



Automatic generation of evolutionary hypotheses using mixed Gaussian phylogenetic models

Venelin Mitov^{a,b,1}, Krzysztof Bartoszek^c, and Tanja Stadler^{a,b}

^aComputational Evolution Group, Department of Biosystems Science and Engineering, ETH Zürich, 4058 Basel, Switzerland; ^bComputational Evolution Group, Swiss Institute of Bioinformatics (SIB), 4058 Basel, Switzerland; and ^cDivision of Statistics and Machine Learning, Department of Computer and Information Science, Linköping University, 58183 Linköping, Sweden

Edited by Scott V. Edwards, Harvard University, Cambridge, MA, and approved July 1, 2019 (received for review August 13, 2018)

Phylogenetic comparative methods are widely used to understand and quantify the evolution of phenotypic traits, based on phylogenetic trees and trait measurements of extant species. Such analyses depend crucially on the underlying model. Gaussian phylogenetic models like Brownian motion and Ornstein–Uhlenbeck processes are the workhorses of modeling continuous-trait evolution. However, these models fit poorly to big trees, because they neglect the heterogeneity of the evolutionary process in different lineages of the tree. Previous works have addressed this issue by introducing shifts in the evolutionary model occurring at inferred points in the tree. However, for computational reasons, in all current implementations, these shifts are “intramodel,” meaning that they allow jumps in 1 or 2 model parameters, keeping all other parameters “global” for the entire tree. There is no biological reason to restrict a shift to a single model parameter or, even, to a single type of model. Mixed Gaussian phylogenetic models (MGPMs) incorporate the idea of jointly inferring different types of Gaussian models associated with different parts of the tree. Here, we propose an approximate maximum-likelihood method for fitting MGPMs to comparative data comprising possibly incomplete measurements for several traits from extant and extinct phylogenetically linked species. We applied the method to the largest published tree of mammal species with body- and brain-mass measurements, showing strong statistical support for an MGPM with 12 distinct evolutionary regimes. Based on this result, we state a hypothesis for the evolution of the brain–body–mass allometry over the past 160 million y.

correlated quantitative traits | selection | evolutionary regimes | clustering | nonultrametric tree

Life is extremely diverse as the result of the dynamic change in evolutionary forces driving speciation and phenotypic evolution (1). Gaussian phylogenetic models, such as Brownian motion (BM) and Ornstein–Uhlenbeck (OU) processes, have become a standard tool in the comparative analysis of quantitative traits (2, 3). Among many applications, these models have been used for more appropriately correcting for phylogeny in comparative regression analyses of morphological or pathogen traits (4–12) and for testing hypotheses about the evolutionary forces that have led to observable patterns in the traits of modern taxa (2, 3, 13). With ever-growing tree size and scope of the phylogenetic analysis, it is unlikely that a single regime of evolution described by a single model could have driven the changes in the traits across the entire tree. Such a model would have too low of a resolution to accommodate the inherent heterogeneity in the evolutionary process. Even worse, fitting a misspecified model to a large phylogeny is prone to inferring statistically significant, but strongly biased parameter values, due to their tendency to “compensate” for the modeling error (12, 14). There is no biological reason to constrain the change of a model regime to a single or a few model parameters, nor is there any reason to restrict the change to a single type of model. However, to the best of our knowledge, all current implementations inferring phylogenetic models with shifts impose such restrictions, motivated mainly by computability issues (15–23).

In this work, we propose a method for overcoming the computational complexity of fitting jointly a set of different model types with independent parameter sets to phylogenetically linked comparative data. Our approach relies on a subfamily, hereby denoted \mathcal{G}_{LInv} , of the Gaussian phylogenetic models, with the transition density exhibiting the properties that the expectation depends linearly on the ancestral trait value and the variance is invariant with respect to the ancestral value. In a related work, we have shown that the likelihood of such models can be calculated in time proportional to the number of nodes in the tree (24). Here, we generalize this fast likelihood calculation algorithm to mixed phylogenetic models over the \mathcal{G}_{LInv} family, which we denote mixed Gaussian phylogenetic models (MGPMs). We develop an algorithm for fast maximum-likelihood search of an optimal MGPM fitted to a dataset of possibly incomplete measurements from several traits of present-day species and/or fossilized specimens, annotating the tips of a time-calibrated tree.

A prominent example with a long history in evolutionary biology is the comparative analysis of brain- and body-mass data from mammals (25, 26). In the quest for the origin of intelligence,

Significance

Phylogenetic comparative methods (PCMs) are used to study the evolution of various biological species, ranging from microorganisms to animals and plants. These methods combine trait measurements, such as body masses measured in a set of species, with the species’ phylogenetic tree, to quantify the trait’s evolution along the tree. Here, we show that current PCMs fail to reproduce the patterns of evolution of brain and body mass in mammals, because they use mathematical models that cannot represent the heterogeneity of the evolutionary processes acting in different lineages of the tree. As a solution, we propose mixed Gaussian phylogenetic models allowing one to infer changes in the type and magnitude of evolutionary forces occurring on specific branches of the tree.

Author contributions: V.M. designed research; V.M. performed research; V.M. and K.B. contributed new reagents/analytic tools; V.M. analyzed data; and V.M. and T.S. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

Data deposition: The recursive clade partition algorithm for mixed Gaussian phylogenetic model inference was implemented in the R-package PCMFit and has been deposited in GitHub (<https://github.com/venelin/PCMFit>). The analysis of the mammal data has been implemented in the R-package MGPMmammals and has been deposited in GitHub (<https://github.com/venelin/MGPMmammals>). The simulation tests have been implemented in the R-package MGPMsimulations and have been deposited in GitHub (<https://github.com/venelin/MGPMsimulations>).

¹To whom correspondence may be addressed. Email: vmitov@gmail.com.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1813823116/-DCSupplemental.

Published online August 2, 2019.

it has been shown that, in mammals, brain mass has a negative allometric relationship with body mass, meaning that brain mass tends to scale at lower proportions with respect to body mass (25–28). Many studies have compared this allometry between separate mammal clades (e.g., refs. 27 and 28 and references therein). However, the choice of the groups to be compared in these studies has been driven mainly by the established taxonomic ranking (i.e., order, family, genus) and by the researcher's intuition about which groups "could" be different. Moreover, most of the studies in the past have neglected the phylogenetic relationship between the species within a group—a known source of bias in comparative regression analysis (4). More recent works did take the phylogeny into account but restricted the model to a single BM process over the entire tree (28). How far from the reality is such a "global" BM assumption? Can we make a data-driven choice of the groups to compare, based on the patterns of distinct macroevolutionary regimes? Can we infer the ancestral values of body and brain mass as well as their allometry from extant species data?

To address these questions, we performed an MGPM maximum-likelihood (ML) fit to body- and brain-mass data from 629 extant mammal species representative of 21 orders, extracted from the previous works of refs. 28 and 29. This revealed a trend of gradually decreasing brain–body-mass allometry, for a large paraphyletic group of nearly 400 species. Deviations from this trend were identified in 10 smaller groups manifesting different and, sometimes, contrasting patterns.

This article is organized as follows. In *Approaches*, we formulate the so-called intermodel shift problem, i.e., the optimization problem aiming at finding the optimal model shifts in a phylogenetic tree with multivariate trait measurements associated with its terminal nodes (tips). Then, we briefly describe our proposed solution based on the MGPM. In *Results*, we report the analysis of the mammal data. In *Discussion*, we interpret the results and discuss our methods in the light of existing methods and tools. A detailed description of the methods is provided in *Materials and Methods* and in *SI Appendix*. In *SI Appendix*, sections I–K, we report additional results from validation tests based on simulated data.

Approaches

The Intermodel Shift Problem. Given the number of traits k , a tree \mathcal{T} representing the evolutionary relationship of N species (tips), and a family of k -variate phylogenetic models \mathcal{M} , a mixed phylogenetic model on \mathcal{T} is defined as a configuration of shift points and mapped models, $S = \{<0, m_0>, <s_1, m_1>, \dots, <s_R, m_R>\}$, where $<0, m_0>$ denotes the initial model $m_0 \in \mathcal{M}$ starting from the root (0) and modeling the trait evolution on the descending lineages until reaching a tip from \mathcal{T} or another shift from S ; each other shift $<s_i, m_i>$ denotes a point s_i on a branch of \mathcal{T} and a model $m_i \in \mathcal{M}$, assuming the trait values at the point s_i as the initial state, and again modeling the evolution on the subtree with root s_i , \mathcal{T}_{s_i} , until reaching a tip or a shift. We call "shift-point configuration" the set of points where the shifts occur, i.e., $\{0, s_1, \dots, s_R\}$. We denote by $\mathcal{S}(\mathcal{T}, \mathcal{M})$ the family of all mixed phylogenetic models over \mathcal{T} and \mathcal{M} , with mixed referring to several models on a single tree. The "intermodel shift problem" is the problem of finding the mixed phylogenetic model $S^* \in \mathcal{S}(\mathcal{T}, \mathcal{M})$ that fits "best" to data \mathcal{X} consisting of trait values at the tips of \mathcal{T} . We call S^* the best intermodel shift configuration.

Defining "best fit" in the statistical sense is not straightforward, due to the notorious problem of "overfitting" coming along with complex parametric models. In this work, we use the Akaike information criterion (AIC) as a score function penalizing the ML fit of a model, based on the number of free parameters. We note, however, that there is no general agreement on a best scoring function, in particular, for small datasets, where the com-

monly used AIC and AICc have been shown to be biased toward more complex models (30, 31).

Dealing with the Computational Complexity. With a few exceptions (32), maximizing the likelihood of a mixed phylogenetic model is a multivariate nonconvex optimization task involving numerous calculations of the model likelihood for the given tree and data. Furthermore, searching for the best intermodel shift configuration is hard, because the number of possibilities to choose the branches for R shift points grows exponentially with respect to the number of tips in the tree. Our approach to this complexity is 2-fold:

The \mathcal{G}_{LInv} family of models. In particular, we restrict \mathcal{M} to a subfamily of the Gaussian phylogenetic models, denoted \mathcal{G}_{LInv} .

Definition. We say that a phylogenetic trait model belongs to the \mathcal{G}_{LInv} family if it satisfies the following:

- 1) After branching the traits evolve independently in the 2 descending lineages.
- 2) The distribution of the trait \vec{X} , at time t conditional on the value at time $s < t$, is Gaussian with the mean and variance satisfying

a) the expectation of $\vec{X}(t)$ conditional on $\vec{X}(s)$ is a linear function of $\vec{X}(s)$, i.e.,

$$E[\vec{X}(t)|\vec{X}(s)] = \vec{\omega}_{s,t} + \Phi_{s,t}\vec{X}(s);$$

b) the variance of $\vec{X}(t)$ conditional on $\vec{X}(s)$ is invariant with respect to (does not depend on) $\vec{X}(s)$, i.e.,

$$\text{Var}[\vec{X}(t)|\vec{X}(s)] = \mathbf{V}_{s,t},$$

for some vector $\vec{\omega}_{s,t}$ and matrices $\Phi_{s,t}, \mathbf{V}_{s,t}$, which may depend on s and t , but do not depend on $\vec{X}(\cdot)$.

The \mathcal{G}_{LInv} family includes the multivariate BM and OU processes, as well as many of their variants widely used in phylogenetic comparative methods (24). In ref. 24, we have proved that for any tree and any phylogenetic model satisfying the *Definition*, it is possible to calculate the likelihood of the model, given multi-trait data for the tips with some tips possibly missing some trait values, through a pruning algorithm, based on analytical integration over the unobserved trait values at the internal nodes of the tree. Here, we have extended this algorithm to support mixed phylogenetic models over the \mathcal{G}_{LInv} family, meaning the type of model may change at intermodel shift points.

Fast model selection. As a next step, we implemented a parallel recursive clade partition (RCP) algorithm solving the intermodel shift problem by returning an (approximate) optimal intermodel shift configuration for a given tree and multivariate trait data at the tips. This algorithm relies on several "heuristics" aiming to reduce 1) the number of candidate shift-point configurations and 2) the number of possible model type mappings to a given shift-point configuration. For the following sections, it is important to mention 2 of these heuristics:

- 1) We assume that a shift point can only occur at the beginning of a branch and we call the end node of such a branch a "shift node."
- 2) We introduce a threshold, q , on the minimal number of tips "visible" from an ancestor shift node, with visibility meaning that there are no other shifts on the paths from this shift node to any of the visible tips. This heuristic has a 2-fold benefit: First, it accelerates the search by dramatically reducing the number of candidate shift-point configurations. More importantly, as we show in simulations, this heuristic effectively

reduces the risk of model overfitting (*SI Appendix, section I*). The drawback of this heuristic is that shifts in clades and paraphyletic groups smaller than q tips will not be detected.

A detailed description of the RCP algorithm is provided in *SI Appendix, section A and Algorithm S1*.

Results

An MGPM Analysis of the Brain–Body Allometry in Mammals. We performed an MGPM fit to the biggest publicly available phylogenetic tree of mammal species with available body- and brain-mass measurements (Fig. 1). This is a subtree of 629 extant species with ancestral nodes spanning 166 Ma, which were extracted from the time-calibrated mammal tree published in ref. 29. Body- and brain-mass data were available in the form of mean estimates from finite samples, provided from previous works (ref. 28 and references therein). We used the available sample sizes and sample SDs for 144 body-mass and 87 brain-mass measurements to estimate SEs. For the species and traits, where no sample size and SD were available, we imputed the SE using linear regression on the corresponding body- and brain-mass mean estimates (*SI Appendix, section H*).

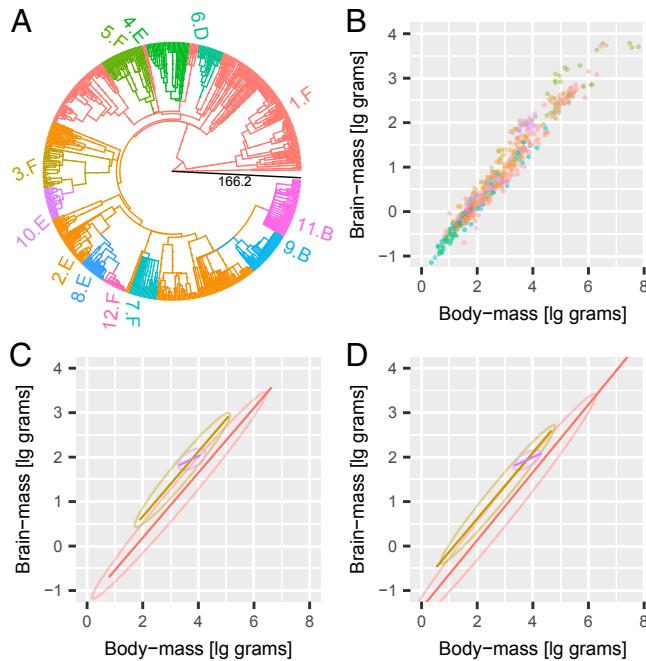


Fig. 1. An MGPM model of phylogenetically linked body- and brain-mass data from mammal species. (A) A tree of 629 extant species representative of 21 mammal orders (subsampled from ref. 29). (B) Body and brain masses measured as log₁₀-transformed mean values from finite samples of individual organisms from each species (curated measurements available from ref. 28). In A, a colored number followed by an uppercase letter denotes the regime number and model type selected for each of the 12 model regimes found in MGPM*. (C) “Standard” estimates for the 95% contours and linear regression line of brain mass on body mass for 3 regimes—1, 3, and 10. These estimates ignore the phylogenetic relationship, assuming independence of the data points in each group. (D) Expected 95% contours and regression lines for regimes 1, 3, and 10, according to MGPM*. Under the hypothesis that the inferred MGPM is the true model, the distributions in D represent the expectation at the present time for samples of species that have evolved independently from the root to an arbitrary tip in the corresponding regime following the regime shifts on that path. Thus, the MGPM* expectations correct for possible biases due to phylogenetic relationship. We observed an agreement between the standard estimates and the MGPM* expectations for most of the 12 groups (*SI Appendix, Fig. S2* and *Discussion*).

Table 1. Competing model fits to the mammal data

Model	q	R	p	$\ell\ell$	AIC	ΔAIC
Global BM_A	n.a.	1	4	−540.79	1,089.58	1,321.28
Global BM_B	n.a.	1	5	30.60	−51.19	180.51
Global OU_C	n.a.	1	8	−540.79	1,097.58	1,329.28
Global OU_D	n.a.	1	9	30.60	−43.19	188.51
Global OU_E	n.a.	1	10	47.62	−75.24	156.46
Global OU_F	n.a.	1	11	62.89	−103.77	127.93
SURFACE OU	20	1	8	−540.83	1,097.66	1,329.37
SCALAR OU	20	6	38	98.37	−120.74	110.97
RATEMATRIX BM	20	9	37	116.15	−158.30	73.40
MGPM* (A–F)	20	12	115	230.85	−231.70	0.00

q , minimal number of tips visible from a shift node; R , number of inferred regimes; p , number of parameters; $\ell\ell$, log-likelihood (higher values are better); AIC, Akaike information criterion (higher values are worse); ΔAIC , difference with respect to the best AIC score (higher values are worse); n.a., not applicable. The optimal parameter values of the models are described in *SI Appendix, section M and Tables S1–S10*. We note that, up to small error of the numerical optimization, the models BM_A , OU_C , and SURFACE OU converged to the same BM_A model (*SI Appendix, Tables S1, S3, and S7*). The SCALAR OU model was the third best fit to the data. This fit converged to a BM_B model with shifts (*SI Appendix, Table S8*). The fit of the RATEMATRIX BM model (which is also a BM_B model with shifts; *Materials and Methods*) resulted in the second-best AIC score.

Inference was done using the log-10-transformed trait values. For the MGPM, we searched for shifts over 6 candidate model types ranging from a model of neutrally and independently evolving traits to a complex model of evolution under selection and causal relationship between the traits. All of these model types were defined as specifications of the BM and the OU models (*Materials and Methods*). We denote these model types by BM_A , BM_B , OU_C , OU_D , OU_E , and OU_F or by the letters A–F.

The best MGPM fit found by the RCP algorithm had $AIC^* = -231.7$, log-likelihood $\ell\ell^* = 230.85$, and a total of $p = 115$ parameters specifying 11 shift points and 12 regimes. Further, we use the notation MGPM* to denote this model. We compared MGPM* to fits of competing models including global BM_A , ..., OU_F (no shifts), a SURFACE OU model (18), a SCALAR OU model (23), and a RATEMATRIX BM model (22). Since the SURFACE OU, the SCALAR OU, and the RATEMATRIX BM are in G_{LM} , we implemented these fits using the RCP algorithm, specifying the same setting for the threshold q as for MGPM* (*Materials and Methods*). This confirmed a significant advantage for MGPM* ($\Delta\text{AIC} > 73.40$ for all tested competing methods; Table 1).

To assess the confidence of MGPM*, we performed a model parametric bootstrap. In particular, we generated 50 bootstrap datasets, by simulating MGPM* on the mammal tree with the inferred shift points (Fig. 1A and *SI Appendix, section H*). Then, we reran the RCP algorithm over the models BM_A , ..., OU_F and the tree (without providing the shift-point configuration), for each simulated dataset (*SI Appendix, Figs. S5–S9*).

Using MGPM* and the MGPMs from the parametric bootstrap, we reconstructed the evolution of body mass, brain mass, and their allometric relationship since the root of the tree dated 166.2 Ma ago (Fig. 2). To that end, we first discretized the time interval into epochs at each 2 Ma from the root to the present time. Then, for each epoch, we inserted singleton nodes on all branches of the tree intersecting with this epoch. In doing this, we preserved the regime assignment (coloring) of the trees, both for MGPM* (Fig. 1A) and for the colored trees resulting from the parametric bootstrap inferences (*SI Appendix, Figs. S5–S7*). Finally, based on the inferred model parameters, for MGPM* and for each bootstrap MGPM, we calculated the expected body mass, brain mass, their expected variance–covariance

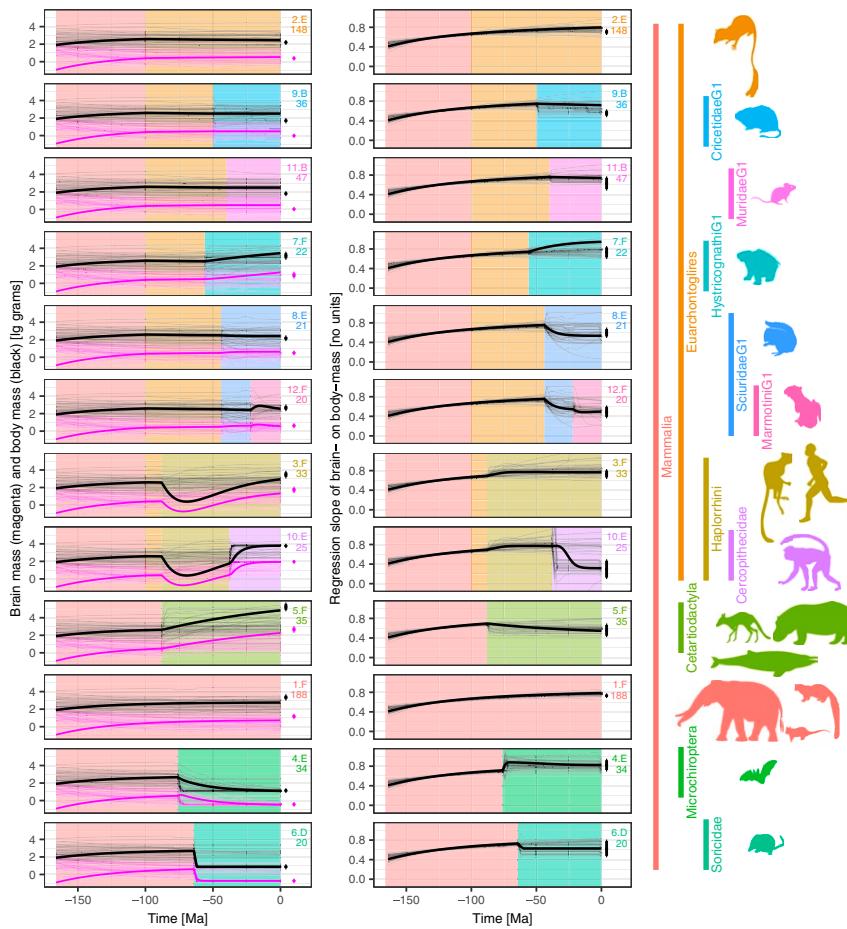


Fig. 2. An MGPM reconstruction of the evolution of body mass and brain mass and their allometric relationship in mammals. (Left and Center) Inferred evolution of body mass and brain mass and brain–body-mass regression slope for each lineage in the mammal tree starting from the root (166.2 Ma ago) and ending at a random tip (extant species) in each of the 12 regimes in MGPM*. The allometry between brain mass and body mass is quantified as the deviation from 1 of the regression slope—increasing regression slope corresponds to decreasing allometry. The thicker lines represent the expected evolution for the mean trait value and the regression slope in each of the 12 regimes assuming the hypothesis that the model MGPM* is the true model; each thinner line represents the corresponding expectation from an MGPM fit to 1 of 50 “parametric bootstraps” datasets—these datasets were generated by simulating MGPM* on the mammal tree (Fig. 1A and *SI Appendix*, section H). The background colors correspond to the 12 inferred regimes in the tree according to MGPM* (Fig. 1A). The error bars on white background on the right side of each plot denote the standard estimates with 95% confidence intervals from the extant species in each regime, ignoring the phylogenetic relationship; for each of the 12 regimes, the selected model type in MGPM* and the number of extant species are written in the top right corner of each plot. (Right) Silhouette images courtesy of Phylopic/T. Michael Keesey, Joseph Wolf, Natasha Vitek, Daniel Jaron, Catherine Yasuda, Allis Markham, Gareth Monger, Jan A. Venter, Herbert H. T. Prins, David A. Balfour, Rob Slotow, C. De Muizon, Scott Hartman, Michael Scroggie, Yan Wong, and Becky Barnes (see also *SI Appendix*, section G for full credit details).

matrix, and the regression slope (*SI Appendix*, sections D, E, and H).

Method Validation. We conducted an extensive simulation study analyzing the MGPM on 1,152 simulated datasets. These simulations confirmed that the MGPM inference correctly identifies clusters in the tree associated with different evolutionary regimes and accurately discriminates between OU and BM regimes. A comparison against previously published models with shifts (18, 23) revealed a crucial advantage for the MGPM with respect to 9 performance criteria (*SI Appendix*, section I and Figs. S11–S61). Following the general requirements for phylogenetic comparative methods (PCMs) (31), we estimated the type I error rate against single-regime BM simulations to ~15% (*SI Appendix*, section J and Fig. S62). Finally, we evaluated the invariance of the MGPM inference to rigid transformations of the data, with “invariance” meaning that the optimal MGPM fits before and after the transformation have equal scores, as well as matching shift-point configurations and model type assignments (31). Our analysis revealed that the invariance does not hold in general for

MGPMs over \mathcal{G}_{LInv} , due to including candidate model types that restrict the between-trait variance–covariance matrix (e.g., BMA and OU_C ; *Materials and Methods* and *SI Appendix*, sections H.5 and K, Fig. S10, and Tables S11–S21).

Discussion

The Mammal Data Have a Strong Statistical Support for an MGPM (A–F) Model. Based on a significant AIC difference ($\Delta AIC > 73$), we confirm that the mammal data have a strong statistical support for a complex MGPM with, predominantly, OUE and OU_F regimes, relative to models assuming trait independence, single-regime models, and simpler BM or scalar OU models with shifts (Table 1 and *SI Appendix*, section H). These results undermine the use of methods assuming global correlation or selection patterns. In fact, a phylogenetic model neglecting the possibility for differing correlation and selection patterns between different taxons can be just as misleading as any standard statistical method that is completely ignorant of the phylogeny. For example, consider the original work providing the mammal data for this study (28). Boddy *et al.* (28) reported an estimate

of the encephalization quotient (EQ) in *Homo sapiens* of 5.72, on the basis of “the standard log brain mass vs. log body mass regression line” (ref. 28, p. 984); compared with 1.16, on the basis of the phylogenetic independent contrasts (PIC); and 12.6, on the basis of the phylogenetic generalized least-squares (PGLS) methods (28). To “retain the same biological context as other encephalization studies” (ref. 28, p. 984) in the main text, Boddy *et al.* (28) reported the standard EQ estimates. However, they used the PIC- and PGLS-derived estimates for all statistical tests (28). The MGPM analysis clarifies this confusion. In particular, it shows that the regression line differs significantly between different taxons (Figs. 1 C and D and 2 and *SI Appendix*, Fig. S2). Therefore, any phylogenetic model assuming a global regression line for all species provides a “consensus” regression line, with weights of the different taxons depending on the assumed stochastic process. Remarkably, if we free the phylogenetic model from the assumption of a single process (of a specific type) covering the entire tree, we observe a general agreement between the standard and the “phylogenetically correct” regression lines in most of the 12 regimes (Fig. 1 C and D and *SI Appendix*, Fig. S2).

The MGPM Enables a Data-Driven Choice of Groups for Analysis. A most interesting example is the clade of Haplorrhini (dry-nose primates) and its subclade of the Cercopithecidae (Old World monkeys) showing significantly different regression slopes (regimes 3 and 10; Figs. 1 C and D and 2). Excluding Cercopithecidae from its parent clade of Haplorrhini reveals parallel regression lines for Haplorrhini and the major mammal group (regimes 3 and 1; Fig. 1 C and D). Hence, these 2 regimes differ solely by the intercept. This confirms the previous observation that Haplorrhini exhibit significantly higher encephalization compared with other primates (28). Boddy *et al.* (28) compared the mean EQ in several sister clades within Haplorrhini (including Anthropoidea vs. Tarsiiformes and Catarrhini vs. Platyrhini) but did not identify any significant difference. In contrast, our analysis revealed a shift at the root of the Cercopithecidae clade. This was present in the 2 models with best AIC, MGPM*, and RATEMATRIX BM and was detected in 36 of the 50 parametric bootstrap datasets (*SI Appendix*, Figs. S5–S7 and S10). Occupying a narrow niche in the phenotype space and exhibiting far more pronounced allometry, Cercopithecidae might be subject to stronger selective pressures relative to other primates. Future studies exploring larger samples of species in this clade should test this hypothesis.

The Ancestral Levels of Brain–Body-Mass Allometry Can Be Inferred with High Confidence. Looking back in time, the inferred model suggests the hypothesis that, with slope = 0.4, the brain–body-mass allometry has been far more pronounced in the mammal ancestors 160 Ma ago (Fig. 2). This slope has increased gradually through time until reaching nowadays levels of ≈ 0.75 for all species in regimes 1, 2, and 3 (Fig. 2). We observe a remarkable bootstrap support for this trend, contrasting with considerably lower support for the estimates of ancestral trait values (compare thin transparent lines in Fig. 2, Left and Center). The poor bootstrap support for the ancestral values of brain mass and body mass manifests a well-known issue of identifiability for OU models (20, 23, 30). For example, ref. 30 showed analytically that in an OU model on an ultrametric tree, it is not possible to infer both the root value \vec{X}_0 and the long-term optimum $\vec{\theta}$. Conversely, the apparent strong signal for the ancestral allometry has, in our view, not been appreciated and represents an appealing subject for future theoretical and empirical studies.

Related Previous Approaches. The idea of jointly fitting different types of Gaussian models dates back at least to the work

of Slater (33), where he measured the statistical support for a shift from an OU to a BM process in the evolution of mammal body size occurring at the end of the Mesozoic (but see ref. 34). Later, Clavel *et al.* (35) implemented a nonpruning algorithm for multivariate likelihood calculation for shifts between BM, OU, and the early burst (EB) model of adaptive radiation in their R-package mvMorph. These works assume a known point in time where a global shift occurs on all lineages of the tree. Moreover, in its current version mvMorph is restricted to trees of moderate size, because it uses a slow likelihood calculation algorithm. Many authors have proposed methods for finding local intramodel shifts in some of the parameters of the OU model and under various simplifying assumptions including tree ultrametricity (i.e., all species have been sampled at the present time), a single trait or independently evolving multiple traits, and shared or fixed parameter values between model regimes (e.g., a scalar OU model with a global [scalar diagonal] selection strength matrix and drift matrix for all regimes) (2, 15–21, 23, 36). In *SI Appendix*, section I, we discuss several of these tools and implement a simulation-based comparison of the MGPM method to existing implementations of phylogenetic comparative models with shifts.

The ambitious task of finding “local” intermodel shifts occurring on individual branches has, to our knowledge, not been addressed. Our main goal here is to propose a solution for this lack of generality in the existing methods and tools. The MGPM provides a unified computationally efficient and extensible framework for a large family of models and for any type of tree.

Materials and Methods

The OU Process. The k -variate OU process is defined by the stochastic differential equation

$$d\vec{X}(t) = \mathbf{H}(\vec{\theta} - \vec{X}(t))dt + \Sigma_u dW(t), \quad [1]$$

where $\vec{X}(t)$ is a k -dimensional real vector, \mathbf{H} is a $k \times k$ -dimensional eigen-decomposable real matrix, $\vec{\theta}$ is a k -dimensional real vector, Σ_u is a $k \times k$ -dimensional real positive definite matrix, and $W(t)$ denotes the k -dimensional standard Wiener process. The branching process, where each branching event gives rise to 2 independent instances of the process (Eq. 1), starting from the value of \vec{X} at the branching point, is a \mathcal{G}_{Linv} process (*SI Appendix*, section C).

Biologically, $\vec{X}(t)$ denotes the mean values of k traits in a species at a time t from the root, the parameter $\Sigma = \Sigma_u \Sigma_u^T$ defines the magnitude and shape of the momentary fluctuations in the mean vector due to genetic drift, and the matrix \mathbf{H} and the vector $\vec{\theta}$ specify the trajectory of the population mean through time. When \mathbf{H} is the 0 matrix, the process is equivalent to BM and the parameter $\vec{\theta}$ is irrelevant. When \mathbf{H} has strictly positive eigenvalues, the population mean converges in the long term to $\vec{\theta}$, although the trajectory of this convergence can be complex. In all candidate model types, we restrict \mathbf{H} to have nonnegative eigenvalues—a negative eigenvalue of \mathbf{H} transforms the process into repulsion with respect to $\vec{\theta}$, which, while biologically plausible, is not identifiable in an ultrametric tree.

MGPM (A–F). The 6 candidate model types BM_A, \dots, OU_F were defined as specifications of the OU process as follows:

- BM_A ($\mathbf{H} = 0$, diagonal Σ): BM, uncorrelated traits.
- BM_B ($\mathbf{H} = 0$, symmetric Σ): BM, correlated traits.
- OU_C (diagonal \mathbf{H} , diagonal Σ): OU, uncorrelated traits.
- OU_D (diagonal \mathbf{H} , symmetric Σ): OU, correlated traits, but simple (diagonal) selection strength matrix.
- OU_E (symmetric \mathbf{H} , symmetric Σ): An OU with nondiagonal symmetric \mathbf{H} and nondiagonal symmetric Σ .
- OU_F (asymmetric \mathbf{H} , symmetric Σ): An OU with nondiagonal asymmetric \mathbf{H} and nondiagonal symmetric Σ .

Other Models. For comparison, we implemented 3 previously published models with shifts (all of which belong to \mathcal{G}_{Linv}):

- SURFACE OU (18): This model assumes traits following univariate OU processes with shared shift points for the long-term optima. Formally, it is equivalent to a k -variate OU process with global diagonal \mathbf{H} and Σ and regime-specific $\vec{\theta}$.

- SCALAR OU (23): This is an OU model with shifts in both $\vec{\theta}$ and Σ . While this model accounts for coevolution of the traits (symmetric Σ), its main restriction is the assumption that the matrix H is scalar diagonal and is shared by all regimes (23).
- RATEMATRIX BM (22): This is equivalent to a BM_β model with shifts.

Implementation. The likelihood calculation was implemented in the R-package PCMBBase (<https://github.com/venelin/PCMBBase>), using internal calls to its Rcpp companion PCMBBaseCpp (<https://github.com/venelin/PCMBBaseCpp>) (24) and the SPLITT C++ library (<https://github.com/venelin/SPLITT>) (37). The RCP algorithm was implemented in the R-package PCMFit (<https://github.com/venelin/PCMFit>) (38). Further details on the implementation, as well as the used third-party libraries and resources, are

- provided in *SI Appendix, sections A–G*. The analysis of the mammal data has been implemented in the R-package MGPMAMmals (<https://github.com/venelin/MGPMAMmals>) (39) (see also *SI Appendix, section H*). The simulation tests have been implemented in the R-package MGPMSSimulations (<https://github.com/venelin/MGPMSSimulations>) (40) (see also *SI Appendix, sections I and J*).
- ACKNOWLEDGMENTS.** V.M. and T.S. thank ETH Zürich for funding. K.B.'s research is supported by Vetenskapsrådets Grant 2017-04951. We thank Prof. Dr. Jörg Stelling for providing the analyzed mammal data including the taxonomic labels for the internal nodes of the tree and for valuable suggestions. We thank Dr. Joëlle Barido-Sottani and 4 anonymous reviewers for valuable suggestions.
1. M. J. Benton, B. C. Emerson, How did life become so diverse? The dynamics of diversification according to the fossil record and molecular phylogenetics. *Palaeontology* **50**, 23–40 (2007).
 2. M. A. Butler, A. A. King, Phylogenetic comparative analysis: A modeling approach for adaptive evolution. *Am. Nat.* **164**, 683–695 (2004).
 3. M. W. Pennell, L. J. Harmon, An integrative view of phylogenetic comparative methods: Connections to population genetics, community ecology, and paleobiology. *Ann. New York Acad. Sci.* **1289**, 90–105 (2013).
 4. J. Felsenstein, Phylogenies and the comparative method. *Am. Nat.* **125**, 1–15 (1985).
 5. E. P. Martins, T. F. Hansen, Phylogenies and the comparative method: A general approach to incorporating phylogenetic information into the analysis of interspecific data. *Am. Nat.* **149**, 646–667 (1997).
 6. E. A. Housworth, E. P. Martins, M. Lynch, The phylogenetic mixed model. *Am. Nat.* **163**, 84–96 (2004).
 7. S. Alizon et al., Phylogenetic approach reveals that virus genotype largely determines HIV set-point viral load. *PLoS Pathog.* **6**, e1001123 (2010).
 8. G. Shirreff et al., How effectively can HIV phylogenies be used to measure heritability? *Evol. Med. Public Health* **2013**, 209–224 (2013).
 9. E. Hodcroft et al., The contribution of viral genotype to plasma viral set-point in HIV infection. *PLoS Pathog.* **10**, e1004112 (2014).
 10. F. Blanquart et al., Viral genetic variation accounts for a third of variability in HIV-1 set-point viral load in Europe. *PLoS Biol.* **15**, e2001855 (2017).
 11. F. Bertels et al., Dissecting HIV virulence: Heritability of setpoint viral load, CD4+ T cell decline and per-parasite pathogenicity. *Mol. Biol. Evol.* **35**, 27–37 (2017).
 12. V. Mitov, T. Stadler, A practical guide to estimating the heritability of pathogen traits. *Mol. Biol. Evol.* **35**, 756–772 (2018).
 13. T. F. Hansen, E. P. Martins, Translating between microevolutionary process and macroevolutionary patterns: The correlation structure of interspecific data. *Evolution* **50**, 1404 (1996).
 14. N. Cooper, G. H. Thomas, C. Venditti, A. Meade, R. P. Freckleton, A cautionary note on the use of Ornstein-Uhlenbeck models in macroevolutionary studies. *Biol. J. Linn. Soc.* **118**, 64–77 (2015).
 15. B. C. O'Meara, C. Ané, M. J. Sanderson, P. C. Wainwright, Testing for different rates of continuous trait evolution using likelihood. *Evolution* **60**, 922–933 (2006).
 16. J. M. Eastman, M. E. Alfaro, P. Joyce, A. L. Hipp, L. J. Harmon, A novel comparative method for identifying shifts in the rate of character evolution on trees. *Evolution* **65**, 3578–3589 (2011).
 17. J. M. Beaulieu, D. C. Jhuweng, C. Boettiger, B. C. O'Meara, Modeling stabilizing selection: Expanding the Ornstein-Uhlenbeck model of adaptive evolution. *Evolution* **66**, 2369–2383 (2012).
 18. T. Ingram, D. L. Mahler, SURFACE: Detecting convergent evolution from comparative data by fitting Ornstein-Uhlenbeck models with stepwise Akaike information criterion. *Methods Ecol. Evol.* **4**, 416–425 (2013).
 19. J. C. Uyeda, L. J. Harmon, A novel Bayesian method for inferring and interpreting the dynamics of adaptive landscapes from phylogenetic comparative data. *Syst. Biol.* **63**, 902–918 (2014).
 20. M. Khabbazian, R. Kriebel, K. Rohe, C. Ané, Fast and accurate detection of evolutionary shifts in Ornstein-Uhlenbeck models. *Methods Ecol. Evol.* **7**, 811–824 (2016).
 21. P. Bastide, M. Mariadassou, S. Robin, Detection of adaptive shifts on phylogenies by using shifted stochastic processes on a tree. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **79**, 1067–1093 (2017).
 22. D. S. Caetano, L. J. Harmon, ratematrix: An R package for studying evolutionary integration among several traits on phylogenetic trees. *Methods Ecol. Evol.* **8**, 1920–1927 (2017).
 23. P. Bastide, C. Ané, S. Robin, M. Mariadassou, Inference of adaptive shifts for multivariate correlated traits. *Syst. Biol.* **113**, 2158–680 (2018).
 24. V. Mitov, K. Bartoszek, G. Asimomitis, T. Stadler, Fast likelihood calculation for multivariate phylogenetic comparative methods: The PCMBBase R package. arXiv:1809.09014 (24 September 2018).
 25. O. Snell, "Das Gewicht des Gehirnes und des Hirnmantels der Säugerthiere in Beziehung zu deren geistigen Fähigkeiten" in *Sitzungsberichte der Gesellschaft für Morphologie und Psychologie in München* (Society for Morphology and Physiology, 1891), vol. 7, pp. 90–94.
 26. H. Jerison, *Evolution of The Brain and Intelligence* (Academic Press, Inc., New York, NY, 1973).
 27. S. H. Montgomery, I. Capellini, R. A. Barton, N. I. Mundy, Reconstructing the ups and downs of primate brain evolution: Implications for adaptive hypotheses and Homo floresiensis. *BMC Biol.* **8**, 9 (2010).
 28. A. M. Boddy et al., Comparative analysis of encephalization in mammals reveals relaxed constraints on anthropoid primate and cetacean brain scaling. *J. Evol. Biol.* **25**, 981–994 (2012).
 29. O. R. P. Bininda-Emonds et al., The delayed rise of present-day mammals. *Nature* **446**, 507–512 (2007).
 30. L. S. T. Ho, C. Ané, Intrinsic inference difficulties for trait evolution with Ornstein-Uhlenbeck models. *Methods Ecol. Evol.* **5**, 1133–1146 (2014).
 31. D. C. Adams, M. L. Collyer, Multivariate phylogenetic comparative methods: Evaluations, comparisons, and recommendations. *Syst. Biol.* **67**, 14–31 (2018).
 32. P. Zwierink, C. Uhler, D. Richards, Maximum likelihood estimation for linear Gaussian covariance models. arXiv:1408.5604 (24 August 2014).
 33. G. J. Slater, Phylogenetic evidence for a shift in the mode of mammalian body size evolution at the Cretaceous-Palaeogene boundary. *Methods Ecol. Evol.* **4**, 734–744 (2013).
 34. G. J. Slater, Correction to 'Phylogenetic evidence for a shift in the mode of mammalian body size evolution at the Cretaceous-Palaeogene boundary', and a note on fitting macroevolutionary models to comparative paleontological data sets. *Methods Ecol. Evol.* **5**, 714–718 (2014).
 35. J. Clavel, G. Escarguel, G. Merceron, mvMorph: An R package for fitting multivariate evolutionary models to morphometric data. *Methods Ecol. Evol.* **6**, 1311–1319 (2015).
 36. J. C. Uyeda, M. W. Pennell, E. T. Miller, R. Maia, C. R. McClain, The evolution of energetic scaling across the vertebrate tree of life. *Am. Nat.* **190**, 185–199 (2017).
 37. V. Mitov, T. Stadler, "Parallel likelihood calculation for phylogenetic comparative models: The SPLITTC++ library" in *Methods in Ecology and Evolution*, T. Münkemüller, Ed. (John Wiley & Sons Ltd., 2018), pp. 2041–210X.13136.
 38. V. Mitov, PCMFit: An R-package for statistical inference of phylogenetic comparative models. Version v1.0.0. Zenodo. <https://venelin.github.io/PCMFit/>. Deposited 18 July 2019.
 39. V. Mitov, MGPMAMmals: Data and R-code for the analysis of the mammal dataset. Version v1.0.0. Zenodo. <https://venelin.github.io/MGPMAMmals/>. Deposited 18 July 2019.
 40. V. Mitov, MGPMSSimulations: Data and R-code for the simulation study. Version v1.0.0. Zenodo. <https://venelin.github.io/MGPMSSimulations/>. Deposited 18 July 2019.