Received: 23 April 2024 | Accepted: 3 April 2025

DOI: 10.1111/2041-210X.70044

Methods in Ecology and Evolution

RESEARCH ARTICLE

PhyBaSE: A Bayesian structural equation model approach to causal inference in phylogenetic comparative analyses

Achaz von Hardenberg[1,2] | Alejandro Gonzalez-Voyer[3]

[1]Conservation Biology Research Group, Department of Biological Sciences, University of Chester, Chester, UK

[2]Department of Earth and Environmental Sciences, University of Pavia, Pavia, Italy

[3]Instituto de Ecología, Universidad Nacional Autónoma de México, Ciudad de México, Mexico

Correspondence
Achaz von Hardenberg
Email: achaz.hardenberg@unipv.it

Alejandro Gonzalez-Voyer
Email: alejandro.gonzalez@iecologia.unam.mx

Funding information
Swedish Research Council, Grant/Award Number: 2013/5064; Programa de Intercambio Académico, Universidad Nacional Autonóma de Mexico; Santander Universities

Handling Editor: Daniele Silvestro

Abstract

1. One of the main limitations of phylogenetic comparative analyses is that associations between traits can only be interpreted as correlations. Here, we present a novel Bayesian structural equation model (PhyBaSE) which allows us to disentangle direct from indirect relationships among variables to propose potential causal hypotheses while accounting for phylogenetic non-independence.
2. Compared with the existing maximum-likelihood based approach, PhyBaSE models are more flexible, allowing the inclusion of trait and phylogenetic uncertainty, as well as non-continuous variables. To facilitate the application of the method, we provide worked examples, data and code.
3. We exemplify the method both with simulated as well as empirical data. Our analyses with simulated data indicate that PhyBaSE models have higher power than classic Phylogenetic Path Analysis to discriminate between competing models. As an example of PhyBaSE using empirical data, we revisit different hypotheses proposed to explain the relationship between relative brain size and group size in Bovids. Our results challenge the previously supported social brain hypothesis and provide support for an allometric effect of body size on social group size and an effect of brain size on life span, as predicted by the cognitive buffer hypothesis.
4. The flexibility of PhyBaSE models will allow researchers to explore more complex hypotheses on the evolution of behavioural, ecological and life history traits at a macroevolutionary level and how these are linked to anthropogenic drivers of biodiversity loss and extinction, taking full advantage of the increasing number of publicly available species-specific datasets.

KEYWORDS
Bayesian statistics, causal inference, path analysis, phylogenetic comparative methods, structural equation models

Achaz von Hardenberg and Alejandro Gonzalez-Voyer equal contribution.

1 | INTRODUCTION

> *"To find the causes for the existing characteristics and particularly adaptations of organisms is the main preoccupation of the evolutionary biologist."*
>
> - Ernst Mayr, 1961

Evolutionary biology, at its very roots, is concerned with finding the ultimate causes explaining 'Why' traits are in the form we can observe them now (Mayr, 1961). Controlled and randomized experiments (Fisher, 1935), the golden standard in causal inference, can be used to study microevolutionary processes, but their use is limited when the interest lies in explaining the macroevolutionary processes behind the variability in traits among species (Freckleton, 2009). Phylogenetic comparative methods (reviewed in e.g. Garamszegi, 2014; O'Meara, 2012; Revell & Harmon, 2022), and in particular, the Phylogenetic Generalized Least Squares approach (PGLS: Grafen, 1989; Martins & Hansen, 1997) allow researchers to investigate the evolutionary correlation between phenotypic traits and other life history, ecological, morphological, behavioural and physiological covariates, controlling for the non-independence of data points due to common ancestry (Felsenstein, 1985). While such phylogenetic linear models and extensions thereof have greatly advanced our understanding of macroevolutionary processes, they suffer from the same limitations of correlations and linear models, which, taken on their own, have predictive value and provide information about the association between traits but cannot enlighten us regarding the causality of the relationships, an important limitation for macroevolutionary analyses (Martins, 2000; von Hardenberg & Gonzalez-Voyer, 2013).

New structural equation modelling (SEM) approaches, originating from the Path Analysis method developed by evolutionary biologist Sewall Wright (1921) and based on Graph Theory (Pearl, 2009) provide formal methods to specify and compare complex models of the relationships among ecological variables, to disentangle direct and indirect causal effects when only observational data are available (Laland et al., 2011). While SEM and Path Analysis methods have been part of the ecologist's toolbox for at least 25 years (Shipley, 1997), their use within a phylogenetic comparative approach in macroevolution was, until recently, hampered by the problem of explicitly accounting for the underlying covariance in the residuals due to phylogenetic non-independence, which in a path analysis framework can lead to high inflation of type I error (Gonzalez-Voyer & von Hardenberg, 2014; von Hardenberg & Gonzalez-Voyer, 2013).

These limitations have been overcome by Phylogenetic Path Analysis (PPA, von Hardenberg & Gonzalez-Voyer, 2013) which combines PGLS with the d-sep method for Path Analysis developed by Shipley (2016), introducing causal inference in the modern phylogenetic comparative methods toolbox. Previous work (Gonzalez-Voyer & von Hardenberg, 2014; von Hardenberg & Gonzalez-Voyer, 2013) has shown how this approach allows evolutionary biologists to formulate causal models of hypothesized direct and indirect relationships between life history, ecological and morphological traits, translating them to alternative probabilistic models, testing the implied conditional independencies and comparing them with well-established information theoretic approaches. Evolutionary biologists have employed PPA to tackle a host of different evolutionary questions that required disentangling direct from indirect relationships (see e.g. Dale et al., 2015; Gonzalez-Voyer et al., 2016; Jiménez-Ortega et al., 2020; Krüger et al., 2014; Santini et al., 2019). Although the use of PPA has broadened the type of questions that can be addressed in comparative biology, an endeavour facilitated by the development of the phylopath R package that automates part of the work (van der Bijl, 2018), we are aware that the method has limitations, such as the inability to account for uncertainty in the phylogeny and within-species variation, the failure to jointly impute missing values (which can greatly reduce the sample size), or not being able to handle latent variables, among others. Previous work has shown that accounting for phylogenetic uncertainty results in more precise estimates of regression model parameters than including a single consensus tree and enables more precise estimates of credibility intervals (de Villemereuil et al., 2012). Phylogenetic uncertainty can be included in the form of a sample from the posterior distribution of a Bayesian inference, or from alternative phylogenetic reconstructions. For example, Krüger et al. (2014) controlled for phylogenetic uncertainty in their analysis of sexual size dimorphism in pinnipeds by repeating all analyses using different phylogenetic hypotheses. However, an advantage of doing so in a Bayesian framework is that the error around estimates of model parameters would include that due to phylogenetic uncertainty. Uncertainty in species trait values can also influence estimates of phylogenetic co-variance, which will in turn influence model parameter estimates (de Villemereuil et al., 2012).

Here, we present a novel Bayesian approach to causal inference in phylogenetic comparative analyses (Phylogenetic Bayesian Structural Equation model: PhyBaSE) and show how it relaxes some of the assumptions and overcomes some of the limitations of the maximum likelihood-based PPA method (von Hardenberg & Gonzalez-Voyer, 2013). An alternative phylogenetic structural equation modelling (PSEM) approach has been recently proposed by Thorston et al. (2023) and Thorston and van der Bijl (2023). However, while these previous PSEMs used causal assumptions as represented in a graphical model and based upon scientific judgement (Bollen & Pearl, 2013), they did not validate these assumptions by checking for conditional independence as we propose here for PhyBaSE models as well as previously for PPA (von Hardenberg & Gonzalez-Voyer, 2013). We point readers to the Methods section where we provide more details on the importance of doing so.

Firstly, we generate simulated data under a given model and use the simulated data to validate the Bayesian implementation of the phylogenetic structural equation model and show that PhyBaSE can accurately identify the model used to generate the data among a set of candidate models. We compare the results obtained from PhyBaSE with those from PPA and show that PhyBaSE has more power to discriminate between alternative models than PPA. Using

the simulated data, we present an example of an analysis incorporating measurement error either as multiple observations per species or including an estimate of the standard error around species mean values. Finally, as an empirical example of the applicability of our PhyBaSE method, we revisit some hypotheses proposed to explain interspecific differences in relative brain size using ungulates as a model system. We analysed the relationships between brain size, body size, longevity and sociality in ungulates, translating proposed hypotheses into formal path models explicitly incorporating phylogenetic uncertainty. We also use this example to show how we can include binomial variables in the model.

2 | METHODS

2.1 | Simulated data

As a simple dataset on which to present the steps involved in the implementation of a PhyBaSE model we use simulated life history data on the fictitious mammal order of the Rhinogrades (Stümpke & Steiner, 1967). Gonzalez-Voyer and von Hardenberg (2014) used the same dataset to illustrate the implementation of PPA and we refer readers to that source for more details on this fictitious group of mammals. The simulated dataset aimed to analyse the factors influencing the evolution of range size in this peculiar group of mammals, in which the nasal appendage evolved to be used for locomotion. Hypotheses proposed an allometric relationship between range size and body mass; in addition, species with larger range sizes were found to have larger litter sizes due to higher resource availability, although whether this relationship is causal or correlational remains unclear. The relationship between range size and nose length also remains unclear, as nose length could influence range size due to the effect on long-distance displacements, although a reverse relationship has also been proposed, as dispersal distance is related to nose length. The traits simulated for this example were body mass = BM, litter size = LS, nose length = NL, dispersal distance = DD and range size = RS.

Briefly, we simulated a phylogenetic tree under a pure-birth model with 100 species using the *ape* package in R (Paradis & Schliep, 2019). We then simulated the evolution of five traits on said tree, previously transformed using $\lambda = 0.8$, with varying degrees of correlation among traits given by a specific causal model. Variables that were directly linked in the pre-specified path model presented correlations of 0.5, while variables with indirect links presented correlations with decreasing proportionality based on the number of variables that separated them, with correlations decreasing by half for each variable in the indirect path linking the two traits (Gonzalez-Voyer & von Hardenberg, 2014). All the data files and R scripts used to implement the PPA models in Gonzalez-Voyer and von Hardenberg (2014), including a tutorial on how to generate the simulated Rhinograd tree and data, are available online on Zenodo (Gonzalez-Voyer & von Hardenberg, 2025) making the comparison between the PhyBaSE and the PPA approaches straightforward. We compared the fit of nine alternative models to the simulated data. For details on the tested models, we refer readers to Figure 8.7 in Gonzalez-Voyer & von Hardenberg (2014).

2.2 | Bovids trait data

For the empirical example revisiting hypotheses proposed to explain the relationship between brain size and longevity in Bovids, we collated phenotypic trait data from different sources. We obtained data on length of the gestation period (in days), adult body weight (in grams) and maximum longevity (in years) from the AnAge database available online (De Magalhaes & Costa, 2009). We collected data on brain size (brain volume in mL) from Finarelli (2011) including both estimates of brain size from endocranial volumes as well as estimates derived from external measurements of the cranium. Note that Finarelli (2011) showed the external measures of the cranium provide a highly accurate estimate of endocranial volume ($R^2 = 0.949$) and by combining the two types of measures, we were able to increase the sample size. Prior to analyses, we transformed brain volume to brain mass by multiplying by the density of fresh brain tissue (1.036 g/mL; see Iwaniuk & Nelson, 2002). Finally, we obtained estimates of group size from Pérez-Barbería and Gordon (2005). All data were log-transformed prior to analyses to fulfil assumptions of the model of phenotypic evolution.

2.3 | Bovids phylogeny

We downloaded phylogenetic trees of Certiodactyla from the 10kTrees web site (https://10ktrees.nunn-lab.org Version 1; Arnold et al., 2010). We downloaded 500 trees from the posterior distribution provided on the website, which are sampled in proportion to their likelihood using Markov chain Monte Carlo. We sampled 100 phylogenies from the downloaded trees to reduce computation time for the models. The trees were pruned to include only species for which phenotypic data were available using the *drop.tip()* function in the *ape* package (Paradis & Schliep, 2019) in R (R Core Team, 2015).

2.4 | Implementation of PhyBaSE models

We show how a Phylogenetic Bayesian Structural Equation (PhyBaSE) model can be implemented using JAGS (Plummer, 2003). We use JAGS because of its seamless integration within the R language (R Core Team, 2015) through the *rjags* (Plummer et al., 2016) and *R2jags* (Su et al., 2015) packages, which greatly simplifies the modelling workflow. PPA is based on the sequential estimation of separate phylogenetic regressions, and the fit of the full model is tested combining the *p*-values of the conditional independencies implied in the causal model using Fisher's *C* statistics (Shipley, 2016). Instead, in a PhyBaSE model, a set of structural equations

describing the hypothesized direct relationships among the variables is defined, and the parameters of these structural equations are jointly estimated in the same inferential procedure using a Gibbs sampler in MCMC. The advantage lies in the fact that all parameters are estimated in the same model, rather than separately as in PPA, and thus the estimated parameters in one structural equation are informed by the estimates of the parameters in the other structural equations composing the hypothesized causal model. In synthesis, the steps involved in fitting a PhyBaSE model (see also Figures S1 and S4 for a graphical summary) are analogous to those for PPA (Gonzalez-Voyer & von Hardenberg, 2014; von Hardenberg & Gonzalez-Voyer, 2013) using Bayesian graphical models implemented with the JAGS language: (1) Define the causal model by means of a Directed Acyclic Graph (DAG). While it has been shown that cyclic causal models still imply conditional independencies under some conditions (Spirtes, 1995) we recommend using PhyBaSE only with causal models which can be represented as a DAG. Indeed, cyclic causality (such as feedback loops) is harder to reconcile in an evolutionary context in which traits, and thus the relationships among traits, evolve linearly over time. In a comparative framework, it is difficult (if not virtually impossible) to collect the necessary data to test cyclic causality as it implies datapoints obtained at different points in time for a large number of species; (2) Find the minimum set of conditional independencies that must be fulfilled for the model to be considered a potential causal model, which can be calculated by hand as shown in Shipley (2016) and in previous works presenting phylogenetic path analysis (Gonzalez-Voyer & von Hardenberg, 2014; von Hardenberg & Gonzalez-Voyer, 2013). Alternatively, automated functions available in the *phylopath* R package (van der Bijl, 2018) can be used to identify the conditional independencies; (3) Verify the conditional independencies implied by the model, checking that 0 is included within the Bayesian 95% credible intervals of all estimates of the slopes of the linear relationships among the variables assumed to be conditionally independent by the causal model; (4) If all the conditional independencies are fulfilled, carry out the inference following the model using MCMC; (5) Compare the relative fit of the causal models complying with the assumptions in steps 2. and 3. using Bayesian Model selection methods such as the Deviance Information Criterion (DIC, Spiegelhalter et al., 2002) or the Watanabe-Aikaike Information Criterion (WAIC, Watanabe & Opper, 2010; see a graphical representation of the workflow in Figure S1). Both DIC and WAIC can be considered measures of model fit or adequacy, penalized for model complexity, analogous to AIC (Akaike, 1973) commonly used for maximum likelihood models. While DIC has seen a wide popularity as a measure of model selection in Bayesian analysis, it has been criticized for not being a fully Bayesian measure of fit, for its lack of consistency and for its poor theoretical justification (Spiegelhalter et al., 2014). A more robust, fully Bayesian criterion for model selection appears to be WAIC (Watanabe & Opper, 2010). In the provided code we show how this criterion can be extracted from the models along with DIC. While in this article, we chose to use only WAIC and DIC for model selection, because of their ease of implementation and interpretation, other Bayesian model selection methods, such as Bayes Factors through bridge sampling, could be used. We direct readers to Hooten and Hobbs (2015) for a thorough discussion on the advantages and disadvantages of various Bayesian model selection approaches.

We advocate for proposed models to reflect researchers' uncertainty regarding relationships among traits; in other words, if uncertain about given relationships among traits, test them within the model comparison step, as path analysis is a hypothesis-testing, not hypothesis-generating approach (Shipley, 2016).

We would like to stress that ensuring that the minimum set of conditional independencies of the proposed causal models is met is a crucial step, as a given model can only be considered a potential causal explanation if and only if the minimum set of conditional independencies is found to be fulfilled (Shipley, 2016). Researchers could be misled if model comparison is done prior to selecting only potential causal models. We cannot stress enough the need to carefully consider the set of models to be compared, regarding each as a graphical representation of hypothesized relationships among variables, ensuring each causal link reflects a biologically plausible relationship, and avoiding any shortcuts or temptation to test a large number of possible (but potentially biologically implausible) models. If given careful consideration, and basing models on theory, rejected models can be as informative as those that are accepted as potential causal explanations.

The Bayesian structural equation model approach we use follows Lee (2007) and Lee and Song (2012). To ensure conditional conjugacy and thus facilitate the Gibbs sampling process, we used a normal prior for the intercepts and slopes (with a mean of 0 and a precision [1/standard deviation] of 1.0^{-6}), and a gamma prior (with shape and scale values of 1) for the parameter tau (the precision of the multivariate normal likelihoods), which is equivalent to an inverse gamma prior on the variance (Gelman et al., 2013). To account for phylogenetic non-independence, we include an estimate of the evolutionary parameter λ (Freckleton et al., 2002; Pagel, 1999) using a uniform prior (range 0-1) following de Villemereuil et al. (2012). In S3 (Appendix 1), we provide a step-by-step tutorial, as well as all the code, for the implementation of a simple causal model using the Rinogradentia data.

2.5 | Bayesian coverage analysis

To validate the robustness of PhyBaSE, we conducted a Bayesian coverage analysis following the guidelines provided by Mendes et al. (2024). We simulated 100 datasets using the best fitting model for the Rhinograd data (model 8) in JAGS. The alpha and beta parameters were simulated from Normal priors centred at 0 and 0.5, respectively, and variances of 1. For the lambda parameters, ranging from 0 to 1, we used a Normal prior with mean 0.8 and variance = 0.04, truncated to the interval [0, 1]. The variance of 0.04 was chosen to ensure the distribution was sufficiently

concentrated within the [0, 1] interval, minimizing truncation effects while still allowing for reasonable uncertainty around the mean. For each simulated dataset, we fitted our Bayesian model using the same priors and calculated the proportion of times the parameter values fell within the 95% HPD intervals of the posterior estimates.

2.6 | Variability in traits

When repeated measures for traits within the same species are available, a PhyBaSE measurement error model accounting for intraspecific variability can easily be implemented following the example code presented in de Villemereuil et al. (2012) to which we refer readers for details. Previous work has shown how to incorporate within-species variation in comparative analyses and the importance of doing so (e.g. de Villemereuil et al., 2012; Ives et al., 2007; Silvestro et al., 2015). In S3 (Appendix 2), we provide a working example within the PhyBaSE context based on simulated Rhinograd data.

The model proposed by de Villemereuil et al. (2012) assumes multiple observations for each trait for each species are available, which are rather uncommon in currently available species-level databases commonly used for phylogenetic comparative analyses (e.g. PanTHERIA; Jones et al., 2009). Although most of these databases generally include only point estimates of the mean, estimating a measure of variability around these point estimates from the original literature may be a more achievable task. When only average values for each trait of each species are available (μ_x), but standard error (se_x) is obtainable from the original sources, assuming that the average values come from a normal distribution (following the Central Limit Theorem) each value of a trait x for individual i (x_i) can be sampled from a normally distributed informative prior with mean $= \mu_x$ and se $= se_x$.

2.7 | Including phylogenetic uncertainty

To include uncertainty in the phylogenetic tree (e.g. multiple phylogenies sampled from a Bayesian tree inference; Arnold et al., 2010), we use the approach proposed by de Villemereuil et al. (2012). We exemplify the approach in our models using the Bovids data. All necessary code to reproduce the ananlyses is provided on Zenodo (von Hardenberg & Gonzalez-Voyer, 2025)

2.8 | Alternative hypotheses of the evolution of Bovid brain size

As a case study using empirical data, we revisit alternative hypotheses on the evolution of brain size in Bovids. The first hypothesis, known as the *cognitive buffer hypothesis* (reviewed in Sol, 2009), posits that larger brains favour longer lifespans (Jiménez-Ortega et al., 2020), because the greater behavioural flexibility associated with larger brains allows to reduce extrinsic mortality. Thus, an explanation for the correlation between brain size and group living, based on the cognitive buffer hypothesis, is that larger brains favour longer lifespans and enable species to live in groups (depicted in model CB1 in Figure 1). Longer lifespan could be influenced by brain size while group size is influenced by body size, if small species gain a bigger advantage through reduced predation risk (depicted in model CB2); alternatively, longer lifespans could enable species to develop larger brains (depicted in model CB3). An alternative hypothesis is that group living poses high cognitive challenges for individuals and thus favours increased relative brain size in social species. This is an extension of the *social brain hypothesis* originally proposed to explain brain size differences in primates (Dunbar & Shultz, 2007). A previous study found support for the social brain hypothesis in the form of a positive correlation between group size and brain size, controlling for body size (Pérez-Barbería & Gordon, 2005). The directionality of the causal relationship is unclear as larger brains could enable species to live in groups, or alternatively, larger group sizes could select for larger brain sizes, alternatives depicted in models SB1 and SB2 (Figure 1). The aforementioned models also propose that group size influences longevity, as a previous study found that lifespan increases with group size in both sexes (Bro-Jørgensen, 2012). Finally, we must also consider the possibility that the relationships are driven by allometry; for example, body size may directly influence sociality, as mentioned previously, because smaller species benefit more from reduced predation pressure, and that larger species generally live longer lives (model LHA1, Figure 1). Finally, longer lifespan could be a result of a slower life history, as previous work has found that larger brain size is associated with slower pace of life in mammals (Barton & Capellini, 2011; Gonzalez-Voyer et al., 2016), and group size could be influenced by allometry (model LHA2, Figure 1). The only means by which these three alternative hypotheses to explain the correlation between brain size and group living can be contrasted is by means of path analysis allowing to disentangle direct from indirect relationships. We note this is an example of the use of PhyBaSE with non-simulated data and not meant as an exhaustive analysis of models to explain possible causal relationships among the traits. All variables were standardized before modelling to facilitate model convergence and to obtain standardized path coefficients.

2.9 | Inclusion of binomial data as a dependent variable

We show how to include binomial data as a dependent variable, fitting the model with the highest support on Bovids considering Gregariousness as a binomial trait (1 = Gregarious, 0 = Non gregarious) rather than as a continuous variable. For the sake of this demonstration, we arbitrarily considered a species as gregarious if group

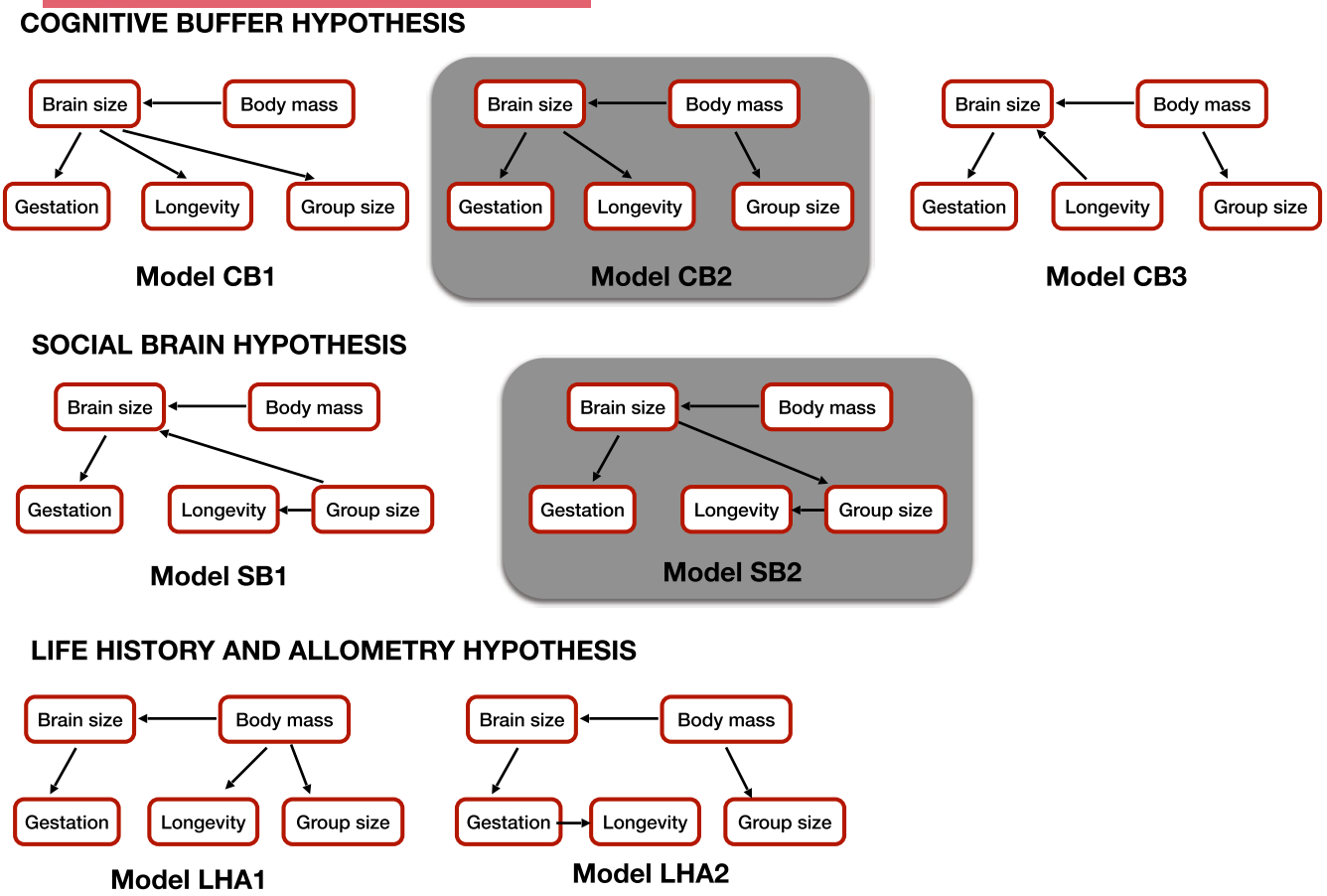

FIGURE 1 Graphical representation of the directed acyclic graphs (DAG) representing the three alternative hypotheses proposed to explain the relationship between body mass, brain size, life history (using gestation length as a proxy), longevity and group size. The hypotheses are represented by a number of alternative DAGs which represent proposed alternatives. Names of the DAGs follow the names of the JAGS models which test them (see Methods for details).

size was ≥ 6 and not gregarious if < 6. The model is based on the logit-linked binomial GLMM proposed by Ives and Garland (2014, 2010). The response variable is modelled on a logit scale, with an error term ε that follows a normal distribution and contains the phylogenetic information (Ives & Garland, 2014). The co-variance is modelled as following a variant of Brownian motion with the λ parameter multiplying the off-diagonal of the variance–covariance matrix (Freckleton et al., 2002). Code is available on Zenodo (von Hardenberg & Gonzalez-Voyer, 2025)

3 | RESULTS

3.1 | Comparison of PhyBaSE models with corresponding PPA models

The comparison between model selection using classical PPA models (based on the CICc Information Criterion; von Hardenberg & Gonzalez-Voyer, 2013) on the Rhinograd data (Gonzalez-Voyer & von Hardenberg, 2014) and model selection using the corresponding PhyBaSE models (based on WAIC and DIC information criteria of the fitted model, rather than the conditional independencies) is presented in Table 1. It is worth noting that as for PPA, the PhyBaSE approach was able to identify the 'correct' model (model 8, under which the data was generated). In addition, with PhyBaSE, using WAIC or DIC as a model selection criterion, model 8 is unambiguously identified as the best fitting model compared with all the others in the set (ΔWAIC = 367.6, ΔDIC = 359.8 to the next model, see Table 1). Model selection using CIC_c in PPA was instead less clear-cut in the discrimination between models 8, 6 and 4, all being within $\Delta CIC_c < 2$ (but see the discussion in Gonzalez-Voyer & von Hardenberg, 2014 about why models 4 and 6 might not be as supported and competitive as the best fitting model 8). Median posterior estimates of the regression coefficients (slopes) between predictor and response variables (DAG nodes connected by an edge) were similar to the corresponding estimates from PPA (Figure 2). In Table 1, we also provide the runtimes of all tested models when run on an Apple Macbook with M2 processor and 24GB RAM. In S4 (Appendix 3), we provide a comparison of model fitting and runtimes for Model 8 using PhyBase, *phylopath* (van der Bijl, 2018), *phylosem* (Thorston & van der Bijl, 2023) and *brms* (Bürkner, 2017).

TABLE 1 Comparison of model fit for the path models adjusted to simulated data using PPA, where model fit is based on the C-statistic, and the same models adjusted using PhyBaSE.

Model	CICc	ΔCICc	*q*	WAIC	ΔWAIC	pWAIC	DIC	ΔDIC	pD	Runtime (min)
8	27.7	0.0	9	881.8	0.0	6.3	887.5	0.0	12.1	6.221
4	29.1	1.4	10	1239.4	357.6	8.0	1248.5	361.0	17.0	21.498
6	28.9	1.2	10	1240.1	358.3	8.2	1248.3	360.8	16.2	8.306
7	31.0	3.3	11	1240.1	358.3	8.2	1250.0	362.5	17.0	8.405
9	29.8	2.1	10	1240.2	358.4	8.2	1247.7	360.2	15.8	21.513
5	30.3	2.6	11	1240.7	358.9	8.7	1249.9	362.4	17.9	8.423
3	49.0	21.3	9	1255.3	373.5	7.5	1262.9	375.4	14.9	21.383
1	83.8	56.1	9	1292.9	411.1	7.5	1299.9	412.2	14.7	21.488
2	85.2	57.5	10	1294.0	412.2	8.0	1301.9	414.4	15.8	21.505

Note: Shown are model numbers, the number of estimated parameters (*q*), the CIC_c and ΔCIC_c scores for PPA models, and for PhyBaSE models, the WAIC, ΔWAIC, as well as DIC and ΔDIC (for comparison). pWAIC and pD are the effective number of parameters for WAIC and DIC respectively. Runtime (in minutes) of the Bayesian models was calculated running the models on a MacBook Air with Apple M2 2022 processor and 24 GB memory.

3.2 | Bayesian coverage analysis

All the simulated alpha, beta and lambda parameters showed close to 95% coverage confirming the robustness of our approach. The coverage percentages for the parameters are as follows: alpha parameters (Intercepts): DD = 95%, LS = 93%, NL = 98%; Beta Parameters (slopes): BM = 96%, BM2 = 97%, NL = 97%, RS = 96%; Lambda Parameters: DD = 95%, LS = 93%, NL = 96%. We present posterior means of prior parameters plotted against their simulated values in Figure S2.

3.3 | Accounting for intraspecific variability in traits

Our models including individual variability in traits performed well when tested on simulated data both when repeated measures are available (Table 2) as well as when only mean values and standard errors are available (Table 3). Note that the latter model performed in a comparable way to the model for which individual repeated values were available, because in this example for simplicity, we calculated the standard error for each trait from the sample with multiple observations per species.

3.4 | Brain size, group size and longevity in Bovids revisited

Only two of the seven proposed models fulfilled all conditional independencies (Models CB2 and LHA1). The two models appear to be fundamentally equivalent after model selection.

Model CB2 (ΔWAIC = 0.0, ΔDIC = 0.0) suggests that brain size influences the length of gestation and maximum longevity, while body size influences group size. Model LHA1 (ΔWAIC = 2.0, ΔDIC = 1.9) is very similar to the previous one, except that here it is body size which influences longevity as well as group size. Table 4 shows the median standardized path coefficient estimates with 95% credibility intervals for the marginally better fitting model (CB2).

3.5 | Inclusion of binomial data as dependent variable

When we repeated the most supported model including gregariousness as a binary response (to exemplify how binary traits can be included in PhyBase) we found very similar results, as expected. The model shows adequate convergence with high effective sample size and r-hat values close to 1.0 (see Table 5). As for the model with group size, the model with gregariousness as a binary variable shows a positive association between body mass and gregariousness, with a 95% credibility interval that does not span zero, indicating that the slope (expressed as log-odds) is different from zero.

4 | DISCUSSION

We present a novel implementation of Bayesian structural equation models (SEM) which explicitly accounts for phylogenetic non-independence and can accommodate estimates of within-species variation in trait values, an extension of previously developed phylogenetic path analyses (von Hardenberg & Gonzalez-Voyer, 2013). An advantage of Bayesian SEM over maximum likelihood approaches such as those implemented in AMOS (Collier, 2020) or Lavaan (Rosseel, 2012), besides the possibility of accounting for phylogenetic non-independence as shown here, is that the implied conditional independencies of each SEM model can be easily tested, allowing for a more consistent causal interpretation of the results. In the examples we presented, all models were run including an estimate of the λ parameter; thus, the implied model of evolution was a variant of the

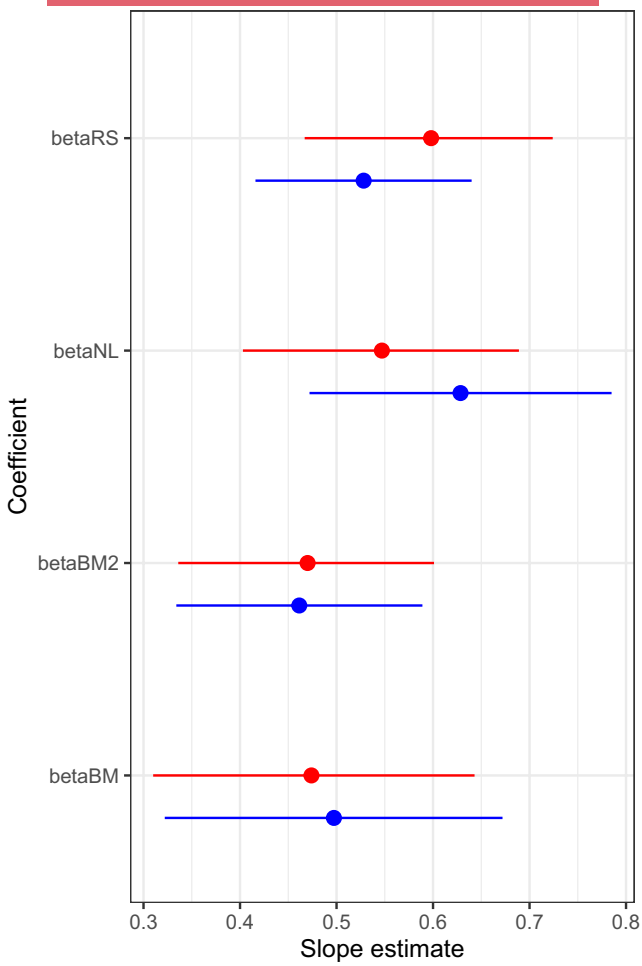

FIGURE 2 Comparison between median estimated path coefficients for the Rhinograds model 8 estimated with PPA (blue) and with PhyBaSE (red). Intervals around point estimates are confidence intervals and credibility intervals, respectively. Estimated path coefficients (beta) correspond to the following causal paths in the model: BetaBM = BM -> BR; betaBM2 = BM -> NL; betaNL = NL -> DD; betaRS = RS -> NL where BM = Body Mass; NL = Nose Length; RS = Range Size; DD = Dispersal Distance; LS = Litter Size.

widely used Brownian motion model (Felsenstein, 1985; Freckleton et al., 2002), although other models of phenotypic evolution can be implemented. We refer readers to recent discussions regarding the use of different models of trait evolution (e.g. Cooper et al., 2015; Freckleton et al., 2011; Grabowski et al., 2023) and highlight the need for careful consideration about the biological implications of the implemented models of trait evolution; for example, combining different models within a given analysis is nonsensical. Furthermore, the λ parameter was estimated separately for each causal link, and therefore allowed to differ between causal links. We prefer to run analyses with λ parameters estimated for each causal link because we think this allows for more precise adjustments for phylogenetic non-independence of residuals and therefore more precise parameter estimates (see Revell, 2010). However, if preferred, models can also be run using a single multivariate estimate of λ. Model fit can be assessed using the deviance information criterion (DIC) or a Bayesian variant of the Akaike Information Criterion (WAIC; Hooten & Hobbs, 2015). These approaches allow for model comparison and model averaging approaches to be implemented.

4.1 | Comparison of PhyBaSE models with corresponding PPA models

We revisited examples based on simulated data developed to present the previous phylogenetic path analysis method (PPA, von Hardenberg & Gonzalez-Voyer, 2013). PhyBaSE can find the model that generated the data, as was PPA; however, our results show that PhyBaSE has higher power to discriminate between the best model and competing causal models, with some confidence.

An important difference with PPA (von Hardenberg & Gonzalez-Voyer, 2013) is that model fit in PhyBaSE is assessed based on the posterior distribution of the proposed model generating the observed data, while in PPA model fit is based on the C-statistic and therefore the degree to which conditional independencies are met. This difference seems to confer greater power to PhyBaSE to

Coef.	2.5%	25%	Median	75%	97.5%	Rhat	n.eff
betaBM	0.401	0.448	0.473	0.497	0.544	1.001	2900
betaBM2	0.381	0.424	0.446	0.469	0.518	1.001	3000
betaNL	0.552	0.601	0.624	0.647	0.692	1.001	3000
betaRS	0.570	0.614	0.638	0.662	0.708	1.002	1000
lambdaBM	0.021	0.251	0.508	0.744	0.973	1.004	1300
lambdaDD	0.023	0.071	0.101	0.132	0.195	1.001	3000
lambdaLS	0.210	0.260	0.291	0.323	0.380	1.001	3000
lambdaNL	0.357	0.425	0.461	0.499	0.572	1.001	2300
lambdaRS	0.027	0.252	0.514	0.768	0.979	1.001	3000

Note: Estimated path coefficients (beta) correspond to the following causal paths in the model: BetaBM = BM -> BR; betaBM2 = BM -> NL; betaNL = NL -> DD; betaRS = RS -> NL where BM = Body Mass; NL = Nose Length; RS = Range Size; DD = Dispersal Distance; LS = Litter Size. Rhat is the potential scale reduction factor which is = 1.000 at convergence. N.eff = effective sample size.

TABLE 2 Median estimated path coefficients, lambda values and corresponding credibility intervals for the Rhinograds Model 8 including simulated variability in the traits (10 simulated measures for each trait for each species).

TABLE 3 Median estimated path coefficients, lambda values and corresponding credibility intervals for the Rhinograds Model 8 when variability in the traits is available as average and standard error.

Coef.	2.5%	25%	Median	75%	97.5%	Rhat	n.eff
betaBM	0.296	0.401	0.457	0.513	0.622	1.001	3000
betaBM2	0.336	0.424	0.468	0.510	0.597	1.001	3000
betaNL	0.389	0.493	0.546	0.600	0.697	1.004	590
betaRS	0.476	0.559	0.601	0.643	0.724	1.001	3000
lambdaBM	0.025	0.230	0.484	0.741	0.969	1.006	390
lambdaDD	0.104	0.299	0.416	0.528	0.707	1.001	3000
lambdaLS	0.420	0.607	0.696	0.767	0.871	1.001	3000
lambdaNL	0.552	0.707	0.773	0.831	0.916	1.001	2600
lambdaRS	0.031	0.258	0.504	0.749	0.971	1.002	3000

Note: Estimated path coefficients (beta) correspond to the following causal paths in the model: BetaBM = BM -> BR; betaBM2 = BM -> NL; betaNL = NL -> DD; betaRS = RS -> NL where BM = Body Mass; NL = Nose Length; RS = Range Size; DD = Dispersal Distance; LS = Litter Size. Rhat is the potential scale reduction factor which is = 1.000 at convergence. N.eff = effective sample size.

TABLE 4 Median estimated standardized path coefficients, lambda values and corresponding credibility intervals for the best fitting Bovids model (CB2).

Coef.	2.5%	25%	Median	75%	97.5%	Rhat	N.eff
betaBM	0.799	0.876	0.916	0.954	1.024	1.001	6000
betaBM2	0.414	0.580	0.666	0.752	0.923	1.001	6000
betaBR	0.285	0.436	0.512	0.588	0.727	1.001	6000
betaBR2	0.343	0.506	0.590	0.675	0.846	1.001	6000
lambdaBR	0.153	0.435	0.575	0.695	0.868	1.002	1800
lambdaG	0.463	0.701	0.807	0.900	0.987	1.002	6000
lambdaL	0.058	0.377	0.579	0.736	0.922	1.001	6000
lambdaS	0.227	0.543	0.715	0.853	0.979	1.001	6000

Note: Estimated path coefficients (beta) correspond to the following causal paths in the model: BetaBM = BM -> BR; betaBM2 = BM -> S; betaBR = BR -> G; betaBR2 = BR -> L where BM = Body Mass; BR = Brain Size; S = Group Size; G = Gestation length. Rhat is the potential scale reduction factor which is = 1.000 at convergence. N.eff = effective sample size.

TABLE 5 Median estimated path coefficients, lambda values and corresponding credibility intervals for the best fitting Bovids model (CB2) with Group Size (S) transformed into the binary variable Gregariousness (GR, with species with S > 6 coded as 1 = Gregarious and species with S < 6 as 0 = non-Gregarious).

Coef.	2.5%	25%	Median	75%	97.5%	Rhat	n.eff
betaBM	0.796	0.875	0.915	0.954	1.025	1.001	6000
betaBM2	1.198	1.856	2.267	2.750	3.914	1.002	1300
betaBR	0.287	0.435	0.514	0.590	0.743	1.002	3300
betaBR2	0.332	0.505	0.591	0.677	0.834	1.001	3300
lambdaBR	0.146	0.436	0.579	0.699	0.876	1.001	6000
lambdaG	0.461	0.699	0.814	0.901	0.988	1.001	3100
lambdaL	0.060	0.379	0.576	0.733	0.925	1.002	4100
lambdaS	0.115	0.508	0.722	0.878	0.988	1.001	6000

Note: Estimated path coefficients (beta) correspond to the following causal paths in the model: BetaBM = BM -> BR; betaBM2 = BM -> GR; betaBR = BR -> G; betaBR2 = BR -> L where BM = Body Mass; BR = Brain Size; GR = Gregariousness; G = Gestation length. Rhat is the potential scale reduction factor which is = 1.000 at convergence. N.eff = effective sample size. The Beta coefficients for Gregariousness are expressed as log-odds.

discriminate between alternative causal models, that is, those where conditional independencies are met. We validated our best fitting model (model 8) with a Bayesian coverage analysis and our results indicate strong coverage for the alpha and beta parameters, as well as the λ parameter. This suggests that PhyBase performs well for estimating alpha- and beta-parameters, as well as the λ parameter.

4.2 | Accounting for intraspecific variability in traits

PhyBaSE models can also easily incorporate within-species variation, either by means of repeated measures or by means of the standard error for the species mean value for a given trait. The aim

in the examples presented here was to show an implementation of models accounting for within-species variation, rather than test the effects of including it on parameter estimates (for the later we refer readers to de Villemereuil et al., 2012). If multiple values for a given trait for all species are available, these are included in analyses constructing a hierarchical model, where repeated measures are nested within each species (as in de Villemereuil et al., 2012). However, such repeated measures are rarely available for many, let alone all species, and it might be more common that an estimate of the standard error around the mean species value is available. In such cases, we show how this estimate of within-species variation can be used to inform the prior on the variable, by incorporating explicitly the estimated within-species variation into the model.

4.3 | Brain size, group size and longevity in Bovids revisited

As an empirical example of the application of our PhyBaSE method, we revisited the relationship between brain size and group size in Bovids, including phylogenetic uncertainty in the analysis. We proposed seven different path models to contrast the three hypotheses proposed to explain relationships among these two traits of interest: the cognitive buffer hypothesis, the social brain hypothesis, and the life-history and allometry hypothesis. The results, although tentative as we did not do an exhaustive analysis of possible models, nonetheless allow us to draw some conclusions regarding the contrasted hypotheses. Firstly, our path models allow us to call into question the social brain hypothesis as a potential explanation for the correlation between brain size and group size, as proposed by Pérez-Barbería and Gordon (2005), since both models based on the social brain hypothesis, including a causal link between brain size and group size, did not meet the required conditional independencies to be considered potential causal explanations. The two models for which conditional independencies were met (CB2 and LHA1) included instead a direct causal link from body size to group size but no causal link between brain size and group size. The difference in WAIC between our two accepted models (CB2 and LHA1) is too small (ΔWAIC = 2) to discriminate between them; thus, it is unclear whether longevity is influenced by brain size, supporting the cognitive buffer hypothesis (Allman et al., 1993; Sol, 2009), or rather whether longevity is influenced by body size, as larger species generally live longer lives (*Life History and Allometry Hypothesis*), but given the hypotheses are not mutually exclusive, this is not necessarily surprising and could point to a synergistic effect of both traits.

4.4 | Inclusion of binomial data as dependent variable

Finally, we presented an example of analyses with a non-continuous response (child) variable. An obvious advantage of the Bayesian implementation of phylogenetic structural equation models is the flexibility of the approach regarding the distribution of the residuals, which can accommodate binomial, categorical, as well as continuous variables. In the example presented here, the child node (group size) was artificially transformed to a binary trait (for merely illustrative purposes). The results from this model are very similar to those of the model including group size as a continuous trait, and the model showed adequate convergence and a high effective sample size for all parameters.

5 | CONCLUSIONS

In conclusion, we showed how phylogenetic path analysis can be implemented in a Bayesian framework and that our proposed method appears to have a higher discriminating power when comparing competing causal models than maximum likelihood-based PPA (von Hardenberg & Gonzalez-Voyer, 2013). Our Bayesian approach also offers higher flexibility allowing for the inclusion of variability in traits (either as repeated measurements for the same trait or as standard error around the mean value), phylogenetic uncertainty, and non-continuous traits. The flexible JAGS language on which our models are based allows for further future extension. Shipley and Douma (2021) recently showed how latent variables can be included in piecewise structural equation models and therefore tested with the d-sep approach. Analogously, this could be implemented with the Bayesian approach in PhyBaSE adding a prior on the unmeasured latent variable. This could potentially be useful to better account for allometric relationships among traits—that is, including an unmeasured latent variable as causal parent of the traits assumed to be correlated because of allometry—rather than assuming direct causal links among them. Another important problem in path analysis is that of missing values. The typical approach to dealing with missing data is to eliminate species for which even a single data point is missing, which may cause the sample size to rapidly decrease with an increasing number of variables included in a model. Furthermore, missingness of trait values can lead to biased parameter estimates, particularly when data are not missing at random (Nakagawa & Freckleton, 2011), which has stimulated the development of approaches for trait imputation informed by phylogenetic relationships among species (e.g. Goolsby et al., 2017), and more recently an approach accounting for both phylogenetic relationships and trait covariances (Thorston et al., 2023). While we did not implement this in our examples as it would require a much longer treatise beyond the scope of this article, it would be possible to impute the missing data jointly with model fit in the phylogenetic structural equation model, providing priors to the missing values and informing the imputation both with the phylogenetic signal in the residuals and the relationship with the parent node (predictor variable). While PhyBaSE offers a flexible Bayesian framework for phylogenetic structural equation modelling, it is important to acknowledge the computational trade-offs. Compared with existing frequentist approaches such as PPA implemented in phylopath (van der Bijl, 2018) and phylosem (Thorston & van der Bijl, 2023), the Bayesian implementation in PhyBaSE runs

approximately 300–1000 times slower (see comparison in S4). This increase in runtime reflects the added complexity of sampling from posterior distributions and estimating parameter uncertainty within a fully Bayesian context. As a result, users should carefully consider the computational demands, especially when dealing with large datasets or complex models. Despite these limitations, PhyBaSE provides advantages in its flexibility, the ability to incorporate prior information, and quantifying uncertainty—features that may outweigh the computational cost for certain research questions.

AUTHOR CONTRIBUTIONS

Both Achaz von Hardenberg and Alejandro Gonzalez-Voyer contributed equally to all aspects of the manuscript.

ACKNOWLEDGEMENTS

This work was partly funded by a Swedish Research Council (Vetenskapsrådet) Young Researcher grant to AGV (2013-5064), and by funding from Programa de Intercambio Académico, Universidad Nacional Autónoma de México to AGV and AvH. AvH was funded by an International Research Excellence Grant by the University of Chester funded by Santander Universities. We thank Anthony Ives and Jarrod Hadfield for their valuable suggestions on implementing the Bayesian version of the phylogenetic binomial GLMM and two anonymous reviewers for their constructive comments, which improved the manuscript. Open access publishing facilitated by Universita degli Studi di Pavia, as part of the Wiley - CRUI-CARE agreement.

CONFLICT OF INTEREST STATEMENT

The authors have no conflict of interest to declare.

PEER REVIEW

The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/10.1111/2041-210X.70044.

DATA AVAILABILITY STATEMENT

All data and code presented in this article is available via https://doi.org/10.5281/zenodo.15175097 (von Hardenberg & Gonzalez-Voyer, 2025). Data files and R scripts used to simulate the Rhinograd data and implement the PPA models in Gonzalez-Voyer and von Hardenberg (2014) are available via https://doi.org/10.5281/zenodo.15183209 (Gonzalez-Voyer & von Hardenberg, 2025).

ORCID

Achaz von Hardenberg https://orcid.org/0000-0002-9899-1687

REFERENCES

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov & F. Csaki (Eds.), *International symposium on information theory* (pp. 267–281). Akademiai Kiado, Budapest.

Allman, J., McLaughlin, T., & Hakeem, A. (1993). Brain weight and lifespan in primate species. *Proceedings of the National Academy of Sciences of the United States of America*, *90*(1), 118–122.

Arnold, C., Matthews, L. J., & Nunn, C. L. (2010). The 10kTrees website: A new online resource for primate phylogeny. *Evolutionary Anthropology: Issues, News, and Reviews*, *19*(3), 114–118.

Barton, R. A., & Capellini, I. (2011). Maternal investment, life histories, and the costs of brain growth in mammals. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(15), 6169–6174.

Bollen, K. A., & Pearl, J. (2013). Eight myths about causality and structural equation models. In S. L. Morgan (Ed.), *Handbook of causal analysis for social research* (pp. 301–328). Springer Netherlands.

Bro-Jørgensen, J. (2012). Longevity in bovids is promoted by sociality, but reduced by sexual selection. *PLoS One*, *7*(9), e45769. https://doi.org/10.1371/journal.pone.0045769

Bürkner, P.-C. (2017). Brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28. https://doi.org/10.18637/jss.v080.i01

Collier, J. (2020). *Applied structural equation modeling using AMOS: Basic to advanced techniques*. Routledge.

Cooper, N., Thomas, G. H., Venditti, C., Meade, A., & Freckleton, R. P. (2015). A cautionary note on the use of Ornstein Uhlenbeck models in macroevolutionary studies. *Biological Journal of the Linnean Society*, *118*, 64–77.

Dale, J., Dey, C. J., Delhey, K., Kempenaers, B., & Valcu, M. (2015). The effects of life history and sexual selection on male and female plumage colouration. *Nature*, *527*(7578), 367–370.

De Magalhaes, J., & Costa, J. (2009). A database of vertebrate longevity records and their relation to other life-history traits. *Journal of Evolutionary Biology*, *22*(8), 1770–1774. https://doi.org/10.1111/j.1420-9101.2009.01783.x

de Villemereuil, P., Wells, J. A., Edwards, R. D., & Blomberg, S. P. (2012). Bayesian models for comparative analysis integrating phylogenetic uncertainty. *BMC Evolutionary Biology*, *12*(1), 1–16. https://doi.org/10.1186/1471-2148-12-102

Dunbar, R. I., & Shultz, S. (2007). Evolution in the social brain. *Science*, *317*(5843), 1344–1347.

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

Finarelli, J. A. (2011). Estimating endocranial volume from the outside of the skull in artiodactyla. *Journal of Mammalogy*, *92*(1), 200–212.

Fisher, R. A. (1935). *The design of experiments*. Oliver, Boyd.

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

Freckleton, R. P., Cooper, N., & Jetz, W. (2011). The dangers of ignoring an evolutionary model. *American Naturalist*, *178*, E10–E17.

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

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

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian Data Analysis* (3rd ed.). Chapman and Hall/CRC.

Gonzalez-Voyer, A., González-Suárez, M., Vilà, C., & Revilla, E. (2016). Larger brain size indirectly increases vulnerability to extinction in mammals. *Evolution*, *70*(6), 1364–1375.

Gonzalez-Voyer, A., & von Hardenberg, A. (2014). An introduction to phylogenetic path analysis. In *Modern phylogenetic comparative methods and their application in evolutionary biology* (pp. 201–229). Springer.

Gonzalez-Voyer, A., & von Hardenberg, A. (2025). Online practice material for Gonzalez-Voyer & von Hardenberg (2014) (v1.0.0.1). *Zenodo*. https://doi.org/10.5281/zenodo.15183209

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

Grabowski, M., Pienaar, J., Voje, K. L., Andersson, S., Fuentes-Gonzalez, J., Kopperud, B. T., Moen, D. S., Tsuboi, M., Uyeda, J., & Hansen, T. F. (2023). A cautionary note on "A cautionary note on the use of Ornstein Uhlenbeck models in macroevolutionary studies". *Systematic Biology*, *72*, 955–963.

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

Hooten, M. B., & Hobbs, N. T. (2015). A guide to Bayesian model selection for ecologists. *Ecological Monographs*, *85*(1), 3–28.

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

Ives, A. R., & Garland, T. (2014). Phylogenetic regression for binary dependent variables. In T. Garland (Ed.), *Modern phylogenetic comparative methods and their application in evolutionary biology* (pp. 231–261). Springer.

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

Iwaniuk, A. N., & Nelson, J. E. (2002). Can endocranial volume be used as an estimate of brain size in birds? *Canadian Journal of Zoology*, *80*(1), 16–23.

Jiménez-Ortega, D., Kolm, N., Immler, S., Maklakov, A. A., & Gonzalez-Voyer, A. (2020). Long life evolves in large-brained bird lineages. *Evolution*, *74*(12), 2617–2628.

Jones, K. E., Bielby, J., Cardillo, M., Fritz, S. A., O'Dell, J., Orme, C. D. L., Safi, K., Sechrest, W., Boakes, E. H., Carbone, C., Connolly, C., Cutts, M. J., Foster, J. K., Grenyer, R., Habib, M., Plaster, C. A., Price, S. A., Rigby, E. A., Rist, J., … Purvis, A. (2009). PanTHERIA: A species-level database of life history, ecology, and geography of extant and recently extinct mammals: Ecological archives E090-184. *Ecology*, *90*(9), 2648. https://doi.org/10.1890/08-1494.1

Krüger, O., Wolf, J. B., Jonker, R. M., Hoffman, J. I., & Trillmich, F. (2014). Disentangling the contribution of sexual selection and ecology to the evolution of size dimorphism in pinnipeds. *Evolution*, *68*(5), 1485–1496.

Laland, K. N., Sterelny, K., Odling-Smee, J., Hoppitt, W., & Uller, T. (2011). Cause and effect in biology revisited: Is Mayr's proximate-ultimate dichotomy still useful? *Science*, *334*(6062), 1512–1516.

Lee, S. Y. (2007). *Structural equation modeling: A Bayesian approach*. John Wiley & Sons.

Lee, S.-Y., & Song, X.-Y. (2012). *Basic and advanced bayesian structural equation modeling: With applications in the medical and behavioral sciences*. John Wiley & Sons.

Martins, E. P. (2000). Adaptation and the comparative method. *Trends in Ecology & Evolution*, *15*(7), 296–299. https://doi.org/10.1016/s0169-5347(00)01880-2

Martins, E. P., & 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*(4), 646–667.

Mayr, E. (1961). Cause and effect in biology: Kinds of causes, predictability, and teleology are viewed by a practicing biologist. *Science*, *134*(3489), 1501–1506.

Mendes, F. K., Bouckaert, R., Carvalho, L. M., & Drummond, A. J. (2024). How to validate a Bayesian evolutionary model. *Systematic Biology*, *72*(3), 616–634.

Nakagawa, S., & Freckleton, R. P. (2011). Model averaging, missing data and multiple imputation: A case study for behavioural ecology. *Behavioral Ecology and Sociobiology*, *65*(1), 103–116.

O'Meara, B. C. (2012). Evolutionary inferences from phylogenies: A review of methods. *Annual Review of Ecology, Evolution, and Systematics*, *43*(267), 267–285.

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

Paradis, E., & Schliep, K. (2019). ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, *35*, 526–528.

Pearl, J. (2009). *Causality*. Cambridge University Press.

Pérez-Barbería, F. J., & Gordon, I. J. (2005). Gregariousness increases brain size in ungulates. *Oecologia*, *145*(1), 41–52.

Plummer, M. (2003). JAGS: A program for analysis of bayesian graphical models using gibbs sampling. *Proceedings of the 3rd International Workshop on Distributed Statistical Computing*, 124, 1–10.

Plummer, M., Stukalov, A., & Denwood, M. (2016). Rjags: Bayesian graphical models using MCMC. In *R package version* (No. 6; Vol. 4).

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

Revell, L. J., & Harmon, L. J. (2022). *Phylogenetic comparative methods in r*. Princeton University Press.

Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*, 1–36.

Santini, L., González-Suárez, M., Russo, D., Gonzalez-Voyer, A., von Hardenberg, A., & Ancillotto, L. (2019). One strategy does not fit all: Determinants of urban adaptation in mammals. *Ecology Letters*, *22*(2), 365–376. https://doi.org/10.1111/ele.13199

Shipley, B. (1997). Exploratory path analysis with applications in ecology and evolution. *The American Naturalist*, *149*(6), 1113–1138.

Shipley, B. (2016). *Cause and correlation in biology: A user's guide to path analysis, structural equations and causal inference with R* (2nd. edn). Cambridge University Press.

Shipley, B., & Douma, J. C. (2021). Testing piecewise structural equations models in the presence of latent variables and including correlated errors. *Structural Equation Modeling: A Multidisciplinary Journal*, *28*, 582–589.

Silvestro, D., Kostikova, A., Litsios, G., Pearman, P. B., & Salamin, N. (2015). Measurement error should always be incorporated into phylogenetic comparative analysis. *Methods in Ecology and Evolution*, *6*, 340–346.

Sol, D. (2009). Revisiting the cognitive buffer hypothesis for the evolution of large brains. *Biology Letters*, *5*(1), 130–133.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B: Statistical Methodology*, *64*(4), 583–639.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van der Linde, A. (2014). The deviance information criterion: 12 years on. *Journal of the Royal Statistical Society, Series B: Statistical Methodology*, *76*(3), 485–493. https://doi.org/10.1111/rssb.12062

Spirtes, P. (1995). Directed cyclic graphical representation of feedback models. In *Proceedings of the 11th Conference on uncertainty in artificial intelligence* (pp. 491–498). Morgan Kaufmann.

Stümpke, H., & Steiner, G. (1967). *The snouters: Form and life of the rhinogrades*. American Museum of Natural History.

Su, Y.-S., Yajima, M., Su, M. Y.-S., & SystemRequirements, J. (2015). Package "R2jags". In R package version 0.03–08, http://CRAN.R-project.org/package=R2jags

Team, R. C. (2015). *R: A language and environment for statistical computing*. R foundation for statistical computing.

Thorston, J. T., Maureaud, A. A., Frelat, R., Mérigot, B., Bigman, J. S., Friedman, S. T., Palomares, M. L. D., Pinsky, M. L., Price, S. A., & Wainwright, P. (2023). Identifying direct and indirect associations among traits by merging phylogenetic comparative methods and structural equation models. *Methods in Ecology and Evolution*, *14*(5), 1259–1275.

Thorston, J. T., & van der Bijl, W. (2023). 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.

van der Bijl, W. (2018). Phylopath: Easy phylogenetic path analysis in R. *PeerJ*, *6*, e4718.

von Hardenberg, A., & Gonzalez-Voyer, A. (2013). Disentangling evolutionary cause-effect relationships with phylogenetic confirmatory path analysis. *Evolution: International Journal of Organic Evolution*, *67*(2), 378–387.

von Hardenberg, A., & Gonzalez-Voyer, A. (2025). PhyBaSE: data and code (v1.0.1). *Zenodo*. https://doi.org/10.5281/zenodo.15175097

Watanabe, S., & Opper, M. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, *11*(12), 3571–3594.

Wright, S. (1921). Correlation and causation. *Journal of Agricultural Research*, *20*(7), 557–585.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

Figure S1. Depiction of the steps involved in a phylogenetic structural equation analysis using the Bovids models as an example.

Figure S2. Posterior means of prior parameters plotted against their simulated values. The bars represent 95% HPD intervals, colored blue if they include the true value and red if they do not.

S3 Appendix 1. Implementation of a simple PhyBaSE model with JAGS/Winbugs.

Appendix 2. Variability in trait model.

S4 Appendix 3. Comparison of PhyBaSE with phylopath, phylosem and brms.

How to cite this article: von Hardenberg, A., & Gonzalez-Voyer, A. (2025). PhyBaSE: A Bayesian structural equation model approach to causal inference in phylogenetic comparative analyses. *Methods in Ecology and Evolution*, *00*, 1–13. https://doi.org/10.1111/2041-210X.70044