# Promoting the use of phylogenetic multinomial generalised mixed-effects model to understand the evolution of discrete traits

**Ayumi Mizuno[1,2], Szymon M Drobniak[2,3], Coralie Williams[2,4], Malgorzata Lagisz[1,2], and Shinichi Nakagawa[1,2]**

[1]**Department of Biological Sciences, Faculty of Science, The University of Alberta, Edmonton, Canada**
[2]**Evolution and Ecology Research Centre, School of Biological, Earth and Environmental Sciences, The University of New South Wales, Sydney, New South Wales, Australia**
[3]**Institute of Environmental Sciences, Faculty of Biology, Jagiellonian University, Kraków, Poland**
[4]**School of Mathematics and Statistics, The University of New South Wales, Sydney, Australia**

## FUNDING INFORMATION

SN, CW, and ML were supported by the Australian Research Council (ARC) Discovery Project Grants (DP210100812 and DP230101248). Also, AM and SN were supported by the Canada Excellence Research Chair (CERC-2022-00074). CW was supported by the Statistical Society of Australia top-up scholarship. SMD was supported by the Polish National Science Centre (NCN) grant (UMO-2020/39/B/NZ8/01274) and by NCN & CHIST-ERA project FAIRBiRDS (2022/04/Y/NZ8/00184).

# ABSTRACT

Phylogenetic comparative methods (PCMs) are fundamental tools for understanding trait evolution across species. While linear models are widely used for continuous traits in ecology and evolution, their application to discrete traits - particularly ordinal and nominal traits - remains limited. Researchers sometimes recategorise such traits into binary traits (0 or 1 data) to make them more manageable. However, this risks distorting the original data structure and meaning, potentially reducing the information it initially contained. This paper promotes the use of phylogenetic generalised linear mixed-effects models (PGLMMs) as a flexible framework for analysing the evolution of discrete traits. We introduce the theoretical foundations of PGLMMs and demonstrate how univariate and multivariate versions of binary PGLMMs, which might be more familiar to evolutionary biologists, can be conceptually extended to model ordinal and nominal traits. Specifically, we describe ordered and unordered multinomial PGLMMs for ordinal and nominal traits, respectively. We then explain how to interpret regression coefficients and (co)variance components, including associated statistics (e.g., phylogenetic heritability and correlation) from PGLMMs for discrete traits. Using real-world examples from avian datasets, we illustrate the practical implementation of PGLMMs to reveal evolutionary patterns in discrete traits. We also provide online tutorials to guide researchers through the application of these models using Bayesian implementations in R. By making complex models more accessible, we aim to facilitate a more precise and insightful understanding of the evolution and function of discrete traits, which has received relatively limited attention in evolutionary biology so far.

# 1 INTRODUCTION

As children, while flipping through the pages of visual dictionaries, we may have wondered why Earth is home to such a vast diversity of living things. Today, it is estimated that at least 8.7 million species inhabit our planet, with new species continually being discovered (Mora et al., 2011; Scheffers et al., 2012). These extant species flourish in a wide range of environments and exhibit species-specific traits. They have evolved unique traits while adapting to their surrounding environment. Even limited to the avian class, species live in areas ranging from polar regions to deserts and also urban environments and have evolved specific characteristics, such as body shapes, foraging methods, and reproductive systems. Phylogenetic inertia or constraints, rooted in evolutionary history, as well as shared genetic background, can lead closely related species to show similar traits. At the same time, distantly related species may have convergent traits adapted to similar environments. Understanding the factors that influence the evolution of these traits and speciation is critically important in evolutionary biology and ecology - topics that have long fascinated researchers in these fields.

However, until the 1980s, when Felsenstein (1985) published the first paper on phylogenetic comparative methods (PCMs), the evolution of organisms could not be tested using statistically formalised approaches. PCMs have rapidly advanced in recent years (Lynch, 1991; Pagel, 1997; Martins and Hansen, 1997; Housworth et al., 2004; Ives et al., 2007; Hadfield, 2010; Hadfield and Nakagawa, 2010; Felsenstein, 2012; O'Meara, 2012; Garamszegi, 2014; Cornwell and Nakagawa, 2017; Harmon, 2018; Cornwallis and Griffin, 2024), offering powerful frameworks for revealing patterns and processes of organism evolution while accounting for phylogenetic non-independence among species (Felsenstein, 1985; Grafen, 1989; Martins and Hansen, 1997). By incorporating phylogenetic information, PCMs can estimate common **ancestral states**, speciation rates, and evolutionary relationships between various types of traits and related factors. Such traits encompass behavioural, ecological, morphological, and physiological characteristics and life histories, along with ecological and environmental factors, which are often evolutionary drivers of these traits (e.g., Felsenstein, 1985).

Two major classes of models dominate phylogenetic comparative methods (PCMs) for trait evolution: **linear models** and **Markov models**. Linear models, such as phylogenetic generalized least squares (PGLS) (Martins and Hansen, 1997; Pagel, 1997; Garland and Ives, 2000; Rohlf, 2001), are commonly used to analyse continuous traits (e.g., body size, lifespan). These models offer considerable flexibility by allowing researchers to examine relationships between one or more response variables and explanatory variables, whether continuous or discrete (e.g., environmental factors, species-specific traits). By contrast, Markov models (e.g., the Mk model) (Pagel, 1994; Lewis, 2001) are predominantly used to model the evolution of discrete (categorical) traits (here defined to include binary, ordinal [ordered multinomial], and nominal [unordered multinomial] data, but excluding count [Poisson] data; see figure 1) and estimate the probabilities of transitioning between different trait states over evolutionary time.

Yet, discrete response variables need not be confined to Markov models; linear models can also handle categorical outcomes (e.g., Grizzle et al., 1969; Anderson, 1984; Sloane and Morgan, 1996). Binary traits, being a special case of discrete traits, have been frequently analysed via linear models and are widely familiar to evolutionary biologists. Nonetheless, the use of linear models for ordinal and nominal traits remains uncommon. This situation is likely due to challenges in both implementation and interpretation, stemming from the relative unfamiliarity of extending linear models to these types of data (i.e., ordinal and nominal traits). Despite these hurdles, one of the

greatest advantages of linear models is their unified framework, which can seamlessly accommodate both continuous and discrete traits. However, as far as we know, there is no accessible guide or tutorial to investigate ordinal and nominal traits under the linear model framework (cf. Hadfield and Nakagawa, 2010).

To fill this gap, we demonstrate how linear models, particularly phylogenetic generalised linear mixed models (PGLMMs; Lynch, 1991; Housworth et al., 2004, Hadfield and Nakagawa, 2010), can be applied to study the evolution of discrete traits (see figure 1). PGLMMs are also well-suited for modelling data with varying levels of statistical non-independence, such as multiple observations per species by using random effects. Additionally, PGLMMs can accommodate multiple response variables (i.e., multiple traits), which can shed light on, for example, coevolution among traits. Importantly, Bayesian packages such as the `brms` package (Bürkner, 2017) and the `MCMCglmm` package (Hadfield, 2010; Hadfield and Nakagawa, 2010) make these models accessible for diverse research applications. Here, our focus is on ordinal and nominal traits, which have posed challenges due to the complexities in implementation and interpretation. We introduce the theoretical foundation of PGLMMs and then provide practical applications using real-world examples. Finally, we offer online tutorials (link) that guide users through implementing these models with Bayesian approaches in `R`. We aim to equip evolutionary biologists with a robust framework for analysing discrete traits. By combining theoretical foundations with practical applications, we will make complex models more accessible, thus supporting a more precise and insightful understanding of trait evolution studies.

# 2 THEORY

To build upon what evolutionary biologists are likely to be familiar with, we first describe PGLMMs for continuous traits (using the Gaussian link function; Hadfield, 2010; Ives and Helmus,2011; Pottier et al., 2024). By doing so, we introduce the concept of **phylogenetic signal** and **heritability** (note that these are not always interchangeable - for example, while Pagel's $\lambda$ can be interpreted as a measure of phylogenetic heritability, other metrics, such as Blomberg's $K$ (Blomberg et al., 2003), quantify different aspects of phylogenetic signal and do not directly reflect heritability). We also present the essential idea of **univariate and multivariate (multi-response) models**. The introduction of multivariate models is important since models for nominal traits inherently function as multivariate models. Then, we extend PGLMMs to discrete traits, including binary, ordinal, and nominal traits (using the probit and logit link functions). Here, we show how the PGLMM for binary traits, which might be more familiar to evolutionary biologists, provides a basis for the GLMMs for ordinal and nominal traits, which are less familiar to evolutionary biologists (figure 1). Finally, we also touch upon how to account for within-species variation in PGLMMs (i.e., replicates within species) as an area of future development (Cornwallis and Griffin, 2024).

## 2.1 PGLMM for continuous traits

The phylogenetic (linear) mixed-effects model (PMM) was first introduced by (Lynch, 1991) (note that a PMM is a PGLMM with the Gaussian link function). Despite its versatility, PMMs are less frequently used than popular PGLS (phylogenetic generalised least square) models (Martins and Hansen, 1997; Pagel, 1997; Garland and Ives, 2000; Rohlf, 2001). The PMM is a type of linear mixed-effects model that accounts for the non-independence of species due to their shared evolutionary history (i.e., phylogeny; Felsenstein, 1985; Grafen, 1989; Martins and Hansen, 1997). The simplest (univariate) PMM can be written as:

$$y_i = \beta_0 + a_i + e_i, \tag{1}$$

$$a_i = \mathbf{a} \sim \mathcal{N}(\mathbf{0}, \sigma_a^2 \mathbf{A}), \tag{2}$$

$$e_i = \mathbf{e} \sim \mathcal{N}(\mathbf{0}, \sigma_e^2 \mathbf{I}), \tag{3}$$

where $y_i$ is the trait value of the $i$th species, $\beta_0$ is the intercept (across-species overall mean), $a_i$ is the random effect of the $i$th species, $e_i$ is the residual error, $\mathbf{a}$ and $\mathbf{e}$ are vectors of random effects and residuals, respectively, the random effect $a_i$ is assumed to be normally distributed with a mean of zero and a variance-covariance of $\sigma_a^2\mathbf{A}$ (where **phylogenetic correlation matrix A** is a square matrix derived from an ultra-metric phylogenetic tree where the distance from the root to every tip is the same, assuming a constant rate of evolution (Gavryushkin and Drummond, 2016)), and the residual error $e_i$ is assumed to be normally distributed with a mean of zero and a variance-covariance of $\sigma_e^2\mathbf{I}$ (the identity matrix $\mathbf{I}$ is a square matrix with ones on the diagonal and zeros elsewhere). The variance component $\sigma_a^2$ shows how much of the variation is caused by phylogenetic signals (Pagel, 1999; Freckleton et al., 2002). In contrast, $\sigma_e^2$ represents the variation that is not related to phylogeny. This could include factors like species-specific ecological differences (e.g., adaptations to their environment; $\sigma_e^2$ may also include measurement errors or variation within studies, but here, we assume it mainly reflects non-phylogenetic variation at the species level). Note that the construction of $\mathbf{A}$ can be adjusted according to the mode of evolution one assumes; usually, a linear decline of phylogenetic relationship (dependence) from the common ancestors (ancestral node) and extant species (tips) is assumed, and this is known as the Brownian model of evolution (Felsenstein, 1973, 1985).

Importantly, phylogenetic heritability ($H^2$) can be calculated as the proportion of the total variance that is due to the variance associated with phylogeny:

$$H^2 = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_e^2} \tag{4}$$

The PMM can be extended to include fixed effects (e.g., other species-specific traits and environmental variables).

$$y_i = \beta_0 + \beta_1 x_{1i} + ... + \beta_h x_{hi} + a_i + e_i \tag{5}$$

where $x_{1i}, x_{2i}, ...x_{hi}$ are the fixed effects (i.e., $x_{hi}$ is $h$th fixed effect of the $i$th species).

Rather than fitting species traits as another fixed effect, we can fit it as another response using a multivariate PMM. A bivariate PMM can be written as (see Housworth et al., 2004; Adams and Collyer, 2024):

$$\mathbf{y_i} = \begin{pmatrix} y_i^{(1)} = \beta_0^{(1)} + a_i^{(1)} + e_i^{(1)} \\ y_i^{(2)} = \beta_0^{(2)} + a_i^{(2)} + e_i^{(2)} \end{pmatrix}, \tag{6}$$

where $\mathbf{y_i}$ is a vector consisting of $y_i^{(1)}$ and $y_i^{(2)}$ which are two different traits (e.g., the body mass and wing length) of the $i$th species, respectively, $\beta_0^{(1)}$ and $\beta_0^{(2)}$ are the intercepts (across-species overall means), $a_i^{(1)}$ and $a_i^{(2)}$ are the phylogenetic effects of the $i$th species, and $e_i^{(1)}$ and $e_i^{(2)}$ are the residuals (assuming to be mostly non-phylogenetic effects).

These phylogenetic effects are distributed as follows:

$$\begin{pmatrix} a_i^{(1)} \\ a_i^{(2)} \end{pmatrix} = \begin{pmatrix} \mathbf{a}^{(1)} \\ \mathbf{a}^{(2)} \end{pmatrix} \sim \mathcal{N}\left( \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \Sigma_a \otimes \mathbf{A} \right), \tag{7}$$

$$\Sigma_a \otimes \mathbf{A} = \begin{pmatrix} \sigma_{a1}^2 \mathbf{A} & \rho_{a12}\sigma_{a1}\sigma_{a2}\mathbf{A} \\ \rho_{a12}\sigma_{a1}\sigma_{a2}\mathbf{A} & \sigma_{a2}^2 \mathbf{A} \end{pmatrix}, \tag{8}$$

where $\Sigma_a \otimes \mathbf{A}$ is the **variance-covariance matrix** of the phylogenetic effects (the symbol $\otimes$ denotes the Kronecker product), $\mathbf{A}$ is the phylogenetic correlation matrix, $\sigma_{a1}^2$ and $\sigma_{a2}^2$ are the variances of the phylogenetic effects of the first and second traits, respectively, and $\rho_{a12}$ is the correlation between the phylogenetic effects of the two traits. A positive $\rho_{a12}$ indicates that the traits have tended to evolve in the same direction due to shared evolutionary history (i.e., correlated phylogenetic effects). In contrast, a negative $\rho_{a12}$ indicates that the traits have evolved in opposite directions phylogenetically. It is important to mention that $\rho_{a12}$ refers specifically to phylogenetic correlations. We may find that overall correlations between traits differ if strong non-phylogenetic or residual effects obscure these phylogenetic patterns.

Similarly, the residuals (non-phylogenetic effects) are distributed as follows:

$$\begin{pmatrix} e_i^{(1)} \\ e_i^{(2)} \end{pmatrix} = \begin{pmatrix} \mathbf{e}^{(1)} \\ \mathbf{e}^{(2)} \end{pmatrix} \sim \mathcal{N}\left( \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \Sigma_e \otimes \mathbf{I} \right), \tag{9}$$

$$\Sigma_e \otimes \mathbf{I} = \begin{pmatrix} \sigma_{e1}^2 \mathbf{I} & \rho_{e12}\sigma_{e1}\sigma_{e2}\mathbf{I} \\ \rho_{e12}\sigma_{e1}\sigma_{e2}\mathbf{I} & \sigma_{e2}^2 \mathbf{I} \end{pmatrix}, \tag{10}$$

where $\Sigma_e \otimes \mathbf{I}$ is the variance-covariance matrix of the non-phylogenetic effects (residuals), $\sigma_{e1}^2$ and $\sigma_{e2}^2$ are the variances of the residuals of the first and second traits, respectively, and $\rho_{e12}$ is the correlation between the non-phylogenetic effects of the two traits. Such a correlation can represent convergence or divergence due to ecological factors. For example, shared environmental pressures, such as temperature and precipitation, can drive traits toward similar adaptations, while predation or resource competition may lead to trait differentiation. Using 4, phylogenetic heritability ($H^2$) can be obtained separately for the first and second traits.

Also, as mentioned above, this bivariate model can be extended by including fixed effects (e.g., other species traits and environmental variables).

$$\mathbf{y_i} = \begin{pmatrix} y_i^{(1)} = \beta_0^{(1)} + \beta_1^{(1)} x_{1i} + ... + \beta_h^{(1)} x_{hi} + a_i^{(1)} + e_i^{(1)} \\ y_i^{(2)} = \beta_0^{(2)} + \beta_1^{(2)} x_{1i} + ... + \beta_h^{(2)} x_{hi} + a_i^{(2)} + e_i^{(2)} \end{pmatrix} \tag{11}$$

## 2.2 PGLMM for discrete traits 1: binary traits

The simplest phylogenetic generalised linear mixed-effects model (PGLMM) for binary traits can be written using either the **logit or probit link functions** (figure 2):

$$y_i \sim \text{Bernoulli}(p_i) = \text{Binomial}(1, p_i) \tag{12}$$

$$P(y_i = 1) = p_i = \exp(l_i)/\left(1 + \exp(l_i)\right) = \text{logit}^{-1}(l_i) \tag{13}$$

$$l_i = \ln\left(\frac{p_i}{1 - p_i}\right) \tag{14}$$

$$P(y_i = 1) = p_i = \Phi^{-1}(l_i) = \text{probit}^{-1}(l_i) \tag{15}$$

$$l_i = \Phi(p_i) \tag{16}$$

$$l_i = \beta_0 + a_i \tag{17}$$

where $\mathbf{y_i}$ represents the binary trait value of the $i$th species, indicating whether a specific trait is present (e.g., maternal care). The probability of observing the trait, denoted as $p_i$, follows a Bernoulli distribution, which is a special case of the binomial distribution with a size parameter of 1. The probability $p_i$ is transformed onto a latent scale using either a logit or probit link function. In the logit case, the transformation is defined as (13), where $\mathbf{l_i}$ is the **latent linear predictor**. Alternatively, in the probit case, the transformation uses the inverse **cumulative distribution function** $\Phi^{-1}$ of the standard normal distribution (14). The $\mathbf{l_i}$ is related to the probability $p_i$ and can be expressed differently depending on the link function used. In the logit case, $\mathbf{l_i}$ represents the log-odds ratio (14). In the probit case, $\mathbf{l_i}$ corresponds directly to the transformed probability, which is the cumulative standard normal probability for quantile $\mathbf{p_i}$ (15). The $\mathbf{l_i}$ is modelled as a linear combination of the intercept $\beta_0$ (the overall mean across species on the latent scale) and $a_i$, which represents the phylogenetic effect of the $i$th species (17). The inverse transformations, $\text{logit}^{-1}$ (also known as the inv-logit transformation) and $\Phi^{-1}$, are used to map the latent scale back to the probability scale. Note that this model lacks the residuals $\mathbf{e_i}$; this is because a Bernoulli distribution does not have any overdispersion, or errors are determined by an underlying distribution (Hadfield, 2010). Using this property, phylogenetic heritability ($\mathbf{H^2}$) can be calculated for binary traits as follows for the logit and probit link functions, respectively (Hadfield, 2010; Nakagawa and Schielzeth, 2010, 2013):

$$H^2 = \frac{\sigma_a^2}{\sigma_a^2 + \pi^2/3} \tag{18}$$

$$H^2 = \frac{\sigma_a^2}{\sigma_a^2 + 1} \tag{19}$$

where $\pi^2/3$ is the variance for the logistic distribution while 1 is the variance for the standard normal distribution. As mentioned above, this model can be extended by including fixed effects:

$$l_i = \beta_0 + \beta_1 x_{1i} + ... + \beta_h x_{hi} + a_i. \tag{20}$$

Also, this model can be extended to a bivariate model with fixed effects:

$$\mathbf{l_i} = \begin{pmatrix} l_i^{(1)} = \beta_0^{(1)} + \beta_1^{(1)} x_{1i} + ... + \beta_k^{(1)} x_{hi} + a_i^{(1)} \\ l_i^{(2)} = \beta_0^{(2)} + \beta_1^{(2)} x_{1i} + ... + \beta_k^{(2)} x_{hi} + a_i^{(2)} \end{pmatrix}, \tag{21}$$

where $\mathbf{l_i}$ is a vector consisting of $l_i^{(1)}$ and $l_i^{(2)}$, which are the linear predictors of the first and second binary traits (e.g., the presence of maternal and paternal care), respectively, $\beta_0^{(1)}$ and $\beta_0^{(2)}$ are the intercepts (across-species overall means), $\beta_1^{(1)}$ and $\beta_1^{(2)}$ are the fixed effects of the first and second traits, respectively, and $a_i^{(1)}$ and $a_i^{(2)}$ are the phylogenetic effects of the first and second traits, respectively and the phylogenetic effects are distributed as with 8 and 9.

We note that binary traits can have a natural order (e.g., egg-laying vs live-born, solitary living vs social living) or can be two independent categories (e.g., sex or habitat). The point is that the PGLMM for binary traits can be considered as both the GLMM for ordinal traits (ordered multinomial traits) and the GLMM for nominal traits (unordered multinomial traits). Indeed, a binomial distribution is the simplest case of a multinomial distribution as well as a categorical distribution:

$$y_i \sim \text{Binomial}(1, p_i) \leftrightarrow \mathbf{y}_i' \sim \text{Multinomial}(1, (p_{1i}, p_{2i})) \leftrightarrow y_i^{(k)} \sim \text{Categorical}(p_{1i}, p_{2i}), \tag{22}$$

where $\mathbf{y}_i{}'$ is a multinomial distribution with two categories ($\mathbf{y}_i{}' = (1,0)$ and $\mathbf{y}_i{}' = (0,1)$ with the size of 1) $y_i^{(k)}$ is a categorical distribution with k categories/levels and $p_{1i}$ and $p_{2i}$ are the probabilities of the first and second categories, respectively (i.e., $p_{1i} = 1 - p_i$ and $p_{2i} = p_i$; $y_i^{(1)} = \text{A} \leftrightarrow \mathbf{y}_i{}' = (1,0)$ and $p_{2i} = p_i$ and $y_i^{(2)} = \text{B} \leftrightarrow \mathbf{y}_i{}' = (0,1)$; see also figure 2a). Each element of $\mathbf{y}_i{}'$ corresponds to the probability of one of the two possible outcomes in the multinomial distribution. The linear predictors $l_i$, therefore, can be re-written for the logit and probit link functions, respectively:

$$l_i = \ln\left(\frac{p_i}{1 - p_i}\right) = \ln\left(\frac{p_{2i}}{p_{1i}}\right), \tag{23}$$

$$l_i = \Phi(p_i) = \Phi(p_{2i}) = \Phi(1 - p_{1i}). \tag{24}$$

Also, for the logic and probit link functions, the probabilities (of a binary variable) can be re-written with $\mathbf{p_i}$ being a vector of probabilities of the two categories for $i$th species, denoted as A and B, can be expressed as follows:

$$\mathbf{p_i} = \begin{pmatrix} P(y_i^{(1)} = \mathrm{A}) = p_{1i} = 1/\left(1+\exp(l_i)\right) = \mathrm{logit}^{-1}(-l_i) \\ P(y_i^{(2)} = \mathrm{B}) = p_{2i} = \exp(l_i)/\left(1+\exp(l_i)\right) = \mathrm{logit}^{-1}(l_i) \end{pmatrix}, \tag{25}$$

$$\mathbf{p_i} = \begin{pmatrix} P(y_i^{(1)} = \mathrm{A}) = p_{1i} = \Phi(c_0 - l_i) = \Phi(-l_i) \\ P(y_i^{(2)} = \mathrm{B}) = p_{2i} = 1 - \Phi(c_0 - l_i) = \Phi(l_i) \end{pmatrix}, \tag{26}$$

where $c_0$ is a cut-point (here defined as $c_0 = 0$) used in ordinal models (which is explained in detail later).

In the next two sections, we describe the PGLMM for ordinal and nominal traits by building up on the PGLMM for binary traits. It is important to note that, traditionally, the logit link function is more commonly used for nominal variables, while the probit link function is more commonly used for ordinal variables (Hadfield, 2015).

### 2.3 PGLMM for discrete traits 2: ordinal traits

Ordinal variables (e.g., plumage colour gradation, migration level, mating system, and social hierarchy) can be modelled as so-called "**threshold models**" with the probit link function, where thresholds are usually referred to as cut-points (figure 2b). Importantly, in such a model, **latent values** of the ordinal traits are assumed to be continuous following the standard normal distribution. Here, we define an ordinal trait with three levels, whose threshold model (ordered multinomial or categorical PGLMM) can be defined as:

$$\mathbf{y}_i{}' \sim \mathrm{Mulitnomial}(1, (p_{1i}, p_{2i}, p_{3i})) \leftrightarrow y_i^{(k)} \sim \mathrm{Categorical}(p_{1i}, p_{2i}, p_{3i}). \tag{27}$$

where $\mathbf{y}_i{}'$ is a vector of probabilities for $i$th species to belong to one of three levels (e.g., A, B and C; $\mathbf{y}_i{}' = (1,0,0)$, $\mathbf{y}_i{}' = (0,1,0)$ and $\mathbf{y}_i{}' = (0,0,1)$), $p_{1i}$, $p_{2i}$, and $p_{3i}$ are the probabilities of the three levels, respectively. The probabilities of the three levels are calculated as the cumulative probabilities of the standard normal distribution (Hadfield, 2015). Using these probabilities, the linear predictor $l_i$ can be written as:

$$\mathbf{p_i} = \begin{pmatrix} P(y_i^{(1)} = \mathrm{A}) = p_{1i} = \Phi(c_0 - l_i) \\ P(y_i^{(2)} = \mathrm{B}) = p_{2i} = \Phi(c_1 - l_i) - \Phi(c_0 - l_i) \\ P(y_i^{(3)} = \mathrm{C}) = p_{3i} = 1 - \Phi(c_1 - l_i) \end{pmatrix} \tag{28}$$

$$l_i = \beta_0 + a_i + e_i \tag{29}$$

$$\mathbf{e} \sim \mathcal{N}(0, \mathbf{I}) \tag{30}$$

where $\mathbf{p}_1{}'$ is a vector of probabilities for $i$th species to belong to one of three levels (e.g., A, B and C; $\mathbf{p}_1{}' = (0.2, 0.5, 0.3)$), $c_0$ and $c_1$ are cut-points (thresholds) that separate the three levels of the ordinal trait (e.g., whether a species is migratory, partially or not; colour gradients such as white, grey and black or yellow, orange and red) and these cut-points divide the continuous range of values (e.g., migration distances, colour gradients) into categories. For example, species with latent variable above $c_1$ would belong to category C (0, 0, 1), species with latent variable between $c_0$ and $c_1$ would belong to category B (0, 1, 0), and those with latent variable below $c_0$ would belong to category A (1, 0, 0). The probabilities of the three levels are calculated as the cumulative probabilities of the standard normal distribution(Hadfield, 2015). Note that the first category/level (here, A) is the **reference level**. In addition, unlike a binary GLMM (17), we have the residual variance, which is distributed with the mean of 0 and the variance of 1 (**I**, i.e., the standard normal distribution). Phylogenetic heritability on the latent scale can be obtained using 19 (but see for original scale calculation R package *QGglmm*; de Villemereuil et al., 2016).

Extending this model to take an ordinal variable with four levels is not difficult:

$$\mathbf{p_i} = \begin{pmatrix} P(y_i^{(1)} = \mathrm{A}) = p_{1i} = \Phi(c_0 - l_i) \\ P(y_i^{(2)} = \mathrm{B}) = p_{2i} = \Phi(c_1 - l_i) - \Phi(c_0 - l_i) \\ P(y_i^{(3)} = \mathrm{C}) = p_{3i} = \Phi(c_2 - l_i) - \Phi(c_1 - l_i) \\ P(y_i^{(4)} = \mathrm{D}) = p_{4i} = 1 - \Phi(c_2 - l_i) \end{pmatrix} \tag{31}$$

where $c_0$, $c_1$, and $c_2$ are cut-points that separate the four levels of the ordinal trait (e.g., whether the plumage colours of a bird species are classified as yellow, light orange, orange, or red; figure 2b). The probabilities of the four levels are calculated as the cumulative probabilities of the standard normal distribution (Hadfield, 2015; see also **Box 2**).

This model can include fixed effects to explain variation in the response variable:

$$l_i = \beta_1 x_i ... + \beta_h x_{hi} + a_i + e_i \tag{32}$$

where $l_i$ represents the linear predictor for species $i$. $x_{hi}$ represent $h$th explanatory variables associated with individual $i$, each with a corresponding regression coefficient $\beta_h$ . The term $a_i$ indicates the random effect specific to species $i$, and $e_i$ represents the residual variance, which is assumed to follow the standard normal distribution as in Equation 30. It combines the effects of explanatory variables and phylogenetic effects to produce a latent variable that determines the probability of $y_i$ categorising a specific ordinal category (here, A, B, C, or D). This formulation allows the model to account for both continuous variables (figure 3a) and categorical variables (figure 3b; with *k-1* dummy variables for *k* levels) as explanatory variables, making it versatile for analysing realistic data. Detailed guidance on applying this model and interpreting its results can be found in the section "3. worked examples" (*3.1 ordinal model*) and our online tutorial.

### 2.4 PGLMM for discrete traits 3: nominal traits

We use unordered multinomial PGLMM to model nominal traits. Multinomial (nominal) PGLMMs are multivariate (multi-response) models where with $k$ levels (categories), a set of $k-1$ binary PGLMMs are fitted (Agresti et al. (2000); Hadfield and Nakagawa (2010); figure 2c). Here, we define a nominal trait with four levels (e.g., B = blue, G = green, R = red, and W = white; figure 2c) whose multinomial model can be defined as (so comparable to running trivariate binary PGLMMs):

$$y_i^{(k)} \sim \text{Categorical}(p_{1i}, p_{2i}, p_{3i}, p_{4i}). \tag{33}$$

where $y_i^{(k)}$ is the trait value of the $i$th species for the $k$th level (e.g., B = blue, G = green, R = red, and W = white), $p_{1i}$, $p_{2i}$, $p_{3i}$, and $p_{4i}$ are the probabilities of the four levels, respectively. The probabilities of the four levels and the vector of associated three linear predictors ($\mathbf{l_i}' = (l_i^{(1)}, l_i^{(2)}, l_i^{(3)})$) are defined as follows:

$$\mathbf{p_i} = \begin{pmatrix} P(y_i^{(1)} = \text{B}) = p_{1i} = 1/\left(1 + \exp(l_{2i}) + \exp(l_{3i}) + \exp(l_{4i})\right) \\ P(y_i^{(2)} = \text{G}) = p_{2i} = \exp(l_{2i})/\left(1 + \exp(l_{2i}) + \exp(l_{3i}) + \exp(l_{4i})\right) \\ P(y_i^{(3)} = \text{R}) = p_{3i} = \exp(l_{3i})/\left(1 + \exp(l_{2i}) + \exp(l_{3i}) + \exp(l_{4i})\right) \\ P(y_i^{(4)} = \text{W}) = p_{4i} = \exp(l_{4i})/\left(1 + \exp(l_{2i}) + \exp(l_{3i}) + \exp(l_{4i})\right) \end{pmatrix} \tag{34}$$

$$\mathbf{l_i} = \begin{pmatrix} l_i^{(1)} = \ln\left(\frac{p_{2i}}{p_{1i}}\right) = \beta_0^{(1)} + \beta_1^{(1)} x_{1i} + ... + \beta_h^{(1)} x_{hi} + a_i^{(1)} \\ l_i^{(2)} = \ln\left(\frac{p_{3i}}{p_{1i}}\right) = \beta_0^{(2)} + \beta_1^{(2)} x_{1i} + ... + \beta_h^{(2)} x_{hi} + a_i^{(2)} \\ l_i^{(3)} = \ln\left(\frac{p_{4i}}{p_{1i}}\right) = \beta_0^{(3)} + \beta_1^{(3)} x_{1i} + ... + \beta_h^{(3)} x_{hi} + a_i^{(3)} \end{pmatrix}, \tag{35}$$

where $\mathbf{p}_i$ is a vector of probabilities for the $i$th species (e.g., $\mathbf{p}_i = (0.2, 0.3, 0.4, 0.1)$), $\mathbf{l_i}$ a vector consisting of $l_i^{(1)}$, $l_i^{(2)}$, and $l_i^{(3)}$, which are the linear (latent) transformation of the probability of the second, third and fourth traits in relation to the first trait (e.g., the probably of G = green in relation to reference category of B = blue), respectively, $\beta_0^{(1)}$, $\beta_0^{(2)}$, and $\beta_0^{(3)}$ are the intercepts of three models based on pair-wise comparisons (see also **Appendix**). The three phylogenetic effects $a_i^{(1)}$, $a_i^{(2)}$, and $a_i^{(3)}$ are distributed as follows:

$$\begin{pmatrix} a_i^{(1)} \\ a_i^{(2)} \\ a_i^{(3)} \end{pmatrix} = \begin{pmatrix} \mathbf{a}^{(1)} \\ \mathbf{a}^{(2)} \\ \mathbf{a}^{(3)} \end{pmatrix} \sim \mathcal{N}\left(\mathbf{0}, \Sigma_a \otimes \mathbf{A}\right) \tag{36}$$

$$\Sigma_a \otimes \mathbf{A} = \begin{pmatrix} \sigma_{a1}^2 \mathbf{A} & \rho_{a12}\sigma_{a1}\sigma_{a2}\mathbf{A} & \rho_{a13}\sigma_{a1}\sigma_{a3}\mathbf{A} \\ \rho_{a12}\sigma_{a1}\sigma_{a2}\mathbf{A} & \sigma_{a2}^2 \mathbf{A} & \rho_{a23}\sigma_{a2}\sigma_{a3}\mathbf{A} \\ \rho_{a13}\sigma_{a1}\sigma_{a3}\mathbf{A} & \rho_{a23}\sigma_{a2}\sigma_{a3}\mathbf{A} & \sigma_{a3}^2 \mathbf{A} \end{pmatrix} \tag{37}$$

where $\Sigma_a \otimes \mathbf{A}$ is the variance-covariance matrix of the phylogenetic effects (the symbol $\otimes$ denotes the Kronecker product), $\mathbf{A}$ is the phylogenetic correlation matrix, $\sigma_{a1}^2$, $\sigma_{a2}^2$, and $\sigma_{a3}^2$ are the variances of the phylogenetic effects of the first, second, and third traits, respectively, and $\rho_{a12}$, $\rho_{a13}$, and $\rho_{a23}$ are the correlations between the phylogenetic effects of the three traits. For this model, one can obtain three heritability estimates (on the latent scale) for each phylogenetic variance component using 18. This allows the model to consider both continuous variables (figure 3c) and categorical explanatory variables (figure 3d; with $k$-1 dummy variables for $k$ levels). The effect of each explanatory variable on the response variable is evaluated by comparing the reference category with each non-reference category. You can find detailed guidance on applying this model and interpreting its results in the section "3. worked examples" (*3.2 nominal model*) and our online tutorial.

## 2.5 PGLMM with non-phylogenetic species effect

An increasing number of comparative datasets now include more than one individual per species; this is especially so for continuous traits (e.g., weight and height). Therefore, one can model the within-species effect into PGLMM. For the simplest such model with the Gaussian link function can be written as:

$$y_{ij} = \beta_0 + a_i + s_i + e_j, \tag{38}$$

$$s_i = \mathbf{s} \sim \mathcal{N}(\mathbf{0}, \sigma_s^2 \mathbf{I}), \tag{39}$$

$$e_j = \mathbf{e} \sim \mathcal{N}(\mathbf{0}, \sigma_e^2 \mathbf{I}), \tag{40}$$

where $y_{ij}$ is the trait value of the $j$th individual of the $i$th species, $\beta_0$ is the intercept (across-species overall mean), $a_i$ is the phylogenetic effect of the $i$th species, $s_i$ is the non-phylogenetic effect of the $i$th species, and $s$ is a vector of non-phylogenetic effects. $e_i$ is the within-species effect (residuals, which could include measurement errors), $\mathbf{e}$ is a vector of residuals, $\sigma_e^2$ is the variance of the residuals. $\mathbf{I}$ is the identity matrix and other symbols as above. The phylogenetic heritability ($H^2$) can be obtained by the same formula (4).

The bivariate version of this model can be written as:

$$\mathbf{y_{ij}} = \begin{pmatrix} y_{ij}^{(1)} = \beta_0^{(1)} + a_i^{(1)} + s_i^{(1)} + e_{ij}^{(1)} \\ y_{ij}^{(2)} = \beta_0^{(2)} + a_i^{(2)} + s_i^{(2)} + e_{ij}^{(2)} \end{pmatrix}, \tag{41}$$

$$\begin{pmatrix} \mathbf{s}^{(1)} \\ \mathbf{s}^{(2)} \end{pmatrix} \sim \mathcal{N}\left( \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \Sigma_s \otimes \mathbf{I} \right), \tag{42}$$

$$\Sigma_s \otimes \mathbf{I} = \begin{pmatrix} \sigma_{s_1}^2 \mathbf{I} & \rho_{s12}\sigma_{s_1}\sigma_{s_2}\mathbf{I} \\ \rho_{s12}\sigma_{s_1}\sigma_{s_2}\mathbf{I} & \sigma_{s_2}^2 \mathbf{I} \end{pmatrix}, \tag{43}$$

$$\begin{pmatrix} \mathbf{e}^{(1)} \\ \mathbf{e}^{(2)} \end{pmatrix} \sim \mathcal{N}\left( \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \Sigma_e \otimes \mathbf{I} \right), \tag{44}$$

$$\Sigma_e \otimes \mathbf{I} = \begin{pmatrix} \sigma_{e_1}^2 \mathbf{I} & \rho_{e12}\sigma_{e_1}\sigma_{e_2}\mathbf{I} \\ \rho_{e12}\sigma_{e_1}\sigma_{e_2}\mathbf{I} & \sigma_{e_2}^2 \mathbf{I} \end{pmatrix}, \tag{45}$$

where $\mathbf{y_{ij}}$ is a vector consisting of $y_{ij}^{(1)}$ and $y_{ij}^{(2)}$, which are the trait values of the $j$th individual of the $i$th species for the first and second traits, respectively, $\beta_0^{(1)}$ and $\beta_0^{(2)}$ are the intercepts (across-species overall means), $a_i^{(1)}$ and $a_i^{(2)}$ are the phylogenetic effects of the $i$th species for the first

and second traits, respectively, $s_i^{(1)}$ and $s_i^{(2)}$ are the non-phylogenetic effects of the *i*th species for the first and second traits, respectively, $e_j^{(1)}$ and $e_j^{(2)}$ are the residuals (within-species effect) of the *j*th individual of the *i*th species for the first and second traits, respectively, $\mathbf{e}^{(1)}$ and $\mathbf{e}^{(2)}$ are vectors of residuals (within-species effects) for the first and second traits, respectively, $\Sigma_e \otimes \mathbf{I}$ is the variance-covariance matrix of the within-species effect, and $\sigma_e^2$ is the variance of the residuals. Also, phylogenetic heritability ($H^2$) can be obtained by the same formula (4; probably more accurately as the non-phylogenetic variance $\sigma_s$ is directly estimated and not confounded by the within-species effect and other sources of variance).

The equivalent bivariate binary PGLMM with fixed effects can be written as:

$$\mathbf{l_{ij}} = \begin{pmatrix} l_{ij}^{(1)} = \beta_0^{(1)} + \beta_1^{(1)} x_{1ij} + ... + \beta_h^{(1)} x_{hij} + a_i^{(1)} + s_i^{(1)} \\ l_{ij}^{(2)} = \beta_0^{(2)} + \beta_1^{(2)} x_{1ij} + ... + \beta_h^{(2)} x_{hij} + a_i^{(2)} + s_i^{(2)} \end{pmatrix}, \tag{46}$$

where $\mathbf{l_{ij}}$ is a vector consisting of $l_{ij}^{(1)}$ and $l_{ij}^{(2)}$, which are the linear predictors of the *j*th individual of the *i*th species for the first and second binary traits, respectively. The fixed effects, $x_{1ij}$ and $x_{2ij}$, correspond to the first and second traits at the individual level of the *j*th individual of the *i*th species. These fixed effects can be defined either at the species (between-species) level or the individual (within-species) level, depending on how the model is specified.

Now, we can model the non-phylogenetic effects explicitly, as shown in 9, enabling the estimation of non-phylogenetic variances for the first and second binary traits ($\sigma_{s1}^2$, $\sigma_{s2}^2$), as well as non-phylogenetic correlations ($\rho_{s12}$). This approach can overcome limitations in earlier bivariate binary models (21), where such detailed modelling was not possible. Although already mentioned, it is important to understand that the positive values of phylogenetic correlation ($\rho_{a12}$) and non-phylogenetic correlation ($\rho_{s12}$) can indicate co-evolution and convergence, respectively, while negative values suggest evolutionary trade-offs and ecological divergence.

Based on what was described above, one can extend ordered and unordered multinomial PGLMMs to include within-species variation. However, such datasets appear to be relatively uncommon in current comparative biology. Nonetheless, there are traits, such as colour and behavioural polymorphisms within species, where this approach could be applied. Some examples of such PGLMMs are available in the online tutorials (link).

# 3 WORKED EXAMPLES

To demonstrate the application of the PGLMMs using **Bayesian approaches**, we present some examples using the *MCMCglmm* function from package `MCMCglmm` v.2.36 (Hadfield, 2010; Hadfield and Nakagawa, 2010) and the *brm* function from `brms` v.2.21.0 (Bürkner, 2017) in `R` v.4.4.2 (R Core Team, 2024) using real-world data. The dataset used here was sourced from AVONET (Tobias et al., 2022), a comprehensive global database of avian traits. To account for phylogenetic dependency, we incorporated ultrametric phylogenetic trees obtained from BirdTree.org (Jetz et al., 2012) as a random effect in our models. Our examples here cover ordinal and nominal models, and each example provides a summary of the key results. We also offer an online tutorial (link). that provides comprehensive details on the implementation of PGLMMs, using `MCMCglmm` and `brms` packages. This includes data preparation, model fitting, and diagnostic checks for both ordinal and nominal models, as well as Gaussian and binary models.

### 3.1 Ordinal (ordered multinomial) model

We analysed the relationships between migration level (an ordinal response variable), body mass (a continuous predictor variable), and habitat density (a categorical predictor variable) across 136 species of birds of prey (figure 4a). The migration level was categorised into three distinct levels: sedentary, partially migratory, and migratory. Adult body mass (average of two sexes) was included as a predictor variable to capture potential associations with migratory capacity, reflecting physiological and energetic constraints on long movement. Habitat density (dense, semi-open, open) was also considered, with species classified based on their predominant habitat type (detailed definitions of all variables can be found in Tobias et al., 2022).

We used an intercept-only model to calculate the phylogenetic signal from the variance components because it provides an estimate of the variance attributable to phylogeny without the influence of additional explanatory variables. Phylogenetic heritability (aka phylogenetic signal), measured as the proportion of variance explained by the phylogenetic random effect, was estimated at 0.54 (95% CI: 0.00, 0.81) in `MCMCglmm` and 0.42 (95% CI: 0.00, 0.84) in `brms`. These values suggest that closely related species exhibit more similar migratory behaviours than distantly related species, although non-phylogenetic factors also play a substantial role.

Differences were observed in the threshold (cut-point) estimates between `MCMCglmm` and `brms`, arising from the packages' distinct modelling frameworks (the true-intercept model in `MCMCglmm` and the zero-intercept model in `brms`; see **Box 2** for a detailed explanation). In `MCMCglmm`, a single cut-point was estimated at 0.99 (95% CI: 0.66, 1.34) and model intercept (estimate = 0.32, 95% CI (-0.80, 1.45)). In contrast, `brms` provided two cut-points as regression coefficients: -0.32 (95% CI: -1.36, 0.61) and 0.66 (95% CI: 0.24, 1.75). These thresholds define cut-off points on an underlying continuous latent scale that determines observed ordinal categories (see the Theory section). Here, the migration level (sedentary, partially migratory, and migratory) is represented as a continuous latent variable. Species with a latent value below the first threshold are categorised as sedentary; those between the first and second thresholds are categorized as partially migratory; and values exceeding the second threshold correspond to migratory species. As explained in the Theory section (PGLMM for discrete traits 2: ordinal traits), these values can be used to calculate the proportion of species in each migration category by converting them into cumulative probabilities using the cumulative distribution function. The proportion in each category is determined by the difference between consecutive thresholds' probabilities. For instance, we can obtain that 37.5% were sedentary, 37.5% were partially migratory, and 25.0% were migratory.

We then assessed the effects of body mass and habitat density as explanatory variables. A slight negative effect of body mass on migration level was observed in both models (`MCMCglmm`: estimate = -0.233, 95% CI: -0.51, 0.02; `brms`: estimate = -0.26, 95% CI: -0.57, 0.02; figure 4a). These estimates suggest that larger species may exhibit lower migratory tendencies, although the effects are relatively modest and overlap with zero. As mentioned above, habitat density was categorised as dense (reference), open, and semi-open. Species living in open environments (e.g., grasslands) had a higher likelihood of being migratory compared to those in dense habitats (e.g., dense thickets) (`MCMCglmm`: estimate = 1.15, 95% CI: 0.52, 1.86; `brms`: estimate = 1.24, 95% CI: 0.57, 2.05; figure 4a). In contrast, species in semi-open environments (e.g., open shrublands) did not show a significant difference in migratory behaviour compared to those in dense habitats (`MCMCglmm`: estimate = 0.32, 95% CI: -0.26, 0.89; `brms`: estimate = 0.37, 95% CI: -0.20, 0.99; figure 4a).

### 3.2 Nominal (unordered multinomial) model

We used the predominant locomotory niche of 173 thrush bird species data as the nominal response variable (figure 4b). This variable consists of three categories: ground-dwelling, perching, and generalist. The ground-dwelling category includes species that primarily spend their time on the ground, foraging by walking or hopping. The perching category refers to species that spend much of their time perched on raised surfaces, such as tree branches, rocks, buildings, posts, or wires. The generalist category includes species that do not predominantly fit into a single lifestyle and utilise a variety of niches. The explanatory variables include two models: the first uses tail length as a single predictor, and the second includes both tail length and diet as predictors. Tail length may be associated with manoeuvrability and locomotion strategies, which can vary across ground-dwelling, perching, and generalist lifestyles. Diet reflects foraging behaviour and habitat use, which can also be key factors shaping locomotory niches. Diet originally contained three categories (carnivore, herbivore, omnivore), but for ease of interpretation, it was recoded into a binary variable distinguishing between omnivores and non-omnivores (figure 4b). In our model, the "generalist" category is the reference category because it appears first alphabetically. When discrete (categorical) variables are included in a model, the reference category is usually the first alphabetical unless you expressly set another category as the reference.

When the phylogenetic signal is strong, it indicates that closely related species are more likely to belong to the same trait category (for example, both being ground-dwelling). This suggests that the trait is relatively stable over evolutionary time and strongly influenced by phylogenetic history. Conversely, closely related species are less likely to share the same category when the phylogenetic signal is weak. They may fall into different categories (such as one species being ground-dwelling while another being perching). This implies that the trait changes frequently and is less constrained by phylogenetic relationships. The phylogenetic heritability, measured as the proportion of variance explained by the phylogenetic random effect, was estimated at 0.83 (95% CI:0.55, 0.91) in `MCMCglmm` and 0.96 (95% CI: 0.78, 0.99) in `brms` for perching. Ground-dwelling phylogenetic heritability was estimated at 0.83 (95% CI: 0.60, 0.92) in `MCMCglmm` and 0.97 (95% CI: 0.83, 1.00) in `brms`. These values suggest that the "perching" and "ground-dwelling" are both strongly influenced by phylogenetic relationships, indicating a high level of evolutionary conservatism. This implies that their distribution among species is not random. Phylogenetic correlation, which is a measure of the degree to which traits are correlated due to shared evolutionary history, was 0.06 (95% CI: -0.12, 0.12) in `MCMCglmm` and -0.27 (95% CI: -0.88, 0.50) in `brms`. The 95% CIs cross 0, so there was no statistically significant relationship between perching and ground-dwelling traits. A positive correlation would suggest that the evolution of one trait (perching) is associated with an increased likelihood of the evolution of the other trait (ground-dwelling), indicating that these traits may tend to evolve together. Conversely, a negative correlation would imply that a strong evolutionary tendency toward one trait (perching) reduces the likelihood of exhibiting the other trait (ground-dwelling), reflecting a potential evolutionary trade-off between the two traits.

We examined the effects of tail length and diet (omnivorous vs. non-omnivorous) on primary avian lifestyles. Tail lengths were log-transformed and centred to facilitate interpretation. Results from both the `MCMCglmm` and `brms` models indicated that tail length did not significantly influence whether species were more likely to adopt perching or ground-dwelling lifestyles compared to generalists (perching - `MCMCglmm`: estimate = -0.21, 95%CI: -3.25, 2.85; `brms`: estimate = -0.88, 95%CI: -4.83, 2.84 / ground-dwelling - `MCMCglmm`: estimate = -1.53, 95%CI: -4.25, 1.27; `brms`: estimate = -1.51, 95%CI: -5.43, 2.28; figure 4b). While omnivory did not have a statistically

significant impact on the likelihood of adopting a perching versus generalist lifestyle (`MCMCglmm`: estimates = -0.29, 95%CI: -1.55, 1.00; `brms`: estimate = 0.13, 95%CI: -1.54, 2.00; figure 4b), omnivorous species were more likely to exhibit diverse lifestyles, such as being generalists rather than ground-dwelling (`MCMCglmm`: estimates = -3.53, 95%CI: -4.86, -2.41; `brms`: estimate = -4.47, 95%CI: -7.13, -2.70; figure 4b). This means that if a species is omnivorous, it is more likely to be generalist than ground-dwelling or perching.

# 4 CONCLUSION

Through this paper, we have provided a theoretical and practical overview of how to implement phylogenetic generalised linear mixed-effects models (PGLMMs) for discrete traits along with continuous counterparts. As far as we know, our work presents the first comprehensive introduction to PGLMMs for discrete traits, showing how binary PGLMMs act as bridges to other types of PGLMMs. Importantly, this theoretical introduction is complemented by an extensive online tutorial, which covers practical difficulties implementing such PGLMMs using Bayesian statistics, for example, 1) setting Bayesian priors, 2) checking MCMC chain convergence and mixing, 3) interpreting regression coefficients and (co)variance components, and 4) simulating toy datasets. We hope this paper demonstrates the utility and flexibility of linear models in analysing discrete traits and encourages their broader adoption in evolutionary and ecological studies.

With the growth of open science, the rapid accumulation of data across biological disciplines has created both challenges and opportunities (e.g., Marx, 2013; Losos et al., 2013; Cushman, 2014; Pal et al., 2020). We can also combine existing research data to uncover new knowledge (e.g., Gallagher et al., 2020). One of the challenges is effectively organising these immense datasets, which needs a proper understanding of appropriate statistical analyses and methodologies. Since Darwin introduced the theory of natural selection (Darwin, 1859), researchers have made remarkable progress in understanding and unravelling biological phenomena. However, significant gaps remain in our understanding of the ecology and evolution of living organisms, and many topics continue to be debated. We believe our work will help researchers handle their data more effectively, fostering further progress in evolutionary and ecological research.

# REFERENCES


Adams, D. C. and Collyer, M. L. (2024). Extending phylogenetic regression models for comparing within-species patterns across the tree of life. *Methods in ecology and evolution*.

Agresti, A., Booth, J. G., Hobert, J. P., and Caffo, B. (2000). 2. random-effects modeling of categorical response data. *Sociological methodology*, 30:27–80.

Anderson, J. A. (1984). Regression and ordered categorical variables. *Journal of the Royal Statistical Society. Series B, Statistical methodology*, 46:1–22.

Blomberg, S. P., Garland, Jr, T., and Ives, A. R. (2003). Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution; international journal of organic evolution*, 57:717–745.

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

Cornwallis, C. K. and Griffin, A. S. (2024). A guided tour of phylogenetic comparative methods for studying trait evolution. *Annual review of ecology, evolution, and systematics*, 55:181–204.

Cornwell, W. and Nakagawa, S. (2017). Phylogenetic comparative methods. *Current biology*, 27:R333–R336.

Cushman, S. A. (2014). Grand challenges in evolutionary and population genetics: the importance of integrating epigenetics, genomics, modeling, and experimentation. *Frontiers in genetics*, 5:197.

Darwin, C. (1859). *Origin of the Species*.

de Villemereuil, P., Schielzeth, H., Nakagawa, S., and Morrissey, M. (2016). General methods for evolutionary quantitative genetic inference from generalized mixed models. *Genetics*, 204:1281–1294.

Felsenstein, J. (1973). Maximum likelihood and minimum-steps methods for estimating evolutionary trees from data on discrete characters. *Systematic biology*, 22:240–249.

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

Felsenstein, J. (2012). A comparative method for both discrete and continuous characters using the threshold model. *The American naturalist*, 179:145–156.

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

Gallagher, R. V., Falster, D. S., Maitner, B. S., Salguero-Gómez, R., Vandvik, V., Pearse, W. D., Schneider, F. D., Kattge, J., Poelen, J. H., Madin, J. S., Ankenbrand, M. J., Penone, C., Feng, X., Adams, V. M., Alroy, J., Andrew, S. C., Balk, M. A., Bland, L. M., Boyle, B. L., Bravo-Avila, C. H., Brennan, I., Carthey, A. J. R., Catullo, R., Cavazos, B. R., Conde, D. A., Chown, S. L., Fadrique, B., Gibb, H., Halbritter, A. H., Hammock, J., Hogan, J. A., Holewa, H., Hope, M., Iversen, C. M., Jochum, M., Kearney, M., Keller, A., Mabee, P., Manning, P., McCormack, L., Michaletz, S. T., Park, D. S., Perez, T. M., Pineda-Munoz, S., Ray, C. A., Rossetto, M., Sauquet, H., Sparrow, B., Spasojevic, M. J., Telford, R. J., Tobias, J. A., Violle, C., Walls, R., Weiss, K. C. B., Westoby, M., Wright, I. J., and Enquist, B. J. (2020). Open science principles for accelerating trait-based science across the tree of life. *Nature ecology & evolution*, 4:294–303.

Garamszegi, L. Z. (2014). *Modern phylogenetic comparative methods and their application inevolutionary biology: Concepts and practice*. Springer, New York, NY.

Garland, Jr, T. and Ives, A. R. (2000). Using the past to predict the present: Confidence intervals for regression equations in phylogenetic comparative methods. *The American naturalist*, 155:346–364.

Gavryushkin, A. and Drummond, A. J. (2016). The space of ultrametric phylogenetic trees. *Journal of theoretical biology*, 403:197–208.

Grafen, A. (1989). The phylogenetic regression. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 326:119–157.

Grizzle, J. E., Starmer, C. F., and Koch, G. G. (1969). Analysis of categorical data by linear models. *Biometrics*, 25:489–504.

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*, 6:706–714.

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

Harmon, L. J. (2018). *Phylogenetic comparative methods: Learning from trees*. Createspace Independent Publishing Platform, North Charleston, SC.

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

Ives, A. R. and Helmus, M. R. (2011). Generalized linear mixed models for phylogenetic analyses

of community structure. *Ecological monographs*, 81:511–525.
Ives, A. R., Midford, P. E., and Garland, Jr, T. (2007). Within-species variation and measurement error in phylogenetic comparative methods. *Systematic biology*, 56:252–270.
Jetz, W., Thomas, G. H., Joy, J. B., Hartmann, K., and Mooers, A. O. (2012). The global diversity of birds in space and time. *Nature*, 491:444–448.
Lewis, P. O. (2001). A likelihood approach to estimating phylogeny from discrete morphological character data. *Systematic biology*, 50:913–925.
Losos, J. B., Arnold, S. J., Bejerano, G., Brodie, 3rd, E. D., Hibbett, D., Hoekstra, H. E., Mindell, D. P., Monteiro, A., Moritz, C., Orr, H. A., Petrov, D. A., Renner, S. S., Ricklefs, R. E., Soltis, P. S., and Turner, T. L. (2013). Evolutionary biology for the 21st century. *PLoS biology*, 11:e1001466.
Lynch, M. (1991). Methods for the analysis of comparative data in evolutionary biology. *Evolution*, 45:1065–1080.
Martins, E. and Hansen, T. F. (1997). Phylogenies and the comparative method: A general approach to incorporating phylogenetic information into the analysis of interspecific data. *The American naturalist*, 149:646–667.
Marx, V. (2013). Biology: The big challenges of big data: Biology. *Nature*, 498:255–260.
Mora, C., Tittensor, D. P., Adl, S., Simpson, A. G. B., and Worm, B. (2011). How many species are there on earth and in the ocean? *PLoS biology*, 9:e1001127.
Münkemüller, T., Lavergne, S., Bzeznik, B., Dray, S., Jombart, T., Schiffers, K., and Thuiller, W. (2012). How to measure and test phylogenetic signal. *Methods in Ecology and Evolution*, 3:743–756.
Nakagawa, S. and Schielzeth, H. (2010). Repeatability for gaussian and non-gaussian data: a practical guide for biologists. *Biological reviews*, 85:no–no.
Nakagawa, S. and Schielzeth, H. (2013). A general and simple method for obtainingR2from generalized linear mixed-effects models. *Methods in ecology and evolution*, 4:133–142.
O'Meara, B. C. (2012). Evolutionary inferences from phylogenies: A review of methods. *Annual review of ecology, evolution, and systematics*, 43:267–285.
Pagel, M. (1994). Detecting correlated evolution on phylogenies: a general method for the comparative analysis of discrete characters. *Proceedings. Biological sciences / The Royal Society*, 255:37–45.
Pagel, M. (1997). Inferring evolutionary processes from phylogenies. *Zoologica scripta*, 26:331–348.
Pagel, M. (1999). Inferring the historical patterns of biological evolution. *Nature*, 401:877–884.
Pal, S., Mondal, S., Das, G., Khatua, S., and Ghosh, Z. (2020). Big data in biology: The hope and present-day challenges in it. *Gene reports*, 21:100869.
Pottier, P., Noble, D. W. A., Seebacher, F., Wu, N. C., Burke, S., Lagisz, M., Schwanz, L. E., Drobniak, S. M., and Nakagawa, S. (2024). New horizons for comparative studies and meta-analyses. *Trends in ecology & evolution*, 39:435–445.
R Core Team (2024). R: A language and environment for statistical computing. R foundation for statistical computing, vienna, austria.
Revell, L. J., Harmon, L. J., and Collar, D. C. (2008). Phylogenetic signal, evolutionary process, and rate. *Systematic biology*, 57:591–601.
Rohlf, F. J. (2001). Comparative methods for the analysis of continuous variables: geometric interpretations. *Evolution; international journal of organic evolution*, 55:2143–2160.
Scheffers, B. R., Joppa, L. N., Pimm, S. L., and Laurance, W. F. (2012). What we know and don't know about earth's missing biodiversity. *Trends in ecology & evolution*, 27:501–510.
Sloane, D. and Morgan, S. P. (1996). An introduction to categorical data analysis. *Annual review of*

*sociology*, 22:351–375.

Tobias, J. A., Sheard, C., Pigot, A. L., Devenish, A. J. M., Yang, J., Sayol, F., Neate-Clegg, M. H. C., Alioravainen, N., Weeks, T. L., Barber, R. A., Walkden, P. A., MacGregor, H. E. A., Jones, S. E. I., Vincent, C., Phillips, A. G., Marples, N. M., Montaño Centellas, F. A., Leandro-Silva, V., Claramunt, S., Darski, B., Freeman, B. G., Bregman, T. P., Cooney, C. R., Hughes, E. C., Capp, E. J. R., Varley, Z. K., Friedman, N. R., Korntheuer, H., Corrales-Vargas, A., Trisos, C. H., Weeks, B. C., Hanz, D. M., Töpfer, T., Bravo, G. A., Remeš, V., Nowak, L., Carneiro, L. S., Moncada R, A. J., Matysioková, B., Baldassarre, D. T., Martínez-Salinas, A., Wolfe, J. D., Chapman, P. M., Daly, B. G., Sorensen, M. C., Neu, A., Ford, M. A., Mayhew, R. J., Fabio Silveira, L., Kelly, D. J., Annorbah, N. N. D., Pollock, H. S., Grabowska-Zhang, A. M., McEntee, J. P., Carlos T Gonzalez, J., Meneses, C. G., Muñoz, M. C., Powell, L. L., Jamie, G. A., Matthews, T. J., Johnson, O., Brito, G. R. R., Zyskowski, K., Crates, R., Harvey, M. G., Jurado Zevallos, M., Hosner, P. A., Bradfer-Lawrence, T., Maley, J. M., Stiles, F. G., Lima, H. S., Provost, K. L., Chibesa, M., Mashao, M., Howard, J. T., Mlamba, E., Chua, M. A. H., Li, B., Gómez, M. I., García, N. C., Päckert, M., Fuchs, J., Ali, J. R., Derryberry, E. P., Carlson, M. L., Urriza, R. C., Brzeski, K. E., Prawiradilaga, D. M., Rayner, M. J., Miller, E. T., Bowie, R. C. K., Lafontaine, R.-M., Scofield, R. P., Lou, Y., Somarathna, L., Lepage, D., Illif, M., Neuschulz, E. L., Templin, M., Dehling, D. M., Cooper, J. C., Pauwels, O. S. G., Analuddin, K., Fjeldså, J., Seddon, N., Sweet, P. R., DeClerck, F. A. J., Naka, L. N., Brawn, J. D., Aleixo, A., Böhning-Gaese, K., Rahbek, C., Fritz, S. A., Thomas, G. H., and Schleuning, M. (2022). AVONET: morphological, ecological and geographical data for all birds. *Ecology letters*, 25:581–597.

**Box 1: Glossary**

**Ancestral states**
Ancestral states refer to the traits of organisms that existed in their common ancestors. We often use it in the context of reconstructing the traits of an ancestral species based on the traits observed in its living descendant species. For example, if we are studying the trait evolution of a certain family, we may be interested in estimating what trait their common ancestral trait is. These estimated traits, such as the plumage colour, diet, or habitat type, are the "ancestral states."

**Linear model**
A linear (regression) model assumes a linear relationship between the response variable(s) and one or more independent variables. The model relies on assumptions such as linearity, independence, homoscedasticity, and normally distributed residuals. In more complex models, such as mixed-effects models (also called hierarchical models), random effects are included to account for variability not captured by the fixed effects. We can include that observations within specific groups (e.g., individuals, locations, or time points) may vary in ways that are different from the overall trend as the random effect. This allows the model to handle data that is organised into groups, where we expect the data points within each group to be more similar to each other than to data points from other groups. If the model does not include any explanatory variable, the model does not explain y based on any predictors. The prediction of $y$ is just the overall mean, with random effects $u_j$ allowing for variations between groups or clusters (if using a mixed-effects model) and $\varepsilon$ accounting for individual random errors. This model is often called a null model and is used as a baseline for comparing models that include explanatory variables.

**Markov model**
A Markov model is a mathematical framework used to describe systems that transition from one state to another in a probabilistic manner that is based on the Markov Property. Markov Property assumes that the future state depends only on the current state and not on the past events. Markov models are widely used in various fields due to their versatility in modelling probabilistic systems (simplifying complex systems while retaining useful predictive power). In evolutionary biology, the Markov k-state model (Mk model) is popular and used in phylogenetics to describe the evolution of discrete traits across a phylogenetic tree. The Mk model is commonly used to study the evolution of discrete traits, such as the gain or loss of morphological features and the evolution of behavioural traits. It assumes that trait changes occur according to a Markov process. The transition rates at which changes occur between these states are modelled as a continuous-time process. All states are equally likely to transition to another state unless otherwise specified. The process is time-reversible, meaning the probability of transitioning from state A to state B over time is the same as transitioning from state B to state A (under the same conditions).

**Log odds-ratio**
Log odds is the natural logarithm of the odds ratio, where the odds ratio represents the ratio of the probability *P* of an event occurring to the probability *1-P* of it not occurring. The relationship between explanatory variables and probability is described as non-linear, making it difficult to model using a simple linear relationship. Additionally, while probabilities are constrained to a range of 0 to 1, linear predictors can take any value, creating a mismatch that must be addressed. By using log odds, the odds are transformed into a scale ranging from $-\infty$ to $+\infty$, allowing for the application of the linear framework (i.e., logistic regression). The log odds can then be back-transformed using the logistic function, enabling the results to be interpreted as probabilities within the range of 0 to 1.

**Phylogenetic heritability/phylogenetic signal**
Phylogenetic signal shows statistical non-independence between trait values and their phylogeny and indicates how closely related species resemble each other (e.g., Revell et al., 2008; Münkemüller et al., 2012). A strong phylogenetic signal indicates that closely related species resemble each other more than expected by chance, reflecting the influence of shared ancestors. Pagel's $\lambda$, one of the phylogenetic signal indices, is also interpreted as phylogenetic heritability.

**Uni- and multivariate model (uni-response and multi-response variable models)**
In our paper, we define the univariate model as a model with a single dependent variable based on one or more independent variables, while the multivariate model is defined as the model that includes multiple dependent variables simultaneously. Multivariate models can consider correlations between dependent variables (detailed definitions of all variables can be found in Tobias et al., 2022).

**Phylogenetic variance-covariance matrix / phylogenetic correlation matrix**
In the context of phylogenetic analysis, the variance-covariance matrix reflects how variables of interest (e.g., traits or behaviours) vary and covary across species. The diagonal components represent the variance of each variable, indicating the extent to which a variable deviates from its mean across species. The off-diagonal components represent the covariance between pairs of variables, showing how two variables co-vary across species. Covariance indicates both the direction and strength of the relationship between two variables, but its value is influenced by the
574 units of measurement, making its magnitude scale-dependent. To remove this scale dependence, covariance is standardised by dividing it by the product of the standard deviations of the two variables (X and Y), resulting in the correlation coefficient ($r$). The correlation matrix contains these correlation coefficients, with off-diagonal components indicating the relationship between pairs of variables.

**Probit and logit model**
Probit model coefficients represent the effect that an independent variable has on a latent dependent variable, which follows a standard normal distribution. This latent variable determines the probability that the observed dependent variable falls into a specific category. Logit model coefficients describe the log odds, showing the effect that a one-unit increase in an independent variable has on the probability that the dependent variable is classified into a specific category (as opposed to the reference category). The logit model is widely used due to its simplicity of calculation and ease of interpretation. Both models often give similar results, so you can use either one depending on the context.

**Link function and linear predictor**
In GLMMs, the linear predictor ($l$) is the combination of fixed effects and random effects. The link function can transform the expected value of the response variable to relate it to the linear predictor. This allows GLMMs to model non-linear relationships between predictors and the response variables. For example, when a response variable is binary traits, the link function can be the logit or probit link function.

**Cumulative Distribution Function (CDF) and Probability Density Function (PDF)**
A Probability Density Function (PDF) describes the probability distribution of continuous random variables. It indicates the density of probability at different values. While the PDF itself does not give the probability at a specific point, the probability over a range can be calculated by integrating the PDF over that range. The total area under the PDF curve equals 1. A Cumulative Distribution Function (CDF) gives the probability that a random variable takes a value less than or equal to a given point. It represents the accumulated probability up to that point, ranging from 0 to 1. The CDF can be derived by integrating the PDF.

**Threshold model**
The threshold regression model is used for binary or ordered response variables. The concept behind the threshold model is that an unobserved continuous latent variable determines the observed discrete (categorical) response variable based on whether it exceeds certain thresholds (cut-points). The model assumes that $y$ is a continuous latent variable, but the response variable $y$ is observed in discrete categories.

**Latent variable (unobserved variable)**
A latent variable ($l$) provides the link function (e.g., logit and probit) to the values of the response variable ($y$). For example, when a response variable is a binary outcome, the latent variable is determined by whether $l$ exceeds a threshold. If $l$ is under the threshold, then $y = 1$; if $l$ is more than the threshold, then $y = 0$.

**Reference category/level (baseline category)**
Both ordered or unordered variables include two or more categories (e.g., colour - blue, red, white). The reference (baseline) category is used to compare with other categories and act as a "reference." In the regression model, the model estimates the coefficients as the difference between the non-reference and reference categories.

**Bayesian statistics (approaches)**
Bayesian approach is a statistical method that combines prior knowledge (prior distribution) with observed data (likelihood) to update beliefs and derive the posterior distribution of unknown parameters in a model. The term "parameters" refers to the unknown values needed to define the model's structure, and the Bayesian approach seeks to estimate the probability distribution of these parameters. Computational methods such as Markov Chain Monte Carlo (MCMC) are used to incorporate uncertainty in the data. Bayesian estimation considers the entire distribution of the parameter, allowing for the reflection of uncertainty in the parameter estimates.

**Box 2: Two ways of parametrising the ordinal (threshold) model** The ordinal (threshold) model has two different parametrisations, *the true-intercept model* and *the zero-intercept model*, with different interpretations and implementations. For example, `MCMCglmm` uses the former, and `brms` uses the latter. The true-intercept model always has the intercept of the regression model equal to the mean of its probability density function (PDF). Zero is always the first cut-point in every model, which is implicit. The intercept is always fixed at zero in the zero-intercept models, but the cut-points differ between regression models. Figure B1 is a conceptual description of the difference between the true- and zero-intercept models. The model has three category-ordered predictor variables (A, B, and C). The latent variable is defined as $l_i = b_0 + a_i$, where $b_0$ represents the intercept and $a_i$ means the phylogenetic random effect, which determines the probability of each category. The blue-shaded PDF illustrates the PDF of the latent variable distribution (i.e., the underlying distribution of the response variable, not influenced by predictor variables). The warm-coloured areas indicate the probability of the observed outcome y being categorised into A, B, or C, depending on where the value of the latent variable $l$ falls relative to the thresholds. The dark-yellow solid line describes the intercept ($b_0$), which is the mean of the latent distribution. The pink dashed lines represent the thresholds (cut-points) that define the boundaries between categories A, B, and C. The true-intercept model usually reports only the second cut-point, with the first cut-point set to 0 (e.g., `MCMCglmm`). The single reported threshold (threshold1) can move and define the boundary between the second and third categories. In contrast, the zero-intercept model (figure B1b) reported two cut-points (outputted as intercepts in `brms`), with the model intercept fixed at 0. The two cut-points can move and correspond to the thresholds between the first and second categories (threshold 0: A and B) and between the second and third categories (threshold 1: B and C), respectively.

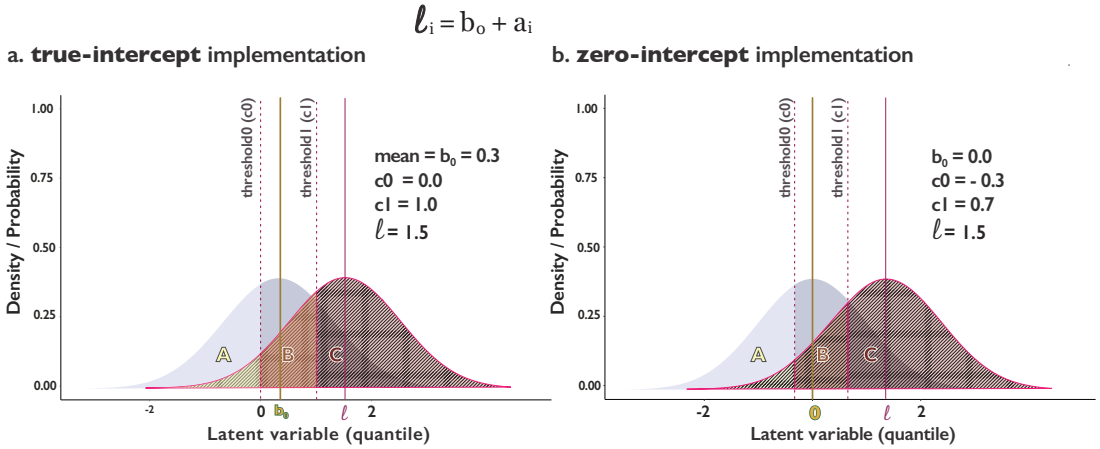


**Figure B1**. Different parametrisation of ordinal threshold models. (a) In the true-intercept model, the model intercept corresponds to the mean of the latent variable distribution and can vary across models, while the first threshold (cut-point) is always fixed at 0. (b) In the zero-intercept model, the model intercept is always fixed at 0, and the first threshold (cut-point) can vary between models.

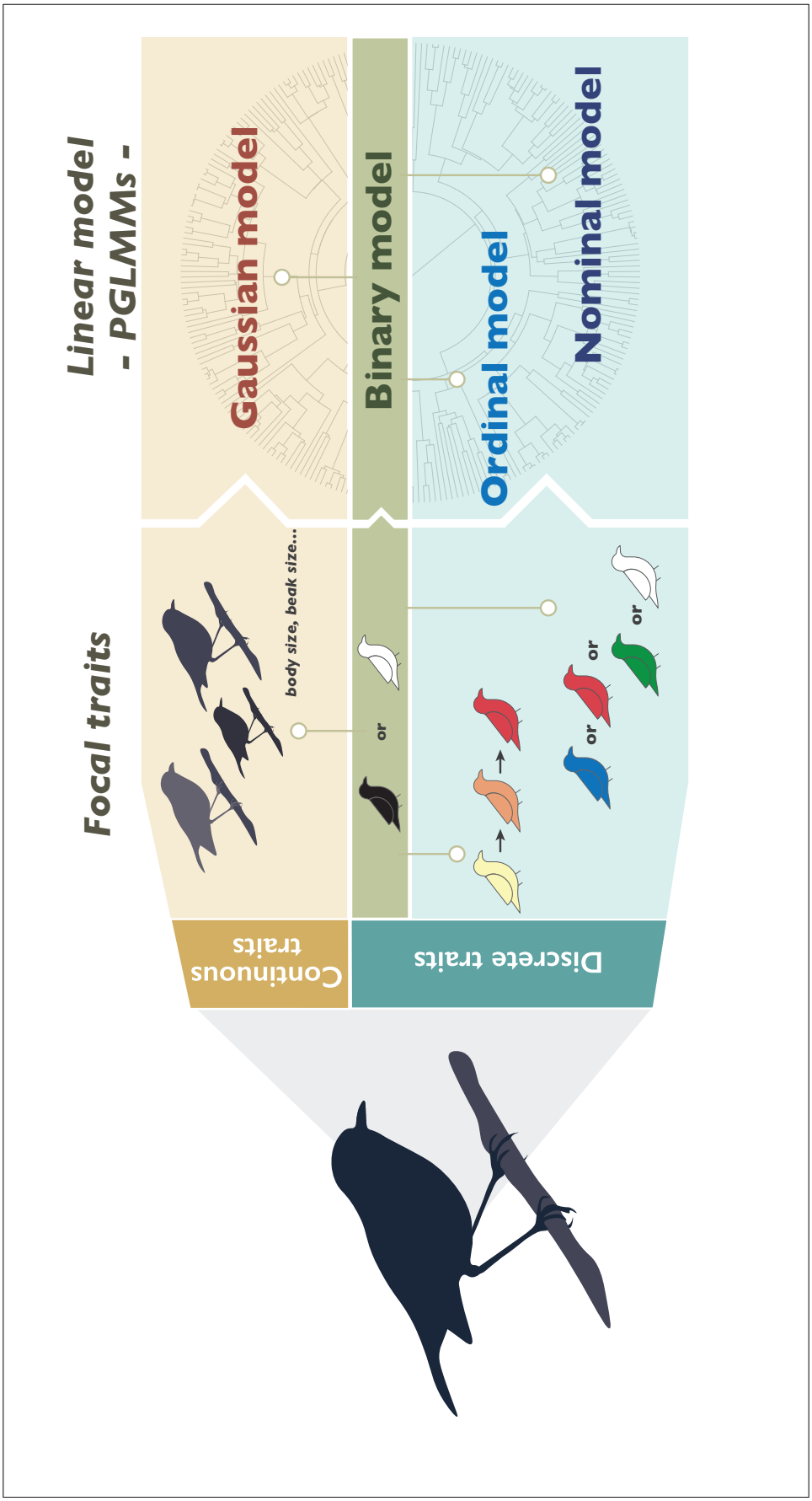
Linear model
- PGLMMs -
Gaussian model
Binary model
Ordinal model
Nominal model
Focal traits
body size, beak size...
or
or
or
or
Continuous traits
Discrete traits

**Figure 1**

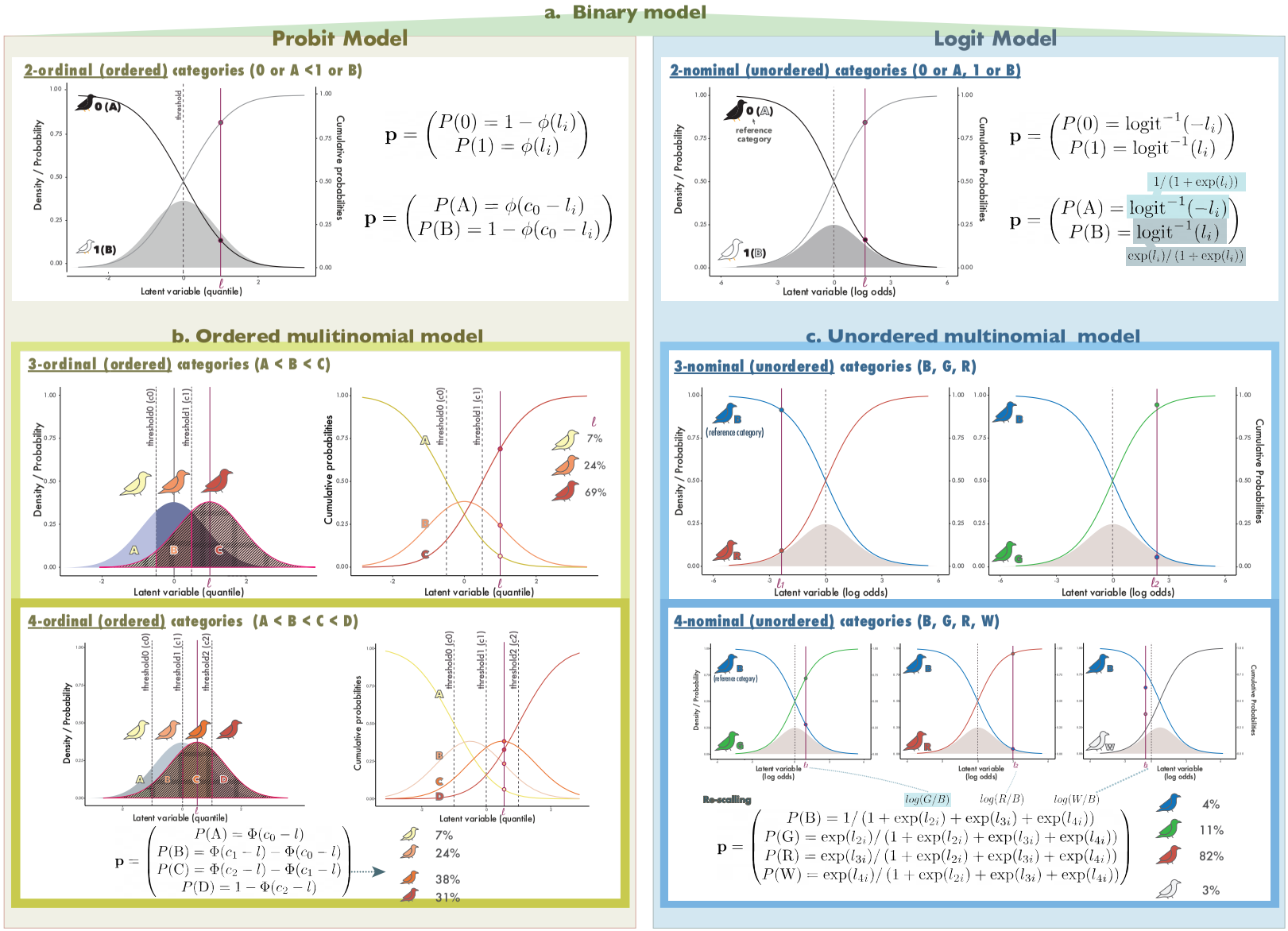
a. Binary model
Probit Model
2-ordinal (ordered) categories (0 or A <1 or B)
0 (A)
1(B)
threshold
Density / Probability
Cumulative probabilities
Latent variable (quantile)
$\mathbf{p} = \begin{pmatrix} P(0) = 1 - \phi(l_i) \\ P(1) = \phi(l_i) \end{pmatrix}$
$\mathbf{p} = \begin{pmatrix} P(\mathrm{A}) = \phi(c_0 - l_i) \\ P(\mathrm{B}) = 1 - \phi(c_0 - l_i) \end{pmatrix}$
Logit Model
2-nominal (unordered) categories (0 or A, 1 or B)
reference category
Cumulative Probabilities
Latent variable (log odds)
$\mathbf{p} = \begin{pmatrix} P(0) = \mathrm{logit}^{-1}(-l_i) \\ P(1) = \mathrm{logit}^{-1}(l_i) \end{pmatrix}$
$1/(1+\exp(l_i))$
$\mathbf{p} = \begin{pmatrix} P(\mathrm{A}) = \mathrm{logit}^{-1}(-l_i) \\ P(\mathrm{B}) = \mathrm{logit}^{-1}(l_i) \end{pmatrix}$
$\exp(l_i)/(1+\exp(l_i))$
b. Ordered mulitinomial model
3-ordinal (ordered) categories (A < B < C)
threshold0 (c0)
threshold1 (c1)
Cumulative probabilities
7%
24%
69%
4-ordinal (ordered) categories (A < B < C < D)
threshold2 (c2)
$\mathbf{p} = \begin{pmatrix} P(\Lambda) = \Phi(c_0 - l) \\ P(\mathrm{B}) = \Phi(c_1 - l) - \Phi(c_0 - l) \\ P(\mathrm{C}) = \Phi(c_2 - l) - \Phi(c_1 - l) \\ P(\mathrm{D}) = 1 - \Phi(c_2 - l) \end{pmatrix}$
7%
24%
38%
31%
c. Unordered multinomial model
3-nominal (unordered) categories (B, G, R)
(reference category)
Latent variable (log odds)
4-nominal (unordered) categories (B, G, R, W)
Re-scaling
$log(G/B)$
$log(R/B)$
$log(W/B)$
$\mathbf{p} = \begin{pmatrix} P(\mathrm{B}) = 1/(1+\exp(l_{2i})+\exp(l_{3i})+\exp(l_{4i})) \\ P(\mathrm{G}) = \exp(l_{2i})/(1+\exp(l_{2i})+\exp(l_{3i})+\exp(l_{4i})) \\ P(\mathrm{R}) = \exp(l_{3i})/(1+\exp(l_{2i})+\exp(l_{3i})+\exp(l_{4i})) \\ P(\mathrm{W}) = \exp(l_{4i})/(1+\exp(l_{2i})+\exp(l_{3i})+\exp(l_{4i})) \end{pmatrix}$
4%
11%
82%
3%

**Figure 2**

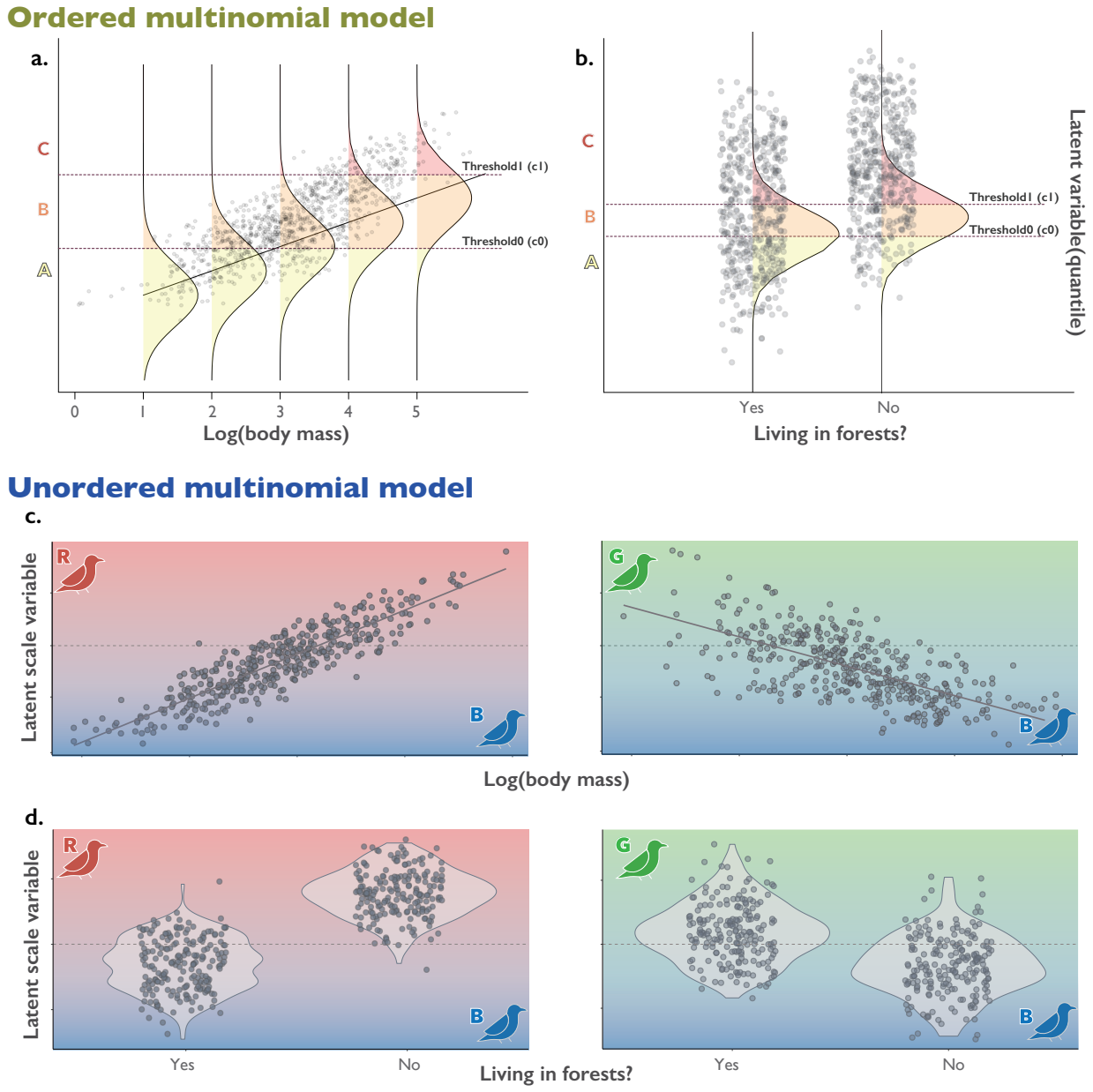
Ordered multinomial model
a.
C
B
A
Threshold1 (c1)
Threshold0 (c0)
0
1
2
3
4
5
Log(body mass)
b.
C
B
A
Threshold1 (c1)
Threshold0 (c0)
Latent variable(quantile)
Yes
No
Living in forests?
Unordered multinomial model
c.
Latent scale variable
R
B
G
B
Log(body mass)
d.
Latent scale variable
R
B
G
B
Yes
No
Yes
No
Living in forests?

**Figure 3**

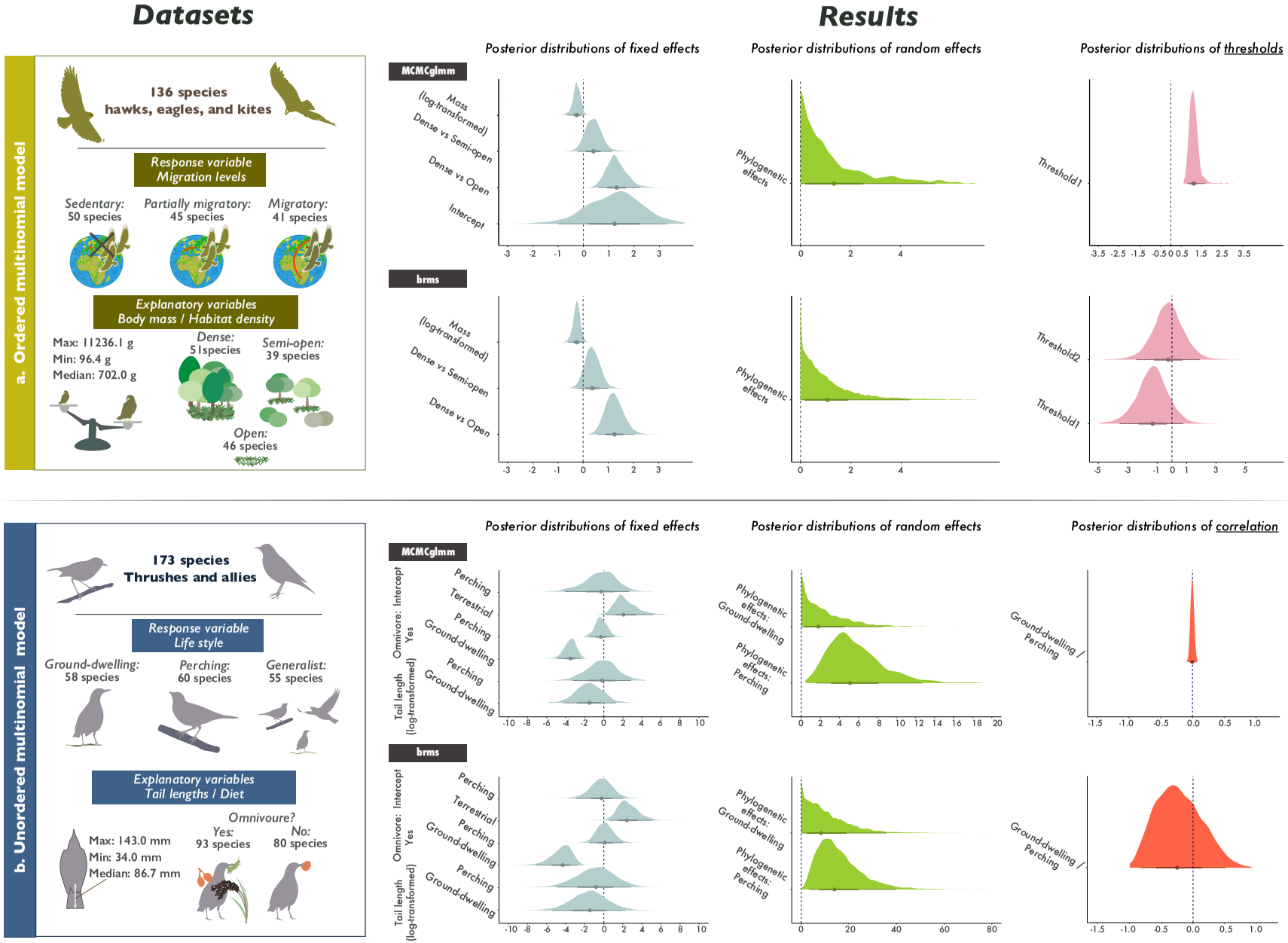
Datasets
Results
a. Ordered multinomial model
136 species
hawks, eagles, and kites
Response variable
Migration levels
Sedentary:
50 species
Partially migratory:
45 species
Migratory:
41 species
Explanatory variables
Body mass / Habitat density
Max: 11236.1 g
Min: 96.4 g
Median: 702.0 g
Dense:
51species
Semi-open:
39 species
Open:
46 species
Posterior distributions of fixed effects
Posterior distributions of random effects
Posterior distributions of thresholds
MCMCglmm
Mass (log-transformed)
Dense vs Semi-open
Dense vs Open
Intercept
Phylogenetic effects
Threshold1
brms
Mass (log-transformed)
Dense vs Semi-open
Dense vs Open
Phylogenetic effects
Threshold2
Threshold1
b. Unordered multinomial model
173 species
Thrushes and allies
Response variable
Life style
Ground-dwelling:
58 species
Perching:
60 species
Generalist:
55 species
Explanatory variables
Tail lengths / Diet
Max: 143.0 mm
Min: 34.0 mm
Median: 86.7 mm
Omnivoure?
Yes:
93 species
No:
80 species
Posterior distributions of fixed effects
Posterior distributions of random effects
Posterior distributions of correlation
MCMCglmm
Omnivore: Yes
Tail length (log-transformed)
Perching
Terrestrial
Perching
Ground-dwelling
Perching
Ground-dwelling
Phylogenetic effects: Ground-dwelling
Phylogenetic effects: Perching
Ground-dwelling / Perching
brms
Omnivore: Yes
Tail length (log-transformed)
Perching
Terrestrial
Perching
Ground-dwelling
Perching
Ground-dwelling
Phylogenetic effects: Ground-dwelling
Phylogenetic effects: Perching
Ground-dwelling / Perching

**Figure 4**

# 5 FIGURE LEGENDS

**Figure 1.** Conceptual diagram of relationships between different types of traits and phylogenetic mixed-effect models (PGLMMs). The figure demonstrates how organism traits can be categorised into continuous and discrete (categorical) traits, showing the relationships between the Gaussian model and discrete models within the framework of PGLMM. The binary model, as one of the discrete models, serves as a conceptual bridge between the Gaussian model and ordinal or nominal models. On the latent scale, the binary model closely aligns with the Gaussian model, reflecting their underlying similarity. It is worth noting that ordinal and nominal models can be understood as extensions of the binary model (for further details, refer to the main text). Notably, this paper does not deal with count data (e.g., frequencies of a certain behaviour, the number of certain morphological patterns), although such data can be considered as discrete traits. Yet, count data are usually not considered categorical traits.

**Figure 2.** Overview of the discrete models. (a) Binary model. The binary model represents the simplest structure and serves as the foundation for more complex models such as (b) ordinal and (c) nominal models. In the binary model, two common link functions are used: the probit and the logit link functions. In the probit model (a: right), the probability density plot illustrates the distribution of the latent variable, which follows a standard normal distribution $N(0,1)$. Most of the data lies within the range between -2 to 2. The curves labelled 0(A) and 1(B) represent cumulative probabilities calculated using the cumulative distribution function (CDF) of the normal distribution ($\Phi$). Curve 0(A) indicates the probability that $y$ belongs to the category "0 or A" (e.g., "non-white plumages or darker plumages"). Curve 1(B) indicates the probability that $y$ belongs to category "1 or B" (e.g., "white plumages or lighter plumages"). For binary data, the threshold is fixed at 0. The latent variable $l$ determines the value of $y$; if $l$ exceeds the threshold, $y = 1$, otherwise $y = 0$. In the logit model (a: left), the probability density plot shows the distribution of the latent variable (log odds), which follows a logistic distribution $\text{Logistic}(0, \pi^2/3)$. The majority of the data falls within the range between -6 to 6. The curves 0 (A) and 1 (B) represent cumulative probabilities calculated using the logistic cumulative distribution function (sigmoid function). The curve 0(A) indicates the probability that $y$ belongs to the category "0 or A" (e.g., "non-white plumages or black plumages"). Curve 1(B) indicates the probability that $y$ belongs to category "1 or B" (e.g., "white plumages"). Both models are capable of describing the binary data (0-1 data). However, binary data can also be treated as ordinal data, in which case an ordered multinomial model may be applied, or as nominal data, where an unordered multinomial model is suitable. (b) Ordered multinomial model. This model extends the binary framework to handle ordinal traits, such as plumage darkness, across multiple ordered categories. Here, examples with three categories (A $<$ B $<$ C) and four categories (A $<$ B $<$ C $<$ D) are illustrated. The left density plots show the distributions of the latent variable. The grey distribution represents a standard normal distribution $N(0,1)$, while the coloured distributions correspond to latent variable values associated with specific categories. The thresholds (c0, c1, and c2) divide the latent variable range, assigning observations to the appropriate categories. For example, in the 3-category case, latent variable $l$ exceeds $c_1$, placing it in category C. In the 4-category case, $l$ lies between $c_1$ and $c_2$, placing it in category C. The cumulative probability curves, calculated using the CDF of the normal distribution ($\Phi$), depict the probability of $y$ falling into each category based on the $l$. These probabilities are derived using threshold values and the $l$. (c) Unordered multinomial model. This model is suited for nominal traits, such as plumage colour, across multiple unordered categories. Examples with three categories (B, G, R) and 4 categories (B, G, R, W) are

presented. Note that many statistical applications select the first category (alphabetically) as the reference category; here, category B is chosen. The plots are structured similarly to the binary logit model, with relative probabilities computed for each category compared to the reference category. In the 3-category case, the probabilities for R and G are expressed relative to B. In the 4-category case, the probabilities for R, G, and W are expressed relative to B. The overall probability of each category is obtained by rescaling the relative probabilities. For example, the probability of category G in four unordered multinomial traits is calculated as, $P(\mathrm{G}) = \frac{\exp(l_{2i})}{1+\exp(l_{2i})+\exp(l_{3i})+\exp(l_{4i})}$, where $l_{2i}$, $l_{3i}$, and $l_{4i}$ represent the latent variables corresponding to the categories G, R, and W, respectively.

**Figure 3.** Conceptual summary of ordered multinomial and unordered multinomial regression models. The examples use body mass as a continuous explanatory variable and forest living status as a binary (discrete) explanatory variable for bird species. In (a) and (b), two horizontal lines mark the thresholds (c0 and c1) that divide categories A and B (threshold0) and categories B and C (threshold1). These categories, A, B, and C, correspond to the classification of plumage brightness, as seen in figure 2b. (a) illustrates that heavier bird species tend to have brighter plumages, suggesting a positive relationship between body mass and plumage colour intensity. This indicates that as body mass increases, so does the likelihood of a bird exhibiting a brighter plumage. In contrast, (b) shows that species inhabiting forested areas generally have duller plumages, which suggests a negative relationship between living environment and plumage brightness. The species living in forests are more likely to have less bright plumages compared to those in other habitats (not forest-living). In (c) and (d), the gradations in the illustrated probabilities of plumage colours. The marginal lines in these plots indicate a 50% probability, representing the point at which there is no clear trend in the plumage colours. The categories B, G, and R, which refer to different colour classifications, are aligned with figure 2c, where category B serves as the reference. (c) demonstrates that species with redder plumages tend to have a higher body mass than those with blue plumages. Conversely, species with green plumages are generally lighter in body mass than those with blue plumages. Finally, (d) illustrates the relationship between forest living and plumage colours, showing that forest species are more likely to exhibit green colours, not red colours.

**Figure 4.** Visualisation of used datasets and posterior distributions of model parameters from ordered and unordered multinomial models (`MCMCglmm` and `brms`) in the worked example section. (a) ordered multinomial model. We used migration data (3 ordered category traits) as a response variable and body mass and habitat density as explanatory variables for 136 species of birds of prey. Body mass was log-transformed before analysis. The posterior distribution results show fixed effects, phylogenetic random effects, and thresholds. The differences in threshold estimates and the presence or absence of intercept between `MCMCglmm` and brms arise from different parametrisations (*the true-intercept model* and *the zero-intercept model*; see **Box 2**). In `MCMCglmm`, the first threshold is 0 (sedentary to partially migratory), and the second threshold (partially migratory to migratory) is shown as threshold 1. In `brms`, thresholds are described as intercepts: threshold1 (intercept1: sedentary to partially migratory) and threshold2 (intercept2: partially migratory to migratory). (b) unordered multinomial model. we used the lifestyle data (3 unordered category traits) as a response variable and tail length and presence or absence of omnivores as explanatory variables of thrushes 173 species. The posterior distribution results represent point estimates and credible intervals for posterior distributions, categorised by fixed effects, phylogenetic random effects, and phylogenetic correlations. The unordered multinomial model in `MCMCglmm` shows results for phylogenetic

random effects and correlations after rescaling (refer to the main text and online tutorial for details). Regarding the graph of the phylogenetic random effect, please note that the x-axis scale differs between `MCMCglmm` and `brms`. In the online tutorial, we provide a clear explanation of the prior settings in the nominal model. You can also find the results of `MCMCglmm` using an uninformative prior. Thick horizontal lines represent 66% credible intervals, and thin horizontal lines illustrate 95% credible intervals in both model results. The points in the centre of each thick line indicate the mean estimates.

# 6 APPENDIX

The multinomial model can be also parametrised in an alternative way, by using so called contrast matrix $\Delta$ which defines the odds ratios by indicating a reference level within categories. In a general case with $n$+1 categories, with the first category being the reference, we have:

$$\Delta = \begin{bmatrix} -\mathbf{1}_n^{'} \\ \mathbf{I}_{n\times n} \end{bmatrix} \tag{47}$$

For the case of a trait with four categories, this would result in:

$$\Delta = \begin{bmatrix} -1 & -1 & -1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{48}$$

This matrix can then be used to project the latent scale estimates into a space of probabilities defined on an $n$-dimensional simplex (a tetrahedron in the 3-dimensional case of 4 categories).

$$\Delta\Delta^{'} = \begin{bmatrix} 3 & -1 & -1 & -1 \\ -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{bmatrix}, \tag{49}$$

$$(\Delta\Delta^{'})^{-1}\Delta = \begin{bmatrix} -\frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} \\ \frac{3}{4} & -\frac{1}{4} & -\frac{1}{4} \\ -\frac{1}{4} & \frac{3}{4} & -\frac{1}{4} \\ -\frac{1}{4} & -\frac{1}{4} & \frac{3}{4} \end{bmatrix} \tag{50}$$

Such a projection results in values that (exponentiated) are proportional to probabilities of observing each of the $n$ categories for a given row (case) of data. To get actual probabilities one would of course have to scale each resulting expectation by their sum (a sum-to-unity constraint):

$$\exp\left((\Delta\Delta^{'})^{-1}\Delta\mathbf{l_i}\right) \propto E(\mathbf{p_i}). \tag{51}$$

It is easy to verify that this formulation yields the same final expressions of expected probabilities as simple calculations based on the arithmetic of odds rations. The non-normalised predictions on log scale are as follows:

$$(\Delta\Delta^{'})^{-1}\Delta\mathbf{l_i} = \begin{bmatrix} \frac{-l_{2i}-l_{3i}-l_{4i}}{4} \\ \frac{3l_{2i}-l_{3i}-l_{4i}}{4} \\ \frac{-l_{2i}+3l_{3i}-l_{4i}}{4} \\ \frac{-l_{2i}-l_{3i}+3l_{4i}}{4} \end{bmatrix}. \tag{52}$$

The scaling term (sum of all exponentiated terms) is:

$$\begin{aligned} &\exp\left(\frac{-l_{2i}-l_{3i}-l_{4i}}{4}\right)+\exp\left(\frac{3l_{2i}-l_{3i}-l_{4i}}{4}\right)+\exp\left(\frac{-l_{2i}+3l_{3i}-l_{4i}}{4}\right)+ \\ &\exp\left(\frac{-l_{2i}-l_{3i}+3l_{4i}}{4}\right)=\left(\exp\left(l_{2i}\right)+\exp\left(l_{3i}\right)+\exp\left(l_{4i}\right)+1\right)\exp\left(\frac{-l_{2i}-l_{3i}-l_{4i}}{4}\right). \end{aligned} \tag{53}$$

Then we can verify, for category no. 1, that the final expected probability for the given case is:

$$\begin{aligned} p_{1i} &= \frac{\exp\left(\frac{-l_{2i}-l_{3i}-l_{4i}}{4}\right)}{\left(\exp\left(l_{2i}\right)+\exp\left(l_{3i}\right)+\exp\left(l_{4i}\right)+1\right)\exp\left(\frac{-l_{2i}-l_{3i}-l_{4i}}{4}\right)}= \\ &= \frac{1}{\exp\left(l_{2i}\right)+\exp\left(l_{3i}\right)+\exp\left(l_{4i}\right)+1}, \end{aligned} \tag{54}$$

which is exactly equivalent to the probability in equation 34. Similarly, for category no. 2 (and by extension for the remaining ones), we have:

$$\begin{aligned} p_{2i} &= \frac{\exp\left(\frac{3l_{2i}-l_{3i}-l_{4i}}{4}\right)}{\left(\exp\left(l_{2i}\right)+\exp\left(l_{3i}\right)+\exp\left(l_{4i}\right)+1\right)\exp\left(\frac{-l_{2i}-l_{3i}-l_{4i}}{4}\right)}= \\ &= \frac{\exp\left(l_{2i}\right)}{\exp\left(l_{2i}\right)+\exp\left(l_{3i}\right)+\exp\left(l_{4i}\right)+1}. \end{aligned} \tag{55}$$