**Supplementary information**

Identifying direct and indirect associations among traits by merging phylogenetic comparative methods and structural equation models

James Thorson, Aurore A. Maureaud, Romain Frelat, Bastien Mérigot, Jennifer S. Bigman, Sarah Friedman, Maria Lourdes D. Palomares, Malin L. Pinsky, Samantha A. Price, Peter Wainwright

**Supplementary Materials A: List of all notation**

Table A1: A summary of mathematical notation used in the main text, listing the symbol, dimension, object “Type” (i.e., classify all objects into different categories where
“dimension” defines the dimensionality of model objects, “index” is an index for looping across dimensions, “Data” is data fitted by the model, “Fixed effect” is a parameter estimated using maximum marginal likelihood, “Random effect” is a coefficient that is marginalized across when calculating the marginal likelihood, “Derived quantity” is an internal object calculated from Data, Fixed effects, or Random effects that clarifies model structure and logic, and “Indicator vector” and “Indicator matrix” contain integers that map between other model objects), and a brief description of its role in the model or presentation.

|  |  |  |  |
| --- | --- | --- | --- |
| **Symbol** | **Dimension** | **Type** | **Description** |
|  | integer | dimension | Number of latent traits, where  |
|  | integer | dimension | Number of modeled taxa including tips and ancestors (but not the root) |
|  | integer | dimension | Number of continuous traits |
|  | integer | dimension | Number of categorical traits |
|  | integer | dimension | Total number of traits,  |
|  | integer | dimension | Total number of trait measurements |
|  | integer | dimension | Number of archetypes specified to classify taxa as a finite mixture of archetypes |
|  | integer | index | Index of taxa,  |
|  | integer | index | Index of latent traits,  |
|  | integer | index | Index of measured traits,  |
|  | integer | index | Index of trait measurements,  |
|  | integer | index | Index of levels for a given categorical trait,  |
|  | integer | index | Index of archetypes,  |
|  |  | Derived quantity | Correlation among any two taxa and for a given trait, created from distance between a taxon and its ancestor, or from distance between the root and the most recent common ancestor of two taxa, as well as a constant representing the maximum root-to-tip distance  |
|  |  | Derived quantity | Correlation among traits for a given taxon, calculated using methods from Structural Equation Modelling (SEM) |
|  |  | Random effect | Latent traits for all taxa, subsetted to extract row-vector for a given taxon, or for a given taxon and trait  |
|  |  | Fixed effects | Estimated trait values for root of tree |
|  |  | Fixed effects | Path matrix representing linkages among traits |
|  |  | Fixed effects | Cholesky of exogenous covariance among traits, where  |
|  |  | Fixed effect | Measurement variance for continuous trait  |
|  |  | Data | Matrix of trait measurements including one column for each continuous or categorical trait |
|  |  | Indicator vector | Integer-vector listing the number of levels for each trait , where continuous trait has and categorical trait has  |
|  |  | Indicator vector | Integer-vector associating each column of latent traits with a column of trait measurements , i.e.,  |
|  |  | Indicator vector | Integer-vector associating each measurements with a taxon  |
|  | Continuous traits: Categorical traits:  | Derived quantity | Submatrix containing columns of associated with trait measurement , where is a column vector for continuous traits, and contains one or more columns for categorical traits |
|  |  | Indicator matrix | Matrix expanding levels for categorical trait ; specifically a matrix with columns where each row contains a 1 in the column corresponding to level and zeros otherwise. |
|  |  | Derived quantity | Probability for each level of levels for categorical trait and taxon , calculated from via a multivariate logistic transformation |
|  |  | Derived quantity | Percent variance explained for each continuous trait  |
|  |  | Derived quantity | Trait value identified for each archetype and latent trait  |
|  |  | Derived quantity | Mixture proportion for each taxon and archetype  |

**Supplementary Materials B: Constructing the correlation matrix for an additive tree**

In the main text, we introduce the matrix of latent traits with rows for each of taxa and columns for each of trait variables. This matrix is treated as a random effect and values are integrated across when calculating the marginal likelihood of fixed effects , , , given the matrix of trait measurements :

|  |  |
| --- | --- |
|  | B1 |

where the probability of data differs somewhat for continuous (Eq. 4) or categorical traits (Eq. 6), and is a hyper-distribution for latent traits (Eq. 1). This marginal likelihood is then maximized to identify maximum-likelihood estimates for parameters:

|  |  |
| --- | --- |
|  | B2 |

In practice, we approximate Eq. B1 using the Laplace approximation, and this involves a sequence of inner and outer optimization iterations that is outlined in detail elsewhere (Kristensen et al., 2016). Briefly, this involves:

1. defining a joint log-likelihood as the sum of the log-likelihood for each individual trait measurement (applying Eq. 4/6 to every value of );
2. proposing an initial value for fixed effects
3. Maximizing with respect to random effects given current values of , and computing the matrix of second derivatives of with respect to (termed the Hessian matrix );
4. Computing the Laplace approximation to the marginal likelihood, , where is the determinant of the Hessian matrix;
5. Computing the gradient of with respect to , and using this gradient to propose a new value of fixed effects ;
6. Repeating steps 3-5 until the gradient in Step-5 is below a specified threshold (we use 0.0001), and defining that value of as the maximum-likelihood estimate;
7. Given the maximum likelihood estimate for , identify the values of that maximize the joint likelihood, and treat these as the “empirical Bayes” predictions for latent trait-values;
8. Computing the matrix of second derivatives of with respect to , inverting this, and treating it as the estimation covariance for fixed effects;

Steps 3-5 are done automatically using the R package TMB (Kristensen et al., 2016), while other steps are done in the R statistical platform (R Core Team, 2021). Steps 6 and 8 involve identifying the value of fixed effects that maximizes the log-likelihood, and also confirms that the negative log-likelihood is positive definite. Satisfying these two criteria ensures that the model is estimable conditional upon available data. As corollary, a non-estimable model can be diagnosed by identifying that the matrix of second derivatives of is rank-deficient (Hunter & Caswell, 2009).

This treatment requires specifying a hyper-distribution for random effects , and we specify a multivariate normal distribution with a separable covariance constructed as the Kroencker production ,where is the matrix representing correlation among taxa for a given trait and is the matrix representing covariance among traits for a given taxon.

We here discuss computational approaches for constructing the correlation matrix . The correlation is typically calculated using a first-order Markov process, including random-walk, Ornstein-Uhlenbeck (OU), or other simple models. For a Brownian motion model, this correlation matrix is typically constructed by calculating the pairwise correlation for every pair of taxa and as:

|  |  |
| --- | --- |
|  | B3 |

where is the distance between the root and the most recent common ancestor of two taxa (Paradis, 2012 modifying Eq. 6.4), and is the maximum root-to-tip distance . Note that we do not model any parameter for the variance of this “Brownian motion” model, because the variance-per-time parameter would be confounded with the scale of and we therefore fix that variance-per-time parameter a priori. Alternatively for an OU model this pairwise correlation can be constructed as:

|  |  |
| --- | --- |
|  | B4 |

where is an estimated parameter representing the autoregressive rate (Paradis, 2012 modifying Eq. 6.5).

 Here, we instead present an alternative formulation of the same process by factoring the joint probability density, , into a series of conditional distributions, , which each represent the change in traits occurring over the branch that connects taxa and its immediate ancestor . Importantly, this conditional form only requires conditioning the distribution of traits for each taxon on the traits of its immediate ancestor , which results from the first-order Markov structure of Brownian-motion, OU and other common models. Using a Brownian motion model, this involves specifying:

|  |  |
| --- | --- |
|  | B5 |

where and is the evolutionary distance between taxon and its ancestor . This specification arises because the Brownian motion model assumes a linear increase in variance with increased evolutionary distance. This specification can then be expressed as a simultaneous autoregressive (SAR) process. This involves constructing the “spatial dependence matrix” of Ver Hoef et al. (2018 Eq. 7), with a for row and column associated with each edge of the specified tree and elsewhere. This SAR process can then be used to construct the inverse-covariance (“precision”) matrix directly, which allows computationally efficient computation (Ver Hoef et al., 2018).

Alternatively, we could instead specify an OU process:

|  |  |
| --- | --- |
|  | B6 |

where and , and we again specify that as to avoid confounding with the scale of (see e.g., Hocking et al., 2018). Or we could specify:

|  |  |
| --- | --- |
|  | B7 |

where

|  |  |
| --- | --- |
|  | B8 |

and where is an indicator that equals 1 if taxon is a tip of the tree and zero otherwise. In this model, is the “Pagel’s lambda” parameter that specifies covariance as a mixture of a Brownian motion model and a star phylogeny with weightings and , respectively. Although we do not implement these methods in the *FishLife* package, they would be relatively simple to implement in future releases. Importantly, the additional parameters or would be estimated jointly with other fixed effects, and the Laplace approximation and subsequent nonlinear maximization routine for identifying fixed effect values that maximize the marginal likelihood would require no further changes.

 Many studies have specified an OU or other autoregressive process to generate a correlation, and then formed a joint covariance for a Gaussian Markov random field via Kroenecker product of the OU correlation and other modeled covariance terms. This process is widely used in spatial statistics (e.g., Hocking et al., 2018; Thorson et al., 2021), and has been used to estimate parameters (and predict latent spatial variables) when specifying 10,000-100,000 latent variables (treated as random effects) and 100-1000 parameters (treated as fixed effects). We acknowledge that fitting a multivariate phylogenetic OU process of this size has not previously been demonstrated, and recommend such a demonstration as topic for future research.

**Works cited**

Fox, J., Nie, Z., & Byrnes, J. (2020). *sem: Structural equation models. R package version 3.1-11*. https://CRAN.R-project.org/package=sem

Hocking, D. J., Thorson, J. T., O’Neil, K., & Letcher, B. H. (2018). A geostatistical state-space model of animal densities for stream networks. *Ecological Applications*, *28*(7), 1782–1796. https://doi.org/10.1002/eap.1767

Holt, S. J. (1958). The evaluation of fisheries resources by the dynamic analysis of stocks, and notes on the time factors involved. *ICNAF Special Publication*, *1*, 77–95.

Hunter, C. M., & Caswell, H. (2009). Rank and Redundancy of Multistate Mark-Recapture Models for Seabird Populations with Unobservable States. In *Modeling Demographic Processes In Marked Populations* (pp. 797–825). Springer, Boston, MA. https://doi.org/10.1007/978-0-387-78151-8\_37

Kristensen, K., Nielsen, A., Berg, C. W., Skaug, H., & Bell, B. M. (2016). TMB: Automatic differentiation and Laplace approximation. *Journal of Statistical Software*, *70*(5), 1–21. https://doi.org/10.18637/jss.v070.i05

Paradis, E. (2012). *Analysis of Phylogenetics and Evolution with R* (Vol. 2). Springer.

Paradis, E., & Schliep, K. (2019). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, *35*(3), 526–528. https://doi.org/10.1093/bioinformatics/bty633

R Core Team. (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Thorson, J. T., Arimitsu, M. L., Barnett, L. A. K., Cheng, W., Eisner, L. B., Haynie, A. C., Hermann, A. J., Holsman, K., Kimmel, D. G., Lomas, M. W., Richar, J., & Siddon, E. C. (2021). Forecasting community reassembly using climate-linked spatio-temporal ecosystem models. *Ecography*, *44*(4), 612–625. https://doi.org/10.1111/ecog.05471

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

Ver Hoef, J. M., Hanks, E. M., & Hooten, M. B. (2018). On the relationship between conditional (CAR) and simultaneous (SAR) autoregressive models. *Spatial Statistics*, *25*, 68–85. https://doi.org/10.1016/j.spasta.2018.04.006

**Supplementary Materials C: Illustrating missing data**

****

Fig. C1: Illustration of data availability for each unique fish species (top panel), genus (middle panel) or family (bottom panel). Each panel shows the proportion of taxa with at least one measurement of a given trait (diagonal cells), or the proportion with each pair of traits (off-diagonal cells), using different color bar legends for each taxonomic level. For clarity, we also list the percentage of taxa with at least one measurement on the diagonal.

**Supplementary Materials D: Comparison with Phylogenetic Comparative Methods**

In the main text, we describe a new method that combines structural equation modelling and phylogenetic comparative methods, while also allowing fit to both continuous and categorical traits. Here, we briefly describe how the method compares with conventional phylogenetic comparative methods (PCM) using R-package *phylolm* (Tung Ho & Ané, 2014) as example of the latter. We first outline theoretical differences, and then provide a simulation experiment to illustrate code-interface differences and explore relative performance.

 We first compare *phylolm* and *FishLife* on a theoretical basis, and start by listing differences in their scope and feature-set:

1. Package *phylolm* assembles the variance-covariance for different evolutionary models, including Brownian motion but also Ornstein-Uhlenbeck (OU), trend, and other models. *FishLife* by contrast only includes an implementation for the Brownian motion model. However, other evolutionary models could be added including OU, Pagel’s lambda and kappa, and rapid burst. The phylogenetic SEM that we present requires calculating the correlation among tips either as a simultaneous or conditional autoregressive process, and parameters that are used to assemble are then estimated via nonlinear minimization of the marginal likelihood function.
2. Package *phylolm* uses the specified evolutionary model, phylogenetic tree, and model parameters to assemble the evolutionary covariance matrix. In then uses characteristics of this structure to efficiently calculate the determinant and quadratic products of the inverse-covariance. By contrast, *FishLife* specifies a conditional or simultaneous autoregressive process (which is available for Brownian motion and OU evolutionary models, and potentially others too) for trait evolution. This CAR or SAR process results in a sparse precision matrix (i.e., is zero for any two taxa that are not immediate ancestors or decedents). It then uses automatic differentiation via R-package TMB to calculate this sparse matrix representing the inverse-correlation matrix (Kristensen et al., 2016). It is then efficient to calculate the determinant and quadratic product from this inverse-covariance directly.
3. Package *phylolm* requires that all predictor and response variables are available for each taxa that is fitted, and drops data for any taxon that does not have all predictor and response variables. By contrast, *FishLife* marginalizes across latent traits for all taxa within the supplied tree, regardless of whether they have complete (or in fact any) data. By doing so, it infers latent traits values jointly while estimating other model parameters. As a result of these design decisions, *phylolm* defines a S3 *predict.phylolm* function that predicts traits for new taxa (or for nodes in the phylogenetic tree) that condition only upon specified predictor variables and not the position of that taxa within the user-specified tree. By contrast, *FishLife* automatically predicts traits for ancestral nodes or new taxa (tips of the phylogenetic tree) that condition upon all data for that taxon, as well as phylogenetic covariance for each trait (i.e., based on the joint covariance in Eq. 1).
4. Both packages include capabilities to estimate the magnitude of measurement errors (i.e., an additive diagonal component to the evolutionary covariance matrix), or to assume that the response is measured without error (i.e., that this diagonal component has variance approaching zero from above).
5. Both packages use R-package *ape* (Paradis & Schliep, 2019)to define the structure of phylogenetic trees, and to allow interoperability with other packages. However, *FishLife* has undergone less development, and therefore does not have a well-defined class structure to leverage this interoperability for use when plotting output or other common tasks in phylogenetic comparative methods.
6. Package *phylolm* defines dependencies using a linear model, which involves defining a directed acyclic graph for the dependencies among variables. By contrast, *FishLife* uses R-package *sem* (Fox et al., 2020) as a user-interface to allow users to specify a RAM that represents the linear dependencies among variables. These dependencies do not need to be acyclic, i.e., can include loops, rings, or other mechanisms involving feedbacks among variables.

This list is not intended to be exhaustive, but instead to emphasize some of the major differences between software platforms.

 We next provide results from a simulation experiment comparing *FishLife* and *phylolm*. To do so, we use R-package apeand function rtree to simulate a new phylogenetic tree with 100 “tips” (i.e., extent taxa with measurable traits) and randomized branch lengths and structure. We then simulate two variables from this tree, using a Brownian motion model for trait evolution and exploring scenarios with complete data for each taxon, or 60% of taxa missing measurements for each trait:

# Settings

Ntree = 100

Nrep = 100

sd\_x = 0.3

sd\_y = 0.3

b0\_x = 1

b0\_y = 2

b\_xy = 1

missing\_rate\_set = c(0, 0.6)[1]

Simulating data from this model involves the following code:

 # Simulate tree

 tree = ape::rtree(n=Ntree)

 # Simulate data

 xfit = x = b0\_x + sd\_x \* phylolm::rTrait(n = 1, phy=tree)

 yfit = y = b0\_y + b\_xy\*x + sd\_y \* phylolm::rTrait(n = 1, phy=tree)

 # Missing at random

 drop\_x = sample( 1:Ntip(tree), replace=FALSE, size=round(Ntip(tree)\*missing\_rate\_set) )

 xfit[drop\_x] = NA

 drop\_y = sample( 1:Ntip(tree), replace=FALSE, size=round(Ntip(tree)\*missing\_rate\_set) )

 yfit[drop\_y] = NA

We then fit these data with package *phylolm*

 # Fit phylolm

 P = phylolm::phylolm(yfit ~ 1 + xfit, phy=tree)

Similarly, we fit the same data using FishLife and identical settings:

 # Run FishLife

 Fit = Fit\_model(

 text = "x -> y, p",

 Database = NULL,

 Y\_ij = data.frame( x=xfit, y=yfit ),

 tree = tree,

 min\_replicate\_measurements = Inf )

We replicate this experiment 500 times for each rate of missing data and record the estimated slope parameter for these two models as well as a conventional linear model with the model formula.

 Results confirm that *FishLife* and *phylolm* give essentially identical estimates when data are complete (Fig. D1, left panel), and that the linear model has degraded performance due to ignoring the known structure for phylogenetic correlations in residuals. By contrast, *FishLife* shows a small improvement in estimation performance when data are missing (Fig. D1, right panel), when *phylolm* has performance intermediate between *FishLife* and the uncorrected linear model. Presumably this improvement occurs because *FishLife* is informed by traits measurements even for taxa where only one is available, and this information improves estimates of trait-values for related species (i.e., jointly estimates missing data and model parameters).

Fig. D1: Results from 500 replicates of a simulation experiment with 0% (left panel) or 60% missing data (right panel), showing the distribution of estimated values for a the impact of the predictor `xfit` on response `yfit` (which has true value 1.0) using three alternative methods (*FishLife, phylolm*, or a conventional linear model), showing the binned count across replicates (y-axis) of estimates (x-axis, see color label at top of plot), where a well-performing model will have a distribution centered at the true value (1.0) and a tight distribution around this value.



**Supplementary Materials E: Detailed results**

Table E1: Estimated mechanisms that jointly define the evolutionary covariance used in phylogenetic trait imputation, where each row corresponds to an edge specified within the structural equation model and maximum likelihood estimate is a estimated value in path matrix .

|  |  |  |
| --- | --- | --- |
| Link | Maximum likelihood estimate | Standard Error |
| temperature -> log(length\_infinity) | -0.017 | 0.003 |
| temperature -> log(growth\_coefficient) | 0.04 | 0.003 |
| temperature -> log(natural\_mortality) | 0.021 | 0.002 |
| temperature -> log(weight\_infinity) | 0.009 | 0.005 |
| log(length\_infinity) -> log(growth\_coefficient) | -0.662 | 0.02 |
| log(length\_infinity) -> log(natural\_mortality) | -0.821 | 0.03 |
| log(length\_infinity) -> log(length\_max) | 0.999 | 0.011 |
| log(length\_infinity) -> log(weight\_infinity) | 2.963 | 0.041 |
| log(natural\_mortality) -> log(length\_maturity) | -1.281 | 0.064 |
| log(natural\_mortality) -> log(age\_maturity) | -0.297 | 0.047 |
| log(natural\_mortality) -> log(age\_max) | -0.851 | 0.039 |
| log(growth\_coefficient) -> log(length\_maturity) | 0.279 | 0.037 |
| log(growth\_coefficient) -> log(age\_maturity) | -0.501 | 0.048 |
| log(weight\_infinity) -> trophic\_level | 0.097 | 0.007 |
| log(weight\_infinity) -> log(fecundity) | 0.709 | 0.023 |
| log(weight\_infinity) -> log(offspring\_size) | 0.061 | 0.01 |
| log(length\_infinity) -> log(aspect\_ratio) | 0.062 | 0.009 |
| log(aspect\_ratio) -> log(max\_body\_width) | -0.134 | 0.009 |
| log(aspect\_ratio) -> log(max\_body\_depth) | 0.223 | 0.012 |
| log(aspect\_ratio) -> log(lower\_jaw\_length) | -0.108 | 0.018 |
| log(aspect\_ratio) -> log(min\_caudal\_pedoncule\_depth) | 0.171 | 0.02 |
| log(length\_infinity) -> spawning\_typeguarders | -3.476 | 0.645 |
| log(length\_infinity) -> spawning\_typebearers | -3.617 | 0.715 |
| log(length\_infinity) -> habitatpelagic | 0.485 | 0.202 |
| log(length\_infinity) -> habitatbenthopelagic | 0.083 | 0.13 |
| log(length\_infinity) -> habitatreefassociated | -0.097 | 0.153 |
| log(length\_infinity) -> habitatbathymetric | 0.578 | 0.247 |
| log(length\_infinity) -> feeding\_modeplanktivorous\_or\_other | -1.758 | 0.373 |
| log(length\_infinity) -> feeding\_modemacrofauna | 0.844 | 0.111 |
| log(length\_infinity) -> body\_shapeelongated | -1.07 | 0.156 |
| log(length\_infinity) -> body\_shapeother | -3.092 | 0.949 |
| log(length\_infinity) -> body\_shapeshort\_and\_or\_deep | -0.396 | 0.207 |
| log(length\_infinity) -> body\_shapeeellike | 0.369 | 0.526 |

Table E2: Predicted “total” effect of an exogenous change in each trait (column) on other traits (rows), including both direct and indirect effects. We list only the total effect for those two traits (columns) that have direct and/or indirect effects on five or more other traits (i.e., nonzero entries in five or more rows rows). See Fig. 2 for details on interpreting categorical trait linkages and names.

|  |  |  |
| --- | --- | --- |
|  | temperature | log(length\_infinity) |
| log(length\_infinity) | -0.017 | 0 |
| log(growth\_coefficient) | 0.051 | -0.662 |
| log(natural\_mortality) | 0.035 | -0.821 |
| log(weight\_infinity) | -0.042 | 2.963 |
| log(length\_max) | -0.017 | 0.999 |
| log(length\_maturity) | -0.031 | 0.867 |
| log(age\_maturity) | -0.036 | 0.576 |
| log(age\_max) | -0.03 | 0.699 |
| trophic\_level | -0.004 | 0.289 |
| log(fecundity) | -0.03 | 2.1 |
| log(offspring\_size) | -0.003 | 0.182 |
| log(aspect\_ratio) | -0.001 | 0.062 |
| log(max\_body\_width) | 0 | -0.008 |
| log(max\_body\_depth) | 0 | 0.014 |
| log(lower\_jaw\_length) | 0 | -0.007 |
| log(min\_caudal\_pedoncule\_depth) | 0 | 0.011 |
| ST: guarders | 0.06 | -3.476 |
| ST: bearers | 0.062 | -3.617 |
| H: pelagic | -0.008 | 0.485 |
| H: benthopelagic | -0.001 | 0.083 |
| H: reefassociated | 0.002 | -0.097 |
| H: bathymetric | -0.01 | 0.578 |
| FM: planktivorous\_or\_other | 0.03 | -1.758 |
| FM: macrofauna | -0.014 | 0.844 |
| BS: elongated | 0.018 | -1.07 |
| BS: other | 0.053 | -3.092 |
| BS: short\_and\_or\_deep | 0.007 | -0.396 |
| BS: eellike | -0.006 | 0.369 |

Table E3: Estimated parameters in matrix , i.e., the Cholesky of the exogenous covariance for the SEM, where the covariance among traits is calculated from and .

|  |  |  |
| --- | --- | --- |
| Link | Par | SE |
| spawning\_typebearers <-> spawning\_typeguarders | 3.11 | 0.607 |
| habitatbenthopelagic <-> habitatpelagic | 1.471 | 0.235 |
| habitatreefassociated <-> habitatpelagic | 0.472 | 0.188 |
| habitatbathymetric <-> habitatpelagic | 0.671 | 0.231 |
| habitatreefassociated <-> habitatbenthopelagic | -0.317 | 0.136 |
| habitatbathymetric <-> habitatbenthopelagic | 0.355 | 0.176 |
| habitatbathymetric <-> habitatreefassociated | -0.412 | 0.176 |
| feeding\_modemacrofauna <-> feeding\_modeplanktivorous\_or\_other | -2.898 | 0.325 |
| body\_shapeother <-> body\_shapeelongated | 0.768 | 0.285 |
| body\_shapeshort\_and\_or\_deep <-> body\_shapeelongated | -0.613 | 0.153 |
| body\_shapeeellike <-> body\_shapeelongated | 1.57 | 0.189 |
| body\_shapeshort\_and\_or\_deep <-> body\_shapeother | 2.2 | 0.794 |
| body\_shapeeellike <-> body\_shapeother | 1.062 | 0.689 |
| body\_shapeeellike <-> body\_shapeshort\_and\_or\_deep | 0.49 | 0.213 |
| log(age\_max) <-> log(age\_max) | 0.257 | 0.01 |
| trophic\_level <-> trophic\_level | 0.305 | 0.005 |
| log(aspect\_ratio) <-> log(aspect\_ratio) | 0.183 | 0.003 |
| log(fecundity) <-> log(fecundity) | 0.887 | 0.022 |
| log(growth\_coefficient) <-> log(growth\_coefficient) | 0.24 | 0.007 |
| temperature <-> temperature | 3.178 | 0.071 |
| log(length\_max) <-> log(length\_max) | 0.089 | 0.006 |
| log(length\_infinity) <-> log(length\_infinity) | 0.382 | 0.005 |
| log(length\_maturity) <-> log(length\_maturity) | 0.104 | 0.011 |
| log(age\_maturity) <-> log(age\_maturity) | 0.186 | 0.01 |
| log(natural\_mortality) <-> log(natural\_mortality) | 0.111 | 0.009 |
| log(weight\_infinity) <-> log(weight\_infinity) | 0.263 | 0.015 |
| log(max\_body\_depth) <-> log(max\_body\_depth) | -0.084 | 0.001 |
| log(max\_body\_width) <-> log(max\_body\_width) | -0.075 | 0.001 |
| log(lower\_jaw\_length) <-> log(lower\_jaw\_length) | 0.172 | 0.002 |
| log(min\_caudal\_pedoncule\_depth) <-> log(min\_caudal\_pedoncule\_depth) | 0.174 | 0.002 |
| log(offspring\_size) <-> log(offspring\_size) | 0.347 | 0.008 |
| spawning\_typeguarders <-> spawning\_typeguarders | 3.358 | 0.326 |
| spawning\_typebearers <-> spawning\_typebearers | 5.488 | 0.61 |
| habitatbathymetric <-> habitatbathymetric | 2.539 | 0.199 |
| habitatbenthopelagic <-> habitatbenthopelagic | 1.704 | 0.107 |
| habitatreefassociated <-> habitatreefassociated | 2.087 | 0.141 |
| habitatpelagic <-> habitatpelagic | 1.895 | 0.202 |
| feeding\_modemacrofauna <-> feeding\_modemacrofauna | 1.184 | 0.079 |
| feeding\_modeplanktivorous\_or\_other <-> feeding\_modeplanktivorous\_or\_other | 2.93 | 0.338 |
| body\_shapeelongated <-> body\_shapeelongated | 1.493 | 0.221 |
| body\_shapeshort\_and\_or\_deep <-> body\_shapeshort\_and\_or\_deep | 2.702 | 0.169 |
| body\_shapeeellike <-> body\_shapeeellike | 4.974 | 0.383 |
| body\_shapeother <-> body\_shapeother | 4.451 | 0.772 |

Fig. E1: Illustration of the estimated trade-off between timing of maturation (, y-axis) and relative mortality (, x-axis), as well as the theoretical prediction (Holt, 1958) assuming .

 

Fig. E2: Path diagram representing specified causal linkages and estimated coefficients linking fish traits when using a subset of available data and a merged version of two separate ultrametric phylogenies constructed independently for bony and cartilaginous fishes (see Fig. 2 for more details)



Fig. E3: Scree plot showing the residual sum of squares (RSS, y-axis) given a specified number of archetypes (x-axis), where we select three archetypes based on the small changes in RSS when adding archetypes past this number.

