Version dated: May 30, 2017

BioRxiv/New Results

Fast and Robust Inference of Phylogenetic Ornstein-Uhlenbeck Models Using Parallel Likelihood Calculation

Venelin Mitov[1,2], Tanja Stadler[1,2]

[1]*Swiss Federal Institute of Technology in Zurich, Switzerland;*

[2]*Swiss Institute of Bioinformatics, Switzerland*

Corresponding authors: Venelin Mitov, Department of Biosystem Sciences and Engineering, Swiss Federal Institute of Technology, Mattenstrasse 26, CH-4058 Basel, Switzerland; E-mail: `vmitov@gmail.com`.

Tanja Stadler, Department of Biosystem Sciences and Engineering, Swiss Federal Institute of Technology, Mattenstrasse 26, CH-4058 Basel, Switzerland; E-mail: `tanja.stadler@bsse.ethz.ch`.

Abstract

Phylogenetic comparative methods have been used to model trait evolution, to test selection versus neutral hypotheses, to estimate optimal trait-values, and to quantify the rate of adaptation towards these optima. Several authors have proposed algorithms calculating the likelihood for trait evolution models, such as the Ornstein-Uhlenbeck (OU)

process, in time proportional to the number of tips in the tree. Combined with gradient-based optimization, these algorithms enable maximum likelihood (ML) inference within seconds, even for trees exceeding 10,000 tips. Despite its useful statistical properties, ML has been criticised for being a point estimator prone to getting stuck in local optima. As an elegant alternative, Bayesian inference explores the entire information in the data and compares it to prior knowledge but, usually, runs in much longer time, even for small trees. Here, we propose an approach to use the full potential of ML and Bayesian inference, while keeping the runtime within minutes. Our approach combines (i) a new algorithm for parallel likelihood calculation; (ii) a previously published method for adaptive Metropolis sampling. In principle, the strategy of (i) and (ii) can be applied to any likelihood calculation on a tree which proceeds in a pruning-like fashion leading to enormous speed improvements. As a showcase, we implement the phylogenetic Ornstein-Uhlenbeck mixed model (POUMM) in the form of an easy-to-use and highly configurable R-package. In addition to the above-mentioned usage of comparative methods, the POUMM allows to estimate non-heritable variance and phylogenetic heritability. Using simulations and empirical data from 487 mammal species, we show that the POUMM is far more reliable in terms of unbiased estimates and false positive rate for stabilizing selection, compared to its alternative - the non-mixed Ornstein-Uhlenbeck model, which assumes a fully heritable and perfectly measurable trait. Further, our analysis reveals that the phylogenetic mixed model (PMM), which assumes neutral evolution (Brownian motion) can be a very unstable estimator of phylogenetic heritability, even if the Brownian motion assumption is only weakly violated. Our results prove the need for a simultaneous account for selection and non-heritable variance in phylogenetic evolutionary models and challenge stabilizing selection hypotheses stated in numerous macro-evolutionary studies.

Keywords: Phylogenetic mixed model, phylogenetic heritability, Brownian motion, measurement error, continuous trait, stabilizing selection, rate of adaptation, environmental contribution, Open MP, Single instruction multiple data, parallel pruning

Introduction

The past decades have seen active developement of phylogenetic comparative models of trait evolution, progressing from null neutral models, such as single-trait

Brownian motion (BM), to complex multi-trait models incorporating selection, interaction between trait values and diversification, and co-evolution of multiple traits (O'Meara 2012; Manceau, Lambert, and Morlon 2016). Recent works have shown that, for a broad family of phylogenetic comparative models, the likelihood of an observed tree and data conditioned on the model parameters can be computed in time proportional to the size of the tree (FitzJohn 2012; Ho and Ané 2014; Goolsby, Bruggeman, and Ané 2016; Manceau, Lambert, and Morlon 2016). This family includes Gaussian models like Brownian motion and Ornstein-Uhlenbeck phylogenetic models as well as some non-Gaussian models like phylogenetic logistic regression (Emmanuel Paradis and Claude 2002; Ives and Garland 2009; Ho and Ané 2014). All of these likelihood calculation techniques rely on post-order tree traversal or 'pruning' as coined by (Felsenstein 1973). Using pruning algorithms for likelihood calculation in combination with a gradient-based optimization method (Boyd and Vandenberghe 2004), maximum likelihood model inference runs within seconds on contemporary computers, even for phylogenies containing many thousands of tips (Ho and Ané 2014). Other important features of the maximum likelihood estimate (MLE) are its simple interpretation as the point in parameter space maximizing the probability of the observed data under the assumed model, and its theoretical properties making it ideal for hypothesis testing and for model selection via likelihood ratio tests and information criteria. However, a major disadvantage of MLE is that, being a point estimate, it does not allow to explore the likelihood surface. Further, gradient based optimization, while fast, is prone to getting stuck in local optima.

As an elegant alternative, Bayesian approaches such as Markov Chain Monte Carlo (MCMC) provide posterior samples and high posterior density (HPD) intervals for the model parameters but require many orders of magnitude more likelihood evaluations. This represents a bottleneck in Bayesian analysis, in particular, when faced with large phylogenies of many thousands of tips, such as transmission trees from large-scale

epidemiological studies, e.g. Hodcroft et al. (2014). While big data provides sufficient statistical power to fit a complex model, the time needed to perform a full scale Bayesian inference often limits the choice to a faster but less informative ML-inference, or a Bayesian inference on a simplified model. Another issue with Bayesian methods is that they require some level of expertise for specifying an apropriate prior, assessing the convergence of the MCMC and interpreting the results.

In this article, we propose a general approach allowing to use ML and Bayesian inference to their full potential, even for complex phylogenetic comparative models and for very large trees exceeding millions of tips, where the limiting factor becomes the available memory and not the calculation time. To achieve this goal, our approach combines two ideas: (i) the pruning algorithm for likelihood calculation can be accelerated by orders of magnitude through parallelization on modern multi-core processors and graphics adapters; (ii) the number of iterations needed for MCMC convergence can be reduced by the use of adaptive Metropolis sampling (Vihola 2012; Scheidegger 2012). Our parallel algorithm relies on a previously unexplored representation of the likelihood function as a quadratic polynomial of the trait value at the root. A nice property of the algorithm is that its parallel efficiency converges to 1 as the number of tips in the tree goes to infinity. Thus, for large trees, the parallel speed-up is practically limited by the number of available processing cores.

Numerous studies have discussed the Ornstein-Uhlenbeck (OU) process (Ornstein and Zernike 1919; Uhlenbeck and Ornstein 1930) as a model for trait adaptation under stabilizing selection, see e.g. Hansen (1997), Beaulieu et al. (2012), L. J. Harmon et al. (2010), Manceau, Lambert, and Morlon (2016) and references therein. However, in a cautionary note, Cooper et al. (2015) questionned the application of OU as a validation model for stabilizing selection and showed through simulations that OU-inferences are prone to overestimating the strength of selection if they do not account for measurement

error. Connecting to these studies and providing a showcase for our parallel pruning algorithm, we implemented the phylogenetic Ornstein-Uhlenbeck mixed model (POUMM). Formally, POUMM can be regarded as adding a white noise term to the non-mixed phylogenetic OU model (POU), also referred to as single stationary peak (SSP) model (L. J. Harmon et al. 2010) and "Hansen" model with a single selection regime (Hansen 1997; Butler and King 2004). This white noise term is interpreted as a non-heritable component, i.e. a contribution to the measured trait-value not explainable by the assumed phylogenetic model, such as a measurement error, an environmental contribution and a model residual. Another way to interpret the POUMM is as an extension of the phylogenetic mixed model (PMM) (Lynch 1991; Housworth, Martins, and Lynch 2004), replacing the BM process with an OU process. The POUMM combines the applications of the above two models and, as we will show, is more reliable in terms of correct estimation of selection strength and phylogenetic heritability. However, currently, there are no software tools supporting fast Bayesian POUMM inference on large non-ultrametric trees. We provide our implementation in the form of a package written in the R language of statistical computing (R Core Team 2013). Based on our simulations, the time for a combined MLE and MCMC-fit on a tree of 10,000 tips, including two parallel MCMC chains of a million iterations is reduced from days to a few minutes. We present the model and its applications, the parallel algorithm for likelihood calculation and simulation results validating the technical correctness of the software and comparing its performance and robustness to alternative models and implementations. The POUMM R-package has already been used in several studies quantifying the heritability of continous traits, such as the "set-point virus load" and the "CD4 cell decline" in large HIV phylogenies with tips sampled through time (Mitov and Stadler (2016), Blanquart et al. (2017), Bertels et al. (2017), Bachmann et al. (2017)). Here, we additionally apply the method to an ultrametric tree and body weight data from 487 extant mammal species, including monophyletic

groups of 227 Rodentia, 138 Chiroptera and 122 Soricomorpha species (Bininda-Emonds et al. 2007; Raia, Carotenuto, and Meiri 2010; Smith et al. 2003). Strikingly, the analysis of this data reveals that the outcome of model selection based on a likelihood ratio (LR) test and the Akaike information criterion (AICc) depends on the inclusion of a non-heritable component in the model. When comparing models with a non-heritable component (PMM vs POUMM), the Brownian-motion based PMM gets selected for the three mammal orders, as well as the combined phylogeny. Conversely, if comparing models assuming full heritablity, the POU model gets selected over non-mixed phylogenetic Brownian motion (PBM). For all trees the PMM model has the best AICc compared to all other models. These results challenge previous hypotheses of stabilizing selection that have been validated through POU models acting at the *class* and *order* levels (e.g. Raia and Meiri (2011)). Before us, others have pointed out this issue in simulation studies (Cooper et al. 2015). But the continuous use of phylogenetic comparative models assuming full phylogenetic heritability shows the low awareness for that issue and the need to provide strong empirical evidence. We revisit this issue in the Discussion section.

Materials and Methods

Through the rest of the article we will rely on the following setup. Given is a rooted phylogenetic tree $\mathcal{T}$ with N tips indexed by $1, ..., N$ and a root node, 0 (Fig. 1). Without restrictions on the tree topology, non-ultrametric trees (i.e. tips have different time-distance from the root) and polytomies (i.e. nodes with any finite number of descendants) are accepted. Internal nodes are indexed by the numbers $N + 1, ...$ Associated with the tips is a N-vector of observed real trait values denoted by $\mathbf{z}$. We denote by $\mathcal{T}_i$ the subtree rooted at node i and by $\mathbf{z}_i$ the set of values at the tips belonging to $\mathcal{T}_i$. For any internal node j, we denote by $Desc(j)$ the set of its direct descendants. Furthermore, for any $i \in Desc(j)$, we denote by t_{ji} the length of the edge connecting j with i and by t_{0i} the sum of

edge-lengths (time-distance) from the root to i. The mean root-tip distance in the tree is denoted by $\bar{t}$. For two tips i and k, we denote by $t_{0(ik)}$ the time-distance from the root to their most recent common ancestor (mrca), and by τ_{ik} the sum of edge-lengths on the path from i to k (also called phylogenetic/patristic distance between i and k).

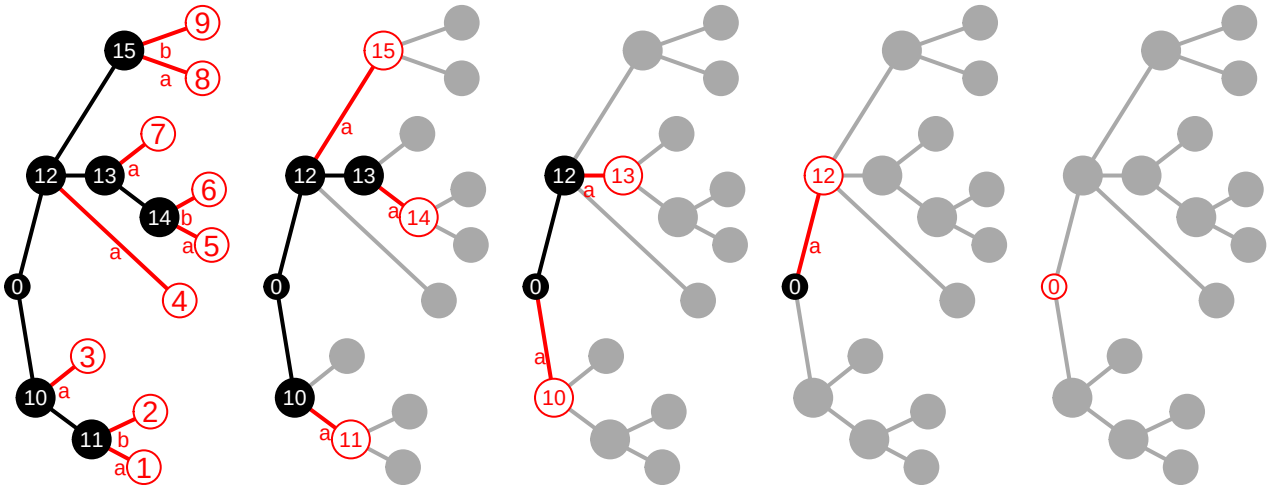


Figure 1: Breadth-first pruning on a tree with $N = 9$ tips. Each tree from left to right depicts one pruning iteration; black: non-tip nodes at a current pruning step; red: tip nodes to be pruned; grey: pruned nodes. Letters 'a' and 'b' next to branches denote the order in which the coefficients a_{ji}, b_{ji}, c_{ji} are added to their parent's a_j, b_j and c_j (algorithm 1).

The Phylogenetic Ornstein-Uhlenbeck Mixed Model

The phylogenetic Ornstein-Uhlenbeck mixed model (POUMM) decomposes the trait value as a sum of a non-heritable component, e, and a genetic component, g, which (i) evolves continuously according to an Ornstein-Uhlenbeck (OU) process along branches; (ii) gets inherited by the branches descending from each internal node. In biological terms, g is a genotypic value (Lynch and Walsh 1998) that evolves according to random drift with stabilizing selection towards a global optimum; e is a non-heritable component, which can be interpreted in different ways, depending on the application, i.e. a measurement error, an environmental contribution, a residual with respect to a model prediction, or the sum of all

these. The OU-process acting on g is parameterized by an initial genotypic value at the root, g_0, a global optimum, θ, a selection strength, $\alpha>0$, and a random drift unit-time standard deviation, σ. Denoting by W_t the standard Wiener process (Grimmett and Stirzaker 2001), the evolution of the trait-value, $z(t)$, along a given lineage of the tree is described by the equations:

$$z(t) = g(t) + e \tag{1}$$

$$dg(t) = \alpha[\theta - g(t)]dt + \sigma dW_t, \tag{2}$$

The stochastic differential equation 2 defines the OU-process, which represents a random walk tending towards the global optimum θ with stronger attraction for bigger difference between $g(t)$ and θ (Ornstein and Zernike 1919; Uhlenbeck and Ornstein 1930). The model assumptions for e are that they are iid normal with mean 0 and standard deviation σ_e at the tips. Any process along the tree that gives rise to this distribution at the tips may be assumed for e. For example, in the case of epidemics, a newly infected individual is assigned a new e-value which represents the contribution from its immune system and this value can change or remain constant throughout the course of infection. In the case of macro-evolution, e may represent the ecological (non-genetic) differences between species. In particular, the non-heritable component e does not influence the behavior of the OU-process $g(t)$. Thus, if we were to simulate trait values z on the tips of a phylogenetic tree $\mathcal{T}$, we could first simlate the OU-process from the root to the tips to obtain g, and then add the white noise e (i.e. an iid draw from a normal distribution) to each simulated g value at the tips.

The POUMM represents an extension of the phylogenetic mixed model (PMM) (Lynch 1991; Housworth, Martins, and Lynch 2004), since, in the limit $\alpha \to 0$, the

Table 1: Population properties at the tips of the phylogeny under POUMM and PMM. μ_i: expected value at tip i; Σ_{ii}: expected variance for tip i; Σ_{ij}: expected covariance of the values of tips i and j; $H^2_{\bar{t}}$: phylogenetic heritability at mean root-tip distance; H^2_{∞}: phylogenetic heritability at long-term equilibrium; H^2_e: time-independent (empirical) phylogenetic heritability. Since the expressions for the expected variance-covariance matrix of the POUMM are only defined for strictly positive α, the expressions for PMM are obtained noting that $\lim_{\alpha\to 0} \alpha/(1-e^{\alpha t}) = -1/t$.

	POUMM	PMM ($\alpha \to 0$)
Θ:	$<g_0,\, \alpha,\, \theta,\, \sigma,\, \sigma_e>$	$<g_0,\, \sigma,\, \sigma_e>$
$\mu_i(\Theta, \mathcal{T})$:	$e^{-\alpha t_{0i}} g_0 + (1 - e^{-\alpha t_{0i}})\, \theta$	g_0
$\Sigma_{ii}(\Theta, \mathcal{T})$:	$\frac{\sigma^2}{2\alpha}\, (1 - e^{-2\alpha t_{0i}}) \,+\, \sigma_e^2$	$\sigma^2\, t_{0i} + \sigma_e^2$
$\Sigma_{ij}(\Theta, \mathcal{T})$:	$\frac{\sigma^2}{2\alpha}\, e^{-\alpha \tau_{ij}} \left(1 - e^{-2\alpha t_{0(ij)}}\right)$	$\sigma^2\, t_{0(ij)}$
$H^2_{\bar{t}}$:	$\frac{\sigma^2\left(1-e^{-2\alpha\bar{t}}\right)}{\sigma^2\left(1-e^{-2\alpha\bar{t}}\right)+2\alpha\sigma_e^2}$	$\bar{t}\sigma^2/(\bar{t}\sigma^2 + \sigma_e^2)$
H^2_{∞}:	$\sigma^2/(\sigma^2 + 2\alpha\sigma_e^2)$	1
H^2_e:	$1 - \sigma_e^2/s^2(\mathbf{z})$	$1 - \sigma_e^2/s^2(\mathbf{z})$

OU-process converges to a Brownian motion (BM) with unit-time standard deviation σ. Both, the POUMM and the PMM, define an expected multivariate normal distribution for the trait values at the tips. The mean vectors and the variance-covariance matrices of these distributions are written in table 1. Note that the trait expectation and variance for a tip i depends on its time-distance from the root (t_{0i}), and the trait covariance for a pair of tips (ij) depends on the time-distance from the root to their mrca ($t_{0(ij)}$), and, in the case of POUMM, on their patristic distance (τ_{ij}) (table 1).

A parallel pruning algorithm for fast likekelihood calculation

For a fixed tree, $\mathcal{T}$, the log-likelihood of the observed data is defined as the function:

$$\ell\ell(\Theta) = \ln(f(\mathbf{z}_0|\mathcal{T}, \Theta)), \tag{3}$$

where f denotes a probability density function (pdf) and $\Theta =< g_0, \alpha, \theta, \sigma, \sigma_e >$. Here, we

propose a parallel variant of the pruning algorithm (Felsenstein 1973). The log-likelihood is calculated by consecutive integration over the unobservable genotypic values, g_i, progressing from the tips to the root. Central for the algorithm is the following theorem:

Theorem 1 (Quadratic polynomial representation of the POUMM log-likelihood)**.** *For $\alpha \geq 0$, a real θ and non-negative σ and σ_e, the POUMM log-likelihood can be expressed as a quadratic polynomial of g_0:*

$$\ell\ell(\Theta) = a_0 g_0^2 + b_0 g_0 + c_0, \tag{4}$$

where $a_0 < 0$, b_0 and c_0 are real coefficients. We denote by $u(\alpha, t)$ the function:

$$u(\alpha, t) := \begin{cases} \alpha/(1 - e^{\alpha t}), & \textit{for } \alpha > 0 \\ -1/t, & \textit{for } \alpha = 0 \end{cases} \tag{5}$$

Then, the coefficients in eq. 4 can be expressed with the following recurrence relation:

1. For $j \in \{1, ..., N\}$ (tips):

$$a_j = -\frac{1}{2\sigma_e^2};\ b_j = \frac{z_j}{\sigma_e^2};\ c_j = -\frac{z_j^2}{2\sigma_e^2} - \ln\sqrt{2\pi\sigma_e^2} \tag{6}$$

2. For $j > N$ (internal nodes) or $j = 0$ (root):

$$
\begin{aligned}
a_j &= \sum_{i \in Desc(j)} \frac{a_i u(\alpha, 2t_{ji})}{u(\alpha, 2t_{ji}) - \alpha + \sigma^2 a_i} \\
b_j &= \sum_{i \in Desc(j)} \frac{u(\alpha, 2t_{ji}) \left[2\theta a_i (e^{\alpha t_{ji}} - 1) + b_i e^{\alpha t_{ji}}\right]}{u(\alpha, 2t_{ji}) - \alpha + \sigma^2 a_i} \\
c_j &= \sum_{i \in Desc(j)} \Bigg\{ c_i + \alpha t_{ji} - \frac{0.25\, b_i^2 \sigma^2}{-\alpha + a_i \sigma^2 + u(\alpha, 2t_{ji})} - \\
&\qquad 0.5 \ln \left(\frac{-\alpha + a_i \sigma^2 + u(\alpha, 2t_{ji})}{u(\alpha, 2t_{ji})} \right) + \\
&\qquad \frac{\alpha\theta \left[a_i \theta - (b_i + a_i \theta) e^{\alpha t_{ji}}\right]}{u(\alpha, t_{ji}) + (-\alpha + a_i \sigma^2)\left(1 + e^{\alpha t_{ji}}\right)} \Bigg\}.
\end{aligned}
\tag{7}
$$

Proof. Induction from the tips to the root of the tree.

- *Basis:* For a tip-node i, $\mathcal{T}_i$ is the trivial tree consisting of this tip-node only and the pdf of $\mathbf{z}_i$, conditioned on the unobservable genotypic value g_i, is given by the normal pdf with mean g_i and standard deviation σ_e. This pdf can be written as:

$$
f(\mathbf{z}_i | g_i; \sigma_e) = \mathcal{N}(z_i; g_i, \sigma_e) = e^{-\frac{1}{2\sigma_e^2} g_i^2 + \frac{z_i}{\sigma_e^2} g_i - \frac{z_i^2}{2\sigma_e^2} - \ln(\sigma_e) - 0.5 \ln(2\pi)} \tag{8}
$$

  By defining $a_i = -\frac{1}{2\sigma_e^2}$, $b_i = \frac{z_i}{\sigma_e^2}$ and $c_i = -\frac{z_i^2}{2\sigma_e^2} - \log(\sigma_e) - 0.5 \log(2\pi)$ and taking the natural logarithm of the pdf we obtain the log-likelihood representation from eq. 4, where $a_0 < 0$, b_0 and c_0 can be calculated from the observed value z_i and the model parameter σ_e.

- *Inductive hypothesis:* Assume that for an internal node j, the statement of the theorem has been proven for all subtrees $\mathcal{T}_i$, $i \in Desc(j)$.

- *Inductive step:* Substituting g_j for g_0 and t_{ji} for t_{0i} in the POUMM expressions for μ_i and Σ_{ii} (Table 1), and integrating over g_i, we can write the pdf of $\mathbf{z}_i$, conditioned on g_j, t_{ji} and Θ:

$$
\begin{aligned}
f(\mathbf{z}_i|\Theta, t_{ji}, g_j) &= \int_{-\infty}^{\infty} f(g_i|\Theta, t_{ji}, g_j) \times e^{a_i g_i^2 + b_i g_i + c_i} dg_i \\
&= \int_{-\infty}^{\infty} \mathcal{N}\left[g_i; e^{-\alpha t_{ji}} g_j + (1 - e^{-\alpha t_{ji}})\theta, (1 - e^{-2\alpha t_{ji}})\frac{\sigma^2}{2\alpha}\right] \times \\
&\quad e^{a_i g_i^2 + b_i g_i + c_i} dg_i \\
&= \int_{-\infty}^{\infty} e^{(p_{ji}+a_i){g_i}^2 + (q_{ji}+b_i)g_i + (r_{ji}+c_i)} dg_i \ , \quad \text{where} \qquad (9) \\
&\quad p_{ji} = -\frac{\alpha e^{2\alpha t_{ji}}}{\sigma^2\left(e^{2\alpha t_{ji}}-1\right)} \\
&\quad q_{ji} = \frac{2\alpha e^{\alpha t_{ji}}\left[g_j + \theta\left(e^{\alpha t_{ji}}-1\right)\right]}{\sigma^2\left(e^{2\alpha t_{ji}}-1\right)} \\
&\quad r_{ji} = -\ \frac{\alpha\left[g_j + \theta\left(e^{\alpha t_{ji}}-1\right)\right]^2}{\sigma^2\left(e^{2\alpha t_{ji}}-1\right)} - \frac{1}{2}\ln\left(\frac{\pi\sigma^2\left(1-e^{-2\alpha t_{ji}}\right)}{\alpha}\right)
\end{aligned}
$$

We notice that p_{ji}, q_{ji} and r_{ji} in eq. 9 are not defined in the case of BM ($\alpha = 0$). In that case, we take the limit for $\alpha \to 0$ represented by the function $u(\alpha, t)$ in the main text (eq. 5). By substituting $u(\alpha,\ t)$ in the expressions for p_{ji}, q_{ji} and r_{ji} (eq. 9) we obtain:

$$
\begin{aligned}
p_{ji} &= \frac{e^{2\alpha t_{ji}} u(\alpha, 2t_{ji})}{\sigma^2} \\
q_{ji} &= -\frac{u(\alpha, 2\,t_{ji})\left[g_j + \theta\left(e^{\alpha t_{ji}}-1\right)\right]}{\sigma^2} \qquad (10) \\
r_{ji} &= \frac{u(\alpha, 2\,t_{ji})\left[g_j + \theta\left(e^{\alpha t_{ji}}-1\right)\right]^2}{\sigma^2} - \frac{1}{2}\ln\left(-\frac{\pi\sigma^2}{u(\alpha, 2t_{ji})e^{2\alpha t_{ji}}}\right).
\end{aligned}
$$

Since $a_i < 0$ and, for positive t and $\alpha \in [0, \infty)$, $u(\alpha,\ t)$ accepts strictly negative

values in the interval $[-1/t, 0)$, the integral in eq. 9 has a closed form solution:

$$\begin{cases} \int\limits_{-\infty}^{\infty} e^{(p_{ji}+a_i)g_i^2+(q_{ji}+b_i)g_i+(r_{ji}+c_i)} dg_i \\ = \exp\left[\frac{-(q_{ji}+b_i)^2}{4(p_{ji}+a_i)} + (r_{ji}+c_i) + \ln\left(\sqrt{\frac{\pi}{-(p_{ji}+a_i)}}\right)\right] \\ = e^{a_{ji}g_j^2+b_{ji}g_j+c_{ji}} \text{ , where} \\ a_{ji} = \frac{a_i u(\alpha,2t_{ji})}{u(\alpha,2t_{ji})-\alpha+\sigma^2 a_i} \\ b_{ji} = \frac{u(\alpha,2t_{ji})(e^{\alpha t_{ji}}(2\theta a_i+b_i)-2\theta a_i)}{u(\alpha,2t_{ji})-\alpha+\sigma^2 a_i} \\ c_{ji} = c_i + \alpha t_{ji} - \frac{0.25\, b_i^2\sigma^2}{-\alpha+a_i\sigma^2+u(\alpha,2t_{ji})} - \\ \qquad 0.5\ln\left(\frac{-\alpha+a_i\sigma^2+u(\alpha,2t_{ji})}{u(\alpha,2t_{ji})}\right) + \frac{\alpha\theta\left[a_i\theta-(b_i+a_i\theta)e^{\alpha t_{ji}}\right]}{u(\alpha,t_{ji})+(-\alpha+a_i\sigma^2)\left(1+e^{\alpha t_{ji}}\right)} \end{cases} \tag{11}$$

In eq. 11 above, $a_{ji} < 0$ because it is a fraction with a positive nominator and a negative denominator (note that $a_i < 0$ by the inductive hypothesis and $u(\alpha, 2t_{ji}) < 0$ by definition). Since the vectors $\mathbf{z}_i$, $i \in Desc(j)$, are conditionally independent given Θ, the conditional pdf of $\mathbf{z}_j$ factorizes as:

$$\begin{aligned} f(\mathbf{z}_j|\Theta, g_j, \mathcal{T}_j) &= \prod_{i\in Desc(j)} f(\mathbf{z}_i|\Theta, t_{ji}, g_j) \\ &= \prod_{i\in Desc(j)} e^{a_{ji}g_j^2+b_{ji}g_j+c_{ji}} \\ &= \exp\left[\left(\sum_{i\in Desc(j)} a_{ji}\right) g_j^2 + \left(\sum_{i\in Desc(j)} b_{ji}\right) g_j + \sum_{i\in Desc(j)} c_{ji}\right]. \end{aligned} \tag{12}$$

By denoting $a_j = \sum_{i\in Desc(j)} a_{ji}$, $b_j = \sum_{i\in Desc(j)} b_{ji}$ and $c_j = \sum_{i\in Desc(j)} c_{ji}$ and noticing that $a_j < 0$ as a sum of negative terms, we have proven the inductive step and, thus, the theorem.

□

It can be shown that current pruning implementations (FitzJohn 2012) rely on equivalent formulations of the above theorem. The parallel pruning algorithm differs from these implementations in the ordering of algebraic operations so that they can be performed in parallel for groups of tips or internal nodes rather than consecutively for individual nodes in order of depth-first traversal. This parallelization scheme can also be applied to the generalizaed 3-point representation of the likelihood described in Ho and Ané (2014), allowing to parallelize the likelihood calculation for some non-Gaussian models such as phylogenetic logistic regression and phylogenetic Poisson regression.

Algorithm 1: Parallel pruning

Data: $\mathcal{T}$, $\mathbf{z}$; g_0, α, θ, σ, σ_e
Result: $\ell\ell(g_0, \alpha, \theta, \sigma, \sigma_e; \mathbf{z}, \mathcal{T})$
initialization:
for tips $i \in \{1, ..., N\}$, set a_i, b_i, c_i (eq. 6);
for nodes $j > N$ or $j = 0$, set a_j, b_j, c_j to 0;
set $\{< ji >\}$ to the set of edges $< ji >$ in $\mathcal{T}$, where $i \in \{1, ..., N\}$;
while $\{< ji >\} \neq \phi$ **do**
  for $< ji > \in \{< ji >\}$ **do**
    // Iterations can be executed in parallel
    set $a_{<ji>}$, $b_{<ji>}$, $c_{<ji>}$ to the sub-summands in eq. 7 ;
    add $a_{<ji>}$, $b_{<ji>}$, $c_{<ji>}$ to a_j, b_j, c_j (see branch labels on Fig. 1);
  end
  pruning: set $\mathcal{T}$ to the tree obtained upon removal of $i \in \{< ji >\}$;
  set $\{i\}$ to to the subset of parent nodes in $\{< ji >\}$, which have become tips after the pruning (Fig. 1);
  set $\{< ji >\}$ to the edges leading to $\{i\}$;
end
set $\ell\ell(\Theta) := a_0 g_0^2 + b_0 g_0 + c_0$.

Applications of the POUMM model

From a modeling perspective, the POUMM can be regarded as a combination of two of its widely used nested models:

- The non-mixed phylogenetic Ornstein-Uhlenbeck (POU) model, (also referred to as the 'Hansen model' (Butler and King 2004)), which has been used to infer local phenotypic optima in different phylogenetic clades or across discrete categories of phylogeneticall related species (Butler and King 2004), and to prove the presence of stabilizing selection towards a single stationary peak (SSP) (L. J. Harmon et al. 2010).
- The phylogenetic mixed model (PMM) (Lynch 1991; Housworth, Martins, and Lynch 2004), which has been used to measure phylogenetic heritability (Hodcroft et al. 2014);

Thus, the POUMM combines the applications of the above two models and, as we will show in Results, it is much more accurate and robust in estimating the corresponding parameters. Besides inferring its parameters, the POUMM has several useful properties helping the interpretation of the data and allowing to make predictions about the future trait evolution of the considered population. The properties which we consider represent bijective functions of some of the POUMM parameters. Thus, it is possible to reparametrize the POUMM, so that the model inference is done directly on properties of interest, e.g. the phylogenetic heritability. This is particularly useful for Bayesian inference, since for Bayesian inference priors should be specified for the properties of interest rather than the default POUMM parameters. We call a *parametrization* any numerical bijective function mapping its argument into the default POUMM parameter-space $(< g_0, \alpha, \theta, \sigma, \sigma_e >)$.

Trait distribution at equilibrium.—

An interesting property of the POUMM is that, in the limit $t_{0,(ij)} \to \infty$, it defines a stationary normal distribution for the heritable component (g) at the tips with mean θ and a variance-covariance matrix:

$$\begin{aligned} \Sigma_{ii} = \sigma^2_\infty &= \tfrac{\sigma^2}{2\alpha} \\ \Sigma_{ij} = \Sigma_{ij,\infty} &= \sigma^2_\infty \, e^{-\alpha\tau_{ij}} \end{aligned} \tag{13}$$

The above property proves useful when there is a prior knowledge that the observed population is at equilibrium, because one can use the trait variance in the population, $\sigma^2_z = \sigma^2_\infty + \sigma^2_e$ as model parameter. The corresponding parametrization is:

$$< g_0, \alpha, \theta, \sigma^2_z, \sigma_e > \rightarrow < g_0, \alpha, \theta, \sigma = \sqrt{2\alpha(\sigma^2_z - \sigma^2_e)}, \sigma_e > \tag{14}$$

With this parametrization, one can specify an informed prior for σ^2_z based on empirical estimates on similar data.

Another important aspect of the above property (eq. 13) is that it helps to better understand the selection stregnth parameter α. As it turns out, α can have two different biological interpretations. Considering the expression for μ in table 1, α defines the rate of convergence of the population mean towards the long-term optimum θ. This rate is bigger for bigger values of α and for bigger deviations from θ. Thus, α is considered as selection strength or rate of adaptation under stabilizing selection. Assuming that the majority of the tips and their mrca's are far enough from the root, Σ_{ij} can be viewed as an exponentially decreasing function of the phylogenetic distance τ_{ij} (eq. 13). Seen from that angle, the parameter α can be interpreted as the rate of phenotypic decorrelation between tips, due to genetic drift. When interpreting the results of a model fit, it is important to be aware of this dual interpretation of α. In many cases (e.g. in ultrametric macro-evolutionary tree), the only source of information for inferring α are the observed differences between the tips in the tree. Thus, in the absence of additional evidence, it can

be erronous to assume that the inferred value of α informs stabilizing selection and an adaptation rate towards θ.

A likelihood ratio test between the ML POUMM and PMM fits can be used to test if the inferred parameter α is significantly above 0. As pointed out in the previous paragraph, a significantly positive α does not necessarily imply stabilizing selection towards θ. Further, it is important to note that the value of α can only be interpreted with respect to the time scale of the phylogeny. It can be more intuitive to consider the phylogenetic half-life, $t_{1/2} = \frac{\ln(2)}{\alpha}$, which equals the time it takes for a species entering a new niche to evolve halfway toward its new expected optimum (Hansen 1997).

Phylogenetic heritability.—

The term *phylogenetic heritability*, introduced with the phylogenetic mixed model (PMM) (Housworth, Martins, and Lynch 2004), measures the proportion of phenotypic variance in a population attributable to heritable factors, such as genes, as opposed to non-heritable factors, such as environment and measurement error. Although this concept has been applied mostly in the context of the original PMM, i.e. under the assumption of Brownian motion, the same concept applies to any evolutionary model allowing for the estimation of measurement error (ME) (Hansen and Bartoszek 2012). The *phylogenetic heritability* is defined as the expected proportion of phenotypic variance attributable to g at the tips of the tree, $\sigma^2(g)/\left[\sigma^2(g) + \sigma_e^2\right]$ (Housworth, Martins, and Lynch 2004). This definition is a phylogenetic variant of the definition of broad-sense heritability, H^2, from quantitative genetics (Lynch and Walsh 1998). However, in the case of a trait evolving along a phylogeny, the expected genotypic variance, $\sigma^2(g)$, and, therefore, the phylogenetic heritability, are functions of time. Depending on the applicaiton, the following three types of phylogenetic heritability might all be of interest:

- Expectation at the mean root-tip distance ($\bar{t}$):

$H^2_{\bar{t}} := \left[\sigma^2 \frac{\left(1-e^{-2\alpha\bar{t}}\right)}{2\alpha}\right] / \left[\sigma^2 \frac{\left(1-e^{-2\alpha\bar{t}}\right)}{2\alpha} + \sigma_e^2\right]$. This definition gives rize to three parametrizations where $H^2_{\bar{t}}$ is a free model parameter, while one of the parameters α, σ or σ_e is a function of $H^2_{\bar{t}}$ and the two other parameters. These parametrizations can be expressed as:

$$< g_0, \alpha, \theta, H^2_{\bar{t}}, \sigma_e > \rightarrow < g_0, \alpha, \theta, \sigma = \sqrt{\frac{2\alpha H^2_{\bar{t}} \sigma_e^2}{(1-e^{-2\alpha t})(1-H^2_{\bar{t}})}}, \sigma_e > \tag{15}$$

$$< g_0, H^2_{\bar{t}}, \theta, \sigma, \sigma_e > \rightarrow < g_0, \alpha = \frac{y + W(ye^{-y})}{2\bar{t}}, \theta, \sigma, \sigma_e > \tag{16}$$

$$< g_0, \alpha, \theta, \sigma, H^2_{\bar{t}} > \rightarrow < g_0, \alpha, \theta, \sigma, \sigma_e = \sqrt{\frac{\sigma^2(1-e^{-2\alpha\bar{t}})}{2\alpha}(\frac{1}{H^2_{\bar{t}}} - 1)} >, \tag{17}$$

where $y = \frac{\bar{t}\sigma^2}{\sigma_e^2 H^2_{\bar{t}}} - \frac{\bar{t}\sigma^2}{\sigma_e^2}$ and W is the Lambert-W function.

- Expectation at equilibrium of the OU-process ($t \rightarrow \infty$): $H^2_\infty := \lim_{\bar{t}\rightarrow\infty} H^2_{\bar{t}} = \frac{\sigma^2}{\sigma^2 + 2\alpha\sigma_e^2}$. By taking the limit $\bar{t} \rightarrow \infty$ in equations 15, 16 and 17, we obtain the corresponding parametrizations:

$$< g_0, \alpha, \theta, H^2_\infty, \sigma_e > \rightarrow < g_0, \alpha, \theta, \sigma = \sqrt{\frac{2\alpha H^2_\infty \sigma_e^2}{1 - H^2_\infty}}, \sigma_e > \tag{18}$$

$$< g_0, H^2_\infty, \theta, \sigma, \sigma_e > \rightarrow < g_0, \alpha = \frac{\sigma^2(1-H^2_\infty)}{2H^2_\infty \sigma_e^2}, \theta, \sigma, \sigma_e > \tag{19}$$

$$< g_0, \alpha, \theta, \sigma, H^2_\infty > \rightarrow < g_0, \alpha, \theta, \sigma, \sigma_e = \sqrt{\frac{\sigma^2}{2\alpha}(\frac{1}{H^2_\infty} - 1)} >, \tag{20}$$

- Empirical (time-independent) version of the heritability based on the sample phenotypic variance $s^2(\mathbf{z})$: $H^2_e := 1 - \sigma_e^2 / s^2(\mathbf{z})$. This definition is useful when the tree is non-ultrametric but there is sufficient evidence that the empirical distribution of the trait is stationary along the tree. In this case, $s^2(\mathbf{z})$ coincides with the sum of

the OU variance at equilibrium and σ_e^2. The corresponding parametrization is:

$$< g_0, \alpha, \theta, \sigma, H_e^2 > \rightarrow < g_0, \alpha, \theta, \sigma, \sigma_e = \sqrt{s^2(\mathbf{z})(1 - H_e^2)} > \tag{21}$$

Implementation

Likelihood calculaiton.—

We tested four implementations of algorithm 1 as follows:

- `R on 1 core`: A serial R-implementation based on operations with numerical vectors in R. This implementation can switch transparently between `double` and `Rmpfr` floating point precision (M. Maechler 2014), thus, guaranteeing numerical stability in cases of extreme parameter values, trait values or branch lengths.
- `C++, Armadillo on 1 core`: A C++ implementation using the library Armadillo (Sanderson and Curtin 2016), through the R-package RcppArmadillo (Eddelbuettel and Sanderson 2014);
- `C++, omp-for on X core(s)`: A serial or parallel C++ implementation where vector elementwise operations are written as C++ `for` loops and the Open MP preprocessor directive `"#pragma omp for"` is used to parallelize the iterations;
- `C++, omp-for-simd on X core(s)`: A serial or parallel C++ implementation where vector elementwise operations are programmed as C++ `for` loops and the Open MP preprocessor directive `"#pragma omp for simd"` is used to parallelize and vectorize the iterations.

We compiled the above C++ implementations using version 16.0.0 of the Intel compiler (command icpc with enabled -O3 -march=corei7-avx -mavx optons) on

GNU/Linux OS and performed parallelization benchmarks on up to 10 cores on a processor Intel(R) Xeon(R) CPU E5-2697 v2 @ 2.70GHz (Results).

Possible treatments of g_0.—

Recalling that g_0 is an unknown parameter, $\ell\ell(\Theta)$ is maximized over g_0 by taking $g_0 = -0.5\, b_0/a_0$, which is the maximum of eq. 4. During the Bayesian inference, a prior for the parameter has to be specified and it needs to be sampled like all other parameters. Note that, in many cases, e.g. for long ultrametric trees, the likelihood surface can be nearly flat for the parameter g_0, and maximizing over g_0 may result in extremely high or low values. In these cases, it is better to admit that the data does not inform this parameter, and to exclude it from the free model parameters. This can be done in the following ways:

- assume that g_0 coincides with the long-term optimum θ;
- assume a fixed value for g_0;
- Integrate the log-likelihood over g_0 by assuming that it is sampled from a normal distribution such as the OU equilibrium normal distribution with mean θ and variance $\frac{\sigma^2}{2\alpha}$. We note, that the latter option may be tricky since the OU equilibrium distribution is not defined for $\alpha = 0$.

Fitting the POUMM model.—

Fitting of the POUMM model was implemented as a pipeline including the following steps, where each step can employ any of the four likelihood implementations mentioned above:

1. Perform three MLE searches using the R-function `optim` and the L-BFGS-B method (Byrd et al. 1995), starting from three different points in parameter space;
2. Run three MCMC chains as follows: (i) a chain sampling from the prior distribution; (ii) a chain sampling from the posterior distribution and started from the MLE found

in step 1; (iii) a chain sampling from the posterior distribution and started from a random point in parameters space.

3. If the parameter tuple of highest likelihood sampled in the MCMC has a likelihood higher than the MLE found in step 1, repeat the MLE search starting from that parameter tuple;

To reduce the number of iterations for MCMC convergence, we use adaptive Metropolis sampling with coerced acceptance rate (Vihola 2012; Scheidegger 2012). By using the R-package foreach (Analytics and Weston 2015), our implementation supports running the MCMC chains in parallel. By comparing the posterior samples from two MCMCs initiated from different starting points, it can be assessed whether the MCMCs have converged to the true posterior. We do this quantitatively by the use of the Gelman-Rubin convergence diagnostic (Brooks and Gelman 1998) implemented in the R-package coda (Plummer et al. 2006). Values of the Gelman-Rubin (G.R.) statistic significantly different from 1 indicate that at least one of the two posterior samples deviates significantly from the true posterior distribution. By visual comparison of posterior density with prior desnity plots, it is possible to assess whether the data contains information differring from the prior for a given sampled parameter.

An R-package.—

We provide the model implementation in the form of an R-package called POUMM. Before model fitting, the user can choose from different POUMM parametrizations and prior settings (function `specifyPOUMM`). A set of standard generic functions, such as `plot`, `summary`, `logLik`, `coef`, etc., provide means to assess the quality of a fit (i.e. MCMC convergence, consistence between ML and MCMC fits) as well as various inferred properties, such as high posterior density (HPD) intervals.

In addition, the POUMM package uses several third-party R-packages: ape (E

Paradis, Claude, and Strimmer 2004), data.table (Dowle et al. 2014), coda (Plummer et al. 2006), foreach (Analytics and Weston 2015), ggplot2 (Wickham 2009), GGally (Schloerke et al. 2016), gsl (Hankin 2006) and Matrix (Bates and Maechler 2017).

Results

Simulations

To validate the correctness of the Bayesian POUMM implmentation, we used the method of posterior quantiles (S. R. Cook, Gelman, and Rubin 2006). In this method, the idea is to generate samples from the posterior quantile distributions of selected model parameters (or functions thereof) by means of numerous "replications" of simulation followed by Bayesian parameter inference. In each replication, "true" values of the model parameters are drawn from a fixed prior distribution and trait-data is simulated under the model specified by these parameter values. We perform these simulations on a fixed tree of size $N = 4000$. Then, the to-be-tested software is used to produce a posterior distribution of parameters based on the simulated trait-data. Next, the posterior quantiles of the "true" parameter values (or functions thereof) are calculated from the corresponding posterior samples generated by the to-be-tested software. By running multiple independent replications on a fixed prior, it is possible to generate large samples from the posterior quantile distributions of the individual model parameters, as well as any derived quantities. Assuming correctness of the simulations, any statistically significant deviation from uniformity of these posterior quantile samples indicates an error in the to-be-tested software (S. R. Cook, Gelman, and Rubin 2006).

We compared the accuracy of the POUMM to its two nested models - the PMM which restricts $\alpha = 0$ and infers $<g_0, \sigma, \sigma_e>$, and the non-mixed Ornstein-Uhlenbeck model

(abbreviated as POU) which restricts $\sigma_e = 0$ and infers $<g_0, \alpha, \theta, \sigma>$. Two phylogenetic trees were used for the simulations:

- Ultrametric (BD, $N = 4000$) - an ultrametric birth-death tree of 4000 tips generated using the TreeSim R-package (Stadler et al. 2013,Boskova, Bonhoeffer, and Stadler (2014)) (function call: `sim.bd.taxa(4000, lambda = 2, mu = 1, frac = 1, complete = FALSE)`);
- Non-ultrametric (BD, $N = 4000$) - a non-ultrametric birth-death tree of 4000 tips generated using the TreeSim R-package (Stadler et al. 2013,Boskova, Bonhoeffer, and Stadler (2014)) (function call: `sim.bdsky.stt(4000, lambdasky = 2, deathsky = 1, timesky=0)`).

Simulation scenarios of 2000 replications were run using the prior distribution $< g_0, \alpha, \theta, \sigma, \sigma_e > \sim \mathcal{N}(5, 25) \times \text{Exp}(0.1) \times \mathcal{U}(2, 8) \times \text{Exp}(0.4) \times \text{Exp}(1)$. The goal of using this prior was to explore a large enough subspace of the POUMM parameter space, while keeping MCMC convergence and mixing within reasonable time (runtime up to 30 minutes for two MCMCs of 10^6 adaptive Metropolis iterations at target acceptance rate of 1%). From the above prior, we drew a sample of $n = 2000$ parameter tuples, $\{\Theta^{(1)}, ..., \Theta^{(n)}\}$, which were used as replication seeds in two simulation-modes:

- Simulate POUMM - for a given $\Theta^{(i)}$, simulate genotypic values $\mathbf{g}^{(i)}(\mathcal{T}, \Theta^{(i)})$ according to an OU-branching process with initial value $g_0^{(i)}$ and parameters $\alpha^{(i)}$, $\theta^{(i)}$, $\sigma^{(i)}$. Then add random white noise ($\sim \mathcal{N}(0, \sigma_e^{2(i)})$) to the genotypic values at the tips, to obtain the final trait values $\mathbf{z}^{(i)}$.
- Simulate PMM - for a given $\Theta^{(i)}$, reset $\alpha^{(i)}$ to 0, so that the genotypic values come from a BM process, then, repeat the procedure as in mode "Simulate POUMM". The purpose of implementing this mode was (i) to validate the technical correctness of our PMM implementation by testing for uniformity its posterior quantile distributions;

(ii) to obtain an impression of the robustness of the POUMM method to a prior favoring OU ($\alpha > 0$) in the case of true BM processes ($\alpha = 0$).

Combining the two phylogenies with the two simulation modes we obtained four test-scenarios with a total of $4 \times 2000 = 8000$ replications. The resulting posterior quantile distributions for the PMM and POUMM Bayesian fits in each of these scenarios are shown on Fig. 2 for the non-ultrametric and Fig. S1 for the ultrametric tree.

Technical correctness.—

Both, the PMM and POUMM implementation, generate uniformly distributed posterior quantiles for all relevant parameters when the Bayesian inference has been done on data simulated under the correct simulation mode, i.e. "Simulate PMM" for PMM and "Simulate POUMM" for POUMM. This is confirmed visually by the corresponding histograms (Fig. 2 and Fig. S1), as well as statistically, by a non-significant p-value from a Kolmogorov-Smirnov uniformity test at the 0.01 level. This observation validates the technical correctness of the software.

Robustness to model misspecification.—

Robustness results discussed in this section are all visualized in Fig. 2. When fitting POUMM to data simulated under PMM ($\alpha = 0$), there is a tendency to infer a positive α, overestimating σ and underestimating σ_e. The cause for this is the wrong prior for α pulling it away from 0. Thus, we recommend to always test the hypothesis $\alpha = 0$, e.g. through a likelihood ratio test between the maximum likelihood PMM and POUMM fits. The deviation from uniformity is far less pronounced for $H^2_{\bar{t}}$ (posterior quantiles tending slightly to 1) and H^2_e (posterior quantiles tending slightly to 0).

When fitting PMM to simulations under POUMM, there is a highly significant deviation from uniformity of the posterior quantiles for all parameters and derived quantities. The fact that most posterior quantiles for $H^2_{\bar{t}}$ and H^2_e are at the extremes of the

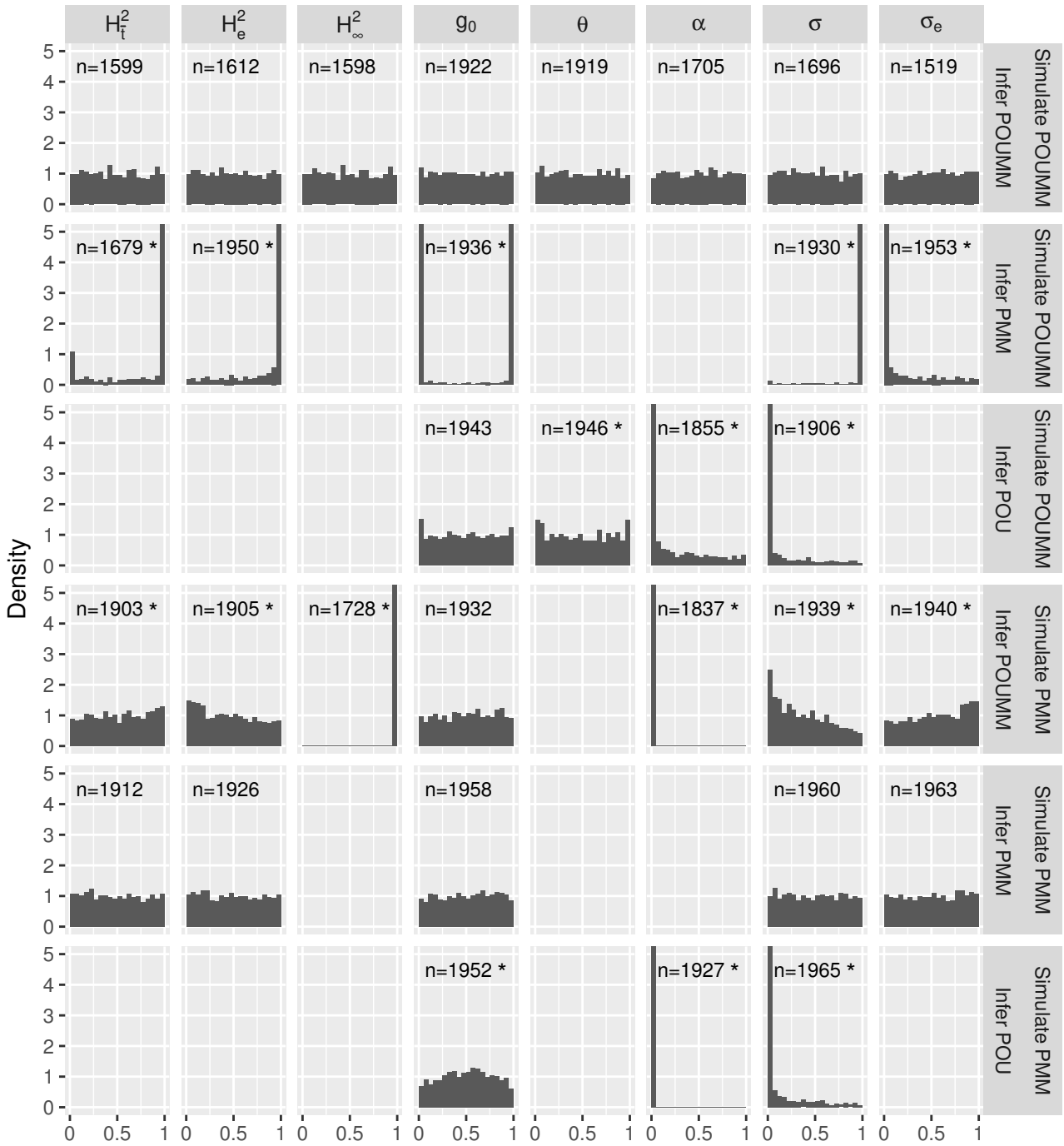


Figure 2: Posterior quantiles from simulation scenarios on a non-ultrametric tree ($N = 4000$). Values tending to 1 indicate that the true value dominates the inferred posterior sample for most of the replications. This means that the model fit tends to underestimate the true parameter. Conversely, values tending to 0 indicate overestimation. The number n at the top of each histogram denotes the number of replications out of 2000 which reached acceptable MCMC convergence and mixing after 10^6 iterations. An asterisk indicates significant uniformity violation (Kolmogorov-Smirnov P-value < 0.01). For the analogous results on an ultrametric tree, see Fig. S1.

histogram is indicative for a systematic negative or positive bias in the inferred parameters. These results indicate that the PMM can be a very unstable erronous estimator of phylogenetic heritability when the data violates the Brownian motion assumption.

When fitting POU ($\sigma_e = 0$) to simulations including environmental contribution (both simulation modes), there is a strong tendency to overestimate α and σ. This indicates that POU-inference ignoring non-heritable contributions is prone to false positive tests for stabilizing selection. The next section provides empirical evidence for this prediction.

Analysis of body weight evolution in mammals

We analysed phylogenetic and body weight data from 227 Rodentia, 138 Chiroptera and 122 Soricomorpha species. An ultrametric tree composed of three monophyletic groups of the above mammal orders was extracted from a previously published supertree of 4510 extant mammal species (Bininda-Emonds et al. 2007) and the body weights have been taken from (Raia, Carotenuto, and Meiri 2010). The orders Rodentia, Chiroptera and Soricomorpha represented the largest groups of species in the supertree with available body weight measures. To extract the phylogeny from the supertree, we used the function "'drop.tip"' from the R-package "'ape"' (E Paradis, Claude, and Strimmer 2004), removing all species of other orders and/or species without available body weight (fig. 3). Further, because all of the phylogenetic models discussed here assume that the trait values at the tips of an ultrametric tree are samples from a normal distribution, we dropped 10 Rodentia, 15 Chiroptera, 4 Soricomorpha species, which appeared as "giants" compared to other species in their corresponding orders and visibly distorted the normality of the data (fig. 3, Discussion). Upon the above filtering, we obtained four trees as follows:

- Rodentia: $N = 217$, $\bar{t} = 85.3$ million years (Myr);
- Chiroptera: $N = 123$, $\bar{t} = 74.9$ Myr;

- Soricomorpha: $N = 118$, $\bar{t} = 84$ Myr;
- All (combining the three above trees): $N = 458$, $\bar{t} = 98.9$ Myr.

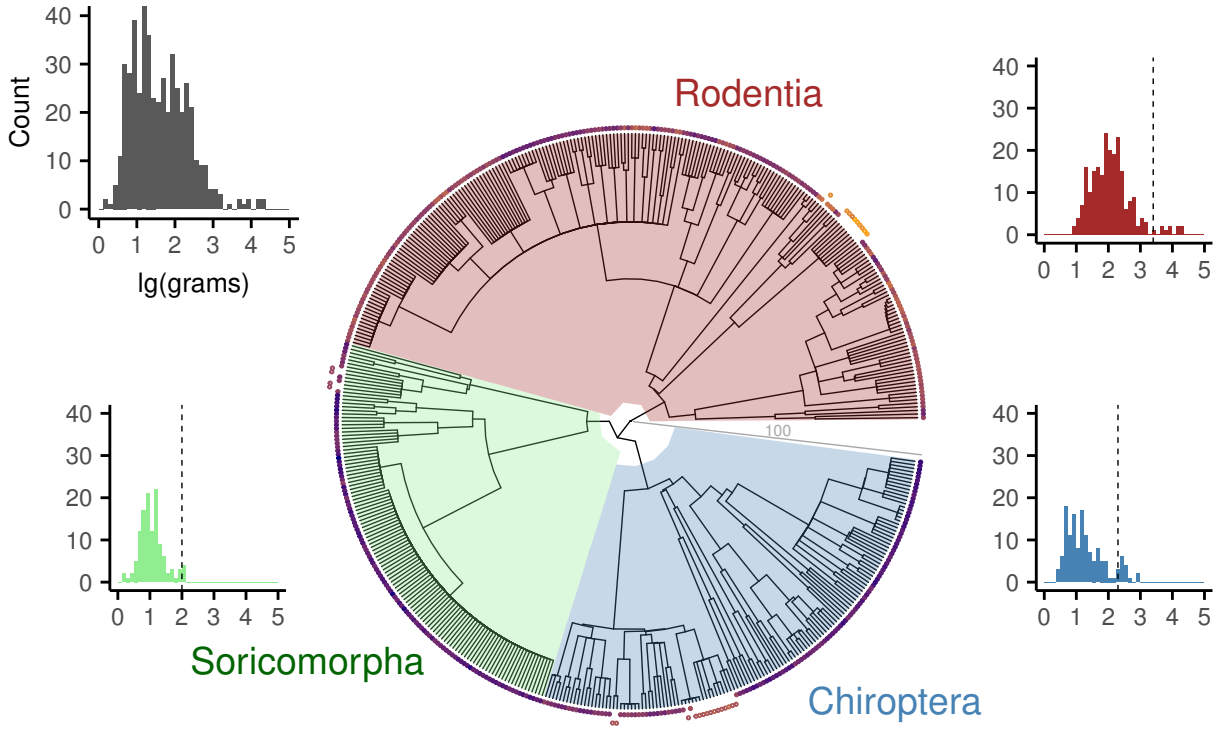


Figure 3: Phylogenetic tree and body weights of 227 Rodentia, 138 Chiroptera and 122 Soricomorpha species (extracted from mammal supertree published in Bininda-Emonds et al. (2007)). Colored bullets represent color-coded lg(body-mass) (blue: low, orange: high, data from Raia, Carotenuto, and Meiri (2010)); outliers ("giants") species are shown as colored circles positionend slightly outwards; body weight distributions are represented as colored histograms (corresponding to each order), a grey histogram in the top-left corner representing the total body weight distribution across the three orders, including outliers. Dashed vertical bars represent the outlier cut-off for each order. A grey line from the root to the tips indicates the time-scale (length) of the tree in million years (Myr).

For these four trees, we compared the maximum likelihood fits of the following five models:

- PBM / brown: Brownian motion assuming full phylogenetic heritability ($\sigma_e = 0$). For this model, we used our implementation based on the POUMM package (parametrization $< \theta, \sigma > \rightarrow < g_0 \equiv \theta, \alpha = 0, \theta, \sigma, \sigma_e = 0 >$) and the implementation

in the R-package `ouch` (function `brown`). The MLEs for the two implementations were matching exactly, except for the tree "All", where the `brown` function returned maximum log-likelihood of $-\infty$.

- POU($\int g_0$) / hansen: Ornstein-Uhlenbeck assuming full phylogenetic heritability ($\sigma_e = 0$). For this model, we tested our implementation based on the POUMM package (parametrization $< \alpha, \theta, \sigma > \rightarrow < \int_{-\infty}^{\infty}, \alpha, \theta, \sigma, \sigma_e = 0 >$) and the implementation in the R-package `ouch v2.9.2`, function `hansen` with one global selection regime (Butler and King 2004). By the notation $\int_{-\infty}^{\infty} g_0$, we mean that the OU-likelihood is defined as the expectation of the conditional likelihood on g_0, assuming that the distribution of g_0 is the long-term equilibrium (stationary) OU-distribution. The log-likelihood values and the MLEs for the two implementations were matching up to the 5th decimal digit (table 2).
- POU: Ornstein-Uhlenbeck assuming full phylogenetic heritability ($\sigma_e = 0$) and substituting the value of θ for the parameter g_0. While the only difference of this model with the Hansen model above is that it replaces integration over g_0 with a concrete value, this model includes PBM/brown as a special case ($\alpha = 0$) and, therefore, one can use a likelihood-ratio test for model selection. For this model, we tested our implementation based on the POUMM package (parametrization $< \alpha, \theta, \sigma > \rightarrow < g_0 \equiv \theta, \alpha, \theta, \sigma, \sigma_e = 0 >$). The resulting maximum log-likelihoods and MLEs were very close to the `hansen` estimates (table 2).
- PMM: The phylogenetic mixed model (Lynch 1991; Housworth, Martins, and Lynch 2004), i.e. Brownian motion plus error ($\sigma_e \geq 0$). For this model, we used our implementation based on the POUMM package with parametrization $< g_0, H^2_{\bar{t},\sigma_e} > \rightarrow < g_0 \equiv \theta, \alpha = 0, \theta, \sigma, \sigma_e = 0 >$, where the parameter σ is calculated from eq. 15, after setting α to 0.
- POUMM: The phylogenetic Ornstein-Uhlenbeck mixed model. For this model, we

used our implementation based on the POUMM package with parametrization $< g_0, H^2_{\bar{t},\sigma_e} > \rightarrow < g_0 \equiv \theta, \alpha = 0, \theta, \sigma, \sigma_e = 0 >$, where the parameter σ is calculated from eq. 15.

The results from fitting the above models to the four mammal trees are written in table 2. According to best (lowest) Akaike information criterion corrected for finite sample size (AICc), all four trees are unanimous about the best fit being the PMM model. The POUMM method MLE is very similar to the PMM MLE in all cases, but is the second-best model in terms of AICc, because it gets an extra penalty for the extra-parameter α, which's MLE is approximately 0 in all trees. This reveals stronger support for neutral drift evolution (i.e. BM) compared to stabilizing selection acting at the *class* and *order* levels. However, if the model would not account for a non-heritable component in the trait and would assume $H^2_{\bar{t}} = 1$, i.e. the PBM/brown, Hansen and POU models, the model selection would turn in favour of the OU-models. This suggests that the conclusion on whether or not stabilizing selection acts on the *class* and *order* levels depends strongly on the inclusion of a non-heritable component in the model, even if this component explains only a small relative proportion of the total phenotypic variance (PMM estimates of phylogenetic heritablity above 93% for all trees, except Soricomorpha).

Table 2: Model selection criteria and MLEs for five phylogenetic models fitted to mammal body-weight data. The abbreviation LR refers to a likelihood ratio test statistics against a nested null-model (PBM or PMM); asterisks denoting significant p-values: $* < 0.05$, $** < 0.01$, $*** < 0.001$. The best (the lowest) AICc values for each tree are shown in bold. Fixed parameters are written in grey. The parameters g_0 and σ_e are not shown, since g_0 is fixed or integrated over (see text) and the positive value of σ_e can be calculated from the phylogenetic heritability $H_{\bar{t}}^2$. The half-life $t_{1/2}$ is given in Millions of years (Myr).

Model	N	dof	log-lik.	LR vs PBM	LR vs PMM	AICc	MLE: $< \alpha \times 100, \theta, \sigma \times 10, H_{\bar{t}}^2, t_{1/2} >$
Rodentia							
PBM / brown	217	2	-66.33	0	-	136.72	$< 0, 2.04, 0.81, 1, \infty >$
POU($\int g_0$)/hansen	217	3	-64.61	-	-	135.32	$< 1.37, 2.03, 0.91, 1, 50.75 >$
POU	217	3	-64.18	4.3*	-	134.48	$< 1.23, 2.03, 0.9, 1, 56.46 >$
PMM	217	3	-63.68	5.3*	0	**133.48**	$< 0, 2.04, 0.71, 0.96, \infty >$
POUMM	217	4	-63.46	5.75	0.45	135.11	$< 0.57, 2.04, 0.77, 0.97, 121.19 >$
Chiroptera							
PBM/brown	123	2	2.13	0	-	-0.16	$< 0, 1.15, 0.59, 1, \infty >$
POU($\int g_0$)/hansen	123	3	5.11	-	-	-4.02	$< 1.8, 1.12, 0.68, 1, 38.58 >$
POU	123	3	5.5	6.74**	-	-4.79	$< 1.67, 1.13, 0.67, 1, 41.61 >$
PMM	123	3	11.89	19.53***	0	**-17.59**	$< 0, 1.14, 0.43, 0.93, \infty >$
POUMM	123	4	11.89	19.53***	0	-15.45	$< 0, 1.14, 0.43, 0.93, \infty >$
Soricomorpha							
PBM/brown	118	2	-24.24	0	-	52.58	$< 0, 1.2, 0.74, 1, \infty >$
POU($\int g_0$)/hansen	118	3	-23.24	-	-	52.68	$< 1.72, 1.2, 0.84, 1, 40.3 >$
POU	118	3	-23.05	2.37	-	52.32	$< 1.54, 1.2, 0.83, 1, 44.97 >$
PMM	118	3	-22.19	4.09*	0	**50.6**	$< 0, 1.2, 0.46, 0.8, \infty >$
POUMM	118	4	-22.19	4.09	0	52.74	$< 0, 1.2, 0.46, 0.8, \infty >$
All							
PBM/brown	458	2	-98.36	0	-	200.74	$< 0, 1.56, 0.74, 1, \infty >$
POU($\int g_0$)/hansen	458	3	-96.42	-	-	198.9	$< 0.8, 1.55, 0.79, 1, 86.78 >$
POU	458	3	-95.65	5.41*	-	197.35	$< 0.74, 1.55, 0.79, 1, 94.25 >$
PMM	458	3	-90.41	15.89***	0	**186.88**	$< 0, 1.56, 0.63, 0.96, \infty >$
POUMM	458	4	-90.4	15.9***	0.01	188.9	$< 0.07, 1.53, 0.64, 0.96, 1041.29 >$

Performance

Likelihood calculation time.—

We measured the time for one likelihood calculation on random ultrametric and non-ultrametric trees of up to 100,000 tips running on an Intel(R) Xeon(R) CPU E5-2697 v2 at 2.70GHz processor with $X \in \{1, 2, 4, 6, 8, 10\}$ parallel cores. We compared the serial

and parallel pruning implementatoins of the POUMM package to several serial pruning implementations provided in the R-packages geiger (Pennell et al. 2014), and diversitree (FitzJohn 2012). To measure the likelihood calculation times we used the R-package microbenchmark (Mersmann 2015) with argument `times` set to 100. All of the above implementations were compiled from source-code using the R-command `install.packages('package-directory', repos=NULL, type='source')`, and the same C++ compiler and compiler arguments (version 16.0.0 of the Intel compiler, command `icpc` with options `-O3 -march=corei7-avx -mavx`). The resulting average times in milliseconds are shown on fig. 4. On small trees of 100 tips, the fastest implementation was the serial implementation from the package diversitree (0.05 ms) followed by the POUMM `omp-for-simd-on-1-core` (0.07 ms). On trees of 1000 tips, the fastest implementation was the POUMM `omp-for-simd-on-1-core` (0.12 ms) followed by the parallel POUMM implementation `omp-for-simd-on-X-cores` and `omp-for-on-X-cores` (below 0.2 ms). With $N > 1000$ tips, the simd and multicore parallelization resulted in up to 13 times speed-up compared to non-simd serial POUMM C++ implementations (Fig. 5) and speed-ups of up to two orders of magnitude when comparing to serial pruning implementations (Fig. 4 and Fig. 5). These results show that the parallel efficiency tends to increase with N, so that on big trees, or in cases of smaller trees but numerous traits, a parallel pruning implementation could potentially achieve 100% parallel efficiency. Also noteworthy is that, thanks to the single instruction multiple data (simd) techology, parallelization is also possible on a single core. This is why, the time for the omp-for-simd implementation on a single core (5 ms for $N = 100,000$) is several times shorter than the time for the omp-for and the Armadillo based implementations. This also explains why the parallel speed-up is bigger than the number of cores when comparing the `omp-for-simd on X cores` to the non-simd (`omp-for on 1 core`) implementations.

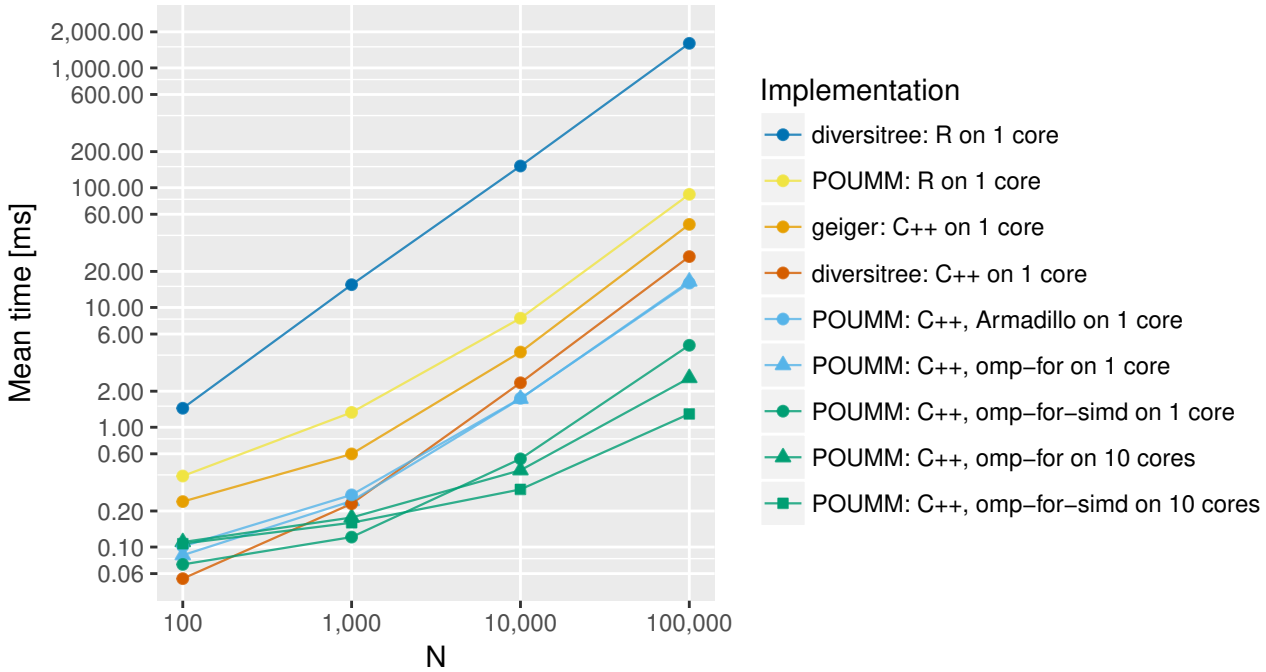


Figure 4: Likelihood calculation times for R and C++ implementations of the pruning algorithm.

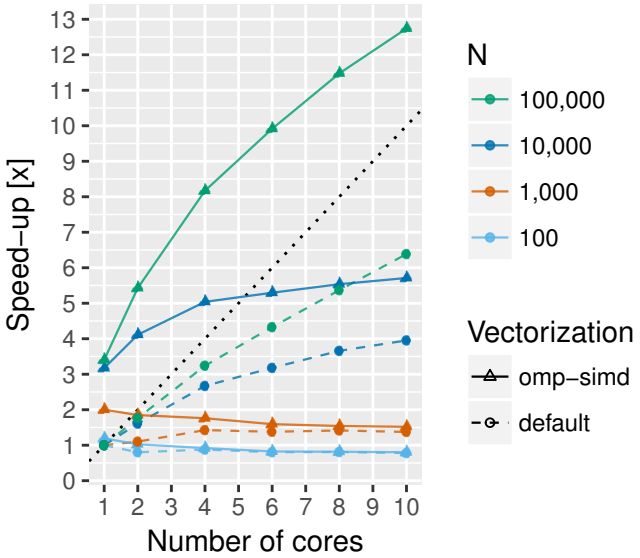


Figure 5: Parallel pruning speed-up on multiple CPU cores. Dashed lines represent the measured parallel speed-up of omp-for implementations with respect to the serial non-simd (omp-for) implementation running on one core; continuous lines represent the parallel speed-up of omp-for-simd implementation with respect to the serial non-simd (omp-for) implementation.

Improved MCMC convergence and MLE inference through adaptive Metropolis sampling.—

To measure the MCMC convergence speed-up from the adaptive Metropolis sampling, we reran one simulation scenario (2000 replications on a non-ultrametric tree of 4000 tips) with disabled adaptation. As a criterion for convergence, we used the absolute difference from 1 of the Gelman-Rubin convergence diagnostic (Brooks and Gelman 1998) (the closer $|G.R. - 1|$ is to 0, the better the convergence). When enabling adaptive Metropolis sampling, more than 1600 (80%) of the 2000 replications had reached $|G.R. - 1| < 0.01$ after a million iterations. Conversely, when disabling adaptive Metropolis sampling, less than 300 (15%) of the replications had reached $|G.R. - 1| < 0.01$ after a million iterations (the 80% quantile of $|G.R. - 1|$ was equal to 0.57, indicating very power convergence). We also noticed that 1455 out of 2000 replications (73%) of the POUMM inferences with enabled adaptative Metropolis sampling resulted in an improved MLE after running the MCMC chains, compared to 1045 (50%) when disabling adaptation. These observations show that adaptive Metropolis sampling considerably accelerates the MCMC convergence towards the posterior distribution and can be used to improve the MLE inference when using a weak prior or a prior that does not strongly contradict with the evidence (likelihood).

Discussion

The OU process has been applied as a model for stabilizing selection in macro-evolutionary studies (LANDE 1976; Felsenstein 1988; Hansen 1997; L. J. Harmon et al. 2010; Raia and Meiri 2011). Most of these studies assume that the whole trait evolves according to an OU process, usually disregarding the presence of a biologically relevant non-heritable component. When modelling species trait evolution, a non-heritable component including ecological contribution, measurement error, model residual error and possible inaccucies in the phylogenetic tree may be well justified and may in fact be

important to understand the full evolutionary process. In our empirical example, we have shown that the evolution of body weight in three orders of mammals is best described by the sum of a Brownian motion process and a non-heritable white noice (PMM). Except for the Soricomorpha tree, in which most of the species are in an unresolved polytomy descending from a single ancestor (12.6 Myr), the relative contribution of white noice to the trait variance is estimated between 3% to 7%. Strikingly, it is the inclusion or not of this small contribution in the model that decides the choice between a BM or an OU model of evolution.

Recently, the OU process has been applied in a micro-evolutionary context as a model for the evolution of pathogen traits, such as set-point virus load (spVL) during HIV infection (Mitov and Stadler (2016), Blanquart et al. (2017), Bertels et al. (2017), Bachmann et al. (2017)). When modeling pathogen evolution, the branching points in the tree represent transmission events, and the non-heritable component includes the contribution of the host immune system, the environment and the measurement and phylogenetic error. Thus, for pathogens, it is crucial to incorporate e in the model in order to quantify the importance of host- versus pathogen factors in trait formation (Alizon et al. 2010; Shirreff et al. 2013). In contrast to our macro-evolutionary example, based on AICc, the above studies have selected an OU process with added white noice (POUMM) as the best model for the evolution of spVL and CD4.

In agreement with our simulations, the above empirical examples provide evidence that mixed models are more appropriate than PBM and POU for modeling the evolution of continuous traits in epidemiology and macroevolution. Note also that the concept of model mixing and phylogenetic heritability is applicable to any phylogenetic model, beyond BM and OU.

Our R-package joins a growing collection of tools implementing phylogenetic OU inference. Among others, these include the R-packages `ape v4.0` (E Paradis, Claude, and

Strimmer 2004), `ouch v2.9-2` (Butler and King 2004), `GLSME v1.0.3` (Hansen and Bartoszek 2012), `diversitree v0.9-9` (FitzJohn 2012), `geiger v2.0.6` (Pennell et al. 2014), `surface v0.4-1` (Ingram and Mahler 2013), `mvMORPH v1.0.8` (Clavel, Escarguel, and Merceron 2015), `bayou v1.0.0` (Uyeda, Eastman, and Harmon 2015), `OUwie v1.50` (Beaulieu and OMeara 2016), `phylolm v2.5` (Ho and Ané 2014), `RPANDA v1.1` (Manceau, Lambert, and Morlon 2016). It may come as a surprise that most of the above package versions do not implement inference of a non-heritable component, i.e. a parameter σ_e. To our knowledge, of the above mentioned package versions only geiger and GLSME allow σ_e to be estimated from the phylogeny. However geiger does not seem to support non-ultrametric trees (mismatching likelihood values on non-ultrametric trees), and GLSME is much slower than the other packages. Most of the above packages allow the specification of a fixed standard measurement error before the model is fit to the data. This seems useful in macro-evolutionary studies where the standard measurement error can be estimated from the observed empirical variance within a species and the finite sample size. However, estimating a parameter σ_e from the data is still necessary, because measurement error is only one of many non-heritable factors contributing to the trait.

The idea to infer phylogenetic heritability assuming that g follows an OU process along the phylogeny has so far been discouraged mainly for interpretational and practical reasons: (i) in biology, individuals get selected based on their whole trait values, rather than the genotypic component g; (ii) small ultrametric macro-evolutionary trees do not contain sufficient signal for a simultaneous inference of the OU-and environmental variance (Housworth, Martins, and Lynch 2004). We argue that modeling an OU process on the whole trait value rather than g comes at the cost of additional parameters and reduced statistical power, because it necessitates to account for jumps in the trait value at the branching points as well as the unobserved speciation/transmission events along the tree. Conversely, assuming that the OU process acts directly on g is mathematically more

convenient, because it allows the inference of a single continuous OU-process along the tree, while adding e only at the tips of the tree. The implications of these simplified assumptions must be validated through simulations as done, e.g. in toy model simulations in Mitov and Stadler (2016).

Finally, we have shown the gain in speed performance from parallelizing the likelihood calculation of a univariate Ornstein-Uhlenbeck model. A main advantage of the parallel pruning algorithm with respect to sequential pruning implementations, e.g. diversitree (FitzJohn 2012) and geiger (Pennell et al. 2014), is that most of the algebraic calculations are done on vectors instead of single numbers and can be executed in parallel on contemporary SIMD and multicore systems. Languages such as Matlab and R are optimized for vector operations. This also explains why the serial POUMM pruning implementation in R is nearly as fast as serial pruning implementations written in C++ (Fig. 4). The parallelization scheme described here applies to all phylogenetic models where pruning can be used for likelihood calculation. Previously, it has been shown that this class of models spans over all multivariate Gaussian and some non-Gaussian models (Ho and Ané 2014). It is noteworthy that the performance benefit from parallelization increases with the complexity of the model (i.e. number of observable variables) and the size of the data (number of observations, N).

Another feature of the parallel pruning alogrithm is that it can be applied when parameters of the model change through time or across clades, e.g. when the values of the parameters are functions of geological time, geographic location or environment. Following this approach, the lineages of the tree can be cut into segments associated with different model regimes. These applications suggests that the parallel pruning algorithm has the potential to meet the challenges of increasing model complexity and volumes of data in comparative phylogenetic analysis.

Supplementary Material

Supplementary figures are available online. Data from the performance benchmarks, simulations and the analysis of mammal body weight is available on the dryad database. The POUMM package and user guide is available at `https://CRAN.R-project.org/package=POUMM`.

Funding

V.M. and T.S. thank ETH Zürich for funding. T.S. is supported in part by the European Research Council under the 7th Framework Programme of the European Commission (PhyPD: Grant Agreement Number 335529).

Acknowledgements

We thank Dr. Krzysztof Bartoszek for valuable insights on the Ornstein-Uhlenbeck process.

References

Alizon, Samuel, Viktor von Wyl, Tanja Stadler, Roger D Kouyos, Sabine Yerly, Bernard Hirschel, Jürg Böni, et al. 2010. "Phylogenetic approach reveals that virus genotype largely determines HIV set-point viral load." *PLoS Pathogens* 6 (9): e1001123.

Analytics, Revolution, and Steve Weston. 2015. "foreach: Provides Foreach Looping Construct for R."

Bachmann, Nadine, Teja Turk, Claus Kadelka, Alex Marzel, Mohaned Shilaih, Jürg Böni, Vincent Aubert, et al. 2017. "Parent-offspring regression to estimate the heritability

of an HIV-1 trait in a realistic setup." *Retrovirology* 14 (1): 33.

Bates, D, and M Maechler. 2017. "Matrix: Sparse and Dense Matrix Classes and Methods." *R Package Version 0999375-43*.

Beaulieu, Jeremy M, and Brian OMeara. 2016. "OUwie: Analysis of Evolutionary Rates in an OU Framework."

Beaulieu, Jeremy M, Dwueng-Chwuan Jhwueng, Carl Boettiger, and Brian C O'Meara. 2012. "MODELING STABILIZING SELECTION: EXPANDING THE ORNSTEIN-UHLENBECK MODEL OF ADAPTIVE EVOLUTION." *Evolution; International Journal of Organic Evolution* 66 (8): 2369–83.

Bertels, Frederic, Alex Marzel, Gabriel Leventhal, Venelin Mitov, Jacques Fellay, Huldrych F Günthard, Jürg Böni, et al. 2017. "Dissecting HIV Virulence: Heritability Of Setpoint Viral Load, CD4+ T Cell Decline And Per-Parasite Pathogenicity."

Bininda-Emonds, Olaf R P, Marcel Cardillo, Kate E Jones, Ross D E MacPhee, Robin M D Beck, Richard Grenyer, Samantha A Price, Rutger A Vos, John L Gittleman, and Andy Purvis. 2007. "The delayed rise of present-day mammals." *Nature* 446 (7135): 507–12.

Blanquart, François, Chris Wymant, Marion Cornelissen, Astrid Gall, Margreet Bakker, Daniela Bezemer, Matthew Hall, et al. 2017. "Viral genetic variation accounts for a third of variability in HIV-1 set-point viral load in Europe." *Plos Biology* In press (personal communication).

Boskova, Veronika, Sebastian Bonhoeffer, and Tanja Stadler. 2014. "Inference of Epidemiological Dynamics Based on Simulated Phylogenies Using Birth-Death and Coalescent Models." *PLoS Computational Biology (PLOSCB) 10(4)* 10 (11): e1003913.

Boyd, Stephen P, and Lieven Vandenberghe. 2004. *Convex Optimization*. Cambridge University Press.

Brooks, S P, and A Gelman. 1998. "General methods for monitoring convergence of

iterative simulations." *Journal of Computational and Graphical Statistics* 7 (4): 434–55.

Butler, M A, and A A King. 2004. "Phylogenetic comparative analysis: A modeling approach for adaptive evolution." *American Naturalist* 164 (6): 683–95.

Byrd, Richard H, Peihuang Lu, Jorge Nocedal, and Ci You Zhu. 1995. "A limited memory algorithm for bound constrained optimization." *SIAM Journal on Scientific Computing* 16 (5): 1190–1208.

Clavel, Julien, Gilles Escarguel, and Gildas Merceron. 2015. "mvmorph: an r package for fitting multivariate evolutionary models to morphometric data." *Methods in Ecology and Evolution* 6 (11): 1311–9.

Cook, Samantha R, Andrew Gelman, and Donald B Rubin. 2006. "Validation of Software for Bayesian Models Using Posterior Quantiles." *Journal of Computational and Graphical Statistics* 15 (3): 675–92.

Cooper, Natalie, Gavin H Thomas, Chris Venditti, Andrew Meade, and Rob P Freckleton. 2015. "A cautionary note on the use of Ornstein Uhlenbeck models in macroevolutionary studies." *Biological Journal of ...* 118 (1): 64–77.

Dowle, Matthew, T Short, S Liangolou, and A Srinivasan. 2014. "data.table: Extension of data.frame," July, 9.

Eddelbuettel, Dirk, and Conrad Sanderson. 2014. "RcppArmadillo - Accelerating R with high-performance C++ linear algebra." *Computational Statistics & Data Analysis* 71: 1054–63.

Felsenstein, J. 1973. "Maximum-likelihood estimation of evolutionary trees from continuous characters." *American Journal of Human Genetics* 25 (5): 471–92.

———. 1988. "Phylogenies And Quantitative Characters." *Annual Review of Ecology and Systematics* 19 (1): 445–71.

FitzJohn, Richard G. 2012. "Diversitree: comparative phylogenetic analyses of

diversification in R." *Methods in Ecology and Evolution* 3 (6): 1084–92.

Goolsby, Eric W, Jorn Bruggeman, and Cécile Ané. 2016. "Rphylopars: fast multivariate phylogenetic comparative methods for missing data and within-species variation." *Methods in Ecology and Evolution* 8 (1): 22–27.

Grimmett, Geoffrey, and David Stirzaker. 2001. *Probability and Random Processes*. Oxford University Press.

Hankin, RKS. 2006. "Special functions in R: introducing the gsl package." *R News*.

Hansen, Thomas F. 1997. "Stabilizing Selection and the Comparative Analysis of Adaptation." *Evolution; International Journal of Organic Evolution* 51 (5): 1341–51.

Hansen, Thomas F, and Krzysztof Bartoszek. 2012. "Interpreting the evolutionary regression: the interplay between observational and biological errors in phylogenetic comparative studies." *BioRxiv* 61 (3): 413–25.

Harmon, Luke J, Jonathan B Losos, T Jonathan Davies, Rosemary G Gillespie, John L Gittleman, W Bryan Jennings, Kenneth H Kozak, et al. 2010. "Early bursts of body size and shape evolution are rare in comparative data." *Evolution; International Journal of Organic Evolution* 64 (8): 2385–96.

Ho, Lam si Tung, and Cécile Ané. 2014. "A linear-time algorithm for Gaussian and non-Gaussian trait evolution models." *BioRxiv* 63 (3): 397–408.

Hodcroft, Emma, Jarrod D Hadfield, Esther Fearnhill, Andrew Phillips, David Dunn, Siobhan O'Shea, Deenan Pillay, Andrew J Leigh Brown, on behalf of the UK HIV Drug Resistance Database, and the UK CHIC Study. 2014. "The Contribution of Viral Genotype to Plasma Viral Set-Point in HIV Infection." *PLoS Pathogens* 10 (5): e1004112.

Housworth, Elizabeth A, Emília P Martins, and Michael Lynch. 2004. "The phylogenetic mixed model." *The American Naturalist* 163 (1): 84–96.

Ingram, Travis, and D Luke Mahler. 2013. "SURFACE: detecting convergent evolution from comparative data by fitting Ornstein-Uhlenbeck models with stepwise

Akaike Information Criterion." *Methods in Ecology and Evolution* 4 (5): 416–25.

Ives, A R, and T Garland. 2009. "Phylogenetic Logistic Regression for Binary Dependent Variables." *BioRxiv* 59 (1): 9–26.

LANDE, R. 1976. "Natural-Selection and Random Genetic Drift in Phenotypic Evolution." *Evolution; International Journal of Organic Evolution* 30 (2): 314–34.

Lynch, Michael. 1991. "Methods for the Analysis of Comparative Data in Evolutionary Biology." *Evolution; International Journal of Organic Evolution* 45 (5): 1065–80.

Lynch, Michael, and Bruce Walsh. 1998. *Genetics and Analysis of Quantitative Traits*. Sinauer Associates Incorporated.

Maechler, Martin. 2014. "Rmpfr: R MPFR - Multiple Precision Floating-Point Reliable."

Manceau, Marc, Amaury Lambert, and Helene Morlon. 2016. "A unifying comparative phylogenetic framework including traits coevolving across interacting lineages." *BioRxiv*, December, syw115.

Mersmann, Olaf. 2015. "microbenchmark: Accurate Timing Functions."

Mitov, Venelin, and Tanja Stadler. 2016. "The heritability of pathogen traits - definitions and estimators." *BioRxiv*, June, 058503.

Ornstein, L S, and F Zernike. 1919. "The theory of the Brownian Motion and statistical mechanics." *Proceedings of the Koninklijke Akademie Van Wetenschappen Te Amsterdam* 21 (1/5): 109–14.

O'Meara, Brian C. 2012. "Evolutionary Inferences from Phylogenies: A Review of Methods." *Annual Review of Ecology, Evolution, and Systematics* 43 (1): 267–85.

Paradis, E, J Claude, and K Strimmer. 2004. "APE: Analyses of Phylogenetics and Evolution in R language." *Bioinformatics* 20 (2): 289–90.

Paradis, Emmanuel, and Julien Claude. 2002. "Analysis of comparative data using

generalized estimating equations." *Journal of Theoretical Biology* 218 (2): 175–85.

Pennell, Matthew W, Jonathan M Eastman, Graham J Slater, Joseph W Brown, Josef C Uyeda, Richard G FitzJohn, Michael E Alfaro, and Luke J Harmon. 2014. "geiger v2.0: an expanded suite of methods for fitting macroevolutionary models to phylogenetic trees." *Bioinformatics* 30 (15): 2216–8.

Plummer, Martyn, Nicky Best, Kate Cowles, and Karen Vines. 2006. "CODA: Convergence Diagnosis and Output Analysis for MCMC." *R News* 6: 7–11.

R Core Team. 2013. "R: A Language and Environment for Statistical Computing."

Raia, Pasquale, and Shai Meiri. 2011. "The tempo and mode of evolution: body sizes of island mammals." *Evolution; International Journal of Organic Evolution* 65 (7): 1927–34.

Raia, Pasquale, Francesco Carotenuto, and Shai Meiri. 2010. "One size does not fit all: no evidence for an optimal body size on islands." *Global Ecology and ...* 25 (4): 913–484.

Sanderson, Conrad, and Ryan Curtin. 2016. "Armadillo: a template-based C++ library for linear algebra." *Journal of Open Source Software* 1 (2).

Scheidegger, Andreas. 2012. "adaptMCMC: Implementation of a generic adaptive Monte Carlo Markov Chain sampler."

Schloerke, Barret, Jason Crowley, Di Cook, Francois Briatte, Moritz Marbach, Edwin Thoen, Amos Elberg, and Joseph Larmarange. 2016. "GGally: Extension to 'ggplot2'."

Shirreff, George, Samuel Alizon, Anne Cori, Huldrych F Günthard, Oliver Laeyendecker, Ard van Sighem, Daniela Bezemer, and Christophe Fraser. 2013. "How effectively can HIV phylogenies be used to measure heritability?" *Evolution, Medicine, and Public Health* 2013 (1): 209–24.

Smith, Felisa A, S Kathleen Lyons, S K Morgan Ernest, Kate E Jones, Dawn M

Kaufman, Tamar Dayan, Pablo A Marquet, James H Brown, and John P Haskell. 2003. "BODY MASS OF LATE QUATERNARY MAMMALS." *Ecology* 84 (12): 3403–3.

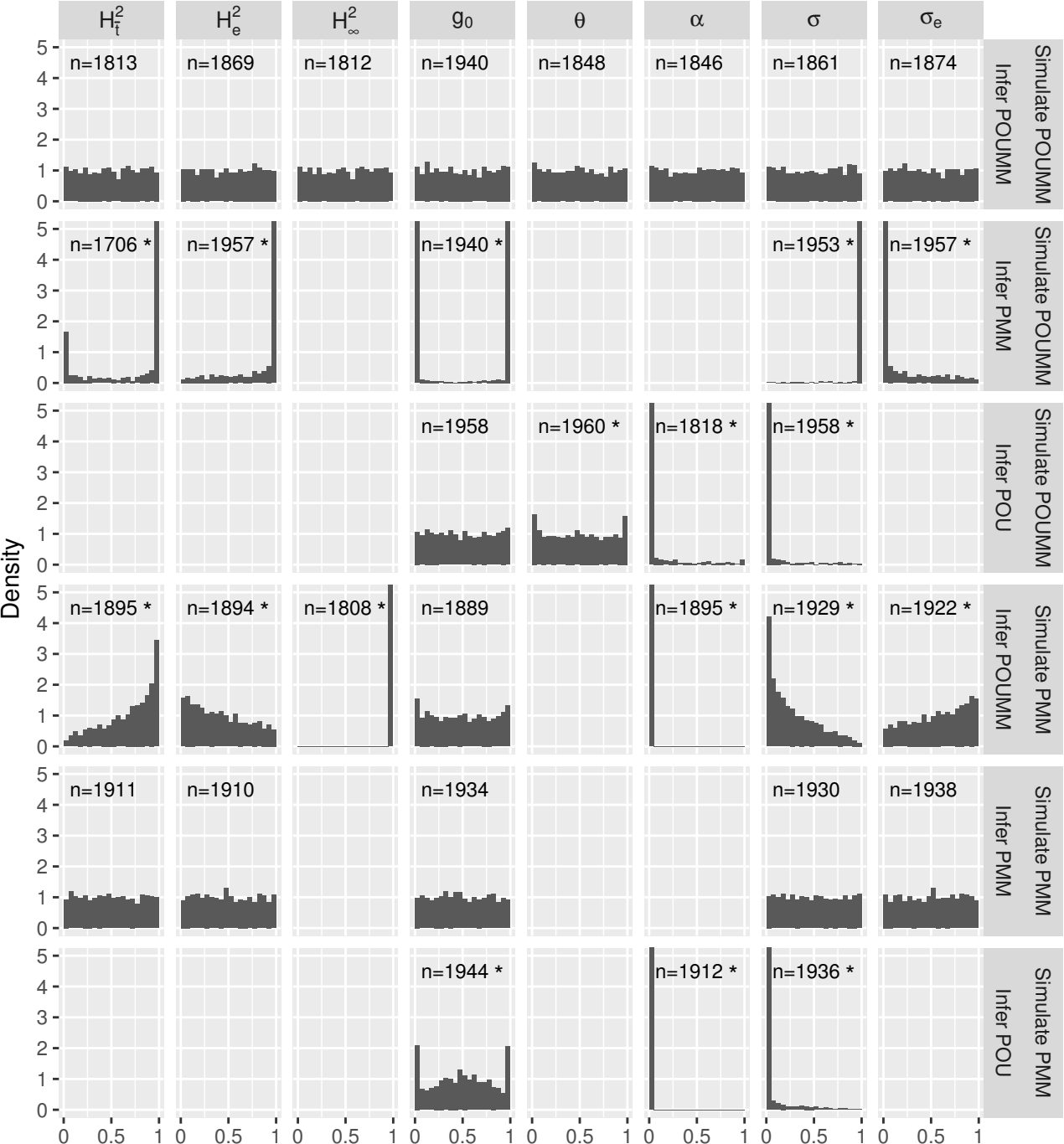
Stadler, Tanja, Denise Kühnert, Sebastian Bonhoeffer, and Alexei J Drummond. 2013. "Birth-death skyline plot reveals temporal changes of epidemic spread in HIV and hepatitis C virus (HCV)." *PNAS* 110 (1): 228–33.

Uhlenbeck, G E, and L S Ornstein. 1930. "On the Theory of the Brownian Motion." *Physical Review* 36 (5): 823–41.

Uyeda, Josef C, Jon Eastman, and Luke Harmon. 2015. "bayou: Bayesian Fitting of Ornstein-Uhlenbeck Models to Phylogenies."

Vihola, Matti. 2012. "Robust adaptive Metropolis algorithm with coerced acceptance rate." *Statistics and Computing* 22 (5): 997–1008.

Wickham, Hadley. 2009. "ggplot2 - Elegant Graphics for Data Analysis." *Use R*.


H_t^2
H_e^2
H_∞^2
g_0
θ
α
σ
σ_e
Density
Simulate POUMM
Infer POUMM
n=1813
n=1869
n=1812
n=1940
n=1848
n=1846
n=1861
n=1874
Simulate POUMM
Infer PMM
n=1706 *
n=1957 *
n=1940 *
n=1953 *
n=1957 *
Simulate POUMM
Infer POU
n=1958
n=1960 *
n=1818 *
n=1958 *
Simulate PMM
Infer POUMM
n=1895 *
n=1894 *
n=1808 *
n=1889
n=1895 *
n=1929 *
n=1922 *
Simulate PMM
Infer PMM
n=1911
n=1910
n=1934
n=1930
n=1938
Simulate PMM
Infer POU
n=1944 *
n=1912 *
n=1936 *
0
1
2
3
4
5
0.5