Evolutionary Ecology Research, 2017, **18**: 41–59

Bayesian estimation of multiple clade competition from fossil data

Daniele Silvestro[1,2,3], Mathias M. Pires[4], Tiago B. Quental[4] and Nicolas Salamin[2,3]

[1]*Department of Biological and Environmental Sciences, University of Gothenburg, Gothenburg, Sweden,*
[2]*Department of Ecology and Evolution, University of Lausanne, Lausanne, Switzerland,*
[3]*Swiss Institute of Bioinformatics, Lausanne, Switzerland and*
[4]*Department of Ecology, University of São Paulo, São Paulo, SP, Brazil*

ABSTRACT

Background: The diversification dynamics of clades is governed by speciation and extinction processes and is likely affected by multiple biotic, abiotic, and stochastic factors. Using quantitative methods to analyse fossil occurrence data, one may infer rates of speciation and extinction in a Bayesian framework. Moreover, Silvestro *et al.* (2015a) recently developed a Multiple Clade Diversity Dependence birth–death model (MCDD) to determine whether diversification dynamics can be explained by positive or negative effects of interactions within or between co-existing clades. However, the power and accuracy of this model and its general applicability have yet to be thoroughly investigated.

Aims: Explore the properties of the existing MCDD implementation, which is based on Bayesian variable selection. Introduce an alternative parameterization based on the Horseshoe prior and show the properties of this approach for Bayesian shrinkage in complex models. Test the ability of the model to correctly identify within and between diversification interference under a suite of different diversification scenarios.

Methods: Use simulations to assess and compare the power and accuracy of the two algorithms.

Results: Diversity dependence within and between clades can be inferred with confidence in a wide range of scenarios using the MCDD model. The two implementations provide comparable results, but the new Horseshoe prior estimator appears to be more reliable, albeit slightly more conservative. The MCDD model is a powerful framework to analyse the putative effects of ecological interactions on macroevolutionary dynamics using fossil data and provides a sound statistical basis for future method developments.

Keywords: Bayesian shrinkage, diversity dependence, fossils, macroevolution.

INTRODUCTION

Biodiversity patterns are largely influenced by ecological interactions at the local and regional scales. Although this is well accepted in the environmental sciences in general, the role of biotic interactions in shaping diversification over longer temporal and spatial scales

Correspondence: D. Silvestro, Department of Biological and Environmental Sciences, University of Gothenburg, 40530 Gothenburg, Sweden. email: daniele.silvestro@bioenv.gu.se

is often questioned (Benton, 1996, 2009). There has been renewed interest recently in the role of biotic components in determining changes in biodiversity over evolutionary and geologic time scales (Quental and Marshall, 2013; Barraclough, 2015; Liow *et al.*, 2015; Voje *et al.*, 2015; Ezard *et al.*, 2016). Work on the macroevolutionary dynamics of different groups, including mammalian carnivores (Silvestro *et al.*, 2015a) and marine invertebrates (Castiglione *et al.*, 2017), suggests that ecological interactions among different lineages can shape diversification by affecting the rates of origination and extinction. When many ecologically similar taxa co-occur in a resource-limited environment, populations ought to be smaller, increasing the risk of extinction and decreasing the likelihood that a newly originating taxon establishes. This is the principle that links evolutionary radiations to the availability of ecological opportunities (Gavrilets and Losos, 2009) and underlies the notion that we should expect the net diversification to decrease as diversity builds up (Jablonski, 2008). Although finding evidence of interactions between extinct species in the fossil record is challenging (if not impossible) for most groups, the rise and fall of clades through time may allow us to infer the probability that the diversification of one lineage interfered with the diversification of other lineages that overlap in space and time (Sepkoski, 1996b; Van Valkenburgh, 1999). The main challenge palaeoecologists and evolutionary biologists have to face when studying the effect of lineages on each other's diversification is to uncover the relationships between lineages while taking into account the sampling incompleteness inherent to the fossil record. Further difficulties are linked to the uncertainty in dating fossil occurrences.

A recently developed method, PyRate (Silvestro *et al.*, 2014a), allows speciation and extinction rates to be estimated from the fossil record under a probabilistic analytical framework that takes into account fossil preservation (Silvestro *et al.*, 2014b). The introduction of the PyRate probabilistic framework opened the possibility for a range of new analyses and reconstructions of macroevolutionary processes using palaeontological data. For instance, this method can reveal speciation and extinction dynamics driving the rise and decline of entire clades (Pires *et al.*, 2015), assess correlations between the evolution of a continuous trait (e.g. body size) and changes in diversification rates (Silvestro *et al.*, 2014a), and infer the strength and selectivity of mass extinctions through time (Silvestro *et al.*, 2015b).

Recent methodological developments expanded the suite of diversification models available in PyRate by introducing birth–death models with diversity dependence in an attempt to determine whether evidence of competition could be detected from fossil data (Silvestro *et al.*, 2015a). Evidence of diversity dependence (Alroy, 1996, 2008; Ezard *et al.*, 2011; Liow and Finarelli, 2014; Silvestro *et al.*, 2015a) has been interpreted as the result of ecological mechanisms that could either limit the opportunity for speciation or increase the chance of extinction (Rosenzweig, 1975; Sepkoski, 1978; Levinton, 1979; Walker and Valentine, 1984; Yoder *et al.*, 2010). It is important to note that a simple decline in speciation rate, albeit congruent with, is not enough evidence in favour of mechanisms of ecological saturation (Cornell, 2013; Moen and Morlon, 2014; Marshall and Quental, 2016). Conversely, diversity trajectories other than a stationary number of species can also be explained by diversity dependence mechanisms (Cornell, 2013; Quental and Marshall, 2013; Rabosky, 2013; Marshall and Quental, 2016). Therefore, effective discrimination of diversity dependence from simple decreases in diversification rates may be difficult from the analysis of phylogenies of extant taxa (Bokma, 2009; Marshall and Quental, 2016).

Under the typical diversity dependence scenario (Etienne *et al.*, 2012), the net diversification of a clade declines as its standing diversity increases. Going beyond the assumption that competitive effects only take place within clades (typical in phylogenetic comparative methods), a recently described extension of the PyRate framework allows testing of

whether the diversification dynamics of a lineage respond to the interference, either by negative (e.g. competitive interactions) or positive effects (e.g. co-evolution and diversification of host and parasites or symbiotic organisms), from its own standing diversity or the diversity of co-occurring clades (Silvestro *et al.*, 2015a). This model, named Multiple Clade Diversity Dependence (MCDD), assumes that speciation and extinction rates of a clade can vary through time as a function of diversity changes. It was initially inspired by the possibility that a given clade could actively drive another to extinction (Sepkoski, 1996a) or to infer the effect of an incumbent lineage on the diversification dynamics of younger clades (Rosenzweig and McCord, 1991; Benton, 1996). Given the potential, but yet disputed, significance of such macroevolutionary scenarios, a statistical model that could properly detect such mechanisms is desirable. The original proposition of this method was tested through simulations focusing on the frequency of false positives (Silvestro *et al.*, 2015a). However, these simulations were limited in that they were designed to reflect the nature and size of the empirical data set of eight clades analysed in that study, but lacked a more general and thorough assessment of the method on a broad range of datasets. Furthermore, the power of the method to recover true diversity dependence effects was not addressed due to difficulties in simulating data under the assumptions of the MCDD model.

Here, we present a broad set of simulations aiming to address the properties of the MCDD model in a general context. We explore the ability of the MCDD model to correctly identify diversity dependence effects and interference within and between lineages on diversification dynamics. Furthermore, we propose a novel implementation of the MCDD model under a different parameterization using a 'Horseshoe prior' [HSP (Carvalho *et al.*, 2010)] on the diversity dependence parameters as an alternative to the Bayesian variable selection (BVS) algorithm proposed in the original implementation. We demonstrate that the MCDD model, and in particular its new implementation based on HSP, can robustly infer diversity dependence within and between clades from fossil data in a wide spectrum of diversification scenarios.

METHODS

The birth–death model

Let us consider a set of N fossil species, each with a lifespan defined by its time of speciation s and time of extinction e. In our notation, the ages of all events are measured as time before the present, so that $s > e$ and extant species are indicated with $e = 0$. Previous work has shown that times of speciation and extinction of taxa sampled in the palaeontological record can be estimated based on fossil occurrence data by modelling the preservation process (Silvestro *et al.*, 2014b). The temporal distribution of the times of speciation and extinction ($\mathbf{s} = \{s_1, \ldots, s_N\}$, $\mathbf{e} = \{e_1, \ldots, e_N\}$) is modelled as the result of an underlying birth–death process, with parameters λ and μ indicating the expected number of speciation and extinction events per lineage per unit of time [e.g. Myr (Silvestro *et al.*, 2014b)]. Within any given time frame $\tau = \{t_i, t_{i+1}\}$ (where $t_i > t_{i+1}$), the likelihood of a birth-death process is:

$$P(\mathbf{s}, \mathbf{e} \mid \lambda, \mu, \tau) \propto \lambda^{B_\tau} \mu^{D_\tau} e^{-(\lambda+\mu)S_\tau}, \quad (1)$$

where B_τ and D_τ are the number of speciation and extinction events occurring within the time frame τ and S_τ represents the sum of species life spans within the time frame (Keiding, 1975; Silvestro *et al.*, 2014b):

$$S_\tau = \sum_{n=1}^{N} \left[\min(s_n, t_i) - \max(e_n, t_{i+1}) \right]. \quad (2)$$

We can incorporate temporal variation of speciation and extinction rates in the model by introducing different time frames and assigning independent rates to each of them. Time frames can be fixed *a priori* (Silvestro *et al.*, 2015b) or estimated from the data (Silvestro *et al.*, 2014b). For a given set of T time frames, the likelihood of a birth–death process is:

$$P(\mathbf{s}, \mathbf{e} \mid \lambda, \mu, T) \propto \prod_{\tau}^{T} \left[\lambda_\tau^{B_\tau} \exp(-\lambda_\tau S_\tau) \right] \prod_{\tau}^{T} \left[\mu_\tau^{D_\tau} \exp(-\mu_\tau S_\tau) \right], \quad (3)$$

where λ_τ and μ_τ are the speciation and extinction rates in each time frame (Silvestro *et al.*, 2014b).

Multiple clade diversity dependence

Let us define as δ_i the diversity trajectory of a clade i, so that $\delta_i(t)$ represents its standing diversity at time t. The MCDD model uses a time varying birth–death model (equation 3) in which the time frames of rate change are defined by changes in the diversity trajectories of the clades considered. Diversity dependence is modelled by a linear correlation between speciation and extinction rates and a diversity trajectory (Etienne *et al.*, 2012; Silvestro *et al.*, 2015a). The diversity trajectories of multiple clades are jointly analysed to assess the existence of diversity dependence effects. Thus, in a set of C clades with diversity trajectories $\mathcal{D} = \{\delta_1, \delta_2, \ldots, \delta_C\}$, the speciation and extinction rates for a given clade i at time t are obtained by the following transformations:

$$\lambda_i(t) = \max\left\{0, \lambda_i - \sum_{j=1}^{C} \lambda_i [\delta_j(t) g_{ij}^{\lambda}]\right\} \quad (4a)$$

and

$$\mu_i(t) = \max\left\{0, \mu_i - \sum_{j=1}^{C} \mu_i [\delta_j(t) g_{ij}^{\mu}]\right\}, \quad (4b)$$

where λ_i, μ_i are baseline speciation and extinction rates of clade i, and g_{ij}^{λ}, g_{ij}^{μ} are the diversity dependence parameters transforming speciation and extinction rates, respectively. The baseline speciation and extinction rates (λ_i, μ_i) represent the rates at which clade i diversifies when not affected by any diversity dependence effects. Since $i \in C$, the within-clade diversity dependence is quantified by g_{ij}, where $i = j$. It is emphasized that the reciprocal interactions between two clades are modelled by two independent parameters (e.g. g_{ij}, g_{ji}), thus providing the directionality of diversity dependence and allowing asymmetric effects.

Bayesian variable selection

The original implementation of the MCDD model used Bayesian variable selection (BVS) to deal with the large number of parameters and tease apart noise from signal. Under the BVS implementation of the model, the diversity dependence parameters are replaced by auxiliary variables so that $g_{ij} = k_{ij} I_{ij}$, where $k \in \mathbb{R}$ is the effect size representing the intensity of

competition ($k > 0$) or positive interaction ($k < 0$). The indicator variable (I_{ij}) can only take values equal to 0 or 1, and determines the presence or absence of a diversity dependence effect of intensity equal to k_{ij}. We treat the auxiliary variables as independent and estimate them from the data (Kuo and Mallick, 1998; Silvestro *et al.*, 2015a). We use uniform priors on the effective sizes, $P(k) \sim \mathcal{U}(-0.3, 0.3)$. Hence, the addition of one species in the diversity trajectory of clade j can decrease or increase the speciation and extinction rates of clade i by up to 30% of its baseline rates, based on equation (4) (Silvestro *et al.*, 2015a). Since the indicators can only take two values, we used a Bernoulli prior distribution on I:

$$P(I_{ij} \mid \eta_i) = \begin{cases} \eta_i & \text{for} \quad I_{ij} = 0 \\ 1 - \eta_i & \text{for} \quad I_{ij} = 1 \end{cases}, \tag{5}$$

where the probability η_i is an unknown hyperparameter with uniform prior probability between 0 and 1, i.e. a flat beta distribution $P(\eta) \sim \mathcal{B}(1, 1)$, and is estimated from the data. The resulting prior on the diversity dependence parameters (g) under this BVS model presents a spike at zero with mass equal to η while the remaining probability $1 - \eta$ is uniformly distributed according to the prior on the effect size $P(k)$. For the baseline speciation and extinction rates, we use an exponential prior with rate parameter l, which is in turn assigned a weak gamma hyper-prior $P(l) \sim \Gamma(1, 0.1)$ and estimated from the data.

The BVS method provides a direct way to estimate the probability of a positive (or negative) diversity dependence effect, which is simply the frequency of positive (or negative) diversity dependence parameter g in the posterior samples. For instance, the marginal probability that clade i is under competition from clade j is given by

$$P(g_{ij} > 0) = \frac{\Sigma_M z_{ij}^{+}}{M}, \tag{6}$$

where M is the number of posterior samples resulting from the MCMC and $z_{ij}^{+} = 1$ when $g_{ij} > 0$ and $z_{ij}^{+} = 0$ when $g_{ij} \leq 0$. A similar equation can be used to estimate the probability that clade i experiences positive interaction from clade j, $P(g_{ij} < 0)$. Such probabilities can be calculated separately for speciation and extinction, by considering g_{ij}^{λ} and g_{ij}^{μ}, respectively. A standard threshold to assess the significance of a diversity dependence effect is 0.5 (Gelman *et al.*, 2013), thus any diversity dependence parameter with a probability >0.5 of being greater (or smaller) than 0 can be interpreted as signal, while probabilities <0.5 are interpreted as noise.

MCMC implementation of the BVS method

The baseline speciation and extinction rates (λ, μ), indicators (I) and effect sizes (k) were sampled via Markov chain Monte Carlo (MCMC) using standard random updates and Metropolis-Hastings acceptance ratio (Metropolis *et al.*, 1953; Hastings, 1970). To improve convergence, proposals for k were bounded within the allowed range of [–0.3, 0.3] using reflection at the boundaries (Ronquist *et al.*, 2007). Values of η and l were drawn directly from their respective conjugate posterior distributions:

$$\eta_i \sim \mathcal{B}\left(1 + 2C - \sum_{j=1}^{C} I_{ij},\ 1 + \sum_{j=1}^{C} I_{ij}\right)$$

and

$$l \sim \Gamma(3, 0.1 + \lambda + \mu),$$

where $\mathcal{B}(a, b)$ is a beta distribution with shape parameters a and b, and $\Gamma(a, b)$ is a gamma distribution with shape parameter a and rate parameter b.

The MCDD model using BVS is implemented in the Python program 'PyRateMCDD', available as part of the open source package PyRate (https:github.com/dsilvestro/PyRate).

Horseshoe prior

An alternative approach to the BVS method described above is provided by the 'horseshoe prior' (HSP), originally described by Carvalho *et al.* (2010) to address problems with large numbers of parameters and sparse signal. Under the HSP, the diversity dependence parameters g_{ij} are assigned a normal prior distribution centred in 0 and with variance given by two hyper-parameters ε_{ij} and τ so that:

$$P(g_{ij} \mid \varepsilon_{ij}, \tau) \sim \mathcal{N}(0, \varepsilon_{ij}^2\tau^2), \tag{7}$$

The hyper-parameters are estimated from the data and assigned half-Cauchy prior distributions: $\varepsilon_{ij} \sim \mathcal{C}^+(0, 1)$ and $\tau \sim \mathcal{C}^+(0, 1)$ (Carvalho *et al.*, 2010; Scott, 2011). The HSP distribution is characterized by an infinitely tall spike at zero, yielding the shrinkage of noise parameters (i.e. negligible diversity dependence effects), and by Cauchy-like heavy tails, allowing signals (i.e. significant diversity dependence effects) of potentially strong positive or negative intensity. The key aspect of the HSP is that it incorporates two shrinkage components: the local shrinkage parameters $\varepsilon = \varepsilon_{i1}, \ldots, \varepsilon_{iC}$, which shrink (or release) each individual diversity dependence parameter $g = g_{i1}, \ldots, g_{iC}$ and a global shrinkage parameter τ, which is estimated from the average signal density (Carvalho *et al.*, 2010). This parameterization has been shown to accurately distinguish signal from noise and to yield posterior parameter estimates that are remarkably similar to those obtained through gold standard Bayesian model averaging methods (Carvalho *et al.*, 2010). In our implementation, we assumed independent local shrinkage parameters for each diversity dependence effect on speciation and extinction (e.g. $\varepsilon_{i1}^{\lambda}, \ldots, \varepsilon_{iC}^{\lambda}$ for $g_{i1}^{\lambda}, \ldots, g_{iC}^{\lambda}$ and $\varepsilon_{i1}^{\mu}, \ldots, \varepsilon_{iC}^{\mu}$ for $g_{i1}^{\mu}, \ldots, g_{iC}^{\mu}$), while we used a single global shrinkage parameter τ for all diversity dependence effects and for both speciation and extinction.

From the estimated values of local and global shrinkage, we can calculate for each diversity dependence parameter g_{ij} the respective shrinkage coefficient $\kappa_{ij} = 1/(1 + \tau^2\varepsilon_{ij}^2)$, where $\kappa_{ij} \approx 1$ indicates that g_{ij} is shrunk near 0 (i.e. no diversity dependence effect), whereas $\kappa_{ij} \approx 0$ indicates that g_{ij} is basically unshrunk to either positive or negative values (Carvalho *et al.*, 2010). Unlike with the BVS implementation, the HSP does not allow a direct estimation of the probability of a positive or negative diversity dependence effect. However, using the estimated shrinkage coefficients we can calculate shrinkage weights, defined as $w_{ij} = 1 - \kappa_{ij}$, which provide an alternative way to assess the significance of the effect. Greater shrinkage weights indicate signal (i.e. evidence for diversity dependence), while weights close to 0 indicate strong shrinkage applied to noise parameters, and a threshold of 0.5 can be used to distinguish between noise and signal (Carvalho *et al.*, 2010).

Despite some obvious differences in parameterization, both BVS and HSP are valid ways to determine which and how many parameters contribute significantly to explaining the data and to suppress those that only represent background noise in the model. The results

obtained from the two methods are expected to be similar in several statistical contexts (O'Hara and Sillanpää, 2009; Carvalho *et al.*, 2010), but it is difficult to predict their performances specifically within the MCDD analytical framework. Here, we took the approach of analysing the same data sets under both algorithms to assess the respective degree of power and accuracy under a range of scenarios (see 'Simulations' below).

MCMC implementation of the HSP method

We used standard random updates for baseline speciation and extinction rates (λ, μ) and for the diversity dependence parameters (g) with Metropolis-Hastings MCMC to sample them from their posterior distribution. To improve convergence and comparability with the BVS implementation, proposals for g were bounded within a range of [–0.3, 0.3] using reflection at the boundaries. In order to update the shrinkage parameters $\varepsilon = \varepsilon_{i1}, \ldots, \varepsilon_{iC}$ and τ, we used a slice-sampling algorithm to draw new parameter values directly from their conditional posterior distribution (Damien *et al.*, 1999; Scott, 2010, 2011). After defining $\eta = 1/\tau_2$ and $\theta = \theta_{i1}, \ldots, \theta_{iC}$, where $\theta_{ij} = g_{ij}/\varepsilon_{ij}$, we used the two following steps to update the global shrinkage parameter τ:

- sample $u \sim \mathcal{U}(0, 1/(1 + \eta))$
- sample

$$\eta' \sim \mathrm{T}\Gamma\left(\frac{2C+1}{2}, \frac{2}{\sum_{j=1}^{C} \theta_{ij}^2}, \frac{1-u_{ij}}{u_{ij}}\right) \tag{8}$$

where $\mathrm{T}\Gamma(a, s, t)$ is a gamma distribution with shape parameter a and scale parameter s truncated at t (i.e. with zero probability outside of the range $[0, t]$), and $2C$ is the number of local shrinkage parameters in a data set of C clades.

We then re-parameterized back to the τ-scale to obtain a posterior draw of the global shrinkage parameter $\tau' = \sqrt{(1/\eta')}$. Similarly, for the local shrinkage parameters we define $\eta_{ij} = 1/\varepsilon_{ij}^2$ and $\mu_{ij} = g_{ij}/\tau$ and

- sample $u_{ij} \sim \mathcal{U}(0, 1/(1 + \eta_{ij}))$
- sample $\eta'_{ij} \sim \mathrm{TExp}(2/\mu_{ij}^2, (1 - u_{ij})/u_{ij})$, where $\mathrm{TExp}(s, t)$ is an exponential distribution with scale parameter s truncated at t.

As done previously, we then transform back to the ε-scale to obtain a posterior draw of the local shrinkage parameter $\varepsilon' = \sqrt{(1/\eta'_{ij})}$ (Scott, 2010, 2011).

The HSP implementation of the MCDD model is available as 'PyRateMCDD-HSP' and included in the open source package PyRate (https:github.com/dsilvestro/PyRate).

Simulations

We tested and compared the performance of the two implementations of the MCDD model by analysing a range of simulated data sets. Assessing the accuracy and power of complex macroevolutionary models using simulations is a necessary practice given that such models may be prone to biases and inaccurate inferences (e.g. Davis *et al.*, 2013; Rabosky and Goldberg, 2015). Owing to the difficulty of simulating data sets under multiple diversity dependence effects,

we simulated cases in which diversity dependence occurs within clades or between two clades (e.g. $g_{ij} \neq 0$). We focused on data sets with either no diversity dependence ($g = 0$) or competitive effects ($g > 0$) and assessed whether the MCDD model can infer the correct diversification scenario. We simulated data sets of 5, 10, and 20 clades under six scenarios (detailed settings are given in Table 1).

I. All clades diversify under a constant birth–death process without diversity dependence effects, thus diversification is time homogeneous and unbounded and clades do not interfere with each other (the typical null hypothesis in phylogenetic studies).
II. All clades diversify under birth–death with random rate shifts through time (see below), but without explicit diversity dependence effects. The diversification process is therefore non-homogeneous, but rate changes are purely stochastic. This scenario is somewhat analogous to the null model used in several studies of biodiversity changes in palaeobiology (Raup and Gould, 1974; Gould *et al.*, 1977).
III. The clade of interest diversifies under self diversity-dependent speciation rate and constant extinction rates, while other clades diversify under random birth–death models with rate shifts as in scenario II. Within-clade diversity dependence has been inferred for a number of clades (Phillimore and Price, 2008) and can be related to the expectations of evolutionary radiations driven by the saturation of ecological opportunities (Gavrilets and Losos, 2009).

Table 1. Summary of the simulation settings

Simulation	Root ages	(Baseline) speciation	(Baseline) extinction	No. of rate shifts	Diversity dependence parameters
I	$r \sim \mathcal{U}(30, 15)$	$\lambda \sim \mathcal{U}(0.01, 1)$	$\mu \sim \mathcal{U}(0.01, 1)$	$s = 0$	
II	$r \sim \mathcal{U}(30, 15)$	$\lambda \sim \mathcal{U}(0.01, 1)$	$\mu \sim \mathcal{U}(0.01, 1)$	$s \sim \mathcal{P}(1)$	
III	$r \sim \mathcal{U}(30, 15)$	$\lambda_i \sim \mathcal{U}(0.75, 1)$, $\lambda \sim \mathcal{U}(0.01, 1)$	$\mu_i \sim \mathcal{U}(0.05, 0.1)$, $\mu \sim \mathcal{U}(0.01, 1)$	$s_i = 0$, $s \sim \mathcal{P}(1)$	$g^{\lambda}_{ii} \sim \mathcal{U}(0.025, 0.05)$
IV	$r \sim \mathcal{U}(30, 15)$	$\lambda \sim \mathcal{U}(0.75, 1)$	$\mu \sim \mathcal{U}(0.05, 0.1)$	$s = 0$	$g^{\lambda}_{yy} \sim \mathcal{U}(0.025, 0.05)$ for $y \in 1, \ldots, C$
V	$r_i \sim \mathcal{U}(30, 25)$, $r_j \sim \mathcal{U}(30, 25)$, $r \sim \mathcal{U}(30, 15)$	$\lambda_i \sim \mathcal{U}(0.75, 1)$, $\lambda_j = [0.3, 0.1]$, $\lambda \sim \mathcal{U}(0.01, 1)$	$\mu_i \sim \mathcal{U}(0.05, 0.1)$, $\mu_j = [0.01, 0.3]$, $\mu \sim \mathcal{U}(0.01, 1)$	$s_i = 0$, $s_j = 1$, $s \sim \mathcal{P}(1)$	$g^{\lambda}_{ii} \sim \mathcal{U}(0.025, 0.05)$ $g^{\mu}_{ij} \sim \mathcal{U}(0.1, 0.3)$
VI	$r_i \sim \mathcal{U}(30, 20)$, $r_j \sim \mathcal{U}(20, 15)$, $r \sim \mathcal{U}(30, 15)$	$\lambda_i \sim \mathcal{U}(0.5, 0.75)$, $\lambda_j = \mathcal{U}(0.75, 1)$, $\lambda \sim \mathcal{U}(0.01, 1)$	$\mu_i \sim \mathcal{U}(0.025, 0.05)$, $\mu_j = \mathcal{U}(0.05, 0.1)$, $\mu \sim \mathcal{U}(0.01, 1)$	$s_i = 0$, $s_j = 0$, $s \sim \mathcal{P}(1)$	$g^{\lambda}_{ij} \sim \mathcal{U}(0.025, 0.05)$ $g^{\lambda}_{jj} \sim \mathcal{U}(0.025, 0.05)$ $g^{\mu}_{ij} \sim \mathcal{U}(0.1, 0.3)$

Note: Parameters without a subscript (e.g. r, λ) refer to values applied to all clades unless otherwise specified. Random values for root ages and (baseline) speciation and extinction rates were drawn from uniform distributions ($\mathcal{U}$) and the number of random shifts in speciation and extinction were drawn from Poisson distributions ($\mathcal{P}$) unless otherwise specified. In simulation V, the birth–death rates were fixed with one shift in speciation rate at time 20 and one shift in extinction rate at time 15. All diversity dependence parameters were equal to 0, unless otherwise specified. Clades were conditioned on having between 20 and 250 (extinct and extant) lineages.

IV. All clades diversify under diversity-dependent speciation rate and constant extinction rates. We simulate here bounded diversification for all clades where the diversification of each clade is only regulated by their own diversity. This scenario is consistent with the view that diversity dependence acts primarily on speciation rates within a given lineage (Alroy, 1996).
V. The clade of interest *i* diversifies under self diversity dependence effects on speciation rates and under competitive effects on extinction rates driven by a competing clade *j*, while clade *j* independently rises and declines in diversity (more details below). Clades *i* and *j* originate roughly at the same time, but clade *j* undergoes a rapid diversification followed by a decline in diversity (independently of competition; see Table 1). Because of the competitive effects of *j*, clade *i* diversifies at higher net rates only with the decline of diversity in clade *j*. Thus, this simulation scenario represents a case of passive replacement, with *j* being the incumbent clade and *i* profiting from its demise (Benton, 1996; Van Valkenburgh, 1999).
VI. The clade of interest *i* diversifies under competitive effects on both speciation and extinction from clade *j*, while clade *j* diversifies under self diversity-dependent speciation rates and constant extinction rates. This scenario reproduces a 'double-wedge' pattern linked with active displacement of clades (Benton, 1996; Sepkoski, 1996a; Van Valkenburgh, 1999), whereby the diversification of the focal clade *i* is negatively affected by the origination and diversification of a younger competing clade *j*.

These six scenarios thus encompass fundamental diversification dynamics studied in palaeobiology and macroevolution. Moreover, they include elements (randomness, within and between diversity dependence, effects on one versus all clades) that could confound the detection of signals of clade interference, potentially affecting the power and accuracy of the MCDD model. Therefore, this set of simulations allows testing of the accuracy of the proposed method to detect interference between clades under a suite of plausible diversification scenarios. Both scenarios I and II are null models, in which diversification occurs under independent birth–death processes with no interactions within or between clades. The rate variation introduced by random rate shifts in simulation II potentially generates a signal that could be interpreted as diversity dependence, since the diversity trajectory of a clade can randomly correlate with rate changes in another clade (Silvestro *et al.*, 2015a). Both scenarios III and IV represent cases of competitive effects within the clade of interest, but, in simulation IV, similar diversity dependence occurs also within the other clades, potentially allowing a better chance of detecting false diversity dependence effects among clades. Scenarios V and VI introduce interactions (diversity dependence) between clades where competitive effects occur between different lineages. The result of competition is clade replacement in both cases, but in V competition is preventing diversification, whereas in VI competition is causing diversity decline.

Data analysis

We simulated 100 data sets under each scenario and clade number (a total of 1800 data sets), and analysed them using the BVS and the HSP implementations of the MCDD model. We ran 5,000,000 MCMC iterations sampling every 5000 to approximate the posterior distribution of the parameters, which we summarized by calculating the mean and 95% highest posterior density (95% HPD). We inspected MCMC convergence using Tracer (Rambaut *et al.*, 2014) and discarded the initial 1,000,000 iterations as burn-in.

We then assessed the significance of all diversity dependence effects based on their estimated probability under BVS (equation 6) and based on their shrinkage weights (w) under HSP. We assessed the frequency of true and false positives as the probability that a clade is correctly (or wrongly) identified as having a significant diversity dependence effect on the clade of interest based on the standard threshold of 0.5 (Carvalho *et al.*, 2010; Gelman *et al.*, 2013) and we explored the effect of higher thresholds. As an additional threshold rule, we looked at the 95% HPD of the diversity dependence parameters (g^{λ}, g^{μ}) and considered the effect as significant only when zero was outside of the 95% HPD interval. Finally, we also estimated the probability that at least one clade, regardless of which one, is found to significantly affect the diversification of the clade of interest. These probabilities in simulations where none of the clades have diversity dependence effects (e.g. simulation I), represent the false positive rate for the entire data set (of 5, 10, or 20 clades) and are therefore expected to increase as the number of clades analysed increases.

RESULTS

We structure the description of the results from the simulations by first assessing the ability of the algorithms to correctly detect diversity dependence effects under different thresholds and evolutionary scenarios. Then, we describe the posterior estimates of the MCDD parameters across simulations and compare them between the BVS and HSP algorithms.

Identification of diversity dependence effects

Our MCDD analyses show that diversity dependence can be correctly identified in most of the simulated scenarios (Tables S1–S3; www.evolutionary-ecology.com/data/3010Appendix.pdf). Under constant rate birth–death models (scenario I), the frequency of false positives using a threshold of 0.5 was between 0.08 and 0.17 under BVS and between 0 and 0.07 under the HSP (Fig. 1A). The false positive rates were consistently smaller under the HSP than under the BVS implementation. The number of clades included in the analysis did not influence strongly these results, although there is a slight trend towards fewer false positives for extinction rates with increasing number of clades. When using a more conservative threshold for significance (0.6), the frequencies of false positives drop to 0.05–0.10 for BVS analyses and below 0.02 for HSP analyses (Fig. 1B). The frequency of false positives falls to 0–0.01 when applying the 95% HPD rule to assess significance (Fig. 1C). Under the BVS implementation, the probability of finding at least one false positive among the analysed clades increases, as expected, with the number of clades included in the data set, from 0.49 in data sets of 5 clades to 0.89 in data sets of 20 clades (0.33 and 0.66 respectively using a threshold of 0.6). Consistent with the lower false positive rates reported above, the probability of finding at least one false positive among the analysed clades was much lower under the HSP implementation (Tables S1–S3). Unexpectedly, false positive rates under HSP decrease (though only slightly) with increasing number of clades, from 0.19 in data sets of 5 clades to 0.13 in data sets of 20 clades. These values further reduce to 0.045 and 0.035 when using a threshold of 0.6.

As expected based on previous simulations (Silvestro *et al.*, 2015a), the introduction of random rate variation (scenario II) increases the chances of inferring significant diversity dependence effects, even if none was used to simulate the data. Under random variable birth–death and a threshold of 0.5, the frequency of false positives inferred by BVS ranged between

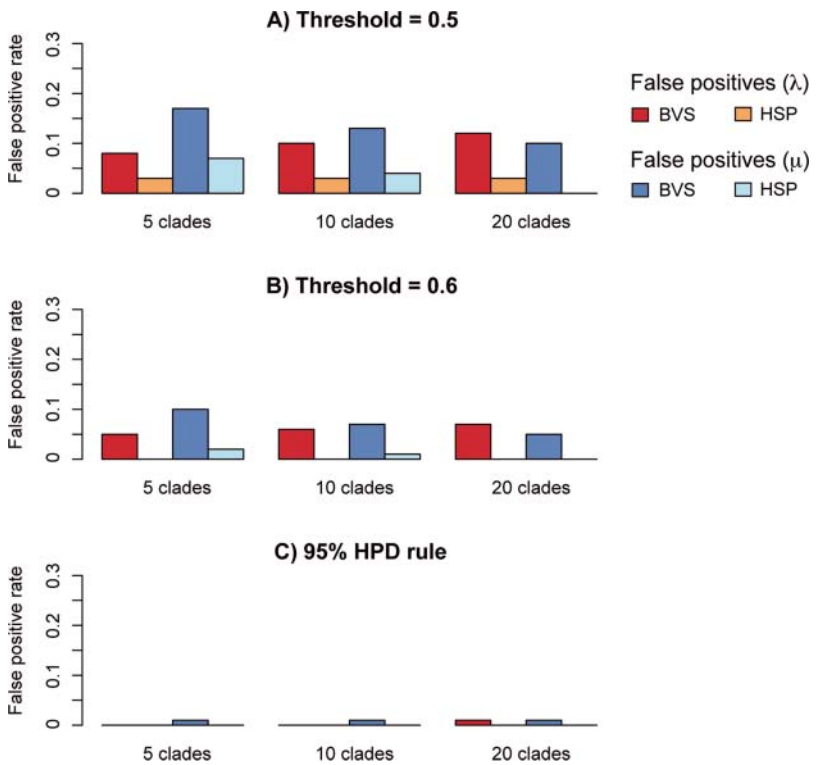

Fig. 1. Frequencies of false positives across simulations (scenario I) under the standard 0.5 threshold and a more conservative threshold of 0.6.

0.09 and 0.21, with no significant variation linked with the number of clades analysed (Tables S1–S3). The frequencies decreased to 0.07–0.14 under a 0.6 threshold. In contrast, HSP analyses yielded false positive frequencies around 0.15–0.18 in data sets of 5 clades, but false positives decreased to 0.10–0.15 with 10 clades, and to 0.06–0.07 with 20 clades. False positives further decreased from 0.07–0.08 (5 clades) to 0.02–0.03 (20 clades) under a 0.6 threshold. Under the 95% HPD rule, false positives ranged between 0 and 0.03 across simulation II.

The analysis of data sets simulated under a model of diversity dependence within the clade (scenarios III and IV) of interest suggests that the model succeeds in identifying diversity dependence. Under BVS, the frequencies of false positives (i.e. the frequency of spurious diversity dependence effects beyond the simulated ones) were generally below 0.09, but reached 0.22 for speciation in the case of 5 clades, all undergoing self diversity dependence speciation (scenario IV). In contrast, the frequencies of false positives under the HSP were always very low and ranged between 0 and 0.03 across simulations III and IV (Tables S1–S3). Both implementations showed high power to correctly identify diversity dependence within the clade of interest. The rates of true positives were slightly higher under BVS (0.91–1) than under HSP (0.85–0.99). The 95% HPD rule yielded low false positive rates for both BVS and HSP (0–0.04) while maintaining reasonably high power (true positive rates of 0.71–0.97).

In simulations with both speciation and extinction undergoing clade competition (scenarios V and VI), the average power to correctly identify diversity dependence under BVS is 0.77 in data sets of 5 clades, decreasing to 0.52 in data sets of 20 clades (Tables S1–S3; Fig. 2). Similarly, the HSP implementation yielded frequencies of true positives ranging from 0.79 (5 clades) to 0.48 (20 clades). Diversity dependence was usually found with more confidence for speciation rates than for extinction (Fig. 2). In simulations of scenario V, the false positive rates are similar to those observed for scenarios I–IV (Fig. 2), but they were substantially higher under scenario VI, reaching 0.13 under BVS and 0.23 under HSP in data sets of 5 clades. False positive rates, however, decreased to 0.10 (BVS) and 0.06 (HSP) in data sets of 20 clades (Tables S1–S3). When using the 95% HPD rule, false positive rates were small for both scenarios V and VI, but at the cost of a lower power, especially for extinction diversity dependence (Fig. 2C; Tables S1–S3).

Posterior parameter estimates

The posterior estimates of the diversity dependence parameters (g^{λ}, g^{μ}) are found to be centred around 0 in cases in which they were indeed set to 0 (i.e. no diversity dependence) when generating the data (Fig. 3E, F; Figs. S1–S18, 3010Appendix). The posterior estimates

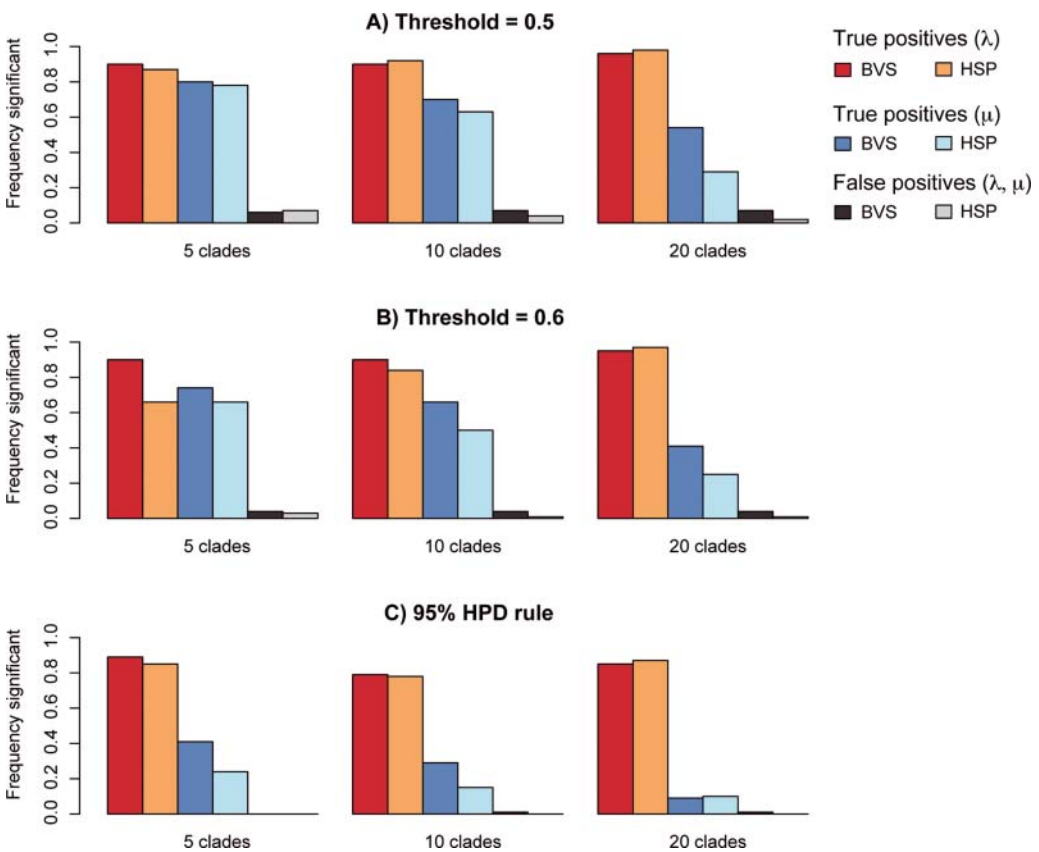

Fig. 2. Frequencies of true and false positives across simulations (scenario V) under the standard 0.5 threshold and a more conservative threshold of 0.6.

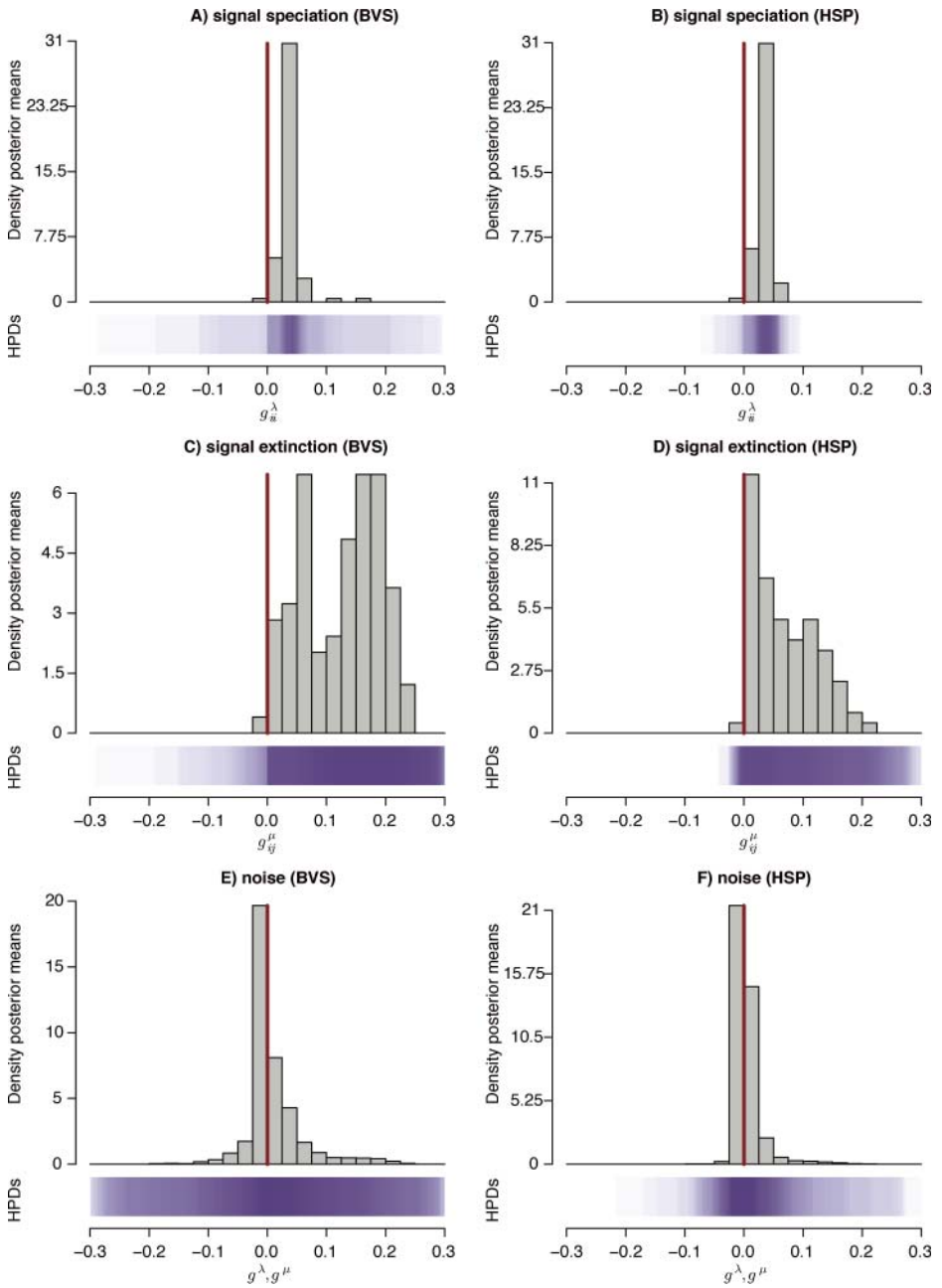

Fig. 3. Posterior estimates of the diversity dependence parameters (g^{λ}, g^{μ}) shown as mean (histograms) and 95% HPDs (shaded areas) across 100 simulations generated under scenario V with 10 clades.

of these parameters represent the 'noise' in the data sets and obtained similar levels of accuracy across all simulations. Estimates are much more strongly concentrated around 0 under the HSP than under the BVS implementation, suggesting that the horseshoe prior applies a stronger shrinkage over the noise parameter (e.g. Fig. 3E, F, Fig. S5). This pattern is also reflected in the size and distribution of the respective HPDs, which are narrow around 0 in HSP analyses, whereas in BSV runs the HPDs are much wider and tend to span the entire prior range [–0.3, 0.3]. Under HSP, we found a consistent decrease in false positive rates with increasing number of clades analysed (Fig. 1), which appears to be linked to decreasing values of the global shrinkage parameter τ (Fig. 4). The global shrinkage parameter reflects the amount of background noise in the data, which likely increases with increasing number of clades analysed. This leads to smaller values of τ, which, in turn, yield a stronger shrinkage of the diversity dependence parameters and, $\varepsilon_{i1}, \ldots, \varepsilon_{iC}$ being equal, lower shrinkage weights ($w_{i1}, \ldots, w_{iC}$,).

The simulated competition effects acting on speciation and extinction rates ($g^{\lambda} > 0$, $g^{\mu} > 0$) were accurately detected in most simulations under both BVS and HSP analyses (Fig. 3A–D; Figs. S7–S18). Large effects, however, tended to be shrunk towards smaller values under HSP (Fig. 3D) and the shrinkage was stronger with increasing number of clades (e.g. Figs. S13–S14), likely a consequence of smaller estimated values of the global shrinkage parameter τ (Fig. 4). As observed for noise parameters, the HPDs of positive diversity dependence effects were substantially narrower in HSP estimates compared with BVS estimates, suggesting higher precision in the posterior distribution (Fig. 3A–D).

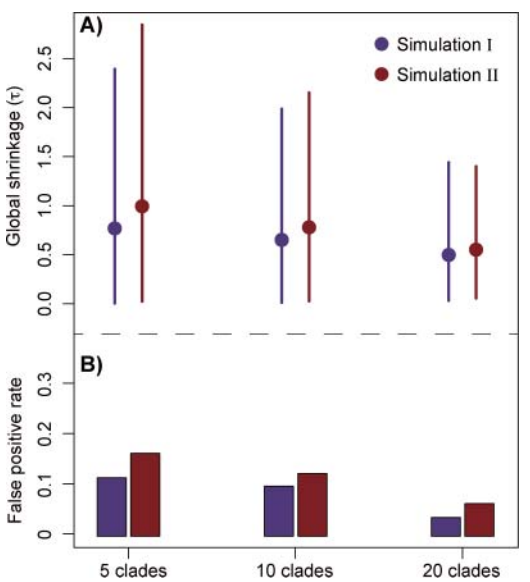

Fig. 4. Posterior estimates of the global shrinkage parameter (τ) posterior mean and 95% HPD averaged over 100 replicates of simulation settings I and II (A). The global shrinkage decreases with increasing number of clades analysed and is associated with a decline in the frequency of false positives (here averaged over speciation and extinction using a 0.5 threshold; B).

DISCUSSION

Inferring diversity dependence from the fossil record is challenging, especially at the species level, because it requires a very good fossil record (Marshall and Quental, 2016). Although there are some analyses of diversity dependence within a given clade (Alroy, 1996, 2008; Ezard *et al.*, 2011; Liow and Finarelli, 2014), the task is even harder when considering interference between separate lineages (but see Silvestro *et al.*, 2015a). The central assumption that guided most research on diversity dependence is that species within the same lineage are more likely to affect each other (Marshall and Quental, 2016), as they should have similar requirements and use similar resources (Darwin, 1859). Nevertheless, niche overlap among distantly related taxa, and thus the effects they exert upon each other, might be as large as that between species that show close phylogenetic relatedness (Schluter, 1986; Diamond, 1987; Englund *et al.*, 1992). In fact, several previous studies suggest that competitive interactions among lineages may occur in a wide range of organisms [e.g. in plants (Knoll, 1986; Schneider *et al.*, 2004), invertebrates (Lidgard *et al.*, 1993; Sepkoski *et al.*, 2000; Liow *et al.*, 2015), and vertebrates (Van Valen and Sloan, 1966; Rosenzweig and McCord, 1991; Van Valkenburgh, 1999; Silvestro *et al.*, 2015a)], although it is usually assumed that competition will preferentially play a role through incumbency effects rather than through active displacement (Rosenzweig and McCord, 1991; Benton, 1996; Jablonski, 2008).

Here we present a method that allows quantification of the effects of diversity dependence within and between lineages in a rigorous way. We have shown that clade competition can be inferred with confidence in a wide range of scenarios using the MCDD model. The two implementations of the MCDD model provide comparable results, but also display some clear differences, linked with the way parameters are shrunk around 0 when identified as noise. Both BVS and HSP are able, though with variable power and accuracy, to tease apart signal from noise through a joint estimation of all parameters. This is a clear advantage over traditional model testing approaches in that the extremely large numbers of possible diversity dependence scenarios are tested in a single run, while noise shrinkage prevents over-parameterization.

Under our BVS algorithm, there is total shrinkage of parameters when identified as noise (i.e. $g_{ij} = 0$ when the respective indicator equals 0) and no shrinkage (uniform prior) when they are identified as signal. This leads to posterior distributions that are usually difficult to summarize due to a spike at exactly 0. The HSP implementation provides instead a continuous transition in the shrinkage applied to noise and signal, whereby all parameters are shrunk through a single global parameter and through local parameters. Owing to the infinite spike at zero and the heavy tails of the horseshoe distribution, noise parameters are shrunk near 0 much more strongly than signal parameters (Carvalho *et al.*, 2010). The difference in these two parameterizations result in much wider credible intervals under BVS, where parameters have essentially no prior constraints (besides the minimum and maximum boundaries) whenever the respective indicators equal 1. In contrast, credible intervals are much narrower under HSP (Fig. 3B, F). However, the stronger overall shrinkage applied by the HSP can result in an underestimation of the absolute value of the diversity dependence parameter when its true value is large (Fig. 3D).

Although the probabilities of clade diversity dependence estimated by BVS are not equivalent to the shrinkage weights inferred from HSP analyses, both metrics can be used to assess the significance of the estimated positive or negative effects. This similarity has been previously shown in different implementations (Carvalho *et al.*, 2010) and is also observed in our MCDD analyses. The use of more conservative thresholds for significance (i.e. 0.6 instead

of the standard 0.5) clearly reduces the risk of wrongly identifying diversity dependence effects, though at the cost of decreasing the power of the inference. We highlight, however, that the posterior mean values of the g^{λ}, g^{μ} parameters obtained from MCDD analyses are effectively the result of model averaging (Carvalho *et al.*, 2010; Gelman *et al.*, 2013), and thus can be interpreted beyond the binary and biologically simplistic choice between 'significant' and 'non-significant' effects. This is particularly evident when we consider the diversity dependence effects estimated under HSP. Despite the presence of some false positives (depending on the threshold rule applied), the posterior values of noise parameters are consistently narrowly centred around 0, indicating that no biologically meaningful effect is detected (e.g. Fig. 3F, Figs. S1–S6). Thus, we argue that robust evolutionary interpretation of the MCDD model can be simply derived from the posterior mean of the HSP estimates, without strictly relying on a threshold rule. Indeed, while shrinkage weights (and probabilities obtained from BVS) provide valuable measures of the statistical support for diversity dependence hypotheses, their values are also directly reflected in the *a posteriori* values of g^{λ}, g^{μ}, which therefore summarize both the probability and the intensity of diversity dependence.

Because the detection of diversity dependence relies on correlations between the birth–death dynamics and diversity trajectories, the ability of the MCDD model to identify competitive or positive interactions between clades depends on the amount of temporal overlap between lineages. Thus, when the amount of temporal overlap between clades is small, the power of the analysis decreases. This is the case of simulation VI, where the clade of interest diversifies under no constraints for the first part of its existence and only *c.* 7.5 Myr later starts to feel the negative effects of a younger competing clade (Table 1). The increased difficulty of identifying diversity dependence effects under this scenario is reflected by a lower power of the analysis and a higher uncertainty around which clade is responsible for rate changes observed in the clade of interest (Tables S1–S3). A similarly high degree of uncertainty around which clade is actively displacing an older clade has been observed in previous empirical analyses of active displacement (Silvestro *et al.*, 2015a).

Previous analyses of the fossil record indicate both active displacement and passive replacement between distantly related lineages may have occurred several times (Rosenzweig and McCord, 1991; Sepkoski, 1996b; Van Valkenburgh, 1999; Sepkoski *et al.*, 2000; Silvestro *et al.*, 2015a). The MCDD model enables both phenomena to be identified from fossil data and may help us unravel whether the effects of ecological interactions can indeed scale up, driving diversification dynamics at geological times scales. Although the general sense has been that active displacement was a rather rare phenomenon (Benton, 1996), we suggest that the development of our method might allow further scrutinization of this perception. We should note, however, that in the most extreme case of incumbency, where the radiation of the later clade is only possible after a complete removal of the incumbent clade, the temporal overlap between clades might be small and the method proposed here is unlikely to detect diversity dependence signals.

The current implementation of MCDD is based on variable rate birth–death models and represents a process-based approach to detect positive or negative interactions within and between clades (where birth and death are the processes generating diversity patterns), as compared with other pattern-based methods that look at simpler correlations between diversity trajectories (Benton, 1996; Sepkoski, 1996b). The possibility to explicitly test diversity dependence hypotheses in a statistically rigorous way may help to understand the relevance of diversity dependence in more general terms. The approaches that we have

developed in this study contribute to fuelling the debate on the existence and strength of limits to species diversification. More generally, it provides a new direction in current attempts to test models in comparative methods (e.g. Maddison and FitzJohn, 2015; Rabosky and Goldberg, 2015) and could be more widely used to assess the power and accuracy of these methods to infer the mechanisms driving macroevolutionary patterns. The inclusion of additional information concerning species ecology, behaviour, phenotypic traits, and geographic distribution can potentially increase the power of fossil-based analyses and improve our understanding of the mechanisms driving diversification dynamics and their responses to biotic interactions.

ACKNOWLEDGEMENTS

We thank P. Smith, J.G. Scott, and X. Meyer for input on the HSP and discussion, and P. Raia, M. Fortelius, and two anonymous reviewers for excellent feedback on the manuscript. All analyses were run at the High-performance Computing Center (Vital-IT) of the Swiss Institute of Bioinformatics. D.S. received funding from the Swedish Research Council (2015-04748) and from a Sinergia grant (CRSII3-147630; Swiss National Science Foundation to N.S.). Both M.M.P. and T.B.Q. are supported by the São Paulo Research Foundation (FAPESP grants nos. 2013/22016-6; 2012/04072-3). N.S. received funding from the Swiss National Science Foundation (CR32I3-143768).

REFERENCES

Alroy, J. 1996. Constant extinction, constrained diversification, and uncoordinated stasis in North American mammals. *Palaeogeogr. Palaeoclimatol. Palaeoecol.*, **127**: 285–311.

Alroy, J. 2008. Dynamics of origination and extinction in the marine fossil record. *Proc. Natl. Acad. Sci. USA*, **105**: 11536–11542.

Barraclough, T.G. 2015. How do species interactions affect evolutionary dynamics across whole communities? *Annu. Rev. Ecol. Evol. Syst.*, **46**: 25–48.

Benton, M.J. 1996. On the nonprevalence of competitive replacement in the evolution of tetrapods. In *Evolutionary Paleobiology* (D. Jablonski, D.H. Erwin and J.H. Lipps, eds.), pp. 185–210. Chicago, IL: University of Chicago Press.

Benton, M.J. 2009. The red queen and the court jester: species diversity and the role of biotic and abiotic factors through time. *Science*, **323**: 728–732.

Bokma, F. 2009. Problems detecting density-dependent diversification on phylogenies. *Proc. R. Soc. Lond. B: Biol. Sci.*, **276**: 993–994.

Carvalho, C.M., Polson, N.G. and Scott, J.G. 2010. The horseshoe estimator for sparse signals. *Biometrika*, **97**: 465–480.

Castiglione, S., Mondanaro, A., Carotenuto, F., Passaro, F., Fortelius, M. and Raia, P. 2017. The many shapes of diversity: ecological and evolutionary determinants of biodiversity through time. *Evol. Ecol. Res.*, **18**: 25–39.

Cornell, H.V. 2013. Is regional species diversity bounded or unbounded? *Biol. Rev.*, **88**: 140–165.

Damien, P., Wakefield, J. and Walker, S. 1999. Gibbs sampling for Bayesian non-conjugate and hierarchical models by using auxiliary variables. *J. R. Stat. Soc. B: Stat. Meth.*, **61**: 331–344.

Darwin, C. 1859. *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. London: John Murray.

Davis, M.P., Midford, P.E. and Maddison, W. 2013. Exploring power and parameter estimation of the BiSSE method for analyzing species diversification. *BMC Evol. Biol.*, **13**: 38.

Diamond, J.M. 1987. Competition among different taxa. *Nature*, **326**: 241.

Englund, G., Johansson, F. and Olsson, T.I. 1992. Asymmetric competition between distant taxa: poecilid fishes and water striders. *Oecologia*, **92**: 498–502.

Etienne, R.S., Haegeman, B., Stadler, T., Aze, T., Pearson, P.N., Purvis, A. *et al.* 2012. Diversity-dependence brings molecular phylogenies closer to agreement with the fossil record. *Proc. R. Soc. Lond. B: Biol. Sci.*, **279**: 1300–1309.

Ezard, T.H.G., Aze, T., Pearson, P.N. and Purvis, A. 2011. Interplay between changing climate and species' ecology drives macroevolutionary dynamics. *Science*, **332**: 349–351.

Ezard, T.H.G., Quental, T.B. and Benton, M.J. 2016. The challenges to inferring the regulators of biodiversity in deep time. *Phil. Trans. R. Soc. Lond. B: Biol. Sci.*, **371**: 20150216.

Gavrilets, S. and Losos, J.B. 2009. Adaptive radiation: contrasting theory with data. *Science*, **323**: 732–737.

Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A. and Rubin, D.B. 2013. *Bayesian Data Analysis*, 3rd edn. Boca Raton, FL: CRC Press.

Gould, S.J., Raup, D.M., Sepkoski, J.J., Schopf, T.J.M. and Simberloff, D.S. 1977. The shape of evolution: a comparison of real and random clades. *Paleobiology*, **3**: 23–40.

Hastings, W.K. 1970. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, **57**: 97–109.

Jablonski, D. 2008. Biotic interactions and macroevolution: extensions and mismatches across scales and levels. *Evolution*, **62**: 715–739.

Keiding, N. 1975. Maximum likelihood estimation in the birth–death process. Ann. *Stat.*, **3**: 363–372.

Knoll, A.H. 1986. Patterns of change in plant communities through geological time. In *Community Ecology* (J. Diamond and T.J. Case, eds.), pp. 126–141. New York: Harper & Row.

Kuo, L. and Mallick, B. 1998. Variable selection for regression models. *Sankhya Ser. B*, **60**: 65–81.

Levinton, J.S. 1979. A theory of diversity equilibrium and morphological evolution. *Science*, **204**: 335–336.

Lidgard, S., Mckinney, F.K. and Taylor, P.D. 1993. Competition, clade replacement, and a history of cyclostome and cheilostome bryozoan diversity. *Paleobiology*, **19**: 352–371.

Liow, L.H. and Finarelli, J.A. 2014. A dynamic global equilibrium in carnivoran diversification over 20 million years. *Proc. R. Soc. Lond. B: Biol. Sci.*, **281**: 20132312.

Liow, L.H., Reitan, T. and Harnik, P.G. 2015. Ecological interactions on macroevolutionary time scales: clams and brachiopods are more than ships that pass in the night. *Ecol. Lett.*, **18**:1030–1039.

Maddison, W.P. and FitzJohn, R.G. 2015. The unsolved challenge to phylogenetic correlation tests for categorical characters. *Syst. Biol.*, **64**: 127–136.

Marshall, C.R. and Quental, T.B. 2016. The uncertain role of diversity dependence in species diversification and the need to incorporate time-varying carrying capacities. Phil. *Trans. R. Soc. Lond. B: Biol. Sci.*, **371**: 20150217.

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.W. and Teller, E. 1953. Equations of state calculations by fast computing machines. *J. Chem. Phys.*, **21**: 1087–1091.

Moen, D. and Morlon, H. 2014. Why does diversification slow down? *Trends Ecol. Evol.*, **29**: 190–197.

O'Hara, R.B. and Sillanpää, M.J.. 2009. A review of Bayesian variable selection methods: what, how and which. *Bayesian Anal.*, **4**: 85–117.

Phillimore, A.B. and Price, T.D. 2008. Density-dependent cladogenesis in birds. *PLoS Biol.*, **6**: e71.

Pires, M.M., Silvestro, D. and Quental, T.B. 2015. Continental faunal exchange and the asymmetrical radiation of carnivores. *Proc. R. Soc. Lond. B: Biol. Sci.*, **282**: 20151952.

Quental, T.B. and Marshall, C.R. 2013. How the red queen drives terrestrial mammals to extinction. *Science*, **341**: 290–292.

Rabosky, D.L. 2013. Diversity-dependence, ecological speciation, and the role of competition in macroevolution. *Annu. Rev. Ecol. Evol. Syst.*, **44**: 481–502.

Rabosky, D.L. and Goldberg, E.E. 2015. Model inadequacy and mistaken inferences of trait-dependent speciation. *Syst. Biol.*, **64**: 340–355.

Rambaut, A., Suchard, M.A., Xie, D. and Drummond, A.J. 2014. *Tracer* v.1.6 [available from: http://beast.bio.ed.ac.uk/tracer].

Raup, D.M. and Gould, S.J. 1974. Stochastic simulation and evolution of morphology – towards a nomothetic paleontology. *Syst. Zool.*, **23**: 305–322.

Ronquist, F., van der Mark, P. and Huelsenbeck, J.P. 2007. Bayesian phylogenetic analysis using MrBayes. In *The Phylogenetic Handbook* (P. Lemey, M. Salemi and A.-M. Vandamme, eds.), pp. 210–266. Cambridge: Cambridge University Press.

Rosenzweig, M.L. 1975. On continental steady states of species diversity. In *The Ecology of Species Communities* (M.L. Cody and J.M. Diamond, eds.), pp. 121–140. Cambridge, MA: Belknap Press.

Rosenzweig, M. and McCord, R. 1991. Incumbent replacement: evidence for long-term evolutionary progress. *Paleobiology*, **17**: 202–213.

Schluter, D. 1986. Character displacement between distantly related taxa? Finches and bees in the Galapagos. *Am. Nat.*, **127**: 95–102.

Schneider, H., Schuettpelz, E., Pryer, K.M., Cranfill, R., Magallon, S.and Lupia, R. 2004. Ferns diversified in the shadow of angiosperms. *Nature*, **428**: 553–557.

Scott, J.G. 2010. Parameter expansion in local-shrinkage models. *arXiv*, 1010.5265.

Scott, J.G. 2011. Bayesian estimation of intensity surfaces on the sphere via needlet shrinkage and selection. *Bayesian Anal.*, **6**: 307–327.

Sepkoski, J.J., Jr. 1978. A kinetic model of Phanerozoic taxonomic diversity: I. Analysis of marine orders. *Paleobiology*, **4**: 223–251.

Sepkoski, J.J., Jr. 1996a. Competition in macroevolution: the double wedge revisited. In *Evolutionary Paleobiology* (D. Jablonski, D.H. Erwin and J.H. Lipps, eds.), pp. 211–255. Chicago, IL: University of Chicago Press.

Sepkoski, J.J., Jr. 1996b. Patterns of Phanerozoic extinction: a perspective from global databases. In Global Events and Event Stratigraphy (O.H. Walliser, ed.), pp. 35–51. Berlin: Springer.

Sepkoski, J.J., Jr., McKinney, F.K. and Lidgard, S. 2000. Competitive displacement among post-Paleozoic cyclostome and cheilostome bryozoans. *Paleobiology*, **26**: 7–18.

Silvestro, D., Salamin, N. and Schnitzler, J. 2014a. PyRate: a new program to estimate speciation and extinction rates from incomplete fossil record. *Methods Ecol. Evol.*, **5**: 1126–1131.

Silvestro, D., Schnitzler, J., Liow, L.H., Antonelli, A. and Salamin, N. 2014b. Bayesian estimation of speciation and extinction from incomplete fossil occurrence data. *Syst. Biol.*, **63**: 349–367.

Silvestro, D., Antonelli, A., Salamin, N. and Quental, T.B. 2015a. The role of clade competition in the diversification of North American canids. *Proc. Natl. Acad. Sci. USA*, **112**: 8684–8689.

Silvestro, D., Cascales-Miñana, B., Bacon, C.D. and Antonelli, A. 2015b. Revisiting the origin and diversification of vascular plants through a comprehensive Bayesian analysis of the fossil record. *New Phytol.*, **207**: 425–436.

Van Valen, L. and Sloan, R.E. 1966. The extinction of the multituberculates. *Syst. Zool.*, **15**: 261–278.

Van Valkenburgh, B. 1999. Major patterns in the history of carnivorous mammals. *Annu. Rev. Earth Planet. Sci.*, **27**: 463–493.

Voje, K.L., Nolen, O.H., Liow, L.H. and Stenseth, N.C. 2015. The role of biotic forces in driving macroevolution: beyond the red queen. *Proc. R. Soc. Lond. B: Biol. Sci.*, **282**: 20150186.

Walker, T.D. and Valentine, J.W. 1984. Equilibrium models of evolutionary species diversity and the number of empty niches. *Am. Nat.*, **124**: 887–899.

Yoder, J.B., Clancey, E., Des Roches, S., Eastman, J.M., Gentry, L., Godsoe, W. *et al.* 2010. Ecological opportunity and the origin of adaptive radiations. *J. Evol. Biol.*, **23**: 1581–1596.