Version dated: August 7, 2017

RH: ESTIMATING MACROEVOLUTIONARY LANDSCAPES

ONLINE APPENDIX FOR: A GENERAL MODEL FOR ESTIMATING MACROEVOLUTIONARY LANDSCAPES

Florian C. Boucher[1,2], Vincent Démery[3], Elena Conti[1], Luke J. Harmon[4], and Josef Uyeda[4]

[1]*Department of Systematic and Evolutionary Botany (ISEB), University of Zurich, Zurich, Switzerland;*

[2]*Department of Botany and Zoology, University of Stellenbosch, Stellenbosch, South Africa;*

[3]*Gulliver, CNRS, ESPCI Paris, PSL Research University, 10 rue Vauquelin, Paris, France*

[4]*Department of Biological Sciences and Institute for Bioinformatics and Evolutionary Studies (IBEST), University of Idaho, Moscow, Idaho, USA*

Corresponding author: Florian C. Boucher, Department of Botany and Zoology, University of Stellenbosch, Private Bag X1, Matieland 7602, South Africa; E-mail: flofloboucher@gmail.com.

Online Appendix I: Characteristic time of the FPK model

General definition

The Fokker-Planck-Kolmogorov equation (Eq. 1 in main text) for the probability density is linear in $p(x,t)$, and can thus be written as $\frac{\partial p}{\partial t}(x,t) = \mathcal{L}p(x,t)$, where $\mathcal{L}$ is a functional operator. The operator $\mathcal{L}$ is self-adjoint, so it can be diagonalized and it has real eigenvalues. We denote its eigenvalues $\lambda_0 \geq \cdots \geq \lambda_n \geq \ldots$, and $q_k(x)$ the vector corresponding to λ_k. The largest one corresponds to the stationary state, $\lambda_0 = 0$, and all others are strictly negative.

We can write any probability density $p(x,t)$ as a time dependent linear combination of eigenvectors,

$$p(x,t) = \sum_{k=0}^{\infty} a_k(t) q_k(x). \tag{1}$$

The time derivative of this expression involves the time derivative of the coefficient $a_k(t)$:

$$\frac{\partial p}{\partial t}(x,t) = \sum_{k=0}^{\infty} \frac{da_k}{dt}(t) q_k(x). \tag{2}$$

On the other hand, inserting the linear combination in the FPK evolution equation (Eq. 6 in main text) leads to

$$\frac{\partial p}{\partial t}(x,t) = \sum_{k=0}^{\infty} a_k(t) \mathcal{L} q_k(x) = \sum_{k=0}^{\infty} a_k(t) \lambda_k q_k(x). \tag{3}$$

Equating Eqs. (2) and (3), and identifying the coefficients of the eigenvectors, we get

$$\frac{da_k}{dt}(t) = \lambda_k a_k(t), \tag{4}$$

hence $a_k(t) = a_k(0)e^{\lambda_k t}$ and

$$p(x,t) = \sum_{k=0}^{\infty} a_k(0)e^{\lambda_k t} q_k(x). \tag{5}$$

Since the probability density converges to the equilibrium distribution $p^*(x)$, the highest eigenvalue is $\lambda_0 = 0$, and all the others are negative. The eigenvector corresponding to λ_0 is $q_0(x) = p^*(x)$.

All the coefficients $a_k(0)^{\lambda_k t}$ for $k > 0$ decay to zero, and the coefficient with the slowest decay corresponds to $k = 1$. This coefficient defines the characteristic time of the process:

$$T_c = \frac{1}{|\lambda_1|}. \tag{6}$$

Computation with the discretized Fokker-Planck-Kolmogorov equation

As shown in Online Appendix II, the Fokker-Planck-Kolmogorov equation (Eq. 1 in main text) can be discretized, leading to Eq. (7) below. This discretization allows a practical computation of the characteristic time: the eigenvalues of the FPK operator $\mathcal{L}$ are the eigenvalues of the matrix M entering Eq. (7), up to the factor $(n-1)^2/[2\tau]$. It is noteworthy that while the eigenvalues depend in principle on the number of points n on the grid, λ_2 converges as $n \to \infty$.

Characteristic time of the OU process

As seen in the main text of this article, the Ornstein-Uhlenbeck process is a special case of the FPK model with $V(x) = (\alpha/\sigma^2)x^2 - (2\alpha\theta/\sigma^2)x$, with $\alpha > 0$. If we make a translation to center the process around the optimum of the OU process, θ, we can write the potential as $V(x') = (\alpha/\sigma^2)x'^2$, where $x' = x - \theta$. The eigenvalues of the functional operator $\mathcal{L}$ are given by $\lambda_k = -k\alpha$, with $k \in \mathbb{N}$, and the corresponding eigenvectors involve products of the stationary distribution with Hermite polynomials.

Thus, for the OU model, $T_c = 1/|\lambda_1| = 1/\alpha$. This value is proportional to what is usually known as the *phylogenetic half life* of the OU process, which is $\ln(2)/\alpha$.

Online Appendix II: Discretization of the diffusion equation for calculating the likelihood of the FPK model

To compute the likelihood of FPK we discretize the trait interval by considering only a *grid* of n points equally spaced between two extreme values, $B_{\min}$ and $B_{\max}$. The continuous evolution equation for the probability density (Eq. 1 in the main text of this article) can then be cast in a matrix form:

$$\frac{dP}{dt}(t) = \frac{(n-1)^2}{2\tau} M P(t), \tag{7}$$

where $\tau = (B_{\max} - B_{\min})^2/\sigma^2$, $P(t)$ is a matrix whose components $P_{ij}(t)$ describe the probability of the trait being in position i on the grid at time t starting from position j at time 0, and M is a transition matrix specifying the probability for the trait to evolve from one of these n values on the grid to another during an infinitesimal time step according to the FPK model. We will now show how to relate the coefficients of M to the continuous diffusion equation (Eq. 1 in the main text of this article).

During an infinitesimal time step the trait can only move between neighboring sites, leading to our first condition: $M_{ij} = 0$ if $|i-j| > 1$. The next condition comes from the fact that the probability flux between two sites vanishes for the stationary distribution, which reads in its discrete form $P_i^* = \mathcal{N}\exp(-V_i)$. The probability flux between the neighboring sites i and $i+1$ is $M_{i+1,i}P_i^* - M_{i,i+1}P_{i+1}^*$, leading to a second condition for the coefficients of M:

$$\frac{M_{i+1,i}}{M_{i,i+1}} = \exp(V_i - V_{i+1}) \quad \forall i. \tag{8}$$

We must also ensure that M conserves probabilities, *i.e.* that the probabilities for a trait to evolve from site i to any of the sites on the grid sums to one. This can be written as $1 = \sum_i P_{ij}(t) = U^{\mathrm{T}} P_j(t)$, where U is a vector whose components are all equal to 1 ($\forall i$,

$U_i = 1$) and $P_j(t)$ is a vector whose components are $[P_j(t)]_i = P_{ij}(t)$. Taking the time derivative of $U^{\mathrm{T}}P_j(t) = 1$ and using the evolution equation 7 leads to $U^{\mathrm{T}}MP_j(t) = 0$. This relation being satisfied for all j and t, we end up with a second general condition on M, $U^{\mathrm{T}}M = 0$ or $\sum_i M_{ij} = 0$, which can be used to set the last condition:

$$M_{ii} = -(M_{i-1,i} + M_{i+1,i}) \quad \forall i. \tag{9}$$

It is straightforward to show that

$$M_{ij} = \exp\left(\frac{V_j - V_i}{2}\right) \quad \text{for} \quad |i - j| = 1, \tag{10}$$

satisfies the conditions obtained above. Eq. 9 is used to determine M_{ii} and all others terms are null.

Online Appendix III: Transformation of the actual trait interval to $[-1.5, +1.5]$

The choice of the $[-1.5, +1.5]$ interval to define the potential is only used for convenience when making calculations: it ensures that all shapes of macroevolutionary landscapes presented in Figure 2 of the Main text can be inferred. The actual width of this interval is arbitrary: any symmetric interval would insure that the same shapes of macroevolutionary landscapes can be obtained.

Since the range of traits spanned by species in the clade under study will never actually be $[-1.5, +1.5]$, observed trait values x_i are first converted to temporary variables x'_i distributed on the interval $[-1.5, +1.5]$ following an affine transformation: $x'=1.5 \cdot (2 \cdot \frac{x-B_{\min}}{B_{\max}-B_{\min}} - 1)$. All likelihood calculations are then done on a standardized potential V_{st} defined on $[-1.5, +1.5]$.

When this is done (*i.e.*, when the maximum of the likelihood function has been found, or when a sample of the posterior distribution has been taken), the potential is then translated back to $[B_{\min}, B_{\max}]$: the actual potential $V(x)$ experienced on $[B_{\min}, B_{\max}]$ is obtained as $V(x) = V_{st}(x')$.

Online Appendix IV: Parameter inference

Ancestral value of the trait at the root of the tree

The following figures shown results of parameter estimation for x_0. Figure 1 shows results for simulations of the pure FPK model, corresponding to scenarios a-d. Figure 2 shows results for simulations of the BBMV model, corresponding to scenarios e-h. The y-axis shows the ML value of x_0. The dashed horizontal line shows the actual used for simulations. Each boxplot summarizes results over 20 simulations. In both cases, estimation accuracy increases with tree size and has lower variance for small values of T_c.

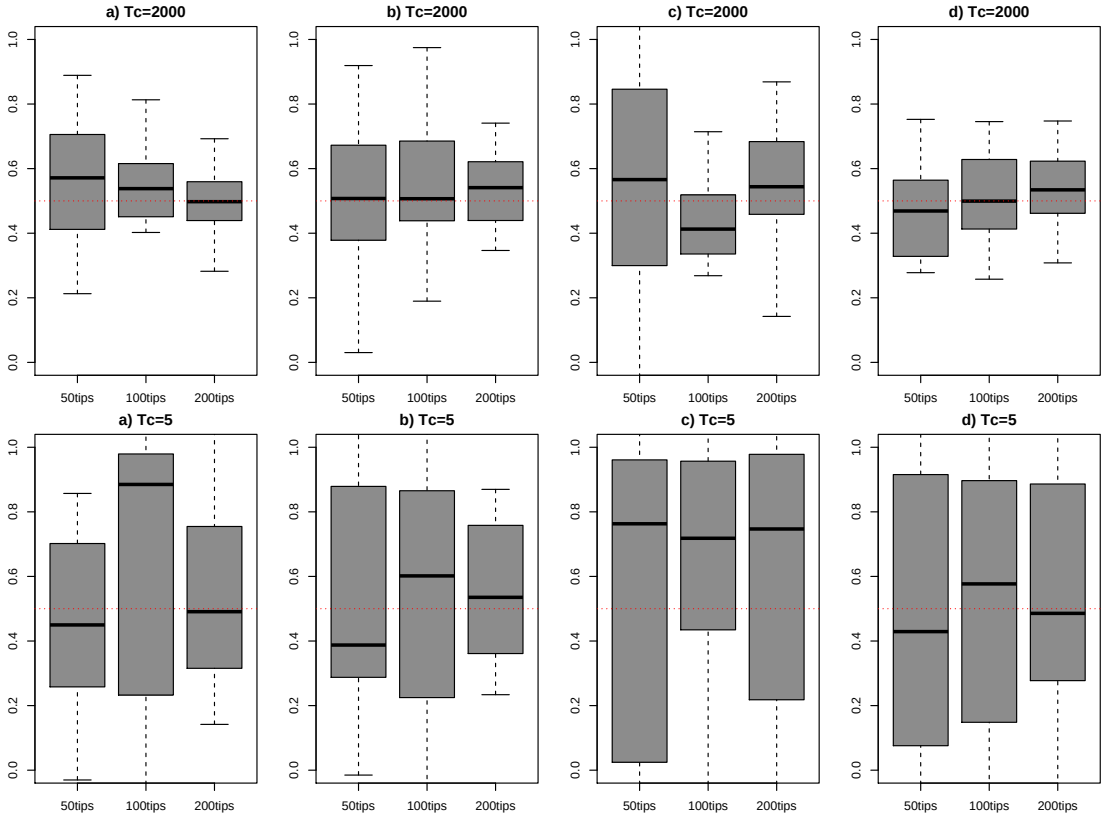

Figure 1: Estimation of x_0 in the FPK model. Values have been truncated between 0 and 1, which were the actual hard bounds used for simulations, but some estimates exceeded these values.

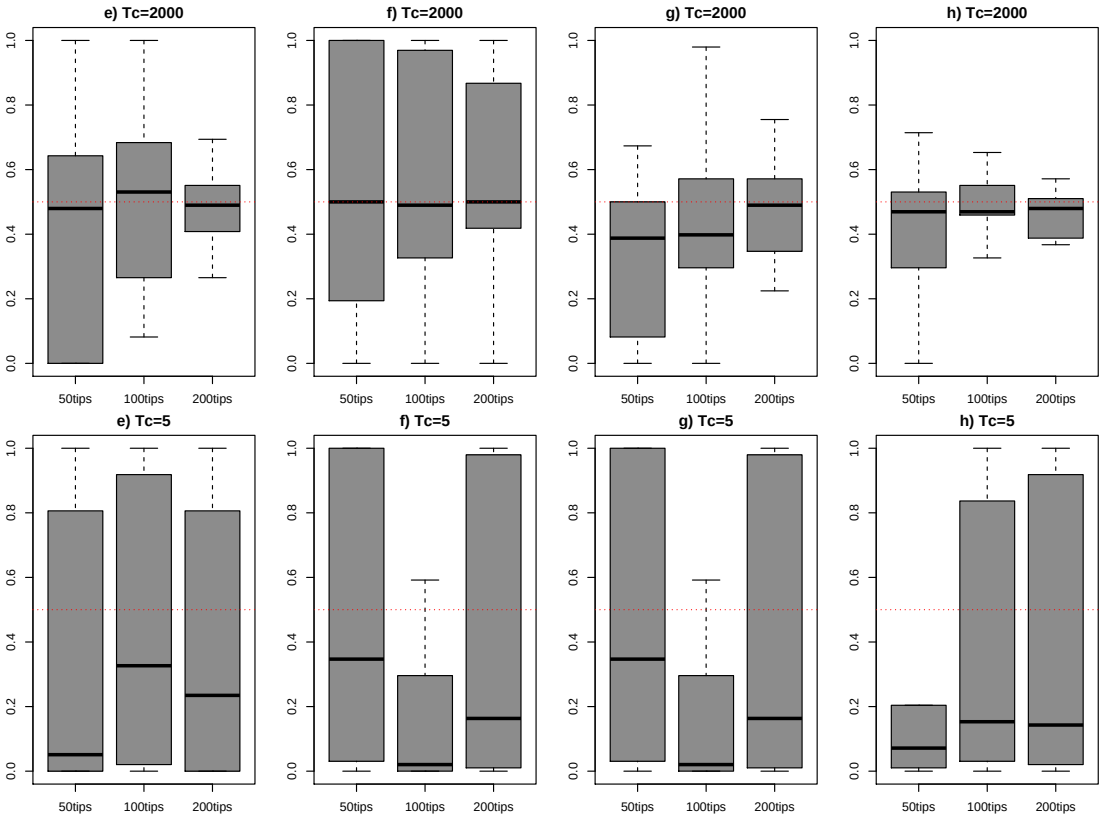

Figure 2: Estimation of x_0 in the BBMV model.

Rate of evolution

The following figures shown results of parameter estimation for σ^2. Figure 3 shows results for simulations of the pure FPK model, corresponding to scenarios a-d. Figure 4 shows results for simulations of the BBMV model, corresponding to scenarios e-h. The y-axis shows the logarithm of the ratio of the value of σ^2 that was estimated to the value that was simulated. The dashed horizontal line thus shows estimations exactly matching the simulated value. Each boxplot summarizes results over 20 simulations. In both cases, estimation accuracy increases with tree size and is higher for high values of T_c.

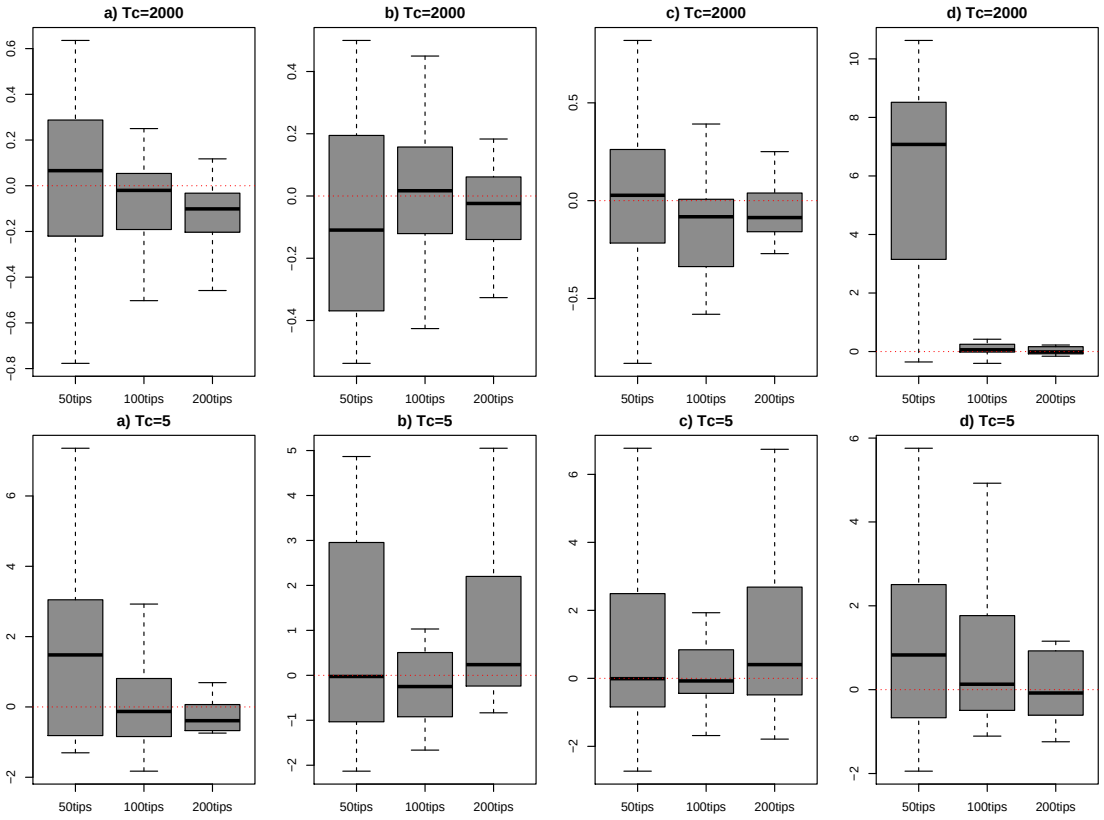

Figure 3: Estimation of σ^2 in the FPK model.

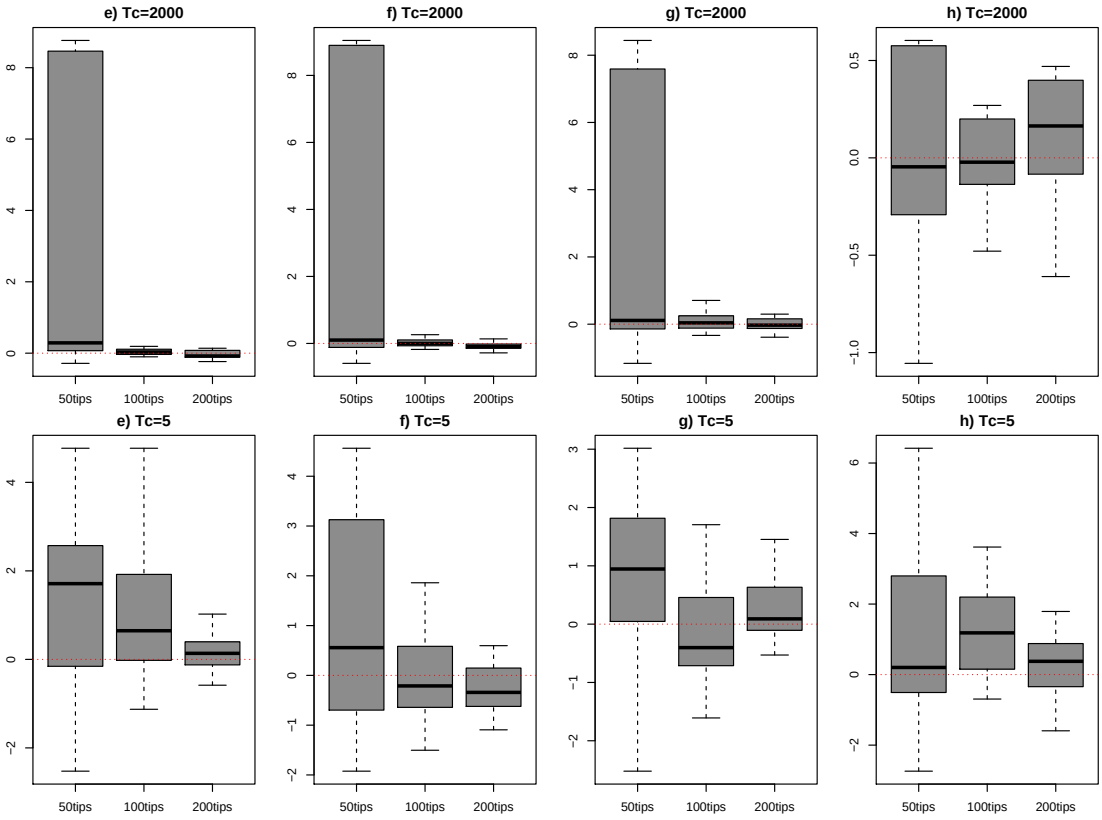

Figure 4: Estimation of σ^2 in the BBMV model.

Shape of the adaptive landscape

The following figures shown results for the estimation of the adaptive landscape in simulations of the FPK (Figures 5 & 6) and BBMV (Figures 7 & 8) models. In each case, thin lines show the macroevolutionary landscapes estimated in 20 simulations, each one in a different color, while the simulated macroevolutionary landscape is shown by the thick black line. Only results for trees with 50 and 200 tips are shown, results for trees of 100 tips are shown in the main text of the article. In each figure, the top row shows simulations with $T_c = 2,000$, in which stationarity was not reached, and the bottom row shows simulations with $T_c = 5$, in which stationarity was reached. In Figures 7 & 8, results for simulations of the flat landscape (BBM, scenario e) are not shown since the macroevolutionary landscape is fixed in this case.

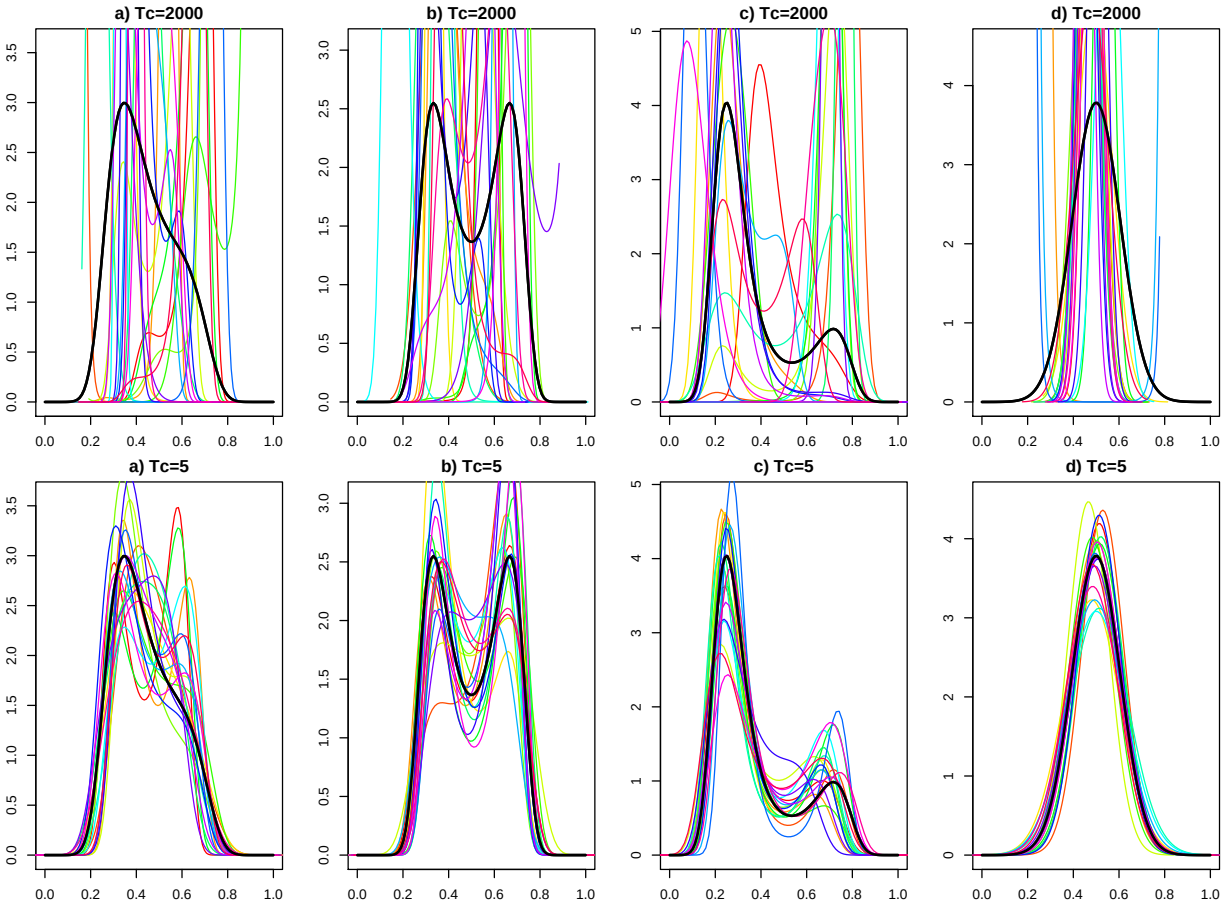

Figure 5: Estimation of the macroevolutionary landscape of the FPK model for trees of 50 tips.

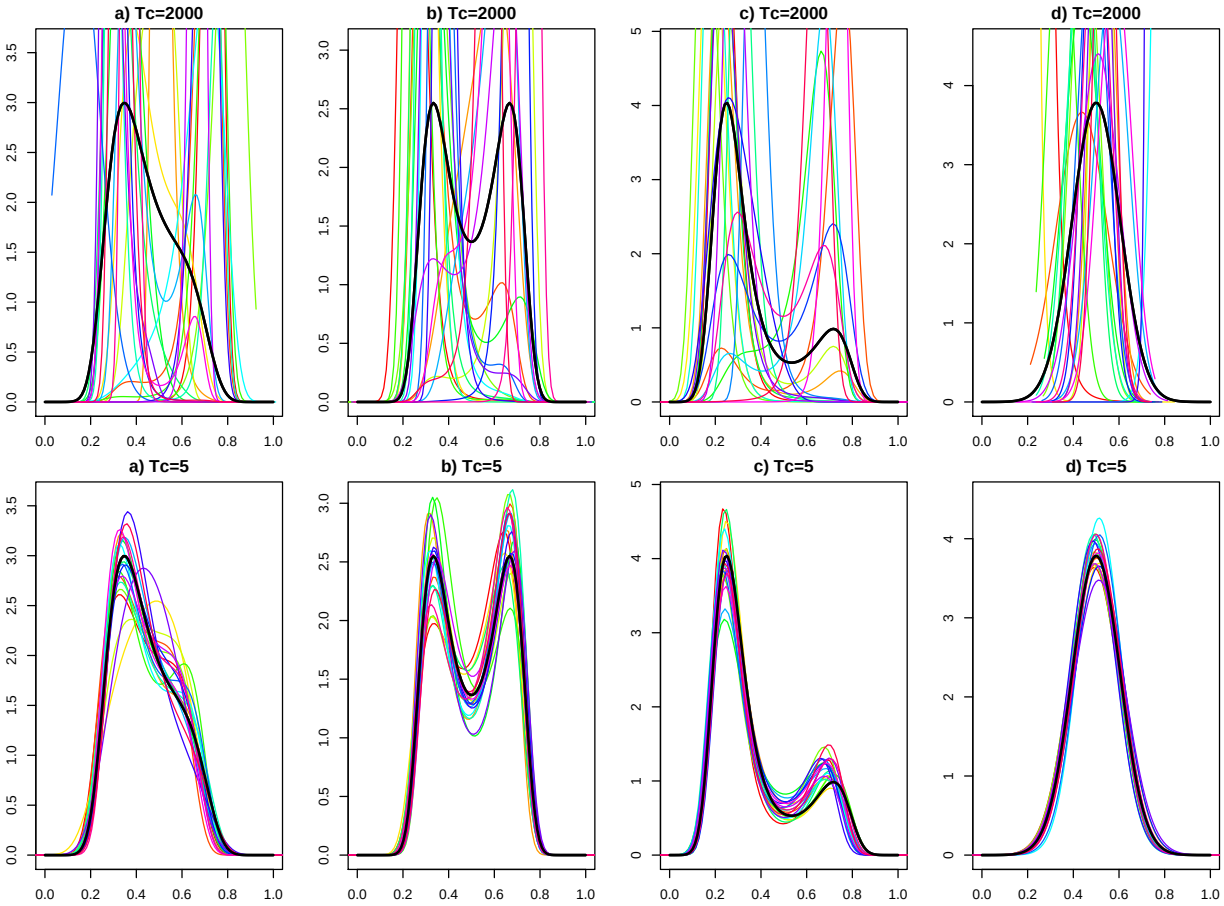

Figure 6: Estimation of the macroevolutionary landscape of the FPK model for trees of 200 tips.

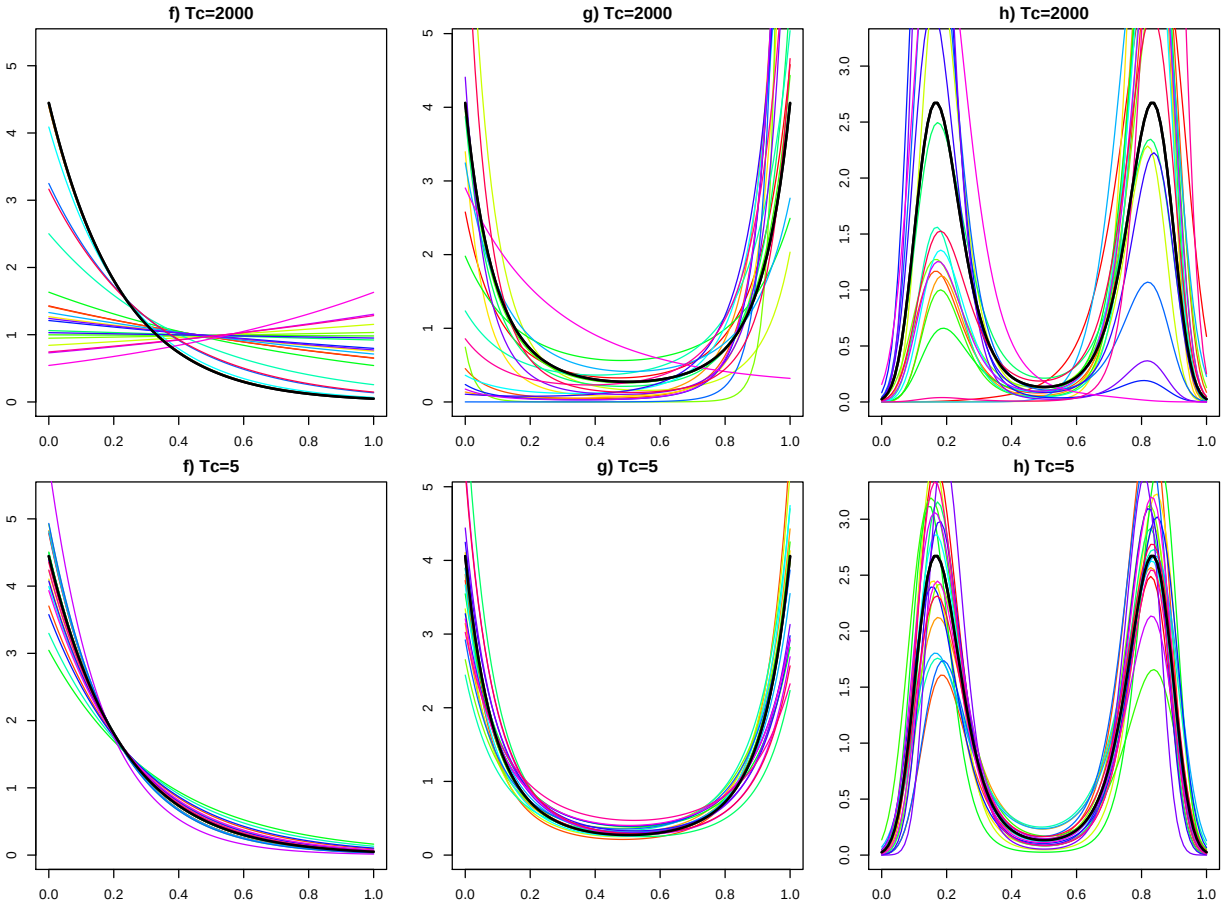

Figure 7: Estimation of the macroevolutionary landscape of the BBMV model for trees of 50 tips.

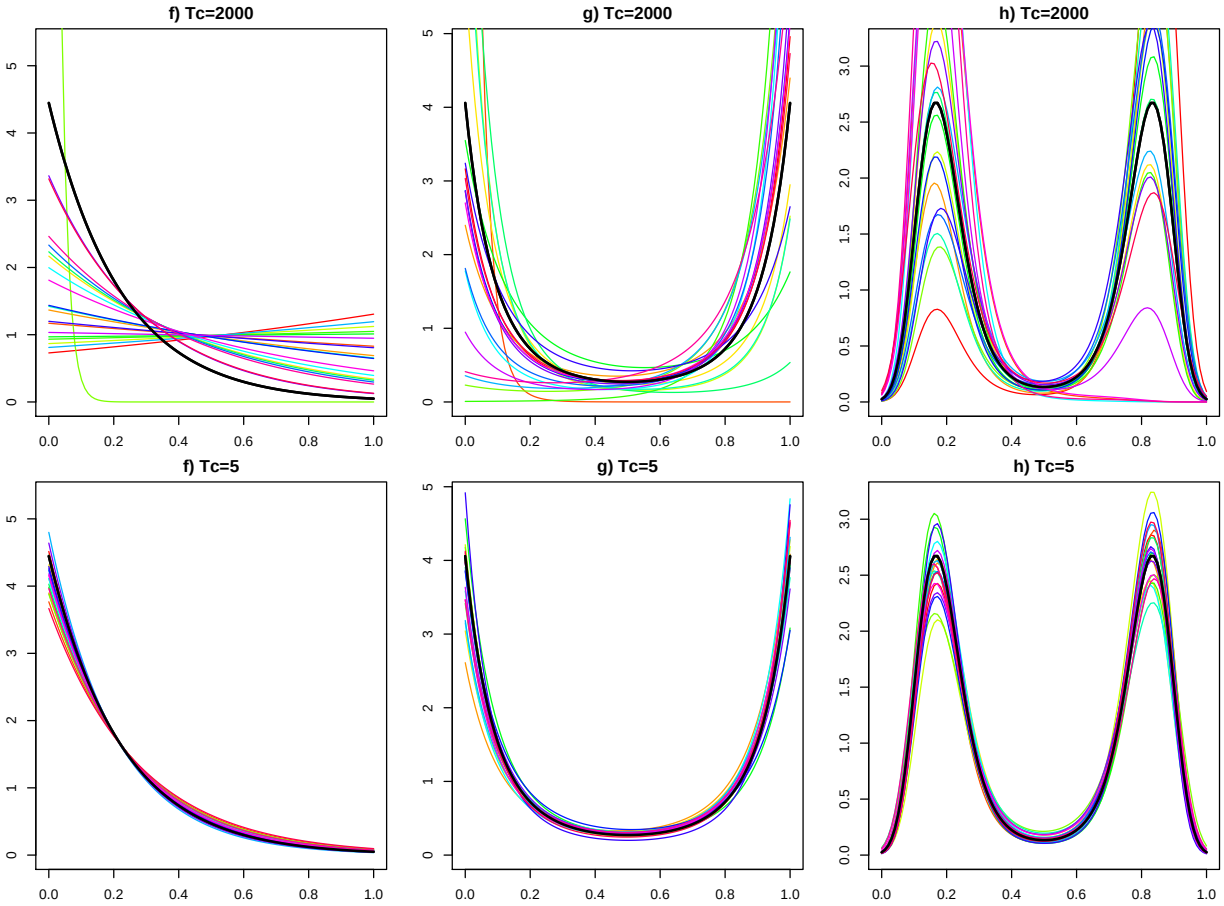

Figure 8: Estimation of the macroevolutionary landscape of the BBMV model for trees of 200 tips.

Online Appendix V: Inference of body size evolution in North-American watersnakes (tribe Thamnophiini) using the FPK model

Datasets

We used the FPK model to investigate the evolution of total length in North-American watersnakes (tribe Thamnophiini). As noted by Burbrink & Myers the distribution of total length in watersnakes is right-skewed and seems to have two modes (see Fig. 9). A dated phylogeny and total length data for 45 species were obtained from Burbrink & Myers 2014 (Glob. Ecol. Biogeo. 23:490-503, full reference in Main text) and can be seen on Figure 10.

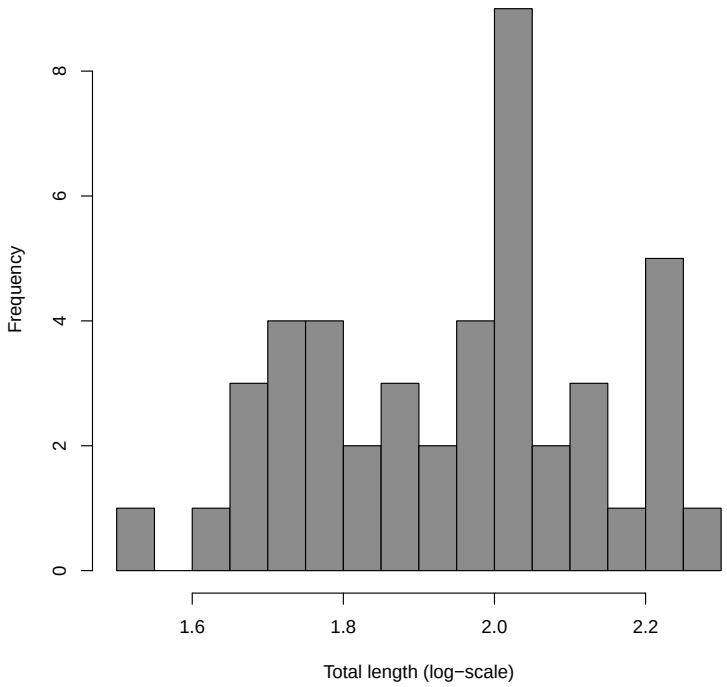

Figure 9: Distribution of log-transformed total length across Thamnophiini.

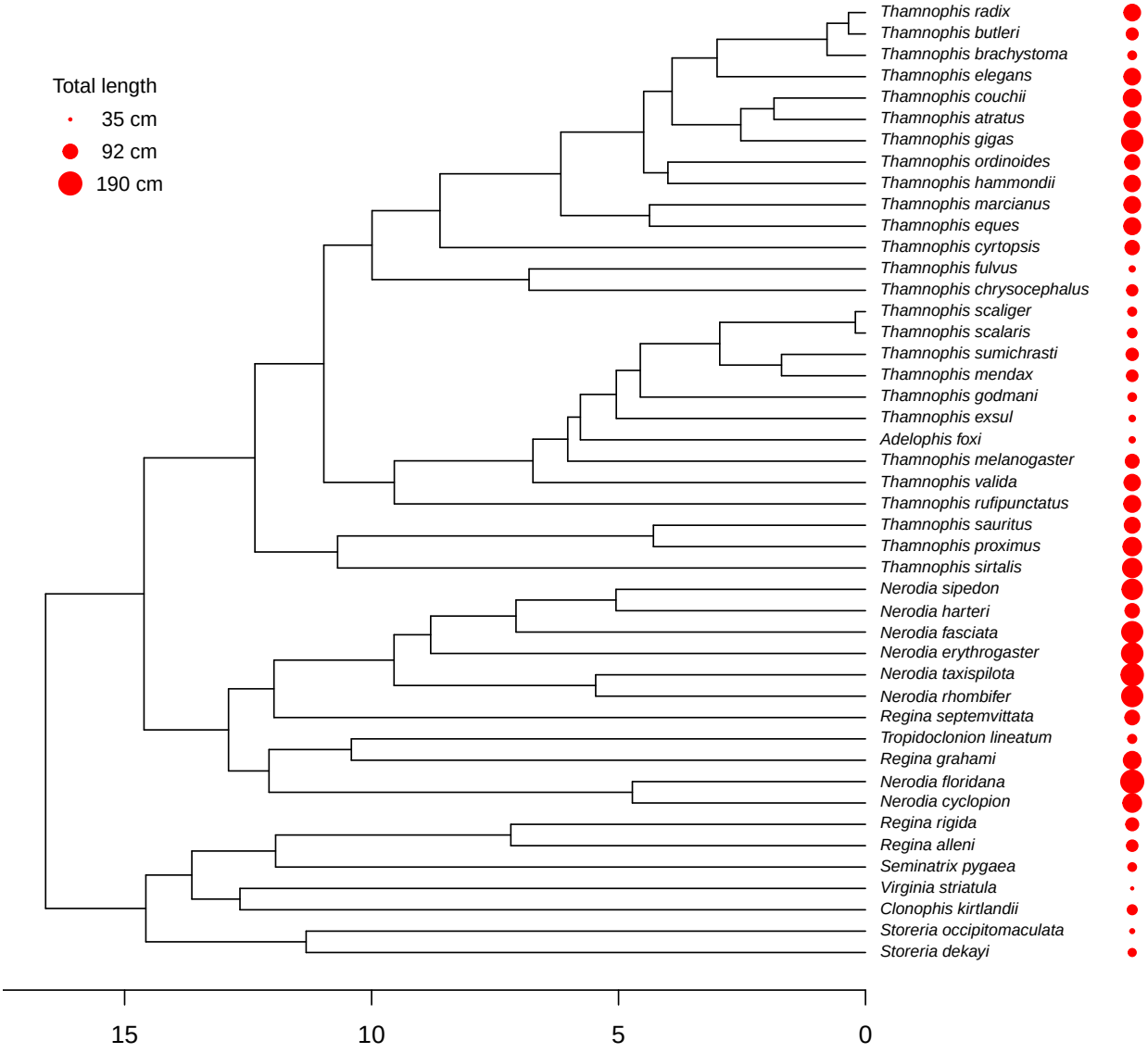

Figure 10: Phylogeny of Thamnophiini and distribution of total length across the group. Red circles in front of each tip of the phylogeny have a size proportional to the log-transformed total length of the species. The minimum, median, and maximum total lengths are shown on the top left panel. The timescale of the phylogeny is given below, in million years.

MCMC inference

We inferred parameters of the full FPK model with three polynomial terms describing the shape of the potential: $V(x) = ax^4 + bx^2 + cx$. Below is the R command used to run one of the MCMC chains:

```
chain=MH_MCMC_FPK(tree=tree_nat,trait=log_TL_nat,bounds=c(1.175290,2.640447),
    Nsteps=500000,record_every=100,plot_every=100,Npts=50,pars_init=c
    (3,-10,3,6,1),prob_update=c(0.2,0.2,0.2,0.2,0.2),verbose=TRUE,plot=TRUE,
    save_to='./results/MCMC_FPK_Natricideae.Rdata',save_every=100,type_priors=
    c(rep('Normal',4),'Uniform'),shape_priors=list(c(0,10),c(0,20),c(0,20),c
    (0,20),NA),proposal_type='Uniform',proposal_sensitivity=c
    (0.5,0.5,0.5,0.5,1),prior.only=F)
```

As shown in the R command above, two independent MCMC chains of 500,000 steps were run (parameter *Nsteps* in the function) and the trait interval was discretized into 50 regularly spaced points (parameter *Npts* in the function). Both MCMC chains were run using normal priors (parameter *type_priors* in the function) with wide standard deviations parameter *shape_priors* in the function) for all continuously varying parameters.

Convergence of each MCMC run was assessed by measuring the effective sample size (ESS) of all parameters of the model, as well as that of the prior, likelihood and posterior. For both runs, ESS was higher than 100 for c, x_0, the likelihood, and the posterior of the model, while ESS was higher than 50 for a, b, and the prior. Convergence of the two independent MCMC chains on the same posterior distribution was then assessed using Gelman and Rubin's convergence diagnostic, as implemented in the *gelman.diag* function of the R package *coda*. We then combined both runs together to produce an estimate of the posterior distribution, which yielded ESS$>$100 for all parameters as well as the likelihood.

Interpretation of results

As can be seen in Figure 11 (which corresponds to Fig. 5 in the main text), we inferred that the macroevolutionary landscape on which total length in watersnakes evolved has two distinct peaks.

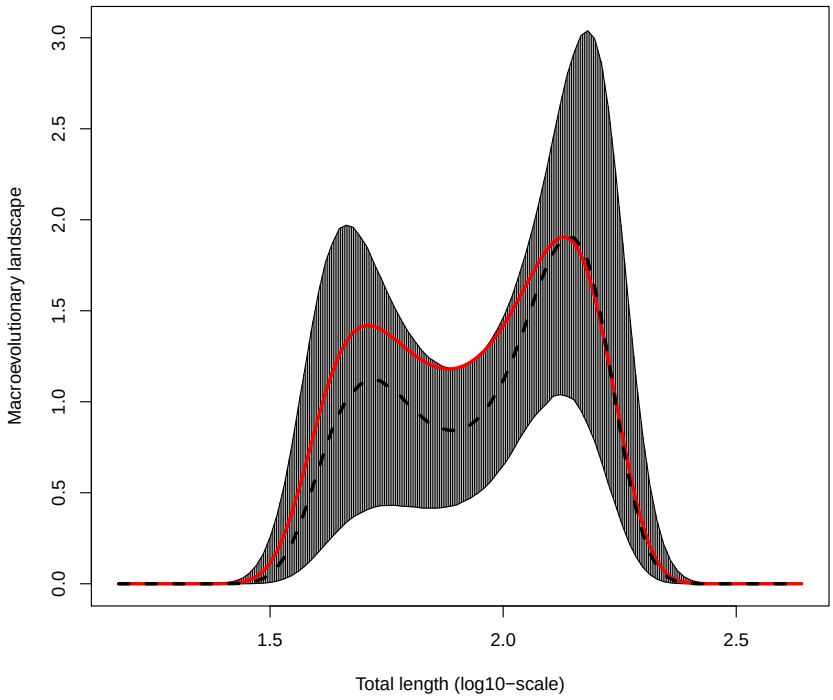

Figure 11: Posterior distribution of the macroevolutionary landscape estimated for body length evolution in watersnakes. This posterior distribution was obtained by concatenating the two MCMC chains after the first 10% of samples were discarded as burnin (800,000 MCMC steps in total). The figure shows the value of the macroevolutionary landscape ($\mathcal{N}\exp(-V(x))$) on the y-axis as a function of $\log_{10}(total\ length)$ measured in centimeters. The dashed black line shows the median value of the macroevolutionary landscape over the posterior, while the grey area ranges from the 25% to the 75% quantiles. The solid red line shows the maximum-likelihood estimate of the macroevolutionary landscape.

In order to get a better idea of the behavior of the FPK model that was inferred in this particular case, we have simulated traits evolving on the watersnake phylogeny under

the FPK model. Figure 12 shows six sets of simulations using maximum-likelihood estimates of parameters, with the value of the trait at the root of the trait randomly drawn from the density estimated on the real dataset. In this case, $T_{tot}/T_c = 0.74$. As can be seen from these six simulations, the outcome of trait evolution is highly stochastic and simulations generally show weak bimodality in the distribution of the trait at the tips of the tree, if any.

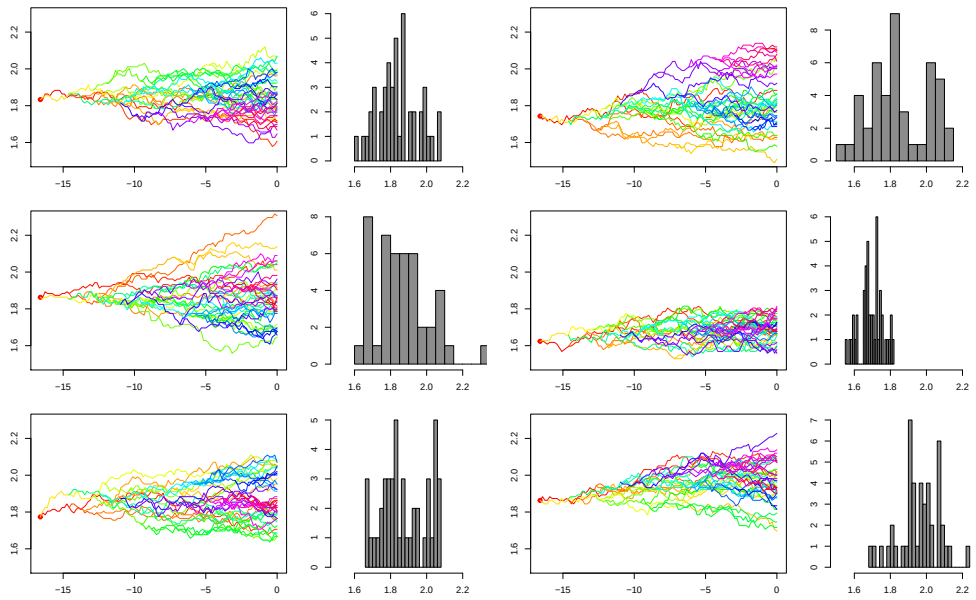

Figure 12: Simulations of trait evolution under the FPK model that was inferred for total length evolution in watersnakes. The tree used for simulations is the actual tree of watersnakes and trait evolution on different branches is shown in different colors. On the right of each of the six simulations, the histogram shows the distribution of total length at the tips of the tree.

As a comparison, Figure 13 show six simulations of the same FPK process, evolving on the same phylogeny, but with a rate of evolution (σ) that is 5x faster than the maximum-likelihood estimate. Since T_c is inversely proportional to σ^2, this corresponds to $T_{tot}/T_c = 18.5$. In this case the process is much closer to stationarity and all simulations show marked bimodality in the distribution of the trait at the tips of the tree. Sharp eyes might see that many switches between the two alternative peaks occur on single branches of the tree.

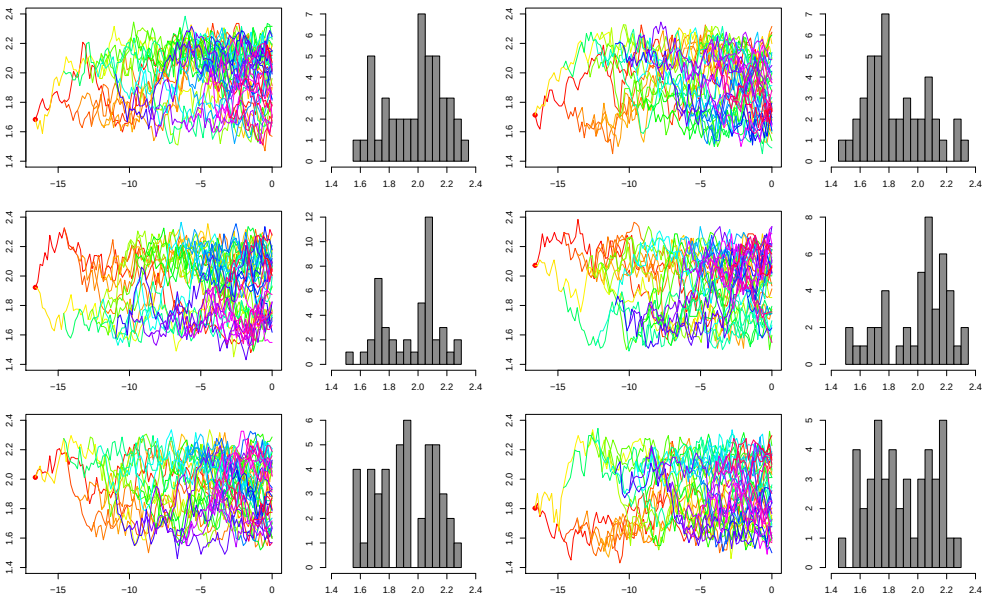

Figure 13: Simulations of trait evolution under the FPK model that was inferred for total length evolution in watersnakes, with a 5-fold increase of the rate of evolution.

Online Appendix VI: Phylogenetic signal in the FPK model

Since it has a strong deterministic component, the FPK model does not retain much phylogenetic signal (PS). Figures 14 & 15 show the values of the λ index of PS (Pagel 1997) estimated in simulations of the FPK and BBMV models. This index ranges from 0 when there is no PS, to 1, when PS is equal to the what is expected under BM. Each boxplot summarizes results over 20 simulations.

Under all scenarios simulated, PS is extremely high when stationarity in the FPK process was not reached (*i.e.*, when $T_c = 2,000$): this is normal since in this case the macroevolutionary landscape has been poorly explored yet and deterministic forces thus played only a minor role. On the contrary, when stationarity was reached ($T_c = 5$), simulated datasets showed almost no PS.

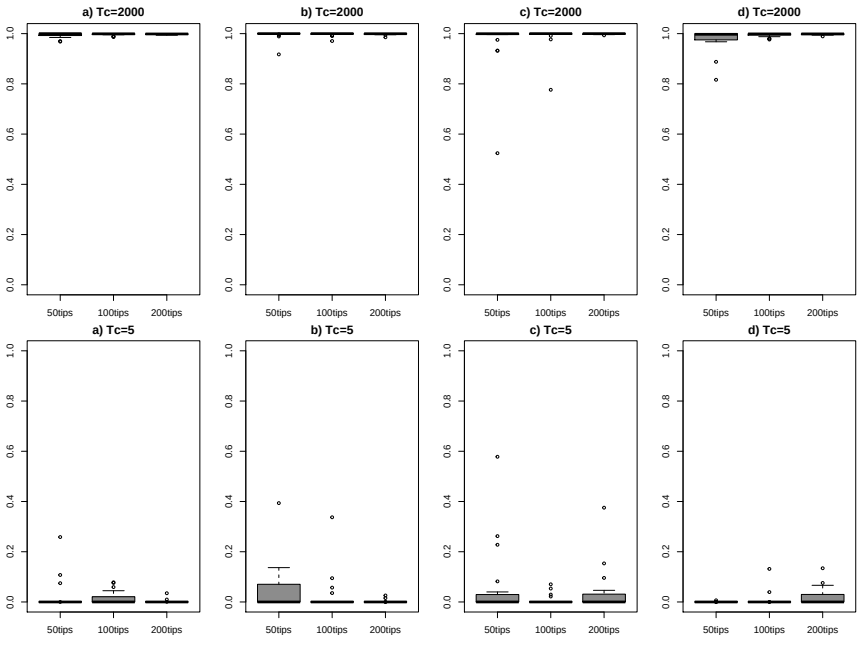

Figure 14: Estimated values of the λ index of PS in simulations of the FPK model.

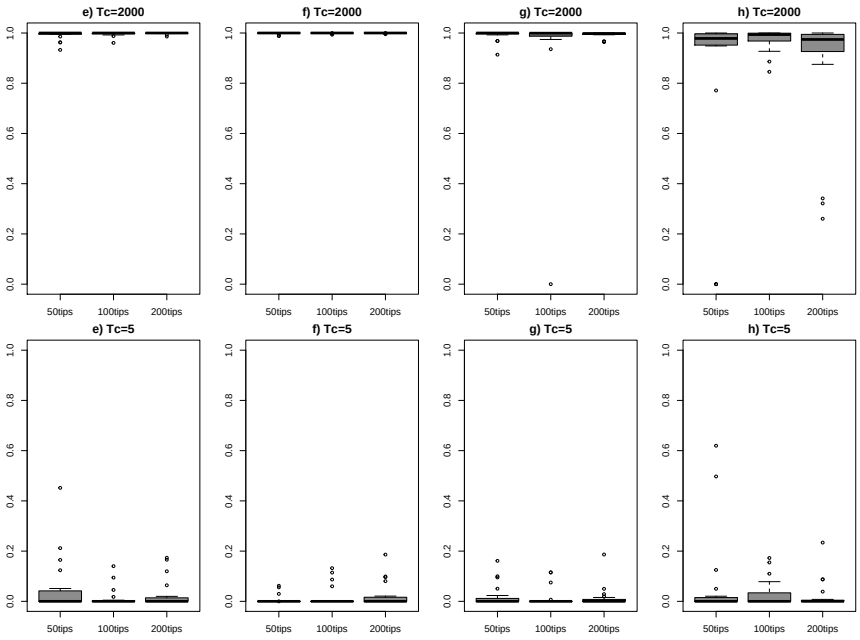

Figure 15: Estimated values of the λ index of PS in simulations of the BBMV model.

However, the FPK model can still be easily discriminated statistically from other models of evolution when PS is high. Our simulations indeed include examples in which FPK had not reached stationarity and thus still showed high PS but could nonetheless easily be discriminated from simpler models (e.g., OU) or even models with high PS (e.g., BM): this was the case for scenarios c, d, g, and h that we simulated (see Fig. 2 in the main text).

Even more interesting is the case of body size evolution in watersnakes (Online Appendix V). In this empirical case study, high phylogenetic signal was estimated (λ=0.84) while a FPK model with two peaks clearly provided a better fit than the OU model (ΔAIC=13.2) or even BM (ΔAIC=15.1). Furthermore, despite the relatively high PS, uncertainty in the shape of the macroevolutionary landscape was rather limited, whether estimated from maximum-likelihood confidence intervals (see main text) or bayesian credibility intervals (see Fig. 11). This suggests that relatively good accuracy in the estimation of the macroevolutionary landscape can be achieved in some cases even in the presence of PS.