Biological Reviews

Multi-Response Phylogenetic Mixed Models: Concepts and Application

Ben Halliwell[1,2,3], Barbara R. Holland [1,2], and Luke A. Yates[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.v6.0.0" available at https://zenodo.org/records/11188867
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, generalised 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) PMM to introduce key concepts and outline the limitations of this approach for characterizing patterns of trait co-evolution. We emphasise MR-PMM as a preferable approach to analyses involving multiple species traits, due to the explicit decomposition of trait covariance across levels in the model hierarchy. We discuss multilevel distributional models, meta-analyses, multivariate models of evolution, and extensions to non-Gaussian response traits. We highlight techniques for causal inference using precision matrices, as well as advanced topics including prior specification and latent factor models. Using simulated data and visual examples, we discuss interpretation, prediction, and model validation. We implement many of the techniques discussed in example analyses of plant functional traits to demonstrate the general utility of MR-PMM in handling complex real world datasets. 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 **9**

2.1 Mixed Models 9

2.2 Phylogenetic Mixed Models 11

2.2.1 Phylogenetic signal 12

2.2.2 Models of evolution 13

2.2.3 Limitations of single response models 14

2.3 Multi-Response Phylogenetic Mixed Models (MR-PMM) 15

2.3.1 Implications 17

3 MR-PMM - extensions to the basic model **18**

3.0.1 Multilevel models 18

3.0.2 Meta-analyses 20

3.0.3 Distributional models 21

3.0.4 Multivariate models of evolution 23

3.0.5 Non-Gaussian response traits 24

3.0.6 Considerations for fixed effects 26

3.0.7 Partial correlations 27

4 Interpretation **28**

4.1 Phylogenetic (co)variances 29

4.2 Pattern versus process 29

4.3 Visualisations 30

4.4 Relevance for biological hypotheses 32

5 Prediction **32**

5.1 Predicting to new or missing observations 32

5.2 Model validation using posterior predictive checks 33

5.3 Predictive assessment using cross validation 34

6 Example Analysis - leaf traits in Eucalyptus **35**

6.1 Methods 35

6.2 Results 38

7 Extended Topics **41**

7.1 Priors for multivariate normal distributions 41

7.2 Latent factor methods for dimension reduction 42

8 Discussion **44**

8.1 Summary 44

8.2 Strengths and Weaknesses 45

8.3 Recommendations and Future directions 47

8.4 Conclusions 49

9 Tutorial **50**

10 Acknowledgements **51**

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 flies (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 4.1). 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 (e.g., parameters of species reaction norms) (Pottier et al., 2024), as well as for disentangling 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 response correlations are not of specific interest; fitting multiple correlated response variables allows for the inclusion of partially missing data and may improve predictive accuracy due to the sharing of information across common grouping variables (see Section 5, also see Riley et al. 2017; Pottier et al. 2024).

Despite these benefits, and the example of pioneering researchers, MR-PMM remains underutilized, 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 framework with general utility for comparative biology, although maximum likelihood implementations do exist for certain special cases (Mitov et al., 2020; Butler et al., 2023). The capacity for MR-PMM to provide a more informative analysis of trait (co)evolution compared with SR-PMM, is well appreciated (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, doubts about data requirements, and practical barriers to implementation, continue to prevent common usage of MR-PMM among ecologists and evolutionary biologists. Thus, we begin section 2 with a brief conceptual introduction to mixed models. We define the basic SR-PMM, the different parametrizations used to quantify phylogenetic signal, and the 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 that are modeled in MR-PMM (e.g., phylogenetic and residual), how these covariance structures are specified using covariance matrices, and how these matrices are parameterised to estimate correlations

between species traits at different levels in the model hierarchy. We discuss extensions to the basic MR-PMM including multilevel distributional models, meta-analyses, multivariate models of trait evolution, non-Gaussian response traits, and techniques for causal inference using precision matrices and graphical models. Finally, we explore a number of extended topics to showcase emerging techniques, including prior specification for shrinkage of target parameters, as well as latent factor methods for dimension reduction. In section 4 we discuss interpretation and explore data simulated from MR-PMM to clarify the distinction between phylogenetic and non-phylogenetic components of trait correlation. In Section 5 we cover predictive assessment, including methods based on posterior predictive distributions and leave-one-out (LOO) cross-validation (CV). In section 6, we present an example analysis on leaf functional traits in Eucalyptus from the AusTraits database (Falster et al., 2021). Finally, in section 8 we discuss the strengths and weaknesses of MR-PMM, highlight emerging syntheses facilitated by the method, and provide practical advice for analysts.

To support our exposition, we provide extensive GitHub tutorials, including annotated code and mathematical appendices. These tutorials demonstrate how to simulate multivariate trait data containing phylogenetic structure, and fit corresponding MR-PMM in two popular R packages, `MCMCglmm` (Hadfield, 2010) and `brms` (Bürkner, 2017). They also cover many common tasks and challenges in the MR-PMM workflow, including data cleaning and manipulation, via example analyses of publicly available, real world datasets.

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 recognized 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. In a simple case, the hierarchical structure may be 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:

$$\texttt{y} \sim \texttt{x} + (\texttt{1}|\texttt{group}),$$

where `y` is the observed response, the measured covariate `x` specifies a fixed effect, and (1|`group`) specifies a random intercept (a random effect that adjusts the baseline mean) at the group level. 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.

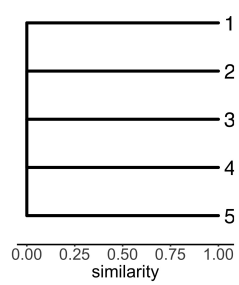
A more general mathematical form for mixed models is given by,

$$\mathbf{y} \;=\; \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e} \tag{1}$$

where $\mathbf{y}$ is a vector of observations. The design matrices $\mathbf{X}$ and $\mathbf{Z}$ relate fixed and random predictors to the data, the corresponding parameter vectors $\boldsymbol{\beta}$ and $\mathbf{u}$ contain the fixed and random effects to be estimated, and $\mathbf{e}$ is a vector of residual errors.

For the coded model formula, the random intercept (1|`group`) assumes the distribution of the random effect $\mathbf{u}$ is characterised by an identity matrix $\mathbf{I}$ (a matrix with 1's along the diagonal and 0's in all off-diagonals) equal in dimension to the number of groups (where $\mathbf{I}_{\mathrm{N_{group}}}$ is of dimension $\mathrm{N_{group}} \times \mathrm{N_{group}}$), and scaled by a group level variance, $\sigma^2_{\mathrm{group}}$. For example, for a study including data on 5 different species,

$$\mathbf{u} \sim \mathcal{N}(0, \sigma^2_{\mathrm{species}}\mathbf{I}_{\mathrm{N_{species}}}) \tag{2}$$



$$\rightarrow \mathbf{I}_{\mathrm{N_{species}}} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix},$$

where the notation $\sim$ means "is distributed as" and $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes a (multivariate) normal distribution with mean $\boldsymbol{\mu}$ and (co)variance $\boldsymbol{\Sigma}$. The off-diagonal elements of $\mathbf{I}_{\mathrm{N_{species}}}$ are all zero which implies independence between different levels of the grouping factor, such that no pair of species is expected to produce more similar observations than any other. Graphically, an identity matrix can be represented as a comb where tips (group levels) have no shared

edges (2). The residual errors $\mathbf{e}$ are also assumed to be independent *a priori*, capturing observation-specific deviations from the modelled mean,

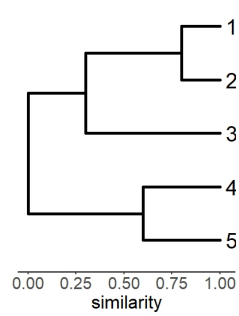
$$\mathbf{e} \sim \mathcal{N}(0, \sigma^2_{\text{res}} \mathbf{I}_{\text{N}_{\text{obs}}}), \tag{3}$$

Mixed models permit the specification of complex variance structures using multiple random effects which may be given as nested (e.g., populations within species), crossed to act independently, or encoded with additional structure such as spatial or temporal relatedness. For example, if species are not uniformly distributed across the landscape, we may consider modeling spatial autocorrelation in species effects. This is achieved by adjusting our assumptions about the distribution of random effects $\mathbf{u}$. For example, by substituting the identity matrix with a covariance matrix derived from the pairwise distances between sampling sites. Structured random effects are employed in this way to model dependencies arising from a range of data-generating processes, including evolutionary processes of species diversification.

2.2 Phylogenetic Mixed Models

In analyses of inter-species data, dependence arises from phylogenetic signal: the tendency for closely related species to resemble each other due to the hierarchical evolutionary history of life (Felsenstein, 1985; Pagel, 1999). This phylogenetic structure is often represented by a phylogenetic correlation matrix $\mathbf{C} = (c_{ij})$—an $\text{N}_{\text{species}} \times \text{N}_{\text{species}}$ matrix that encodes our expectation of similarity among species phenotypes, where for a pair of species i and j, the expected correlation is given by the off-diagonal element c_{ij}. In PMM, $\mathbf{C}$ enters the model as the expected correlation of a species-level random effect scaled by the phylogenetic variance σ^2_{phy}. For example, given a phylogeny of 5 species, we may write

$$\mathbf{u} \sim \mathcal{N}(0, \sigma^2_{\text{phy}} \mathbf{C}) \tag{4}$$



$$\rightarrow \mathbf{C} = \begin{bmatrix} 1 & 0.8 & 0.3 & 0 & 0 \\ 0.8 & 1 & 0.3 & 0 & 0 \\ 0.3 & 0.3 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0.6 \\ 0 & 0 & 0 & 0.6 & 1 \end{bmatrix}.$$

Importantly, $\mathbf{C}$ is derived by assuming a particular model of evolutionary trait change along the branches of a phylogeny (Eq. 4 assumes Brownian Motion (BM), but see Section 2.2.2). $\mathbf{C}$ is typically supplied by the user, although joint inference of the phylogenetic tree and the phylogenetic mixed model is possible in some instances, e.g., using `BEAST` (Hassler et al. 2023; Bouckaert et al. 2019). In a Bayesian setting, phylogenetic uncertainty can also be incorporated by combining posterior samples across a set of candidate topologies (Villemereuil et al. 2012; also see Nakagawa and de Villemereuil 2019).

To summarise, PMM is a special case of a linear mixed model (1) that incorporates phylogenetic information via an associated correlation matrix, $\mathbf{C}$. Species trait values $\mathbf{y}$ are modelled as the sum of fixed effects, a phylogenetic random effect, and residual error (see Section 3.0.5 for non-Gaussian extensions). Additional random effects are permitted to model other sources of dependence among observations, such as study effects and within species replication (see Section 3.0.1).

2.2.1 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, phylogenetic signal (λ) describes the proportion of variance in $\mathbf{y}$ attributable to the phylogenetic structure and is expressed in terms of the estimated variance components as

$$\lambda = \frac{\sigma^2_{\text{phy}}}{\sigma^2_{\text{phy}} + \sigma^2_{\text{res}}}. \tag{5}$$

Additional variance components may be included in the denominator of (5) when random effects capture meaningful components of within- and between-species phenotypic variance, e.g., trait (co)variance owing to (shared) environmental effects (Kruuk and Hadfield, 2007). When fixed or random variance components are excluded from the denominator, (5) expresses phylogenetic signal conditional on those factors. This is generally desirable for covariates, which are often introduced to account for known biases or aspects of experimental design and thus do not contribute to the variation we seek to decompose (de Villemereuil et al., 2018). For example, when analysing data on morphometric measurements of preserved

museum specimens, we may include storage duration as a fixed effect covariate to account for differences in tissue shrinkage among specimens in the collection. Such sources of variation may be considered extraneous to the estimation of λ, and thus excluded from the denominator of (5).

2.2.2 Models of evolution

Values of $\lambda < 1$ imply a component of variation in $\mathbf{y}$ that cannot be explained by phylogenetic dependence, conditional on a phylogenetic tree (topology and branch lengths) and a specified model of trait evolution. The default model of evolution in PMM is Brownian Motion (BM)—a continuous-time stochastic process characterised by a random walk of Gaussian distributed increments. The process branches at each node in the phylogeny with 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; Felsenstein 1985).

In the original presentation of phylogenetic generalised least squares (PGLS) (Grafen, 1989), a scalar multiple of the phylogenetic correlation matrix is taken to represent the total covariance among observations, $\Sigma = \sigma^2\mathbf{C}$, which assumes traits have evolved via BM on the tree with no additional variation. This assumption is equivalent to fixing $\sigma^2_{\text{res}} = 0$ in (5). With subsequent developments, $\mathbf{C}$, 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). For a model with no additional random effects, this transformed phylogenetic correlation matrix $\mathbf{C}(\boldsymbol{\theta})$, is then multiplied by the scalar variance parameter σ^2 to define

$$\text{var}(\mathbf{u}) = \sigma^2\mathbf{C}(\boldsymbol{\theta}). \tag{6}$$

A common instance is when $\mathbf{C}(\boldsymbol{\theta})$ is a weighted sum of $\mathbf{C}$ and an identity matrix $\mathbf{I}$ (equal in dimension to $\mathbf{C}$) characterised by a single parameter, $\boldsymbol{\theta} = \lambda$ (Pagel, 1999),

$$\mathbf{C}(\lambda) = \lambda\mathbf{C} + (1 - \lambda)\mathbf{I}. \tag{7}$$

In this popular formulation of PGLS, the additive terms express the relative proportions,

scaled by λ, of the total variance attributable to phylogenetic and residual effects. For the special case of a single observation per species (e.g., a species mean), (6) and (7) can be combined to obtain equivalence with the model given by (2) and (3) for $\sigma^2 = \sigma^2_{\text{phy}} + \sigma^2_{\text{res}}$ and $\lambda = \frac{\sigma^2_{\text{phy}}}{\sigma^2_{\text{phy}} + \sigma^2_{\text{res}}}$ (also see Housworth et al. 2004; Cinar et al. 2022). This equivalence clarifies the relationships to other common methods: The basic PMM is equivalent to the original presentation of PGLS when $\Sigma = \sigma^2_{\text{phy}}\mathbf{C}$ (i.e., $\sigma^2_{\text{res}} = 0$) and to ordinary least squares when $\Sigma = \sigma^2_{\text{res}}\mathbf{I}$ (i.e., $\sigma^2_{\text{phy}} = 0$) (Westoby et al., 2023; Blomberg et al., 2012). The phylogenetic threshold model of Felsenstein (2005) is equivalent to PMM with a binary response trait and probit link (Hadfield 2015; see section 3.0.5 for details).

In addition to (7), 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). PMM supports several popular models (e.g., δ, κ, early burst (EB), Ornstein-Uhlenbeck (OU), accelerating-decelerating (ACDC)) by appropriate transformations to $\mathbf{C}$, as in (6). This extends the estimation of phylogenetic random effects to a range of evolutionary models besides BM (see Tung Ho and Ané 2014; Blomberg et al. 2003), for which model selection techniques can be used to discriminate among alternative evolutionary hypotheses (Revell and Harmon, 2022; Goolsby et al., 2017). Unfortunately however, neither `MCMCglmm` nor `brms` currently provide options for estimating the parameters of evolutionary models other than BM.

2.2.3 Limitations of single response models

PMM provides an elegant solution to the statistical problem of phylogenetic dependence in inter-species data. However, univariate implementations 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 can be compromised when predictor traits themselves contain phylogenetic signal (Westoby et al., 2023).

Suppose we observe a phenotypic correlation between two traits, $\mathbf{y}_1$ and $\mathbf{y}_2$, 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 $\mathbf{y}_1$ and $\mathbf{y}_2$ across species is due to a phylogenetically conserved relationship between these traits (e.g., phylogenetic niche conservatism, Losos 2008; Wiens et al. 2010), with the remainder due to causes that are independent of phylogeny (e.g., within-species trait covariance). If we model these data using a single-response model (1), treating $\mathbf{y}_2$ as a fixed predictor of $\mathbf{y}_1$ 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 $\mathbf{y}_2$ is confounded with (phylogenetic components of) the residual covariance structure (Warton 2022; also see Wilson 2008; Rausher 1992 and Marques et al. 2022 for analogous issues in quantitative genetics and spatial statistics). In a worst-case scenario, these two correlation components may cancel each other out, 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 $\mathbf{y}_1$ and $\mathbf{y}_2$ as response variables. That is, move all traits to the left-hand side of the model equation (9), into a stacked column-vector $\mathbf{y} = (\mathbf{y_1}, \mathbf{y_2})^{\mathrm{T}}$. 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 and which the predictors, 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 precision in the prediction of missing and/or partially observed new data. We argue this approach is usually more appropriate and more meaningful for comparative analyses including multiple species traits (also see Section 4).

2.3 Multi-Response Phylogenetic Mixed Models (MR-PMM)

In MR-PMM, all species traits are modelled jointly as response variables, which means both the random effects $\mathbf{u}$ and residual errors $\mathbf{e}$ associated with each response must also be modelled jointly. For Gaussian response traits, this is achieved directly using multivariate normal distributions (but see Section 3.0.5 for non-Gaussian response types). Due to the compact form of the mathematical notation, the multi-response model is written identically

to the single-response case,

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e}, \tag{8}$$

where $\mathbf{y} = (\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_{\mathrm{N_{traits}}})^\mathsf{T}$ contains observations for all traits and all species, stacked in a single column vector. To illustrate, we represent the linear predictor for the bivariate case by placing each trait in a separate row,

$$\mathbf{y} = \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{Z}_1\mathbf{u}_1 + \mathbf{e}_1 \\ \mathbf{X}_2\boldsymbol{\beta}_2 + \mathbf{Z}_2\mathbf{u}_2 + \mathbf{e}_2 \end{pmatrix} \tag{9}$$

with

$$\begin{aligned} \mathbf{u} &= (\mathbf{u}_1, \mathbf{u}_2)^\mathsf{T} \sim \mathcal{N}(0, \boldsymbol{\Sigma}^{\mathrm{phy}} \otimes \mathbf{C}) \\ \mathbf{e} &= (\mathbf{e}_1, \mathbf{e}_2)^\mathsf{T} \sim \mathcal{N}(0, \boldsymbol{\Sigma}^{\mathrm{res}} \otimes \mathbf{I}_{\mathrm{N_{obs}}}), \end{aligned} \tag{10}$$

where $\mathbf{X}_1\boldsymbol{\beta}_1$ and $\mathbf{Z}_1\mathbf{u}_1$ are the linear predictors for the fixed and random effects, respectively, for response trait $\mathbf{y}_1$, and similarly for $\mathbf{y}_2$. The notation adopted in (9) emphasizes the joint distribution of effects for each response trait, where between-trait covariances across both phylogenetic $\mathbf{u}$ and residual $\mathbf{e}$ components are modeled with the trait covariance matrices

$$\boldsymbol{\Sigma}^{\mathrm{phy}} = \begin{pmatrix} \Sigma_{11}^{\mathrm{phy}} & \Sigma_{12}^{\mathrm{phy}} \\ \Sigma_{21}^{\mathrm{phy}} & \Sigma_{22}^{\mathrm{phy}} \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Sigma}^{\mathrm{res}} = \begin{pmatrix} \Sigma_{11}^{\mathrm{res}} & \Sigma_{12}^{\mathrm{res}} \\ \Sigma_{21}^{\mathrm{res}} & \Sigma_{22}^{\mathrm{res}} \end{pmatrix}, \tag{11}$$

where the matrix elements $\Sigma_{ij}^{\mathrm{phy}}$ and $\Sigma_{ij}^{\mathrm{res}}$ are scalar covariance components between traits i and j, representing the phylogenetic and residual components of covariation, respectively. The Kronecker product $\otimes$ is a type of matrix multiplication which, using the phylogenetic component as an example, can be written as

$$\boldsymbol{\Sigma}^{\mathrm{phy}} \otimes \mathbf{C} = \begin{pmatrix} \Sigma_{11}^{\mathrm{phy}}\mathbf{C} & \Sigma_{12}^{\mathrm{phy}}\mathbf{C} \\ \Sigma_{21}^{\mathrm{phy}}\mathbf{C} & \Sigma_{22}^{\mathrm{phy}}\mathbf{C} \end{pmatrix}. \tag{12}$$

The Kronecker product has the effect of introducing phylogenetic dependence, via $\mathbf{C}$, into the covariance structure of each trait and each pairwise trait relationship (see tutorial for fully worked examples). The total variance of $\mathbf{y}$ is the sum of the variances of the phylogenetic and residual components, transformed to the dimension of the multivariate data vector using

the random effects design matrix $\mathbf{Z}$,

$$\text{var}(\mathbf{y}) = \mathbf{Z}(\mathbf{\Sigma}^{\text{phy}} \otimes \mathbf{C})\mathbf{Z}^{\text{T}} + \mathbf{\Sigma}^{\text{res}} \otimes \mathbf{I}_{\text{N}_{\text{obs}}}. \tag{13}$$

Despite the large dimension of var($\mathbf{y}$), the number of estimated (co)variance parameters in MR-PMM is $\text{N}_{\text{trait}}(\text{N}_{\text{trait}} + 1)$ (e.g., only 6 parameters for the bivariate case). For the general case, these parameters can be expressed in terms of the standard deviations σ_i and the correlations ρ_{ij} used to characterise the matrix elements of both $\mathbf{\Sigma}^{\text{phy}} = (\Sigma_{ij}^{\text{phy}})$ and $\mathbf{\Sigma}^{\text{res}} = (\Sigma_{ij}^{\text{res}})$ written as

$$\Sigma_{ij}^{\text{phy}} = \sigma_i^{\text{phy}} \sigma_j^{\text{phy}} \rho_{ij}^{\text{phy}} \qquad \text{and} \qquad \Sigma_{ij}^{\text{res}} = \sigma_i^{\text{res}} \sigma_j^{\text{res}} \rho_{ij}^{\text{res}}, \tag{14}$$

for $i, j = 1, .., \text{N}_{\text{traits}}$.

2.3.1 Implications

Moving from a SR to a MR model facilitates more refined inferences, improves prediction, and allows users to include all available data, not just complete cases. In terms of inference, the Kronecker products in (10) provides a means to include both phylogenetic and non-phylogenetic structures in the modelled (co)variance term for each trait and trait pair. Specifically, for the basic MR-PMM presented in (9) and (10), the estimated quantity $\widehat{\Sigma}_{ij}^{\text{phy}} > 0$ 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 $\widehat{\Sigma}_{ij}^{\text{res}} > 0$ provides evidence for a component of correlation that is independent of phylogeny (see Figure 2 for an illustration). In contrast, the SR model fails to account for signal in 'predictor' traits, leading to a confound between phylogenetic and residual effects (Westoby et al. 2023, also see Section 2.2.3).

For predictive goals, MR-PMM is both more precise and more accurate than univariate linear models due to sharing of information across levels that accounts for phylogenetic relationships. Prediction to missing trait values leverages all data across traits and species, weighting their contributions by the estimated phylogenetic and non-phylogenetic covariances (see Section 5). Meaningful reductions in predictive variance compared to SR models may require that traits are at least moderately correlated (Riley et al., 2017; Jackson et al., 2011). This

may apply even more so at the phylogenetic level, where estimates of covariance are generally more uncertain (see Tutorial). However, conserved trait correlations are not uncommon in nature (Westoby et al., 2023), with comparative studies often investigating hypotheses about coordination (e.g., trade-offs, co-selection) among species traits that display strong phylogenetic signal (e.g., Kelly et al. 2021; Sanchez-Martinez et al. 2020; Blomquist 2019). Indeed, recent work highlights the utility of MR-PMM for analysing function-valued traits such as performance curves, growth trajectories, or reaction norms (Pottier et al., 2024). Parameters that describe these functions (e.g., intercepts, slopes) are often correlated, and the strength and direction of these correlations may be biologically meaningful, providing opportunities to test hypotheses about the functional form of trait relationships (e.g., Pettersen et al. 2023; Kontopoulos et al. 2020).

Finally, the structure of the MR model allows us to flexibly include missing data. This means that users can retain observations where only a subset of the response traits have been observed, which improves inference (e.g., observations reporting values for just two response traits out of the complete set improve the estimation and decomposition of the (co)variances for these two traits). Including incomplete records usually means that more data (observations and species) can be included in analyses, especially as the number of included response variables increases and incomplete cases may comprise a substantial fraction of the data (see Tutorial).

3 MR-PMM - extensions to the basic model

In this section, we explore several useful extensions to the basic MR-PMM that expand the utility for comparative biology.

3.0.1 Multilevel models

Data for phylogenetic comparative analyses are typically compiled from multiple sources. This often yields multiple observations per species, with individual studies contributing observations on only a subset of focal traits for a subset of species. Mixed models like

MR-PMM offer a general framework for dealing with these complex, cross-classified effect structures via multilevel models. Benefits include the ability to partition variation among multiple hierarchical levels (e.g., between-species, within-species, within-individual), account for variation in sampling effort, and fit models to data containing a mixture of record types (McNeish 2021; Nakagawa and Santos 2012, also see Figure 4 in Nakagawa et al. (2017) for a helpful visualization). Furthermore, accounting for multilevel structure may be necessary 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).

Equation (9) can be extended to account for multilevel structure in the data by modelling additional random effect components in $\mathbf{u}$ that capture the sampling hierarchy. For example, to account for study effects, we include a random effect at the study level $\mathbf{u}^{\text{study}}$, which models variability between studies. To account for multiple observations per species, we include an additional, unstructured random effect at the species level $\mathbf{u}^{\text{ind}}$, which models variability between species that is independent of phylogeny. Each component term of $\mathbf{u}$ is associated with a (co)variance structure that specifies dependence within that component term (15). However, the effects of different component terms are assumed to be independent *a priori*,

$$\mathbf{u} = \begin{bmatrix} \mathbf{u}^{\text{phy}} \\ \mathbf{u}^{\text{ind}} \\ \mathbf{u}^{\text{study}} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \mathbf{\Sigma}^{\text{phy}} \otimes \mathbf{C} & 0 & 0 \\ 0 & \mathbf{\Sigma}^{\text{ind}} \otimes \mathbf{I}_{\text{N}_{\text{species}}} & 0 \\ 0 & 0 & \mathbf{\Sigma}^{\text{study}} \otimes \mathbf{I}_{\text{N}_{\text{studies}}} \end{bmatrix} \right), \tag{15}$$

where $\mathbf{\Sigma}^{\text{study}} = \text{diag}((\sigma_1^{\text{study}})^2, (\sigma_2^{\text{study}})^2, ..., (\sigma_{\text{Ntraits}}^{\text{study}})^2)$. Modelling these additional components in $\mathbf{u}$ has important consequences for the way trait variation is partitioned, affecting the interpretation of model parameters. For example, the random effects specified in (15) represent a shift from modelling all non-phylogenetic sources of trait correlation in the residual error $\mathbf{e}$ as in (9), to modelling non-phylogenetic between-species trait correlation explicitly, via $\mathbf{u}^{\text{ind}}$, study effects via $\mathbf{u}^{\text{study}}$, and any remaining trait (co)variance as residual error in $\mathbf{e}$. Residual correlation for such a model represents within-species trait correlation, which may require very large datasets to be robustly estimated (Zhou et al., 2022). In practice, this means residual errors are often treated as independent in multilevel MR-PMM (i.e., $\mathbf{\Sigma}^{\text{res}}$ is constrained to a diagonal matrix). For example, by setting `rescor = F` and `idh(trait):units` in `brms` and `MCMCglmm`, respectively (see tutorial).

Additional components could, of course, be added to $\mathbf{u}$ (15) to model other sources of

covariance between observations, such as spatial or temporal effects (e.g. Markovski et al. 2023; Gomes et al. 2023; also see Freckleton and Jetz 2009). This flexibility of multi-level models makes MR-PMM well-suited for large-scale comparative studies including complex hierarchically structured data.

3.0.2 Meta-analyses

Data compiled for phylogenetic comparative analyses often contain a mixture of record types, including individual-level observations as well as species means, with or without associated sampling variances. In some instances, effect size measures such as test statistics or model parameters with associated errors are specified as response variables. Integrating data with heterogeneous errors casts MR-PMM as a meta-analysis in which observations are weighted according to their uncertainty. Multi-response 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 (Pottier et al., 2024).

We emphasise that it is always preferable to fit models to observation-level data rather than statistical summaries of those data, such as means with associated sampling variances. In particular, models fit to species mean trait values neglect within-species variation, which can bias results (e.g., estimates of phylogenetic signal, Cinar et al. 2022) and ultimately obscure the relative contribution of biological and methodological drivers of the observed trait variation. However, meta-analytic implementations of MR-PMM may be desirable when summary statistics are the only available data for some species. In practice, observation-specific sampling variances s^2_{ij}, $i = 1, ...\mathrm{N}_{\mathrm{obs}}$, $j = 1, ...\mathrm{N}_{\mathrm{traits}}$ are often modelled as additive components of the residual variance,

$$\mathrm{e}_{ij} \sim \mathcal{N}(0, (\sigma^{\mathrm{res}}_j)^2 + s^2_{ij}), \tag{16}$$

for example, using the `se()` and `mev()` arguments in `brms` and `MCMCglmm`, respectively. However, this approach is limited for data with a mixture of records, because individual-level observations are rarely reported with observational error, meaning only the sampling variances of species mean observations can be modelled. A straightforward approach is to weight each observation proportional to its reported sample size. This is implemented via the `weights` argument in `brms`, though currently unavailable in `MCMCglmm`. One workaround

is to produce a vector of inverse model weights to be passed into the model as sampling variances (e.g., $w^{-1} = 1/n$). However, this leads to scale issues, where the sampling variance added to individual-level observations ($n = 1$) may be large relative to the residual error.

3.0.3 Distributional models

Distributional models allow separate predicition for each parameter of the response distribution, e.g., the mean and variance of a Gaussian distribution (also known as a location-scale model). Biologically, these models distinguish between the central tendency (location) and variability (scale) of a trait, both of which can be influenced by genetic, environmental, and evolutionary factors (O'Dea et al., 2022; Westneat et al., 2015). Multilevel distributional models that include random intercepts in both the mean and variance, for a common grouping variable, introduce covariance and heteroskedasticity (non-constant variance), respectively—so called double-hierarchical models.

To illustrate, consider the multilevel MR-PMM discussed in Section 3.0.1. The $\mathbf{u}$ terms in (15) are *location* random effects, because each level in the hierarchy (e.g., $\mathbf{u}^{\text{phy}}$) adds a normally distributed random intercept to the model for the *mean*, which captures covariance attributable to the grouping variable. However, the model remains homoskedastic across species—all observations, irrespective of their phylogenetic effects, share a common residual variance—unless the model for the residual variance also depends on the group structure. To implement a group-structured variance model, we can specify

$$\begin{aligned} e_{ij} &\sim \mathcal{N}(0, (\sigma_{ij}^{\text{res}})^2) \\ \log \sigma_{ij}^{\text{res}} &= \alpha_j + \mathrm{t}_{j,\,\text{phy[i]}}, \end{aligned} \tag{17}$$

where, $(\sigma_{ij}^{\text{res}})^2$ is the residual variance of the ith observation for the jth trait, α_j is the fixed intercept for the variance of the jth trait, and $\mathrm{t}_{j,\,\text{phy[i]}}$ is a trait-specific random phylogenetic effect. To include the possibility of correlations between the random phylogenetic effects appearing in both the mean and variance, and thus share information across the distributional

parameters, it is preferable to specify a joint prior distribution (Bürkner, 2019),

$$\begin{bmatrix} \mathbf{u}_j^{\text{phy}} \\ \mathbf{t}_j^{\text{phy}} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} (\sigma_j^{\text{phy}})^2 & \sigma_j^{\text{phy}} \phi_j^{\text{phy}} \rho_j^{\text{phy}} \\ \sigma_j^{\text{phy}} \phi_j^{\text{phy}} \rho_j^{\text{phy}} & (\phi_j^{\text{phy}})^2 \end{bmatrix} \otimes \mathbf{I}_{\text{Nspecies}} \right). \tag{18}$$

While distributional models are increasingly used to understand variation in intra-individual behavioural variability (Mitchell et al., 2021; O'Dea et al., 2022), models such as (18) appear to be severely underutilised in phylogenetic comparative analyses (Horvath et al., 2023; Cleasby and Nakagawa, 2011). Indeed, the mean and variance of biological traits are often positively correlated (a special case of heteroskedasticity), routinely threatening the assumption of homogeneous variances in common statistical tests. In some instances, this issue is addressed by the mean-variance relationship determined by the choice of statistical distribution and link function (e.g., Poisson with log link), or else using variance-stabilising transformations (e.g., log transforming response data). However, heteroskedasticity may remain in spite of these efforts, and this 'variance of the variance' may in fact correspond to interesting biological effects that we would like to model explicitly. As with our model for the mean, we may include an arbitrary number of predictors for the variance, including fixed effects (e.g., methodological or environmental covariates), and random effects that contain further hierarchical structure. For example, a reasonable assumption for many phylogenetic comparative analyses is that residual variances should differ between species and that variation in residual variance should itself contain phylogenetic structure (i.e., closely related species should show similar amounts of residual variability). Furthermore, estimating correlations between phylogenetic random effects (as in 18) provides a method to test for conserved mean-variance relationships within and between traits.

For studies attempting to understand drivers of phenotypic variability, explicitly partitioning trait variance across known hierarchical structures has the potential to be extremely informative (O'Dea et al., 2022; McNeish, 2021; Westneat et al., 2015). Distributional models also have a clear statistical motivation; heteroskedasticity violates, by definition, the assumption of homogeneous variances, leading to inefficient parameter estimation and biased errors, with potential consequences for inference and hypothesis tests.

3.0.4 Multivariate models of evolution

In 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). Models of evolution comprise both stochastic and deterministic components and are naturally expressed as stochastic differential equations where the rate of change of trait values over time depends on current values. 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 more easily obtained (though still present convergence issues even for the most sophisticated algorithms (Butler et al., 2023)) since the covariance is proportional to $\mathbf{C}$ (12) up to a re-scaling of the off-diagonal elements (13), equivalent to a simple branch-length transformation. For more complex models of evolution, the covariance matrix cannot generally be obtained via 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,

$$\mathrm{d}\mathbf{y}(t) = \mathbf{A}\big(\mathbf{y}(t) - \boldsymbol{\eta}\big)\mathrm{d}t + \boldsymbol{\Lambda}\mathrm{d}\boldsymbol{W}(t), \tag{19}$$

where, $\mathbf{y}$ is an $\mathrm{N}_{\mathrm{traits}}$-vector of trait values, $\mathbf{A}$ is a $\mathrm{N}_{\mathrm{traits}} \times \mathrm{N}_{\mathrm{traits}}$ strength of selection (or rate of adaptation) matrix, $\boldsymbol{\eta}$ is an $\mathrm{N}_{\mathrm{traits}}$-vector of trait optima values, and $\boldsymbol{W}$ is an $\mathrm{N}_{\mathrm{traits}}$-dimensional Brownian process with diffusion matrix $\boldsymbol{\Lambda}$. Viewed in discrete time as a vector autoregressive model, the differential equation (19) provides a rule to update trait values based on their current values together with a (Brownian) stochastic step. Each off-diagonal element of $\mathbf{A}$ models the deterministic effect of one trait value on the update of another trait, permitting a test of Granger causality—the notion that historical values of one trait improve the prediction of another (Shojaie and Fox, 2022). Although Granger causality provides a more direct inference of causation than a correlation coefficient, especially given $\mathbf{A}$ in (19) need not be symmetric, it is not necessarily more meaningful than the trait correlations that we can derive from $\boldsymbol{\Lambda}$. Indeed, estimated correlations in the stochastic step may also capture causal relationships, just not those associated with adaptation toward trait optima.

For the MVOU model (19), the $\mathrm{N}_{\text{traits}} \times \mathrm{N}_{\text{traits}}$ 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),

$$\text{(20)} \qquad \mathrm{cov}(\mathbf{y}_i, \mathbf{y}_j) = e^{-\mathbf{A}(c_{ii}-c_{ij})} \left(\int_0^{c_{ij}} e^{-\mathbf{A}\nu} \mathbf{\Lambda}\mathbf{\Lambda}^{\mathsf{T}} e^{-\mathbf{A}^{\mathsf{T}}\nu} \, d\nu \right) e^{-\mathbf{A}^{\mathsf{T}}(c_{jj}-c_{ij})}.$$

where the phylogeny enters the computation via the matrix elements of $\mathbf{C} = (c_{ij})$. As intimidating as it appears, (20) can be evaluated directly provided the matrix $\mathbf{A}$ admits an eigendecomposition (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 $\mathbf{A}$ goes to zero (i.e., no attraction to optima), (20) reduces to the standard multivariate BM where $\mathbf{\Sigma}^{\text{phy}} = \mathbf{\Lambda}\mathbf{\Lambda}^{\mathsf{T}}$. Many common MVOU variants are characterised by strong constraints on the form of $\mathbf{A}$ which further simplify the evaluation of (20). For example, when attraction to optima is assumed to be equal and independent for all traits; i.e., when $\mathbf{A}$ is a multiple of the identity matrix (Goolsby et al., 2017; Tung Ho and Ané, 2014).

In principle, analytic computation of $\mathrm{cov}(\mathbf{y}_i, \mathbf{y}_j)$ extends MR-PMM to a broad class of multivariate Gaussian evolutionary models (e.g., for multivariate OU, it may be sufficient to use (20) in place of $\mathbf{\Sigma}^{\text{phy}} \otimes \mathbf{C}$ in (9)). 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 (Bouckaert et al., 2019; Mitov et al., 2020; Clavel et al., 2015; Goolsby et al., 2017). 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; together with modern algorithms for computing tree-structured multivariate likelihoods (Mitov et al., 2020; Hassler et al., 2022; Bastide et al., 2021). Progress in this direction would yield a powerful modelling framework, capable of synthesising model selection among alternate multivariate evolutionary models with the generalisable multilevel framework of MR-PMM.

3.0.5 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 approach models between-trait covariance in the usual way, via a multivariate Gaussian distribution, while at the same time permitting trait-specific probability distributions and associated link functions (Hadfield, 2010). For traits $i = 1, .., \mathrm{N}_{\mathrm{trait}}$, the latent-variable formulation is written

$$\boldsymbol{l} \;=\; \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e}, \tag{21}$$

where $\boldsymbol{l} = (\boldsymbol{l}_1, \boldsymbol{l}_2, ..., \boldsymbol{l}_{\mathrm{N}_{\mathrm{traits}}})^{\mathsf{T}}$ are latent variables associated with the observed traits and the corresponding (latent-trait) covariances are given by matrices $\boldsymbol{\Sigma}^{\mathrm{phy}}$ and $\boldsymbol{\Sigma}^{\mathrm{res}}$ which characterise the distributions of $\mathbf{u}$ and $\mathbf{e}$ as in (10); the latter are now 'pseudo' rather than true residuals introduced as observation-level random effects. A model for the observations $\mathbf{y}_i$ is given in terms of a trait-specific probability distribution f_i and link function η_i

$$\mathbf{y}_i \sim f_i(\eta_i^{-1}(\boldsymbol{l}_i), \boldsymbol{\phi}_i), \tag{22}$$

where $\boldsymbol{\phi}_i$ includes any distribution-specific parameters (if required). Depending on the choice of distribution f_i, additional constraints may need to be imposed on associated variance terms to ensure parameter identifiability.

Variance partitioning to define an analogue of λ (5) is possible for the latent-variable models given by (21) and (22), 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). Variance terms are usually more easily interpreted when they are expressed on the scale of trait observations rather than that of the latent variables (de Villemereuil, 2018). For multi-response models involving two or more distribution types, there are generally no closed-form solutions for the observation-level variances and associated statistics, however numerical methods are readily applied given point estimates or a set of posterior samples of the model parameters (de Villemereuil et al., 2016). A further complication for interpretation, however, is that the non-linear transformations required to transform covariances from the latent to the observation scale can change the bounds of the correlation coefficient. For example, the minimum correlation of two exponentiated random normal variables is approximately -0.37 and the correlation bounds of a random normal and an exponentiated random normal variable are approximately ± 0.76.

An alternative to specifying trait-specific probability distributions, is to model discrete or constrained trait observations as censored measurements of continuous latent variables. This idea generalises the threshold model (Albert and Chib, 1993) where the absence of non-linear link functions means that correlations between the continuous variables, both observed and imputed, can be directly interpreted on the scale of observed data (Clark et al., 2017). Implementation of this approach requires sampling from truncated multivariate normal distributions for which recently developed computational methods have substantially increased efficiency, especially for high-dimensional problems which were previously intractable (Zhang et al., 2023).

The software packages `MCMCglmm` and `brms` incorporate non-Gaussian traits using the approach given by (21) and (22), whereas `BEAST` implements the generalised probit model using the new efficient sampling approaches.

3.0.6 Considerations for fixed effects

There are two common 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 related 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. Equivalent procedures are used to control for environmental variation in multivariate species co-occurrence models (Ovaskainen et al., 2017; Warton et al., 2015). However, in MR-PMM, a preferable approach will often be to treat biological covariates as response traits and assess conditional dependencies via partial correlations.

3.0.7 Partial correlations

The practice of assessing conditional relationships between traits (i.e., relationships 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. In MR-PMM, we estimate conditional relationships from a joint model, where both focal traits and covariates are modelled as response variables. This is preferable 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). Thus, treating covariates as response variables can avoid potential confounds between fixed and phylogenetic effects (see section 2.2.3). Partial correlations between response traits in MR-PMM are derived from elements of the inverse trait-covariance matrices, called precision matrices $\mathbf{\Omega} = (\Omega_{ij})$; that is,

$$\mathbf{\Omega}^{\text{phy}} = (\mathbf{\Sigma}^{\text{phy}})^{-1} \qquad \text{and} \qquad \mathbf{\Omega}^{\text{res}} = (\mathbf{\Sigma}^{\text{res}})^{-1}, \tag{23}$$

for which the corresponding partial correlation coefficients are $-\Omega_{ij}/\sqrt{\Omega_{ii}\Omega_{jj}}$.

The estimation of partial correlations via precision matrices casts MR-PMM in the framework of Gaussian graphical models (Popovic et al., 2019; Epskamp et al., 2018; Yuan and Lin, 2007; Magwene, 2001), which have great potential for clarifying assumptions of phylogenetic comparative methods (Uyeda et al., 2018). Elements of these precision matrices relate directly to the existence of edges in a graphical causal network (Figure 1). For example, suppose we obtain data on three co-varying species traits x, y and z. In the precision view (Figure 1 right panel), the absence of an edge between x and y signifies conditional independence: that x and y are independent, given z. This does not imply that the correlation between x and y is zero (Figure 1 left panel), only that the partial correlation between these traits is zero. Precision matrices are often sparser than their corresponding covariance matrices (i.e., contain more off-diagonal elements that are effectively zero), which focuses our inference on a reduced set of candidate causal relationships. For example, Halliwell et al. (2024) used MR-PMM to estimate partial correlations between social behaviour (binary response) and a range of species traits and environmental niche variables in a global sample of 1696 lizard species. The authors used partial correlations to disentangle direct climate

effects on the evolution of social behaviour from indirect effects driven by adaptations to climate (i.e., traits correlated with climate) that go on to promote the evolution of social behaviour.

Outside of controlled experiments, there is always the strong possibility that partial correlations 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; also see tutorial), and are readily obtained from a fitted MR-PMM.

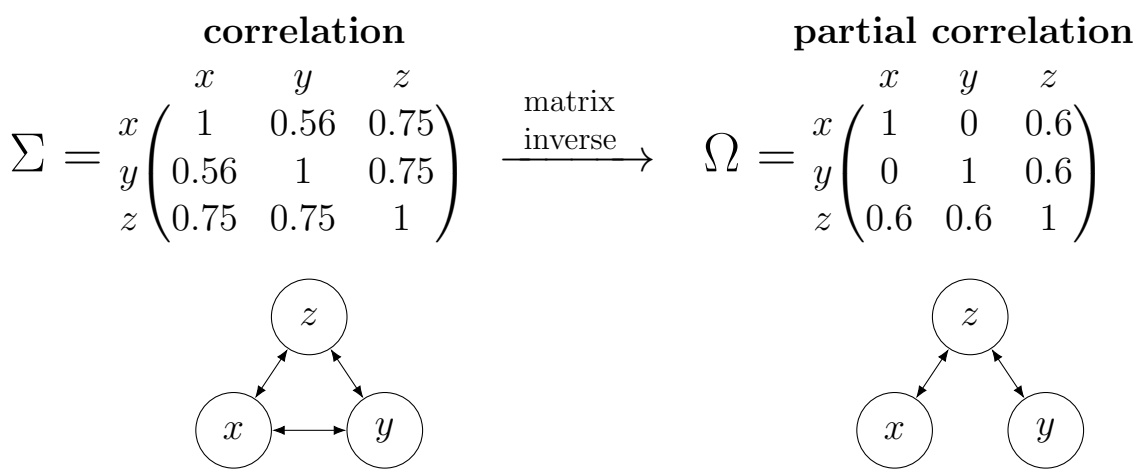


Figure 1: Correlation (Σ) and partial correlation ($\Omega = \Sigma^{-1}$) matrices between species traits x, y and z. The undirected network graphs below provide a qualitative representation of trait relationships, where the absence of an edge between two traits corresponds to a zero off-diagonal element in the matrix above. 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. In this example, x and y are strongly correlated, however this relationship is fully explained by a shared response to the covariate z, i.e., x and y are independent, conditional on z.

4 Interpretation

The biological interpretation of phylogenetic models and their 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.

4.1 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 stochastic processes, or reflect constraints and adaptive responses to selection, will vary for different traits, between clades, and across phylogenetic scales. These forces are also likely to fluctuate throughout the course of evolutionary history, complicating the interpretation of trait covariance (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". Expressed another way, the phylogenetic component of trait (co)variance in PMM models variation attributable to the phylogenetic relationships among taxa. Examples include variation arising from genetic differences between clades that have accumulated over evolutionary time, as well as non-genetic (e.g., spatial, environmental or developmental) effects that are phylogenetically structured for one reason or another.

4.2 Pattern versus process

There are inherent limitations to the inferences we can make about evolutionary processes from data observed only 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 (Thorson and van der Bijl, 2023; Uyeda et al., 2018; Quental and Marshall, 2010; Slater et al., 2012; Uyeda and Harmon, 2014).

Because phylogenetic and non-phylogenetic correlations can arise from multiple distinct

causal processes, we argue for caution around mechanistic interpretations. Two traits may be phylogenetically correlated 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 that involves phylogenetic niche conservatism (Westoby et al., 2023; Wiens et al., 2010). Phylogenetic correlations may also reflect indirect relationships between traits, due to similar selective pressures acting on multiple traits independently throughout evolutionary history. As above, such hypotheses should not be considered mutually exclusive (Losos, 2011; Revell et al., 2008). Deriving conditional dependencies can help evaluate evidence for alternative mechanistic hypotheses in MR-PMM (see Section 3.0.7). Ultimately, however, it is rarely the pleasure of a comparative biologist to declare causation. Rather, to uncover meaningful patterns of variation, test predictions from theory, and generate hypotheses for future comparative and experimental research.

4.3 Visualisations

To compare and contrast phylogenetic and non-phylogenetic (co)variance components, and clarify their biological interpretation, we simulated data from a simple MR-PMM for two Gaussian response traits y_1 and y_2, as in (9). Figure 2 shows scatter plots of simulated data together with the tree used to derive the phylogenetic correlation matrix C. The ellipses within the scatter plot provide a visual representation of the simulation conditions for each panel. A heat map of trait data plotted against the phylogeny highlights the distinct signature each source of covariance leaves in the data (see Figure 2 caption for details).

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.

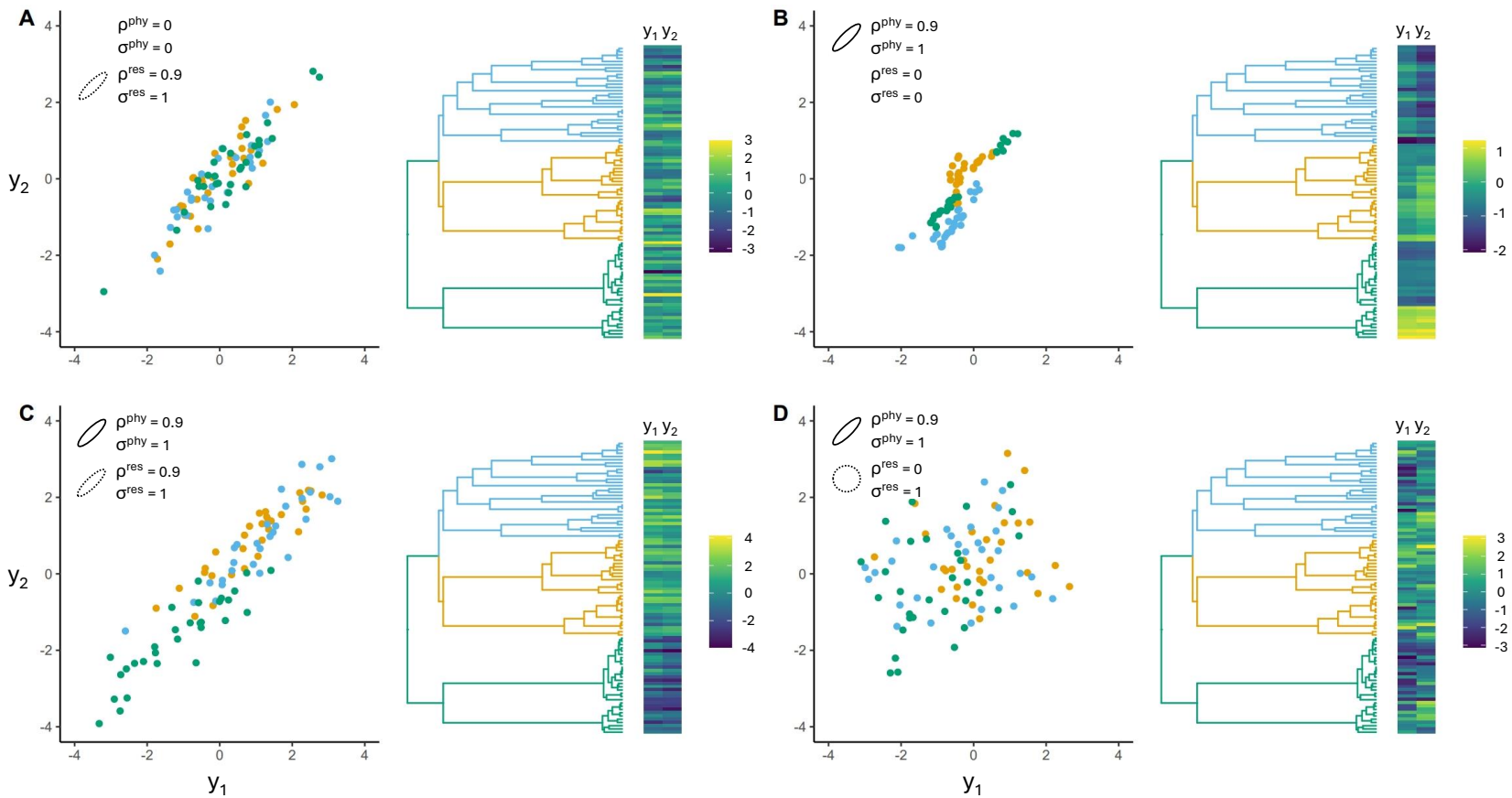


Figure 2: Bivariate trait data (y_1, y_2) simulated from a basic MR-PMM (9) containing different levels of phylogenetic (ρ^{phy}) and residual (ρ^{res}) correlation (**A**-**D**). Simulation conditions for each panel are inset at the top left of each scatterplot. Data are plotted in scatterplots, as well as heatmaps arranged against the generating phylogeny. In **A**, y_1 and y_2 have no phylogenetic signal ($\sigma^{\text{phy}} = 0$), but a strong positive residual correlation ($\rho^{\text{phy}} = 0, \rho^{\text{res}} = 0.9$). Clades overlap completely in the scatter plot and the heatmap shows bands of colour across y1 and y2 that appear random with respect to phylogeny. **B** shows the opposing situation, y_1 and y_2 are positively correlated but entirely with respect to phylogeny ($\rho^{\text{phy}} = 0.9, \rho^{\text{res}} = 0$) with no residual variation in either trait ($\sigma^{\text{res}} = 0$). The scatter plot shows clearly distinguishable clades and 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 promoting separation of clades along the major axis of co-variation. The heatmap shows phylogenetic structure weakening 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 **C**, y_1 and y_2 have equal phylogenetic and residual variances, and a strong positive correlation operating on both levels. This scenario, where phylogenetic and residual correlations are similar in sign and magnitude, is likely to be common for many biological traits. In **D**, both traits contain phylogenetic and residual variance, but correlation is only present on the phylogenetic level. This shows how easily conserved correlations are obscured when residual sources of variation contribute substantially to trait variance. Notably, **D** represents a set of conditions for which SR-PMM, such as PGLS, will typically fail to detect a significant association between y_1 and y_2 (Westoby et al., 2023).

4.4 Relevance for biological hypotheses

We see the flexible framework and nuanced interpretation of MR-PMM as having at least four strong usage cases for biologists: 1) to provide a more meaningful decomposition of trait (co)variances, including phylogenetic and non-phylogenetic components, among both Gaussian and non-Gaussian response traits; 2) to test for partial correlations, consistent with functional relationships between species traits, in a rigorous phylogenetic framework; 3) to test theory and generate new hypotheses about the ecological and evolutionary drivers of trait variation; 4) to predict species vulnerability to environmental change based on phylogenetically structured trait-trait and trait-environment relationships. 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. 2024; Westoby et al. 2023; Gallinat and Pearse 2021, also see Abrego et al. 2020).

5 Prediction

While the estimation of trait covariance, rather than trait prediction *per se*, is the focus of MR-PMM, the estimated covariance structure can be fully exploited for predictive goals. This approach 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.

5.1 Predicting to new or missing observations

Missing data is a perennial problem in comparative biology (Freckleton, 2009; Nakagawa and Freckleton, 2008). For many traits of interest, data collection remains too expensive,

technically demanding, or logistically challenging to keep pace with the expansion of species phylogenies and functional trait databases. To illustrate, we consider another example from the field of plant hydraulic ecophysiology. Physiological resistance to xylem cavitation is central to water stress tolerance in plants and thus a key metric for evaluating species and ecosystem vulnerability to climate change (Brodribb et al., 2020; Choat et al., 2012). While new methods allow direct quantification of cavitation resistance in diverse species (e.g., P_{50}, the water potential at which 50% of xylem conductivity is lost to cavitation (Brodribb et al., 2016, 2017)), observations require specialised equipment, technical proficiency, and are time consuming to obtain. However, because cavitation resistance forms part of a coordinated growth strategy involving conserved trait networks (Sanchez-Martinez et al., 2020; Skelton et al., 2021; Liu et al., 2024), phylogenetically structured trait correlations are likely to enhance our ability to predict this crucial, yet labor-intensive, trait where sufficient data on correlated traits are available. Recent work highlights strong correlations between xylem anatomical traits and P_{50} (Lens et al., 2022), suggesting functional indicator traits, some of which are cheaply and easily measured at scale (e.g., petiole XLA, Blackman et al. 2024). An MR-PMM specifying P_{50} (poor species coverage), functional traits (moderate species coverage), and environmental indices (good species coverage) as joint response variables could leverage all available data, phylogenetic relationships, as well as trait-trait and trait-environment correlations to impute missing P_{50} values for unobserved species.

In this multivariate mixed modelling context, we emphasise that prediction is both more precise and more accurate, relative to univariate linear models, because information is shared across species in a way that respects evolutionary history. The MR-PMM predictive model for missing trait values learns from all observed data across traits and species, weighting their predictive contributions according to the estimated phylogenetic and non-phylogenetic covariances, given the chosen model of evolution.

5.2 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 as-

sessments (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 3 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?

5.3 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, validation, or 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 typical assessments of model adequacy which concern 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 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 overfitting (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).

6 Example Analysis - leaf traits in Eucalyptus

To demonstrate potential applications of MR-PMM, we present an example analysis using data on leaf traits for 457 species of Eucalyptus from the AusTraits database (Falster et al., 2021). Our intentions for this analysis were to 1) decompose the variance in each trait and evaluate phylogenetic signal; 2) estimate phylogenetic and non-phylogenetic between-species trait correlations while accounting for multilevel structure in the data; and 3) evaluate partial correlations. Full details of this analysis, as well as additional examples exploring other model extensions with this dataset, are provided in the tutorial.

6.1 Methods

We derived the phylogenetic correlation matrix $\mathbf{C}$ from the maximum likelihood time-calibrated 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 ($\delta\,^{13}\mathbf{C}$). All responses were log-transformed, zero-centred and scaled to unit variance prior to analysis. The random effects were specified as in (15), to estimate phylogenetic and non-phylogenetic between-species trait correlations while accounting for study effects.

For model validation, we performed posterior predictive checks. The proportion of the data falling in the predictive intervals was close to the nominal quantiles, indicating that the model is adequately calibrated (although the proportions for $\delta\,^{13}\mathbf{C}$ were a little low), and visual inspection of the plotted distributions verified the capacity of the model to gener-

ate plausible data (Figure 3). The lower coverage of the posterior predictive intervals for $\delta^{13}\mathbf{C}$ may indicate that the model is missing relevant covariates. Indeed, between-study effects accounted for the largest proportion of variance in $\delta^{13}\mathbf{C}$, suggesting methodological differences between studies or local environmental factors may be a contributing factor to poor predictive performance. Further, the chosen probability distribution (Gaussian), or the log-transformation, may be less well suited to this trait. For predictive assessment using approximate LOO-CV (Bürkner et al., 2021), see the tutorial.

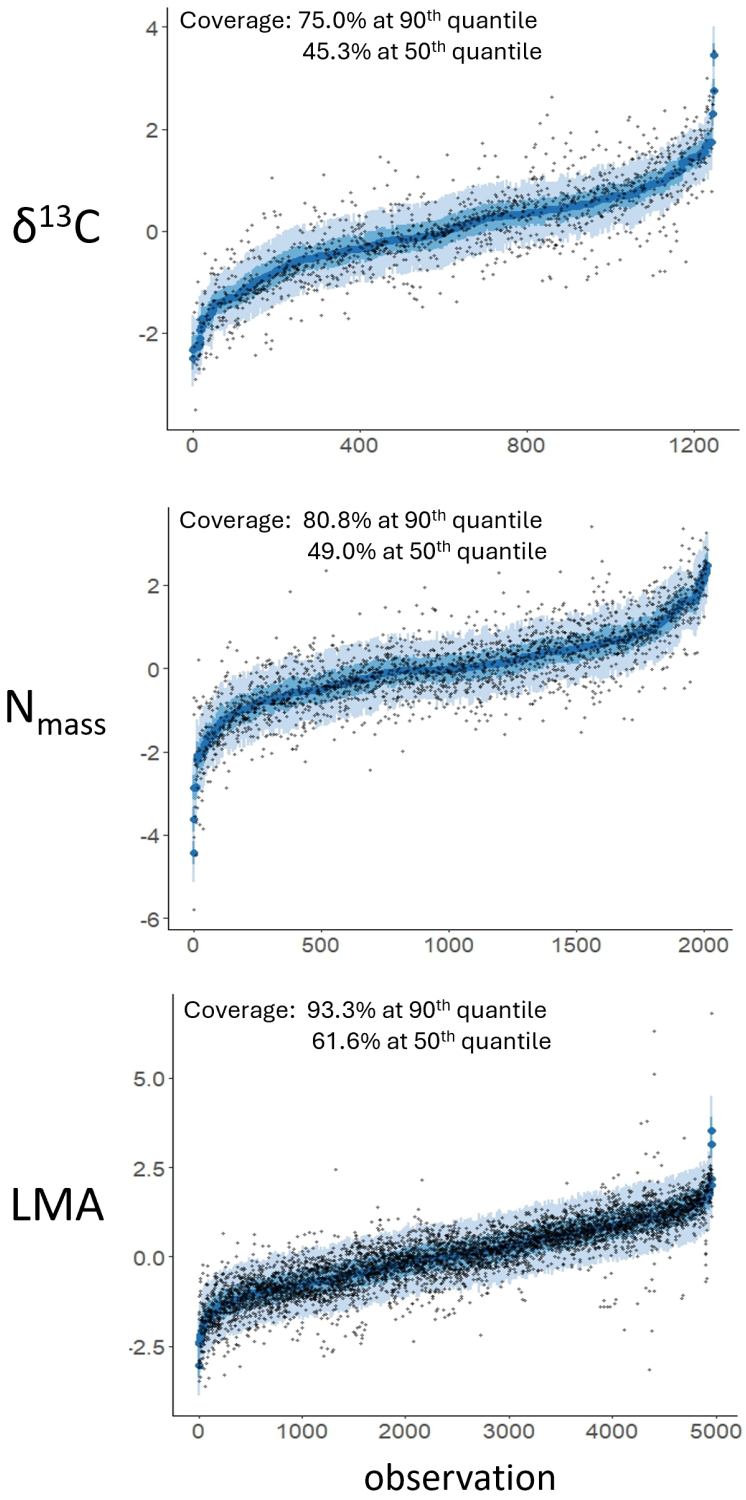


Figure 3: Posterior predictive checks from a fitted MR-PMM of leaf traits across 457 species of Eucalyptus. For each trait, black points represent observed values, while five-point summaries show the median (dark blue points), 0.5 CI (blue bars), and 0.95 CI (light blue bars) of the posterior predictive distribution for each observation. Predictions are made by conditioning on all random effects, with observations ordered by the predictive mean.

6.2 Results

Phylogenetic signal, and hence the tendency for similar values among closely related species (Figure 5), varied considerably between traits (Figure 4**A**, posterior median for $\hat{\lambda}$: LMA = 0.75, $\mathbf{N}_{\text{mass}} = 0.49$, $\delta^{13}\mathbf{C} = 0.27$).

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 (Figure 4**B**), 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 4**B**). This situation highlights an important strength of MR-PMM; the capacity to disentangle trait relationships operating at different levels in the model hierarchy. In line with simulations, consistent correlations across levels produce clear trends in inter-species data, while opposing correlations tend to obscure these relationships (Figure 5; Figure 2D; see tutorial).

Finally, while all non-phylogenetic trait relationships were retained as partial correlations, only the negative relationship between LMA and $\mathbf{N}_{\text{mass}}$ was retained at the phylogenetic level (Figure 4**C**). LMA and $\mathbf{N}_{\text{mass}}$ also showed the strongest relationships overall (Figure 4**B**), suggesting a deeply conserved functional integration between these traits.

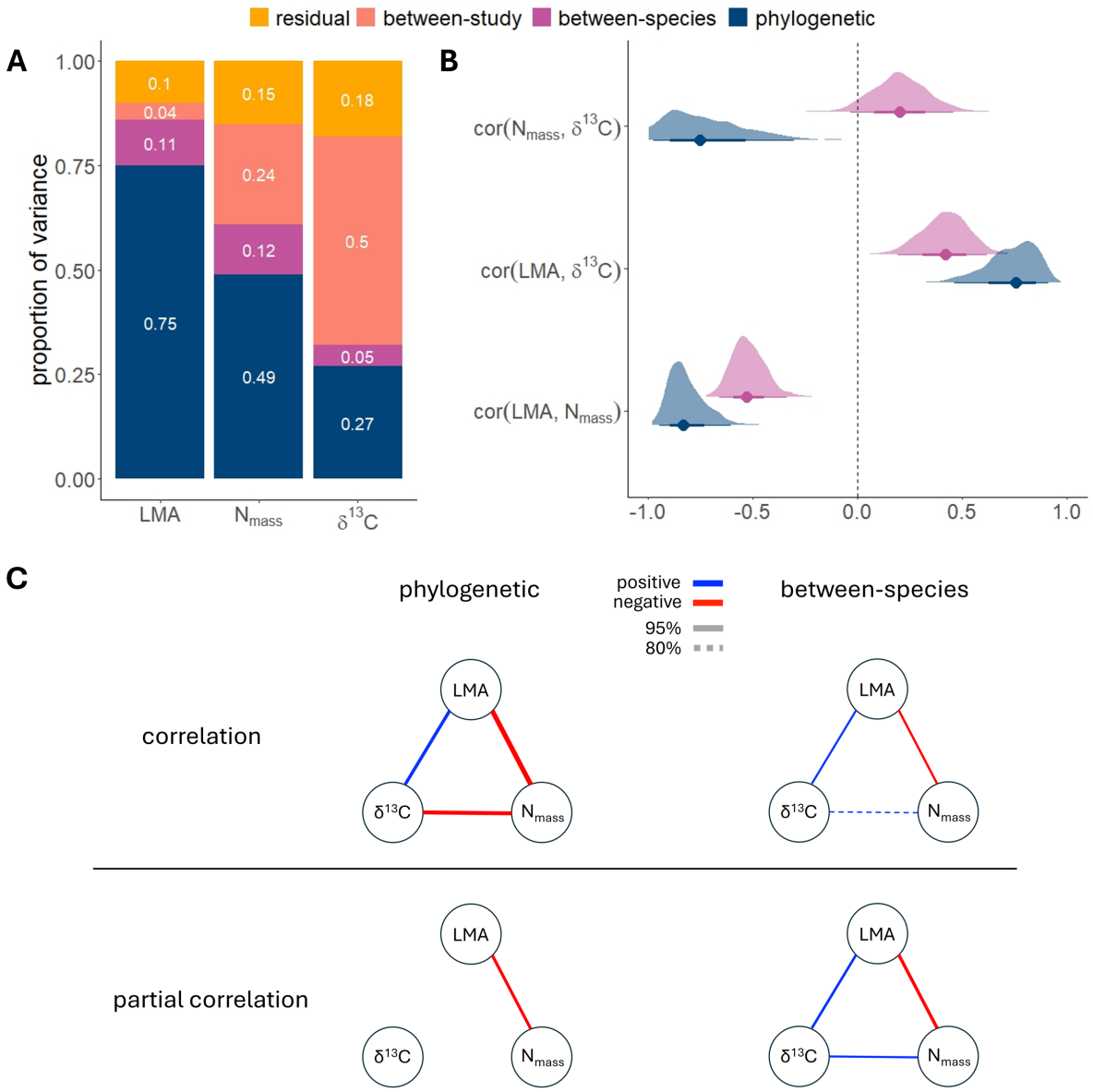


Figure 4: Results from an MR-PMM of leaf traits across 457 species of Eucalyptus. **A** variance decomposition reveals large differences between response traits in phylogenetic signal, as well as the relative contribution of different sources of variance; **B** phylogenetic and non-phylogenetic between-species correlation coefficient estimates. Points represent the posterior median for each estimate, with 50% and 95% CIs represented by heavy and light wicks respectively; **C** correlations and partial correlations between traits represented as network graphs. Edge widths are proportional to the posterior median of each coefficient estimate, with linetype indicating significance at the 95% (solid) and 80% (dashed) CI.

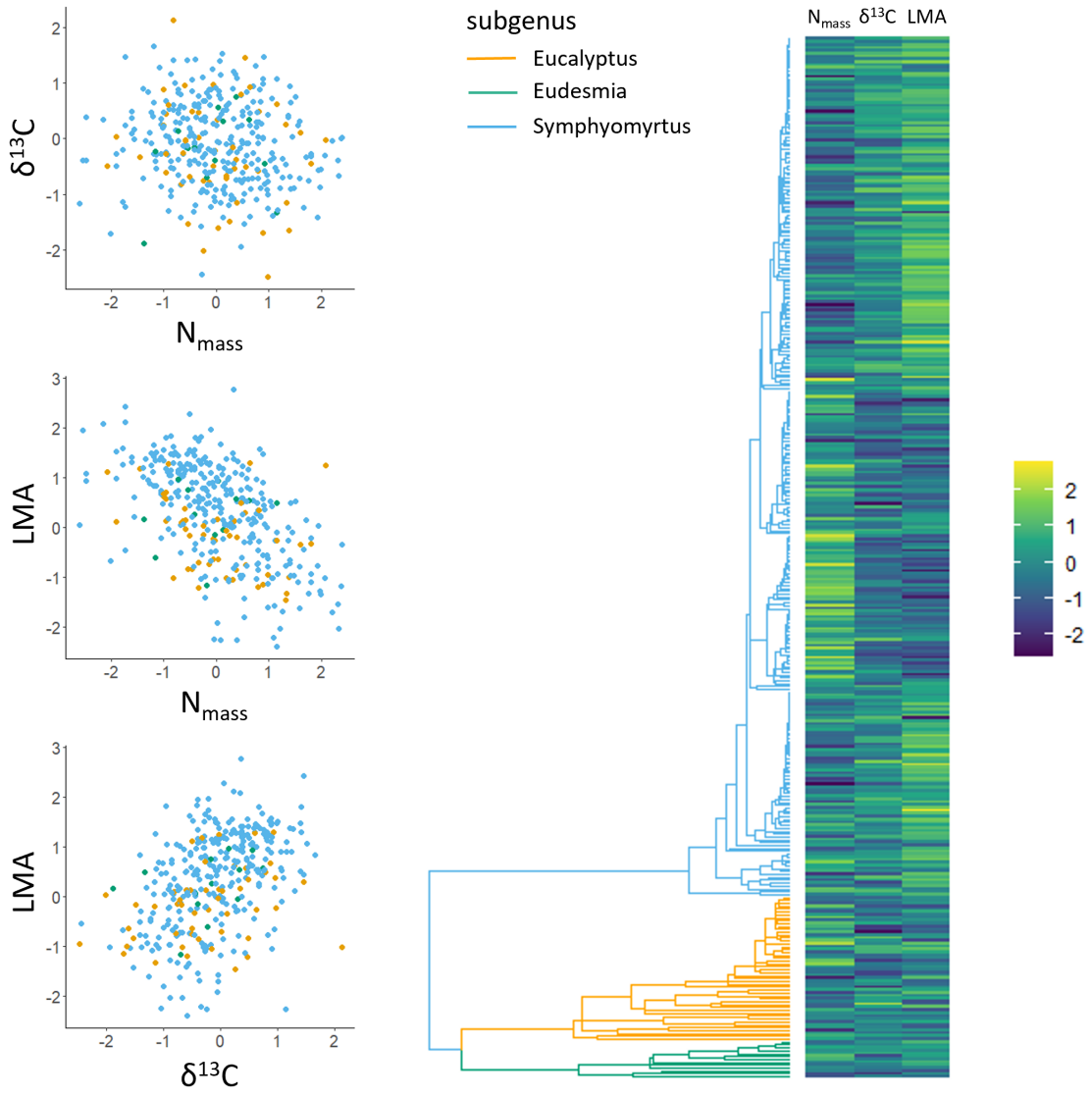


Figure 5: Scatter-plots (left) and heat-maps (right) of three leaf traits across 361 species of Eucalyptus (data filtered to complete cases). Trait values have been log-transformed and scaled. For heat maps (right), trait values are aligned with the corresponding species in the phylogeny (centre). For $\delta^{13}\mathbf{C}$ and $\mathbf{N}_{\text{mass}}$ (top left), opposing phylogenetic and non-phylogenetic correlations reported by the model (Figure 4) are obscured at the level of species phenotypes.

7 Extended Topics

The following section explores technical, yet important aspects of fitting MR-PMMs. These topics are particularly relevant in the case of high-dimensional and/or sparse trait data, where issues such as prolonged fitting time, model convergence difficulties, or parameter identifiability require the application of specialised computational methodologies.

7.1 Priors for multivariate normal distributions

As discussed in Section 3.0.7, trait covariance matrices and their inverses are often targets for inference; for example, to estimate conditional dependencies between traits or compute phylogenetic signal. However, it can be challenging to estimate these matrices as they generally have a large number of parameters; there are $\frac{1}{2}N(N+1)$ parameters in each of $\mathbf{\Sigma}^{\text{phy}}$ and $\mathbf{\Sigma}^{\text{ind}}$, where $N = N_{\text{trait}}$ denotes the number of traits throughout this section. In the MR-PMM framework, the covariance matrices $\mathbf{\Sigma}^{\text{phy}}$ and $\mathbf{\Sigma}^{\text{ind}}$ each parameterise a multivariate normal distribution which are priors for the vectors $\mathbf{u}_{\text{phy}}$ and $\mathbf{u}_{\text{ind}}$, respectively (see (15)). In a fully Bayesian setting, it is also necessary to specify priors for the covariance matrices themselves, for which there are many different options to choose from depending on the way the model is parameterized. The choice of prior impacts on the model-fitting strategy and efficiency, as well as the degree and type of shrinkage.

The classical choices of parameterisation for the multivariate normal are the covariance and precision forms. To each of these choices, there is a conjugate prior, the inverse Wishart and the Wishart distributions, respectively, characterized by a target matrix and degree of belief parameter. Conjugate priors are an algebraically convenient choice because they provide analytic solutions for certain steps of the model fitting which can greatly reduce estimation time in a Gibbs sampling approach (the covariance form with inverse Wishart prior is implemented in `MCMCglmm`). Departing from conjugacy, regularising priors such as the zero-centred Laplacian (two-sided exponential) and Gaussian distributions can be used to 'shrink' the off-diagonal elements of the covariance or precision matrix. In particular, the use of the Laplacian prior for the precision matrix, called the *graphical lasso*, is able to shrink matrix elements all the way to zero, generating a sparse network of conditional

dependencies. Further examples of sparsity-inducing priors include the adaptive lasso (Zou, 2006) and the horseshoe (Carvalho et al., 2009) which contain global terms to control the total amount of shrinkage and local terms to flexibly allocate shrinkage on a per-parameter basis. These types of sparsity-inducing priors have been used extensively for the general problem of variable selection, and hold enormous potential for inferring trait relationships in the MR-PMM setting, especially when data are sparse relative to the total number of candidate correlations.

A shortcoming of specifying priors for the covariance (or the precision) matrix is that the variances and correlations (or the precisions and partial correlations) are not treated independently. For example, we may wish to apply shrinkage to the correlation coefficients while maintaining an uninformative prior on the variances, or vice versa. To address this issue, the covariance matrix can be decomposed as $\mathbf{\Sigma} = \mathbf{SRS}^\mathsf{T}$ where $\mathbf{S} = \text{diag}(\sigma_1, \sigma_2, ..., \sigma_N)$ is the diagonal matrix of standard deviations and $\mathbf{R}$ is a correlation matrix (Barnard et al., 2000). This decomposition allows priors for the standard deviations to be specified separately from those of the correlation matrix. The matrix $\mathbf{R}$ can be further factored into the form $\mathbf{R} = \mathbf{LL}^\mathsf{T}$ where $\mathbf{L}$ is a lower-triangular matrix called the Cholesky factor. The latter parameterisation is implemented in `brms` where the default priors are half Student-t for the standard deviations and the Lewandowski-Kurowicka-Joe (LKJ, Lewandowski et al. (2009)) prior for $\mathbf{L}$. The LKJ prior has a single parameter η which for $\eta = 1$ specifies a uniform distribution in the space of correlation matrices and for $\eta > 1$ regularises estimates toward the identity where each correlation coefficient ρ_{ij} has marginal density $\text{Beta}(\eta + \frac{N-2}{2}, \eta + \frac{N-2}{2})$.

Use of the Cholesky factor is computationally advantageous for multivariate normal models, as it simplifies evaluation of the likelihood and allows an unconstrained parameterisation of $\mathbf{R}$ in terms of correlation angles via a certain geometric representation (Forrester and Zhang, 2020). The latter are efficient to sample and allow rapid evaluation of the LKJ prior while guaranteeing the validity of the corresponding correlation matrix.

7.2 Latent factor methods for dimension reduction

Although regularising priors can reduce the effective number of parameters for an estimated covariance matrix, for large numbers of traits it can be necessary to reduce the actual number

of parameters, however regularised they may be. A commonly used technique for dimension reduction in high-dimensional multivariate statistics is latent factor analysis (also called factor analytic), which is based on a rank-reduced representation of a covariance or correlation matrix. To explain the method, we first reparametrise the full MR-PMM (8) as

$$
\begin{aligned}
\mathbf{y} &= \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}(\mathbf{S}^{\text{phy}}\mathbf{L}^{\text{phy}} \otimes \mathbf{L}^{\mathbf{C}})\boldsymbol{\gamma} + (\mathbf{S}^{\text{res}}\mathbf{L}^{\text{res}} \otimes \mathbf{I}_{N_{\text{obs}}})\boldsymbol{\epsilon} \\
\boldsymbol{\gamma} &\sim N(0, \mathbf{I}_{N_{\text{traits}} \times N_{\text{species}}}) \\
\boldsymbol{\epsilon} &\sim N(0, \mathbf{I}_{N_{\text{traits}} \times N_{\text{obs}}}),
\end{aligned} \tag{24}
$$

where $\mathbf{S}^{\text{phy}}$ and $\mathbf{S}^{\text{res}}$ are diagonal matrices of standard deviations and $\mathbf{L}$ are lower triangular Cholesky factors satisfying $\mathbf{R}^{\text{phy}} = \mathbf{L}^{\text{phy}}(\mathbf{L}^{\text{phy}})^{\mathsf{T}}$, $\mathbf{R}^{\text{res}} = \mathbf{L}^{\text{res}}(\mathbf{L}^{\text{res}})^{\mathsf{T}}$, and $\mathbf{C} = \mathbf{L}^{\mathbf{C}}(\mathbf{L}^{\mathbf{C}})^{\mathsf{T}}$, with $\boldsymbol{\Sigma}^{\text{phy}} = \mathbf{S}^{\text{phy}}\mathbf{R}^{\text{phy}}(\mathbf{S}^{\text{phy}})^{\mathsf{T}}$ and $\boldsymbol{\Sigma}^{\text{res}} = \mathbf{S}^{\text{res}}\mathbf{R}^{\text{res}}(\mathbf{S}^{\text{res}})^{\mathsf{T}}$; see Section 7.1 for definitions of $\mathbf{S}$ and $\mathbf{R}$.

For the model (24), the rank of the correlation matrix $\mathbf{R}^{\text{phy}}$ (or $\mathbf{R}^{\text{res}}$) can be reduced from N_{traits} to k by replacing the $N_{\text{traits}} \times N_{\text{traits}}$ lower triangular square matrix $\mathbf{L}^{\text{phy}}$ with the rectangular $N_{\text{traits}} \times k$ lower triangular matrix $\mathbf{L}^{\text{phy}(k)}$ and $\boldsymbol{\gamma}$ with $\boldsymbol{\gamma}^{(k)} \sim N(0, \mathbf{I}_{k \times N_{\text{species}}})$, for a choice of integer $k < N_{\text{traits}}$. The elements of $\mathbf{L}^{\text{phy}(k)}$ and $\boldsymbol{\gamma}^{(k)}$ are called the *loadings* and *latent factors*, respectively.

It is important to note a few technical aspects of fitting latent variable models. Here we have chosen the k-column matrix of loadings to be a lower triangular matrix. It must also satisfy the constraints that the Euclidean row norms (i.e., the sum of squares of elements in each row) equals one and that the diagonal elements are positive valued. These constraints ensure that $\mathbf{L}^{(k)}(\mathbf{L}^{(k)})^{\text{T}}$ is a valid correlation matrix and that each matrix element is identifiable. These constraint choices are not unique and it is common to relax the lower-triangular restriction and require only that the columns be orthogonal. The latter allows all traits to contribute loadings to each factor which can aid direct interpretation of the factors. This choice, however, comes at the cost of sampling efficiency and the introduction of additional technical issues such as label switching (Hassler et al., 2022). A possible solution is to retain the lower triangular parameterisation for efficient model fitting and to compute the leading k eigenvectors of the estimated $\mathbf{L}^{(k)}(\mathbf{L}^{(k)})^{\text{T}}$ which are orthogonal and comprise contributions from all traits. As for the full rank case discussed in the previous section, regularising priors can be placed on factor loadings (Runcie and Mukherjee, 2013).

Latent factor models provide an effective means to reduce the complexity of high-dimensional multivariate models. In the MR-PMM setting, these models can be fit with the R package `asreml` (Butler et al., 2023) using residual maximum likelihood or using the Julia package `PhylogeneticFactorAnalysis`(Hassler et al., 2022) which acts as an interface for the Bayesian phylogenetic inference software `BEAST` (Suchard et al., 2018). Both packages offer parameterisation options and the latter will automatically select the number of factors k using cross validation. Neither `MCMCglmm` nor `brms`, the two packages on which we focus for tutorial, are able to fit latent factor models at this stage, although this class of models can be coded explicitly in the `stan` (Stan Development Team, 2024) statistical modelling language, which is the platform underpinning `brms`.

8 Discussion

8.1 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 phylogenetic correlation matrix. Different evolutionary models are fit via transformations of this matrix, or else by deriving expected pairwise phylogenetic correlations under the chosen model. 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 detailed and informative breakdown of trait covariances. We review these models, with a focus on statistical concepts, model specification, parameter interpretation, and practical implementation. We discuss recent advances in methodology and propose a synthesis integrating multilevel distributional models with 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.

8.2 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. 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). Most notably, BM does not explicitly account for adaptive evolution (e.g. stabilizing and directional selection). At first glance, this 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). Thus, even Brownian phylogenetic (co)variance is relevant for understanding divergence in ecological strategy between clades, a major contributor to the functional diversity of ecosystems.

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 likelihood algorithm in `PCMbase` (Mitov et al., 2020). Like MR-PMM, many of 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 (though see Hassler et al. 2022) and do not permit an arbitrary multilevel decomposition of trait covariances, which is a desirable and informative simplification for many biological questions.

Another important limitation of the MR-PMM specified in (9), is that it assumes the chosen evolutionary model is both constant through time and homogeneous across the tree, making it vulnerable to miss-specification under certain evolutionary scenarios, e.g., rare events (Uyeda et al., 2018; Clavel et al., 2015). Some alternative methods allow for shifts, meaning parameters of the evolutionary model (e.g. σ^2_{phy}), 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, 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).

One strength of the mixed model approach is their extension to generalised linear models that permit multivariate (Gaussian) random effects on the link-transformed mean of traits that are themselves not Gaussian distributed (Nakagawa and Schielzeth, 2010; Hadfield, 2010). Many evolutionary hypotheses invoke causal relationships between continuous and discrete traits. However, currently MR-PMM is one of few methods available to estimate correlations between continuous and discrete traits in a fully phylogenetic framework (see also Haba and Kutsukake 2019). Despite complications arising from non-linear latent variable transformations (see Section 3.0.5), this has enabled several long-standing 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 models that appropriately capture correlated hierarchical 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. 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).

8.3 Recommendations and Future directions

One challenge many analysts face when attempting a phylogenetic comparative analysis is that not all available species with trait data are featured in a suitable published phylogeny. A powerful workaround is to use the best available species-level phylogeny to derive a phylogeny for a higher taxonomic rank (e.g., series, section, or genus). This approach has two tangible benefits: 1) computations involving the phylogenetic correlation matrix $\mathbf{C}$ are a major bottleneck in the evaluation of the model likelihood. Therefore, reducing the dimension of $\mathbf{C}$ by refocusing to a higher taxonomic rank comes with substantial reductions in model fitting time; 2) It is often possible to include data from more species, as those missing from species-level phylogenies can instead be treated as replicates of a higher taxonomic rank that does feature in the phylogeny. This method has the obvious drawback of ignoring phylogenetic structure close to the tips (e.g., sister species relationships), modelling only those effects owing to deeper splits within the tree. However, in cases such as the Eucalypts, where taxonomic series often represent freely hybridising species groups (Pfeilsticker et al., 2023; Larcombe et al., 2015), modelling phylogenetic effects above the species level may in fact be desirable, as it relaxes the assumption that recent divergences can be expressed in terms of a strictly bifurcating evolutionary process.

A more general pitfall in phylogenetic comparative analyses is the failure to adequately account for multilevel structure. Assuming sufficient replication, the preferred approach is always to model species traits at the lowest possible level (e.g., observations on individual organisms), using a multilevel model. Partitioning variation across the model hierarchy utilises all available information, while facilitating evaluation of more interesting biological questions. For example, separation of within- and between-group variance components (e.g., species, population, individual) (Garamszegi and Møller, 2017; Westneat et al., 2015); separation of phylogenetic and spatial effects (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). Detailed variance partitioning may, in fact, be necessary to identify drivers of selection in natural populations. For example, partitioning variance in anti-predator behaviour of 254 bird species supported the view that within- and between-population variances are driven by different selective forces (Garamszegi and Møller, 2017). Similarly, within-species variances account for a consid-

erable proportion of community-level variation in plant functional traits, particularly for whole-plant (e.g. plant height), rather than organ-level (e.g. leaf size), traits (Siefert et al., 2015). As species differ not only in mean trait values but also in the magnitude and structure of trait variation, these studies highlight the enormous potential of multilevel distributional models to advance our understanding of the processes producing and maintaining biological variation.

Integrating multilevel distributional MR-PMM with sophisticated evolutionary models remains a considerable methodological challenge (see section 3.0.4). In particular, software implementations are limited and modern comparative analyses including many thousands of taxa pose genuine computational barriers 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. Of the available options, BM is easily the most tractable, a virtue so often repeated it risks being overlooked (Blomberg et al., 2020). Nonetheless, large-scale studies, particularly in the field of plant functional ecology, continue to be published without directly accounting for phylogeny (e.g., Joswig et al. 2022; 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). However, with the inexorable expansion of biological trait databases (Falster et al., 2021; Kattge et al., 2020), even models assuming BM can become computationally burdensome as the number of species and traits included in analyses grows. 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). A particularly powerful approach 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.

Another 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 response distributions, the relative magnitude of phylogenetic and non-phylogenetic (co)variances, as well as phylogenetic uncertainty and sample sizes. 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). Reduction of model complexity, for example via shrinkage (Section 7.1) or rank-reduction of the covariance matrix (Section 7.2), may be required for data sets with low $n:p$ ratios (e.g., geometric morphometrics), for which specialised methods are also available (Clavel and Morlon 2020; Adams and Collyer 2019, 2018; Collyer et al. 2015; Adams 2014; Hassler et al. 2022, also see Runcie et al. 2021 for methods relevant to genomic prediction). However, a formal examination of power and sensitivity, provided by simulation studies, is necessary to clarify the inherent limitations of estimation and inference in an MR-PMM framework. Fortunately, the ability to impute missing response values often means that more data can be included in analyses. Even relatively small data sets can potentially yield useful inferences by using sparsity-inducing priors (Section 7.1) to identify a small subset of parameters that escape shrinkage to zero.

8.4 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 inter-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 4).

3. Univariate approaches to cross-species analysis 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 4, 6).

4. The capacity to fit multilevel distributional models (3.0.1, 3.0.3), continuous and discrete

response variables (3.0.5), estimate partial correlations (3.0.7), impose different models of trait evolution (3.0.4), and employ conditional predictions for imputation and model validation (5), 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. We argue that careful parameterisation of MR-PMM (Sections 7.1, 7.2) may offer a scalable framework for analysing these large complex datasets.

6. Currently, software implementations of the various methods discussed are available piecemeal across several different applications. In practice, we found this led to different use cases for the two R packages we explored for fitting MR-PMM. The analytic solutions leveraged by a Gibbs sampling approach, give `MCMCglmm` a considerable computational advantage when all responses variables are Gaussian with the restriction that conjugate priors be used. In contrast, the Hamiltonian Monte Carlo sampler used by `brms` is likely to outperform for non-Gaussian error distributions and allows flexible prior specification. `brms` also allows for distributional models, benefits from a user-friendly syntax and integrates well with other useful packages such as `tidyverse` and `loo`.

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

9 Tutorial

Tutorial material relevant to this article can be found here.

10 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, and Jarrod Hadfield and Pierre de Villemereuil for detailed reviews that greatly 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.

Albert, J. H. and S. Chib (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association 88*(422), 669–679.

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

Barnard, J., R. McCulloch, and X.-L. Meng (2000). Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage. *Statistica Sinica 10*(4), 1281 – 1311.

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.

Bastide, P., L. S. T. Ho, G. Baele, P. Lemey, and M. A. Suchard (2021). Efficient Bayesian inference of general Gaussian models on large phylogenetic trees. *The Annals of Applied Statistics 15*(2), 971 – 997.

Björklund, M. (1997). Are 'comparative methods' always necessary? *Oikos 80*(3), 607–612.

Blackman, C. J., B. Halliwell, G. E. Hartill, and T. J. Brodribb (2024). Petiole xla (xylem to leaf area ratio) integrates hydraulic safety and efficiency across a diverse group of eucalypt leaves. *Plant, Cell & Environment 47*(1), 49–58.

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.

Bouckaert, R., T. G. Vaughan, J. Barido-Sottani, S. Duchêne, M. Fourment, A. Gavryushkina, J. Heled, G. Jones, D. Kühnert, N. De Maio, et al. (2019). Beast 2.5: An advanced software platform for bayesian evolutionary analysis. *PLoS computational biology 15*(4), e1006650.

Brodribb, T. J., M. Carriqui, S. Delzon, and C. Lucani (2017). Optical measurement of stem xylem vulnerability. *Plant Physiology 174*(4), 2054–2061.

Brodribb, T. J., J. Powers, H. Cochard, and B. Choat (2020). Hanging by a thread? forests and drought. *Science 368*(6488), 261–266.

Brodribb, T. J., R. P. Skelton, S. A. McAdam, D. Bienaimé, C. J. Lucani, and P. Marmottant (2016). Visual quantification of embolism reveals leaf vulnerability to hydraulic failure. *New Phytologist 209*(4), 1403–1409.

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. (2019). Bayesian item response modeling in r with brms and stan. *arXiv preprint arXiv:1905.09501*.

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 (2023). *ASReml-R Reference Manual Version 4.2*. Hemel Hempstead, HP2 4TP, UK: VSN International Ltd.

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

Carvalho, C. M., N. G. Polson, and J. G. Scott (2009, 16–18 Apr). Handling sparsity via the horseshoe. In D. van Dyk and M. Welling (Eds.), *Proceedings of the Twelth International Conference on Artificial Intelligence and Statistics*, Volume 5 of *Proceedings of Machine Learning Research*, Hilton Clearwater Beach Resort, Clearwater Beach, Florida USA, pp. 73–80. PMLR.

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.

Choat, B., S. Jansen, T. J. Brodribb, H. Cochard, S. Delzon, R. Bhaskar, S. J. Bucci, T. S. Feild, S. M. Gleason, U. G. Hacke, et al. (2012). Global convergence in the vulnerability of forests to drought. *Nature 491*(7426), 752–755.

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.

Clark, J. S., D. Nemergut, B. Seyednasrollah, P. J. Turner, and S. Zhang (2017). Generalized joint attribute modeling for biodiversity analysis: Median-zero, multivariate, multifarious data. *Ecological Monographs 87*(1), 34–56.

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.

Cleasby, I. R. and S. Nakagawa (2011). Neglected biological patterns in the residuals: a behavioural ecologist's guide to co-operating with heteroscedasticity. *Behavioral Ecology and Sociobiology 65*, 2361–2372.

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., M. B. Morrissey, S. Nakagawa, and H. Schielzeth (2018). Fixed-effect variance and the estimation of repeatabilities and heritabilities: Issues and solutions. *Journal of evolutionary biology 31*(4), 621–632.

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.

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.

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.

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.

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.

Forrester, P. J. and J. Zhang (2020). Parametrising correlation matrices. *Journal of Multivariate Analysis 178*, 104619.

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.

Halliwell, B., E. A. O'Connor, T. Uller, S. Meiri, B. Holland, C. K. Cornwallis, and G. While (2024). Reproductive and ecological adaptations to climate underpin the evolution of sociality in lizards. *bioRxiv*, 2024–05.

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.

Hassler, G. W., B. Gallone, L. Aristide, W. L. Allen, M. R. Tolkoff, A. J. Holbrook, G. Baele, P. Lemey, and M. A. Suchard (2022). Principled, practical, flexible, fast: A new approach to phylogenetic factor analysis. *Methods in Ecology and Evolution 13*(10), 2181–2197.

Hassler, G. W., A. F. Magee, Z. Zhang, G. Baele, P. Lemey, X. Ji, M. Fourment, and M. A. Suchard (2023). Data integration in bayesian phylogenetics. *Annual review of statistics and its application 10*(1), 353–377.

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.

Horvath, G., L. Z. Garamszegi, and G. Herczeg (2023). Phylogenetic meta-analysis reveals system-specific behavioural type–behavioural predictability correlations. *Royal Society Open Science 10*(9), 230303.

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.

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.

Joswig, J. S., C. Wirth, M. C. Schuman, J. Kattge, B. Reu, I. J. Wright, S. D. Sippel, N. Rüger, R. Richter, M. E. Schaepman, et al. (2022). Climatic and soil factors explain the two-dimensional spectrum of global plant trait variation. *Nature Ecology & Evolution 6*(1), 36–50.

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.

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.

Larcombe, M. J., B. Holland, D. A. Steane, R. C. Jones, D. Nicolle, R. E. Vaillancourt, and B. M. Potts (2015). Patterns of reproductive isolation in eucalyptus—a phylogenetic perspective. *Molecular Biology and Evolution 32*(7), 1833–1846.

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.

Lens, F., S. M. Gleason, G. Bortolami, C. Brodersen, S. Delzon, and S. Jansen (2022). Functional xylem characteristics associated with drought-induced embolism in angiosperms. *New Phytologist 236*(6), 2019–2036.

Lewandowski, D., D. Kurowicka, and H. Joe (2009). Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis 100*(9), 1989–2001.

Liu, H., Q. Ye, M. R. Lundgren, S. N. Young, X. Liu, Q. Luo, Y. Lin, N. Ye, and G. Hao (2024). Phylogeny and climate explain contrasting hydraulic traits in different life forms of 150 woody fabaceae species. *Journal of Ecology 112*(4), 741–754.

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.

Magwene, P. M. (2001). New tools for studying integration and modularity. *Evolution 55*(9), 1734–1745.

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.

McNeish, D. (2021). Specifying location-scale models for heterogeneous variances as multilevel sems. *Organizational Research Methods 24*(3), 630–653.

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.

Mitchell, D. J., C. Beckmann, and P. A. Biro (2021). Understanding the unexplained: The magnitude and correlates of individual differences in residual variance. *Ecology and Evolution 11*(12), 7201–7210.

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. and P. de Villemereuil (2019). A general method for simultaneously accounting for phylogenetic and species sampling uncertainty via rubin's rules in comparative analysis. *Systematic Biology 68*(4), 632–641.

Nakagawa, S. and R. P. Freckleton (2008). Missing inaction: the dangers of ignoring missing data. *Trends in ecology & evolution 23*(11), 592–596.

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.

O'Dea, R. E., D. W. Noble, and S. Nakagawa (2022). Unifying individual differences in personality, predictability and plasticity: a practical guide. *Methods in Ecology and Evolution 13*(2), 278–293.

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

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.

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.

Pfeilsticker, T. R., R. C. Jones, D. A. Steane, R. E. Vaillancourt, and B. M. Potts (2023). Molecular insights into the dynamics of species invasion by hybridisation in tasmanian eucalypts. *Molecular Ecology 32*(11), 2913–2929.

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, S. Burke, M. Lagisz, L. E. Schwanz, S. M. Drobniak, and S. Nakagawa (2024). New horizons for comparative studies and meta-analyses. *Trends in Ecology & Evolution*.

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.

Rausher, M. D. (1992). The measurement of selection on quantitative traits: biases due to environmental covariances between traits and fitness. *Evolution 46*(3), 616–626.

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. and L. J. Harmon (2022). *Phylogenetic comparative methods in R*. Princeton University Press.

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. and S. Mukherjee (2013, 07). Dissecting High-Dimensional Phenotypes with Bayesian Sparse Factor Analysis of Genetic Covariance Matrices. *Genetics 194*(3), 753–767.

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.

Shojaie, A. and E. B. Fox (2022). Granger causality: A review and recent advances. *Annual Review of Statistics and Its Application 9*(1), 289–319.

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.

Skelton, R. P., L. D. Anderegg, J. Diaz, M. M. Kling, P. Papper, L. J. Lamarque, S. Delzon, T. E. Dawson, and D. D. Ackerly (2021). Evolutionary relationships between drought-related traits and climate shape large hydraulic safety margins in western north american oaks. *Proceedings of the National Academy of Sciences 118*(10), e2008987118.

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 evolution 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.

Stan Development Team (2024). Stan Modeling Language Users Guide and Reference Manual. Version 2.32.6.

Suchard, M. A., P. Lemey, G. Baele, D. L. Ayres, A. J. Drummond, and A. Rambaut (2018, 06). Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evolution 4*(1), vey016.

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.

Thorson, J. T. and W. van der Bijl (2023). phylosem: A fast and simple r package for phylogenetic inference and trait imputation using phylogenetic structural equation models. *Journal of Evolutionary Biology 36*(10), 1357–1364.

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.

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.

Villemereuil, P. d., J. A. Wells, R. D. Edwards, and S. P. Blomberg (2012). Bayesian models for comparative analysis integrating phylogenetic uncertainty. *BMC Evolutionary Biology 12*(1), 1–16.

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.

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.

Westneat, D. F., J. Wright, and N. J. Dingemanse (2015). The biology hidden inside residual within-individual phenotypic variation. *Biological Reviews 90*(3), 729–743.

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.

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.

Zhang, Z., A. Nishimura, N. S. Trovão, J. L. Cherry, A. J. Holbrook, X. Ji, P. Lemey, and M. A. Suchard (2023, August). Accelerating bayesian inference of dependency between mixed-type biological traits. *PLOS Computational Biology 19*(8), e1011419.

Zhou, J., E. Cieraad, and P. M. van Bodegom (2022). Global analysis of trait–trait relationships within and between species. *New Phytologist 233*(4), 1643–1656.

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association 101*(476), 1418–1429.