

RH: PCM AIC_c

Supplementary material

Model Selection Performance in Phylogenetic Comparative Methods under multivariate Ornstein–Uhlenbeck Models of Trait Evolution

KRZYSZTOF BARTOSZEK¹, JESUALDO FUENTES GONZALEZ², VENELIN MITOV³, JASON
PIENAAR⁴, MARCIN PIWCZYŃSKI⁵, RADOSŁAW PUCHAŁKA⁵, KRZYSZTOF SPALIK⁶ AND
KJETIL LYSNE VOJE⁷

¹*Department of Computer and Information Science, Linköping University, Linköping, Sweden*

²*Department of Biological Sciences, Florida International University, Miami, FL, 33199, USA*

³*IntiQuan GmbH, Basel, Switzerland*

⁴*Department of Biological Sciences and the Institute of Environment, Florida International
University, Miami, FL, 33199, USA*

⁵*Department of Ecology and Biogeography, Nicolaus Copernicus University in Toruń, Toruń,
Poland*

⁶*Institute of Evolutionary Biology, Faculty of Biology, Biological and Chemical Research Centre,
University of Warsaw*

⁷ *Natural History Museum, University of Oslo, Oslo, Norway*

Corresponding author: Krzysztof Bartoszek, Department of Computer and Information
Science, Linköping University, Linköping, Sweden ; E-mail: krzysztof.bartoszek@liu.se,
krzbar@protonmail.ch . Author ordering is alphabetical.

APPENDIX SA: **A** MEASUREMENT ERROR

In all statistical analyses measurement error is an important factor to take into account. In the case of phylogenetically correlated data, ignoring it can lead to e.g., biases in parameter estimates (Hansen and Bartoszek 2012). In **mvSLOUCH** the user passes a measurement error structure, that can be a single number, vector, matrix or list of matrices. If it is a single number, then every measurement is assumed to have this error attached to it; a vector indicates a diagonal measurement error matrix equal for all species; a matrix indicates a measurement error matrix common for all species; and a list of matrices indicates species—specific measurement error. It is not possible, like it was in the original implementation, to pass a measurement error matrix allowing for arbitrary dependencies between measurement errors from different species. Such a structure is impossible to include in the “fast” computations framework because it would be impossible for *PCMBase::PCMLik()* to handle it and furthermore it would break the linearity.

If we consider tip i , then from a mathematical perspective, the measurement error will be included through adding it to the variance-covariance matrix of the transition along the pendant branch leading to it. More specifically, denote by Σ_E^i the measurement error variance-covariance matrix for species i ’s traits, and by $\mathbf{V}(t_i)$ the variance-covariance matrix of the transition along the pendant branch (of length t_i) leading to tip i . Then, \mathbf{V}_i , the variance-covariance matrix of tip i conditional on the (unknown) trait value at the start of the pendant branch leading to tip i is (see also Eq. (19) Mitov et al. 2020)

$$\mathbf{V}_i = \mathbf{V}(t_i) + \Sigma_E^i.$$

PCMBase’s interface allows for direct inclusion of measurement error for each species as the upper triangular Cholesky decomposition of the measurement error variance-covariance matrix, Σ_E . However, in **mvSLOUCH** we employed a different

approach. For each tip branch with measurement error attached to it a new regime (in the **PCMBase** sense) was defined with Σ_e equalling the upper triangular Cholesky decomposition of the measurement error variance-covariance matrix for the given species, i.e., $\Sigma_E = \Sigma_e \Sigma_e^T$. This seemingly more complicated procedure was chosen as it requires the implementation of a mechanism to specify varying parameters over the tree. In future releases of **mvSLOUCH** it will be possible to utilize it to have e.g., clade specific values parameters like **A** or Σ_{yy} .

mvSLOUCH only requires that the user provides the measurement error but how to exactly calculate the measurement error variance-covariance from a sample of multiple measurements on each species is another question. Garamszegi (2014) extensively discusses this topic but the most common way seems to be taking the standard error of the mean for each species (see also discussion by Mitov et al. 2020), where n_i is the number of observed individuals observed for species i ,

$$\hat{\Sigma}_{E_i} = \frac{1}{n_i^2} \sum_{j=1}^{n_i} (\vec{x}_{ij} - \bar{x}_i) (\vec{x}_{ij} - \bar{x}_i)^T$$

as the estimate of the variance of the sample mean of n_i independent observations. If further one assumes that the measurement error in each trait is independent, then $\hat{\Sigma}_{E_i}$ will be a diagonal matrix with the r -th (trait r) diagonal entry equalling

$$\hat{\Sigma}_{E_i}[r, r] = \frac{1}{n_i^2} \sum_{j=1}^{n_i} (\vec{x}_{ijr} - \bar{x}_{ir})^2.$$

The **slouch** package considers, another estimator (see also Garamszegi 2014; Grabowski et al. 2016) of Σ_{E_i} , for the measurement error of species i . The estimator is presented (see **slouch**'s vignette on CRAN) in the single trait case, but it can be directly generalized to

multiple traits. First calculate

$$\mathbf{\Sigma}_W = \frac{\sum_{i=1}^n (n_i - 1) \mathbf{\Sigma}_{W_i}}{\sum_{i=1}^n (n_i - 1)}, \quad (\text{S.1})$$

where $\mathbf{\Sigma}_{W_i}$ is the sample variance for species i . Then, take the measurement error variance-covariance matrix for species i as $\mathbf{\Sigma}_{E_i} = \mathbf{\Sigma}_W / n_i$. Alternatively, if one was very worried that the observations of individuals inside a species are strongly dependent, one could take the sample variance for each species.

APPENDIX SB: \mathbf{A} MATRIX PARAMETRIZATIONS

The choice of \mathbf{A} matrix class seems to be critical for formulating biological hypotheses, as well as increasing the stability of the numerical optimization and the ability to identify models and parameters. Here we discuss all possible parametrizations offered by **mvSLOUCH** for $\mathbf{A}_{\text{Eq. (1)}}$. The same ones can be used with respect to $\mathbf{A}_{\text{Eq. (2)}}$ (the interpretations are essentially the same, we will point out differences).

The **mvSLOUCH** package supports the following types of matrices for \mathbf{A} , passed through the parameter *Atype* for both model types, OUOU *mvSLOUCH::ouchModel()* and OUBM *mvSLOUCH::mvslouchModel(): "SingleValueDiagonal", "Diagonal", "UpperTri", "LowerTri", "Symmetric", "SymmetricPositiveDefinite", "DecomposablePositive", "DecomposableNegative", "DecomposableReal", "Invertible", "Any", "TwoByTwo"*. The parametrizations of these matrices are described in detail by Bartoszek (2011). We will start by describing the *"DecomposablePositive"* parametrization first as it is, most probably, the most relevant one from an evolutionary biology perspective.

The parametrization *DecomposablePositive* means that \mathbf{A} has an eigendecomposition with real strictly positive eigenvalues. This can be interpreted as

adaptation towards the optimum vector $\vec{\psi}(t)$. The speed of this adaptation is controlled by the magnitude of the eigenvalues. However, it has to be remembered that each eigenvalue (and corresponding half-life) does not say anything about the speed of adaptation (or loss of ancestral signal) in the trait space, only in the direction of the corresponding eigenvector.

The parametrization *DecomposableNegative* means that \mathbf{A} has an eigendecomposition with real strictly negative eigenvalues. Negative eigenvalues do not have such an elegant biological interpretation as positive ones do. Mathematically it means that the traits vector is being repulsed from the optimum vector $\vec{\psi}(t)$. Although this can be naïvely thought of as character displacement (Bartoszek et al. 2012), any application would require a careful interpretation in terms of well-thought out biological hypotheses

The parametrizations given by *"SingleValueDiagonal"*, *"Diagonal"*, and *"SymmetricPositiveDefinite"* can be considered to be special cases of *"DecomposablePositive"*, given that \mathbf{A} 's diagonal is positive. *"Diagonal"* means that \mathbf{A} is a diagonal matrix. The eigenvalues of \mathbf{A} are just the values on the diagonal and the eigenvector matrix is the identity matrix, $\mathbf{P} = \mathbf{I}$. This matrix can be thought of as a situation where all the traits adapt to their joint primary optimum independently. No trait has any influence on any other's adaptation path. The traits can still be correlated, through Σ_{yy} mixing the Wiener noise processes, but this will only influence the traits' oscillations around their expected value trajectories. *"SingleValueDiagonal"* is the same except that all the values on \mathbf{A} 's diagonal are the same. This means that all traits evolve/adapt at the same speed, which requires a biological justification. Lastly, *"SymmetricPositiveDefinite"* implies that all eigenvalues are real and strictly positive (property of symmetric positive definite matrices) but also that effects are symmetric. How trait i affects trait j 's adaptation is the same as trait j affects trait i 's adaptation. This could in general be problematic as it firstly requires that a unit change in one trait corresponds to a unit change in the other trait. Secondly, symmetry of effects should

require a biological justification. On the other hand, a symmetric positive definite \mathbf{A} is straightforward to parametrize by a triangular matrix as it can be factorized through the Cholesky decomposition. Furthermore, its eigendecomposition is $\mathbf{A} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}^T$, i.e., there is no need to invert \mathbf{P} . Therefore, it is mathematically and numerically friendly.

"UpperTri" (*"LowerTri"*) means that \mathbf{A} is an upper-triangular (lower-triangular) matrix. This implies that there is a hierarchy of influences of adaptations/trait's primary optimum in terms of other traits. The trait corresponding to the bottom (top) row adapts independently, then the second row from the bottom (top) trait's adaptation and primary optimum are only influenced by the bottom (top) row trait. In general the adaptation and primary optimum of the i -th row trait are only influenced by traits corresponding to rows $k, \dots, i+1$ ($1, \dots, i-1$). The eigenvalues of \mathbf{A} are equal to the values on \mathbf{A} 's diagonal and (assuming the eigenvectors are in the order corresponding to the diagonal, i.e., *not* according to the magnitude of the eigenvalues/diagonal entries) \mathbf{P} is upper (lower)-triangular. This means that the first (last) eigenvalue (entry of diagonal) corresponds to the half-life of the first (last) trait, the second (second-last) eigenvalue to the half-life in the direction of a linear combination of the first two traits and, in general the i -th eigenvalue (entry of diagonal) corresponds to the half-life in the direction of a linear combination of the first (last) i traits. We remind again that independent adaptation does not imply independent evolution. Correlations can still occur due to a non-diagonal Σ_{yy} matrix.

The remaining matrix parametrizations *"Symmetric"*, *"DecomposableReal"*, *"Invertible"*, *"Any"*, and *"TwoByTwo"* do not make any specific assumptions about the eigenvalues of \mathbf{A} and hence about the adaptation of the traits. *"Symmetric"* has been described under *"SymmetricPositiveDefinite"* (bar the adaptation).

"DecomposableReal" assumes that the eigenvalues are real; *"Invertible"*, parametrized by its QR decomposition, only assumes invertibility and eigendecomposability. This is a

very general setting and is not recommended. From our experience it results in long running times and the estimation procedure can get easily stuck at a local maximum. However, it was chosen as the default setting for *Atype* as it makes the fewest possible biological assumptions about the traits and should motivate the user to carefully think out the appropriate adaptive hypotheses. *"Any"*, as the name suggests, means any matrix, parametrized just by its entries. It was provided for completeness of parametrization of **A** but in the original implementation of the package it was not practical, as there is no guarantee that it will be invertible (a requirement of the original implementation). As **PCMBase**'s computational engine can handle a singular **A**, this parametrization comes in handy when combined with the newly implemented possibility of specifying individual entries of **A** (and any other model parameter) to predefined values. The final *"TwoByTwo"* parametrization implies that $\mathbf{A} \in \mathbb{R}^{2 \times 2}$ and is equivalent to *"Any"* in this situation. The two-by-two case is special as there is a general closed form formula for the matrix exponential in terms of the matrix's entries only (Bernstein and So 2013).

Concerning the **A** matrix, there is also the function parameter *diagA*. It allows forcing the diagonal to be positive or negative (using exponentiation). No conflicts with the *Atype* specifications are considered though, and hence it should be used with caution. For example, it is not needed for the *"SymmetricPositiveDefinite"* setting. In this case the diagonal will be positive anyway. It is unclear what effect exponentiating it could have for the optimization, e.g., in the *DecomposablePositive*, *DecomposableNegative*, *"DecomposableReal"* or *"Invertible"* setting. On the one hand it will force the diagonal to have the desired sign, but on the other it could interfere with *optim()*'s path over parametrization of the eigenvalues, eigenvectors or QR decomposition.

APPENDIX SC: GENERAL SIMULATION-REESTIMATION STUDY; SETUP AND RESULTS

Parameter estimation under phylogenetic OU models is difficult. Cressler et al. (2015) and Ho and Ané (2014) studied the issue of estimating parameters, also the drift and diffusion parameters, in the one-dimensional phylogenetic OU case. We should not expect the situation to improve when one moves to multiple dimensions. In fact Bartoszek et al. (2012) did observe, in the simulation study, high variability of parameter estimates (Tab. S.1 therein). However, histograms (Figs. S.1–S.7 therein) showed that key for the process’ dynamics summary statistics (deterministic optima, evolutionary and optimal regression, phylogenetic half-lives) tend to cluster around the true value, despite very large outliers sometimes. However, this simulation study was only conducted for one given OUBM model and six particular phylogenies with 32, 64 and 128 tips (each size had a non-ultrametric and ultrametric version). Such a restricted setup was naturally due to limited computational and processor power at the time, and the implementation of the “slow” likelihood evaluation algorithm. As computational power increased with time, more complex studies became possible. For example, Adams and Collyer (2018) looked into model selection capabilities of multivariate phylogenetic software. They simulated phenotypic data for 32, 64, and 128 tips with 4, 8, and 12 independent traits, finding that in some situations model selection by Akaike’s information criterion (AIC Akaike 1974) preferred the more complex OU model. Given the improved estimation possibilities due to the fast likelihood evaluation algorithm, we are able to execute a simulation-reestimation study of far greater complexity. We simulate and reestimate under various BM, OUOU and OUBM setups. In particular we are interested in how do various assumptions of the drift and diffusion matrices effect estimation and which setups are distinguishable. Furthermore, we must remember that there are no analytical results, that we are aware of, concerning directions of potential bias in parameter or model identifiability. Model selection is done by Akaike’s information criterion corrected for sample size (AIC_c , Hurvich and Tsai 1989),

$$\text{AIC} = 2r - 2 \log \mathcal{L}, \quad \text{AIC}_c = \text{AIC} + \frac{2r(r+1)}{K-r-1}.$$

In the above, r is the number of free parameters of the model, \mathcal{L} the value of the likelihood function evaluated at the estimates of the parameters, and K the number of observed data points. If there are no missing observations, then $K = nk$, where n is the number of tips and k the number of variables measured. It should be noted that the AIC_c was derived for K independent data points, which of course does not hold here, due to the common evolutionary history. However, Bartoszek (2016) suggested that this might not be a serious issue in the context of model selection with enough tip species.

We must remember that the general results of consistency of maximum likelihood estimation cannot be directly invoked here. We have a correlated observations setup and, depending on the model of tree growth and particular parameters of the SDE evolving on the tree, required conditions might or might not hold and rates of convergence are unknown in general. In fact both analytical and simulation studies (Ané 2008; Ané et al. 2016; Bartoszek 2016; Sagitov and Bartoszek 2012) show that there are setups where even with the growth of the number of observations, there is a limit on the amount of available information, due to the tree-induced dependencies. We consider a wider variety of tree sizes, 32, 64, 128, 256, 512, 1024 and 2048. It is important to point out that due to running times, it was only possible to carry out this simulations study thanks to the **PCMBaseC++** C++ backend. Without it, using the pure R implementation of **PCMBase**, a pilot (for $n \in \{32, 64, 128, 256, 512, 1024\}$) version of the simulation study presented in this work took about a year. On the other hand with the **PCMBaseC++** backend, an extended version, also with $n = 2048$, took a little under two months. The last month was solely devoted to the analysis of the 2048 tips case. We do not present specific running times as they would not be representative—other tasks (also computationally heavy) were

also being carried out in parallel on the machine running the analyses. All simulations were carried out using version 2.6.2 of **mvSLOUCH** in R version 3.6.1 running on an openSUSE 42.3 (*x86_64*) box with a 3.50GHz Intel[®] Xeon[®] CPU. Below, in Tab. S.1, we present different parameter sets under which four dimensional data were simulated and then, in Tab. S.3 the model identifiability capabilities for the different simulation model-reestimation model pairs (true and competing model). We point out that due to time constraints it was not possible to consider all possible pairs. In Tab. S.2 we explain the different abbreviations for the estimated models. It has to be noted that the actual number of simulation-reestimation repeats is higher, see Tab. S.4, as some runs ended in error when estimating under the correct model.

Table S.1: Models under which simulation of a four dimensional trait was performed. The parameters are in terms of the general Eq. (1). This means that for the BM model $\vec{Y}(0) = \vec{X}(0)$ and $\Sigma_{yy} = \Sigma_{xx}$. For the OUBM models $\mathbf{A}_{\text{Eq. (2)}} \vec{Y}(0) = [\vec{Y}(0)_{\text{Eq. (2)}}^T \vec{X}(0)^T]^T$, \mathbf{A} contains both $\mathbf{A}_{\text{Eq. (2)}}$ and \mathbf{Q} , Σ_{yy} contains both $\Sigma_{yy_{\text{Eq. (2)}}}$ and $\Sigma_{xx_{\text{Eq. (2)}}}$ while $\vec{\psi}$ contains $\vec{\psi}_{\text{Eq. (2)}}$ and 0s (for the “BM” traits).

Abbreviation	$\vec{Y}(0)$	$\vec{\psi}$	\mathbf{A}	Σ_{yy}	comment
BM	$\vec{0}$	—	—	\mathbf{I}_4	independent traits
OUOUs1	$\vec{0}$	$\vec{0}$	$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 4 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$	
OUOUs2	$\vec{0}$	$\vec{0}$	$\begin{bmatrix} 1 & 0.2 & 0.3 & 0.4 \\ 0.4 & 2 & 0.2 & 0.1 \\ 0.1 & 0.3 & 3 & 0.4 \\ 0.1 & 0.1 & 0.1 & 4 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$	
OUBMs1	$\vec{0}$	$\vec{0}$	$\begin{bmatrix} 1 & 0 & 5 & -5 \\ 0 & 2 & -5 & 5 \\ \hline 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \hline 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix}$	
OUBMs2	$\vec{0}$	$\vec{0}$	$\begin{bmatrix} 1 & 0.1 & 5 & -5 \\ 0.3 & 2 & -5 & 5 \\ \hline 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \hline 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix}$	

Table S.2: Abbreviations for models under which estimation was performed, ML stands for maximum likelihood.





Abbreviation	"Atype"	"Syytype"	"diagA"	comment
OYOU1	"SingleValueDiagonal "	"UpperTri "	"Positive"	independent traits
OYOU2	"Diagonal "	"Diagonal "	"Positive"	
OYOU2U	"Diagonal "	"UpperTri "	"Positive"	
OYOU3	"DecomposablePositive"	"Diagonal "	NULL	
OYOU3P	"DecomposablePositive"	"Diagonal "	"Positive"	
OYOU4	"DecomposablePositive"	"UpperTri "	NULL	mimics original setup of Adams and Collyer (2018)
OYOU4P	"DecomposablePositive"	"UpperTri "	"Positive"	
OYOU5	"Invertible"	"UpperTri "	"Positive"	
OUBM1ML	"SingleValueDiagonal "	"UpperTri "	"Positive"	Q estimated by ML
OUBM2ML	"Diagonal "	"Diagonal "	"Positive"	Q estimated by ML
OUBM2UML	"Diagonal "	"UpperTri "	"Positive"	Q estimated by ML
OUBM3ML	"DecomposablePositive"	"Diagonal "	NULL	Q estimated by ML
OUBM3PML	"DecomposablePositive"	"Diagonal "	"Positive"	Q estimated by ML
OUBM4ML	"DecomposablePositive"	"UpperTri "	NULL	Q estimated by ML
OUBM4PML	"DecomposablePositive"	"UpperTri "	"Positive"	Q estimated by ML






Table S.3: Model identifiability capabilities of **mvSLOUCH**. The columns are the true model under which the data was simulated (Tab. S.1, abbreviations described therein). The rows are the models under which estimation was done (abbreviations described in Tab. S.2). The values in the table are the fraction of times the true (under which data was simulated) model was chosen over the alternative (row) model by AIC_c . Each fraction is the result of 100 successful simulation-reestimation repeats. It has to be noted that the actual number of simulation-reestimation repeats is higher, see Tab. S.4 as some runs ended in error when estimating under the correct model. A dash, —, means that that particular combination was not considered. A ♣ means that the simulated and estimated under models are the same, hence a value there is meaningless. If there is an entry in brackets it indicates that estimation under the true model setup was done with the constraint *diagA="Positive"*. For the OUBMs1 and OUBMs2 columns the estimation for **Q** was done by maximum likelihood. The number in bold at the top of each subtable is the number of tips.

32	BM	OUOUs1	OUOUs2	OUBMs1	OUBMs2
BM	♣	(1)	0.21 (0.19)	(1)	0.99 (0.99)
OUOU1	0.76	—	—	—	—
OUOU2	—	♣	0 (0)	(1)	1 (1)
OUOU2U	—	(0.98)	0 (0)	(1)	1 (1)
OUOU3	—	(1)	♣	(1)	1 (1)
OUOU3P	—	(1)	♣	(1)	1 (1)
OUOU4	1	(1)	1 (1)	(1)	1 (1)
OUOU4P	1	(1)	1 (1)	(1)	1 (1)
OUOU5	1	—	—	—	—
OUBM1ML	0.47	—	—	—	—
OUBM2ML	—	(1)	0.04 (0.03)	♣	0.22 (0.15)
OUBM2UML	—	(0.99)	0.04 (0.03)	(0.63)	0.21 (0.11)
OUBM3ML	—	(1)	0.08 (0.07)	(0.86)	♣
OUBM3PML	—	(1)	0.1 (0.08)	(0.89)	♣
OUBM4ML	0.89	(1)	0.12 (0.11)	(0.92)	0.87 (0.76)
OUBM4PML	0.89	(1)	0.13 (0.12)	(0.96)	0.88 (0.83)

64	BM	OUOUs1	OUOUs2	OUBMs1	OUBMs2
BM	♣	(1)	0.75 (0.73)	(1)	1 (1)
OUOU1	0.82	—	—	—	—
OUOU2	—	♣	0 (0)	(1)	1 (1)
OUOU2U	—	(0.97)	0 (0)	(1)	1 (1)
OUOU3	—	(1)	♣	(1)	1 (1)
OUOU3P	—	(1)	♣	(1)	1 (1)
OUOU4	1	(1)	0.99 (0.98)	(1)	1 (1)
OUOU4P	1	(1)	0.99 (0.98)	(1)	1 (1)
OUOU5	1	—	—	—	—
OUBM1ML	0.63	—	—	—	—
OUBM2ML	—	(1)	0.43 (0.39)	♣	0.29 (0.24)
OUBM2UML	—	(1)	0.47 (0.44)	(0.57)	0.34 (0.21)
OUBM3ML	—	(1)	0.53 (0.49)	(0.78)	♣
OUBM3PML	—	(1)	0.55 (0.51)	(0.84)	♣
OUBM4ML	0.95	(1)	0.56 (0.53)	(0.88)	0.82 (0.63)
OUBM4PML	0.96	(1)	0.62 (0.59)	(0.89)	0.88 (0.74)

128	BM	OUOUs1	OUOUs2	OUBMs1	OUBMs2
BM	♣	(1)	0.99 (0.99)	(1)	1 (1)
OUOU1	0.8	—	—	—	—
OUOU2	—	♣	0 (0)	(1)	(1) (1)
OUOU2U	—	(0.95)	0 (0)	(1)	1 (1)
OUOU3	—	(1)	♣	(1)	1 (1)

OUOU3P	—	(1)		(1)	1 (1)
OUOU4	1	(1)	0.99 (0.98)	(1)	1 (1)
OUOU4P	1	(1)	0.99 (0.98)	(1)	1 (1)
OUOU5	1	—	—	—	—
OUBM1ML	0.64	—	—	—	—
OUBM2ML	—	(1)	0.96 (0.93)		0.37 (0.29)
OUBM2UML	—	(1)	0.92 (0.9)	(0.57)	0.28 (0.17)
OUBM3ML	—	(1)	0.94 (0.93)	(0.74)	
OUBM3PML	—	(1)	0.98 (0.98)	(0.83)	
OUBM4ML	0.93	(1)	0.99 (0.98)	(0.79)	0.77 (0.57)
OUBM4PML	0.96	(1)	0.99 (0.99)	(0.88)	0.86 (0.71)

256	BM	OUOU _s 1	OUOU _s 2	OUBM _s 1	OUBM _s 2
BM		(1)	1 (1)	(1)	1 (1)
OUOU1	0.86	—	—	—	—
OUOU2	—		0 (0)	(1)	1 (1)
OUOU2U	—	(0.96)	0 (0)	(1)	1 (1)
OUOU3	—	(1)		(1)	1 (1)
OUOU3P	—	(1)		(1)	1 (1)
OUOU4	1	(1)	0.89 (0.88)	(1)	1 (1)
OUOU4P	1	(1)	0.89 (0.88)	(1)	1 (1)
OUOU5	1	—	—	—	—
OUBM1ML	0.63	—	—	—	—
OUBM2ML	—	(1)	0.99 (0.98)		0.44 (0.33)
OUBM2UML	—	(1)	0.99 (0.98)	(0.52)	0.46 (0.36)

OUBM3ML	—	(1)	1 (1)	(0.73)	♣
OUBM3PML	—	(1)	1 (1)	(0.78)	♣
OUBM4ML	0.92	(1)	1 (1)	(0.79)	0.7 (0.56)
OUBM4PML	0.98	(1)	1 (1)	(0.82)	0.86 (0.73)

512	BM	OUOUs1	OUOUs2	OUBMs1	OUBMs2
BM	♣	(1)	1 (1)	(1)	1 (1)
OUOU1	0.89	—	—	—	—
OUOU2	—	♣	0.01 (0.01)	(1)	1 (1)
OUOU2U	—	(0.99)	0.01 (0.01)	(1)	1 (1)
OUOU3	—	(1)	♣	(1)	1 (1)
OUOU3P	—	(1)	♣	(1)	1 (1)
OUOU4	1	(1)	0.93 (0.93)	(1)	1 (1)
OUOU4P	1	(1)	0.93 (0.93)	(1)	1 (1)
OUOU5	1	—	—	—	—
OUBM1ML	0.72	—	—	—	—
OUBM2ML	—	(1)	1 (1)	♣	0.42 (0.31)
OUBM2UML	—	(1)	1 (1)	(0.64)	0.4 (0.28)
OUBM3ML	—	(1)	1 (1)	(0.7)	♣
OUBM3PML	—	(1)	1 (1)	(0.76)	♣
OUBM4ML	0.97	(1)	1 (1)	(0.78)	0.78 (0.57)
OUBM4PML	0.99	(1)	1 (1)	(0.84)	0.85 (0.72)

1024	BM	OUOUs1	OUOUs2	OUBMs1	OUBMs2
BM	♣	(1)	1 (1)	(1)	1 (1)

OUOU1	0.96	—	—	—	—
OUOU2	—	♣	0.01 (0.01)	(1)	1 (1)
OUOU2U	—	(0.97)	0.02 (0.02)	(1)	1 (1)
OUOU3	—	(0.99)	♣	(1)	1 (1)
OUOU3P	—	(0.99)	♣	(1)	1 (1)
OUOU4	1	(1)	0.81 (0.81)	(1)	1 (1)
OUOU4P	1	(1)	0.81 (0.81)	(1)	1 (1)
OUOU5	1	—	—	—	—
OUBM1ML	0.72	—	—	—	—
OUBM2ML	—	(1)	1 (1)	♣	0.56 (0.38)
OUBM2UML	—	(1)	1 (1)	(0.62)	0.51 (0.34)
OUBM3ML	—	(1)	1 (1)	(0.68)	♣
OUBM3PML	—	(1)	1 (1)	(0.78)	♣
OUBM4ML	0.98	(1)	1 (1)	(0.72)	0.76 (0.54)
OUBM4PML	0.98	(1)	1 (1)	(0.82)	0.85 (0.65)

2048	BM	OUOU _{s1}	OUOU _{s2}	OUBM _{s1}	OUBM _{s2}
BM	♣	(1)	1 (1)	(1)	1 (1)
OUOU1	0.98	—	—	—	—
OUOU2	—	♣	0 (0)	(1)	1 (1)
OUOU2U	—	(0.96)	0 (0)	(1)	1 (1)
OUOU3	—	(1)	♣	(1)	1 (1)
OUOU3P	—	(1)	♣	(1)	1 (1)
OUOU4	1	(1)	0.81 (0.81)	(1)	1 (1)
OUOU4P	1	(1)	0.81 (0.81)	(1)	1 (1)

OOU5	1	—	—	—	—
OUBM1ML	0.67	—	—	—	—
OUBM2ML	—	(1)	1 (1)	♣	0.52 (0.36)
OUBM2UML	—	(1)	1 (1)	(0.52)	0.42 (0.23)
OUBM3ML	—	(1)	1 (1)	(0.73)	♣
OUBM3PML	—	(1)	1 (1)	(0.85)	♣
OUBM4ML	1	(1)	1 (1)	(0.82)	0.76 (0.62)
OUBM4PML	0.97	(1)	1 (1)	(0.89)	0.88 (0.75)

Table S.4: Number of simulation runs performed to obtain 100 (per cell) estimation results that are presented in Tab. S.3. The reruns are the results of errors when estimating under the model under which the data was simulated, hence no comparison was possible.

n	BM	OOUUs1	OOUUs2	OUBMs1	OUBMs2
32	100	100	100	100	100
64	100	100	100	101	100
128	102	100	103	100	101
256	101	101	100	102	101
512	104	102	102	100	103
1024	108	105	109	106	104
2048	123	110	108	110	117

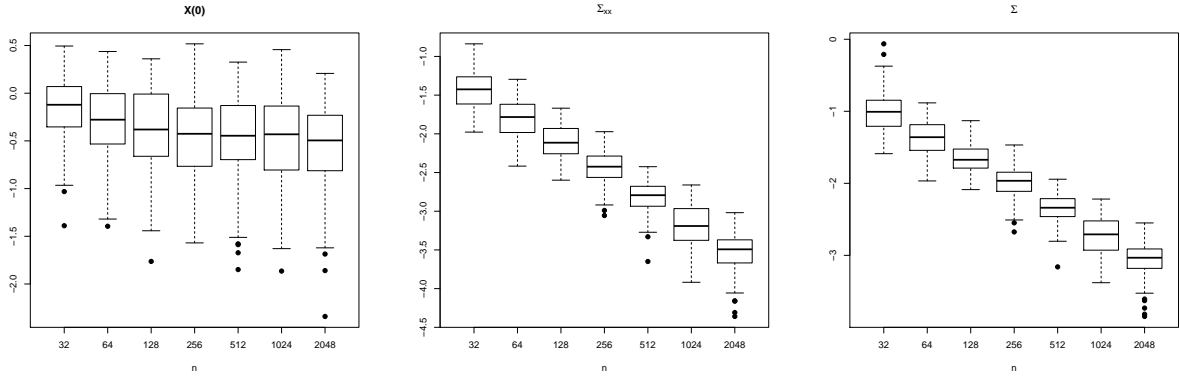


Figure S.1: Quality of parameter estimates under the BM model (Tab. S.1). Boxplots of relative distances (Eq. 4) or Euclidean distances (if the true parameter is $\vec{0}$, $\vec{X}(0)$ case, Eq. 3) between true matrix/vector parameter and estimated ones. The y -axes are on logarithmic scale, x -axes are the number of tips of the phylogeny. The boxplots come from 100 successful simulation-reestimation runs, see Tab. S.4. Left: $\vec{X}(0)$, centre: Σ_{xx} , right: Σ .

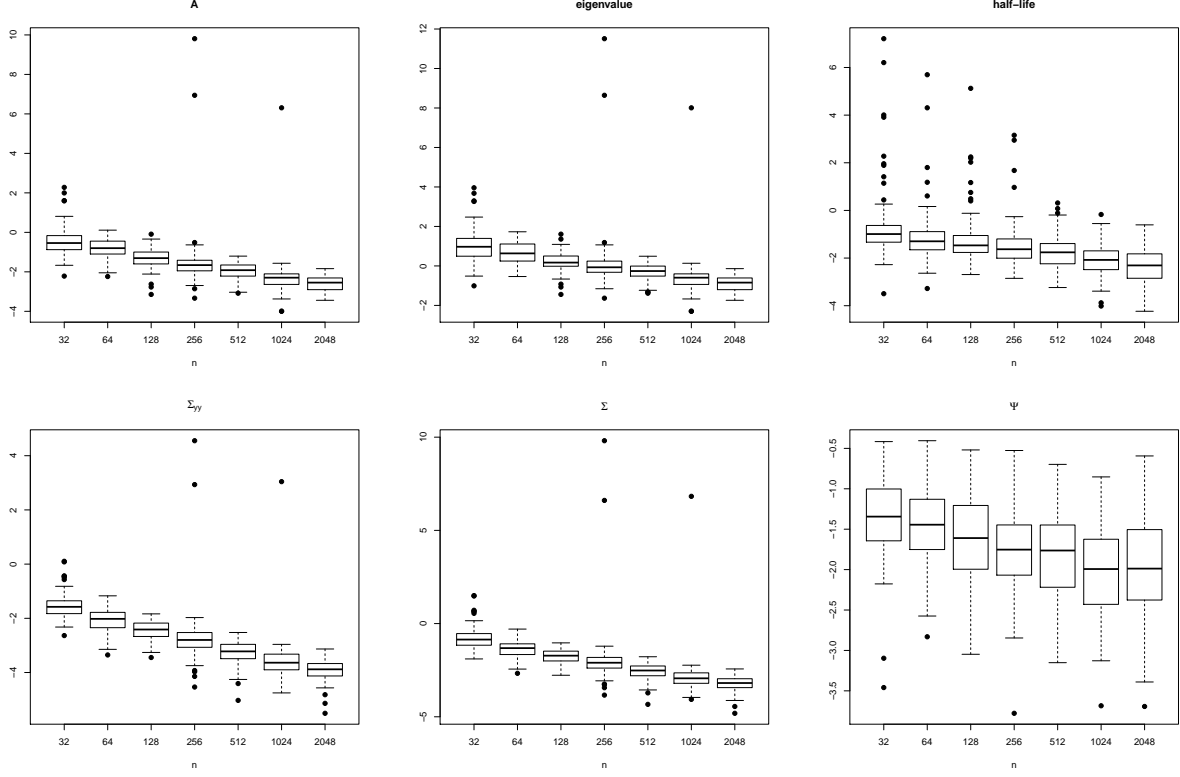


Figure S.2: Quality of parameter estimates under the OUOUS1 model (Tab. S.1). Boxplots of relative distances (Eq. 4) or Euclidean distances (if the true parameter is $\vec{0}$, $\vec{Y}(0)$, $\vec{\psi}$ cases, Eq. 3) between true matrix/vector parameter and estimated ones. The y -axes are on logarithmic scale, x -axes are the number of tips of the phylogeny. The boxplots come from 100 successful simulation-reestimation runs, see Tab. S.4.

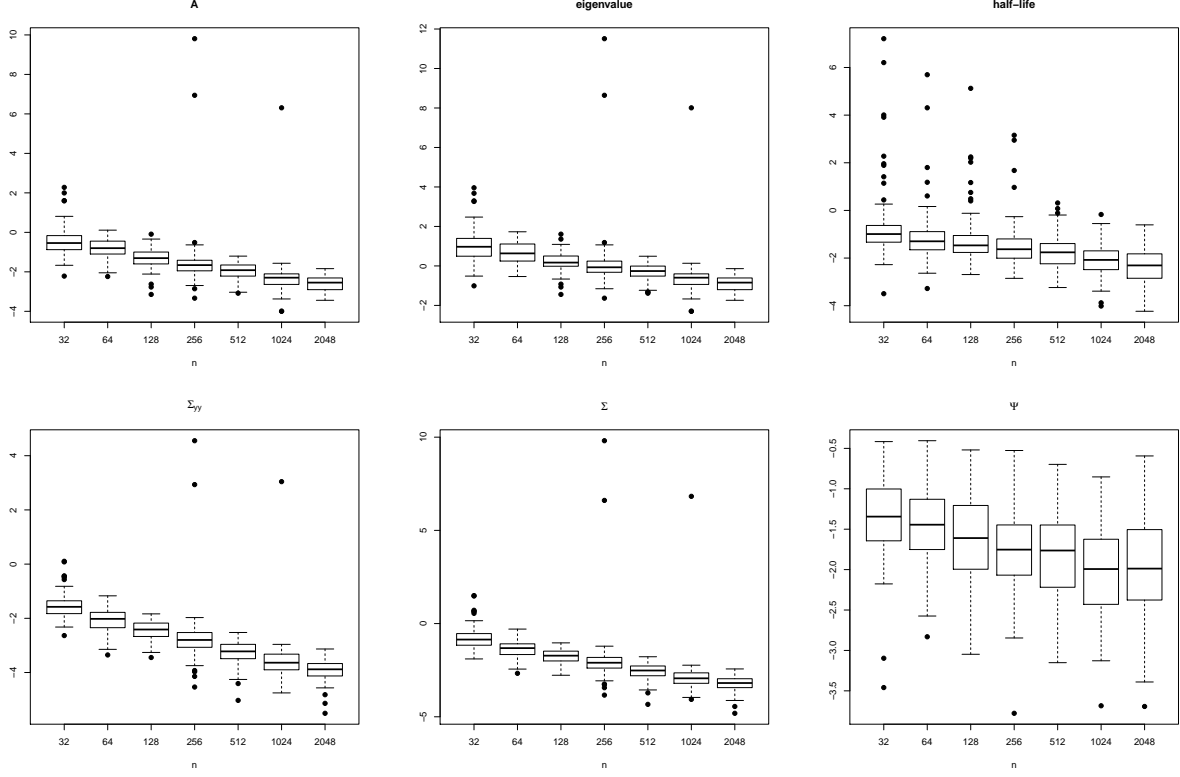


Figure S.3: Quality of parameter estimates under the OUOUS1P model (Tab. S.1). Boxplots of relative distances (Eq. 4) or Euclidean distances (if the true parameter is $\vec{0}$, $\vec{Y}(0)$, $\vec{\psi}$ cases, Eq. 3) between true matrix/vector parameter and estimated ones. The y -axes are on logarithmic scale, x -axes are the number of tips of the phylogeny. The boxplots come from 100 successful simulation-reestimation runs, see Tab. S.4.

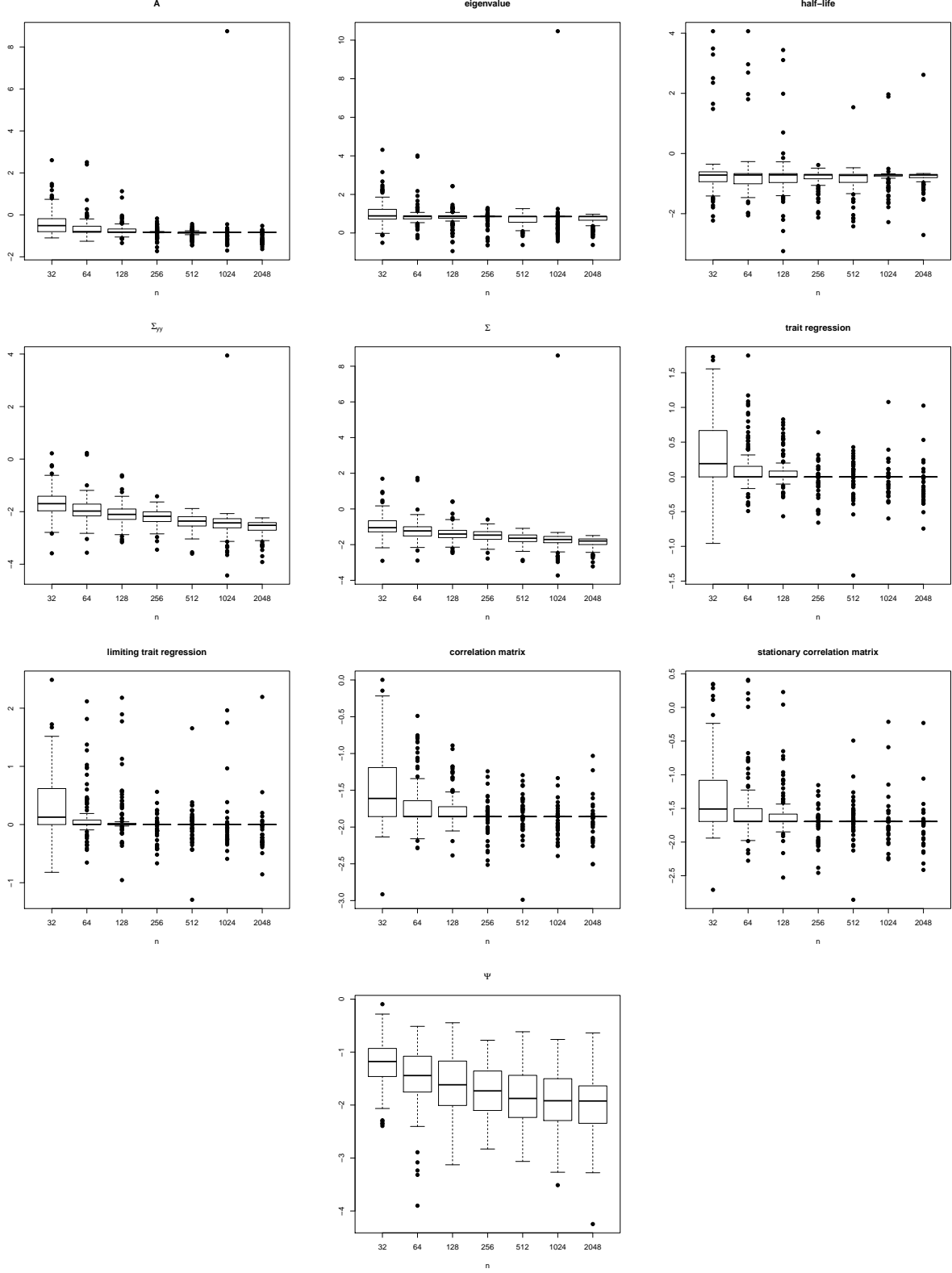


Figure S.4: Quality of parameter estimates under the OUOUS2 model (Tab. S.1). Boxplots of relative distances (Eq. 4) or Euclidean distances (if the true parameter is $\vec{0}$, $\vec{Y}(0)$, $\vec{\psi}$ cases, Eq. 3) between true matrix/vector parameter and estimated ones. The y -axes are on logarithmic scale, x -axes are the number of tips of the phylogeny. The boxplots come from 100 successful simulation-reestimation runs, see Tab. S.4.

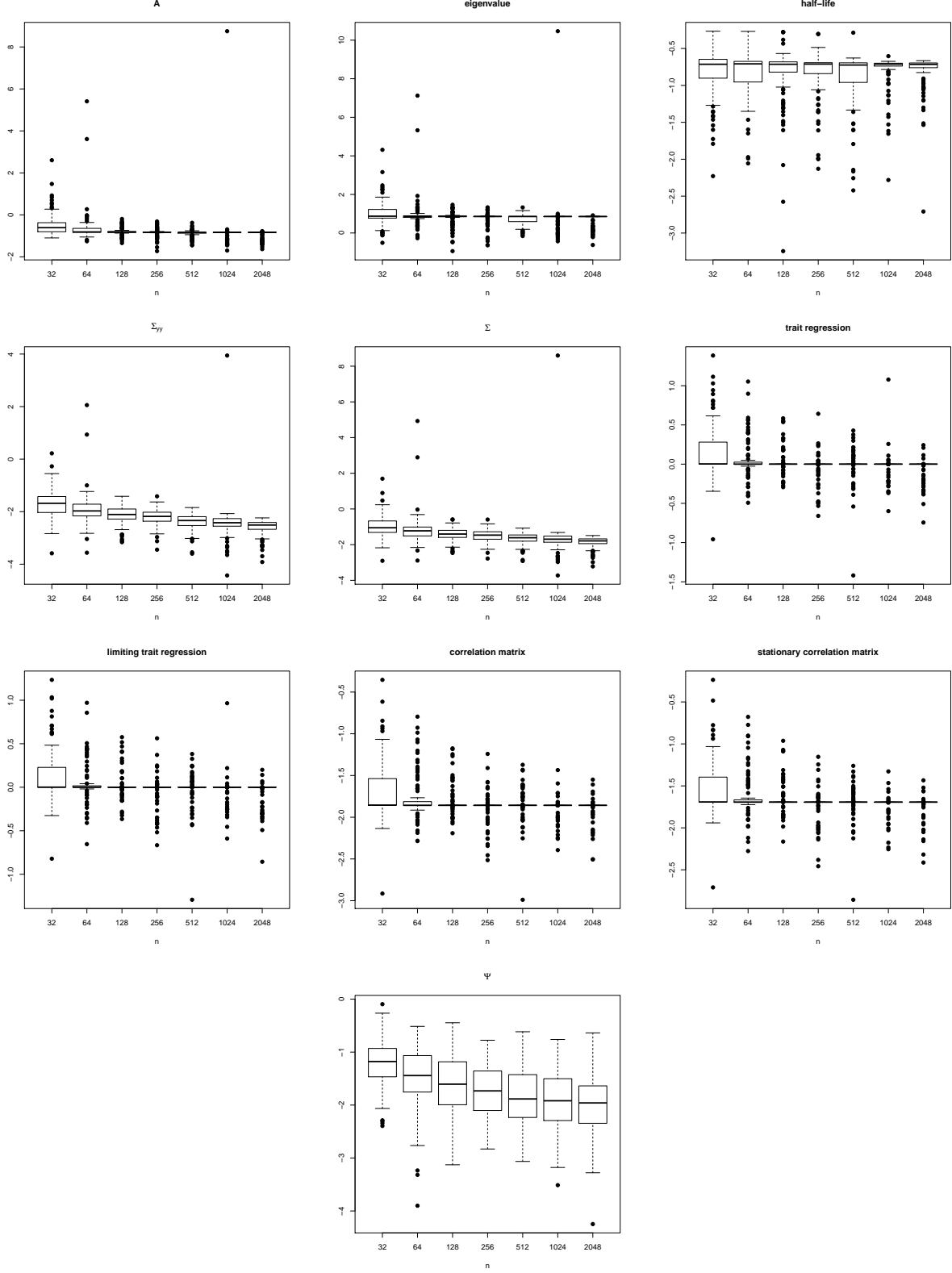


Figure S.5: Quality of parameter estimates under the OUOUS2P model (Tab. S.1). Boxplots of relative distances (Eq. 4) or Euclidean distances (if the true parameter is $\vec{0}$, $\vec{Y}(0)$, $\vec{\psi}$ cases, Eq. 3) between true matrix/vector parameter and estimated ones. The y -axes are on logarithmic scale, x -axes are the number of tips of the phylogeny. The boxplots come from 100 successful simulation-reestimation runs, see Tab. S.4.

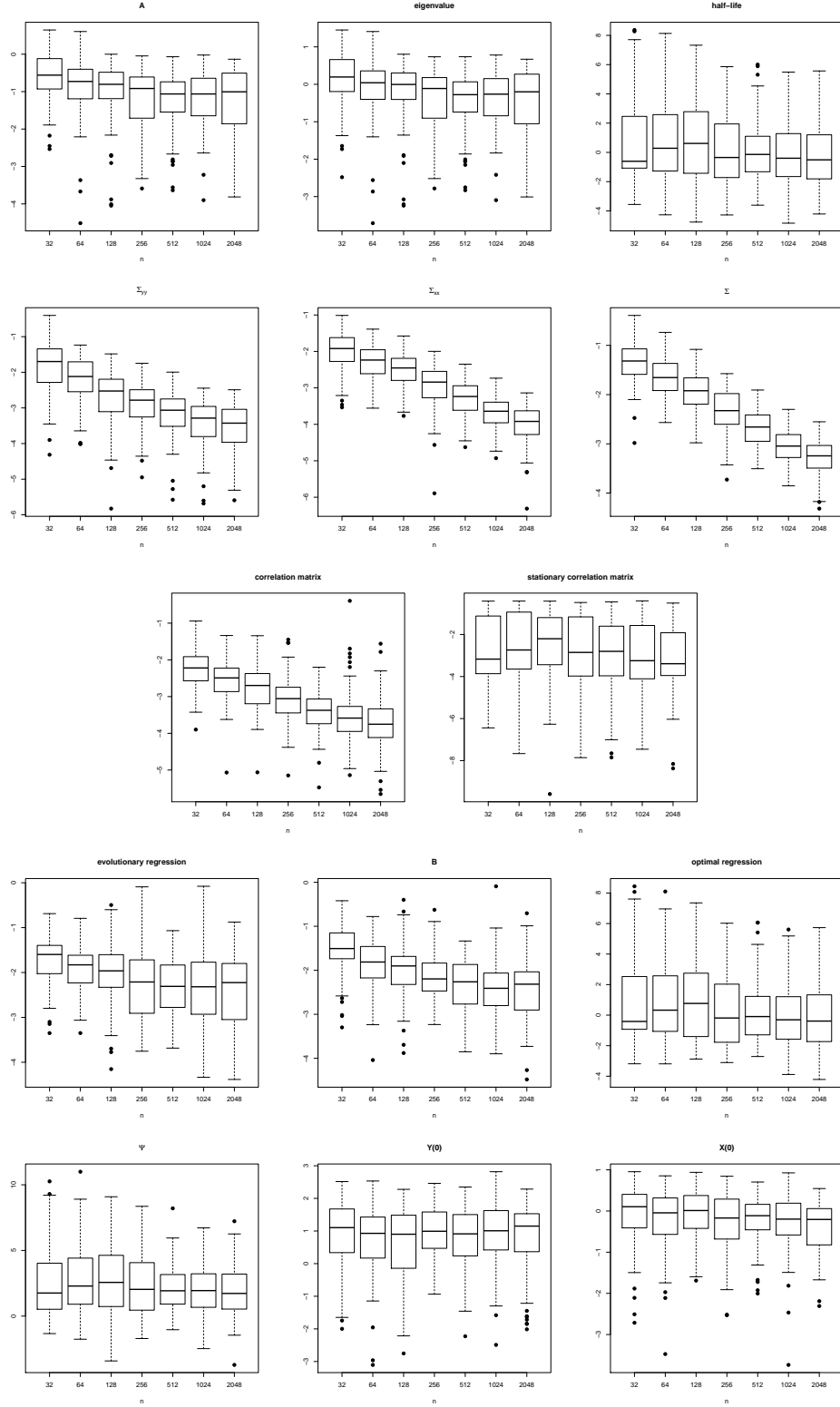


Figure S.6: Quality of parameter estimates under the OUBMs1 model (Tab. S.1) when \mathbf{Q} is estimated by maximum likelihood. Boxplots of relative distances (Eq. 4) or Euclidean distances (if the true parameter is $\vec{0}$, $\vec{Y}(0)$, $\vec{X}(0)$, $\vec{\psi}$ cases, Eq. 3) between true matrix/vector parameter and estimated ones. The y -axes are on logarithmic scale, x -axes are the number of tips of the phylogeny. The boxplots come from 100 successful simulation-reestimation runs, see Tab. S.4.

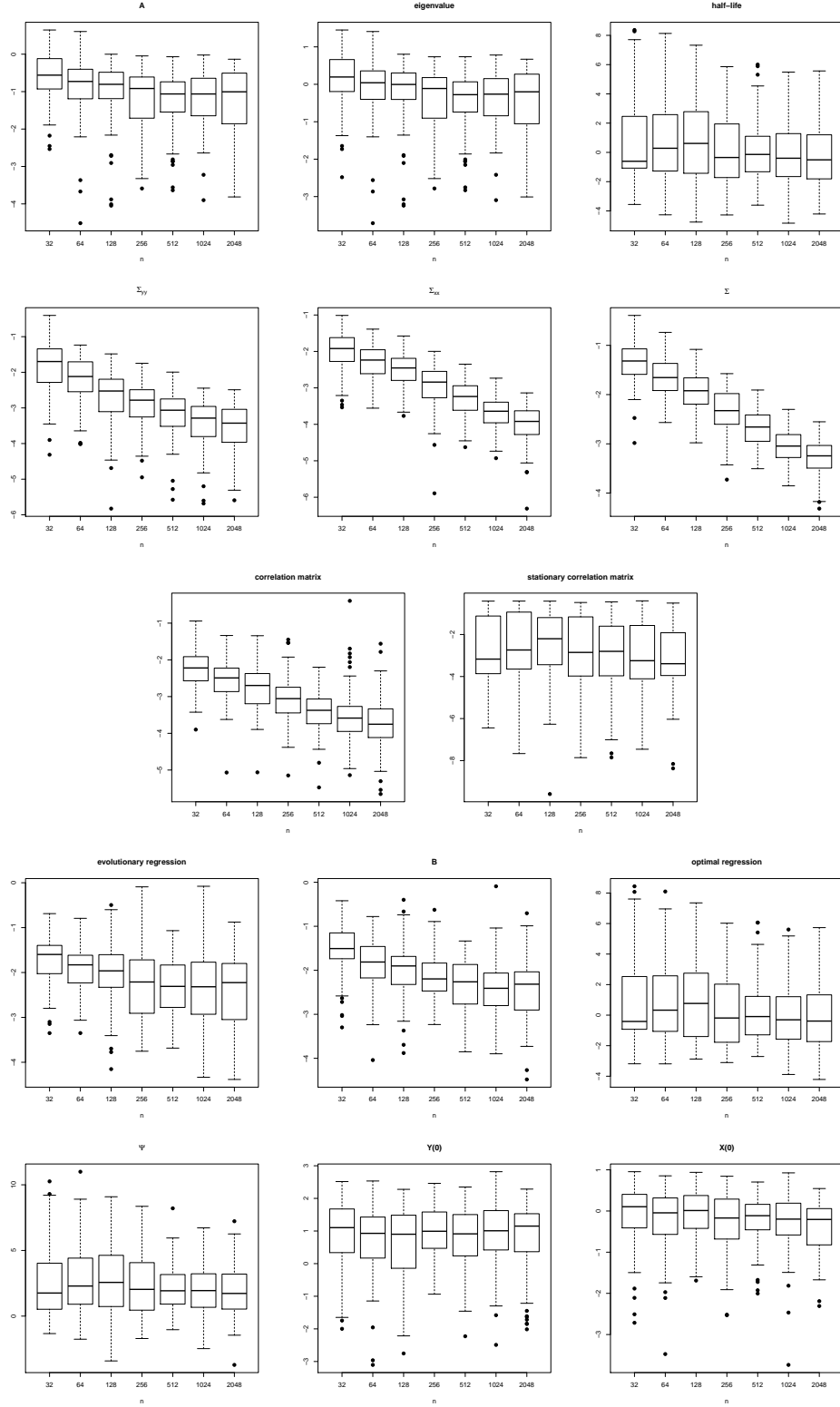


Figure S.7: Quality of parameter estimates under the OUBMs1P model (Tab. S.1) when \mathbf{Q} is estimated by maximum likelihood. Boxplots of relative distances (Eq. 4) or Euclidean distances (if the true parameter is $\vec{0}$, $\vec{Y}(0)$, $\vec{X}(0)$, $\vec{\psi}$ cases, Eq. 3) between true matrix/vector parameter and estimated ones. The y -axes are on logarithmic scale, x -axes are the number of tips of the phylogeny. The boxplots come from 100 successful simulation-reestimation runs, see Tab. S.4.

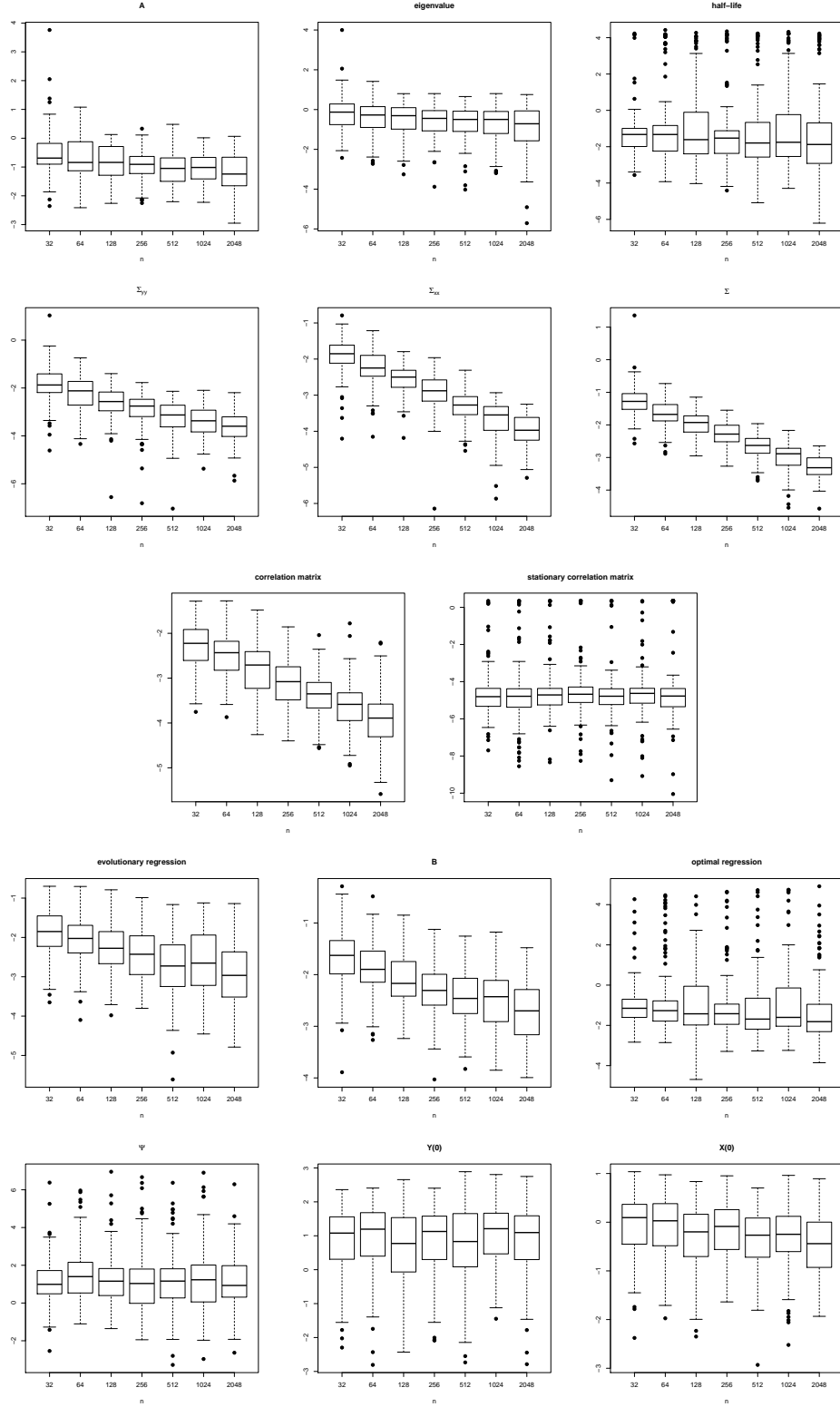


Figure S.8: Quality of parameter estimates under the OUBMs2 model (Tab. S.1) when \mathbf{Q} is estimated by maximum likelihood. Boxplots of relative distances (Eq. 4) or Euclidean distances (if the true parameter is $\vec{0}$, $\vec{Y}(0)$, $\vec{X}(0)$, $\vec{\psi}$ cases, Eq. 3) between true matrix/vector parameter and estimated ones. The y -axes are on logarithmic scale, x -axes are the number of tips of the phylogeny. The boxplots come from 100 successful simulation-reestimation runs, see Tab. S.4.

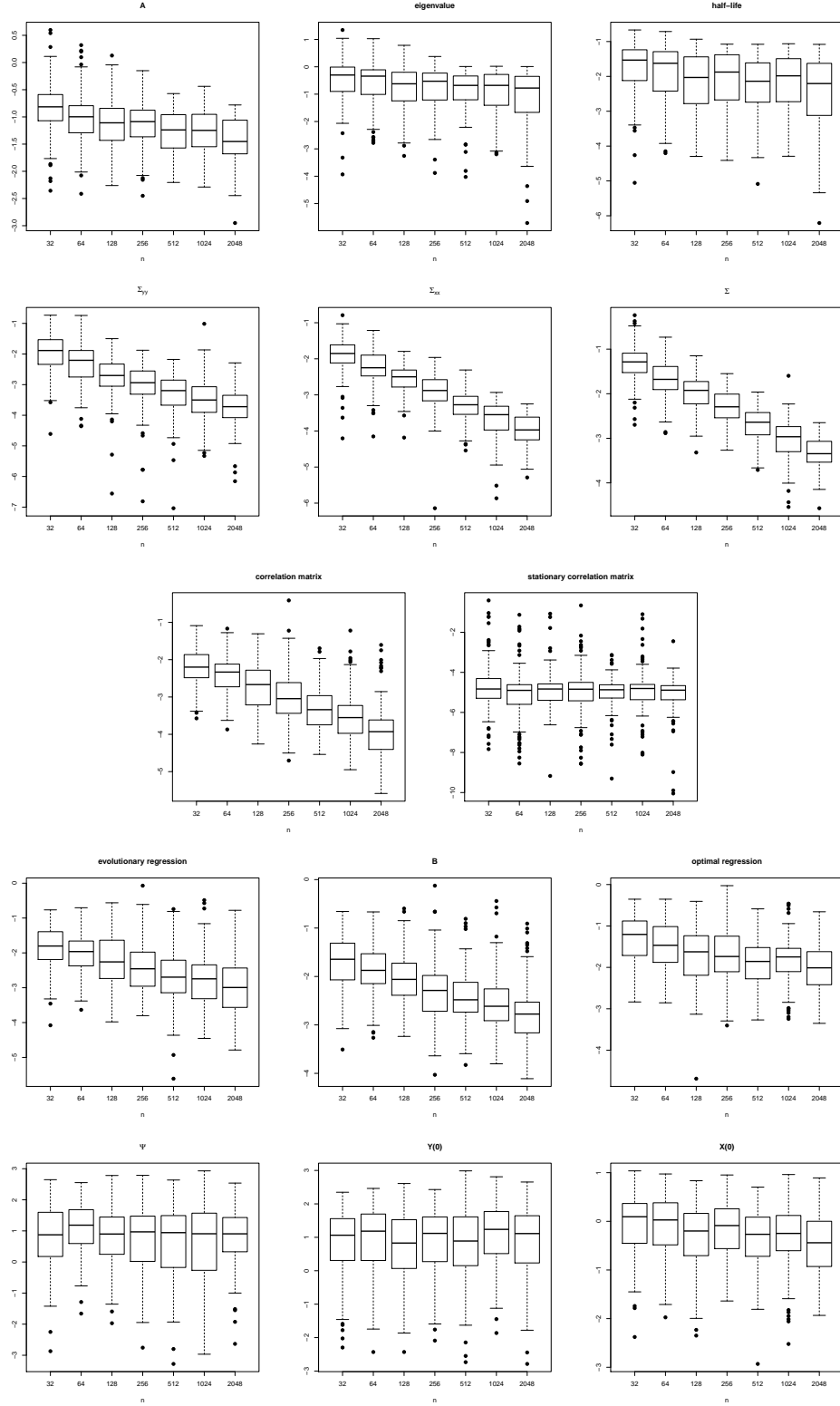


Figure S.9: Quality of parameter estimates under the OUBMs2P model (Tab. S.1) when \mathbf{Q} is estimated by maximum likelihood. Boxplots of relative distances (Eq. 4) or Euclidean distances (if the true parameter is $\vec{0}$, $\vec{Y}(0)$, $\vec{X}(0)$, $\vec{\psi}$ cases, Eq. 3) between true matrix/vector parameter and estimated ones. The y -axes are on logarithmic scale, x -axes are the number of tips of the phylogeny. The boxplots come from 100 successful simulation-reestimation runs, see Tab. S.4.

APPENDIX SD: SIMULATION-REESTIMATION STUDY BASED ON ESTIMATED PARAMETERS FOR FRUIT EVOLUTION IN *Ferula* (APIACEAE); SETUP AND RESULTS

In our “Example analysis 2” we estimated parameters under various models of OU based evolution for fruit evolution in *Ferula* (Apiaceae). Two OUOU models, 4 and 7 (Fig. 3), came out as worth looking into. The difference in AIC_c values is a little over 9 which puts the less supported model 4 (non-diagonal \mathbf{A} matrix) in a region where “evidence ... is inconclusive and value judgments for hypotheses in this region are equivocal” (caption Fig. 2 in, Burnham et al. 2011). Furthermore, there is strong biological intuition behind the second, less favoured model. However, if one considers the parameter estimates, then the estimated off-diagonals of \mathbf{A} in model 4 (\mathbf{A} for model 4U in Tab. S.5) are rather close to 0. This makes the drift matrix very close to that of the best found model 7 (\mathbf{A} for model 7U in Tab. S.5). To the best of our knowledge no study was ever performed to see whether so similar OU models are distinguishable. We therefore perform a simulation study to see if and how well is it possible to distinguish between the two models under the estimated parameters, given the sample size and measurement error. We also investigate the behaviour with growing tree size ($n \in \{128, 256, 512, 1024, 20148\}$), with and without measurement error. In Alg. S.1 we explain how measurement error was added to the simulated sample. In Tab. S.5, we present the various parameter sets under which the data were simulated and then, in Tab. S.6 the model identifiability capabilities for the different simulation model-reestimation model pairs (true and competing models). As in Appendix SC, this study was only possible thanks to the **PCMBaseC++** C++ backend and it took three weeks. All simulations were carried out using version 2.7 of **mvSLOUCH** in R version 3.6.1 running on an openSUSE 42.3 (x86_64) box with a 3.50GHz Intel[®] Xeon[®] CPU. It has to be noted that the actual number of simulation-reestimation repeats is

higher, see Tab. S.7 as some runs ended in error when estimating under the correct model. This is particularly visible when simulation was done under the BM model, where analytical estimators cannot move into a parameter region that mitigates the effect of short tip branches.

Algorithm S.1 Simulating measurement error variance for trait x for the simulation-reestimation study based on the *Ferula* analysis of our “Example analysis 2”.

- 1: $v_x :=$ measurement error vector from all species for variable x
 - 2: $df := 1/\text{mean}(v_x) + 2$
 - 3: **for** i in all tips of tree **do**
 - 4: draw measurement error variance of trait x in tip i as χ^2 with df degrees of freedom
 - 5: **end for**
-

Table S.5: Models estimated from Ferula data in our “Example analysis 2” under which simulation of the five dimensional trait was performed. The parameters are in terms of the general Eq. (1). This means that for the BM model $\vec{Y}(0) = \vec{X}(0)$ and $\Sigma_{yy} = \Sigma_{xx}$. The abbreviations correspond also to the model constraints on models under which estimation is performed in the simulation-reestimation study of this section. For OU models the ancestral state equals the optimum, $\vec{Y}(0) = \vec{\psi}$, hence the $\vec{\psi}$ parameter column is not shown.

Model	$\vec{Y}(0)$	A	Σ_{yy}				
BM	3.74	—	0.723	0	0	0	0
	4.412		0.095	0.537	0	0	0
	2.841		−0.105	0.169	0.93	0	0
	5.064		−0.069	0.297	0.085	0.321	0
	2.62		0.04	0.263	0.651	0.137	0.535
BMm	3.882	—	0.124	0	0	0	0
	4.452		1.003	0.874	0	0	0
	2.909		0.133	0.172	0.083	0	0
	5.056		1.071	0.109	−0.304	0.322	0
	2.606		0.101	0.056	0.043	0.018	0.033

Model	$\vec{Y}(0)$	A	Σ_{yy}
4D	$\begin{bmatrix} 3.912 \\ 4.408 \\ 2.945 \\ 5.049 \\ 2.619 \end{bmatrix}$	$\begin{bmatrix} 2.919 & -1.029 & 0 & 0 & 0 \\ -1.282 & 2.073 & 0 & 0 & 0 \\ 0 & 0 & 2.011 & -0.409 & 0 \\ 0 & 0 & -0.975 & 3.396 & 0 \\ 0 & 0 & 0 & 0 & 3.637 \end{bmatrix}$	$\begin{bmatrix} 1.228 & 0 & 0 & 0 & 0 \\ 0 & 0.648 & 0 & 0 & 0 \\ 0 & 0 & 1.546 & 0 & 0 \\ 0 & 0 & 0 & 0.741 & 0 \\ 0 & 0 & 0 & 0 & 1.867 \end{bmatrix}$
4Dm	$\begin{bmatrix} 3.881 \\ 4.446 \\ 2.908 \\ 5.054 \\ 2.606 \end{bmatrix}$	$\begin{bmatrix} 6.57 & 1.31 & 0 & 0 & 0 \\ -1.969 & 4.895 & 0 & 0 & 0 \\ 0 & 0 & 785.641 & 0.926 & 0 \\ 0 & 0 & -3.677 & 0.027 & 0 \\ 0 & 0 & 0 & 0 & 18.77 \end{bmatrix}$	$\begin{bmatrix} 0.969 & 0 & 0 & 0 & 0 \\ 0 & 1.687 & 0 & 0 & 0 \\ 0 & 0 & 0.611 & 0 & 0 \\ 0 & 0 & 0 & 1.255 & 0 \\ 0 & 0 & 0 & 0 & 0.088 \end{bmatrix}$

Model	$\vec{Y}(0)$	A	Σ_{yy}
4U	3.868	1.509	$S_{yytype} = "UpperTri"$ $\begin{bmatrix} 0.92 & 0.43 & -0.045 & -0.0392 & 0.213 \\ 0 & 0.549 & -0.026 & 0.383 & 0.311 \\ 0 & 0 & 0.811 & -0.047 & 1.143 \\ 0 & 0 & 0 & 0.481 & 0.293 \\ 0 & 0 & 0 & 0 & 1.275 \end{bmatrix}$
	4.414	-0.01	
	2.911	0	
	5.054	0	
	2.624	0	
		0	
4Um	3.881	15314.644	$S_{yytype} = "UpperTri"$ $\begin{bmatrix} 0.026 & 2.61 & 6.215 & 0.436 & 4.826 \\ 0 & 0.7 & 3.856 & 0.049 & 2.256 \\ 0 & 0 & 2.834 & 2.119 & 3.921 \\ 0 & 0 & 0 & 0.03 & 1.408 \\ 0 & 0 & 0 & 0 & 0.393 \end{bmatrix}$
	4.446	-1.103	
	2.908	0	
	5.054	0	
	2.606	0	
		0	

Model	$\vec{Y}(0)$	A	Σ_{yy}
7D	$\begin{bmatrix} 3.853 \\ 4.412 \\ 2.927 \\ 5.057 \\ 2.618 \end{bmatrix}$	$\begin{bmatrix} 1.308 & 0 & 0 & 0 & 0 \\ 0 & 1.194 & 0 & 0 & 0 \\ 0 & 0 & 1.623 & 0 & 0 \\ 0 & 0 & 0 & 3.265 & 0 \\ 0 & 0 & 0 & 0 & 4.35 \end{bmatrix}$	$\begin{bmatrix} 1.08 & 0 & 0 & 0 & 0 \\ 0 & 0.805 & 0 & 0 & 0 \\ 0 & 0 & 1.689 & 0 & 0 \\ 0 & 0 & 0 & 0.843 & 0 \\ 0 & 0 & 0 & 0 & 1.796 \end{bmatrix}$
7Dm	$\begin{bmatrix} 3.881 \\ 4.446 \\ 2.908 \\ 5.053 \\ 2.606 \end{bmatrix}$	$\begin{bmatrix} 13.636 & 0 & 0 & 0 & 0 \\ 0 & 52.581 & 0 & 0 & 0 \\ 0 & 0 & 5.187 & 0 & 0 \\ 0 & 0 & 0 & 2.308 & 0 \\ 0 & 0 & 0 & 0 & 4.487 \end{bmatrix}$	$\begin{bmatrix} 0.876 & 0 & 0 & 0 & 0 \\ 0 & 2.635 & 0 & 0 & 0 \\ 0 & 0 & 0.04 & 0 & 0 \\ 0 & 0 & 0 & 1.26 & 0 \\ 0 & 0 & 0 & 0 & 0.206 \end{bmatrix}$

Model	$\vec{Y}(0)$	A	Σ_{yy}
7U	$\begin{bmatrix} 3.868 \\ 4.415 \\ 2.912 \\ 5.055 \\ 2.624 \end{bmatrix}$	$\begin{bmatrix} 1.509 & 0 & 0 & 0 & 0 \\ 0 & 1.49 & 0 & 0 & 0 \\ 0 & 0 & 1.221 & 0 & 0 \\ 0 & 0 & 0 & 1.202 & 0 \\ 0 & 0 & 0 & 0 & 1.206 \end{bmatrix}$	$\begin{matrix} Syytype="UpperTri" \\ \begin{bmatrix} 0.92 & 0.43 & -0.045 & -0.039 & 0.213 \\ 0 & 0.549 & -0.026 & 0.383 & 0.311 \\ 0 & 0 & 0.811 & -0.047 & 1.143 \\ 0 & 0 & 0 & 0.481 & 0.293 \\ 0 & 0 & 0 & 0 & 1.275 \end{bmatrix} \end{matrix}$
7Um	$\begin{bmatrix} 3.881 \\ 4.447 \\ 2.908 \\ 5.053 \\ 2.606 \end{bmatrix}$	$\begin{bmatrix} 268.692 & 0 & 0 & 0 & 0 \\ 0 & 287.055 & 0 & 0 & 0 \\ 0 & 0 & 3841.808 & 0 & 0 \\ 0 & 0 & 0 & 219.069 & 0 \\ 0 & 0 & 0 & 0 & 509.553 \end{bmatrix}$	$\begin{matrix} Syytype="UpperTri" \\ \begin{bmatrix} 0.874 & 4.703 & 4.611 & 0.219 & -2.239 \\ 0 & 9.663 & 10.438 & 3.382 & -2.72 \\ 0 & 0 & 4.212 & 2.449 & -3.328 \\ 0 & 0 & 0 & 7.915 & 2.902 \\ 0 & 0 & 0 & 0 & 0.195 \end{bmatrix} \end{matrix}$

Table S.6: Model identifiability capabilities of **mvSLOUCH**. The columns are the true model under which the data was simulated (Tab. S.5, abbreviations described therein). The rows are the models under which estimation was done, abbreviations are the same as for the columns. The values in the table are the fraction of times the true (under which data was simulated) model was chosen over the alternative (row) model by AIC_c . Each fraction is the result of successful 100 simulation-reestimation repeats. It has to be noted that the actual number of simulation-reestimation repeats is higher, see Tab. S.7 as some runs ended in error when estimating under the correct model. A ♣ means that the simulated and estimated under models are the same, hence a value there is meaningless. If there is an entry in brackets it indicates that estimation under the true model setup was done with the constraint that the values on **A**'s diagonal are positive. The rows with m concern setups with measurement error present (both for simulation and estimation), and the P means that positive values on the diagonal of **A** were assumed under the alternative (row) model. The number in bold at the top of each subtable is the number of tips. The first subtable, with 78 tips, was done with a fixed tree and measurement error, the original Ferula tree and exactly the same measurement error as in the Ferula data was used. In all the other tables (tree sizes) pure birth trees were simulated and measurement error was simulated according to Alg. S.1.

78	BM/BMm	4D/4Dm	4U/4Um	7D/7Dm	7U/7Um
BM	♣	0.03 (0.53)	0.05 (0.05)	0.01 (0.72)	0.07 (0.07)
BMm	♣	0.89 (0.94)	0.35 (0.45)	0.98 (0.97)	0.7 (0.76)
4D	1	♣	0.93 (0.99)	0.88 (1)	0.92 (0.97)
4Dm	0.07	♣	0.05 (0.05)	0.96 (0.95)	0.04 (0.04)
4D P	1	♣	0.9 (0.96)	0.03 (0.97)	0.89 (0.95)
4Dm P	0.01	♣	0.04 (0.04)	0.96 (0.95)	0.04 (0.04)

4U	1	0.98 (1)	♣	0.97 (1)	0.95 (0.91)
4Um	0.66	0.89 (0.94)	♣	0.98 (0.97)	0.96 (0.96)
4U P	1	0.02 (0.95)	♣	0.15 (1)	0.25 (0.96)
4Um P	0.55	0.89 (0.94)	♣	0.98 (0.98)	0.96 (0.96)
7D	1	0.12 (0.99)	0.93 (0.98)	♣	0.89 (0.95)
7Dm	0.07	0.05 (0.06)	0.02 (0.03)	♣	0.04 (0.03)
7D P	1	0 (0)	0.89 (0.96)	♣	0.95 (0.91)
7Dm P	0.01	0.05 (0.06)	0 (0.01)	♣	0.01 (0.01)
7U	1	0.96 (1)	0.03 (0.78)	0.95 (1)	♣
7Um	0.24	0.88 (0.94)	0.01 (0.02)	0.98 (0.95)	♣
7U P	1	0.01 (0.76)	0.1 (0.03)	0.05 (1)	♣
7Um P	0.18	0.88 (0.94)	0.01 (0.02)	0.98 (0.95)	♣

128	BM/BMm	4D/4Dm	4U/4Um	7D/7Dm	7U/7Um
BM	♣	0.01 (0.63)	0.06 (0.07)	0.03 (0.71)	0.05 (0.06)
BMm	♣	0.25 (0.5)	0.07 (0.09)	0.38 (0.84)	0.26 (0.65)
4D	1	♣	0.96 (1)	0.73 (0.99)	0.95 (0.98)
4Dm	0.82	♣	0.13 (0.51)	0.78 (0.95)	0.1 (0.99)
4D P	1	♣	0.94 (0.98)	0 (0.97)	0.95 (0.98)

4Dm P	0.79	♣	0.07 (0.14)	0.09 (0.75)	0.03 (0.06)
4U	1	0.96 (1)	♣	0.92 (0.99)	1 (0.78)
4Um	0.85	0.89 (0.9)	♣	0.95 (0.96)	0.79 (0.99)
4U P	1	0.01 (0.93)	♣	0.01 (0.97)	0.35 (0.99)
4Um P	0.81	0.2 (0.9)	♣	0.08 (0.94)	0.02 (0.9)
7D	1	0.04 (1)	0.95 (0.99)	♣	0.95 (0.98)
7Dm	0.76	0.19 (0.84)	0.06 (0.34)	♣	0.07 (0.99)
7D P	1	0 (0)	0.94 (0.98)	♣	0.95 (0.98)
7Dm P	0.71	0.07 (0.28)	0.04 (0.08)	♣	0.03 (0.04)
7U	1	0.97 (1)	0.02 (0.71)	0.85 (0.99)	♣
7Um	0.83	0.84 (0.89)	0.31 (0.77)	0.92 (0.96)	♣
7U P	1	0 (0.79)	0.21 (0.01)	0 (0.99)	♣
7Um P	0.81	0.17 (0.8)	0.08 (0.28)	0.03 (0.91)	♣

256	BM/BMm	4D/4Dm	4U/4Um	7D/7Dm	7U/7Um
BM	♣	0.06 (0.66)	0.16 (0.17)	0.02 (0.5)	0.1 (0.11)
BMm	♣	0.39 (0.73)	0.05 (0.07)	0.44 (0.92)	0.39 (0.85)
4D	1	♣	0.8 (0.97)	0.72 (1)	0.86 (0.99)
4Dm	0.77	♣	0.27 (0.68)	0.68 (0.96)	0.29 (0.99)

4D P	1	♣	0.79 (0.92)	0 (1)	0.86 (0.99)
4Dm P	0.67	♣	0.12 (0.39)	0.04 (0.71)	0.03 (0.2)
4U	1	0.97 (1)	♣	0.86 (1)	0.94 (0.73)
4Um	0.77	0.86 (0.99)	♣	0.87 (0.96)	0.69 (1)
4U P	1	0.01 (0.89)	♣	0 (0.99)	0.37 (0.99)
4Um P	0.82	0.1 (0.89)	♣	0.02 (0.94)	0 (0.67)
7D	1	0.16 (1)	0.8 (0.97)	♣	0.85 (0.99)
7Dm	0.71	0.37 (0.92)	0.15 (0.54)	♣	0.19 (1)
7D P	1	0.01 (0.01)	0.8 (0.93)	♣	0.86 (0.99)
7Dm P	0.74	0.05 (0.28)	0.1 (0.21)	♣	0.01 (0.12)
7U	1	0.91 (1)	0.06 (0.69)	0.83 (1)	♣
7Um	0.74	0.76 (0.98)	0.22 (0.68)	0.78 (0.96)	♣
7U P	1	0.01 (0.78)	0.24 (0.03)	0 (1)	♣
7Um P	0.78	0.05 (0.75)	0.12 (0.36)	0.02 (0.89)	♣

512	BM/BMm	4D/4Dm	4U/4Um	7D/7Dm	7U/7Um
BM	♣	0.06 (0.6)	0.24 (0.24)	0.05 (0.33)	0.27 (0.26)
BMm	♣	0.44 (0.72)	0.05 (0.06)	0.46 (0.95)	0.41 (0.87)
4D	1	♣	0.72 (0.92)	0.54 (0.98)	0.73 (0.9)

4Dm	0.81	♣	0.38 (0.63)	0.63 (0.99)	0.34 (1)
4D P	1	♣	0.69 (0.89)	0.02 (0.99)	0.71 (0.87)
4Dm P	0.76	♣	0.17 (0.37)	0.03 (0.63)	0.05 (0.32)
4U	1	0.93 (1)	♣	0.75 (0.99)	0.86 (0.62)
4Um	0.83	0.7 (0.93)	♣	0.79 (0.99)	0.66 (1)
4U P	1	0.02 (0.87)	♣	0.02 (0.98)	0.42 (0.93)
4Um P	0.8	0.07 (0.69)	♣	0.01 (0.84)	0 (0.62)
7D	1	0.1 (1)	0.72 (0.91)	♣	0.73 (0.92)
7Dm	0.81	0.41 (0.91)	0.4 (0.61)	♣	0 (0.98)
7D P	1	0.01 (0.01)	0.7 (0.88)	♣	0.72 (0.87)
7Dm P	0.7	0.01 (0.44)	0.17 (0.3)	♣	0.02 (0.23)
7U	1	0.91 (1)	0.16 (0.58)	0.72 (0.98)	♣
7Um	0.82	0.62 (0.95)	0.48 (0.7)	0.74 (0.99)	♣
7U P	1	0.04 (0.86)	0.41 (0.06)	0.01 (0.98)	♣
7Um P	0.82	0.04 (0.65)	0.25 (0.43)	0.01 (0.83)	♣
<hr/>					
1024	BM/BMm	4D/4Dm	4U/4Um	7D/7Dm	7U/7Um
BM	♣	0.11 (0.59)	0.46 (0.46)	0.1 (0.2)	0.52 (0.52)
BMm	♣	0.49 (0.76)	0.07 (0.08)	0.55 (0.94)	0.56 (0.96)

4D	1	♣	0.56 (0.92)	0.58 (0.96)	0.5 (0.8)
4Dm	0.68	♣	0.46 (0.68)	0.57 (0.98)	0.44 (0.99)
4D P	1	♣	0.5 (0.79)	0.03 (0.96)	0.39 (0.67)
4Dm P	0.63	♣	0.25 (0.45)	0.04 (0.54)	0.05 (0.49)
4U	1	0.93 (0.99)	♣	0.74 (0.99)	0.62 (0.6)
4Um	0.72	0.65 (0.96)	♣	0.61 (0.98)	0.53 (0.99)
4U P	1	0.03 (0.93)	♣	0.06 (0.97)	0.41 (0.87)
4Um P	0.7	0.06 (0.59)	♣	0.01 (0.71)	0.02 (0.61)
7D	1	0.16 (0.99)	0.56 (0.9)	♣	0.46 (0.83)
7Dm	0.66	0.43 (0.93)	0.45 (0.67)	♣	0.43 (0.99)
7D P	1	0.01 (0.01)	0.51 (0.81)	♣	0.39 (0.68)
7Dm P	0.64	0.06 (0.42)	0.25 (0.49)	♣	0.04 (0.34)
7U	1	0.83 (0.98)	0.21 (0.5)	0.74 (0.97)	♣
7Um	0.7	0.72 (0.95)	0.42 (0.64)	0.65 (0.98)	♣
7U P	1	0.04 (0.89)	0.49 (0.14)	0.08 (0.99)	♣
7Um P	0.63	0.07 (0.55)	0.32 (0.54)	0.02 (0.70)	♣
<hr/>					
2048	BM/BMm	4D/4Dm	4U/4Um	7D/7Dm	7U/7Um
BM	♣	0.24 (0.57)	0.82 (0.82)	0.15 (0.17)	0.7 (0.71)

BMm	♣	0.63 (0.83)	0.05 (0.05)	0.62 (0.9)	0.59 (0.9)
4D	1	♣	0.29 (0.74)	0.5 (0.93)	0.31 (0.75)
4Dm	0.65	♣	0.5 (0.62)	0.59 (0.93)	0.41 (0.97)
4D P	1	♣	0.13 (0.6)	0.08 (0.93)	0.23 (0.66)
4Dm P	0.53	♣	0.45 (0.5)	0.02 (0.47)	0.05 (0.44)
4U	1	0.77 (0.93)	♣	0.67 (0.95)	0.6 (0.71)
4Um	0.62	0.62 (0.97)	♣	0.64 (0.93)	0.52 (0.97)
4U P	1	0.1 (0.88)	♣	0.11 (0.94)	0.29 (0.82)
4Um P	0.54	0.05 (0.6)	♣	0.03 (0.61)	0.02 (0.49)
7D	1	0.29 (0.9)	0.24 (0.76)	♣	0.3 (0.75)
7Dm	0.61	0.5 (0.95)	0.55 (0.61)	♣	0.41 (0.97)
7D P	1	0.06 (0.07)	0.11 (0.59)	♣	0.22 (0.64)
7Dm P	0.52	0.02 (0.53)	0.32 (0.37)	♣	0.04 (0.39)
7U	1	0.8 (0.92)	0.45 (0.71)	0.66 (0.95)	♣
7Um	0.58	0.57 (0.99)	0.46 (0.51)	0.61 (0.93)	♣
7U P	1	0.12 (0.89)	0.3 (0.25)	0.14 (0.93)	♣
7Um P	0.52	0.04 (0.6)	0.36 (0.36)	0.03 (0.62)	♣

Table S.7: Number of simulation runs performed to obtain 100 (per cell) estimation results that are presented in Tab. S.6. The reruns are the results of errors when estimating under the model under which the data was simulated, hence no comparison was possible. For $n = 78$ tips, the simulations were done with a fixed tree and measurement error, the original Ferula tree and exactly the same measurement error as in the Ferula data is used. For all other n (tree sizes) pure birth trees are simulated and measurement error was simulated according to Alg. S.1.

n	BM	4D	4U	7D	7U
78	106	100	101	102	101
128	108	101	101	101	102
256	125	100	103	101	101
512	156	103	103	101	103
1024	255	104	106	105	106
2048	881	112	113	113	120

n	BMm	4Dm	4Um	7Dm	7Um
78	100	100	100	100	100
128	100	100	100	100	100
256	100	100	100	100	100
512	100	100	101	100	100
1024	100	101	100	100	100
2048	100	101	101	100	100

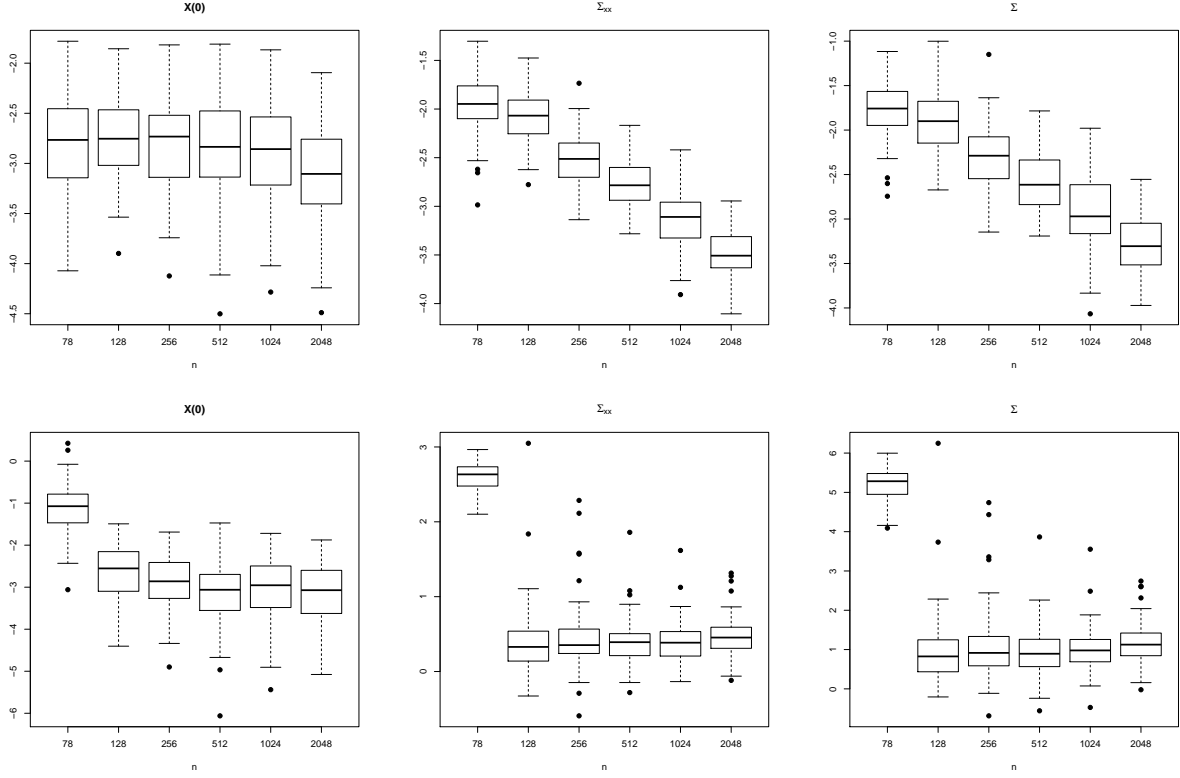


Figure S.10: Quality of parameter estimates under the BM model (Tab. S.5). Boxplots of relative distances (Eq. 4) between true matrix/vector parameter and estimated ones. The y -axes are on logarithmic scale, x -axes are the number of tips of the phylogeny. The boxplots come from 100 successful simulation-reestimation runs, see Tab. S.7. Top: no measurement error in analyses, bottom: with measurement error; left: $\vec{X}(0)$, centre: Σ_{xx} , right: Σ . For $n = 78$ tips, the simulations were done with a fixed tree and measurement error, the original *Ferula* tree and exactly the same measurement error as in the *Ferula* data is used. For all other n (tree sizes) pure birth trees are simulated and measurement error was simulated according to Alg. S.1.

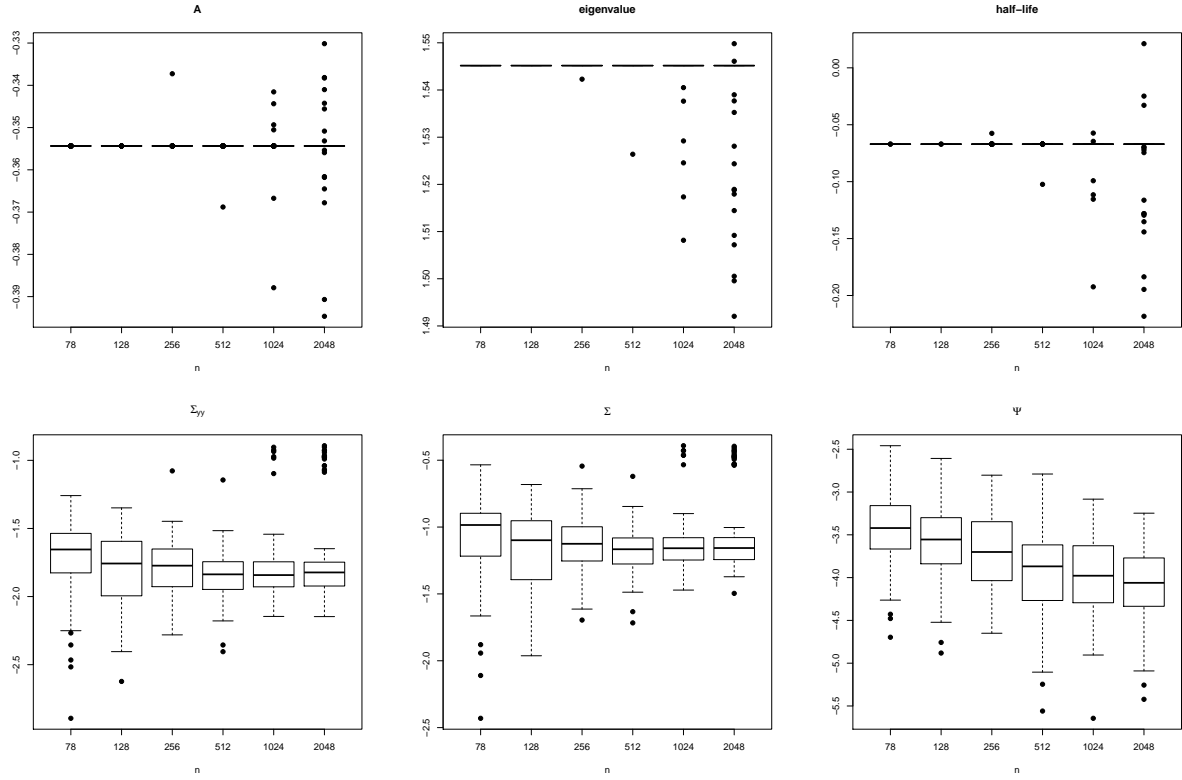


Figure S.11: Quality of parameter estimates under model 4D (model 4, diagonal Σ_{yy} , without measurement error, Tab. S.5). Boxplots of relative distances between true matrix/vector parameter and estimated ones. The y -axes are on logarithmic scale, x -axes are the number of tips of the phylogeny. The boxplots come from 100 successful simulation-reestimation runs, see Tab. S.7. For $n = 78$ tips, the simulations were done with the same tree and measurement error as in the *Ferula* data. For all other n (tree sizes) pure birth trees are simulated and measurement error was simulated according to Alg. S.1.

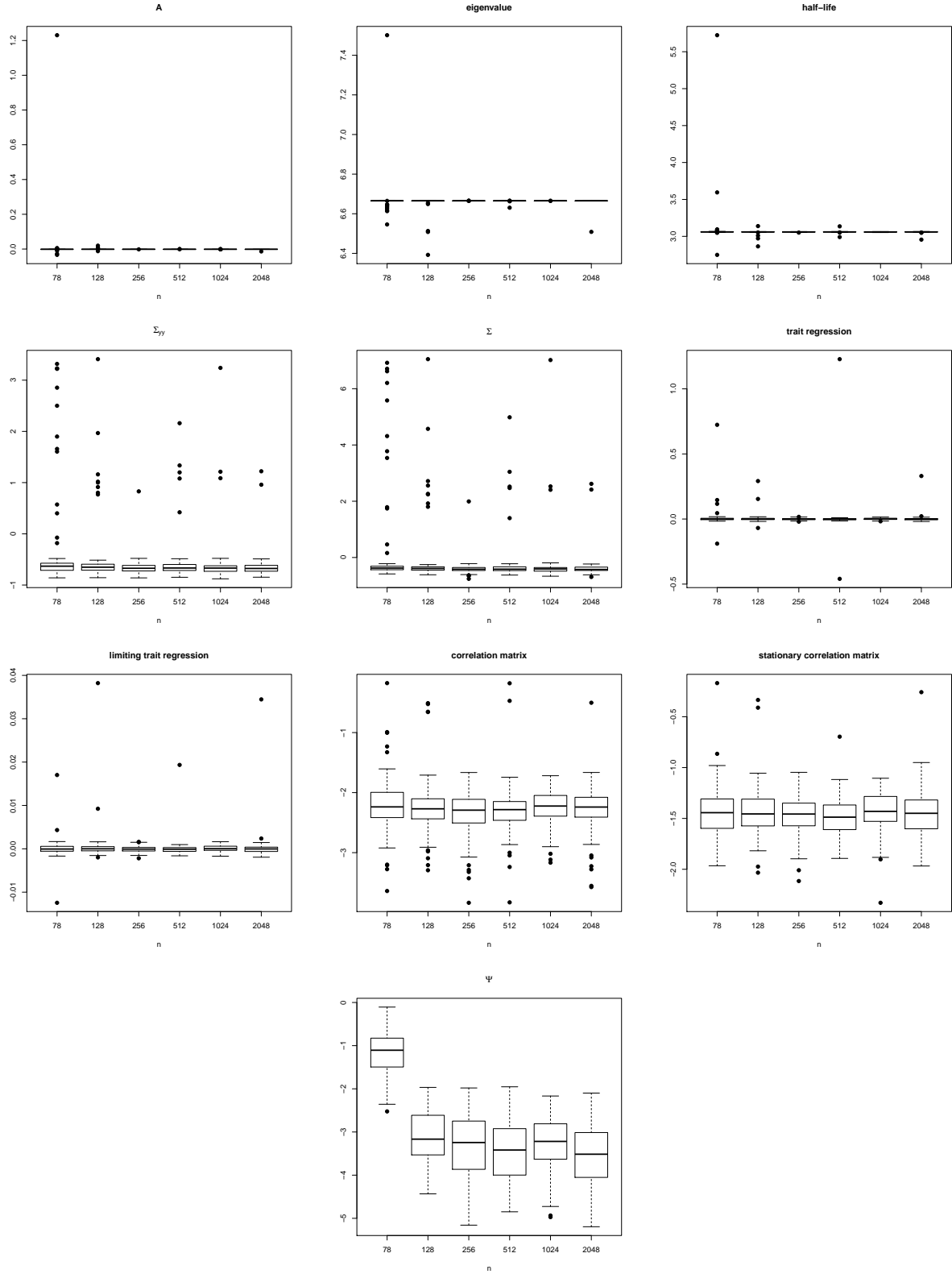


Figure S.12: Quality of parameter estimates under model 4Dm (model 4, diagonal Σ_{yy} , with measurement error, Tab. S.5). Boxplots of relative distances between true matrix/vector parameter and estimated ones. The y -axes are on logarithmic scale, x -axes are the number of tips of the phylogeny. The boxplots come from 100 successful simulation-reestimation runs, see Tab. S.7. For $n = 78$ tips, the simulations were done with the same tree and measurement error as in the *Ferula* data. For all other n (tree sizes) pure birth trees are simulated and measurement error was simulated according to Alg. S.1.

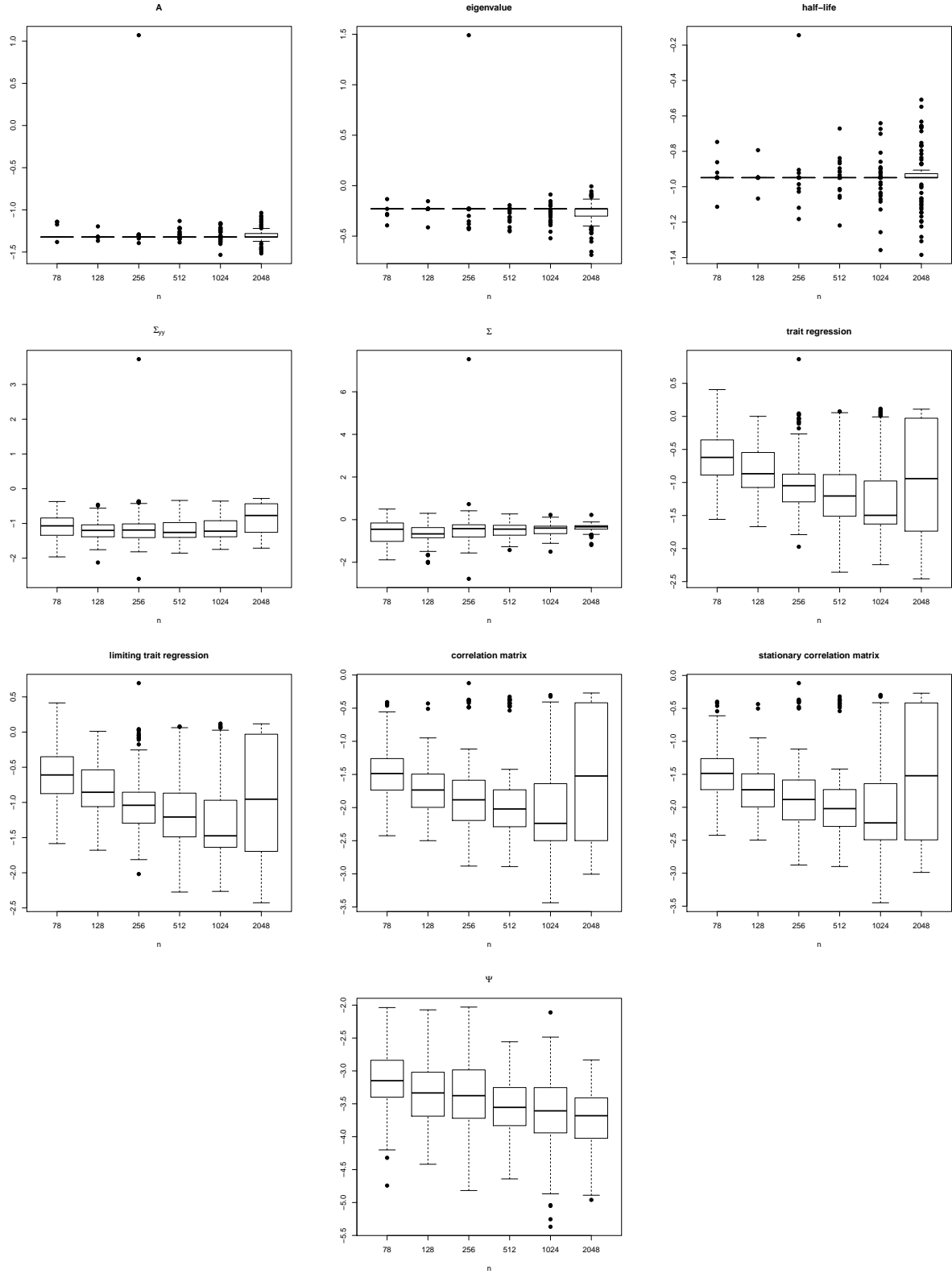


Figure S.13: Quality of parameter estimates under model 4U (model 4, upper triangular Σ_{yy} , without measurement error, Tab. S.5). Boxplots of relative distances between true matrix/vector parameter and estimated ones.⁴⁶ The y -axes are on logarithmic scale, x -axes are the number of tips of the phylogeny. The boxplots come from 100 successful simulation-reestimation runs, see Tab. S.7. For $n = 78$ tips, the simulations were done with the same tree and measurement error as in the *Ferula* data. For all other n (tree sizes) pure birth trees are simulated and measurement error was simulated according to Alg. S.1.

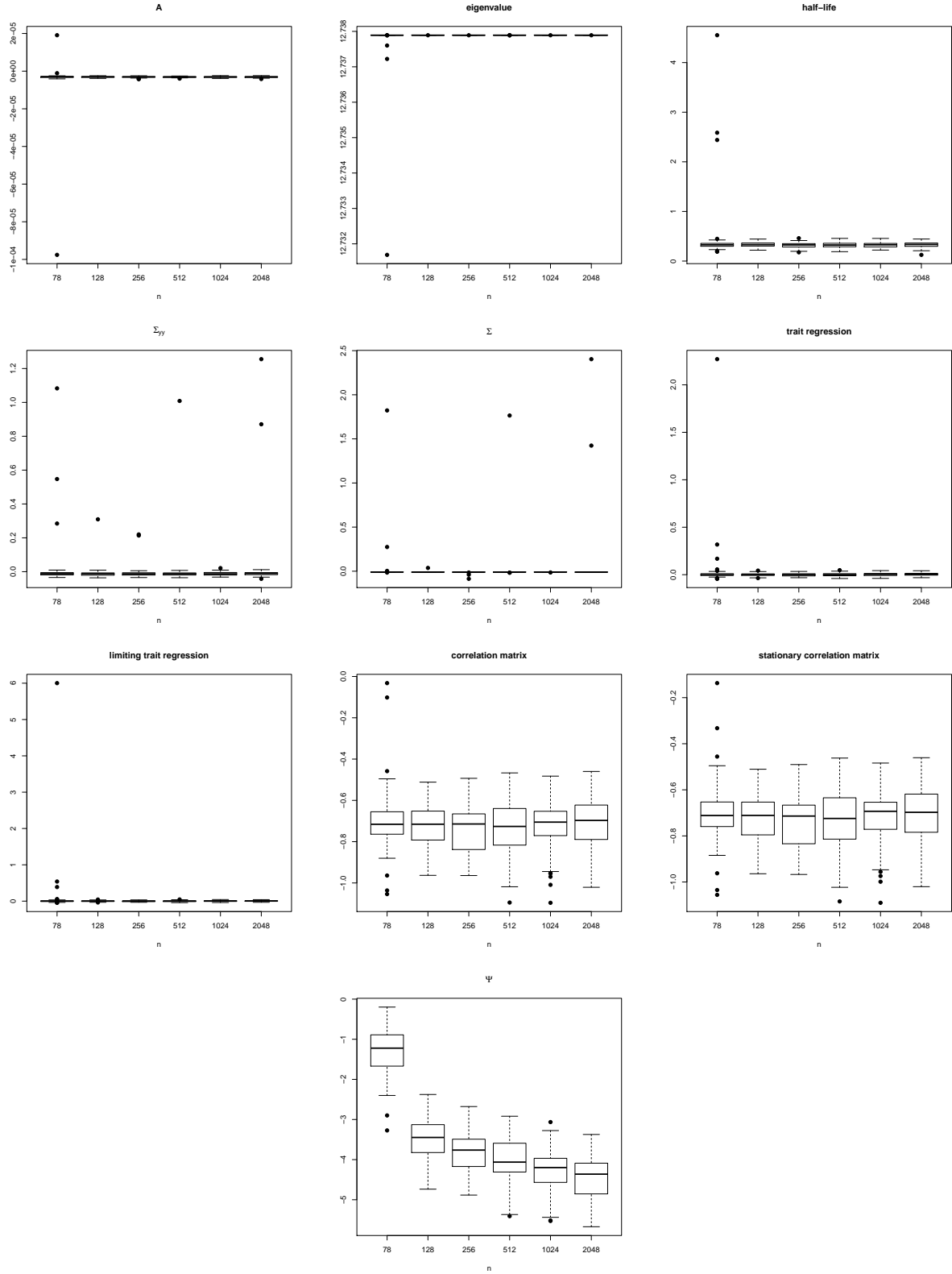


Figure S.14: Quality of parameter estimates under model 4Um (model 4, upper triangular Σ_{yy} , with measurement error, Tab. S.5). Boxplots of relative distances between true matrix/vector parameter and estimated ones. The y -axes are on logarithmic scale, x -axes are the number of tips of the phylogeny. The boxplots come from 100 successful simulation-reestimation runs, see Tab. S.7. For $n = 78$ tips, the simulations were done with the same tree and measurement error as in the *Ferula* data. For all other n (tree sizes) pure birth trees are simulated and measurement error was simulated according to Alg. S.1.

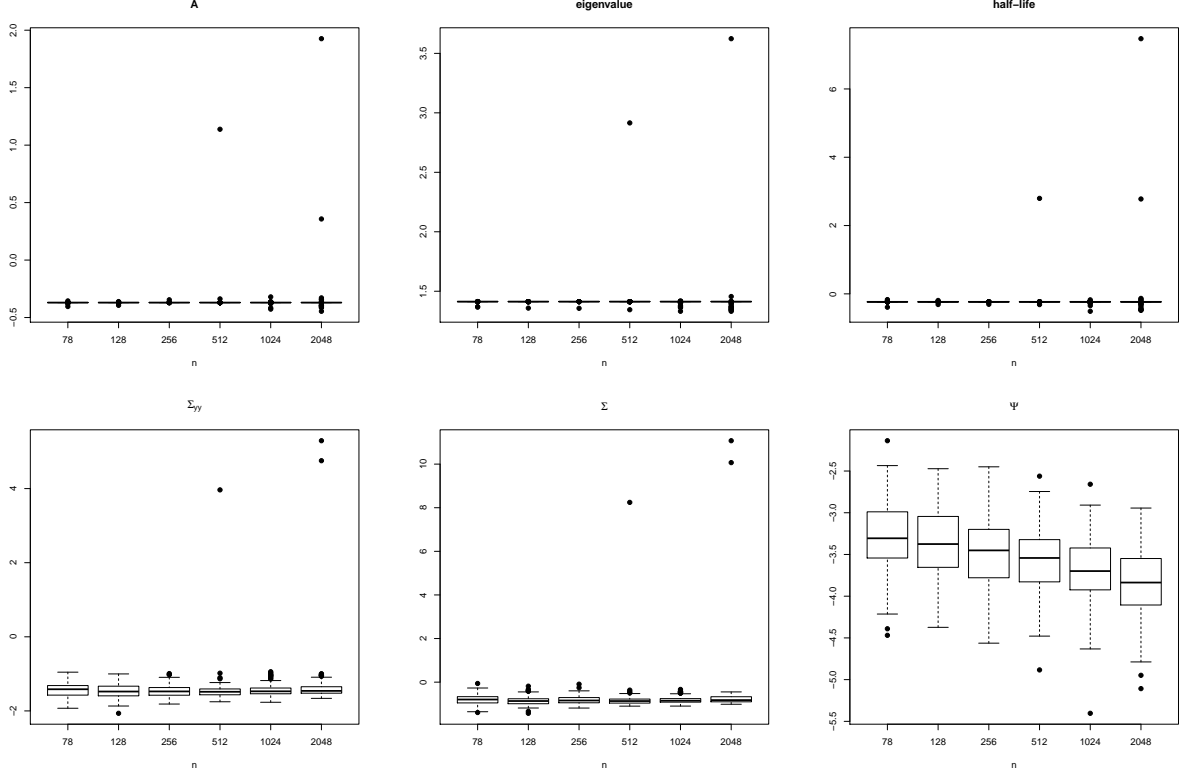


Figure S.15: Quality of parameter estimates under model 7D (model 7, diagonal Σ_{yy} , without measurement error, Tab. S.5). Boxplots of relative distances between true matrix/vector parameter and estimated ones. The y -axes are on logarithmic scale, x -axes are the number of tips of the phylogeny. The boxplots come from 100 successful simulation-reestimation runs, see Tab. S.7. For $n = 78$ tips, the simulations were done with the same tree and measurement error as in the *Ferula* data. For all other n (tree sizes) pure birth trees are simulated and measurement error was simulated according to Alg. S.1.

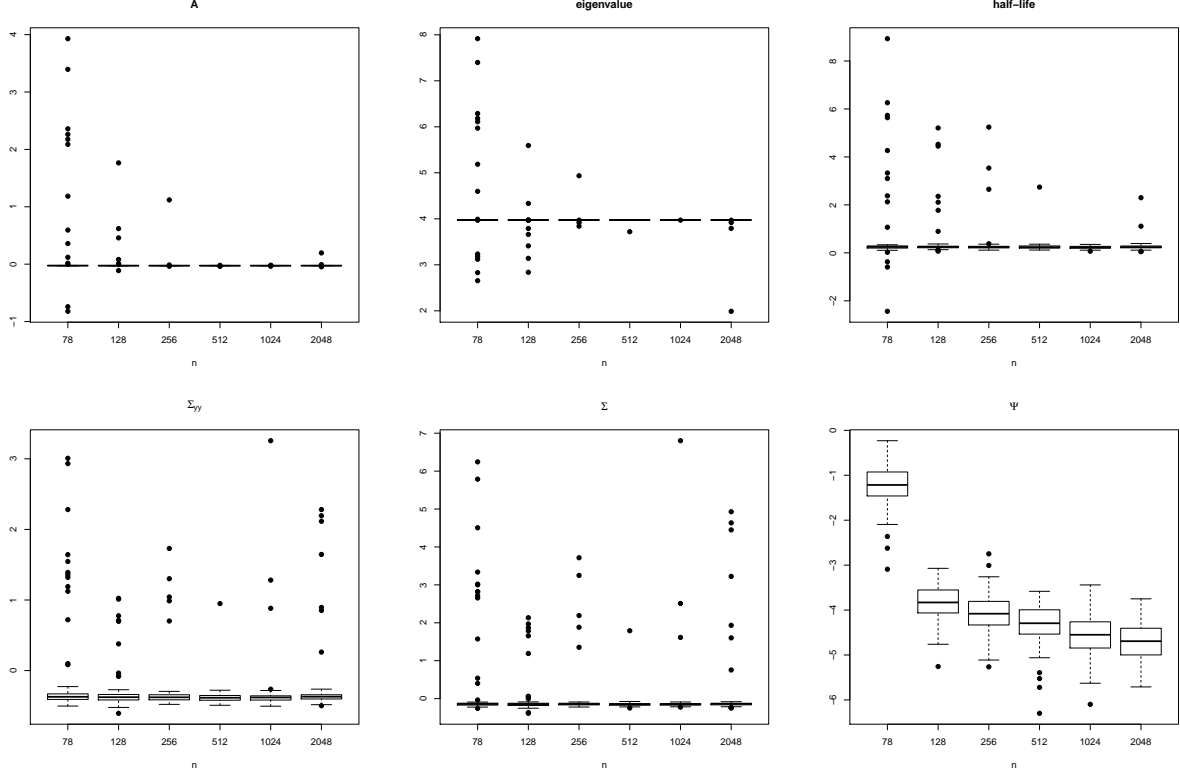


Figure S.16: Quality of parameter estimates under model 7Dm (model 7, diagonal Σ_{yy} , with measurement error, Tab. S.5). Boxplots of relative distances between true matrix/vector parameter and estimated ones. The y -axes are on logarithmic scale, x -axes are the number of tips of the phylogeny. The boxplots come from 100 successful simulation-reestimation runs, see Tab. S.7. For $n = 78$ tips, the simulations were done with the same tree and measurement error as in the *Ferula* data. For all other n (tree sizes) pure birth trees are simulated and measurement error was simulated according to Alg. S.1.

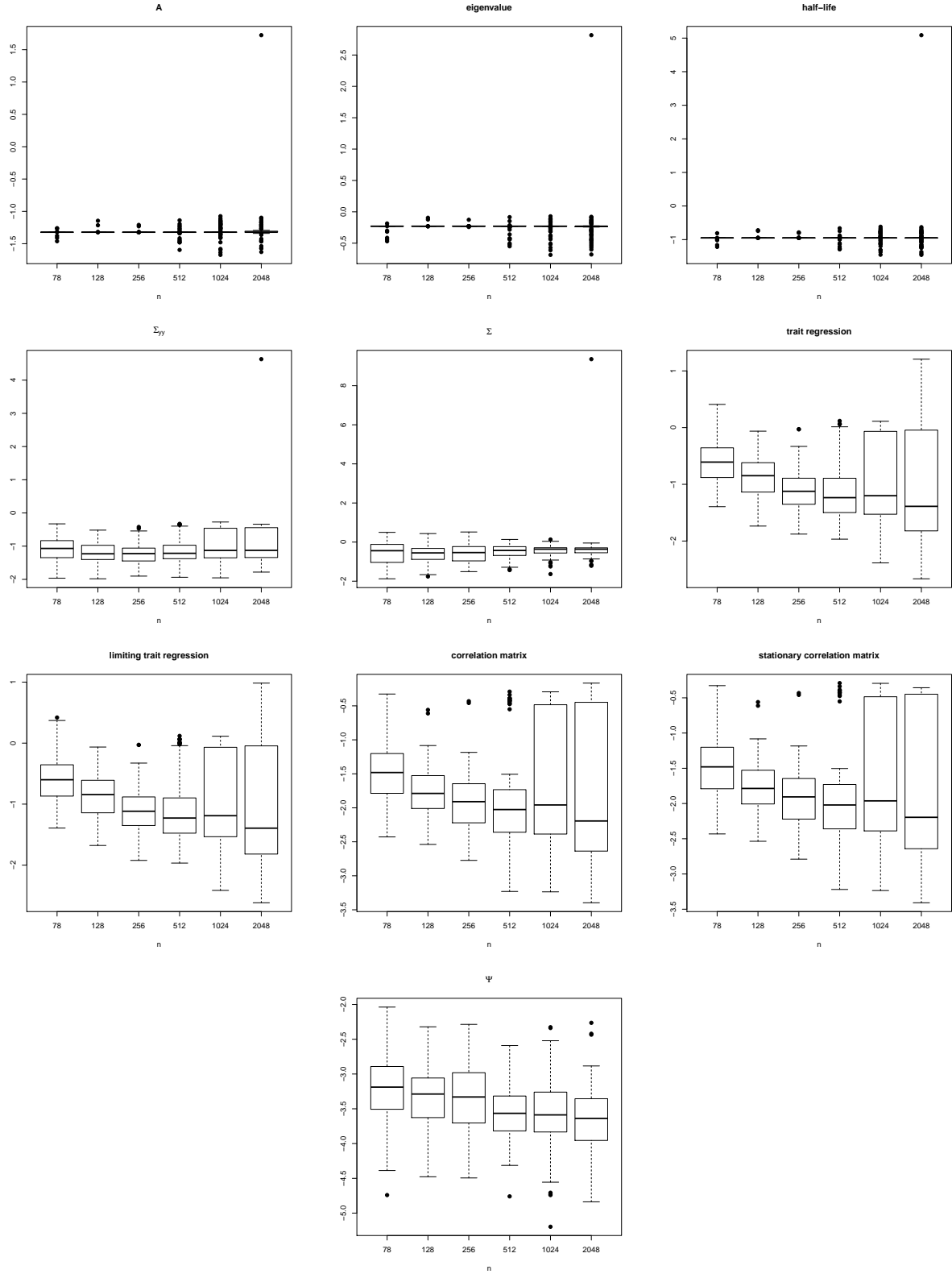


Figure S.17: Quality of parameter estimates under model 7U (model 7, upper triangular Σ_{yy} , without measurement error, Tab. S.5). Boxplots of relative distances between true matrix/vector parameter and estimated ones.⁵⁰ The y -axes are on logarithmic scale, x -axes are the number of tips of the phylogeny. The boxplots come from 100 successful simulation-reestimation runs, see Tab. S.7. For $n = 78$ tips, the simulations were done with the same tree and measurement error as in the Ferula data. For all other n (tree sizes) pure birth trees are simulated and measurement error was simulated according to Alg. S.1.

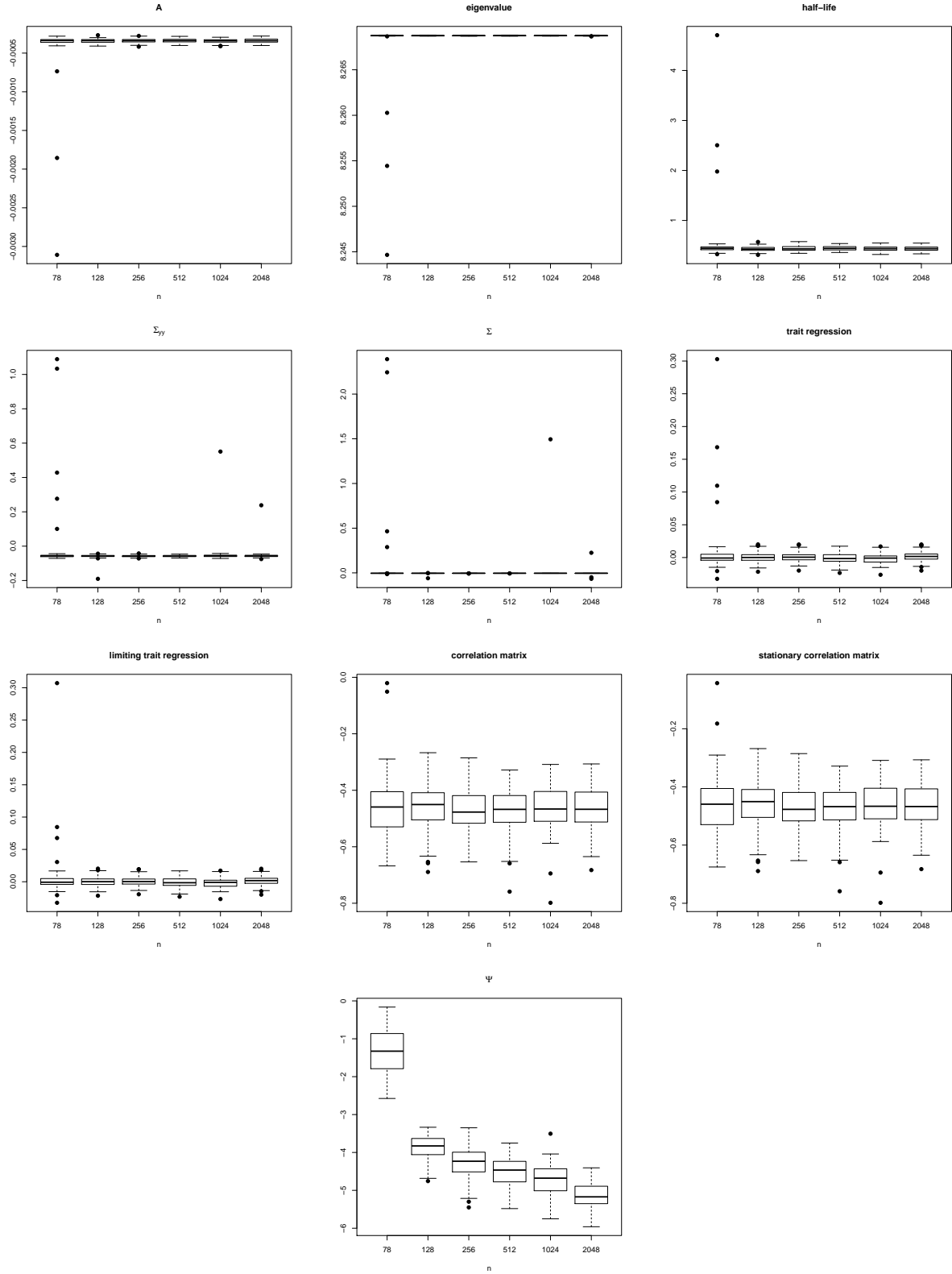


Figure S.18: Quality of parameter estimates under model 7Um (model 7, upper triangular Σ_{yy} , with measurement error, Tab. S.5). Boxplots of relative distances between true matrix/vector parameter and estimated ones. The y -axes are on logarithmic scale, x -axes are the number of tips of the phylogeny. The boxplots come from 100 successful simulation-reestimation runs, see Tab. S.7. For $n = 78$ tips, the simulations were done with the same tree and measurement error as in the *Ferula* data. For all other n (tree sizes) pure birth trees are simulated and measurement error was simulated according to Alg. S.1.

APPENDIX SE: ANALYSES OF UNGULATES

Appendix SE1

Table S.8: Transition matrix conditioning the ancestral state reconstruction of feeding preferences (B = Browser; G = Grazer; M = Mixed feeder) under stochastic character mapping. This matrix corresponds to a symmetric transition rate model, which was significantly better than the other candidates based on likelihood ratio tests (LRTs). The, multiplied by 2, difference of the log-likelihoods was manually computed in R and then, using the asymptotic approximation, the p-value of the χ^2 distribution, with degrees of freedom equalling to the difference in model parameters, was computed using R's *pchisq()* function. While the most complex model (all-rates-different) was not significantly better than the simplest, all-equal rates, (LR= 7.21, degrees of freedom: 5, p-value: 0.21) nor the intermediate, symmetric transition matrix model (LR= -0.26, degrees of freedom: 3, p-value: 1) candidates, the intermediate model explained the history of diet in ungulates better than the all-equal rates (LR= 7.47, degrees of freedom: 2, p-value: 0.02) and was thus selected for the reconstruction. Notice that all three models are nested (simplest is a submodel of intermediate, which in turn is a submodel of the most complex one), so the LRT is permissible.

Diet	B	G	M
B	-1.96	0.00	1.96
G	0.00	-1.05	1.05
M	1.96	1.05	-3.01

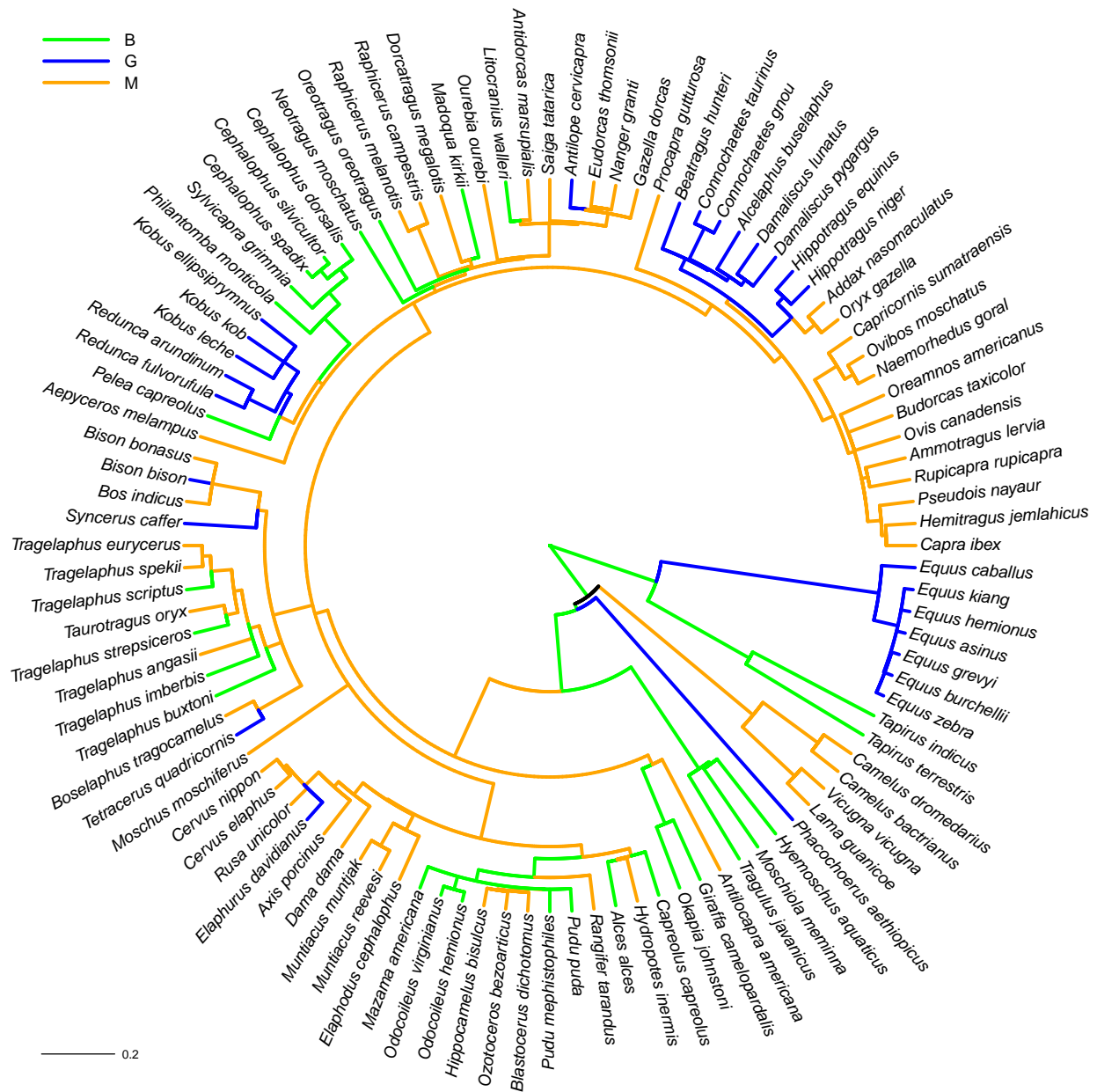


Figure S.19: Ungulate phylogeny with feeding preferences (B = Browser; G = Grazer; M = Mixed feeder) reconstructed under 500 stochastic character mappings with a symmetric transition matrix (Tab. S.8). Character state transitions were assigned to those branches where the highest posterior probability shifted from one diet category to another. Branch lengths are scaled and thus informative of relative time (scale provided at the bottom left).

Appendix SE2

Table S.9: Comparison of best-fitting alternative candidates under each model type based on AIC_c . All these candidates correspond to OUOU models with upper triangular Σ_{yy} and non-diagonal \mathbf{A} (Upper: upper triangular; Lower: lower triangular).

Model	AIC_c	
	Upper	Lower
OU ₁	96.93	96.20
OU _F	63.45	62.78
OU _G	85.44	85.73
OU _B	66.46	66.99

Appendix SE3

In this section we present various parameter estimates for models compared in Tab. 1. For BM, $\vec{X}(0)$ and Σ_{xx} are shown (Tab. S.10). For OU models, \mathbf{A} (Tab. S.11), Σ_{yy} (Tab. S.12), $\vec{\psi}$ (Tab. S.13), and trait regression coefficients (Tab. S.14) are shown.

Table S.10: Σ_{xx} and $\vec{X}(0)$ for BM (left and right sides of the table, respectively).

Variable	Σ_{xx}			$\vec{X}(0)$
	HM ₃	MZW	WM ₃	
HM ₃	0.90	0	0	1.03
MZW	0.64	0.71	0	1.37
WM ₃	0.46	0.39	0.47	0.15

Table S.11: Only diagonal elements of \mathbf{A} are shown for each OU model (in all cases the off-diagonal elements were zero, see main text for details).

Model	HM ₃	MZW	WM ₃
OU ₁	0.37	2.02	1.92
OU _F	1.78	2.98	2.23
OU _G	1.00	2.59	2.06
OU _B	1.45	2.38	1.97

Table S.12: Upper triangular Σ_{yy} for all OU models (see main text for details).

Model		HM ₃	MZW	WM ₃
OU ₁	HM ₃	0.67	0.3	0.59
	MZW	0	0.69	0.94
	WM ₃	0	0	0.91
OU _F	HM ₃	0.67	0.19	0.61
	MZW	0	0.67	0.96
	WM ₃	0	0	0.92
OU _G	HM ₃	0.71	0.27	0.58
	MZW	0	0.68	0.94
	WM ₃	0	0	0.91
OU _B	HM ₃	0.66	0.21	0.61
	MZW	0	0.67	0.94
	WM ₃	0	0	0.91

Table S.13: $\vec{\psi}$ for OU_1 (a single value for each trait), OU_F (one value for each feeding strategy, where B = Browser, M = Mixed feeder, and G = Grazer), OU_G (where B and M are lumped), and OU_B (where G and M are lumped). Note that in all regimes where browsing is specified as a niche (and even the regime where it is lumped with the mixed feeders, OU_G), its optima are the lowest for all variables.

Model	Trait	B	B+M	M	M+G	G
OU_1	HM ₃			0.97		
	MZW			1.28		
	WM ₃			0.06		
OU_F	HM ₃	0.06		1.61		2.59
	MZW	0.86		1.32		2.13
	WM ₃	-0.08		0.03		0.49
OU_G	HM ₃		0.61			3.11
	MZW		1.07			2.15
	WM ₃		-0.04			0.48
OU_B	HM ₃	0.07			2.20	
	MZW	0.88			1.66	
	WM ₃	-0.06			0.20	

Table S.14: Trait regression coefficients for each OU model, which result from calculating the conditional expectation of each response variable (HM₃ and MZW) on all the other traits.

Model	HM ₃		MZW	
	MZW	WM ₃	HM ₃	WM ₃
OU_1	0.67	0.32	0.18	0.82
OU_F	0.31	0.45	0.18	0.77
OU_G	0.50	0.37	0.18	0.77
OU_B	0.37	0.42	0.22	0.78

APPENDIX SF: *Ferula* (APIACEAE) FRUIT'S ANALYSES

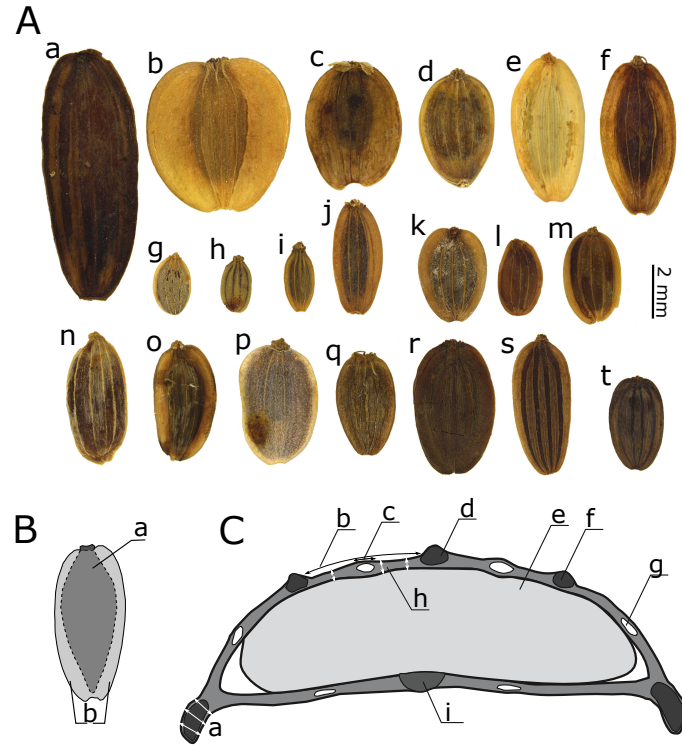


Figure S.20: Morphological and anatomical structure of fruit (mericarp) of representatives of the genus *Ferula*: (A) variation in size and shape of mericarps among species of *Ferula* (a – *Ferula kuhistanica*, b – *F. blanchei*, c – *F. szowitziana*, d – *F. nevskii*, e – *F. schtschurowskiana*, f – *F. hindukushkensis*, g – *F. tuberifera*, h – *F. transiliensis*, i – *Ferula (Leutea) petiolaris*, j – *F. olgae*, k – *F. ammoniacum*, l – *F. aucheri*, m – *F. ovina*, n – *F. kokanica*, o – *F. diversivittata*, p – *F. karelinii*, q – *F. loscosii*, r – *F. glauca*, s – *F. clematidifolia*, t – *F. oopoda*), (B) fruit morphology: a – seed chamber and b – wing area, (C) fruit anatomy: a – wing thickness indicated by three white arrows, b – length of space between median and lateral rib, c – length of oil duct, d – median rib, e – endosperm, f – lateral rib, g – oil duct, h – pericarp thickness indicated by three white arrows, i – funiculus.

Table S.15: Elements of \mathbf{A} matrix for the two best models with measurement error included (models 4 and 7). The abbreviations refer to the following studied fruit traits: RC– the proportion of oil ducts covering the space between median and lateral ribs, PT–periderm thickness, WA–wing area, WT–wing thickness, FM–fruit mass.

Model		RC	PT	WA	WT	FM
4ME Σ_{yy} diagonal	RC	6.57	1.31	0	0	0
	PT	−1.969	4.895	0	0	0
	WA	0	0	785.641	0.926	0
	WT	0	0	−3.677	0.027	0
	FM	0	0	0	0	18.77
7ME Σ_{yy} upper triangular	RC	3.636	0	0	0	0
	PT	0	52.581	0	0	0
	WA	0	0	5.187	0	0
	WT	0	0	0	2.308	0
	FM	0	0	0	0	4.487

Table S.16: Estimated Σ_{yy} for the two best models with measurement error included (models 4 and 7). The abbreviations refer to the following studied fruit traits: RC– the proportion of oil ducts covering the space between median and lateral ribs, PT–periderm thickness, WA–wing area, WT–wing thickness, FM–fruit mass.

Model		RC	PT	WA	WT	FM
4ME	RC	0.969	0	0	0	0
	PT	0	1.687	0	0	0
	WA	0	0	0.611	0	0
	WT	0	0	0	1.255	0
	FM	0	0	0	0	0.088
7ME	RC	0.876	0	0	0	0
	PT	0	2.635	0	0	0
	WA	0	0	0.041	0	0
	WT	0	0	0	1.26	0
	FM	0	0	0	0	0.206

Table S.17: Global optimum, $\vec{\psi}$, for the two best models with measurement error included (models 4 and 7). The ancestral state \vec{Y}_0 is assumed to equal $\vec{\psi}$. The abbreviations refer to the following studied fruit traits: RC– the proportion of oil ducts covering the space between median and lateral ribs, PT–periderm thickness, WA–wing area, WT–wing thickness, FM–fruit mass.

Model	4ME	7ME
RC [ln]	3.881	3.881
PT [ln μm]	4.446	4.446
WA [ln mm^2]	2.908	2.908
WT [ln μm]	5.054	5.053
FM [ln mg]	2.606	2.606

Table S.18: Comparison of half-lives for the two best models with measurement error included (models 4 and 7). The half-lives are given in percentage of tree height. The abbreviations refer to the following studied fruit traits: RC– the proportion of oil ducts covering the space between median and lateral ribs, PT–periderm thickness, WA–wing area, WT–wing thickness, FM–fruit mass.

Model 4ME, Diagonal Σ_{yy}						
Eigenvectors	\vec{e}_1	\vec{e}_2	\vec{e}_3	\vec{e}_4	\vec{e}_5	
RC	0	0	$-0.33 - 0.539i$	$-0.33 + 0.539i$	0	
PT	0	0	0.775	0.775	0	
WA	1	0	0	0	-0.01	
WT	-0.005	-0	0	0	1	
FM	0	1	0	0	0	
Eigenvalues	785.637	18.77	$5.732 + 1.37i$	$5.732 - 1.37i$	0.031	
Half-lives	0.082	3.692	12.092	12.092	2.208	
Model 7ME, Diagonal Σ_{yy}						
Eigenvectors	\vec{e}_1	\vec{e}_2	\vec{e}_3	\vec{e}_4	\vec{e}_5	
RC	1	0	0	0	0	
PT	0	1	0	0	0	
WA	0	0	1	0	0	
WT	0	0	0	1	0	
FM	0	0	0	0	1	
Eigenvalues	52.581	13.636	5.187	4.487	2.308	
Half-lives	1.318	5.083	13.362	15.45	30.038	

Table S.19: Correlation matrices for the two best models with measurement error included (models 4 and 7) with parametric bootstrap 95% confidence intervals for model 4. The confidence intervals are symmetric, they are the 2.5% and 97.5% quantiles from 1000 parametric bootstrap replicates. The abbreviations refer to the following studied fruit traits: RC– the proportion of oil ducts covering the space between median and lateral ribs, PT–periderm thickness, WA–wing area, WT–wing thickness, FM–fruit mass.

Model		RC	PT	WA	WT	FM
4ME	RC	1	−0.133 (−0.757, 0.663)	0	0	0
	PT	−0.133 (−0.757, 0.663)	1	0	0	0
	WA	0	0	1	−0.094 (−0.796, 0.813)	0
	WT	0	0	−0.094 (−0.796, 0.813)	1	0
	FM	0	0	0	0	1
7ME	RC	1	0	0	0	0
	PT	0	1	0	0	0
	WA	0	0	1	0	0
	WT	0	0	0	1	0
	FM	0	0	0	0	1

Table S.20: Regression coefficients between all variables under the two best models with measurement error included (models 4 and 7). with parametric bootstrap 95% confidence intervals for model 4. The confidence intervals are symmetric, they are the 2.5% and 97.5% quantiles from 1000 parametric bootstrap replicates. The abbreviations refer to the following studied fruit traits: RC– the proportion of oil ducts covering the space between median and lateral ribs, PT–periderm thickness, WA–wing area, WT–wing thickness, FM–fruit mass.

Model					
4ME		PT	WA	WT	FM
	RC	−0.068	0	0	0
		(−1.480, 0.882)			
		RC	WA	WT	FM
	PT	−0.257	0	0	0
7ME		RC	PT	WT	FM
	WA	0	0	−0.001	0
				(−1.209, 1.109)	
		RC	PT	WA	FM
	WT	0	0	−7.498	0
4ME		RC	PT	WA	WT
	FM	0	0	0	0
		PT	WA	WT	FM
	RC	0	0	0	0
		RC	PT	WT	FM
	WA	0	0	0	0

	RC	PT	WA	FM
WT	0	0	0	0
	RC	PT	WA	WT
FM	0	0	0	0

References

- Adams, D. C. and M. L. Collyer. 2018. Multivariate phylogenetic comparative methods: Evaluations, comparisons, and recommendations. *Syst. Biol.* 67:14–31.
- Akaike, H. 1974. A new look at the statistical model identification. *IEEE T. Automat. Contr.* 19:716–723.
- Ané, C. 2008. Analysis of comparative data with hierarchical autocorrelation. *Ann. Appl. Stat* 2:1078–1102.
- Ané, C., L. S. T. Ho, and S. Roch. 2016. Phase transition on the convergence rate of parameter estimation under an Ornstein–Uhlenbeck diffusion on a tree. *J. Math. Biol.* 74:355–385.
- Bartoszek, K. 2011. Multivariate Aspects of Phylogenetic Comparative Methods. Licentiate Thesis, University of Gothenburg.
- Bartoszek, K. 2016. Phylogenetic effective sample size. *J. Theor. Biol.* 407:371–386.
- Bartoszek, K., J. Pienaar, P. Mostad, S. Andersson, and T. F. Hansen. 2012. A phylogenetic comparative method for studying multivariate adaptation. *J. Theor. Biol.* 314:204–215.
- Bernstein, D. and W. So. 2013. Some explicit formulas for the matrix exponential. *IEEE Trans. Automat. Contr.* 38:1228–1232.
- Burnham, K. P., D. R. Anderson, and K. P. Huyvaert. 2011. AIC model selection and multimodel inference in behavioral ecology: some background, observations, and comparisons. *Behav. Ecol. Sociobiol.* 65:23–35.

- Cressler, C. E., M. A. Butler, and A. A. King. 2015. Detecting adaptive evolution in phylogenetic comparative analysis using the Ornstein–Uhlenbeck model. *Syst. Biol.* 64:953–968.
- Garamszegi, L. Z. 2014. Uncertainties due to within–species variation in comparative studies: Measurement errors and statistical weights. Pages 157–199 *in* Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology (L. Z. Garamszegi, ed.). Springer, Berlin, Heidelberg.
- Grabowski, M., K. L. Voje, and T. F. Hansen. 2016. Evolutionary modeling and correcting for observation error support a 3/5 brain–body allometry for primates. *J. Hum. Evol.* 94:106–116.
- Hansen, T. F. and K. Bartoszek. 2012. Interpreting the evolutionary regression: the interplay between observational and biological errors in phylogenetic comparative studies. *Syst. Biol.* 61:413–425.
- Ho, L. S. T. and C. Ané. 2014. Intrinsic inference difficulties for trait evolution with Ornstein–Uhlenbeck models. *Meth. Ecol. Evol.* 5:1133–1146.
- Hurvich, C. M. and C. L. Tsai. 1989. Regression and time series model selection in small samples. *Biometrika* 76:297–307.
- Mitov, V., K. Bartoszek, G. Asimomitis, and T. Stadler. 2020. Fast likelihood calculation for multivariate Gaussian phylogenetic models with shifts. *Theor. Pop. Biol.* 131:66–78.
- Sagitov, S. and K. Bartoszek. 2012. Interspecies correlation for neutrally evolving traits. *J. Theor. Biol.* 309:11–19.