_{\mathbf{e}}^{\mathbf{2}} I = \begin{pmatrix} \mathbf{3} & \mathbf{3} & \mathbf{3} & 0 & 0 & 0 \\ \mathbf{3} & \mathbf{3} & \mathbf{3} & 0 & 0 & 0 \\ \mathbf{3} & \mathbf{3} & \mathbf{3} & 0 & 0 & 0 \\ 0 & 0 & 0 & \mathbf{3} & \mathbf{3} & \mathbf{3} \\ 0 & 0 & 0 & \mathbf{3} & \mathbf{3} & \mathbf{3} \\ 0 & 0 & 0 & \mathbf{3} & \mathbf{3} & \mathbf{3} \end{pmatrix} + \begin{pmatrix} \mathbf{2} & 0 & 0 & 0 & 0 & 0 \\ 0 & \mathbf{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & \mathbf{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & \mathbf{2} & 0 & 0 \\ 0 & 0 & 0 & 0 & \mathbf{2} & 0 \\ 0 & 0 & 0 & 0 & 0 & \mathbf{2} \end{pmatrix}$$

**B**




$$\rightarrow \Sigma = \sigma_{\mathbf{b}}^{\mathbf{2}} C + \sigma_{\mathbf{e}}^{\mathbf{2}} I = \begin{pmatrix} \mathbf{4} & \mathbf{3.5} & \mathbf{2} & \mathbf{1.7} & 0 & 0 \\ \mathbf{3.5} & \mathbf{4} & \mathbf{2} & \mathbf{1.7} & 0 & 0 \\ \mathbf{2} & \mathbf{2} & \mathbf{4} & \mathbf{1.7} & 0 & 0 \\ \mathbf{1.7} & \mathbf{1.7} & \mathbf{1.7} & \mathbf{4} & 0 & 0 \\ 0 & 0 & 0 & 0 & \mathbf{4} & \mathbf{2.8} \\ 0 & 0 & 0 & 0 & \mathbf{2.8} & \mathbf{4} \end{pmatrix} + \begin{pmatrix} \mathbf{1} & 0 & 0 & 0 & 0 & 0 \\ 0 & \mathbf{1} & 0 & 0 & 0 & 0 \\ 0 & 0 & \mathbf{1} & 0 & 0 & 0 \\ 0 & 0 & 0 & \mathbf{1} & 0 & 0 \\ 0 & 0 & 0 & 0 & \mathbf{1} & 0 \\ 0 & 0 & 0 & 0 & 0 & \mathbf{1} \end{pmatrix}$$

Figure 1: A) Tree and matrix representations of the total covariance structure $\Sigma$ for data collected using a nested sampling design ($\sigma_b^2 = 3$ and $\sigma_e^2 = 2$) B) Tree and matrix representations of the total covariance structure, $\Sigma$, among species trait values ($\sigma_b^2 = 4$ and $\sigma_e^2 = 1$). Blue and red text used to emphasise the translation of phylogenetic covariance and independent (i.e., non-phylogenetic) error variance respectively to the graphs on the left. Non-phylogenetic variance represents extensions to the terminal branches of the graph (diagonal elements in $I$), reducing the overall expectation of similarity between observations.

**correlation**

$$\begin{array}{c c} & \begin{array}{ccc} x & y & z \end{array} \\ \begin{array}{c} x \\ y \\ z \end{array} & \begin{pmatrix} 1 & 0.56 & 0.75 \\ 0.56 & 1 & 0.75 \\ 0.75 & 0.75 & 1 \end{pmatrix} \end{array}$$

**partial correlation**

$$\begin{array}{c c} & \begin{array}{ccc} x & y & z \end{array} \\ \begin{array}{c} x \\ y \\ z \end{array} & \begin{pmatrix} 1 & 0 & 0.6 \\ 0 & 1 & 0.6 \\ 0.6 & 0.6 & 1 \end{pmatrix} \end{array}$$

Figure 2: Correlations between species traits $x$, $y$ and $z$ (upper left), with the same information expressed qualitatively as a graph (lower left). To evaluate whether the relationship between $x$ and $y$ can be explained by a covariate $z$, we compute partial correlations from the precision matrix of trait covariances, which quantifies the relationship between $x$ and $y$ when controlling for $z$ (upper right). In this example, despite strong covariance at the phenotypic level (upper left), the relationship between $x$ and $y$ is fully explained by a shared response to the covariate $z$.

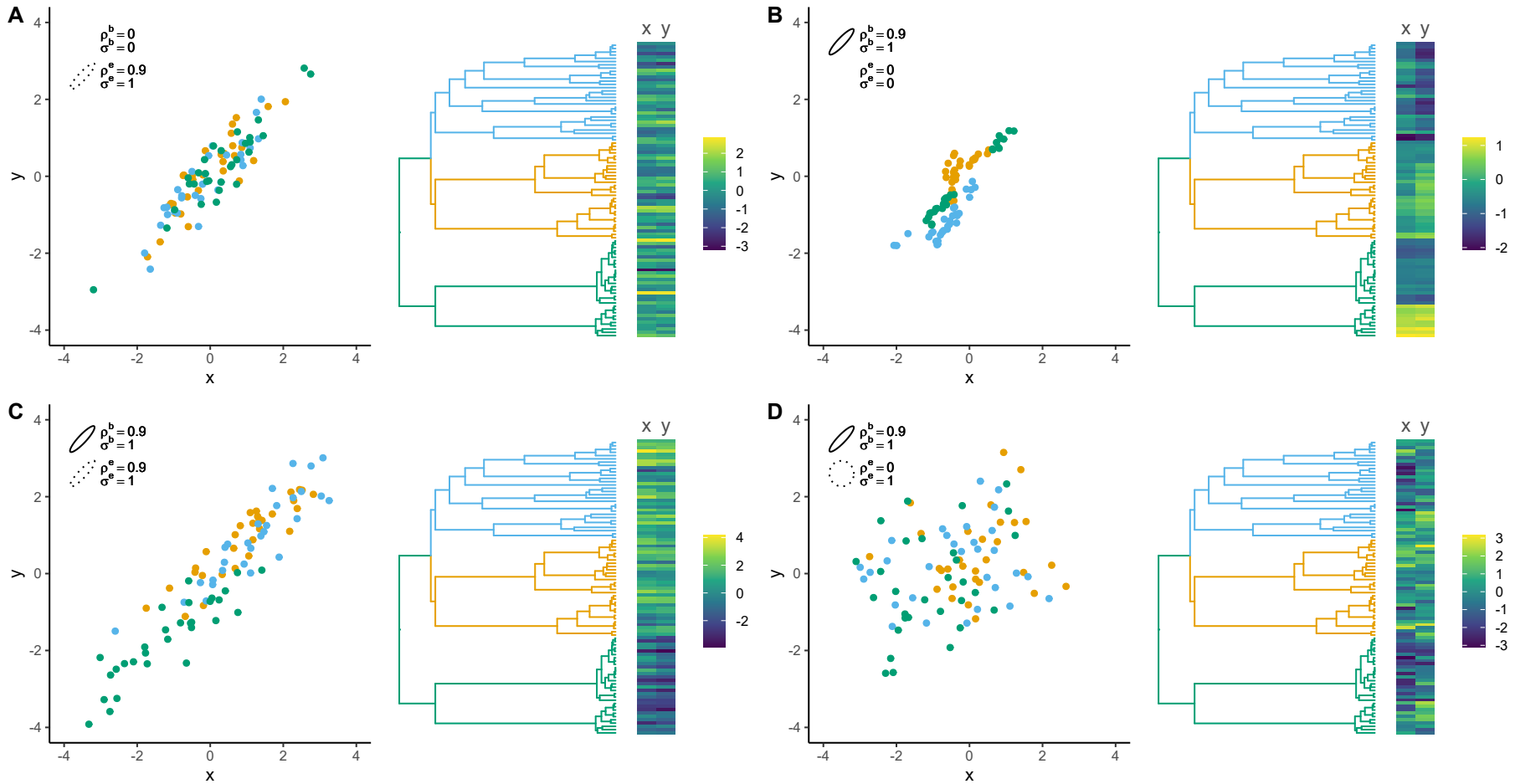



Figure 3: Bivariate trait data simulated from MR-PMM containing different levels of phylogenetic (superscript $b$) and non-phylogenetic (superscript $e$) (co)variance (A-D). Simulation conditions for each data set are inset in scatterplots. Trait data are plotted in scatterplots, as well as heatmaps arranged against the generating phylogeny.

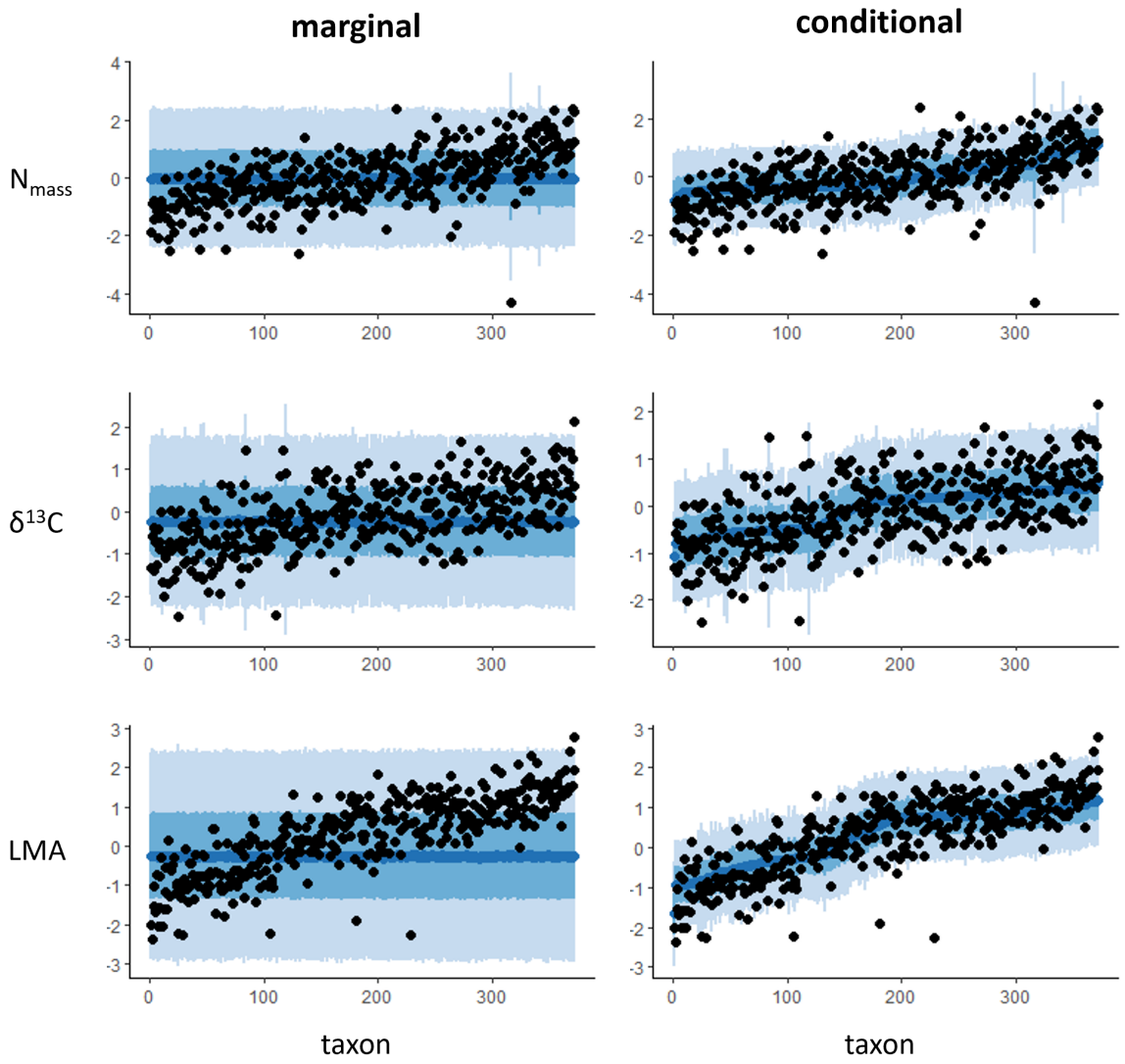



Figure 4: Posterior predictive checks from a fitted MR-PMM of leaf traits in Eucalyptus. Black points represent observed species values. Five-point summaries show the median (dark blue points), 0.5 CI (blue bars), and 0.95 CI (light blue bars) of the posterior predictive distribution for each taxon. For the marginal model, predictions are made at the population level by marginalising over the distribution of phylogenetic random effects. For the conditional model, predictions are made at the taxon level by conditioning on the estimated phylogenetic random effects, which significantly reduces predictive uncertainty. For each trait, taxa are ordered by the predictive mean of the conditional model.

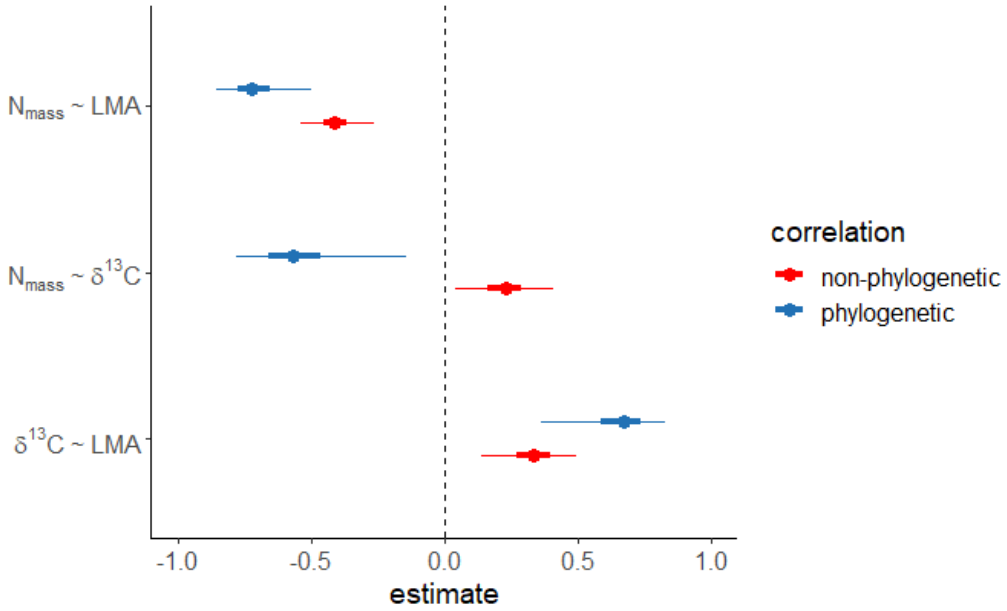



Figure 5: Phylogenetic (blue) and non-phylogenetic (red) correlation coefficient estimates from MR-PMM of leaf traits in Eucalyptus. Points represent the posterior median for each estimate, with 50% and 95% CIs represented by heavy and light wicks respectively.

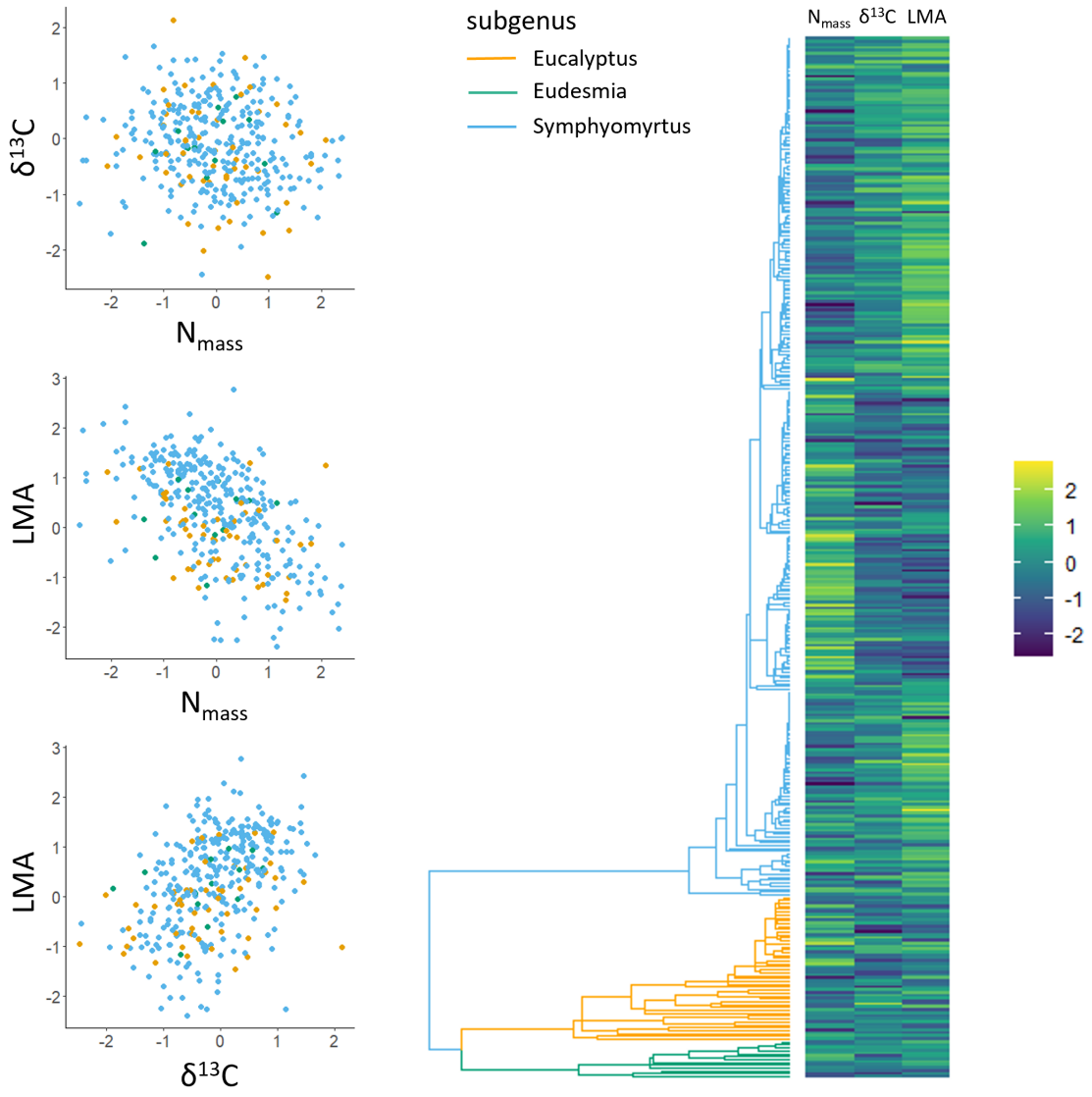



Figure 6: Scatter-plots (left) and heat-maps (right) of three leaf traits across 360 species of Eucalyptus. Trait values have been log-transformed and scaled. For heat maps (right), trait values are aligned with the corresponding species in the phylogeny (centre). For $\delta^{13}\mathbf{C}$ and $\mathbf{N}_{\text{mass}}$ (top left), opposing phylogenetic and non-phylogenetic correlations reported by the model (Figure 5) are obscured at the level of species phenotypes.