COMMENT

A critical comment on the 'multiple variance Brownian motion' model of Smaers *et al.* (2016)

RANDI H. GRIFFIN* and GABRIEL S. YAPUNCICH

Department of Evolutionary Anthropology, Duke University, Durham, NC 27708, USA

Smaers, Mongle & Kandler (2016) (*Biological Journal of the Linnean Society*, 118: 74–98) introduced a new phylogenetic comparative method, multiple-variance Brownian motion (mvBM), for reconstructing ancestral states given a phylogenetic tree and continuous trait data. The authors conducted a simulation study and argued that mvBM outperforms constant variance Brownian motion (BM) when rates of evolution vary across the phylogeny. In this comment, we argue that mvBM is fundamentally a circular analysis that overfits phylogenetic branch lengths to the data. We further argue that the comparison of mvBM to BM under conditions where the assumptions of BM are clearly violated is not an informative performance analysis and that the simulation study of Smaers *et al.* (2016) exaggerates the performance of mvBM by focusing on a narrow range of simulation conditions and reporting aggregated metrics of relative accuracy that obscure inaccuracy and bias in its ancestral state estimates. Our arguments are supported by simulation results. We conclude that mvBM is not a viable phylogenetic comparative method.

ADDITIONAL KEYWORDS: ancestral state reconstruction – Brownian motion – evolutionary biology – simulation study.

INTRODUCTION

Smaers, Mongle & Kandler (2016) introduced a new phylogenetic comparative method (PCM), which they call multiple-variance Brownian motion (mvBM), designed to estimate ancestral states from a phylogeny and continuous trait data. The authors validated their new method with simulations showing that mvBM produces similar or improved ancestral state reconstructions (ASRs) compared to constant variance Brownian motion (BM). The mvBM method is partially based on the independent evolution (IE) method introduced by Smaers & Vinicius (2009). We recently demonstrated several theoretical and statistical problems with IE (Griffin & Yapuncich, 2015), leading us to examine mvBM closely.

In this comment, we argue that mvBM is a nonviable PCM and that the simulation study included in Smaers *et al.* is inadequate. Specifically, we argue that: (1) mvBM is a circular analysis that overfits phylogenetic branch lengths to the data; (2) mvBM has

improved accuracy compared to BM under the burst scenario due to similarities between Smaers *et al.*'s method and phylogenetically independent contrasts (PIC; Felsenstein, 1985); and (3) the aggregated accuracy metrics presented by Smaers *et al.* obscure bias and inaccuracy in mvBM's ASRs. To support our arguments, we performed simulations in R (R Core Team, 2015) with the packages 'ape' (Paradis, Claude & Strimmer, 2004), 'phytools' (Revell, 2012), and 'geiger' (Harmon *et al.*, 2008). Supplementary Information S1 contains a detailed description of our implementation of mvBM. Code for replicating our simulations is available on GitHub (Griffin, 2016).

MVBM IS A CIRCULAR ANALYSIS THAT OVERFITS PHYLOGENETIC BRANCH LENGTHS TO THE DATA

For a phylogeny with n tips, the mvBM method estimates ancestral states in four steps: (1) 'the phenotypic trait values of the $n-1$ internal nodes' (p. 80) are

*Corresponding author. E-mail: randi.griffin@duke.edu

estimated using a phylogeny and trait data for terminal taxa; (2) these values are used to estimate squared change along each branch; (3) each branch is then elongated by an amount proportional to its inferred squared change; and (4) 'the trait values for all internal nodes' (p. 82) are estimated by fitting a standard BM model to the mvBM-transformed phylogeny. Smaers *et al.* say 'the [equations] used for estimating the nodal values can be modified' (p. 81) and 'alternative branch lengthening procedures can be conceived' (p. 82); thus, the specific equations used in Steps 1 and 3 are not essential features of mvBM. Rather, the key feature distinguishing mvBM from other ASR methods is the repeated estimation of internal node values in Steps 1 and 4. In their description, Smaers *et al.* distinguish between the first and final sets of estimated 'phenotypic values at internal nodes' by suggesting that the first set is used to 'parameterize a mvBM model' (p. 80), while the second set represents ASRs 'under a multiple variance BM model of phenotypic evolution' (p. 82). However, it is difficult to interpret the first set of 'phenotypic values at internal nodes' as anything other than ASRs: the first set of nodal values are used to infer change along individual branches as the difference between ancestral and descendent nodal values, which only makes sense if the nodal values represent ASRs for the trait in question. Because the starting and ending points of the algorithm are the same (i.e., estimated trait values at internal nodes), mvBM is fundamentally a circular analysis.

This circularity explains why the likelihood of data under a BM model improves substantially when the original phylogeny is replaced with the mvBM-transformed phylogeny in the final step of the method (Smaers *et al.*, 2016, their Fig. 3). The only way a model can fit data better than the generating model is by *overfitting*, that is, modelling random noise as a meaningful signal (McElreath, 2016). The mvBM method systematically overfits phylogenetic branch lengths to the data by elongating branches such that larger changes are made more likely on branches where larger changes have already been inferred from the initial ASRs. Importantly, mvBM includes no mechanisms to avoid this systematic overfitting (e.g., penalized likelihoods, Bayesian priors, or caps on the number of parameters; reviewed in Ho & Ané, 2014). This presents a major problem for statistical inference: even when evolution follows a single-rate BM process, mvBM *always* estimates $2n-2$ rates and produces a transformed phylogeny that fits the data better than the original phylogeny. However, since this transformed phylogeny largely captures random noise, it does not provide a meaningful insight about the data.

Before discussing Smaers *et al.*'s simulation study, we must point out that mvBM represents a major departure from basic principles underlying modern PCMs. The statistical terminology of Smaers *et al.*

suggests their method relates to existing statistical models of evolution, particularly the multiple-rate BM model of Venditti, Meade & Pagel (2011). However, a true statistical model of evolution can be represented by a likelihood function describing the mathematical relationship among parameters (*sensu* Venditti *et al.*, 2011), permitting parameter optimization and formal model comparison that accounts for the number of estimated parameters (e.g., using Akaike information criterion). The mvBM method includes no likelihood function; therefore, it cannot be used for parameter optimization or formal model comparison, nor can it be accurately described as a 'multiple-rate BM model' (which implies a parametric model of evolution that extends the BM model). Although Smaers *et al.* report likelihoods for mvBM, these likelihoods do not account for the $2n-2$ branch-specific rates estimated in Steps 1–3. Instead, these likelihoods correspond to a BM model with two parameters (mean and variance) fit to the original data, with the overfit mvBM-transformed phylogeny incorporated into the covariance structure. As such, the resulting likelihoods cannot be used for statistical inference.

HOW DOES mvBM PRODUCE MORE ACCURATE ASRs THAN BM UNDER A 'BURST' SCENARIO?

Smaers *et al.* (2016) simulated data under a 'burst' scenario, where the rate of BM evolution is increased by two to four orders of magnitude on one or five branches of the phylogeny. They demonstrated that mvBM produces more accurate ASRs relative to BM when data are generated under the burst scenario. In this section, we show that in practice, mvBM produces ASRs that are roughly intermediate between BM (i.e., global maximum likelihood) and PIC ASRs (i.e., local maximum likelihood; Maddison, 1991). We argue that the improved accuracy of mvBM relative to BM under the burst scenario is a result of the similarity between PIC ASRs and the 'nodal values' computed in the first part of the mvBM algorithm.

In Step 1 of mvBM, Smaers *et al.* traverse the tree from the tips to the root in the same fashion as the PIC algorithm introduced by Felsenstein (1985). At each node, the ASR is computed as a weighted average of a 'global' and a 'local' estimate for the trait value at that node. The 'global' estimate is a weighted mean of all the tips of the phylogeny, with the tips weighted by their inverse-squared patristic distance to the node in question (Smaers *et al.*, 2016: Eq. 2). The 'local' estimate is the average value of the immediate descendants of the node in question, without accounting for phylogenetic distance (Smaers *et al.*, 2016: Eq. 3). Finally, the 'global' and 'local' estimates are averaged (Smaers *et al.*, 2016: Eqns 4 and 5), with the 'local' estimate being given twice the weight of the 'global' estimate. Smaers *et al.* indicate that different weighting

schemes are allowed; thus, the relative weights of the 'global' and 'local' estimates are arbitrary.

We do not believe that this approach is justified based on any valid theoretical model of evolution. The 'global' ASR is nearly identical to the problematic 'adaptive peak' equation of Smaers & Vinicius (2009), although branch lengths are now squared. We have previously highlighted theoretical issues when this equation was used to estimate an 'adaptive peak' (Griffin & Yapuncich, 2015), and Smaers *et al.* (2016) do not provide a convincing justification for why this particular value should now be considered a 'global' estimate of an ancestral state. We also note that the 'local' ASR is nonphylogenetic since it ignores branch lengths. While these criticisms speak to the theoretical weakness of the approach, we instead focus on the fact that it produces similar results to the ASRs computed with Felsenstein's (1985) PIC algorithm. Maddison (1991) showed that PIC ASRs are identical to the local maximum likelihood ASR at each node, whereas BM ASRs correspond to the global maximum likelihood ASRs. Since Smaers *et al.* square the inverse patristic distances in their 'global' estimate, and the 'local' estimate is weighted more heavily, the influence of phylogenetically distant tips is reduced, bringing the initial mvBM ASRs closer to PIC ASRs. These initial ASRs are then used to infer branch-specific rates of change, elongate branches according to the inferred change on the branch, and estimate BM ASRs on the transformed tree. This essentially weights branches such that the initial ASRs become more likely under a BM model. Consequentially, when ASRs are estimated for a second time using a BM model with the mvBM transformed tree, the results are biased towards the initial ASRs, which are similar to PIC ASRs.

The similarity between PIC ASRs and the ASRs computed in Step 1 of mvBM explains why mvBM produces more accurate ASRs than BM under a 'burst' scenario. The ASRs of BM are truly 'global' in the sense that they represent the set of maximum likelihood ASRs under a BM model. As such, when a large burst occurs on a single branch, BM spreads the burst across multiple branches and affects ASRs for all internal nodes. In contrast, PIC produces 'local' ASRs that represent the maximum likelihood ASRs when only the focal node's immediate descendants are considered. Consequently, a burst of evolution will have a more local impact on PIC ASRs.

To illustrate this point, we performed a simple set of simulations. We generated a phylogeny with three tips, such that there is one internal node besides the root (Fig. 1A). We simulated the evolution of 100 continuous traits with mean (root state) 0 and variance 0.01 along the branches leading to the pair of sister taxa (Branches 1–3). On the long branch (Branch 4), we uniformly
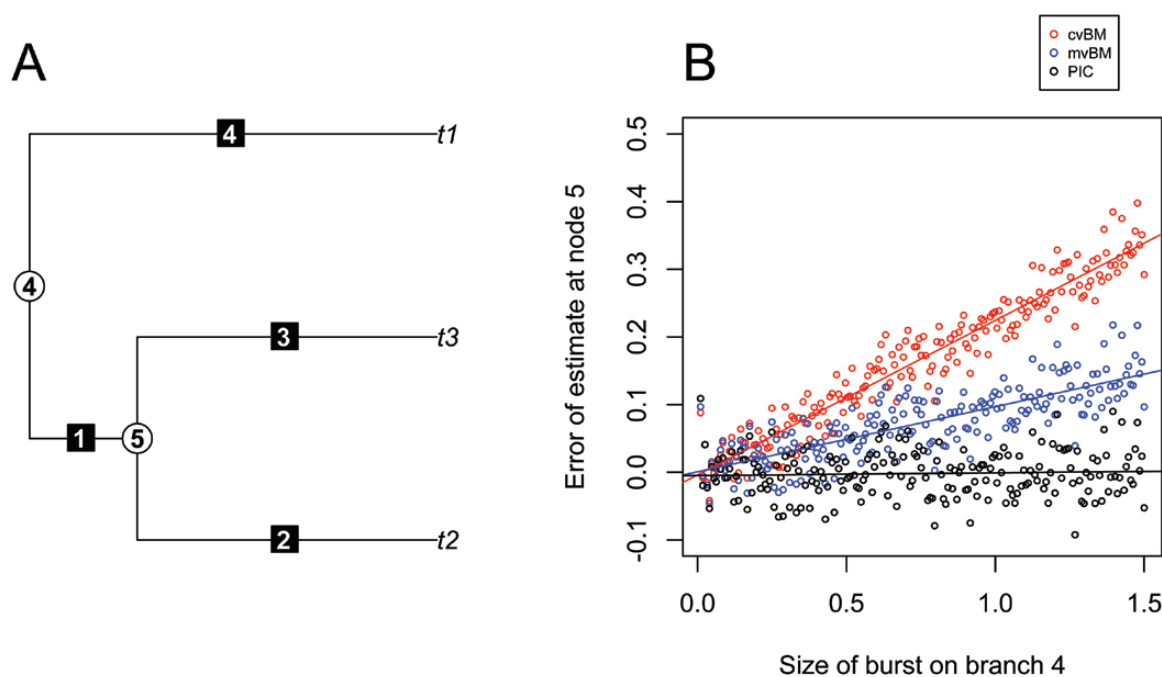

Figure 1: (A) The phylogeny used for simulations demonstrating error and bias in mvBM ancestral state reconstructions. A Brownian motion model is used to simulate 100 traits with variance 0.01 on Branches 1–3, while the trait change on Branch 4 is varied from 0 to 1.5. For each simulation, the ancestral state at Node 5 is estimated using PIC, mvBM, and BM. (B) Relationship between different magnitudes of trait change on Branch 4 and overestimation of the ancestral state at Node 5 based on PIC, mvBM, and BM.

incremented trait change from 0 (representing no change on the long branch) to 1.5 (representing a large burst of evolution on the long branch). We then used BM, mvBM, and PIC to reconstruct the ancestral state at the internal node (Node 5) for each simulated data set and compared the results. Figure 1B shows that BM and mvBM become increasingly biased towards positive values as the evolutionary change along Branch 4 increases from 0 to 1.5, while the accuracy of PIC is unaffected. Since PIC computes local estimates of the ancestral states, the estimate at Node 5 is unaffected by the burst on Branch 4. In contrast, BM spreads the burst across branches such that large positive changes on Branch 4 lead the ASR at Node 5 to be overestimated. Since mvBM is neither entirely local nor entirely global, it produces ASRs that are intermediate between BM and PIC.

Of course, this does not mean that BM is a bad model or that PIC is a better model. Rather, BM is appropriate when the assumption of a constant rate of evolution is reasonable and not when evolutionary rates are increased by orders of magnitude on individual branches. In light of this, we do not believe that Smaers *et al*.'s demonstration that mvBM performs relatively better than BM under simulation conditions where BM is inappropriate constitutes a rigorous performance analysis. Rather, a rigorous performance analysis should demonstrate that a method accurately estimates the parameters it is intended to estimate in an absolute (rather than relative) sense. Below, we present such a performance analysis by replicating the simulation conditions from Smaers *et al*. (2016) and reporting the absolute bias and error in mvBM's ASRs. We also explain why absolute bias and error were not apparent in the simulation results reported by Smaers *et al*. (2016).

BIAS AND ERROR IN MVBM ASRS

Using a 100-taxa phylogeny, Smaers *et al*. simulated 1000 data sets with (1) a burst on one branch, (2) five bursts spread across the tree, and (3) five bursts clustered together. The rate of evolution was increased on 'burst' branches by either 100, 2500, or 10 000 times the baseline rate of evolution. For each simulated data set, they estimated the R^2 of a linear regression of mvBM-estimated ancestral states against true ancestral states. They found that the vast majority of R^2 values were > 0.95 (Figures 4–6 in Smaers *et al*., 2016), suggesting a close fit between mvBM estimates and true ancestral states. They concluded that mvBM produces accurate ASRs even when there are large bursts of evolution on individual branches of the tree. However, Smaers *et al*. did not report ASRs for specific nodes near the bursts, nor did they vary the location of the bursts in the phylogeny. Here, we simulate the single-burst scenario to show that (1) error in ASRs at the base of the 'burst branch' is large and biased; (2) by simulating a large number

of tips relative to bursts and aggregating results across all nodes, Smaers *et al*.'s approach obscures mvBM's inaccuracy near the site of the burst; and (3) the accuracy of mvBM depends on the location of the burst in the phylogeny, but Smaers *et al*.'s simulations focus on branches that tend to produce relatively accurate ASRs.

To examine the accuracy of ASRs near the site of the burst, we replicated the 'burst' simulation conditions used in Smaers *et al*. but varied the location of the burst branch. We generated a pure-birth tree with 100 tips and allowed each branch to be the 'burst branch' for five simulations of a continuous trait, yielding $(2n-2)\times5 = 990$ simulated traits. The baseline rate of evolution was 0.01, while the burst branch had a rate of 1, replicating the smallest burst size considered in Smaers *et al*. Figure 2 shows the relationship between trait change on the burst branch and error in the ASR at the base of the branch. Reconstructions at the base of the burst branch are strongly biased in the direction of the burst, and the degree of bias depends on the size of the burst. In these simulations, the median distance from the largest to the smallest simulated value is only 0.92, so the errors shown in Fig. 2 (ranging from −0.49 to 0.57) are quite large relative to the variation in the data. Because Smaers *et al*.'s accuracy measures are aggregated across the entire tree, their results do not reveal the strength of bias near the site of the burst.

Next, we demonstrate that the R^2 values of estimated versus true ancestral states are much lower when the number of taxa in the tree decreases. We simulated a pure-birth phylogeny with $n = 30$ tips and allowed each branch to be the burst branch for 100 simulations, yielding $(2n-2)\times100 = 5800$ simulated traits. We
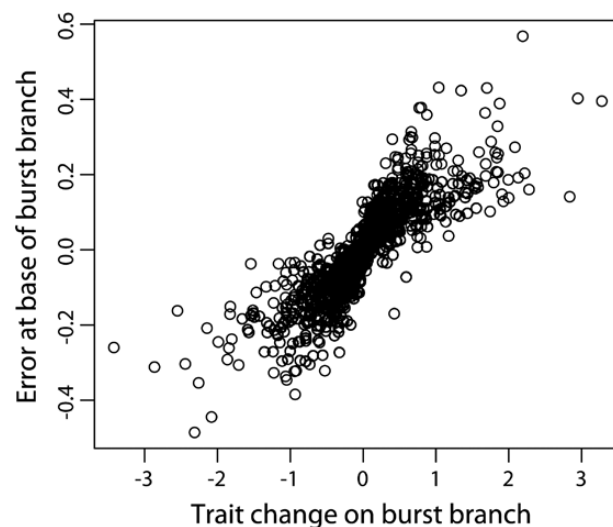

Figure 2. Relationship between error in the estimated ancestral state at the base of the burst branch and the direction and magnitude of trait change on the burst branch.
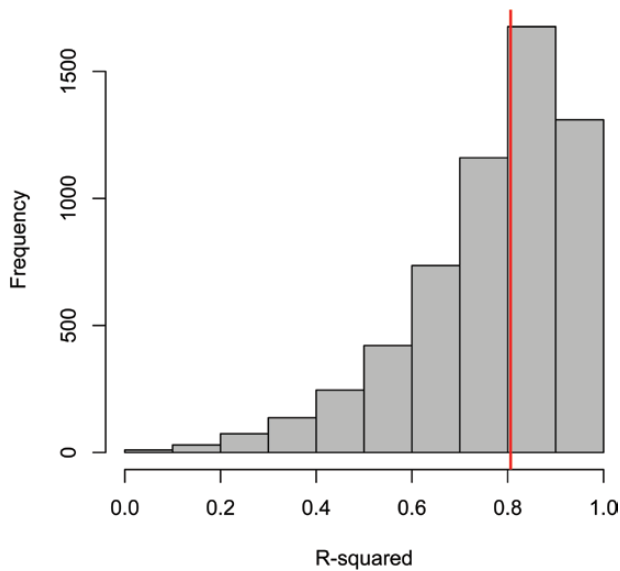
Figure 3. Distribution of R^2 values of estimated versus true ancestral states, given a 30-taxa tree and 5800 simulations in which each of the 58 branches are allowed to be the burst branch for 100 Brownian motion simulations of a continuous trait. The baseline variance of the Brownian motion process is 0.01, while the burst branch has a variance of 1. The solid red line indicates the median R^2.

followed Smaers *et al.* in regressing estimated versus true ancestral states for each simulation to assess accuracy. Figure 3 shows that with the smaller tree, R^2 values are much lower and more variable than the results presented by Smaers *et al.* Rather than the R^2 consistently being > 0.95, it ranges from nearly 0 to 1, and the median is 0.81 (Fig. 3). Smaers *et al.* report higher R^2 values because including more taxa dilutes the effect of the evolutionary burst, resulting in a stronger correlation between estimated and true ancestral states. Thus, the inaccuracy of mvBM ASRs near the site of the burst was obscured by Smaers *et al.*'s use of aggregated accuracy measures across a very large phylogeny relative to the number of bursts.

Additionally, Fig. 4 depicts the median R^2 according to which branch was the burst branch and illustrates that bursts near the tips yield lower median R^2 values relative to bursts deeper in the tree. This is unsurprising since a burst of evolution will have the strongest impact on nodes ancestral to the burst, such that when the burst occurs near the tips, more ancestral nodes will be affected than when the burst occurs near the root. Because Smaers *et al.* only simulated bursts relatively deep in the tree (their Fig. 2), their simulation conditions were favourable towards mvBM.

Supplementary Material S2 shows that the five-burst scenario also produces substantially lower R^2 values given a smaller tree and random burst locations.
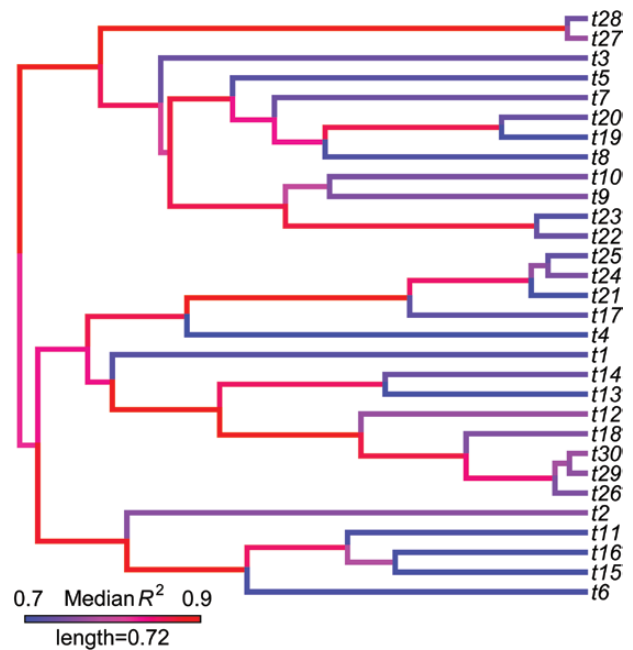

Figure 4. Median R^2 values of estimated versus true ancestral states according to which branch of the phylogenetic tree was the 'burst branch'.

CONCLUSION

We conclude that mvBM has substantial theoretical and statistical problems. Despite its name, mvBM does not incorporate a statistical model of evolution and is not truly a 'multiple-rate BM model'. Our primary concern is that the core of mvBM is a circular procedure that overfits phylogenetic branch lengths to data. We have also demonstrated specific shortcomings in Smaers *et al.*'s simulation study. First, rather than using simulations to test the absolute accuracy of mvBM ASRs, Smaers *et al.* evaluated mvBM relative to BM (a poorly suited method given the simulation conditions) and reported aggregated measures of accuracy rather than accuracy at specific nodes. By reporting results for specific nodes, we demonstrated that mvBM produces strongly biased ASRs near the site of a burst under the same simulation conditions used by Smaers *et al.* Second, outside of the narrow simulation conditions in Smaers *et al.* (i.e., a large number of taxa relative to bursts, a burst branch deep in the tree), mvBM's accuracy declines substantially even according to aggregated measures of accuracy. We conclude that mvBM is not a viable PCM.

ACKNOWLEDGEMENTS

referees for input that greatly improved this manuscript.

REFERENCES

Felsenstein J. 1985. Phylogenies and the comparative method. *American Naturalist* **125:** 1–15.

Griffin RH. 2016. *Evaluating the legitimacy of the mvBM method by Smaers et al., (2016).* Available at: https://github.com/rgriff23/Evaluating_mvBM.

Griffin RH, Yapuncich GS. 2015. The independent evolution method is not a viable phylogenetic comparative method. *PLoS One* **10:** e0144147.

Harmon LJ, Weir JT, Brock CD, Glor RE, Challenger W. 2008. GEIGER: investigating evolutionary radiations. *Bioinformatics (Oxford, England)* **24:** 129–131.

Ho LST, Ané C. 2014. Intrinsic inference difficulties for trait evolution with Ornstein Uhlenbeck models. *Methods in Ecology and Evolution* **5:** 1133–1146.

Maddison WP. 1991. Squared-change parsimony reconstructions of ancestral states for continuous characters on a phylogenetic tree. *Systematic Zoology* **40:** 304–314.

McElreath R. 2016. *Statistical rethinking: a Bayesian course with examples in R and Stan.* Boca Raton, FL: CRC Press, 165–207.

Paradis E, Claude J, Strimmer K. 2004. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics (Oxford, England)* **20:** 289–290.

R Core Team. 2015. *R: a language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing. Available at: URL: http://www.R-project.org/.

Revell LJ. 2012. Phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution* **3:** 217–223.

Smaers JB. 2016. *evomap: R package for the evolutionary mapping of continuous traits: GitHub.* Available at: https://github.com/JeroenSmaers/evomap

Smaers JB, Mongle CS, Kandler A. 2016. A multiple variance Brownian motion framework for estimating variable rates and ancestral states. *Biological Journal of the Linnean Society,* **118:** 74–98, DOI: 10.1111/bij.12765.

Smaers JB, Vinicius L. 2009. Inferring macroevolutionary patterns using an adaptive peak model of evolution. *Evolutionary Ecology Research* **11:** 991–1015.

Venditti C, Meade A, Pagel M. 2011. Multiple routes to mammalian diversity. *Nature* **479:** 393–396.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Supplementary Information S1. Details on our implementation of mvBM
Supplementary Information S2. Five-burst simulations