Article

Syst. Biol. 00(0):114, 2016

DOI:10.1093/sysbio/syvxxx
Advance Access publication Xxx XX, XXXX

Inference of evolutionary jumps in large phylogenies using Lévy processes

PABLO DUCHEN[1], CHRISTOPH LEUENBERGER[1,2], SÁNDOR M. SZILÁGYI[3–5], LUKE HARMON[6], JONATHAN EASTMAN[6], MANUEL SCHWEIZER[7], AND DANIEL WEGMANN[1,*]

[1]*Faculty of Mathematics and Natural Sciences, Department of Biology, University of Fribourg, Chemin du Musée 10, 1700 Fribourg, Switzerland* [2]*Faculty of Mathematics and Natural Sciences, Department of Mathematics, University of Fribourg, Chemin du Musée 23, 1700 Fribourg, Switzerland* [3]*Faculty of Mathematics and Natural Sciences, Department of Informatics, University of Fribourg, Boulevard de Péroles 90, 1700 Fribourg, Switzerland* [4]*Faculty of Sciences and Letters, Department of Informatics, Petru Maior University, Str. N. Iorga Nr. 1, 540088 Tîrgu Mureș, Romania* [5]*Faculty of Electrical Engineering and Informatics, Department of Control Engineering and Information Technology, Budapest University of Technology and Economics, Magyar tudósok krt. 2, H-1117 Budapest, Hungary* [6]*Department of Biological Sciences, University of Idaho, 875 Perimeter Drive MS 3051, 1700 Moscow-Idaho, United States* [7]*Bern Natural History Museum, Bernastrasse 15, 3005 Bern, Switzerland*
[*]*Correspondence to be sent to: Department of Biology, University of Fribourg, Chemin du Musée 10, 1700 Fribourg, Switzerland; E-mail: daniel.wegmann@unifr.ch*

Received 00 Xxxxx 0000; reviews returned 00 Xxxx 0000; accepted 00 Xxxx 0000
Associate Editor: Xxxx Xxxx

Abstract.—While it is now widely accepted that the rate of phenotypic evolution may not necessarily be constant across large phylogenies, the frequency and phylogenetic position of periods of rapid evolution remain unclear. In his highly influential view of evolution, G. G. Simpson supposed that such evolutionary jumps occur when organisms transition into so called new adaptive zones, for instance after dispersal into a new geographic area, after rapid climatic changes, or following the appearance of an evolutionary novelty. Only recently, large, accurate and well calibrated phylogenies have become available that allow testing this hypothesis directly, yet inferring evolutionary jumps remains computationally very challenging. Here, we develop a computationally highly efficient algorithm to accurately infer the rate and strength of evolutionary jumps as well as their phylogenetic location. Following previous work we model evolutionary jumps as a compound process, but introduce a novel approach to sample jump configurations that does not require matrix inversions and thus naturally scales to large trees. We then make use of this development to infer evolutionary jumps in *Anolis* lizards and Loriinae parrots where we find strong signal for such jumps at the basis of clades that transitioned into new adaptive zones, just as postulated by Simpson's hypothesis.

Quantitative traits; Phenotyic evolution; Lévy process; Evolutionary jump; Punctuated equilibrium.

A key goal of evolutionary biology is to understand the mechanisms by which the large phenotypic diversity seen today evolved. Our understanding of these mechanisms is improving rapidly with the advent of increasingly powerful sequencing approaches. For instance, the huge amount of molecular data has led to the resolution of phylogenetic trees encompassing entire orders. Further, methods to reliably identify substitutions that likely resulted from selection, and to accurately place them on a phylogeny have been developed. In contrast, methods to infer events of rapid evolution from phenotypic data have lagged and are mostly restricted to inferring independent evolutionary rates for different clades.

In general, quantitative studies of the evolution of phenotypic/quantitative traits date back to just a few decades. A first attempt was by Edwards et al. (1964) and Cavalli-Sforza and Edwards (1967), who modeled quantitative traits stochastically as "Brownian motion" (BM). However, given the current wealth of molecular data available, a more realistic goal is to only aim at inferring the rates at which quantitative traits evolve, while assuming the underlying phylogeny to be known. This has been successfully done using a BM model in multiple taxa. Freckleton et al. (2002), for instance, used a BM model on a given phylogeny to test if traits showed phylogenetic associations. More recently, Brawand et al. (2011) modeled gene expression evolution as BM and rejected evolution at a constant rate for several genes.

Several extensions to a basic BM model have been proposed. Butler and King (2004) were the first to implement Ornstein-Uhlenbeck (OU) processes with multiple evolutionary optima, as initially described by Hansen (1997), and recently used to describe the evolution of gene expression (e.g. Bedford and Hartl, 2009; Rohlfs et al., 2013). Other extentions to BM allow evolutionary rates to change over time. O'Meara et al. (2006), for instance, contrasted maximum likelihood (ML) estimates of evolutionary rates under BM and showed that major clades of angiosperms vastly differ in their rate of genome size evolution. More recently, Eastman et al. (2011) developed a Bayesian method to jointly infer evolutionary rates in different clades and found evidence for multiple rate shifts in body size evolution in emydid turtles. Shortly after, Slater et al. (2012) have introduced an extension to incompletely sampled phylogenies and trait data using Approximate Bayesian Computation. However, they found no evidence for an elevated rate of body size evolution in pinnipeds in comparison to terrestrial carnivores, despite considerable power. This suggests that the larger body size found

in pinnipeds may be the result of rapid evolutionary changes early in the clade, rather than a change in the rate itself, and hence that models of occasional "evolutionary jumps" may often more accurately explain the evolution of quantitative traits.

According to Simpson (1944), such evolutionary jumps are triggered by shifts of lineages into different adaptive zones, either by dispersal into new geographic areas, the appearance of evolutionary novelties, key innovations, the extinction of lineages leaving niches empty, or by rapid changes in the environment (climatic or ecological). Additionally, the existence of "ecological opportunities" (Losos, 2010) might also trigger such jumps. While OU processes have been proposed to model the dynamics of adaptive landscapes (e.g. Ingram and Mahler, 2013; Uyeda and Harmon, 2014), a promising alternative is to model this type of evolution as a compound process (or Lévy process) consisting of a continuous background process and a discrete jump process. The first implementation of such a model assumed that jumps only occured at speciation events (Bokma, 2008), but Landis et al. (2013) recently described Lévy processes in a much more general way and showed that while the likelihood functions of most of these models are intractable, inference is possible under a Bayesian framework. For instance, when modeling the evolution of quantitative traits as a Poisson compound process, in which traits are assumed to evolve under BM with occasional jumps that occur as a Poisson process on the tree, the likelihood can be calculated analytically when conditioning on a jump configuration (a placement of jumps on the tree). Under the assumption that jump effects are normally distributed, a jump configuration can be seen as simply stretching the branches of the tree on which they occur, and the likelihood is then given by a multivariate normal distribution with the variance-covariance matrix resulting from the stretched tree. The numerical integration is then limited to sampling jump configurations, which is readily done using Markov Chain Monte Carlo (MCMC).

Unfortunately, two computational challenges prohibit the application of this approach to larger trees. First, the space of jump configurations grows exponentially with tree size, leading to very long MCMC chains. Second, the evaluation of the likelihood requires the computation of the inverse of the variance-covariance matrix, which is computationally very demanding since it scales exponentially with tree size (Tung Ho and Ané, 2014). Here, we address these computational issues using an empirical Bayes approach in which we first infer the hierarchical parameters of the Brownian and Poissonian processes using Maximum Likelihood, and then fix those when inferring posterior probabilities on jump locations. This approach allows us to run MCMC chains with fixed hierarchical parameters, for which we find a computationally highly efficient approach that does not require matrix inversions. As a result, this approach readily scales up to very large phylogenies.

We then demonstrate the power and accuracy of our approach with extensive simulations and find that our approach hardly misses any jumps with a meaningful strength. We then illustrate the usefulness of our approach by identifying evolutionary jumps in *Anolis* lizards and Loriini parrots, two well-studied groups for which morphological data is available. We identify few but important evolutionary jumps in both groups, suggesting such periods of rapid evolutionary change to be rare but crucial in shaping the morphological diversity observed today.

THEORY

The null hypothesis: Brownian motion

We first consider a Brownian motion (BM) process on a phylogenetic tree $\mathcal{T}$ with root $\mathcal{O}$ where time is measured in the unit of the branch lengths. The process starts at $\mathcal{O}$ with value $\mu \in \mathbb{R}$ (root state) and then proceeds with variance s_0^2 along the branches. The values of the BM process, as observed at the L leaves, give rise to the random vector

$$\boldsymbol{x} = (x_1,\ldots,x_L)'.$$

Let us fix the notation: The lenghts of the (inner and outer) branches of $\mathcal{T}$ are called $\tau_1,\ldots,\tau_B$ where B is the number of branches. For two leaves i,j we denote by $\boldsymbol{T}_0 = (\tau_{ij})$ the length of their common branch in $\mathcal{T}$ as measured from the root $\mathcal{O}$. Now, under the assumption of a pure BM, and defining $\mathbf{1} = (1,1,\ldots,1)'$, the values $\boldsymbol{x}$ at the leaves have the multivariate normal distribution

$$\boldsymbol{x} \sim \mathcal{N}(\mu\mathbf{1}, s_0^2\boldsymbol{T}_0)$$

or written more conveniently:

$$\boldsymbol{x} = \mu\mathbf{1} + \boldsymbol{\epsilon} \tag{1}$$

with $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, s_0^2\boldsymbol{T}_0)$. Since $\boldsymbol{T}_0$ is positive definite and symmetric, it has a symmetric and positive definite square root $\boldsymbol{Q}$, i.e. $\boldsymbol{Q}^2 = \boldsymbol{T}_0$. Multiplying both sides of (1) with $\boldsymbol{Q}^{-1}$ we get the homoskedastic model

$$\boldsymbol{x}_0 = \mu\boldsymbol{v}_0 + \boldsymbol{\epsilon}_0,$$

where $\boldsymbol{x}_0 = \boldsymbol{Q}^{-1}\boldsymbol{x}$, $\boldsymbol{v}_0 = \boldsymbol{Q}^{-1}\mathbf{1}$, and $\boldsymbol{\epsilon}_0 \sim \mathcal{N}(\mathbf{0}, s_0^2\boldsymbol{I})$. For this we have the usual OLS estimators (see e.g. Davidson and MacKinnon (2004), ch. 3.2)

$$\hat{\mu} = (\boldsymbol{v}_0'\boldsymbol{v}_0)^{-1}\boldsymbol{v}_0'\boldsymbol{x}_0 = \frac{\mathbf{1}'\boldsymbol{T}_0^{-1}\boldsymbol{x}}{\mathbf{1}'\boldsymbol{T}_0^{-1}\mathbf{1}}$$

and

$$\begin{aligned}\hat{s}_0^2 &= \frac{1}{L-1}\boldsymbol{v}_0'\left(\boldsymbol{I} - \boldsymbol{v}_0(\boldsymbol{v}_0'\boldsymbol{v}_0)^{-1}\boldsymbol{v}_0'\right)\boldsymbol{x}_0 \\ &= \frac{1}{L-1}\left(\boldsymbol{x}'\boldsymbol{T}_0^{-1}\boldsymbol{x} - \frac{(\mathbf{1}'\boldsymbol{T}_0^{-1}\boldsymbol{x})^2}{\mathbf{1}'\boldsymbol{T}_0^{-1}\mathbf{1}}\right).\end{aligned}$$

Lévy process

We now extend the BM model by super-imposing an independent Poissonian jump-process with rate λ. The jumps shall be normally distributed with zero mean and variance s_1^2. The (unobservable) random vector

$$\boldsymbol{\nu}=(\nu_1,\ldots,\nu_B)'$$

counts the number of Poisson events (jumps) on each of the B branches. By assumption,

$$\mathbb{P}(\nu_b=n_b)=e^{-\lambda\tau_b}\frac{(\lambda\tau_b)^{n_b}}{n_b!},\quad n_b=0,1,2,\ldots.$$

For a multi-index $\boldsymbol{n}=(n_1,\ldots,n_B)$, we have

$$\mathbb{P}(\boldsymbol{\nu}=\boldsymbol{n})=\prod_{k=1}^{B}e^{-\lambda\tau_b}\frac{(\lambda\tau_b)^{n_b}}{n_b!}. \tag{2}$$

Recall that for two leaves i,j we denote by τ_{ij} the length of their common branch in $\mathcal{T}$ as measured from the root $\mathcal{O}$. In particular, τ_{ii} is the distance (sum of branch lengths) of the leaf i from the root $\mathcal{O}$.

We denote by n_{ij} for two leaves i,j the number of Poisson events along the common branch of length τ_{ij}. Conditional on $\boldsymbol{\nu}=\boldsymbol{n}=(n_1,\ldots,n_B)$, the random vector $\boldsymbol{x}$ is multivariate normal with mean $\mu\mathbf{1}$ and the $L\times L$ variance-covariance matrix $\boldsymbol{\Sigma}(\boldsymbol{n})=(\sigma_{ij}(\boldsymbol{n}))$ where

$$\sigma_{ij}(\boldsymbol{n})=\tau_{ij}s_0^2+n_{ij}s_1^2,\quad 1\le i,j\le L.$$

The conditional density of $\boldsymbol{x}$ given $\boldsymbol{\nu}=\boldsymbol{n}$ is

$$\phi(\boldsymbol{x}|\boldsymbol{n})=\frac{1}{\sqrt{(2\pi)^L\det\boldsymbol{\Sigma}(\boldsymbol{n})}}\cdot\exp\left(-\frac{1}{2}(\boldsymbol{x}-\mu\mathbf{1})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{n})(\boldsymbol{x}-\mu\mathbf{1})\right). \tag{3}$$

The likelihood of $\boldsymbol{x}$ given the four parameters μ (root state), s_0 (Brownian motion) and λ,s_1 (Poissonian jump process) is the mixture distribution

$$f(\boldsymbol{x}|\mu,s_0^2,\lambda,s_1^2)=\sum_{n_1=0}^{\infty}\cdots\sum_{n_B=0}^{\infty}\mathbb{P}(\boldsymbol{\nu}=\boldsymbol{n})\phi(\boldsymbol{x}|\boldsymbol{n}) \tag{4}$$

where we used expressions (2) and (3). It is not hard to show that

$$\mathbb{E}(x_i)=\mu\quad\text{and}\quad\mathrm{Cov}(x_i,x_j)=\tau_{ij}(s_0^2+\lambda s_1^2).$$

Inference under the Lévy process

Here we develop a computationally efficient approach to maximize the likelihood function given in equation (4). While the infinite sums in (4) prohibit an analytical solution, they are readily evaluated using numerical approaches. Landis et al. (2013), for instance, proposed to use an MCMC approach to integrate over jump configurations. Unfortunately, however, such a solution does not scale to large trees, because the calculation of the conditional density values in (3) involves the computation of the inverse of $\boldsymbol{\Sigma}(\boldsymbol{n})$ and its determinant, which are computationally very demanding.

We propose to address this problem by introducing an algorithm to calculate these matrix inversions efficiently under this model. While this algorithm can readily be incorporated into the MCMC approach proposed by Landis et al. (2013), we will then propose an alternative hierarchical Bayes approach that makes even better use of it and leads to a computationally highly efficient inference approach to obtain point estimates of the parameters μ, s_0^2, λ, and s_1^2, as well as posterior probabilities on the location of evolutionary jumps.

Efficient calculation of inverses and determinants For a symmetric non-singular matrix $\boldsymbol{A}$ and a (column) vector $\boldsymbol{a}$, we have

$$(\boldsymbol{A}\pm\boldsymbol{a}\boldsymbol{a}')^{-1}=\boldsymbol{A}^{-1}\mp\frac{1}{1\pm\boldsymbol{a}'\boldsymbol{A}^{-1}\boldsymbol{a}}(\boldsymbol{A}^{-1}\boldsymbol{a})(\boldsymbol{A}^{-1}\boldsymbol{a})' \tag{5}$$

(see Izenman (2008), p. 47) and

$$\det(\boldsymbol{A}\pm\boldsymbol{a}\boldsymbol{a}')=\det\boldsymbol{A}\cdot(1\pm\boldsymbol{a}'\boldsymbol{A}^{-1}\boldsymbol{a}) \tag{6}$$

(see Anderson (2003), Corollary A.3.1). These formulae have recently been shown to speed up the calculation of the likelihodo function under Brownian motion models (Tung Ho and Ané, 2014). Here we use them to develop a fast algorithm applicable to Lévy processes.

Let us first fix some notation: For each branch b, we define the $L\times L$ incidence matrix $\boldsymbol{I}^b=(I^b_{ij})$ by setting $I^b_{ij}=1$ if the branch b is common to the pair of leaves i,j, and $I^b_{ij}=0$ otherwise. Clearly,

$$n_{ij}=\sum_{b=1}^{B}n_bI^b_{ij}.$$

In the following we replace the parameter s_1^2 with the positive factor α given by

$$s_1^2=\alpha s_0^2.$$

Observe that

$$\boldsymbol{\Sigma}(\boldsymbol{n})=s_0^2\boldsymbol{T}(\boldsymbol{n},\alpha)$$

and

$$\det\boldsymbol{\Sigma}(\boldsymbol{n})=s_0^{2L}\det\boldsymbol{T}(\boldsymbol{n},\alpha)$$

where

$$\boldsymbol{T}(\boldsymbol{n},\alpha)=\boldsymbol{T}_0+\alpha\sum_{b=1}^{B}n_b\boldsymbol{I}^b$$

and $\boldsymbol{T}_0=(\tau_{ij})$. Finally, we introduce for $b=1,\ldots,B$ the (column) vectors $\boldsymbol{u}^b$, each one with L components. The i-th component u^b_i is equal to 1 if leaf i is subordinate to branch b (i.e. the path from the root $\mathcal{O}$ to node i contains branch b). Otherwise, if leaf i is not subordinate to branch b, then $u^b_i=0$. It is easy to see that $\boldsymbol{I}^b=\boldsymbol{u}^b(\boldsymbol{u}^b)'$

and thus

$$\boldsymbol{T}(\boldsymbol{n},\alpha)=\boldsymbol{T}_0+\alpha\sum_{b=1}^{B}n_b\boldsymbol{u}^b(\boldsymbol{u}^b)'. \tag{7}$$

We can now apply formulae (5) and (6) to obtain the following iterative scheme for the computation of $\boldsymbol{T}^{-1}(\boldsymbol{n},\alpha)$ and $\det\boldsymbol{T}(\boldsymbol{n},\alpha)$:

First, determine $\boldsymbol{T}_0^{-1}$ and $\det\boldsymbol{T}_0$. Then, for each term with $n_b>0$ in the sum (7), update $\boldsymbol{T}_{b-1}$ to $\boldsymbol{T}_b$ etc. as follows: Let

$$r_b=1+\alpha\ n_b\cdot(\boldsymbol{u}^b)'\ \boldsymbol{T}_{b-1}^{-1}\boldsymbol{u}^b$$

and calculate

$$\begin{aligned}\boldsymbol{T}_b^{-1}&=\boldsymbol{T}_{b-1}^{-1}-\frac{\alpha\ n_b}{r_b}\cdot(\boldsymbol{T}_{b-1}^{-1}\boldsymbol{u}^b)(\boldsymbol{T}_{b-1}^{-1}\boldsymbol{u}^b)',\\ \det\boldsymbol{T}_b&=r_b\cdot\det\boldsymbol{T}_{b-1}.\end{aligned} \tag{8}$$

When all non-zero terms in (7) have been considered, we arrive at $\boldsymbol{T}_B^{-1}=\boldsymbol{T}^{-1}(\boldsymbol{n},\alpha)$ and $\det\boldsymbol{T}_B=\det\boldsymbol{T}(\boldsymbol{n},\alpha)$. Observe that in this scheme, the only matrix inverse that ever has to be determined is $\boldsymbol{T}_0^{-1}$. The number of non-zero n_b will frequently be small compared to B and so will be the number of iterations (8).

Monte Carlo EM algorithm The scheme to calculate the inverse of $\boldsymbol{\Sigma}(\boldsymbol{n})$ allows to find the ML estimates of the parameters μ,s_0^2 and λ by means of a Monte Carlo version of the classical Expectation Maximization (EM) algorithm, in which we treat the random variable $\boldsymbol{\nu}$ as missing (unobserved) data. While this approach does not allow us to find the ML estimate of α, we discuss below how this can be achieved using a simple grid search.

Recall that each iteration of the EM algorithm consists of an estimation (E) and a maximization (M) step. Let us denote the old parameters determined in the previous M-step by $\tilde\theta=(\tilde\mu,\tilde s_0^2,\tilde\lambda,\alpha_0)$, and the new parameters with respect to which the Q-function has to be maximized in the next M-step by $\theta=(\mu,s_0^2,\lambda,\alpha_0)$, where α_0 is a fixed value for α. The two steps of the EM algorithm are then as follows:

Monte Carlo E-step. Simulate stochastically K vectors $\boldsymbol{n}_k$ according to the multi-Poisson distribution $\mathbb{P}(\boldsymbol{\nu}=\boldsymbol{n}|\tilde\lambda)$. Determine the weights

$$\begin{aligned}\pi_k&=\phi(\boldsymbol{x}|\boldsymbol{n}_k,\tilde\mu,\tilde s_0^2,\alpha_0)\\ &=c_k\cdot\exp\left(-\frac{1}{2\tilde s_0^2}\cdot(\boldsymbol{x}-\tilde\mu\boldsymbol{1})'\cdot\boldsymbol{T}^{-1}(\boldsymbol{n}_k,\alpha_0)\cdot(\boldsymbol{x}-\tilde\mu\boldsymbol{1})\right),\end{aligned}$$

with

$$c_k=(2\pi\tilde s_0^2)^{-L/2}\cdot(\det\boldsymbol{T}(\boldsymbol{n}_k,\alpha_0))^{-1/2}.$$

In the M-step we have to maximize the function

$$\begin{aligned}Q(\theta|\tilde\theta)&=\mathbb{E}\left[\log\mathbb{P}(\boldsymbol{x},\boldsymbol{n}|\theta)|x,\tilde\theta\right]\\ &=\sum_{\boldsymbol{n}}\mathbb{P}(\boldsymbol{n}|\boldsymbol{x},\tilde\theta)\log\mathbb{P}(\boldsymbol{x},\boldsymbol{n}|\theta)\end{aligned} \tag{9}$$

with respect to the parameters $\theta=(\mu,s_0^2,\lambda,\alpha_0)$ where

$$\mathbb{P}(\boldsymbol{x},\boldsymbol{n}|\theta)=\phi(\boldsymbol{x}|\boldsymbol{n},\mu,s_0^2,\alpha_0)\mathbb{P}(\boldsymbol{\nu}=\boldsymbol{n}|\lambda)$$

From Bayes' theorem we have

$$\mathbb{P}(\boldsymbol{n}|\boldsymbol{x},\tilde\theta)=\frac{\phi(\boldsymbol{x}|\boldsymbol{n},\tilde\mu,\tilde s_0^2,\alpha_0)\mathbb{P}(\boldsymbol{\nu}=\boldsymbol{n}|\tilde\lambda)}{\mathbb{P}(\boldsymbol{x}|\tilde\theta)}. \tag{10}$$

Thus, according to our Monte Carlo scheme and up to the factor $1/\mathbb{P}(\boldsymbol{x}|\tilde\theta)$, the infinite sum in (9) can be approximated by

$$Q(\theta)\propto\sum_{k=1}^{K}\pi_k\log[\phi(\boldsymbol{x}|\boldsymbol{n}_k)\mathbb{P}(\boldsymbol{\nu}=\boldsymbol{n}_k)] \tag{11}$$

where $\phi(\boldsymbol{x}|\boldsymbol{n}_k)$ and $\mathbb{P}(\boldsymbol{\nu}=\boldsymbol{n}_k)$ are given by (2) and (3), respectively.

M-step. In this step we seek the parameters $\tilde\theta$ which maximize the sum in (11) and which will serve as "old" parameters in the next E-step. We have

$$\mathbb{P}(\boldsymbol{\nu}=\boldsymbol{n}_k)=e^{-\lambda\tau}\lambda^{|\boldsymbol{n}_k|}c(\boldsymbol{n}_k)$$

where $\tau=\sum\tau_i$ is the total length of the tree $\mathcal{T}$, $|\boldsymbol{n}_k|$ denotes the sum of the components of $\boldsymbol{n}_k$, and $c(\boldsymbol{n}_k)$ is a factor that does not depend on any of the parameters θ. From this it is easy to see that

$$\tilde\lambda=\frac{\sum\pi_k|\boldsymbol{n}_k|}{\tau\sum\pi_k}, \tag{12}$$

independently of the values of the other three parameters. Since we assume the value of α to be fixed, we can also give explicit expressions for the values of μ and s_0^2 which maximize $Q(\theta|\alpha=\alpha_0)$. First, determine the matrix

$$\boldsymbol{S}=\sum_{k=1}^{K}\pi_k\boldsymbol{T}^{-1}(\boldsymbol{n}_k,\alpha_0). \tag{13}$$

Standard calculus shows that

$$\tilde\mu=\frac{\mathbf{1}'\boldsymbol{S}\boldsymbol{x}}{\mathbf{1}'\boldsymbol{S}\mathbf{1}} \tag{14}$$

and

$$\begin{aligned}\tilde s_0^2&=\frac{1}{L\sum\pi_k}(\boldsymbol{x}-\tilde\mu\mathbf{1})'\boldsymbol{S}(\boldsymbol{x}-\tilde\mu\mathbf{1})\\ &=\frac{1}{L\sum\pi_k}\left(\boldsymbol{x}'\boldsymbol{S}\boldsymbol{x}-\frac{(\mathbf{1}'\boldsymbol{S}\boldsymbol{x})^2}{\mathbf{1}'\boldsymbol{S}\mathbf{1}}\right).\end{aligned} \tag{15}$$

Simulating $\boldsymbol{n}$ with MCMC In this section we describe how to sample the states $\boldsymbol{n}$ from the probability distribution $\mathbb{P}(\boldsymbol{n}|\boldsymbol{x},\theta)$ using the Metropolis scheme. (To unburden the notation in the description of the MCMC algorithm, we drop the tilde overscript on the parameters.) At each state we will need the inverse matrix $\boldsymbol{T}^{-1}$ of $\boldsymbol{T}(\boldsymbol{n},\alpha_0)$ given by (7). Start the chain e.g. at $\boldsymbol{n}=(0,\ldots,0)$ and with $\boldsymbol{T}_0^{-1}$.

1. Let $\boldsymbol{n}$ denote the current state of the Markov chain and $\boldsymbol{T}^{-1}$ the inverse matrix of $\boldsymbol{T}(\boldsymbol{n},\alpha_0)$.

Choose an index $b=1,\ldots,B$ with equal probability (or with a probability proportional to τ_b) and an increment $\Delta n_b=+1$ or $=-1$ with probability 1/2. The candidate state $\boldsymbol{n}'$ is given by in- or decreasing the b-th index $\boldsymbol{n}$ by Δn_b: $n'_b=n_b+\Delta n_b$.

2. Using (10) and the iteration formula (8) it is not hard to check that the Hastings ratio (proposal probability) can be calculated by

$$h=\min\left[1, r^{-1/2}\left(\frac{\lambda\tau_b}{n_b^+}\right)^{\Delta n_b}\cdot\ldots \cdot\ \exp\left(\frac{\alpha\Delta n_b}{2rs_0^2}\left|(\boldsymbol{x}-\mu\mathbf{1})'\boldsymbol{T}^{-1}\boldsymbol{u}^b\right|^2\right)\right]$$

where $n_b^+=n_b+\max(0,\Delta n_b)$ and

$$r=1+\alpha\Delta n_b\cdot(\boldsymbol{u}^b)'\boldsymbol{T}^{-1}\boldsymbol{u}^b.$$

If the candidate state contains a negative component (i.e. if $n_b=0$ and $\Delta n_b=-1$) then set $h=0$. This ensures that the chain is indeed symmetric.

3. With probability h jump to the candidate state $\boldsymbol{n}'$, otherwise stay at $\boldsymbol{n}$. In the first case, update

$$\boldsymbol{n}\leftarrow\boldsymbol{n}'$$
$$\boldsymbol{T}^{-1}\leftarrow\boldsymbol{T}^{-1}-\frac{\alpha_0\Delta n_b}{r}\cdot(\boldsymbol{T}^{-1}\boldsymbol{u}^b)(\boldsymbol{T}^{-1}\boldsymbol{u}^b)'$$

and go to step 1.

No matrix inverse must ever be calculated in this scheme thanks to the update in step 3. (To counterbalance the accumulation of numerical errors it might however be wise to occasionally calculate $\boldsymbol{T}^{-1}=\boldsymbol{T}^{-1}(\boldsymbol{n},\alpha_0)$ from scratch.)

After the burn-in phase, a fraction $\boldsymbol{n}_1,\ldots,\boldsymbol{n}_M$ of the simulated states will be retained ("thinning out"). These will be used to replace the matrix (13) in the M-step of the EM algorithm by

$$\boldsymbol{S}=\frac{1}{M}\sum_{m=1}^{M}\boldsymbol{T}^{-1}(\boldsymbol{n}_m,\alpha_0). \tag{16}$$

Estimating factor α The Monte Carlo EM algorithm proposed above, while computationally highly efficient, does not allow for the estimation of the factor α. We thus use a numerical approach to iteratively approach the ML estimate of α. Specifically, we start at a value α_0 and then iteratively increase that value such that $\log_{10}\alpha_t=\log_{10}\alpha_{t-1}+\Delta_\alpha$ until the likelihood decreases. The algorithm then turns back by setting $\Delta_\alpha\leftarrow-\Delta_\alpha/e$ and proceeds again until the likelihood decreases. With every switch, the step size gets smaller and the estimate closer to the true MLE value. In each step we use the Monte Carlo EM algorithm described above to calculate the likelihood at the MLE estimates of all other parameters conditioned on that α value. In all application we set $\alpha_0=0.1$ and the initial $\Delta_\alpha=0.1$ and found estimates to be accurate within five switches.

Identifying jump locations To infer the location of jumps on a phylogenetic tree we implement an empirical Bayes approach. As is commonly done in such a setting, we assume the ML estimates $\hat{\mu},\hat{s_0^2},\hat{\lambda}$ and $\hat{\alpha}$ obtained using our Monte Carlo EM scheme are accurate and thus known constants when inferring jump locations. Under this assumption, the MCMC approach introduced above can also be used to sample configurations of jumps $\boldsymbol{n}$ from the probability distribution $\mathbb{P}(\boldsymbol{n}|\boldsymbol{x},\hat{\mu},\hat{s_0^2},\hat{\lambda},\hat{\alpha})$. This allows us to numerically infer for each branch k the posterior probabilities of $\mathbb{P}(n_k=0|\boldsymbol{x},\hat{\mu},\hat{s_0^2},\hat{\lambda},\hat{\alpha})$ and $\mathbb{P}(n_k>0|\boldsymbol{x},\hat{\mu},\hat{s_0^2},\hat{\lambda},\hat{\alpha})$, and thus to identify branches for which there is convincing evidence for an evolutionary jump.

Implementation We implemented the algorithm introduced here in C++ and optimized the code for speed. A user-friendly program to apply it to data is available at our lab website (http://www.unifr.ch/biology/research/wegmann/).

SIMULATIONS

Convergence

Convergence of the MCMC We assessed the convergence of MCMC chains by comparing parameter estimates between two independent and parallel chain runs until 10,000 jump vectors $\boldsymbol{n}$ were sampled. We run a total of 100 such chain pairs for each of two starting locations with values differing ten fold and discarded the first 100 such vectors as burn-in. We also compared two different values to thin the chains: either we sampled every 10th or every 5000th step.

Regardless of the starting values, convergence was reached rather fast but with some variation across parameters (Fig. S1). The parameter to converge fastest was μ, for which the difference in estimates was below 0.01 within 2,000 sampled jump vectors for 90% of all chain pairs. Similarly small differences for s_0^2 and λ were only reached after sampling about 4,000 jump vectors (Fig. S1). Interestingly, a larger thinning did not improve convergence, suggesting that the variance in estimates is dominated by variation in the jump vectors sampled, but not by autocorrelation along the chain. For subsequent analyses, we used a thinning of 10 and sample a total of 5,000 jump vectors.

We next assessed the convergence of the MCMC for the inference of jumps on trees by assessing the difference in posterior probabilities between independent chains (Figure S2). We again run 100 chain pairs, fixed the thinning to 10 and discarded the first 100 jump vectors as burn-in. While we found convergence to be reached within less than 2,000 iterations for branches with very low (<0.05) and very high (>0.95) posterior

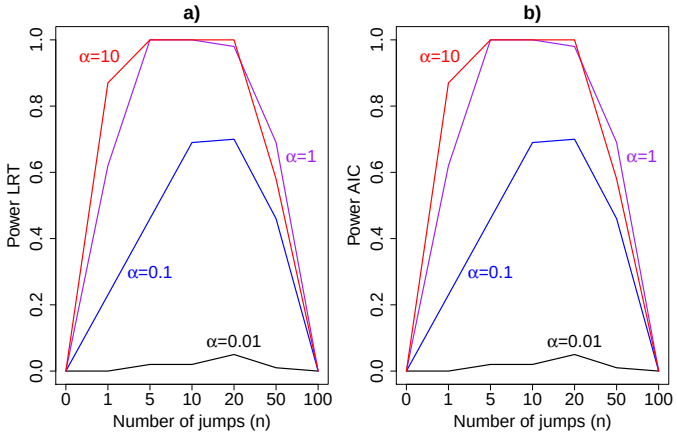

FIGURE 1. Power to reject the null model (Brownian motion) using a likelihood ratio test (LRT) (**a**), or the Akaike information criterion (AIC) (**b**) as a function of the number of simulated jumps n and the jump strengths α.

probabilities, more iterations were required for branches with intermediate posterior probabilities. We found that sampling 5,000 jump vectors gave very consistent results also for inferring the location of jumps.

Convergence of the EM for parameter inference To test if the stochastic EM algorithm converges with the MCMC settings found above (5,000 jump vectors, a burn-in of 100 such vectors and a thinning of 10), we run the EM for a wide range of parameter values for up to 100 iterations. Since the EM algorithm is stochastic, it does not converge onto a single value unless an infinitely large sample of $\boldsymbol{n}$ vectors are used. We thus first inspected obtained patterns visually and found that parameter estimates stabilized after only a few iterations, usually between 10 and 20 (Figure S3).

We then implemented two different measurements to assess convergence more formally: the first is a test statistics assessing the presence of a trend in the parameter estimates, and the second is quantifying the number of slope changes in the individual parameter updates (see Appendix).

Power to reject Brownian motion

To assess the power of our approach to identify Lévy processes and to estimate associated parameters, we run our EM algorithm on data simulated with jumps on trees of 100 leaves, each simulated using a birth-death model (Stadler, 2011) and scaled to a total length of 1. We generated 100 such simulations for many combinations of number of jumps and α values but fixed $\mu=0$ and $s_0^2=1$ since changing these parameters does not affect the inference. We then inferred the MLE estimates for all parameters under both the null model (Brownian motion) and under the alternative Lévy model.

Using both a likelihood ratio test (LRT) or the Akaike information criterion resulted in generally substantial power to reject the null model over a large range of jumps simulated and for many different values of α (Fig. 1). Unsurprisingly, power was much lower if simulated jumps were on the order of the change of the Brownian background process or lower. Here we simulated trees of length 1, and thus the average length of each of the $\sim$200 branches is roughly 0.005. Hence with $\alpha=0.01$, the strength of half of the evolutionary jumps are expected to be smaller or equal to the effect of the background process on an average branch. However, with $\alpha=0.1$, the power to reject the null model was $>80\%$ if multiple jumps were present on the tree.

Interestingly, we also found our approach to regularly fail to reject the null model if the number of jumps was very large, i.e. on the order of the number of branches (50 jumps correspond to a jump on every 4th branch). In such situations, the large variance in traits observed under the Lévy model is also perfectly explained by a pure BM model with larger variance s_0^2 (see below).

In summary, these results show that our method has considerable power to detect Lévy process as long as jumps are meaningfully strong and there are not too many jumps, in which case the Lévy and BM models become indistinguishable from each other.

Accuracy in inferring Lévy parameters

For the cases in which the Lévy model was preferred we next evaluated the power of our approach to infer the associated parameters, starting with the jump strength α. We found that our approach infers α quite accurately over the whole range, but we observed a slight overestimation for lower α values. This is a direct result of the low power to reject a model of Brownian rate at these lower jump strengths such that for simulations that resulted in larger jumps the Brownian model was more easily rejected. But the inferred values for $\alpha \leq 1$ were rarely further from the true value than a factor of 2 if multiple jumps were present (Fig. 2a), while it was unsurprisingly much harder to accurately infer the jump strength in case of a single jump.

We next evaluated the accuracy of our approach in inferring the jump rate λ, again limited to the simulations in which a Lévy model was preferred. As shown in Fig. 2b, our method inferred this parameters very accurately over a large range of jumps simulated and for all values of α, with generally higher accuracy with higher α values.

We then finally evaluated the accuracy in inferring the Brownian background rate s_0^2 (Fig. 2b) and found it to be very accurately inferred whenever the Brownian model was rejected. Interestingly, however, s_0^2 was overstimated whenever the Brownian model could not be rejected but jumps were simulated. This illustares that under certain conditions a Lévy model is indistinguishable from a model of pure Brownian motion with an elevated rate. This is particularly true in the case of weak jumps (small α) or if jumps are very common on the tree.

Jump location

We finally tested the power of our method to infer the location of jumps on the tree. For this we simulated trees with 100 leaves and trait data affected by 20 jumps randomly placed on each tree for different jump strengths α while fixing $s_0^2 = 1$. In each case we then assumed the Lévy parameters to be known and used our MCMC approach to calculate the posterior probability on there being at least one jump for each branch.

We found our method to have a very low false positive rate in identifying jumps in that a posterior probability for jumps >0.5 was never obtained for branches on which we did not simulate any jumps (Fig. 3a), and 90% of all such branches resulted in a posterior probability for jumps below 0.2 even for the weakest jump strengths simulated ($\alpha = 0.1$).

The power to infer true jumps (true positives) was also considerably high, especially for jumps of meaningful strength. For data simulated with $\alpha = 10$, for instance, 90% of all branches on which jumps were simulated resulted in a posterior probability $>50\%$, and 75% even even in a posterior probability $>95\%$. The few branches with jumps for which we did not obtain decisive posterior probabilities in favor of jumps all contained jumps that were considerably weak (Fig. 3b). Such jumps are expected even for large α values since individual jump strengths are assumed to be normally distributed around zero.

A similar pattern was observed when simulating data with smaller α, but even in the case of $\alpha = 0.1$ we obtain posterior probabilities in favor of jumps >0.5 for more than one third of the branches on which jumps were simulated (Fig. 3). At such small α values for a tree of length 1, about 40% of all jumps are expected to have a strength smaller than 10 times the effect of the Brownian process on the same branch. But we note that the difficulty in placing weak jumps did not affect the power to infer the jump rate λ, which was inferred quite accurately even at such low α values (Fig. 2).

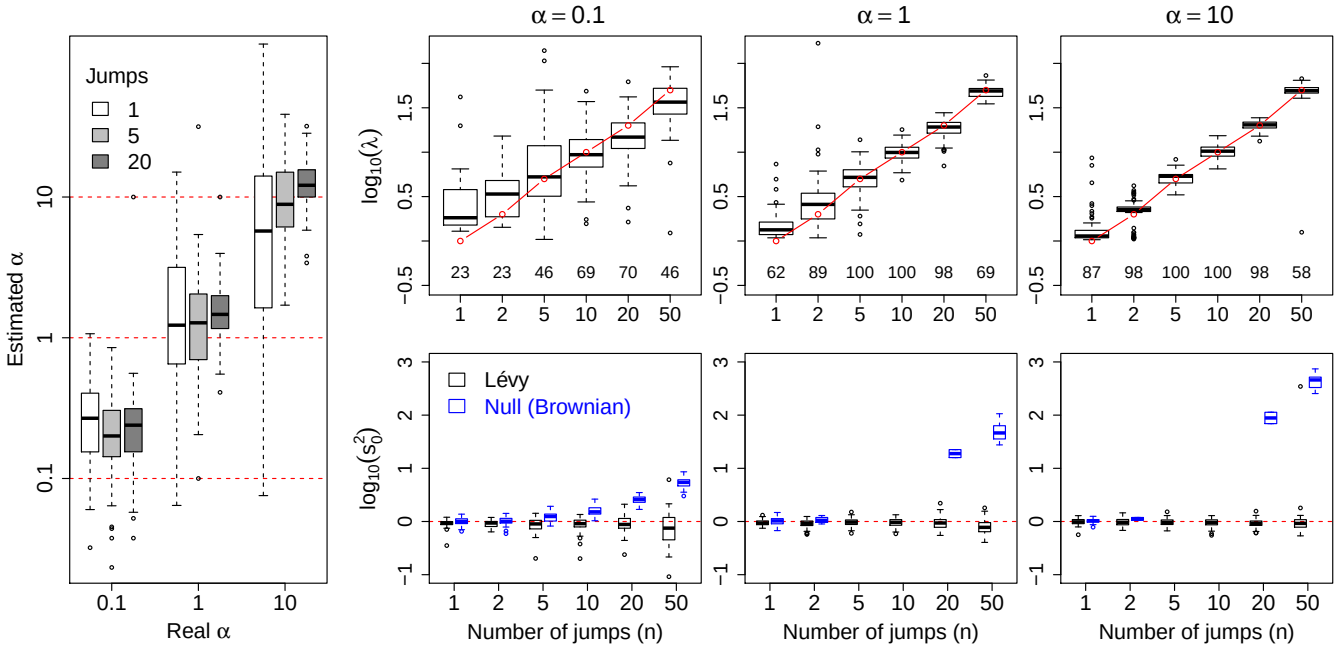

FIGURE 2. Accuracy in infering Lévy parameters. Each boxplot represents the distribution of inferred values across 100 replicates simulated as described in the text for different combinations of jump strengths α and number of simulated jumps n. **a)** Accuracy in inferring factor α. The true α values used in the simulations are indicated with red solid lines. **b)** Top row: distributions of inferred jump rate λ. Connected red open circles represent the true values. The numbers printed below the boxplots indicate the percetage of simulations for which the Brownian model was rejected and are hence included here. Bottom row: distributions of inferred Brownian background rates s_0^2 for simulations in which the Brownian null model was rejected (black) or not rejected (blue).

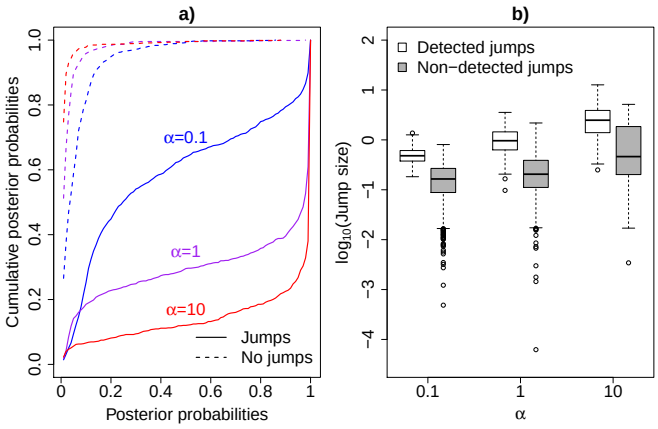

FIGURE 3. Power to detect individual jumps. **a)** Cumulative distribution of jump posterior probabilities on all branches with (solid lines) and without (dashed lines) simulated jumps. Notice that branches with jumps have posterior probabilities that accumulate at 1, whereas branches without jumps accumulate at 0. **b)** Distribution of the absolute strengths of individual jumps that were either detected (white) or not detected (gray).

APPLICATIONS

Quantum evolution in anoles

There have been a few direct tests of Simpsonian jumps between adaptive zones using empirical data (Uyeda et al., 2011). Here, we analyze "evolution by jumps" in the adaptive radiation of anoles, lizards that have adaptively radiated in the Caribbean and South America (Losos, 2009). Following previous work, we focused on anoles on the four islands of the Greater Antilles, as they provide a unique opportunity for testing Simpson's theory of adaptive zones for two reasons. First, there have been repeated dispersal events among islands in the Greater Antilles (Losos et al., 1998; Mahler et al., 2010). These dispersal events represent geographic opportunities, where anole lineages reach a new island and are no longer sympatric with the former set of competitors (Mahler et al., 2010). Second, most anole species can be classified into ecomorphs, habitat specialists that have evolved repeatedly on the four islands of the Greater Antilles (Losos et al., 1998). Transitions between ecomorph categories represent the evolution of key characters in anole lineages that allow them to invade novel habitats (see Losos (2009) for a review).

Anoles have thus repeatedly experienced two conditions under which Simpson expected evolutionary jumps to be observed: dispersal into new geographic areas and the appearance of evolutionary novelties. Importantly, both ecomorph origins and transitions among islands are replicated in the phylogeny of anoles, but are still rare enough that we can estimate the position of transitions on the phylogenetic tree with some confidence (Huelsenbeck et al., 2003; Schluter, 1995).

With this background in mind we tested if a model with evolutionary jumps fits the evolution of body size in anoles better than pure Brownian motion, and if jumps correspond with either of the two factors postulated by Simpson: evolution of key characters and/or geographic dispersal. To address this question, we made use of a recent time-calibrated phylogeny of 170 *Anolis* lizards (Thomas et al., 2009) and analyzed snout-to-vent length (SVL), a standard phenotypic measurement of body size in lizards. This trait is broadly correlated with habitat partitioning in Greater Antillean anoles and represent the primary axes of ecologically driven evolutionary divergence in lizards (Beuttell and Losos, 1999; Losos, 2009; Schoener, 1970). We made use of the sex-specific data of SVL from Thomas et al. (2009) and inferred evolutionary parameters independently for females and males, but excluded five species that lacked information on SVL for one or both sexes (*Anolis darlingtoni*, *A. guamuhaya*, *A. loveridgei*, *A. oporinus*, and *A. polyrhachis*).

We found that the Lévy jump model is preferred over a strict BM model in females, but not in males (Table 1). Evolutionary jumps indicating rapid body size evolution (Figure 4) were found precisely at the basis of the clade comprising the ecomorph "crown giants" Thomas et al. (2009), in which females exhibit particularly large body sizes. The large sexual size dimorphism of this group (Harmon et al., 2005) is also likely explaining why the BM model fits the evolution of male body sizes well. In addition to the clades of crown giants, we also identify evolutionary jumps at the basis of the clade consisting of the species *A. barbatus*, *A. porcus*, and *A. chamaeleonides*. These species, which are known as "false chamaleons" and are part of the former genus *Chameleolis* have been called the "most bizarre West Indian lizards" (Leal and Losos, 2000).

Our analyses support two main conclusions. First, evolutionary change in female anoles is not well described by a uniform random walk. A better description of anole evolution combines a uniform component of change that is punctuated by rapid jumps in trait values. Second, these jumps in body size very well correspond to ecological transitions to novel ecomorphs. The evolution

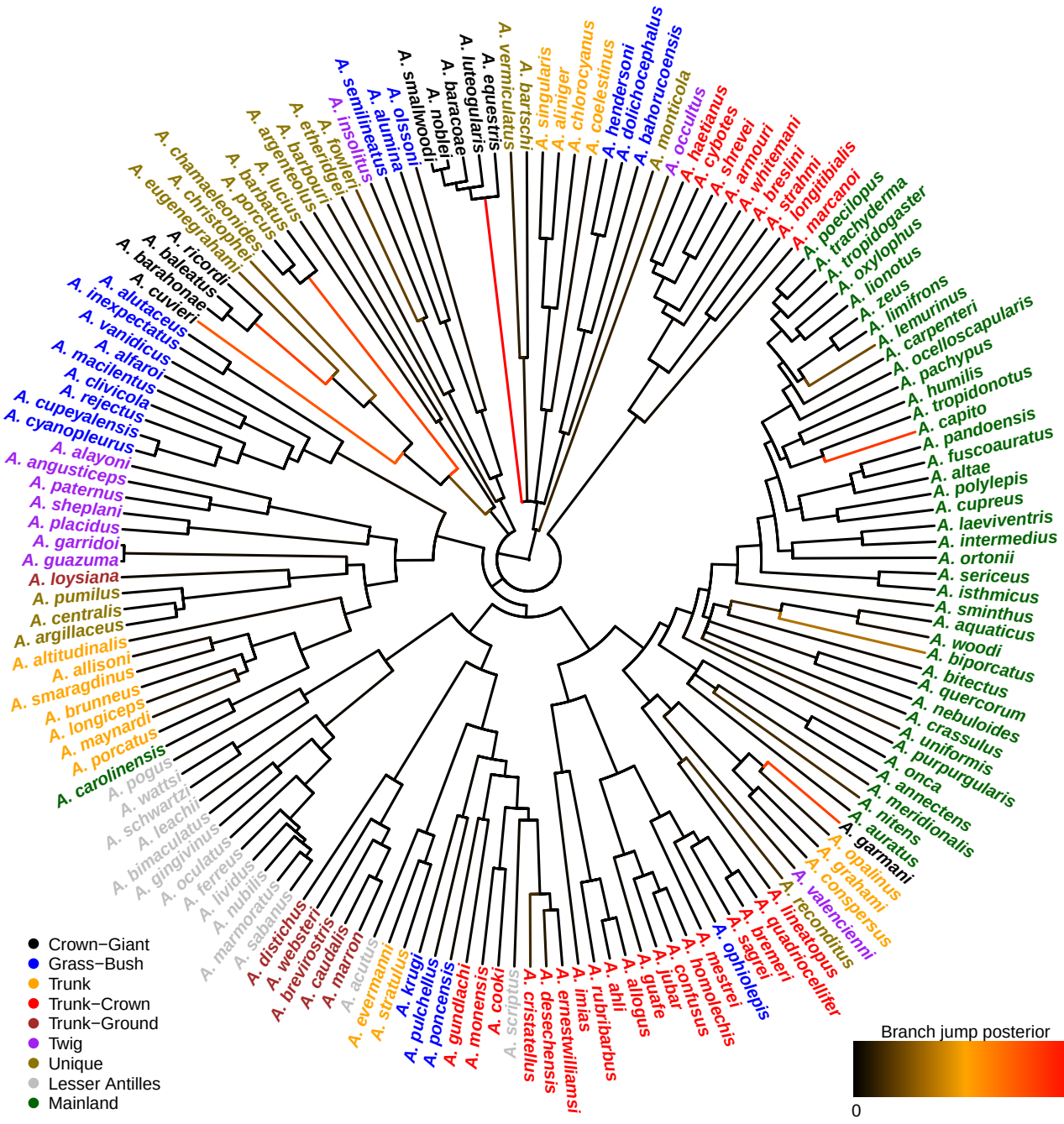

FIGURE 4. Inferred jumps for female body-size evolution on the anoles' tree. The trait measured was the snout-to-vent length (SVL). Branches are colored according to their inferred jump posterior probability (black to red scale going from posterior probability 0 to 1, respectively). Tips are colored according to ecomorphs as defined by Thomas et al. (2009).

of this trait is thus consistent with Simpson's description of evolutionary jumps associated with the entry into new adaptive zones. The fact that we did not find such jumps at the basis of clades of other ecomorphs suggests that body size was not a trait strongly contributing to the ecological transition of those. However, evolutionary jumps might well be found at the basis of those clades when focusing on more relevant traits.

Nectarivory evolution in Loriinae

The Australasian lories belong to the tribe Loriini (Joseph et al., 2012) and are extremely species rich (Schweizer et al., 2011). Their digestive tract is highly adapted to a nectarivorous diet (Güntert, 2012) and Schweizer et al. (2014) has shown quantitatively that a switch in diet to nectarivory might be considered an evolutionary novelty that created an ecological opportunity for species proliferation through

TABLE 1. Inferred Lévy parameters for *Anolis* and Loriinae along with the log-likelihood (ℓ) obtained under the Lévy and BM models and the p-value of a likelihood ratio test (LRT) contrasting these.

	Anolis		Loriini			
	$\log SVL_f$	$\log SVL_m$	log Wgt	PC1	PC2	PC3
μ	3.93	4.18	5.33	-0.062	0.051	0.19
s_0^2	5.06	10.34	44.40	7.68	7.35	5.80
α	0.11	-	-	0.16	0.093	-
λ	11.27	-	-	11.95	7.38	-
ℓ_{BM}	5.03	-15.19	-85.54	-65.00	-60.63	-6.22
$\ell_{Lévy}$	26.61	-13.89	-85.51	-31.99	-24.72	-4.97
LRT p	$4.3 \cdot 10^{-10}$	0.28	0.97	$4.7 \cdot 10^{-15}$	$2.3 \cdot 10^{-16}$	0.29
Preferred model	Lévy	BM	BM	Lévy	Lévy	BM

allopatric partitioning of the same new niche. Using the methodology developed above we now tested if the evolution of the morphology of the digestive tract in parrots as a whole is better characterized by a model of evolutionary jumps or Brownian motion. For this we made use of data from Schweizer et al. (2014), to generate a time-calibrated phylogeny of 78 parrot species using BEAST (Drummond and Rambaut, 2007) implementing a secondary calibration point from Schweizer et al. (2011) for the initial split within parrots. The following 13 measurements of ~~gut morphology~~ were used: the length of intestine, length of esophagus, extension of esophagus glands, length of intermediate zone, length of proventriculus, gizzard height, gizzard width, gizzard depth, maximum gizzard height at main muscles, gizzard thickness at main muscles, gizzard lumen width including koilin layer, gizzard width at the caudoventral thin muscle, maximum gizzard height at the thin muscle, and the maximum gizzard lumen at the thin muscle. Since many of the ~~gut morphology characters~~ considered are both highly correlated with body size as well as among themselves, we first regressed out body mass (Wgt) from each ~~gut morphology~~ trait and then summarized the residuals of all traits using the first three principal components (PCA; see also (Revell, 2009)).

We found that the evolution of both body weight (Wgt) and the first PC axis of ~~gut~~ morphology were much better explained by a model of evolutionary jumps ($p < 10^{-16}$ in both cases) with relatively high rates of jumps (Table 1). Overall, the jumps for PC1 identified with strongest support are both on branches basal to clades of nectarivorous species, particularly at the base of highly specialized nectar feeding Loriini, but also at the base of the genus *Loriculus* (Figure 5). As postulated by Simpson the niche shift to nectarivory especially in Loriini involved a period of rapid evolution reflecting adaptations to feed effectively on nectar (and pollen) (Schweizer et al., 2014). While the jumps within the Neotropical parrots are difficult to interpret in biological terms, the shift along the branch leading to *Psittrichas fulgidus* might be explained by its gizzard morphology similar to that of the Loriini probably reflecting an adaptation to its reportedly mainly frugivorous diet (Schweizer et al., 2014). Some special structures in the digestive tract of the genus *Nestor* have been described in Güntert (2012).

DISCUSSION

While many traits appear to evolve at relatively constant rates over long time periods and across many taxa, some traits seem to undergo periods of rather rapid evolution (see Arnold, 2014). Simpson (1944) postulated that such evolutionary jumps are triggered by a change in selection pressure after lineages transitioned into different adaptive zones, for instance by dispersing into new geographic areas, after the appearance of evolutionary novelties, key innovations, or after rapid climatic or ecological changes of the environment. The appearance of well calibrated phylogenies along with recent statistical developments now allow to test such models on a wide variety of data.

Bokma (2008), for instance, proposed to model evolutionary jumps as a compound process of a continuous background process and a discrete jump process. Recently, Landis et al. (2013) introduced a general framework to infer parameters of such Lévy processes under a Bayesian framework by means of Markov Chain Monte Carlo (MCMC). Unfortunately this approach, while elegant, requires the calculation of the inverse of the variance-covariance matrix describing the correlations between traits as a function of the phylogenetic tree and the jump process, which is computationally prohibitive for large trees.

Here we introduce a computationally highly efficient variant of this approach that naturally scales to large trees. The basis of our approach is an MCMC algorithm in which we can update the inverse of the above mentioned variance-covariance matrix directly without inversion when sampling jump configurations with fixed hierarchical parameters (root state, Brownian rate, jump strength and jump rate). To make use of this development for inference we propose a two-step approach in which the MCMC algorithm is embedded into an Expectation-Maximization (EM) approach to obtain maximum likelihood (ML) estimates of the hierarchical parameters while integrating over jump

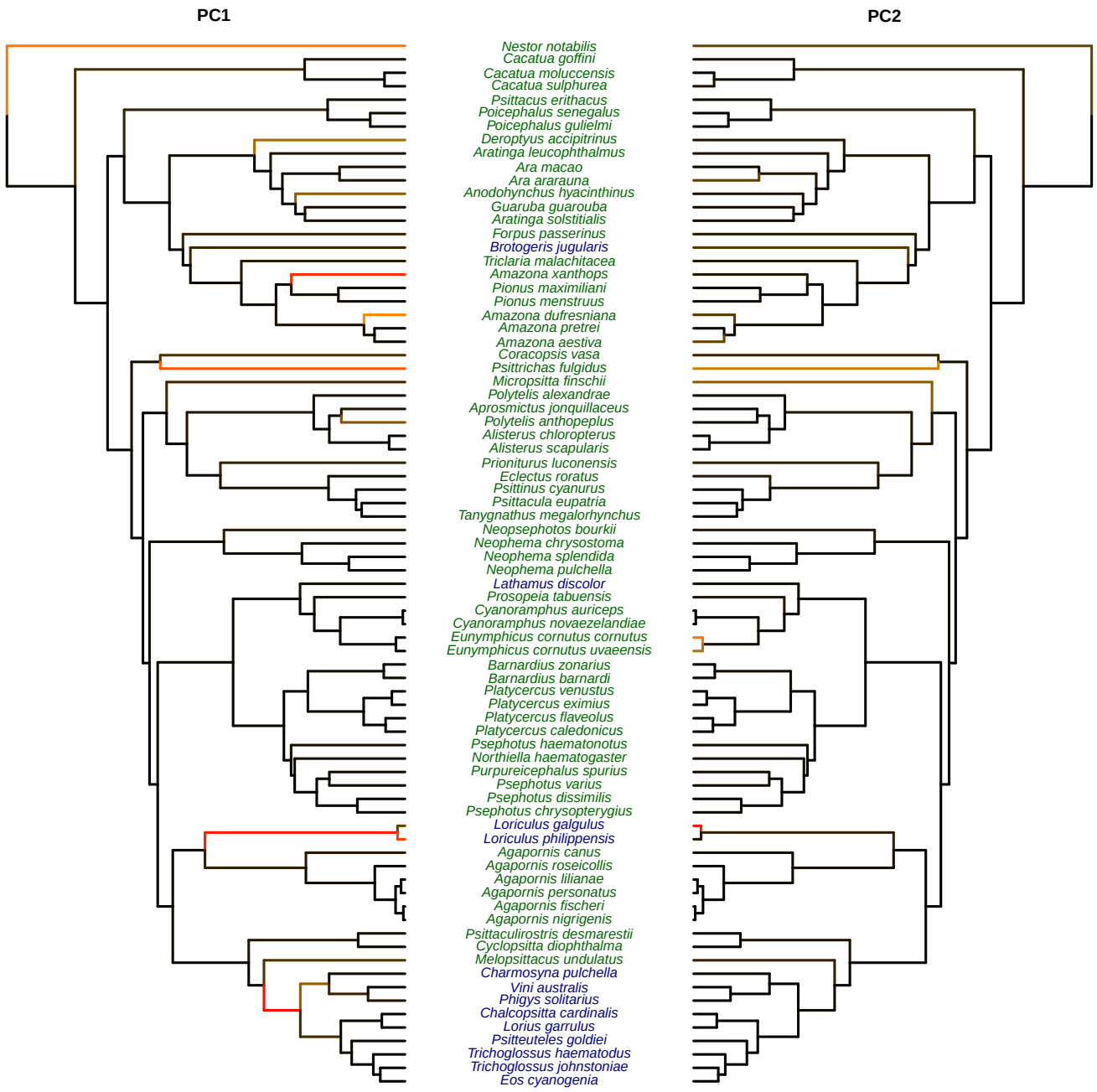

FIGURE 5. Inferred jumps for nectarivory evolution on the Loriinae tree. Results for PC1 are shown on the left phylogeny, and results for PC2 on the right phylogeny. Branches are colored according to their inferred jump posterior probability (black to red scale going from posterior probability 0 to 1, respectively). Species names are colored according to their diet: nectarivorous (blue) and non-nectarivorous (green).

configurations. In a second step, the location of jumps can then be inferred under an empirical Bayes framework in which the hierarchical parameters are fixed to their ML estimate and the developed MCMC algorithm is used to obtain for each branch the posterior probability that a jump occurred at this location.

There are also other methods that deal with the burden of calculating inverses and determinants of variance-covariance matrices. For instance, Freckleton (2012) applied the results of Felsenstein (1973) and Felsenstein (1985) to calculate the likelihood in linear time of a BM model. FitzJohn (2012) also proposed a fast algorithm to calculate BM and OU likelihoods using Gaussian elimination, but this is not applicable to non-gaussian traits. Ho and Ané (2014) proposed a new method, which efficiently calculates likelihoods by

avoiding the calculation of the inverse and determinant of the variance-covariance matrix. Their method requires that this matrix belongs to a class of generalized 3-point structured matrices. Our method, which applies an iterative scheme, differs from the others in the sense that the inverse and determinant of the variance-covariance matrix has to be calculated only once when obtaining the likelihoods, thus obtaining rather fast calculation times.

We demonstrated the applicability of our approach by identifying evolutionary jumps for body size evolution in *Anolis* lizards and the evolution of ~~gut morphology~~, in Australasian lories of the ~~subfamily Loriinae~~. We found strong support for evolutionary jumps in both systems that provide direct support for Simpson's quantum evolutionary hypothesis of adaptive zones. Among the anoles, for instance, we identified evolutionary jumps on the basal lineage leading to crown giants, a group of lizards that transitioned into a novel niche for hunting: the crowns of large tropical trees. Similarly, we identified jumps at the basis of clades of lories that transitioned to nectarivory, an evolutionary novelty that triggered rapid changes in morphology of the digestive system and promoted significant lineage diversification, which was probably mainly non-adaptive after the basal diet shift through allopatric partitioning of the same niche (Schweizer et al., 2014, cf.).

These results also show that the distinction between "gradual" and "punctuated" models of evolution is a false dichotomy; instead, evolution has a gradual component that may be frequently punctuated by periods of rapid change (Levinton, 2001). We further note that in both cases studied here a single jump at the basis of clades is sufficient to explain their trait data, suggesting that the period of rapid evolution was limited to a single branch and that the background rate remained constant. We suggest that future work should follow Simpson's lead and focus on the factors that promote these pulses of evolutionary change.

Although we model evolutionary jumps as instantaneous, we want to be clear that we are not invoking actual instantaneous evolutionary change (e.g. "hopeful monsters") (Charlesworth et al., 1982; Goldschmidt, 1940). Typical microevolutionary processes of selection and drift can cause change that would appear to be instantaneous when viewed over the timescale of macroevolution. Our model is also distinct from punctuated equilibrium, which requires evolutionary jumps to occur only at speciation events (Eldredge and Gould, 1972). The punctuated changes in our model occur along branches in the tree and are not necessarily associated with speciation events. In fact, for the case of anoles, two lines of evidence argue against punctuated equilibrium: first, most speciation events in the tree are not associated with jumps; and second, we know from detailed microevolutionary studies that anole body size can evolve rapidly in response to selection even in the absence of speciation (e.g. Losos et al. (2006)).

FUNDING

This study was supported by Swiss National Foundation grant 31003A_149920 to DW. The work of S. M. Szilágyi was supported by the János Bolyai Fellowship Program of the Hungarian Academy of Sciences.

ACKNOWLEDGMENTS

REFERENCES

Anderson, T. 2003. Multivariate statistical analysis. John Wiley .

Arnold, S. J. 2014. Phenotypic evolution: the ongoing synthesis. The American Naturalist 183:729–746.

Bedford, T. and D. L. Hartl. 2009. Optimization of gene expression by natural selection. Proceedings of the National Academy of Sciences 106:1133–1138.

Beuttell, K. and J. B. Losos. 1999. Ecological morphology of caribbean anoles. Herpetological Monographs Pages 1–28.

Bokma, F. 2008. Detection of punctuated equilibrium by bayesian estimation of speciation and extinction rates, ancestral character states, and rates of anagenetic and cladogenetic evolution on a molecular phylogeny. Evolution 62:2718–2726.

Brawand, D., M. Soumillon, A. Necsulea, P. Julien, G. Csárdi, P. Harrigan, M. Weier, A. Liechti, A. Aximu-Petri, M. Kircher, et al. 2011. The evolution of gene expression levels in mammalian organs. Nature 478:343–348.

Butler, M. A. and A. A. King. 2004. Phylogenetic comparative analysis: a modeling approach for adaptive evolution. The American Naturalist 164:683–695.

Cavalli-Sforza, L. L. and A. W. Edwards. 1967. Phylogenetic analysis. models and estimation procedures. American journal of human genetics 19:233.

Charlesworth, B., R. Lande, and M. Slatkin. 1982. A neo-darwinian commentary on macroevolution. Evolution 36:474–498.

Davidson, R. and J. G. MacKinnon. 2004. Econometric theory and methods vol. 5. Oxford University Press New York.

Drummond, A. J. and A. Rambaut. 2007. Beast: Bayesian evolutionary analysis by sampling trees. BMC evolutionary biology 7:214.

Eastman, J. M., M. E. Alfaro, P. Joyce, A. L. Hipp, and L. J. Harmon. 2011. A novel comparative method for identifying shifts in the rate of character evolution on trees. Evolution 65:3578–3589.

Edwards, A., L. Cavalli-Sforza, and V. Heywood. 1964. Phenetic and phylogenetic classification. Systematics As-soc. Publ Page 67.

Eldredge, N. and S. J. Gould. 1972. Punctuated equilibria: an alternative to phyletic gradualism. Models in paleobiology 82.

Felsenstein, J. 1973. Maximum-likelihood estimation of evolutionary trees from continuous characters. American journal of human genetics 25:471.

Felsenstein, J. 1985. Phylogenies and the comparative method. American Naturalist Pages 1–15.

FitzJohn, R. G. 2012. Diversitree: comparative phylogenetic analyses of diversification in r. Methods in Ecology and Evolution 3:1084–1092.

Freckleton, R. P. 2012. Fast likelihood calculations for comparative analyses. Methods in Ecology and Evolution 3:940–947.

Freckleton, R. P., P. H. Harvey, and M. Pagel. 2002. Phylogenetic analysis and comparative data: a test and review of evidence. The American Naturalist 160:712–726.

Goldschmidt, R. 1940. The material basis of evolution vol. 28. Yale University Press.

Güntert, M. 2012. Morphologische Untersuchungen zur adaptiven Radiation des Verdauungstraktes bei Papageien (Psittaci). Zoologische Jahrbucher. Abteilung für Anatomie und Ontogenie der Tiere 106:471–526.

Hansen, T. F. 1997. Stabilizing selection and the comparative analysis of adaptation. Evolution Pages 1341–1351.

Harmon, L. J., J. J. Kolbe, J. M. Cheverud, and J. B. Losos. 2005. Convergence and the multidimensional niche. Evolution 59:409–421.

Ho, T. and C. Ané. 2014. A linear-time algorithm for gaussian and non-gaussian trait evolution models. Systematic biology 63:397–408.

Huelsenbeck, J. P., R. Nielsen, and J. P. Bollback. 2003. Stochastic mapping of morphological characters. Systematic Biology 52:131–158.

Ingram, T. and D. L. Mahler. 2013. SURFACE: detecting convergent evolution from comparative data by fitting Ornstein-Uhlenbeck models with stepwise Akaike Information Criterion. Methods in Ecology and Evolution 4:416–425.

Izenman, A. 2008. Modern multivariate statistical techniques vol. 1. Springer.

Joseph, L., A. Toon, E. E. Schirtzinger, T. F. Wright, and R. Schodde. 2012. A revised nomenclature and classification for family-group taxa of parrots (Psittaciformes). Zootaxa 3205.

Landis, M. J., J. G. Schraiber, and M. Liang. 2013. Phylogenetic analysis using lévy processes: finding jumps in the evolution of continuous traits. Systematic biology 62:193–204.

Leal, M. and J. B. Losos. 2000. Behavior and ecology of the Cuban "chipojos bobos" *Chamaeleolis barbatus* and *C. porcus*. Journal of Herpetology 34:318–322.

Levinton, J. S. 2001. Genetics, paleontology, and macroevolution. Cambridge University Press.

Losos, J. B. 2009. Lizards in an evolutionary tree: ecology and adaptive radiation of anoles vol. 10. Univ of California Press.

Losos, J. B. 2010. Adaptive radiation, ecological opportunity, and evolutionary determinism. The American Naturalist 175:623–639.

Losos, J. B., R. E. Glor, J. J. Kolbe, and K. Nicholson. 2006. Adaptation, speciation, and convergence: A hierarchical analysis of adaptive radiation in caribbean *Anolis* lizards 1. Annals of the Missouri Botanical Garden 93:24–33.

Losos, J. B., T. R. Jackman, A. Larson, K. de Queiroz, and L. Rodrıguez-Schettino. 1998. Contingency and determinism in replicated adaptive radiations of island lizards. Science 279:2115–2118.

Mahler, D. L., L. J. Revell, R. E. Glor, and J. B. Losos. 2010. Ecological opportunity and the rate of morphological evolution in the diversification of greater antillean anoles. Evolution 64:2731–2745.

O'Meara, B. C., C. Ané, M. J. Sanderson, and P. C. Wainwright. 2006. Testing for different rates of continuous trait evolution using likelihood. Evolution 60:922–933.

Revell, L. J. 2009. Size-correction and principal components for interspecific comparative studies. Evolution 63:3258–3268.

Rohlfs, R. V., P. Harrigan, and R. Nielsen. 2013. Modeling gene expression evolution with an extended ornstein-uhlenbeck process accounting for within-species variation. Molecular biology and evolution Page mst190.

Schluter, D. 1995. Uncertainty in ancient phylogenies. Nature 377:108–110.

Schoener, T. W. 1970. Size patterns in west indian anolis lizards. ii. correlations with the sizes of particular sympatric species-displacement and convergence. American Naturalist Pages 155–174.

Schweizer, M., M. Güntert, O. Seehausen, C. Leuenberger, and S. T. Hertwig. 2014. Parallel adaptations to nectarivory in parrots, key innovations and the diversification of the loriinae. Ecology and evolution 4:2867–2883.

Schweizer, M., O. Seehausen, and S. T. Hertwig. 2011. Macroevolutionary patterns in the diversification of parrots: effects of climate change, geological events and key innovations. Journal of Biogeography 38:2176–2194.

Simpson, G. 1944. Tempo and Mode in Evolution. A Wartime book Columbia University Press.

Slater, G. J., L. J. Harmon, D. Wegmann, P. Joyce, L. J. Revell, and M. E. Alfaro. 2012. Fitting models of continuous trait evolution to incompletely sampled comparative data using approximate bayesian computation. Evolution 66:752–762.

Stadler, T. 2011. Simulating Trees with a Fixed Number of Extant Species. Systematic biology 60:676–684.

Thomas, G. H., S. Meiri, and A. B. Phillimore. 2009. Body size diversification in *Anolis*: novel environment and island effects. Evolution 63:2017–2030.

Tung Ho, L. S. and C. Ané. 2014. A linear-time algorithm for gaussian and non-gaussian trait evolution models. Systematic Biology 63:397–408.

Uyeda, J. C., T. F. Hansen, S. J. Arnold, and J. Pienaar. 2011. The million-year wait for macroevolutionary bursts. Proceedings of the National Academy of Sciences 108:15908–15913.

Uyeda, J. C. and L. J. Harmon. 2014. A novel Bayesian method for inferring and interpreting the dynamics of adaptive landscapes from phylogenetic comparative data. Systematic biology 63:902–918.

APPENDIX

Conditional likelihood

The EM algorithm can be implemented without the MC part if we impose a condition $|\boldsymbol{\nu}| \leq R$ on the likelihood, i.e. if we suppose a priori that there have been only R or less Poisson events on the tree $\mathcal{T}$. In that case, the sum in (9) is over all $\boldsymbol{n}_k$ such that $|\boldsymbol{n}_k| \leq R$. Observe that we have to use the conditional probabilities

$$\mathbb{P}(\boldsymbol{\nu} = \boldsymbol{n}_i | \boldsymbol{\nu} \leq N) = \mathbb{P}(\boldsymbol{\nu} = \boldsymbol{n}_i) / \sum_{k=1}^{K} \mathbb{P}(\boldsymbol{\nu} = \boldsymbol{n}_k).$$

The new Q-function is

$$Q(\theta) = \sum_{k=1}^{K} \pi_k \log[\phi(\boldsymbol{x}|\boldsymbol{n}_k)\mathbb{P}(\boldsymbol{\nu} = \boldsymbol{n}_k | \boldsymbol{\nu} \leq R)] \tag{17}$$

where we can use

$$\pi_k = \phi(\boldsymbol{x}|\boldsymbol{n}_k, \tilde{\mu}, \tilde{s}_0^2, \tilde{\alpha})\mathbb{P}(\boldsymbol{\nu} = \boldsymbol{n}_k | \tilde{\lambda}).$$

In the conditional case, there no longer seems to exists a closed formula like (12) for the optimal $\tilde{\lambda}$. Setting the derivative of (17) w.r.t. λ equal to 0, one can show that $\tilde{\lambda}$ is the root of the following R-th order polynomial:

$$P(\lambda) = \sum_{r=0}^{R} \frac{\tau^r}{r!} \left(\sum \pi_k |\boldsymbol{n}_k| - r \sum \pi_k \right) \lambda^r,$$

i.e. $P(\tilde{\lambda}) = 0$. The estimation of $\tilde{\mu}$ and $\tilde{s}_0$, on the other hand, remains exactly as explained in section 1.4.

Assessing convergence of the EM

We introduce two measures to assess convergence of the Monte Carlo EM algorithm.

Regression criterion We consider a time series $y_1, \ldots, y_n$ and construct a test statistic which allows to reject the null hypothesis that the time series exhibits no trend. For this we estimate the slope $\hat{\beta}$ of the regression line passing through the data points

$$(1, y_1), (2, y_2), \ldots, (n, y_n)$$

and test for the null hypothesis $\beta = 0$ (no trend). Determine the following quantities:

$$\bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i, \quad S_{xx} = \frac{n(n^2-1)}{12}, \quad S_{xy} = -\frac{1}{2} \sum_{i=1}^{n} (n+1-2i) y_i,$$

$$\hat{\beta} = \frac{S_{xy}}{S_{xx}}, \quad \hat{\sigma}^2 = \frac{1}{n-2} \left((\sum_{i=1}^{n} y_i^2) - n\bar{y}^2 - \hat{\beta} S_{xy} \right), \quad se(\hat{\beta}) = \sqrt{\frac{\hat{\sigma}^2}{S_{xx}}}.$$

The test statistic

$$T = \frac{\hat{\beta}}{se(\hat{\beta})}$$

has the Student t-distribution with $n-2$ degrees of freedom. We reject the null hypothesis on the level γ if $|T| \geq t_{\gamma/2, n-2}$. A good rule of thumb (for γ roughly 5% and $n > 15$) is $|T| \geq 2$.

Proportion of slope sign changes We propose a second way of assessing convergence by taking the last n values of the EM algorithm and counting the number of times c there is a change in the sign of the slope between consecutive values. If convergence is reached, we expect the number of slopes with a positive sign to be similar to the number of slopes with a negative sign. We report the test statistic N

$$N = \frac{c}{n-2}$$

where $n-2$ represents the total number of possible sign changes among the last n values.